Springer Proceedings in Mathematics & Statistics

Fred Espen Benth
Giulia Di Nunno *Editors*

Stochastics of Environmental and Financial Economics

Centre of Advanced Study, Oslo, Norway, 2014–2015

Springer Open

Springer Proceedings in Mathematics & Statistics

Volume 138

Springer Proceedings in Mathematics & Statistics

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including operation research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

More information about this series at http://www.springer.com/series/10533

Fred Espen Benth · Giulia Di Nunno
Editors

Stochastics of Environmental and Financial Economics

Centre of Advanced Study, Oslo, Norway, 2014–2015

Springer Open

Editors
Fred Espen Benth
Department of Mathematics
University of Oslo
Oslo
Norway

Giulia Di Nunno
Department of Mathematics
University of Oslo
Oslo
Norway

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-319-23424-3 ISBN 978-3-319-23425-0 (eBook)
DOI 10.1007/978-3-319-23425-0

Library of Congress Control Number: 2015950032

Mathematics Subject Classification: 93E20, 91G80, 91G10, 91G20, 60H30, 60G07, 35R60, 49L25, 91B76

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

Norway is a country rich on natural resources. Wind, rain and snow provide us with a huge resource for clean energy production, while oil and gas have contributed significantly, since the early 1970s, to the country's economic wealth. Nowadays the income from oil and gas exploitation is invested in the world's financial markets to ensure the welfare of future generations. With the rising global concerns about climate, using renewable resources for power generation has become more and more important. Bad management of these resources will be a waste that is a negligence to avoid given the right tools.

This formed the background and motivation for the research group Stochastics for Environmental and Financial Economics (SEFE) at the Centre of Advanced Studies (CAS) in Oslo, Norway. During the academic year 2014–2015, SEFE hosted a number of distinguished professors from universities in Belgium, France, Germany, Italy, Spain, UK and Norway. The scientific purpose of the SEFE centre was to focus on the analysis and management of risk in the environmental and financial economics. New mathematical models for describing the uncertain dynamics in time and space of weather factors like wind and temperature were studied, along with sophisticated theories for risk management in energy, commodity and more conventional financial markets.

In September 2014 the research group organized a major international conference on the topics of SEFE, with more than 60 participants and a programme running over five days. The present volume reflects some of the scientific developments achieved by CAS fellows and invited speakers at this conference. All the 14 chapters are stand-alone, peer-reviewed research papers. The volume is divided into two parts; the first part consists of papers devoted to fundamental aspects of stochastic analysis, whereas in the second part the focus is on particular applications to environmental and financial economics.

We thank CAS for its generous support and hospitality during the academic year we organized our SEFE research group. We enjoyed the excellent infrastructure CAS offered for doing research.

Oslo, Norway
June 2015

Fred Espen Benth
Giulia Di Nunno

Contents

Part I
Foundations

Some Recent Developments in Ambit Stochastics

Ole E. Barndorff-Nielsen, Emil Hedevang, Jürgen Schmiegel and Benedykt Szozda

Abstract Some of the recent developments in the rapidly expanding field of Ambit Stochastics are here reviewed. After a brief recall of the framework of Ambit Stochastics, two topics are considered: (i) Methods of modelling and inference for volatility/intermittency processes and fields; (ii) Universal laws in turbulence and finance in relation to temporal processes. This review complements two other recent expositions.

Keywords Ambit stochastics · Stochastic volatility/intermittency · Universality · Finance · Turbulence · Extended subordination · Metatimes · Time-change

1 Introduction

Ambit Stochastics is a general framework for the modelling and study of dynamic processes in space-time. The present paper outlines some of the recent developments in the area, with particular reference to finance and the statistical theory of turbulence. Two recent papers [8, 36] provide surveys that focus on other sides of Ambit Stochastics.

O.E. Barndorff-Nielsen (✉) · E. Hedevang · B. Szozda
Department of Mathematics, Aarhus University, Ny Munkegade 118,
8000 Aarhus C, Denmark
e-mail: oebn@math.au.dk

E. Hedevang
e-mail: hedevang@math.au.dk

B. Szozda
e-mail: szozda@math.au.dk

J. Schmiegel
Department of Engineering, Aarhus University, Inge Lehmanns Gade 10,
8000 Aarhus C, Denmark
e-mail: schmiegl@eng.au.dk

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_1

A key characteristic of the Ambit Stochastics framework, which distinguishes this from other approaches, is that beyond the most basic kind of random input it also specifically incorporates additional, often drastically changing, inputs referred to as volatility or intermittency.

Such "additional" random fluctuations generally vary, in time and/or in space, in regard to *intensity* (activity rate and duration) and *amplitude*. Typically the volatility/intermittency may be further classified into continuous and discrete (i.e. jumps) elements, and long and short term effects. In turbulence the key concept of energy dissipation is subsumed under that of volatility/intermittency.

The concept of (stochastic) *volatility/intermittency* is of major importance in many fields of science. Thus volatility/intermittency has a central role in *mathematical finance and financial econometrics*, in *turbulence*, in *rain and cloud studies* and other aspects of environmental science, in relation to *nanoscale emitters*, *magnetohydrodynamics*, and to *liquid mixtures of chemicals*, and last but not least in the *physics of fusion plasmas*.

As described here, volatility/intermittency is a *relative* concept, and its meaning depends on the particular setting under investigation. Once that meaning is clarified the question is how to assess the volatility/intermittency empirically and then to describe it in stochastic terms, for incorporation in a suitable probabilistic model. Important issues concern the modelling of propagating stochastic volatility/intermittency fields and the question of predictability of volatility/intermittency.

Section 2 briefly recalls some main aspects of Ambit Stochastics that are of relevance for the dicussions in the subsequent sections, and Sect. 3 illustrates some of the concepts involved by two examples. The modelling of volatility/intermittency and energy dissipation is a main theme in Ambit Stochastics and several approaches to this are discussed in Sect. 4. A leading principle in the development of Ambit Stochastics has been to take the cue from recognised stylised features—or universality traits—in various scientific areas, particularly turbulence, as the basis for model building; and in turn to seek new such traits using the models as tools. We discuss certain universal features observed in finance and turbulence and indicate ways to reproduce them in Sect. 5. Section 6 concludes and provides an outlook.

2 Ambit Stochastics

2.1 General Framework

In terms of mathematical formulae, in its original form [17] (cf. also [16]) an ambit field is specified by

$$Y(x,t) = \mu + \int_{A(x,t)} g(x,\xi,t,s)\sigma(\xi,s)\, L(d\xi\, ds) + Q(x,t) \tag{1}$$

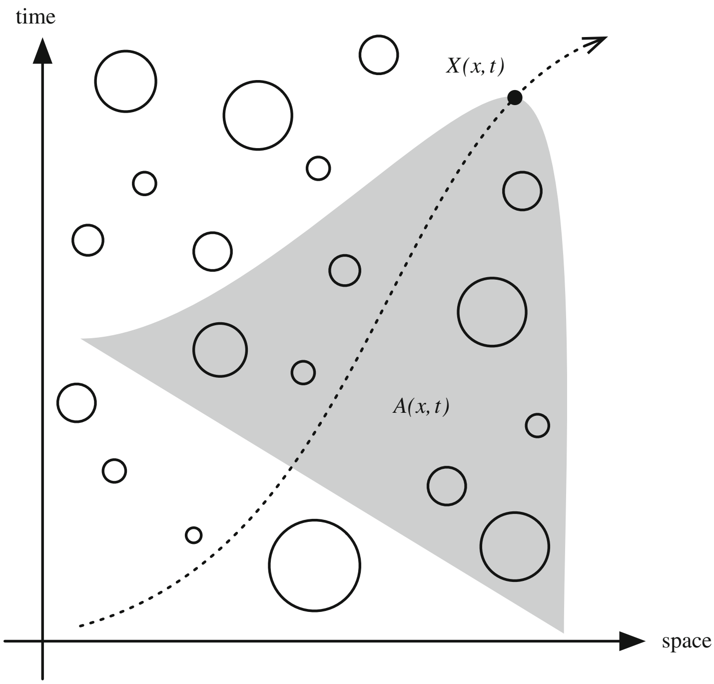

Fig. 1 A spatio-temporal ambit field. The value $Y(x,t)$ of the field at the point marked by the *black dot* is defined through an integral over the corresponding ambit set $A(x,t)$ marked by the shaded region. The *circles* of varying sizes indicate the stochastic volatility/intermittency. By considering the field along the *dotted path* in space-time an ambit process is obtained

where

$$Q(x,t) = \int_{D(x,t)} q(x,\xi,t,s)\chi(\xi,s)\, d\xi\, ds. \tag{2}$$

Here t denotes time while x gives the position in d-dimensional Euclidean space. Further, $A(x,t)$ and $D(x,t)$ are subsets of $\mathbb{R}^d \times \mathbb{R}$ and are termed *ambit sets*, g and q are deterministic weight functions, and L denotes a Lévy basis (i.e. an independently scattered and infinitely divisible random measure). Further, σ and χ are stochastic fields representing aspects of the volatility/intermittency. In Ambit Stochastics the models of the volatility/intermittency fields σ and χ are usually themselves specified as ambit fields. We shall refer to σ as the *amplitude volatility* component. Figure 1 shows a sketch of the concepts.

The development of Y along a curve in space-time is termed an ambit process. As will be exemplified below, ambit processes are not in general semimartingales, even in the purely temporal case, i.e. where there is no spatial component x.

In a recent extension the structure (1) is generalised to

$$Y(x,t) = \mu + \int_{A(x,t)} g(x,\xi,t,s)\sigma(\xi,s)\, L_T(d\xi\, ds) + Q(x,t) \tag{3}$$

where Q is like (2) or the exponential thereof, and where T is a *metatime* expressing a further volatility/intermittency trait. The relatively new concept of metatime is instrumental in generalising subordination of stochastic processes by time change (as discussed for instance in [22]) to subordination of random measures by random measures. We return to this concept and its applications in the next section and refer also to the discussion given in [8].

Note however that in addition to modelling volatility/intermittency through the components σ, χ and T, in some cases this may be supplemented by probability mixing or Lévy mixing as discussed in [12].

It might be thought that ambit sets have no role in purely temporal modelling. However, examples of their use in such contexts will be discussed in Sect. 3.

In many cases it is possible to choose specifications of the volatility/intermittency elements σ, χ and T such that these are infinitely divisible or even selfdecomposable, making the models especially tractable analytically. We recall that the importance of the concept of selfdecomposability rests primarily on the possibility to represent selfdecomposable variates as stochastic integrals with respect to Lévy processes, see [32].

So far, the main applications of ambit stochastics has been to turbulence and, to a lesser degree, to financial econometrics and to bioimaging. An important potential area of applications is to particle transport in fluids.

2.2 Existence of Ambit Fields

The paper [25] develops a general theory for integrals

$$X(x,t) = \int_{\mathbb{R}^d \times \mathbb{R}} h(x,y,t,s)\, M(dx\, dx)$$

where h is a predictable stochastic function and M is a dispersive signed random measure. Central to this is that the authors establish a notion of characteristic triplet of M, extending that known in the purely temporal case. A major problem solved in that regard has been to merge the time and space aspects in a general and tractable fashion. Armed with that notion they determine the conditions for existence of the integral, analogous to those in [37] but considerably more complicated to derive and apply. An important property here is that now predictable integrands are allowed (in the purely temporal case this was done in [23]). Applications of the theory to Ambit Stochastics generally, and in particular to superposition of stochastic volatility models, is discussed.

Below we briefly discuss how the metatime change is incorporated in the framework of [25]. Suppose that $L = \{L(A) \mid A \in \mathscr{B}_b(\mathbb{R}^{d+1}\}$ is a real-valued, homogeneous Lévy basis with associated infinitely divisible law $\mu \in ID(\mathbb{R}^{d+1})$, that is $L([0,1]^{d+1})$ is equal in law to μ. Let (γ, Σ, ν) be the characteristic triplet of μ. Thus $\gamma \in \mathbb{R}$, $\Sigma \geq 0$ and ν is a Lévy measure on $\mathbb{R}$.

Suppose that $\mathbf{T} = \{\mathbf{T}(A) \mid A \in \mathscr{B}(\mathbb{R}^{d+1}\}$ is a random meta-time associated with a homogeneous, real-valued, non-negative Lévy basis $T = \{T(A) \mid A \in \mathscr{B}(\mathbb{R}^{d+1})\}$. That is the sets $\mathbf{T}(A)$ and $\mathbf{T}(B)$ are disjoint whenever $A, B \in \mathscr{B}(\mathbb{R}^{d+1})$ are disjoint, $\mathbf{T}(\cup_{n=0}^{\infty} A_n) = \cup_{n=0}^{\infty} \mathbf{T}(A_n)$ whenever $A_n, \cup_{n=0}^{\infty} A_n \in \mathscr{B}(\mathbb{R}^{d+1})$ and $T(A) = \mathrm{Leb}_{d+1}(\mathbf{T}(A))$ for all $A \in \mathscr{B}(\mathbb{R}^{d+1})$. Here and in what follows, Leb_k denotes the Lebesgue measure on $\mathbb{R}^k$. For the details on construction of random meta-times cf. [11]. Suppose also that $\lambda \in ID(\mathbb{R})$ is the law associated to T and that $\lambda \sim ID(\beta, 0, \rho)$. Thus $\beta \geq 0$ and ρ is a Lévy measure such that $\rho(\mathbb{R}_-) = 0$ and $\int_{\mathbb{R}} (1 \wedge x)\, \rho(dx) < \infty$.

Now, by [11, Theorem 5.1] we have that $L_T = \{L(\mathbf{T}(A)) \mid A \in \mathscr{B}(\mathbb{R}^{d+1})\}$ is a homogeneous Lévy basis associated to $\mu^{\#}$ with $\mu^{\#} \sim ID(\gamma^{\#}, \Sigma^{\#}, \nu^{\#})$ and characteristics given by

$$\begin{aligned} \gamma^{\#} &= \beta\gamma + \int_0^{\infty} \int_{|x|\leq 1} x \mu^s(dx)\rho(ds) \\ \Sigma^{\#} &= \beta\Sigma \\ \nu^{\#}(B) &= \beta\nu(B) + \int_0^{\infty} \mu^s(B)\rho(ds), \qquad B \in \mathscr{B}(\mathbb{R}^{d+1} \setminus \{0\}), \end{aligned}$$

where μ^s is given by $\widehat{\mu^s} = \widehat{\mu}^s$ for any $s \geq 0$.

Finally, suppose that $\sigma(x,t)$ is predictable and that L_T has no fixed times of discontinuity (see [25]). By rewriting the stochastic integral in the right-hand side of (3) as

$$X(x,t) = \int_{\mathbb{R}^{d+1}} H(x,\xi,t,s)\, L_T(d\xi ds),$$

with $H(x,\xi,t,s) = \mathbf{1}_{A(x,t)}(\xi,s) g(x,\xi,t,s)\sigma(\xi,s)$ we can use [25, Theorem 4.1]. Observe that the assumption that σ is predictable is enough as both $A(x,t)$ and $g(x,\xi,t,s)$ are deterministic. This gives us that X is well defined for all (x,t) if the following hold almost surely for all $(x,t) \in \mathbb{R}^{d+1}$:

$$\int_{\mathbb{R}^{d+1}} \left| H(x,\xi,t,s)\gamma^{\#} + \int_{\mathbb{R}} [\tau(H(x,\xi,t,s)y) - H(x,\xi,t,s)\tau(y)]\nu^{\#}(dy) \right| d\xi ds < \infty \tag{4}$$

$$\int_{\mathbb{R}^{d+1}} H^2(x,\xi,t,s)\Sigma^{\#}\, d\xi ds < \infty \tag{5}$$

$$\int_{\mathbb{R}^{d+1}} \int_{\mathbb{R}} (1 \wedge (H(x,\xi,t,s)y)^2\, \nu^{\#}(dy) d\xi ds < \infty. \tag{6}$$

3 Illustrative Examples

We can briefly indicate the character of some of the points on Ambit Stochastics made above by considering the following simple model classes.

3.1 BSS and LSS Processes

Stationary processes of the form

$$Y(t) = \int_{-\infty}^{t} g(t-s)\sigma(s)B_T(ds) + \int_{-\infty}^{t} q(t-s)\sigma(s)^2\, ds. \tag{7}$$

are termed Brownian semistationary processes—or BSS for short. Here the setting is purely temporal and B_T is the time change of Brownian motion B by a chronometer T (that is, an increasing, càdlàg and stochastically continuous process ranging from $-\infty$ to ∞), and the volatility/intermittency process σ is assumed stationary. The components σ and T represent respectively the amplitude and the intensity of the volatility/intermittency. If T has stationary increments then the process Y is stationary. The process (7) can be seen as a stationary analogue of the BNS model introduced by Barndorff-Nielsen and Shephard [14].

Note that in case T increases by jumps only, the infinitesimal of the process B_T cannot be reexpressed in the form $\chi(s)B(ds)$, as would be the case if T was of type $T_t = \int_0^t \psi(u)\, du$ with $\chi = \sqrt{\psi}$.

Further, for the exemplification we take g to be of the gamma type

$$g(s) = \frac{\lambda^{\nu}}{\Gamma(\nu)} s^{\nu-1} e^{-\lambda s} 1_{(0,\infty)}(s). \tag{8}$$

Subject to a weak (analogous to (4)) condition on σ, the stochastic integral in (7) will exist if and only if $\nu > 1/2$ and then Y constitutes a stationary process in time. Moreover, Y is a semimartingale if and only if ν does not lie in one of the intervals $(1/2, 1)$ and $(1, 3/2]$. Note also that the sample path behaviour is drastically different between the two intervals, since, as $t \to 0$, $g(t)$ tends to ∞ when $\nu \in (1/2, 1)$ and to 0 when $\nu \in (1, 3/2]$. Further, the sample paths are purely discontinuous if $\nu \in (1/2, 1)$ but purely continuous (of Hölder index $H = \nu - 1/2$) when $\nu \in (1, 3/2)$.

The cases where $\nu \in (1/2, 1)$ have a particular bearing in the context of turbulence, the value $\nu = 5/6$ having a special role in relation to the Kolmogorov-Obukhov theory of statistical turbulence, cf. [3, 33].

The class of processes obtained by substituting the Brownian motion in (7) by a Lévy process is referred to as the class of Lévy semistationary processes—or LSS processes for short. Such processes are discussed in [8, 24, Sect. 3.7] and references therein.

3.2 Trawl Processes

The simplest non-trivial kind of ambit field is perhaps the *trawl process*, introduced in [2]. In a trawl process, the kernel function and the volatility field are constant and equal to 1, and so the process is given entirely by the ambit set and the Lévy basis. Specifically, suppose that L is a homogeneous Lévy basis on $\mathbb{R}^d \times \mathbb{R}$ and that $A \subseteq \mathbb{R}^d \times \mathbb{R}$ is a Borel subset with finite Lebesgue measure, then we obtain a trawl process Y by letting $A(t) = A + (0, t)$ and

$$Y(t) = \int_{A(t)} L(d\xi\, ds) = \int 1_A(\xi, t-s)\, L(d\xi\, ds) = L(A(t)) \tag{9}$$

The process is by construction stationary. Depending on the purpose of the modelling, the time component of the ambit set A may or may not be supported on the negative real axis. When the time component of A is supported on the negative real axis, we obtain a causal model. Despite their apparent simplicity, trawl processes possess enough flexibility to be of use. If L' denotes the seed[1] of L, then the cumulant function (i.e. the distinguished logarithm of the characteristic function) of Y is given by

$$C\{\zeta \ddagger Y(t)\} = |A| C\{\zeta \ddagger L'\}. \tag{10}$$

Here and later, $|A|$ denotes the Lebesgue measure of the set A. For the mean, variance, autocovariance and autocorrelation it follows that

$$\begin{aligned} \mathbb{E}[Y(t)] &= |A|\, \mathbb{E}[L'], \\ \operatorname{var}(Y(t)) &= |A| \operatorname{var}(L'), \\ r(t) &:= \operatorname{cov}(Y(t), Y(0)) = |A \cap A(t)| \operatorname{var}(L'), \\ \rho(t) &:= \frac{\operatorname{cov}(Y(t), Y(0))}{\operatorname{var}(Y(0))} = \frac{|A \cap A(t)|}{|A|}. \end{aligned} \tag{11}$$

From this we conclude the following. The one-dimensional marginal distribution is determined entirely in terms of the size (not shape) of the ambit set and the distribution of the Lévy seed; given any infinitely divisible distribution there exists trawl processes having this distribution as the one-dimensional marginal; and the autocorrelation is determined entirely by the size of the overlap of the ambit sets, that is, by the shape of the ambit set A. Thus we can specify the autocorrelation and marginal distribution independently of each other. It is, for example, easy to construct a trawl process with the same autocorrelation as the OU process, see [2, 8] for more results and details. By using integer-valued Lévy bases, integer-valued

[1] The Lévy seed $L'(x)$ at x of a Lévy basis L with control measure ν is a random variable with the property that $C\{\zeta \ddagger L(A)\} = \int_A C\{\zeta \ddagger L'(x)\}\, \nu(dx)$. For a homogeneous Lévy basis, the distribution of the seed is independent of x.

trawl processes are obtained. These processes are studied in detail in [5] and applied to high frequency stock market data.

We remark, that $Y(x,t) = L(A + (x,t))$ is an immediate generalisation of trawl processes to trawl fields. It has the same simple properties as the trawl process.

Trawl processes can be used to directly model an object of interest, for example, the exponential of the trawl process has been used to model the energy dissipation, see the next section, or they can be used as a component in a composite model, for example to model the volatility/intermittency in a Brownian semistationary process.

4 Modelling of Volatility/Intermittency/Energy Dissipation

A very general approach to specifying volatility/intermittency fields for inclusion in an ambit field, as in (1), is to take $\tau = \sigma^2$ as being given by a Lévy-driven Volterra field, either directly as

$$\tau(x,t) = \int_{\mathbb{R}^2 \times \mathbb{R}} f(x,\xi,t,s)\, L(d\xi, ds) \tag{12}$$

with f positive and L a Lévy basis (different from L in (1), or in exponentiated form

$$\tau(x,t) = \exp\left(\int_{\mathbb{R}^d \times \mathbb{R}} f(x,\xi,t,s)\, L(d\xi, ds) \right). \tag{13}$$

When the goal is to have stationary volatility/intermittency fields, such as in modelling homogeneous turbulence, that can be achieved by choosing L to be homogeneous and f of translation type. However, the potential in the specifications (12) and (13) is much wider, giving ample scope for modelling inhomogeneous fields, which are by far the most common, particularly in turbulence studies. Inhomogeneity can be expressed both by not having f of translation type and by taking the Lévy basis L inhomogeneous.

In the following we discuss two aspects of the volatility/intermittency modelling issue. Trawl processes have proved to be a useful tool for the modelling of volatility/intermittency and in particular for the modelling of the energy dissipation, as outlined in Sect. 4.1. Section 4.2 reports on a recent paper on relative volatility/intermittency. In Sect. 4.3 we discuss the applicability of selfdecomposability to the construction of volatility/intermittency fields.

4.1 The Energy Dissipation

In [31] it has been shown that exponentials of trawl processes are able to reproduce the main stylized features of the (surrogate) energy dissipation observed for a wide range of datasets. Those stylized features include the one-dimensional marginal distributions and the scaling and self-scaling of the correlators.

The correlator of order (p, q) is defined by

$$c_{p,q}(s) = \frac{\mathbb{E}[\varepsilon(t)^p \varepsilon(t+s)^q]}{\mathbb{E}[\varepsilon(t)^p]\mathbb{E}[\varepsilon(t+s)^q]}. \tag{14}$$

The correlator is a natural analogue to the autocorrelation when one considers a purely positive process. In turbulence it is known (see the reference cited in [31]) that the correlator of the surrogate energy dissipation displays a scaling behaviour for a certain range of lags,

$$c_{p,q}(s) \propto s^{-\tau(p,q)}, \quad T_{\text{small}} \ll s \ll T_{\text{large}}, \tag{15}$$

where $\tau(p, q)$ is the scaling exponent. The exponent $\tau(1, 1)$ is the so-called intermittency exponent. Typical values are in the range 0.1 to 0.2. The intermittency exponent quantifies the deviation from Kolmogorov's 1941 theory and emphasizes the role of intermittency (i.e. volatility) in turbulence. In some cases, however, the scaling range of the correlators can be quite small and therefore it can be difficult to determine the value of the scaling exponents, especially when p and q are large. Therefore one also considers the correlator of one order as a function of a correlator of another order. In this case, self-scaling is observed, i.e., the one correlator is proportional to a power of the other correlator,

$$c_{p,q}(s) \propto c_{p',q'}(s)^{\tau(p,q;p',q')}, \tag{16}$$

where $\tau(p, q; p', q')$ is the self-scaling exponent. The self-scaling exponents have turned out to be much easier to determine from data than the scaling exponents, and like the scaling exponents, the self-scaling exponents have proved to be key fingerprints of turbulence. They are essentially universal in that they vary very little from one dataset to another, covering a large range of the so-called Reynolds numbers, a dimensionless quantity describing the character of the flow.

In [31] the surrogate energy dissipation ε is, more specifically, modelled as

$$\varepsilon(t) = \exp(L(A(t))), \tag{17}$$

where L is a homogeneous Lévy basis on $\mathbb{R} \times \mathbb{R}$ and $A(t) = A + (0, t)$ for a bounded set $A \subset \mathbb{R} \times \mathbb{R}$. The ambit set A is given as

$$A = \{(x, t) \in \mathbb{R} \times \mathbb{R} \mid 0 \le t \le T_{\text{large}}, -f(t) \le x \le f(t)\}, \tag{18}$$

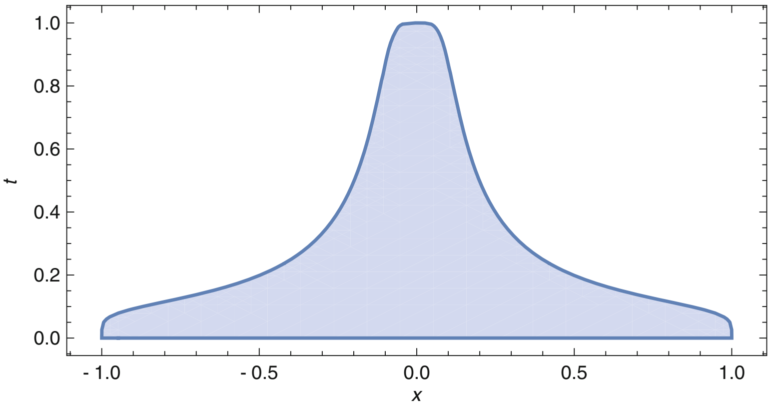

Fig. 2 The shaded region marks the ambit set A from (18) defined by (19) where the parameters are chosen to be $T_{\text{large}} = 1$, $T_{\text{small}} = 0.1$ and $\theta = 5$

where $T_{\text{large}} > 0$. For $T_{\text{large}} > T_{\text{small}} > 0$ and $\theta > 0$, the function f is defined as

$$f(t) = \left(\frac{1 - (t/T_{\text{large}})^{\theta}}{1 + (t/T_{\text{small}})^{\theta}} \right)^{1/\theta}, \quad 0 \le t \le T_{\text{large}}. \tag{19}$$

The shape of the ambit set is chosen so that the scaling behaviour (15) of the correlators is reproduced. The exact values of the scaling exponents are determined from the distribution of the Lévy seed of the Lévy basis. The two parameters T_{small} and T_{large} determine the size of the small and large scales of turbulence: in between we have the inertial range. The final parameter θ is a tuning parameter which accounts for the lack of perfect scaling and essentially just allows for a better fit. (Perfect scaling is obtained in the limit $\theta \to \infty$). See Fig. 2 for an example. Furthermore, self-scaling exponents are predicted from the shape, not location and scale, of the one-point distribution of the energy dissipation alone.

To determine a proper distribution of the Lévy seed of L, it is in [31] shown that the one-dimensional marginal of the logarithm of the energy dissipation is well described by a normal inverse Gaussian distribution, i.e. $\log \varepsilon(t) \sim \text{NIG}(\alpha, \beta, \mu, \delta)$, where the shape parameters α and β are the same for all datasets (independent of the Reynolds number). Thus the shape of the distribution of the energy dissipation is a newly discovered universal feature of turbulence. Thus we see that L should be a normal inverse Gaussian Lévy basis whose parameters are given by the observed distribution of $\log \varepsilon(t)$. This completely specifies the parameters of (17).

4.2 *Realised Relative Volatility/Intermittency/Energy Dissipation*

By its very nature, volatility/intermittency is a relative concept, delineating variation that is relative to a conceived, simpler model. But also in a model for volatility/intermittency in itself it is relevant to have the relative character in mind, as will be further discussed below. We refer to this latter aspect as relative volatility/intermittency and will consider assessment of that by *realised relative volatility/intermittency* which is defined in terms of quadratic variation. The ultimate purpose of the concept of relative volatility/intermittency is to assess the volatility/intermittency or energy dissipation in arbitrary subregions of a region C of space-time relative to the total volatility/intermittency/energy dissipation in C. In the purely temporal setting the *realised relative volatility/intermittency* is defined by

$$[Y_\delta]_t/[Y_\delta]_T \tag{20}$$

where $[Y_\delta]_t$ denotes the realised quadratic variation of the process Y observed with lag δ over a time interval $[0, t]$. We refer to this quantity as RRQV (for realised relative quadratic volatility).

As mentioned in Example 1, in case g is the gamma kernel (8) with $\nu \in (1/2, 1) \cup (1, 3/2]$ then the BSS process (7) is not a semimartingale. In particular, if $\nu \in (1/2, 1)$—the case of most interest for the study of turbulence—the realised quadratic variation $[Y_\delta]_t$ does not converge as it would if Y was a semimartingale; in fact it diverges to infinity whereas in the semimartingale case it will generally converge to the accumulated volatility/intermittency

$$\sigma_t^{2+} = \int_0^t \sigma_s^2 \, ds, \tag{21}$$

which is an object of key interest (in turbulence it represents the coarse-grained energy dissipation). However the situation can be remedied by adjusting $[Y_\delta]_t$ by a factor depending on ν; in wide generality it holds that

$$c\delta^{2(1-\nu)}[Y_\delta]_t \xrightarrow{p} \sigma_t^{2+} \tag{22}$$

as $\delta \longrightarrow 0$, with $x = \lambda^{-1} 2^{2(\nu-1/2)}(\Gamma(\nu) + \Gamma(\nu + 1/2))/\Gamma(2\nu - 1)\Gamma(3/2 - 1)$. To apply this requires knowledge of the value of ν and in general ν must be estimated with sufficient precision to ensure that substituting the estimate for the theoretical value of ν in (22) will still yield convergence in probability. Under relatively mild conditions that is possible, as discussed in [26] and the references therein. An important aspect of formula (20) is that its use does not involve knowledge of ν as the adjustment factor cancels out (Fig. 3).

Convergence in probability and a central limit theorem for the RRQV is established in [10]. Figure 2 illustrates its use, for two sections of the "Brookhaven" dataset,

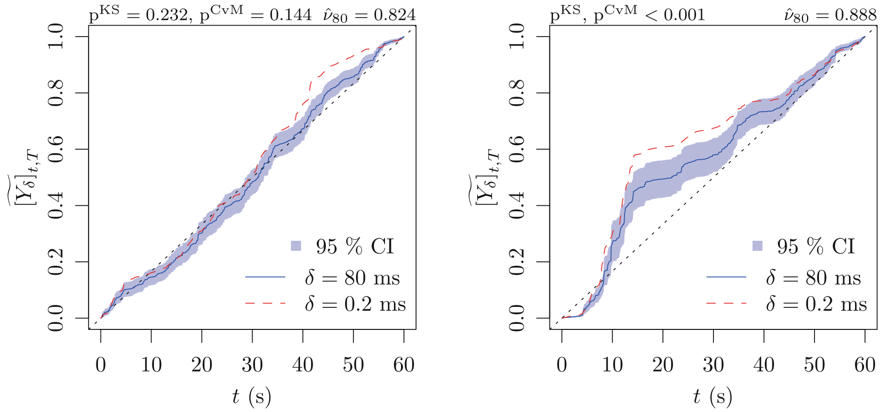

Fig. 3 Brookhaven turbulence data periods 18 and 25—RRQV and 95 % confidence intervals

one where the volatility effect was deemed by eye to be very small and one where it appeared strong. (The "Brookhaven" dataset consist of 20 million one-point measurements of the longitudinal component of the wind velocity in the atmospheric boundary layer, 35 m above ground. The measurements were performed using a hot-wire anemometer and sampled at 5 kHz. The time series can be assumed to be stationary. We refer to [27] for further details on the dataset; the dataset is called no. 3 therein).

4.3 Role of Selfdecomposability

If τ is given by (12) it is automatically infinitely divisible, and selfdecomposable provided L has that property; whereas if τ is defined by (13) it will only in exceptional cases be infinitely divisible.

A non-trivial example of such an exceptional case is the following. The Gumbel distribution with density

$$f(x) = \frac{1}{b} \exp\left(\frac{x-a}{b} - \exp\left(\frac{x-a}{b}\right)\right), \tag{23}$$

where $a \in \mathbb{R}$ and $b > 0$ is infinitely divisible [39]. In [31] it was demonstrated that the one-dimensional marginal distribution of the logarithm of the energy dissipation is accurately described by a normal inverse Gaussian distribution. One may also show (not done here) that the Gumbel distribution with $b = 2$ provides another fit that is nearly as accurate as the normal inverse Gaussian. Furthermore, if X is a Gumbel random variable with $b = 2$, then $\exp(X)$ is distributed as the square of an exponential random variable, hence also infinitely divisible by [39]. Therefore, if the Lévy basis L in (17) is chosen so that $L(A)$ follows a Gumbel distribution with $b = 2$, then $\exp(L(A(t)))$ will be infinitely divisible.

For a general discussion of selfdecomposable fields we refer to [13]. See also [32] which provides a survey of when a selfdecomposable random variable can be represented as a stochastic integral, like in (12). Representations of that kind allow, in particular, the construction of field-valued processes of OU or supOU type that may be viewed as propagating, in time, an initial volatility/intermittency field defined on the spatial component of space-time for a fixed time, say $t = 0$. Similarly, suppose that a model has been formulated for the time-wise development of a stochastic field at a single point in space. One may then seek to define a field on space-time such that at every other point of space the time-wise development of the field is stochastically the same as at the original space point and such that the field as a whole is stationary and selfdecomposable.

Example 1 (*One dimensional turbulence*) Let Y denote an ambit field in the tempo-spatial case where the spatial dimension is 1, and assume that for a preliminary purely temporal model X of the same turbulent phenomenon a model has been formulated for the squared amplitude volatility component, say ω. It may then be desirable to devise Y such that the volatility/intermittency field $\tau = \sigma^2$ is stationary and infinitely divisible, and such that for every spatial position x the law of $\tau(x, \cdot)$ is identical to that formulated for the temporal setting, i.e. ω. If the temporal process is selfdecomposable then, subject to a further weak condition (see [13]), such a field can be constructed.

To sketch how this may proceed, recall first that the classical definition of self-decomposability of a process X says that all the finite-dimensional marginal distributions of X should be selfdecomposable. Accordingly, due to a result by [38], for any finite set of time points $\hat{u} = (u_1, \ldots, u_n)$ the selfdecomposable vector variable $X(\hat{u}) = (X(u_1), \ldots, X(u_n))$ has a representation

$$X(\hat{u}) = \int_0^\infty e^{-\xi}\, L(d\xi, \hat{u})$$

for some n-dimensional Lévy process $L(\cdot, \hat{u})$, provided only that the Lévy measure of $X(\hat{u})$ has finite log-moment. We now assume this to be the case and that X is stationary

Next, for fixed $\hat{u}$, let $\{\tilde{L}(x, \hat{u}) \mid x \in \mathbb{R}\}$, be the n-dimensional Lévy process having the property that the law of $\tilde{L}(1, \hat{u})$ is equal to the law of $X(\hat{u})$. Then the integral

$$X(x, \hat{u}) = \int_{-\infty}^{x} e^{-\xi}\, \tilde{L}(d\xi, \hat{u})$$

exists and the process $\{X(x, \hat{u}) \mid x \in \mathbb{R}\}$ will be stationary—of Ornstein-Uhlenbeck type—while for each x the law of $X(x, \hat{u})$ will be the same as that of $X(\hat{u})$.

However, off hand the Lévy processes $\tilde{L}(\,\cdot\,,\hat{u})$ corresponding to different sets $\hat{u}$ of time points may have no dynamic relationship to each other, while the aim is to obtain a stationary selfdecomposable field $X(x,t)$ such that $X(x,\,\cdot\,)$ has the same law as X for all $x \in \mathbb{R}$. But, arguing along the lines of theorem 3.4 in [9], it is possible to choose the representative processes $\tilde{L}(\,\cdot\,,\hat{u})$ so that they are all defined on a single probability space and are consistent among themselves (in analogy to Kolmogorov's consistency result); and that establishes the existence of the desired field $X(x,t)$. Moreover, $X(\,\cdot\,,\,\cdot\,)$ is selfdecomposable, as is simple to verify.

The same result can be shown more directly using master Lévy measures and the associated Lévy-Ito representations, cf. [13].

Example 2 Assume that X has the form

$$X(u) = \int_{-\infty}^{u} g(u-\xi)\, L(d\xi) \tag{24}$$

where L is a Lévy process.

It has been shown in [13] that, in this case, provided g is integrable with respect to the Lebesgue measure, as well as to L, and if the Fourier transform of g is non-vanishing, then X, as a process, is selfdecomposable if and only if L is selfdecomposable. When that holds we may, as above, construct a selfdecomposable field $X(x,t)$ with $X(x,\,\cdot\,) \sim X(\,\cdot\,)$ for every $x \in \mathbb{R}$ and $X(\,\cdot\,,t)$ of OU type for every $t \in \mathbb{R}$.

As an illustration, suppose that g is the gamma kernel (8) with $\nu \in (1/2, 1)$. Then the Fourier transform of g is

$$\hat{g}(\zeta) = (1 - i\zeta/\lambda)^{-\nu}.$$

and hence, provided that L is such that the integral (24) exists, the field $X(x,u)$ is stationary and selfdecomposable, and has the OU type character described above.

5 Time Change and Universality in Turbulence and Finance

5.1 Distributional Collapse

In [4], Barndorff-Nielsen et al. demonstrate two properties of the distributions of increments $\Delta_\ell X(t) = X(t) - X(t-\ell)$ of turbulent velocities. Firstly, the increment distributions are *parsimonious*, i.e., they are described well by a distribution with few parameters, even across distinct experiments. Specifically it is shown that the four-parameter family of normal inverse Gaussian distributions ($\mathrm{NIG}(\alpha, \beta, \mu, \delta)$) provides excellent fits across a wide range of lags ℓ,

$$\Delta_\ell X \sim \mathrm{NIG}(\alpha(\ell), \beta(\ell), \mu(\ell), \delta(\ell)). \tag{25}$$

Secondly, the increment distributions are *universal*, i.e., the distributions are the same for distinct experiments, if just the scale parameters agree,

$$\Delta_{\ell_1} X_1 \sim \Delta_{\ell_2} X_2 \quad \text{if and only if} \quad \delta_1(\ell_1) = \delta_2(\ell_2), \tag{26}$$

provided the original velocities (not increments) have been non-dimensionalized by standardizing to zero mean and unit variance. Motivated by this, the notion of stochastic equivalence class is introduced.

The line of study initiated in [4] is continued in [16], where the analysis is extended to many more data sets, and it is observed that

$$\Delta_{\ell_1} X_1 \sim \Delta_{\ell_2} X_2 \quad \text{if and only if} \quad \mathrm{var}(\Delta_{\ell_1} X_1) \sim \mathrm{var}(\Delta_{\ell_2} X_2), \tag{27}$$

which is a simpler statement than (26), since it does not involve any specific distribution. In [21], Barndorff-Nielsen et al. extend the analysis from fluid velocities in turbulence to currency and metal returns in finance and demonstrate that (27) holds when X_i denotes the log-price, so increments are log-returns. Further corroboration of the existence of this phenomenon in finance is presented in the following subsection.

A conclusion from the cited works is that within the context of turbulence or finance there exists a family of distributions such that for many distinct experiments and a wide range of lags, the corresponding increments are distributed according to a member of this family. Moreover, this member is uniquely determined by the variance of the increments.

Up till recently these stylised features had not been given any theoretical background. However, in [20], a class of stochastic processes is introduced that exactly has the rescaling property in question.

5.2 A First Look at Financial Data from SP500

Motivated by the developments discussed in the previous subsection, in the following we complement the analyses in [4, 16, 21] with 29 assets from Standard & Poor's 500 stock market index. The following assets were selected for study: AA, AIG, AXP, BA, BAC, C, CAT, CVX, DD, DIS, GE, GM, HD, IBM, INTC, JNJ, JPM, KO, MCD, MMM, MRK, MSFT, PG, SPY, T, UTX, VZ, WMT, XOM. For each asset, between 7 and 12 years of data is available. A sample time series of the log-price of asset C is displayed in Fig. 4, where the thin vertical line marks the day 2008-01-01.

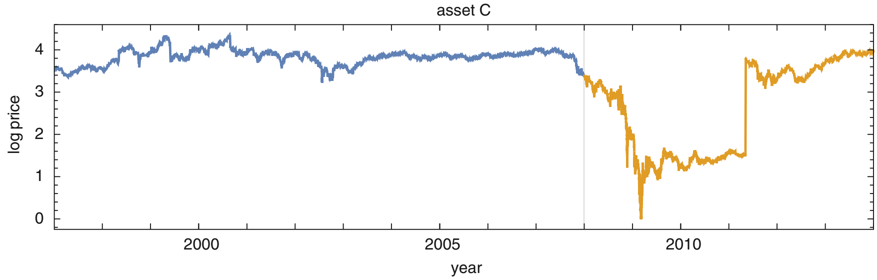

Fig. 4 Time series of the log-price of asset C on an arbitrary scale. The *thin vertical line* marks the day 2008-01-01 and divides the dataset into the two subsets "pre" (*blue*) and "post" (*yellow*)

Asset C is found to be representative of the feature of all the other datasets. Each dataset is divided into two subsets: the "pre" subset consisting of data from before 2008-01-01 and the "post" subset consisting of data from after 2008-01-01. This subdivision was chosen since the volatility in the "post" dataset is visibly higher than in the "pre" dataset, presumably due to the financial crisis. The data has been provided by Lunde (Aarhus University), see also [29].

Figure 5 shows that the distributions of log-returns across a wide range of lags ranging from 1 s to approximately 4.5 h are quite accurately described by normal

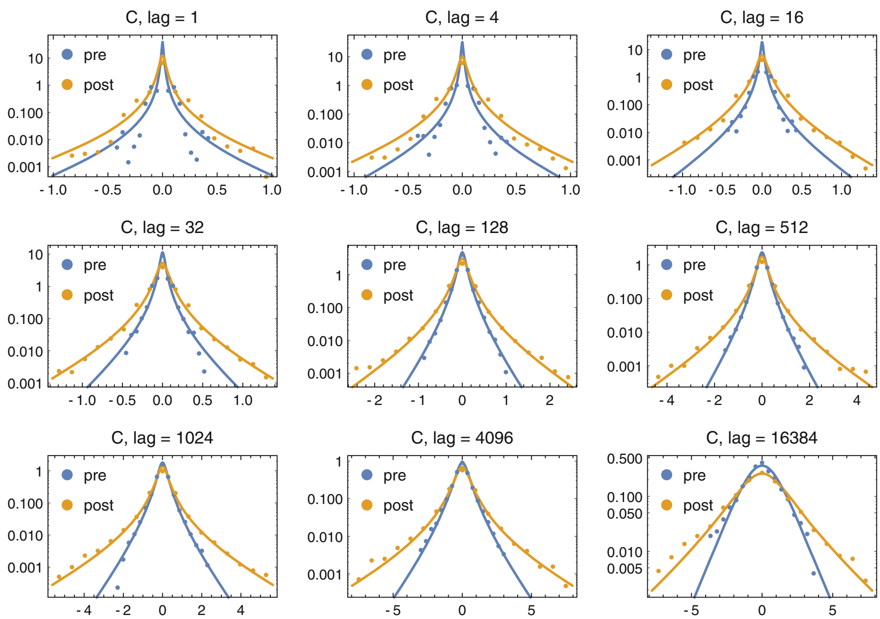

Fig. 5 Probability densities on a log-scale for the log-returns of asset C at various lags ranging from 1 to 16384 s. The *dots* denote the data and the *solid line* denotes the fitted NIG distribution. *Blue* and *yellow* denote the "pre" and "post" datasets, respectively. The log-returns have been multiplied with 100 in order to un-clutter the labeling of the x-axes

inverse Gaussian distributions, except at the smallest lags where the empirical distributions are irregular. We suspect this is due to market microstructure noise. The accuracy of the fits is not surprising given that numerous publications have demonstrated the applicability of the generalised hyperbolic distribution, in particular the subfamily consisting of the normal inverse Gaussian distributions, to describe financial datasets. See for example [1, 14, 15, 28, 40]. We note the transition from a highly peaked distribution towards the Gaussian as the lag increases.

Next, we see on Fig. 6 that the distributions at the same lag of the log-returns for the 29 assets are quite different, that is, they do not collapse onto the same curve. This holds for both the "pre" and the "post" datasets. However, the transition from a highly peaked distribution at small lags towards a Gaussian at large lags hints that a suitable change of time, though highly nonlinear, may cause such a collapse. Motivated by the observations in [21] we therefore consider the variance of the log-returns as a function of the lag. Figure 7 shows how the variance depends on the lag. Except at the smallest lags, a clear power law is observed. The behaviour at the smallest lags is due to market microstructure noise [29]. Nine variances have been selected to represent most of the variances observed in the 29 assets. For each selected variance and each asset the corresponding lag is computed. We note that for the smallest lags/variances this is not without difficulty since for some of the assets the slope approaches zero.

Finally, Fig. 8 displays the distributions of log-returns where the lag for each asset has been chosen such that the variance is the one specified in each subplot. The difference between Figs. 6 and 8 is pronounced. We see that for both the "pre" and the "post" dataset, the distributions corresponding to the same variance tend to be the same. Furthermore, when the "pre" and "post" datasets are displayed together, essentially overlaying the top part of Fig. 8 with the bottom part, a decent overlap is still observed. So while the distributions in Fig. 8 do not collapse perfectly onto the same curve for all the chosen variances, in contrast to what is the case for velocity increments in turbulence (see [4]), we are invariably led to the preliminary conclusion that also in the case of the analysed assets from S&P500, a family of distributions exists such that all distributions of log-returns are members of this family and such that the variance of the log-returns uniquely determines this member. The lack of collapse at the smaller variances may in part be explained by the difficulty in reading off the corresponding lags.

The observed parsimony and in particular universality has implications for modelling since any proper model should possess both features. Within the context of turbulence, BSS-processes have been shown to be able to reproduce many key features of turbulence, see [35] and the following subsection for a recent example. The extent to which BSS-processes in general possess universality is still ongoing research [20] but results indicate that BSS-processes and in general LSS-process are good candidates for models where parsimony and universality are desired features.

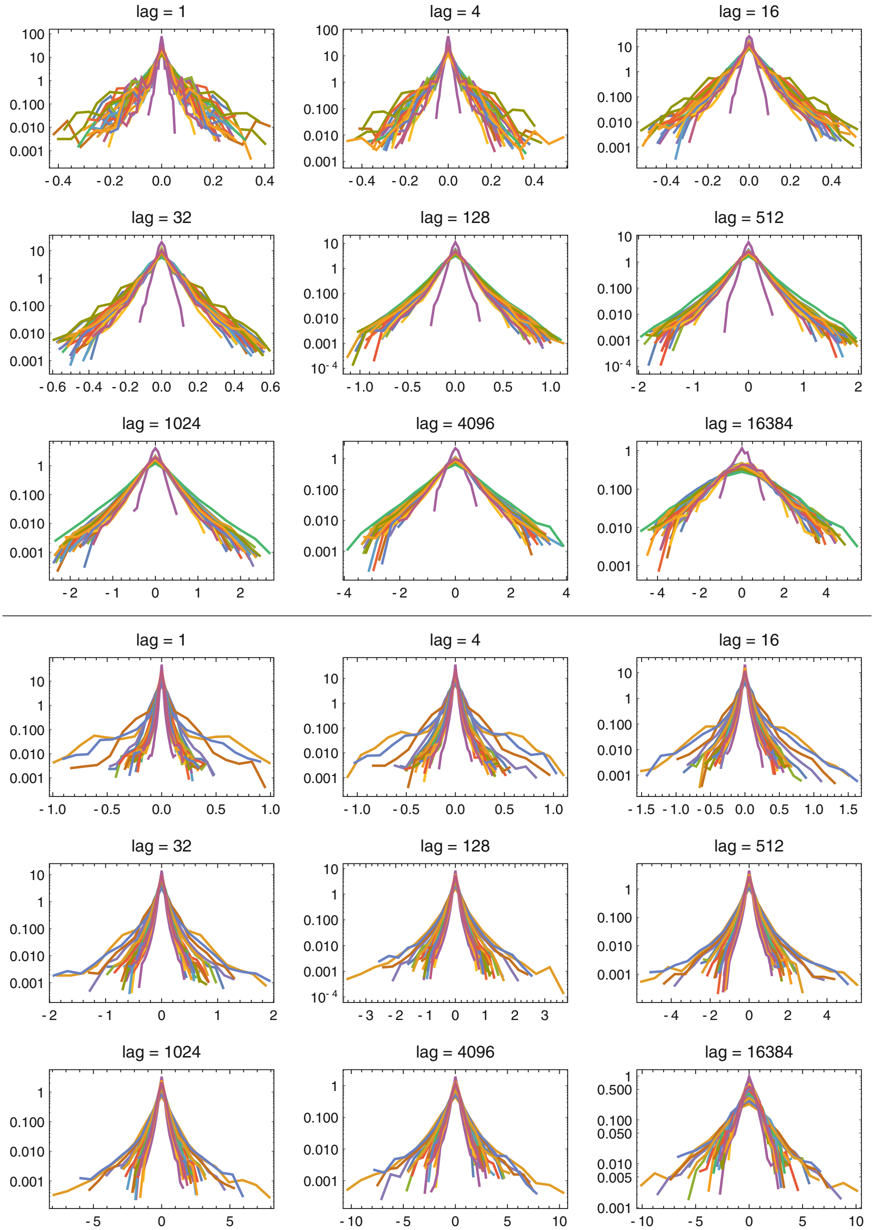

Fig. 6 Probability densities on a log-scale for the log-returns of all 29 assets at various lags ranging from 1 to 16384 s. The *top* and *bottom* halfs represent the "pre" and "post" datasets, respectively. The log-returns have been multiplied with 100 in order to un-clutter the labeling of the x-axes

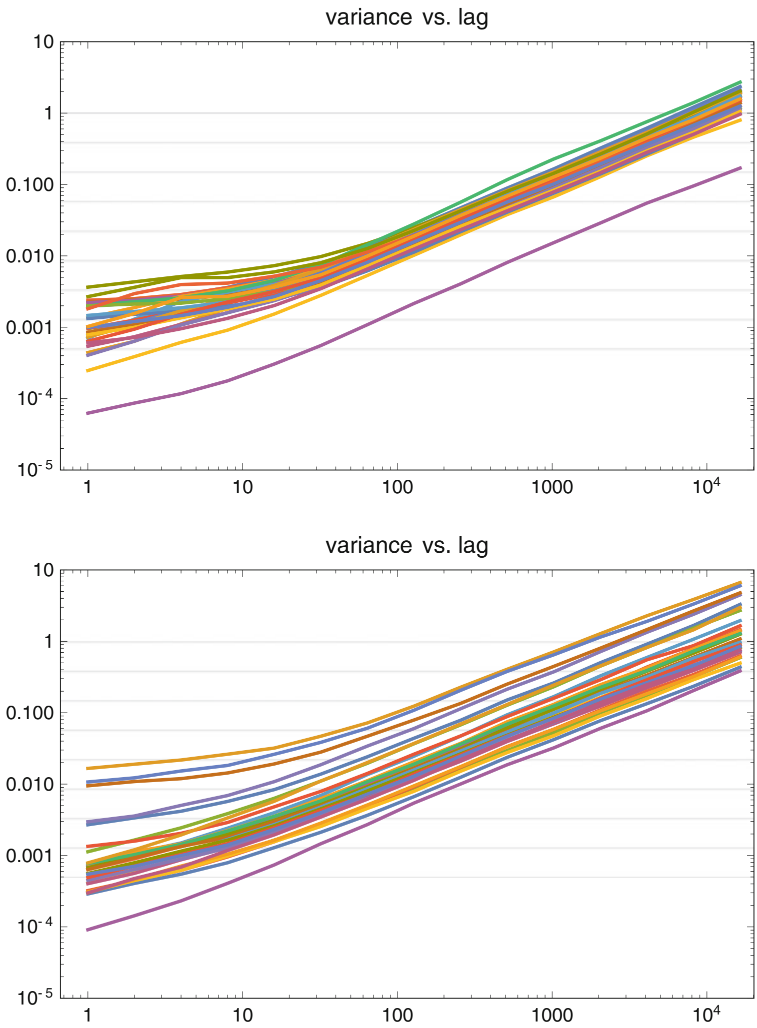

Fig. 7 The variance of the log-returns for the 29 assets as a function of the lag displayed in a double logarithmic representation. The *top* and *bottom* graphs represent the "pre" and "post" datasets, respectively

5.3 Modelling Turbulent Velocity Time Series

A specific time-wise version of (1), called Brownian semistationary processes has been proposed in [18, 19] as a model for turbulent velocity time series. It was shown that BSS processes in combination with continuous cascade models (exponentials of certain trawl processes) are able to qualitatively capture some main stylized features of turbulent time series.

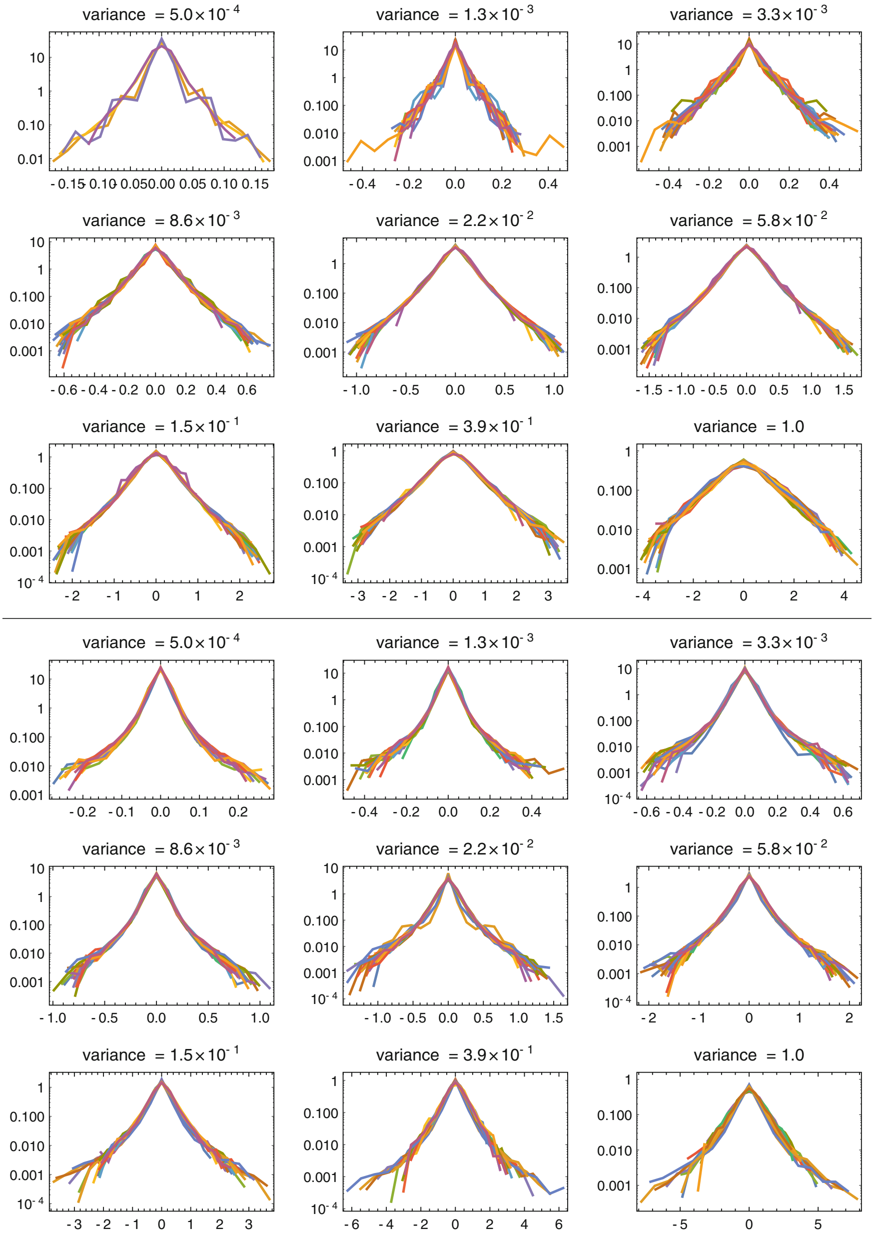

Fig. 8 Probability densities on a log-scale of log-returns where the lag for each asset has been chosen such that the variances of the assets in each subplot is the same. The chosen variances are also displayed in Fig. 7 as *horizonal lines*. For the smallest and largest variances, not all dataset are present since for some datasets those variances are not attained. *Top* "pre", *bottom* "post"

Recently this analysis has been extended to a quantitative comparison with turbulent data [31, 35]. More specifically, based on the results for the energy dissipation oulined in Sect. 4.1, BSS processes have been analyzed and compared in detail to turbulent velocity time series in [35] by directly estimating the model parameters from data. Here we briefly summarize this analysis.

Time series of the main component v_t of the turbulent velocity field are modelled as a BSS process of the specific form

$$v(t) = v(t; g, \sigma, \beta) = \int_{-\infty}^{t} g(t-s)\sigma(s)B(ds) + \beta \int_{-\infty}^{t} g(t-s)\sigma(s)^2\, ds =: R(t) + \beta S(t) \tag{28}$$

where g is a non-negative $L^2(\mathbb{R}_+)$ function, σ is a stationary process independent of B, β is a constant and B denotes standard Brownian motion. An argument based on quadratic variation shows that when $g(0+) \neq 0$, then σ^2 can be identified with the surrogate energy dissipation, $\sigma^2 = \varepsilon$, where ε is the process given by (17). The kernel g is specified as a slightly shifted convolution of gamma kernels [30],

$$\begin{aligned} g(t) &= g_0(t + t_0), \\ g_0(t) &= a t^{\nu_1+\nu_2-1} \exp(-\lambda_2 t)\, {}_1F_1(\nu_1, \nu_1 + \nu_2, (\lambda_2 - \lambda_1)t) 1_{(0,\infty)}(t) \end{aligned}$$

with $a > 0$, $\nu_i > 0$ and $\lambda_i > 0$. Here ${}_1F_1$ denotes the Kummer confluent hypergeometric function. The shift is needed to ensure that $g(0+) \neq 0$.

The data set analysed consists of one-point time records of the longitudinal (along the mean flow) velocity component in a gaseous helium jet flow with a Taylor Reynolds number $R_\lambda = 985$. The same data set is also analyzed in [31] and the estimated parameters there are used to specify $\sigma^2 = \varepsilon$ in (28). The remaining parameters for the kernel g and the constant β can then be estimated from the second and third order structure function, that is, the second and third order moments of velocity increments. In [35] it is shown that the second order structure function is excellently reproduced and that the details of the third order structure function are well captured. It is important to note that the model is completely specified from the energy dissipation statistics and the second and third order structure functions.

The estimated model for the velocity is then succesfully compared with other derived quantities, including higher order structure functions, the distributions of velocity increments and their evolution as a function of lag, the so-called Kolmogorov variable and the energy dissipation, as prediced by the model.

6 Conclusion and Outlook

The present paper highlights some of the most recent developments in the theory and applications of Ambit Stochastics. In particular, we have discussed the existence of the ambit fields driven by metatime changed Lévy bases, selfdecomposability of random fields [13], applications of BSS processes in the modelling of turbulent time series [35] and new results on the distributional collapse in financial data. Some of the topics not mentioned here but also under development are the integration theory with respect to time-changed volatility modulated Lévy bases [7]; integration with respect to volatility Gaussian processes in the White Noise Analysis setting in the spirit of [34] and extending [6]; modelling of multidimensional turbulence based on ambit fields; and in-depth study of parsimony and universality in BSS and LSS processes motivated by some of the discussions in the present paper.

References

1. Barndorff-Nielsen, O.E.: Processes of normal inverse Gaussian type. Finance Stoch. **2**(1), 41–68 (1998)
2. Barndorff-Nielsen, O.E.: Stationary infinitely divisible processes. Braz. J. Probab. Stat. **25**(3), 294–322 (2011)
3. Barndorff-Nielsen, O.E.: Notes on the gamma kernel. Research Report 3, Thiele Centre for Applied Mathematics in Natural Science. (Aarhus University, 2012)
4. Barndorff-Nielsen, O.E., Blæsild, P., Schmiegel, J.: A parsimonious and universal description of turbulent velocity increments. Eur. Phys. J. B **41**, 345–363 (2004)
5. Barndorff-Nielsen, O.E., Lunde, A., Shephard, N., Veraart, A.E.D.: Integer-valued trawl processes: a class of infinitely divisible processes. Scand. J. Stat. **41**(3), 693–724 (2013)
6. Barndorff-Nielsen, O.E., Benth, F.E., Szozda, B.: Stochastic integration for volatility modulated Brownian-driven processes via white noise analysis. Infin. Dimens. Anal. Quantum Probab. **17**(2), 1450011 (2014)
7. Barndorff-Nielsen, O.E., Benth, F.E., Szozda, B.: Integration with respect to time-changed volatility modulated Lévy-driven Volterra processes. (In preparation, 2015)
8. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. Banach Center Publications (To appear). (2015)
9. Barndorff-Nielsen, O.E., Maejima, M., Sato, K.: Infinite divisibility for stochastic processes and time change. J. Theor. Probab. **19**, 411–446 (2006)
10. Barndorff-Nielsen, O.E., Pakkanen, M., Schmiegel, J.: Assessing relative volatility/intermittency/energy dissipation. Electron. J. Stat. **8**, 1996–2021 (2014)
11. Barndorff-Nielsen, O.E., Pedersen, J.: Meta-times and extended subordination. Theory Probab. Appl. **56**(2), 319–327 (2010)
12. Barndorff-Nielsen, O., Pérez-Abreu, V., Thorbjørnsen, S.: Lévy mixing. Lat. Am. J. Probab. Math. Stat. **10**, 921–970 (2013)
13. Barndorff-Nielsen, O.E., Sauri, O., Szozda, B.: Selfdecomposable fields. (In preparation, 2015)
14. Barndorff-Nielsen, O.E., Shephard, N.: Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. J. Royal Stat. Soc. **63**, 167–241 (2001)
15. Barndorff-Nielsen, O.E., Shephard, N., Sauri, O., Hedevang, E.: Levy-driven Models with volatility. (In preparation, 2015)
16. Barndorff-Nielsen, O.E., Schmiegel, J.: Time change and universality in turbulence. In: Research Report 15, Thiele Centre for Applied Mathematics in Natural Science, Aarhus University. (2006)

17. Barndorff-Nielsen, O.E., Schmiegel, J.: Ambit processes; with applications to turbulence and tumour growth. In: Benth, F.E., Nunno, G.D., Linstrøm, T., Øksendal, B., Zhang, T. (eds.) Stochastic Analysis and Applications: The Abel Symposium 2005, pp. 93–124. Springer, Heidelberg (2007)
18. Barndorff-Nielsen, O.E., Schmiegel, J.: A stochastic differential equation framework for the timewise dynamics of turbulent velocities. Theory Probab. Its Appl. **52**, 372–388 (2008)
19. Barndorff-Nielsen, O.E., Schmiegel, J.: Brownian semistationary processes and volatility/intermittency. In: Albrecher, H., Rungaldier, W., Schachermeyer, W. (eds.) Advanced financial modelling. Radon Series on Computational and Applied Mathematics, vol. 8. pp. 1–26. W. de Gruyter, Berlin (2009)
20. Barndorff-Nielsen, O.E., Hedevang, E., Schmiegel, J.: Incremental similarity and turbulence. (Submitted.) (2015)
21. Barndorff-Nielsen, O.E., Schmiegel, J., Shephard, N.: Time change and Universality in turbulence and finance. In: Research Report 18, Thiele Centre for Applied Mathematics in Natural Science, Aarhus University. (2006)
22. Barndorff-Nielsen, O.E., Shiryaev, A.N.: Change of Time and Change of Measure. World Scientific, Singapore (2010)
23. Basse-O'Connor, A., Graversen, S.E., Pedersen, J.: Stochastic integration on the real line. Theory Probab. Its Appl. **58**, 355–380 (2013)
24. Basse-O'Connor, A., Lachièze-Rey, R., Podolskij, M.: Limit theorems for stationary increments Lévy driven moving average processes. (To appear). (2015)
25. Chong, C., Klüppelberg, C.: Integrability conditions for space-time stochastic integrals: theory and applications. Bernoulli. In press (2015)
26. Corcuera, J.M., Hedevang, E., Pakkanen, M., Podolskij, M.: Asymptotic theory for Brownian semi-stationary processes with application to turbulence. Stochast. Proc. Their Appl. **123**, 2552–2574 (2013)
27. Dhruva, B.R.: An experimental study of high Reynolds number turbulence in the atmosphere. PhD Thesis, Yale University (2000)
28. Eberlein, E.: Application of generalized hyperbolic Lévy motions to finance. In: Barndorff-Nielsen, O.E. et al. (eds.) Lévy Processes. Theory and Applications, 319–336. Birkhäuser (2001)
29. Hansen, P.R., Lunde, A.: Realized variance and market microstructure noise. J. Bus. Econ. Stat. **24**(2), 127–161 (2006)
30. Hedevang, E.: Stochastic modelling of turbulence: with applications to wind energy. Ph.D thesis. Aarhus University (2012)
31. Hedevang, E., Schmiegel, J.: A causal continuous-time stochastic model for the turbulent energy cascade in a helium jet flow. J. Turbul. **14**(11), 1–26 (2014)
32. Jurek, Z.J.: The random integral representation conjecture a quarter of a century later. Lith. Math. J. **51**, 362–369 (2011)
33. von Karman, T.: Progress in the statistical theory of turbulence. J. Marine Res. **7**, 252–264 (1948)
34. Lebovits, J.: Stochastic calculus with respect to gaussian processes (2014). arXiv:1408.1020
35. Márquez, J.U., Schmiegel, J.: Modelling turbulent time series by BSS-processes. (Submitted.) (2015)
36. Podolskij, M.: Ambit fields: survey and new challenges. (To Appear in XI Proceedings of Symposium of Probability and Stochastic Processes) (2014)
37. Rajput, B.S., Rosiński, J.: Spectral representations of infinitely divisible processes. Probab. Theory Rel. Fields **82**, 451–487 (1989)
38. Sato, K., Yamazato, M.: Operator selfdecomposable distributions as limits distributions of processes of Ornstein-Uhlenbeck type. Stoch. Process. Appl. **17**(1), 73–100 (1984)
39. Steutel, F.W., van Harn, K.: Infinite Divisibility of Probability Distributions on the Real Line. Marcel Dekker (2004)
40. Veraart, A.E.D., Veraart, L.A.M.: Modelling electricity day-ahead prices by multivariate Lévy semistationary processes. In: Benth, F.E., Kholodnyi, V., Laurence, P. (eds.) Quantitative Energy Finance, pp. 157–188. Springer (2014)

Functional and Banach Space Stochastic Calculi: Path-Dependent Kolmogorov Equations Associated with the Frame of a Brownian Motion

Andrea Cosso and Francesco Russo

Abstract First, we revisit basic theory of functional Itô/path-dependent calculus, using the formulation of calculus via regularization. Relations with the corresponding Banach space valued calculus are explored. The second part of the paper is devoted to the study of the Kolmogorov type equation associated with the so called window Brownian motion, called path-dependent heat equation, for which well-posedness at the level of strict solutions is established. Then, a notion of strong approximating solution, called strong-viscosity solution, is introduced which is supposed to be a substitution tool to the viscosity solution. For that kind of solution, we also prove existence and uniqueness.

Keywords Horizontal and vertical derivative · functional Itô/path-dependent calculus · Banach space stochastic calculus · Strong-viscosity solutions · Calculus via regularization

2010 Math Subject Classification: 35D35 · 35D40 · 35K10 · 60H05 · 60H10 · 60H30

1 Introduction

The present work collects several results obtained in the papers [9, 10], focusing on the study of some specific examples and particular cases, for which an ad hoc analysis is developed. This work is an improved version of [8], trying to explain more

A. Cosso
Laboratoire de Probabilités et Modèles Aléatoires,
CNRS, UMR 7599, Université Paris Diderot, Paris, France
e-mail: andrea.cosso@math.univ-paris-diderot.fr

F. Russo (✉)
Unité de Mathématiques Appliquées, ENSTA ParisTech,
Université Paris-Saclay, 828, Boulevard des Maréchaux, 91120 Palaiseau, France
e-mail: francesco.russo@ensta-paristech.fr

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_2

precisely some details. For example, in [8] a slightly more restrictive definition of strong-viscosity solution was adopted, see Remark 12.

Recently, a new branch of stochastic calculus has appeared, known as *functional Itô calculus*, which results to be an extension of classical Itô calculus to functionals depending on the entire path of a stochastic process and not only on its current value, see Dupire [17], Cont and Fournié [5–7]. Independently, Di Girolami and Russo, and more recently Fabbri, Di Girolami, and Russo, have introduced a stochastic calculus via regularizations for processes taking values in a separable Banach space B (see [12–16]), including the case $B = C([-T, 0])$, which concerns the applications to the path-dependent calculus.

In the first part of the present paper, we follow [9] and revisit functional Itô calculus by means of stochastic calculus via regularization. We recall that Cont and Fournié [5–7] developed functional Itô calculus and derived a functional Itô's formula using discretization techniques of Föllmer [23] type, instead of regularization techniques, which in our opinion, better fit to the notion of derivative. Let us illustrate another difference with respect to [5]. One of the main issues of functional Itô calculus is the definition of the functional (or pathwise) derivatives, i.e., the horizontal derivative (calling in only the past values of the trajectory) and the vertical derivative (calling in only the present value of the trajectory). In [5], it is essential to consider functionals defined on the space of càdlàg trajectories, since the definition of functional derivatives necessitates of discontinuous paths. Therefore, if a functional is defined only on the space of continuous trajectories (because, e.g., it depends on the paths of a continuous process as Brownian motion), we have to extend it anyway to the space of càdlàg trajectories, even though, in general, there is no unique way to extend it. In contrast to this approach, we introduce a new space larger than the space of continuous trajectories $C([-T, 0])$, denoted by $\mathscr{C}([-T, 0])$, which allows us to define functional derivatives. $\mathscr{C}([-T, 0])$ is the space of bounded trajectories on $[-T, 0]$, continuous on $[-T, 0[$ and with possibly a jump at 0. We endow $\mathscr{C}([-T, 0])$ with a topology such that $C([-T, 0])$ is dense in $\mathscr{C}([-T, 0])$ with respect to this topology. Therefore, any functional $\mathscr{U} : [0, T] \times C([-T, 0]) \to \mathbb{R}$, continuous with respect to the topology of $\mathscr{C}([-T, 0])$, admits a unique extension to $\mathscr{C}([-T, 0])$, denoted $u : [0, T] \times \mathscr{C}([-T, 0]) \to \mathbb{R}$. We present some significant functionals for which a continuous extension exists. Then, we develop the functional Itô calculus for $u : [0, T] \times \mathscr{C}([-T, 0]) \to \mathbb{R}$.

Notice that we use a slightly different notation compared with [5]. In particular, in place of a map $\mathscr{U} : [0, T] \times C([-T, 0]) \to \mathbb{R}$, in [5] a family of maps $F = (F_t)_{t\in[0,T]}$, with $F_t : C([0, t]) \to \mathbb{R}$, is considered. However, we can always move from one formulation to the other. Indeed, given $F = (F_t)_{t\in[0,T]}$, where each $F_t : C([0, t]) \to \mathbb{R}$, we can define $\mathscr{U} : [0, T] \times C([-T, 0]) \to \mathbb{R}$ as follows:

$$\mathscr{U}(t, \eta) := F_t(\eta(\cdot + T)|_{[0,t]}), \qquad (t, \eta) \in [0, T] \times C([-T, 0]).$$

Vice-versa, let $\mathscr{U} : [0, T] \times C([-T, 0]) \to \mathbb{R}$ and define $F = (F_t)_{t\in[0,T]}$ as

$$F_t(\tilde{\eta}) := \mathscr{U}(t, \eta), \qquad (t, \tilde{\eta}) \in [0, T] \times C([0, t]), \tag{1}$$

where η is the element of $C([-T, 0])$ obtained from $\tilde{\eta}$ firstly translating $\tilde{\eta}$ on the interval $[-t, 0]$, then extending it in a constant way up to $-T$, namely $\eta(x) := \tilde{\eta}(x + t)1_{[-t,0]}(x) + \tilde{\eta}(-t)1_{[-T,-t)}(x)$, for any $x \in [-T, 0]$. Observe that, in principle, the map $\mathscr{U}$ contains more information than F, since in (1) we do not take into account the values of $\mathscr{U}$ at $(t, \eta) \in [0, T] \times C([-T, 0])$ with η not constant on the interval $[-T, -t]$. Despite this, the equivalence between the two notations is guaranteed; indeed, when we consider the composition of $\mathscr{U}$ with a stochastic process, this extra information plays no role. Our formulation has two advantages. Firstly, we can work with a single map instead of a family of maps. In addition, the time variable and the path have two distinct roles in our setting, as for the time variable and the space variable in the classical Itô calculus. This, in particular, allows us to define the horizontal derivative independently of the time derivative, so that, the horizontal derivative defined in [5] corresponds to the sum of our horizontal derivative and of the time derivative. We mention that an alternative approach to functional derivatives was introduced in [1].

We end the first part of the paper showing how our functional Itô's formula is strictly related to the Itô's formula derived in the framework of Banach space valued stochastic calculus via regularization, for the case of window processes. This new branch of stochastic calculus has been recently conceived and developed in many directions in [12, 14–16]; for more details see [13]. For the particular case of window processes, we also refer to Theorem 6.3 and Sect. 7.2 in [12]. In the present paper, we prove formulae which allow to express functional derivatives in terms of differential operators arising in the Banach space valued stochastic calculus via regularization, with the aim of identifying the building blocks of our functional Itô's formula with the terms appearing in the Itô's formula for window processes.

Dupire [17] introduced also the concept of *path-dependent partial differential equation*, to which the second part of the present paper is devoted. Di Girolami and Russo, in Chap. 9 of [13], considered existence of regular solutions associated with a path dependent heat equation (which is indeed the Kolmogorov equation related to window Brownian motion) with a Fréchet smooth final condition. This was performed in the framework of Banach space valued calculus, for which we refer also to [22]. A flavour of the notion of regular solution in the Banach space framework, appeared in Chap. IV of [30] which introduced the notion of *weak infinitesimal generator* (in some weak sense) of the window Brownian motion and more general solutions of functional dependent stochastic differential equations. Indeed, the monograph [30] by Mohammed constitutes an excellent early contribution to the theory of this kind of equations.

We focus on semilinear parabolic path-dependent equations associated to the window Brownian motion. For more general equations we refer to [9] (for strict solutions) and to [10] (for strong-viscosity solutions). First, we consider regular solution, which we call *strict solutions*, in the framework of functional Itô calculus. We prove a uniqueness result for this kind of solution, showing that, if a strict solution exists, then it can be expressed through the unique solution to a certain backward stochastic differential equation (BSDE). Then, we prove an existence result for strict solutions.

However, this notion of solution turns out to be unsuitable to deal with all significant examples. As a matter of fact, if we consider the path-dependent PDE arising in the hedging problem of lookback contingent claims, we can not expect too much regularity of the solution (this example is studied in detail in Sect. 3.2). Therefore, we are led to consider a weaker notion of solution. In particular, we are interested in a viscosity-type solution, namely a solution which is not required to be differentiable.

The issue of providing a suitable definition of viscosity solutions for path-dependent PDEs has attracted a great interest, see Peng [33] and Tang and Zhang [42], Ekren et al. [18–20], Ren et al. [34]. In particular, the definition of viscosity solution provided by [18–20, 34] is characterized by the fact that the classical minimum/maximum property, which appears in the standard definition of viscosity solution, is replaced with an optimal stopping problem under nonlinear expectation [21]. Then, probability plays an essential role in this latter definition, which can not be considered as a purely analytic object as the classical definition of viscosity solution is; it is, more properly, a probabilistic version of the classical definition of viscosity solution. We also emphasize that a similar notion of solution, called stochastic weak solution, has been introduced in the recent paper [29] in the context of variational inequalities for the Snell envelope associated to a non-Markovian continuous process X. Those authors also revisit functional Itô calculus, making use of stopping times. This approach seems very promising. Instead, our aim is to provide a definition of viscosity type solution, which has the peculiarity to be a purely analytic object; this will be called a *strong-viscosity solution* to distinguish it from the classical notion of viscosity solution. A strong-viscosity solution to a path-dependent partial differential equation is defined, in a few words, as the pointwise limit of strict solutions to perturbed equations. We notice that the definition of strong-viscosity solution is similar in spirit to the *vanishing viscosity* method, which represents one of the primitive ideas leading to the conception of the modern definition of viscosity solution. Moreover, it has also some similarities with the definition of *good solution*, which turned out to be equivalent to the definition of L^p-viscosity solution for certain fully nonlinear partial differential equations, see, e.g., [3, 11, 27, 28]. Finally, our definition is likewise inspired by the notion of *strong solution* (which justifies the first word in the name of our solution), as defined for example in [2, 24, 25], even though strong solutions are required to be more regular (this regularity is usually required to prove uniqueness of strong solutions, which for example in [24, 25] is based on a Fukushima-Dirichlet decomposition). Instead, our definition of strong-viscosity solution to the path-dependent semilinear Kolmogorov equation is not required to be continuous, as in the spirit of viscosity solutions. The term *viscosity* in the name of our solution is also justified by the fact that in the finite dimensional case we have an equivalence result between the notion of strong-viscosity solution and that of viscosity solution, see Theorem 3.7 in [8]. We prove a uniqueness theorem for strong-viscosity solutions using the theory of backward stochastic differential equations and we provide an existence result. We refer to [10] for more general results (when the path-dependent equation is not the path-dependent heat equation) and also for the application of strong-viscosity solutions to standard semilinear parabolic PDEs.

The paper is organized as follows. In Sect. 2 we develop functional Itô calculus via regularization following [9]: after a brief introduction on finite dimensional stochastic calculus via regularization in Sect. 2.1, we introduce and study the space $\mathscr{C}([-T, 0])$ in Sect. 2.2; then, we define the pathwise derivatives and we prove the functional Itô's formula in Sect. 2.3; in Sect. 2.4, instead, we discuss the relation between functional Itô calculus via regularization and Banach space valued stochastic calculus for window processes. In Sect. 3, on the other hand, we study path-dependent PDEs following [10]. More precisely, in Sect. 3.1 we discuss strict solutions; in Sect. 3.2 we present a significant hedging example to motivate the introduction of a weaker notion of solution; finally, in Sect. 3.3 we provide the definition of strong-viscosity solution.

2 Functional Itô Calculus: A Regularization Approach

2.1 Background: Finite Dimensional Calculus via Regularization

The theory of stochastic calculus via regularization has been developed in several papers, starting from [37, 38]. We recall below only the results used in the present paper, and we refer to [40] for a survey on the subject. We emphasize that integrands are allowed to be anticipating. Moreover, the integration theory and calculus appear to be close to a pure pathwise approach even though there is still a probability space behind.

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $T \in]0, \infty[$. Let $\mathbb{F} = (\mathscr{F}_t)_{t\in[0,T]}$ denote a filtration satisfying the usual conditions. Let $X = (X_t)_{t\in[0,T]}$ (resp. $Y = (Y_t)_{t\in[0,T]}$) be a real continuous (resp. $\mathbb{P}$-a.s. integrable) process. Every real continuous process $X = (X_t)_{t\in[0,T]}$ is naturally extended to all $t \in \mathbb{R}$ setting $X_t = X_0$, $t \leq 0$, and $X_t = X_T$, $t \geq T$. We also define a $C([-T, 0])$-valued process $\mathbb{X} = (\mathbb{X}_t)_{t\in\mathbb{R}}$, called the **window process** associated with X, defined by

$$\mathbb{X}_t := \{X_{t+x},\ x \in [-T, 0]\}, \qquad t \in \mathbb{R}. \tag{2}$$

This corresponds to the so-called *segment process* which appears for instance in [43].

Definition 1 Suppose that, for every $t \in [0, T]$, the limit

$$\int_0^t Y_s d^- X_s \ := \ \lim_{\varepsilon\to 0^+} \int_0^t Y_s \frac{X_{s+\varepsilon} - X_s}{\varepsilon} ds, \tag{3}$$

exists in probability. If the obtained random function admits a continuous modification, that process is denoted by $\int_0^\cdot Y d^- X$ and called **forward integral of Y with respect to X**.

Definition 2 A family of processes $(H_t^{(\varepsilon)})_{t\in[0,T]}$ is said to converge to $(H_t)_{t\in[0,T]}$ in the **ucp sense**, if $\sup_{0\le t\le T}|H_t^{(\varepsilon)}-H_t|$ goes to 0 in probability, as $\varepsilon\to 0^+$.

Proposition 1 *Suppose that the limit* (3) *exists in the ucp sense. Then, the forward integral* $\int_0^\cdot Y d^- X$ *of* Y *with respect to* X *exists.*

Let us introduce the concept of covariation, which is a crucial notion in stochastic calculus via regularization. Let us suppose that X, Y are continuous processes.

Definition 3 The **covariation of** X **and** Y is defined by

$$[X,Y]_t \;=\; [Y,X]_t \;=\; \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\int_0^t (X_{s+\varepsilon}-X_s)(Y_{s+\varepsilon}-Y_s)ds, \qquad t\in[0,T],$$

if the limit exists in probability for every $t\in[0,T]$, provided that the limiting random function admits a continuous version (this is the case if the limit holds in the ucp sense). If $X=Y$, X is said to be a **finite quadratic variation process** and we set $[X]:=[X,X]$.

The forward integral and the covariation generalize the classical Itô integral and covariation for semimartingales. In particular, we have the following result, for a proof we refer to, e.g., [40].

Proposition 2 *The following properties hold.*

(i) *Let* S^1, S^2 *be continuous* $\mathbb{F}$*-semimartingales. Then,* $[S^1,S^2]$ *is the classical bracket* $[S^1,S^2]=\langle M^1,M^2\rangle$*, where* M^1 (*resp.* M^2) *is the local martingale part of* S^1 (*resp.* S^2).

(ii) *Let* V *be a continuous bounded variation process and* Y *be a càdlàg process* (*or vice-versa*); *then* $[V]=[Y,V]=0$*. Moreover* $\int_0^\cdot Y d^- V=\int_0^\cdot Y dV$*, is the* **Lebesgue-Stieltjes integral**.

(iii) *If* W *is a Brownian motion and* Y *is an* $\mathbb{F}$*-progressively measurable process such that* $\int_0^T Y_s^2 ds<\infty$*,* $\mathbb{P}$*-a.s., then* $\int_0^\cdot Y d^- W$ *exists and equals the Itô integral* $\int_0^\cdot Y dW$.

We could have defined the forward integral using limits of non-anticipating Riemann sums. Another reason to use the regularization procedure is due to the fact that it extends the Itô integral, as Proposition 2(iii) shows. If the integrand had uncountable jumps (as Y being the indicator function of the rational number in $[0,1]$) then, the Itô integral $\int_0^\cdot Y dW$ would be zero $Y=0$ a.e. The limit of Riemann sums $\sum_i Y_{t_i}(W_{t_{i+1}}-W_{t_i})$ would heavily depend on the discretization grid.

We end this crash introduction to finite dimensional stochastic calculus via regularization presenting one of its cornerstones: Itô's formula. It is a well-known result in the theory of semimartingales, but it also extends to the framework of finite quadratic variation processes. For a proof we refer to Theorem 2.1 of [39].

Theorem 1 *Let $F : [0,T] \times \mathbb{R} \longrightarrow \mathbb{R}$ be of class $C^{1,2}([0,T] \times \mathbb{R})$ and $X = (X_t)_{t\in[0,T]}$ be a real continuous finite quadratic variation process. Then, the following* ***Itô's formula*** *holds, $\mathbb{P}$-a.s.,*

$$\begin{aligned} F(t, X_t) = \; & F(0, X_0) + \int_0^t \partial_t F(s, X_s)ds + \int_0^t \partial_x F(s, X_s)d^- X_s \\ & + \frac{1}{2}\int_0^t \partial^2_{x\,x} F(s, X_s)d[X]_s, \qquad\qquad 0 \le t \le T. \end{aligned} \tag{4}$$

2.1.1 The Deterministic Calculus via Regularization

A useful particular case of finite dimensional stochastic calculus via regularization arises when Ω is a singleton, i.e., when the calculus becomes deterministic. In addition, in this deterministic framework we will make use of the *definite integral* on an interval $[a, b]$, where $a < b$ are two real numbers. Typically, we will consider $a = -T$ or $a = -t$ and $b = 0$.

We start with two conventions. By default, every bounded variation function $f : [a, b] \to \mathbb{R}$ will be considered as càdlàg. Moreover, given a function $f : [a, b] \to \mathbb{R}$, we will consider the following two extensions of f to the entire real line:

$$f_J(x) := \begin{cases} 0, & x > b, \\ f(x), & x \in [a, b], \\ f(a), & x < a, \end{cases} \qquad\qquad f_{\overline{J}}(x) := \begin{cases} f(b), & x > b, \\ f(x), & x \in [a, b], \\ 0, & x < a, \end{cases}$$

where $J :=]a, b]$ and $\overline{J} = [a, b[$.

Definition 4 Let $f, g : [a, b] \to \mathbb{R}$ be càdlàg functions.
(i) Suppose that the following limit

$$\int_{[a,b]} g(s)d^- f(s) \;:=\; \lim_{\varepsilon\to 0^+} \int_{\mathbb{R}} g_J(s)\frac{f_{\overline{J}}(s+\varepsilon) - f_{\overline{J}}(s)}{\varepsilon}ds, \tag{5}$$

exists and it is finite. Then, the obtained quantity is denoted by $\int_{[a,b]} g d^- f$ and called (***deterministic, definite***) ***forward integral of*** g ***with respect to*** f (**on** $[a, b]$).
(ii) Suppose that the following limit

$$\int_{[a,b]} g(s)d^+ f(s) \;:=\; \lim_{\varepsilon\to 0^+} \int_{\mathbb{R}} g_J(s)\frac{f_{\overline{J}}(s) - f_{\overline{J}}(s-\varepsilon)}{\varepsilon}ds, \tag{6}$$

exists and it is finite. Then, the obtained quantity is denoted by $\int_{[a,b]} g d^+ f$ and called (**deterministic, definite**) **backward integral of** g **with respect to** f (**on** $[a, b]$).

The notation concerning this integral is justified since when the integrator f has bounded variation the previous integrals are Lebesgue-Stieltjes integrals on $[a,b]$.

Proposition 3 *Suppose* $f\colon [a,b] \to \mathbb{R}$ *with bounded variation and* $g\colon [a,b] \to \mathbb{R}$ *càdlàg. Then, we have*

$$\int_{[a,b]} g(s)d^- f(s) = \int_{[a,b]} g(s^-)df(s) \ := \ g(a)f(a) + \int_{]a,b]} g(s^-)df(s), \quad (7)$$

$$\int_{[a,b]} g(s)d^+ f(s) = \int_{[a,b]} g(s)df(s) \ := \ g(a)f(a) + \int_{]a,b]} g(s)df(s). \quad (8)$$

Proof *Identity* (7). We have

$$\int_{\mathbb{R}} g_J(s)\frac{f_{\overline{J}}(s+\varepsilon) - f_{\overline{J}}(s)}{\varepsilon}ds = \frac{1}{\varepsilon}g(a)\int_{a-\varepsilon}^{a} f(s+\varepsilon)ds + \int_a^b g(s)\frac{f((s+\varepsilon)\wedge b) - f(s)}{\varepsilon}ds. \quad (9)$$

The second integral on the right-hand side of (9) gives, by Fubini's theorem,

$$\int_a^b g(s)\bigg(\frac{1}{\varepsilon}\int_{]s,(s+\varepsilon)\wedge b]} df(y)\bigg)ds = \int_{]a,b]}\bigg(\frac{1}{\varepsilon}\int_{[a\vee(y-\varepsilon),y]} g(s)ds\bigg)df(y) \overset{\varepsilon\to 0^+}{\longrightarrow} \int_{]a,b]} g(y^-)df(y).$$

The first integral on the right-hand side of (9) goes to $g(a)f(a)$ as $\varepsilon \to 0^+$, so the result follows.

Identity (8). We have

$$\int_{\mathbb{R}} g_J(s)\frac{f_{\overline{J}}(s) - f_{\overline{J}}(s-\varepsilon)}{\varepsilon}ds = \int_{a+\varepsilon}^b g(s)\frac{f(s) - f(s-\varepsilon)}{\varepsilon}ds + \frac{1}{\varepsilon}\int_a^{a+\varepsilon} g(s)f(s)ds. \quad (10)$$

The second integral on the right-hand side of (10) goes to $g(a)f(a)$ as $\varepsilon \to 0^+$. The first one equals

$$\int_{a+\varepsilon}^b g(s)\bigg(\frac{1}{\varepsilon}\int_{]s-\varepsilon,s]} df(y)\bigg)ds \ = \ \int_{]a,b]}\bigg(\frac{1}{\varepsilon}\int_{]y,(y+\varepsilon)\wedge b]} g(s)ds\bigg)df(y) \overset{\varepsilon\to 0^+}{\longrightarrow} \int_{]a,b]} g(y)df(y),$$

from which the claim follows. □

Let us now introduce the deterministic covariation.

Definition 5 Let $f, g \colon [a, b] \to \mathbb{R}$ be continuous functions and suppose that $0 \in [a, b]$. The (**deterministic**) **covariation of** f **and** g (**on** $[a, b]$) is defined by

$$[f, g](x) = [g, f](x) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \int_0^x (f(s+\varepsilon) - f(s))(g(s+\varepsilon) - g(s))ds, \qquad x \in [a, b],$$

if the limit exists and it is finite for every $x \in [a, b]$. If $f = g$, we set $[f] := [f, f]$ and it is called **quadratic variation of** f (**on** $[a, b]$).

We notice that in Definition 5 the quadratic variation $[f]$ is continuous on $[a, b]$, since f is a continuous function.

Remark 1 Notice that if f is a fixed Brownian path and $g(s) = \varphi(s, f(s))$, with $\varphi \in C^1([a, b] \times \mathbb{R})$. Then $\int_{[a,b]} g(s)d^- f(s)$ exists for almost all (with respect to the Wiener measure on $C([a, b])$) Brownian paths f. This latter result can be shown using Theorem 2.1 in [26] (which implies that the deterministic bracket exists, for almost all Brownian paths f, and $[f](s) = s$) and then applying Itô's formula in Theorem 1 above, with $\mathbb{P}$ given by the Dirac delta at a Brownian path f. □

We conclude this subsection with an integration by parts formula for the deterministic forward and backward integrals.

Proposition 4 *Let $f \colon [a, b] \to \mathbb{R}$ be a càdlàg function and $g \colon [a, b] \to \mathbb{R}$ be a bounded variation function. Then, the following **integration by parts formulae** hold:*

$$\int_{[a,b]} g(s)d^- f(s) = g(b)f(b) - \int_{]a,b]} f(s)dg(s), \tag{11}$$

$$\int_{[a,b]} g(s)d^+ f(s) = g(b)f(b^-) - \int_{]a,b]} f(s^-)dg(s). \tag{12}$$

Proof Identity (11). The left-hand side of (11) is the limit, when $\varepsilon \to 0^+$, of

$$\frac{1}{\varepsilon}\int_a^{b-\varepsilon} g(s)f(s+\varepsilon)ds - \frac{1}{\varepsilon}\int_a^b g(s)f(s)ds + \frac{1}{\varepsilon}\int_{b-\varepsilon}^b g(s)f(b)ds + \frac{1}{\varepsilon}\int_{a-\varepsilon}^a g(a)f(s+\varepsilon)ds.$$

This gives

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_{a+\varepsilon}^{b} g(s-\varepsilon)f(s)ds - \frac{1}{\varepsilon}\int_a^b g(s)f(s)ds + \frac{1}{\varepsilon}\int_{b-\varepsilon}^b g(s)f(b)ds + \frac{1}{\varepsilon}\int_{a-\varepsilon}^a g(a)f(s+\varepsilon)ds \\
&= -\int_{a+\varepsilon}^b \frac{g(s) - g(s-\varepsilon)}{\varepsilon} f(s)ds - \frac{1}{\varepsilon}\int_a^{a+\varepsilon} g(s)f(s)ds + \frac{1}{\varepsilon}\int_{b-\varepsilon}^b g(s)f(b)ds \\
&\quad + \frac{1}{\varepsilon}\int_{a-\varepsilon}^a g(a)f(s+\varepsilon)ds.
\end{aligned}$$

We see that

$$\frac{1}{\varepsilon}\int_{b-\varepsilon}^{b} g(s)f(b)ds \overset{\varepsilon\to 0^+}{\longrightarrow} g(b^-)f(b),$$
$$\frac{1}{\varepsilon}\int_{a-\varepsilon}^{a} g(a)f(s+\varepsilon)ds - \frac{1}{\varepsilon}\int_{a}^{a+\varepsilon} g(s)f(s)ds \overset{\varepsilon\to 0^+}{\longrightarrow} 0.$$

Moreover, we have

$$\begin{aligned}-\int_{a+\varepsilon}^{b}\frac{g(s)-g(s-\varepsilon)}{\varepsilon}f(s)ds &= -\int_{a+\varepsilon}^{b} dsf(s)\frac{1}{\varepsilon}\int_{]s-\varepsilon,s]} dg(y)\\ &= -\int_{]a,b]} dg(y)\frac{1}{\varepsilon}\int_{y\vee(a+\varepsilon)}^{b\wedge(y+\varepsilon)} f(s)ds \overset{\varepsilon\to 0^+}{\longrightarrow} -\int_{]a,b[} dg(y)f(y).\end{aligned}$$

In conclusion, we find

$$\begin{aligned}\int_{[a,b]} g(s)d^-f(s) &= -\int_{]a,b]} dg(y)f(y) + (g(b)-g(b^-))f(b) + g(b^-)f(b)\\ &= -\int_{]a,b]} dg(y)f(y) + g(b)f(b).\end{aligned}$$

Identity (12). The left-hand side of (12) is given by the limit, as $\varepsilon \to 0^+$, of

$$\begin{aligned}\frac{1}{\varepsilon}\int_{a}^{b} g(s)f(s)ds - \frac{1}{\varepsilon}\int_{a+\varepsilon}^{b} g(s)f(s-\varepsilon)ds &= \frac{1}{\varepsilon}\int_{a}^{b} g(s)f(s)ds - \frac{1}{\varepsilon}\int_{a}^{b-\varepsilon} g(s+\varepsilon)f(s)ds\\ &= -\int_{a}^{b-\varepsilon} f(s)\frac{g(s+\varepsilon)-g(s)}{\varepsilon}ds + \frac{1}{\varepsilon}\int_{b-\varepsilon}^{b} g(s)f(s)ds\end{aligned}$$

The second integral on the right-hand side goes to $g(b^-)f(b^-)$ as $\varepsilon \to 0^+$. The first integral expression equals

$$\begin{aligned}&-\int_{\mathbb{R}} f_J(s)\frac{g_{\overline{J}}(s+\varepsilon)-g_{\overline{J}}(s)}{\varepsilon}ds + \frac{1}{\varepsilon}f(a)\int_{a-\varepsilon}^{a} g(s+\varepsilon)ds + \int_{b-\varepsilon}^{b} f(s)\frac{g(b)-g(s)}{\varepsilon}ds\\ &\overset{\varepsilon\to 0^+}{\longrightarrow} -\int_{]a,b]} f(s^-)dg(s) - f(a)g(a) + f(a)g(a) + (g(b)-g(b^-))f(b^-),\end{aligned}$$

taking into account identity (7). This gives us the result. □

2.2 *The Spaces* $\mathscr{C}([-T,0])$ *and* $\mathscr{C}([-T,0[)$

Let $C([-T,0])$ denote the set of real continuous functions on $[-T,0]$, endowed with supremum norm $\|\eta\|_\infty = \sup_{x\in[-T,0]}|\eta(x)|$, for any $\eta \in C([-T,0])$.

Remark 2 We shall develop functional Itô calculus via regularization firstly for time-independent functionals $\mathscr{U}: C([-T,0]) \to \mathbb{R}$, since we aim at emphasizing that in our framework the time variable and the path play two distinct roles, as emphasized in the introduction. This, also, allows us to focus only on the definition of horizontal and vertical derivatives. Clearly, everything can be extended in an obvious way to the time-dependent case $\mathscr{U}: [0,T] \times C([-T,0]) \to \mathbb{R}$, as we shall illustrate later. □

Consider a map $\mathscr{U}: C([-T,0]) \to \mathbb{R}$. Our aim is to derive a functional Itô's formula for $\mathscr{U}$. To do this, we are led to define the functional (i.e., horizontal and vertical) derivatives for $\mathscr{U}$ in the spirit of [5, 17]. Since the definition of functional derivatives necessitates of discontinuous paths, in [5] the idea is to consider functionals defined on the space of càdlàg trajectories $\mathbb{D}([-T,0])$. However, we can not, in general, extend in a unique way a functional $\mathscr{U}$ defined on $C([-T,0])$ to $\mathbb{D}([-T,0])$. Our idea, instead, is to consider a larger space than $C([-T,0])$, denoted by $\mathscr{C}([-T,0])$, which is the space of bounded trajectories on $[-T,0]$, continuous on $[-T,0[$ and with possibly a jump at 0. We endow $\mathscr{C}([-T,0])$ with a (inductive) topology such that $C([-T,0])$ is dense in $\mathscr{C}([-T,0])$ with respect to this topology. Therefore, if $\mathscr{U}$ is continuous with respect to the topology of $\mathscr{C}([-T,0])$, then it admits a unique continuous extension $u: \mathscr{C}([-T,0]) \to \mathbb{R}$.

Definition 6 We denote by $\mathscr{C}([-T,0])$ the set of bounded functions $\eta\colon [-T,0] \to \mathbb{R}$ such that η is continuous on $[-T,0[$, equipped with the topology we now describe.

Convergence We endow $\mathscr{C}([-T,0])$ with a topology inducing the following convergence: $(\eta_n)_n$ converges to η in $\mathscr{C}([-T,0])$ as n tends to infinity if the following holds.

(i) $\|\eta_n\|_\infty \leq C$, for any $n \in \mathbb{N}$, for some positive constant C independent of n;
(ii) $\sup_{x\in K} |\eta_n(x) - \eta(x)| \to 0$ as n tends to infinity, for any compact set $K \subset [-T,0[$;
(iii) $\eta_n(0) \to \eta(0)$ as n tends to infinity.

Topology For each compact $K \subset [-T,0[$ define the seminorm p_K on $\mathscr{C}([-T,0])$ by

$$p_K(\eta) \;=\; \sup_{x\in K} |\eta(x)| + |\eta(0)|, \qquad \forall\, \eta \in \mathscr{C}([-T,0]).$$

Let $M > 0$ and $\mathscr{C}_M([-T,0])$ be the set of functions in $\mathscr{C}([-T,0])$ which are bounded by M. Still denote p_K the restriction of p_K to $\mathscr{C}_M([-T,0])$ and consider the topology on $\mathscr{C}_M([-T,0])$ induced by the collection of seminorms $(p_K)_K$. Then, we endow $\mathscr{C}([-T,0])$ with the smallest topology (inductive topology) turning all the inclusions $i_M: \mathscr{C}_M([-T,0]) \to \mathscr{C}([-T,0])$ into continuous maps.

Remark 3 (i) Notice that $C([-T,0])$ is dense in $\mathscr{C}([-T,0])$, when endowed with the topology of $\mathscr{C}([-T,0])$. As a matter of fact, let $\eta \in \mathscr{C}([-T,0])$ and define, for any $n \in \mathbb{N}\backslash\{0\}$,

$$\varphi_n(x) = \begin{cases} \eta(x), & -T \leq x \leq -1/n, \\ n(\eta(0) - \eta(-1/n))x + \eta(0), & -1/n < x \leq 0. \end{cases}$$

Then, we see that $\varphi_n \in C([-T,0])$ and $\varphi_n \to \eta$ in $\mathscr{C}([-T,0])$.
Now, for any $a \in \mathbb{R}$ define

$$\begin{aligned} C_a([-T,0]) &:= \{\eta \in C([-T,0]) \colon \eta(0) = a\}, \\ \mathscr{C}_a([-T,0]) &:= \{\eta \in \mathscr{C}([-T,0]) \colon \eta(0) = a\}. \end{aligned}$$

Then, $C_a([-T,0])$ is dense in $\mathscr{C}_a([-T,0])$ with respect to the topology of $\mathscr{C}([-T,0])$.
(ii) We provide two examples of functionals $\mathscr{U} \colon C([-T,0]) \to \mathbb{R}$, continuous with respect to the topology of $\mathscr{C}([-T,0])$, and necessarily with respect to the topology of $C([-T,0])$ (the proof is straightforward and not reported):

(a) $\mathscr{U}(\eta) = g(\eta(t_1), \ldots, \eta(t_n))$, for all $\eta \in C([-T,0])$, with $-T \leq t_1 < \cdots < t_n \leq 0$ and $g \colon \mathbb{R}^n \to \mathbb{R}$ continuous.
(b) $\mathscr{U}(\eta) = \int_{[-T,0]} \varphi(x) d^- \eta(x)$, for all $\eta \in C([-T,0])$, with $\varphi \colon [0,T] \to \mathbb{R}$ a càdlàg bounded variation function. Concerning this example, keep in mind that, using the integration by parts formula, $\mathscr{U}(\eta)$ admits the representation (11).

(iii) Consider the functional $\mathscr{U}(\eta) = \sup_{x \in [-T,0]} \eta(x)$, for all $\eta \in C([-T,0])$. It is obviously continuous, but it is not continuous with respect to the topology of $\mathscr{C}([-T,0])$. As a matter of fact, for any $n \in \mathbb{N}$ consider $\eta_n \in C([-T,0])$ given by

$$\eta_n(x) = \begin{cases} 0, & -T \leq x \leq -\frac{T}{2^n}, \\ \frac{2^{n+1}}{T}x + 2, & -\frac{T}{2^n} < x \leq -\frac{T}{2^{n+1}}, \\ -\frac{2^{n+1}}{T}x, & -\frac{T}{2^{n+1}} < x \leq 0. \end{cases}$$

Then, $\mathscr{U}(\eta_n) = \sup_{x \in [-T,0]} \eta_n(x) = 1$, for any n. However, η_n converges to the zero function in $\mathscr{C}([-T,0])$, as n tends to infinity. This example will play an important role in Sect. 3 to justify a weaker notion of solution to the path-dependent semilinear Kolmogorov equation. □

To define the functional derivatives, we shall need to separate the "past" from the "present" of $\eta \in \mathscr{C}([-T,0])$. Indeed, roughly speaking, the horizontal derivative calls in the past values of η, namely $\{\eta(x) \colon x \in [-T,0[\}$, while the vertical derivative calls in the present value of η, namely $\eta(0)$. To this end, it is useful to introduce the space $\mathscr{C}([-T,0[)$.

Definition 7 We denote by $\mathscr{C}([-T, 0[)$ the set of bounded continuous functions $\gamma\colon [-T, 0[\to \mathbb{R}$, equipped with the topology we now describe.
Convergence We endow $\mathscr{C}([-T, 0[)$ with a topology inducing the following convergence: $(\gamma_n)_n$ converges to γ in $\mathscr{C}([-T, 0[)$ as n tends to infinity if:

(i) $\sup_{x\in[-T,0[} |\gamma_n(x)| \leq C$, for any $n \in \mathbb{N}$, for some positive constant C independent of n;
(ii) $\sup_{x\in K} |\gamma_n(x) - \gamma(x)| \to 0$ as n tends to infinity, for any compact set $K \subset [-T, 0[$.

Topology For each compact $K \subset [-T, 0[$ define the seminorm q_K on $\mathscr{C}([-T, 0[)$ by

$$q_K(\gamma) = \sup_{x\in K} |\gamma(x)|, \qquad \forall\, \gamma \in \mathscr{C}([-T, 0[).$$

Let $M > 0$ and $\mathscr{C}_M([-T, 0[)$ be the set of functions in $\mathscr{C}([-T, 0[)$ which are bounded by M. Still denote q_K the restriction of q_K to $\mathscr{C}_M([-T, 0[)$ and consider the topology on $\mathscr{C}_M([-T, 0[)$ induced by the collection of seminorms $(q_K)_K$. Then, we endow $\mathscr{C}([-T, 0[)$ with the smallest topology (inductive topology) turning all the inclusions $i_M\colon \mathscr{C}_M([-T, 0[) \to \mathscr{C}([-T, 0[)$ into continuous maps.

Remark 4 (i) Notice that $\mathscr{C}([-T, 0])$ is isomorphic to $\mathscr{C}([-T, 0[) \times \mathbb{R}$. As a matter of fact, it is enough to consider the map

$$\begin{aligned} J\colon \mathscr{C}([-T, 0]) &\to \mathscr{C}([-T, 0[) \times \mathbb{R} \\ \eta &\mapsto (\eta_{|[-T,0[}, \eta(0)). \end{aligned}$$

Observe that $J^{-1}\colon \mathscr{C}([-T, 0[) \times \mathbb{R} \to \mathscr{C}([-T, 0])$ is given by $J^{-1}(\gamma, a) = \gamma 1_{[-T,0[} + a 1_{\{0\}}$.
(ii) $\mathscr{C}([-T, 0])$ is a space which contains $C([-T, 0])$ as a dense subset and it has the property of separating "past" from "present". Another space having the same property is $L^2([-T, 0]; d\mu)$ where μ is the sum of the Dirac measure at zero and Lebesgue measure. Similarly as for item (i), that space is isomorphic to $L^2([-T, 0]) \times \mathbb{R}$, which is a very popular space appearing in the analysis of functional dependent (as delay) equations, starting from [4]. □

For every $u\colon \mathscr{C}([-T, 0]) \to \mathbb{R}$, we can now exploit the space $\mathscr{C}([-T, 0[)$ to define a map $\tilde{u}\colon \mathscr{C}([-T, 0[) \times \mathbb{R} \to \mathbb{R}$ where "past" and "present" are separated.

Definition 8 Let $u\colon \mathscr{C}([-T, 0]) \to \mathbb{R}$ and define $\tilde{u}\colon \mathscr{C}([-T, 0[) \times \mathbb{R} \to \mathbb{R}$ as

$$\tilde{u}(\gamma, a) := u(\gamma 1_{[-T,0[} + a 1_{\{0\}}), \qquad \forall\, (\gamma, a) \in \mathscr{C}([-T, 0[) \times \mathbb{R}. \tag{13}$$

In particular, we have $u(\eta) = \tilde{u}(\eta_{|[-T,0[}, \eta(0))$, for all $\eta \in \mathscr{C}([-T, 0])$.

We conclude this subsection with a characterization of the dual spaces of $\mathscr{C}([-T,0])$ and $\mathscr{C}([-T,0[)$, which has an independent interest. Firstly, we need to introduce the set $\mathscr{M}([-T,0])$ of finite signed Borel measures on $[-T,0]$. We also denote $\mathscr{M}_0([-T,0]) \subset \mathscr{M}([-T,0])$ the set of measures μ such that $\mu(\{0\}) = 0$.

Proposition 5 *Let $\Lambda \in \mathscr{C}([-T,0])^*$, the dual space of $\mathscr{C}([-T,0])$. Then, there exists a unique $\mu \in \mathscr{M}([-T,0])$ such that*

$$\Lambda\eta = \int_{[-T,0]} \eta(x)\mu(dx), \qquad \forall\, \eta \in \mathscr{C}([-T,0]).$$

Proof Let $\Lambda \in \mathscr{C}([-T,0])^*$ and define

$$\tilde{\Lambda}\varphi := \Lambda\varphi, \qquad \forall\, \varphi \in C([-T,0]).$$

Notice that $\tilde{\Lambda}\colon C([-T,0]) \to \mathbb{R}$ is a continuous functional on the Banach space $C([-T,0])$ endowed with the supremum norm $\|\cdot\|_\infty$. Therefore $\tilde{\Lambda} \in C([-T,0])^*$ and it follows from Riesz representation theorem (see, e.g., Theorem 6.19 in [36]) that there exists a unique $\mu \in \mathscr{M}([-T,0])$ such that

$$\tilde{\Lambda}\varphi = \int_{[-T,0]} \varphi(x)\mu(dx), \qquad \forall\, \varphi \in C([-T,0]).$$

Obviously $\tilde{\Lambda}$ is also continuous with respect to the topology of $\mathscr{C}([-T,0])$. Since $C([-T,0])$ is dense in $\mathscr{C}([-T,0])$ with respect to the topology of $\mathscr{C}([-T,0])$, we deduce that there exists a unique continuous extension of $\tilde{\Lambda}$ to $\mathscr{C}([-T,0])$, which is clearly given by

$$\Lambda\eta = \int_{[-T,0]} \eta(x)\mu(dx), \qquad \forall\, \eta \in \mathscr{C}([-T,0]).$$

□

Proposition 6 *Let $\Lambda \in \mathscr{C}([-T,0[)^*$, the dual space of $\mathscr{C}([-T,0[)$. Then, there exists a unique $\mu \in \mathscr{M}_0([-T,0])$ such that*

$$\Lambda\gamma = \int_{[-T,0[} \gamma(x)\mu(dx), \qquad \forall\, \gamma \in \mathscr{C}([-T,0[).$$

Proof Let $\Lambda \in \mathscr{C}([-T,0[)^*$ and define

$$\tilde{\Lambda}\eta := \Lambda(\eta_{|[-T,0[}), \qquad \forall\, \eta \in \mathscr{C}([-T,0]). \tag{14}$$

Notice that $\tilde{\Lambda}\colon \mathscr{C}([-T,0]) \to \mathbb{R}$ is a continuous functional on $\mathscr{C}([-T,0])$. It follows from Proposition 5 that there exists a unique $\mu \in \mathscr{M}([-T,0])$ such that

$$\tilde{\Lambda}\eta \;=\; \int_{[-T,0]} \eta(x)\mu(dx) \;=\; \int_{[-T,0[} \eta(x)\mu(dx) + \eta(0)\mu(\{0\}), \qquad \forall\, \eta \in \mathscr{C}([-T,0]). \tag{15}$$

Let $\eta_1, \eta_2 \in \mathscr{C}([-T,0])$ be such that $\eta_1 1_{[-T,0[} = \eta_2 1_{[-T,0[}$. Then, we see from (14) that $\tilde{\Lambda}\eta_1 = \tilde{\Lambda}\eta_2$, which in turn implies from (15) that $\mu(\{0\}) = 0$. In conclusion, $\mu \in \mathscr{M}_0([-T,0])$ and Λ is given by

$$\Lambda\gamma \;=\; \int_{[-T,0[} \gamma(x)\mu(dx), \qquad \forall\, \gamma \in \mathscr{C}([-T,0[).$$

2.3 Functional Derivatives and Functional Itô's Formula

In the present section we shall prove one of the main result of this section, namely the functional Itô's formula for $\mathscr{U}\colon C([-T,0]) \to \mathbb{R}$ and, more generally, for $\mathscr{U}\colon [0,T] \times C([-T,0]) \to \mathbb{R}$. We begin introducing the functional derivatives, firstly for a functional $u\colon \mathscr{C}([-T,0]) \to \mathbb{R}$, and then for $\mathscr{U}\colon C([-T,0]) \to \mathbb{R}$.

Definition 9 Consider $u\colon \mathscr{C}([-T,0]) \to \mathbb{R}$ and $\eta \in \mathscr{C}([-T,0])$.
(i) We say that u admits the **horizontal derivative** at η if the following limit exists and it is finite:

$$D^H u(\eta) \;:=\; \lim_{\varepsilon \to 0^+} \frac{u(\eta(\cdot)1_{[-T,0[} + \eta(0)1_{\{0\}}) - u(\eta(\cdot - \varepsilon)1_{[-T,0[} + \eta(0)1_{\{0\}})}{\varepsilon}. \tag{16}$$

(i)' Let $\tilde{u}$ be as in (13), then we say that $\tilde{u}$ admits the **horizontal derivative** at $(\gamma, a) \in \mathscr{C}([-T,0[) \times \mathbb{R}$ if the following limit exists and it is finite:

$$D^H \tilde{u}(\gamma, a) \;:=\; \lim_{\varepsilon \to 0^+} \frac{\tilde{u}(\gamma(\cdot), a) - \tilde{u}(\gamma(\cdot - \varepsilon), a)}{\varepsilon}. \tag{17}$$

Notice that if $D^H u(\eta)$ exists then $D^H \tilde{u}(\eta_{|[-T,0[}, \eta(0))$ exists and they are equal; viceversa, if $D^H \tilde{u}(\gamma, a)$ exists then $D^H u(\gamma 1_{[-T,0[} + a 1_{\{0\}})$ exists and they are equal.
(ii) We say that u admits the **first-order vertical derivative** at η if the first-order partial derivative $\partial_a \tilde{u}(\eta_{|[-T,0[}, \eta(0))$ at $(\eta_{|[-T,0[}, \eta(0))$ of $\tilde{u}$ with respect to its second argument exists and we set

$$D^V u(\eta) \;:=\; \partial_a \tilde{u}(\eta_{|[-T,0[}, \eta(0)).$$

(iii) We say that u admits the **second-order vertical derivative** at η if the second-order partial derivative at $(\eta_{|[-T,0[}, \eta(0))$ of $\tilde{u}$ with respect to its second argument, denoted by $\partial^2_{aa}\tilde{u}(\eta_{|[-T,0[}, \eta(0))$, exists and we set

$$D^{VV}u(\eta) \;:=\; \partial^2_{aa}\tilde{u}(\eta_{|[-T,0[}, \eta(0)).$$

Definition 10 We say that $u\colon \mathscr{C}([-T,0]) \to \mathbb{R}$ is of class $\mathscr{C}^{1,2}(\text{past} \times \text{present})$ if

(i) u is continuous;
(ii) $D^H u$ exists everywhere on $\mathscr{C}([-T,0])$ and for every $\gamma \in \mathscr{C}([-T,0[)$ the map

$$(\varepsilon, a) \longmapsto D^H\tilde{u}(\gamma(\cdot - \varepsilon), a), \qquad (\varepsilon, a) \in [0, \infty[\times\mathbb{R}$$

is continuous on $[0,\infty[\times\mathbb{R}$;
(iii) $D^V u$ and $D^{VV}u$ exist everywhere on $\mathscr{C}([-T,0])$ and are continuous.

Remark 5 Notice that in Definition 10 we still obtain the same class of functions $\mathscr{C}^{1,2}(\text{past} \times \text{present})$ if we substitute point (ii) with

(ii') $D^H u$ exists everywhere on $\mathscr{C}([-T,0])$ and for every $\gamma \in \mathscr{C}([-T,0[)$ there exists $\delta(\gamma) > 0$ such that the map

$$(\varepsilon, a) \longmapsto D^H\tilde{u}(\gamma(\cdot - \varepsilon), a), \qquad (\varepsilon, a) \in [0, \infty[\times\mathbb{R} \tag{18}$$

is continuous on $[0, \delta(\gamma)[\times\mathbb{R}$.

In particular, if (ii') holds then we can always take $\delta(\gamma) = \infty$ for any $\gamma \in \mathscr{C}([-T,0[)$, which implies (ii). To prove this last statement, let us proceed by contradiction assuming that

$$\delta^*(\gamma) \;=\; \sup\big\{\delta(\gamma) > 0\colon \text{the map (17) is continuous on } [0,\delta(\gamma)[\times\mathbb{R}\big\} \;<\; \infty.$$

Notice that $\delta^*(\gamma)$ is in fact a *max*, therefore the map (18) is continuous on $[0,\delta^*(\gamma)[\times\mathbb{R}$. Now, define $\bar{\gamma}(\cdot) := \gamma(\cdot - \delta^*(\gamma))$. Then, by condition (ii') there exists $\delta(\bar{\gamma}) > 0$ such that the map

$$(\varepsilon, a) \longmapsto D^H\tilde{u}(\bar{\gamma}(\cdot - \varepsilon), a) = D^H\tilde{u}(\gamma(\cdot - \varepsilon - \delta^*(\gamma)), a)$$

is continuous on $[0,\delta(\bar{\gamma})[\times\mathbb{R}$. This shows that the map (18) is continuous on $[0,\delta^*(\gamma)+\delta(\bar{\gamma})[\times\mathbb{R}$, a contradiction with the definition of $\delta^*(\gamma)$. □

We can now provide the definition of functional derivatives for a map $\mathscr{U}\colon\; C([-T,0]) \to \mathbb{R}$.

Definition 11 Let $\mathscr{U}\colon C([-T,0]) \to \mathbb{R}$ and $\eta \in C([-T,0])$. Suppose that there exists a unique extension $u\colon \mathscr{C}([-T,0]) \to \mathbb{R}$ of $\mathscr{U}$ (e.g., if $\mathscr{U}$ is continuous with respect to the topology of $\mathscr{C}([-T,0])$). Then we define the following concepts.

(i) The **horizontal derivative** of $\mathscr{U}$ at η as:

$$D^H \mathscr{U}(\eta) \;:=\; D^H u(\eta).$$

(ii) The **first-order vertical derivative** of $\mathscr{U}$ at η as:

$$D^V \mathscr{U}(\eta) \;:=\; D^V u(\eta).$$

(iii) The **second-order vertical derivative** of $\mathscr{U}$ at η as:

$$D^{VV} \mathscr{U}(\eta) \;:=\; D^{VV} u(\eta).$$

Definition 12 We say that $\mathscr{U}: C([-T, 0]) \to \mathbb{R}$ is $C^{1,2}$(past$\times$present) if $\mathscr{U}$ admits a (necessarily unique) extension $u: \mathscr{C}([-T, 0]) \to \mathbb{R}$ of class $\mathscr{C}^{1,2}$(past $\times$ present).

Theorem 2 *Let $\mathscr{U}: C([-T, 0]) \to \mathbb{R}$ be of class $C^{1,2}$(past $\times$ present) and $X = (X_t)_{t\in[0,T]}$ be a real continuous finite quadratic variation process. Then, the following* ***functional Itô's formula*** *holds, $\mathbb{P}$-a.s.,*

$$\mathscr{U}(\mathbb{X}_t) \;=\; \mathscr{U}(\mathbb{X}_0) + \int_0^t D^H \mathscr{U}(\mathbb{X}_s)ds + \int_0^t D^V \mathscr{U}(\mathbb{X}_s)d^- X_s + \frac{1}{2}\int_0^t D^{VV} \mathscr{U}(\mathbb{X}_s)d[X]_s, \tag{19}$$

for all $0 \le t \le T$, where the window process $\mathbb{X}$ was defined in (2).

Proof Fix $t \in [0, T]$ and consider the quantity

$$I_0(\varepsilon, t) \;=\; \int_0^t \frac{\mathscr{U}(\mathbb{X}_{s+\varepsilon}) - \mathscr{U}(\mathbb{X}_s)}{\varepsilon} ds \;=\; \frac{1}{\varepsilon}\int_t^{t+\varepsilon} \mathscr{U}(\mathbb{X}_s)ds - \frac{1}{\varepsilon}\int_0^{\varepsilon} \mathscr{U}(\mathbb{X}_s)ds, \qquad \varepsilon > 0.$$

Since the process $(\mathscr{U}(\mathbb{X}_s))_{s\ge 0}$ is continuous, $I_0(\varepsilon, t)$ converges ucp to $\mathscr{U}(\mathbb{X}_t) - \mathscr{U}(\mathbb{X}_0)$, namely $\sup_{0\le t\le T} |I_0(\varepsilon, t) - (\mathscr{U}(\mathbb{X}_t) - \mathscr{U}(\mathbb{X}_0))|$ converges to zero in probability when $\varepsilon \to 0^+$. On the other hand, we can write $I_0(\varepsilon, t)$ in terms of the function $\tilde{u}$, defined in (13), as follows

$$I_0(\varepsilon, t) \;=\; \int_0^t \frac{\tilde{u}(\mathbb{X}_{s+\varepsilon|[-T,0[}, X_{s+\varepsilon}) - \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s)}{\varepsilon} ds.$$

Now we split $I_0(\varepsilon, t)$ into the sum of two terms

$$I_1(\varepsilon, t) = \int_0^t \frac{\tilde{u}(\mathbb{X}_{s+\varepsilon|[-T,0[}, X_{s+\varepsilon}) - \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_{s+\varepsilon})}{\varepsilon} ds, \tag{20}$$

$$I_2(\varepsilon, t) = \int_0^t \frac{\tilde{u}(\mathbb{X}_{s|[-T,0[}, X_{s+\varepsilon}) - \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s)}{\varepsilon} ds. \tag{21}$$

We begin proving that

$$I_1(\varepsilon, t) \xrightarrow[\varepsilon \to 0^+]{\text{ucp}} \int_0^t D^H \mathscr{U}(\mathbb{X}_s) ds. \tag{22}$$

Firstly, fix $\gamma \in \mathscr{C}([-T, 0[)$ and define

$$\phi(\varepsilon, a) \;:=\; \tilde{u}(\gamma(\cdot - \varepsilon), a), \qquad (\varepsilon, a) \in [0, \infty[\times \mathbb{R}.$$

Then, denoting by $\partial_\varepsilon^+ \phi$ the right partial derivative of ϕ with respect to ε and using formula (17), we find

$$\begin{aligned}
\partial_\varepsilon^+ \phi(\varepsilon, a) &= \lim_{r \to 0^+} \frac{\phi(\varepsilon + r, a) - \phi(\varepsilon, a)}{r} \\
&= -\lim_{r \to 0^+} \frac{\tilde{u}(\gamma(\cdot - \varepsilon), a) - \tilde{u}(\gamma(\cdot - \varepsilon - r), a)}{r} \\
&= -D^H \tilde{u}(\gamma(\cdot - \varepsilon), a), \qquad \forall\, (\varepsilon, a) \in [0, \infty[\times \mathbb{R}.
\end{aligned}$$

Since $u \in \mathscr{C}^{1,2}(\text{past} \times \text{present})$, we see from Definition 10(ii), that $\partial_\varepsilon^+ \phi$ is continuous on $[0, \infty[\times \mathbb{R}$. It follows from a standard differential calculus result (see for example Corollary 1.2, Chap. 2, in [32]) that ϕ is continuously differentiable on $[0, \infty[\times \mathbb{R}$ with respect to its first argument. Then, for every $(\varepsilon, a) \in [0, \infty[\times \mathbb{R}$, from the fundamental theorem of calculus, we have

$$\phi(\varepsilon, a) - \phi(0, a) \;=\; \int_0^\varepsilon \partial_\varepsilon \phi(r, a) dr,$$

which in terms of $\tilde{u}$ reads

$$\tilde{u}(\gamma(\cdot), a) - \tilde{u}(\gamma(\cdot - \varepsilon), a) \;=\; \int_0^\varepsilon D^H \tilde{u}(\gamma(\cdot - r), a) dr. \tag{23}$$

Now, we rewrite, by means of a shift in time, the term $I_1(\varepsilon, t)$ in (20) as follows:

$$\begin{aligned}
I_1(\varepsilon, t) = &\int_0^t \frac{\tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s) - \tilde{u}(\mathbb{X}_{s-\varepsilon|[-T,0[}, X_s)}{\varepsilon} ds \\
&+ \int_t^{t+\varepsilon} \frac{\tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s) - \tilde{u}(\mathbb{X}_{s-\varepsilon|[-T,0[}, X_s)}{\varepsilon} ds \\
&- \int_0^\varepsilon \frac{\tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s) - \tilde{u}(\mathbb{X}_{s-\varepsilon|[-T,0[}, X_s)}{\varepsilon} ds.
\end{aligned} \tag{24}$$

Plugging (23) into (24), setting $\gamma = \mathbb{X}_s$, $a = X_s$, we obtain

$$
\begin{aligned}
I_1(\varepsilon, t) = & \int_0^t \frac{1}{\varepsilon}\bigg(\int_0^\varepsilon D^H \tilde{u}(\mathbb{X}_{s-r|[-T,0[}, X_s)dr\bigg)ds \\
& + \int_t^{t+\varepsilon} \frac{1}{\varepsilon}\bigg(\int_0^\varepsilon D^H \tilde{u}(\mathbb{X}_{s-r|[-T,0[}, X_s)dr\bigg)ds \\
& - \int_0^\varepsilon \frac{1}{\varepsilon}\bigg(\int_0^\varepsilon D^H \tilde{u}(\mathbb{X}_{s-r|[-T,0[}, X_s)dr\bigg)ds.
\end{aligned} \tag{25}
$$

Observe that

$$
\int_0^t \frac{1}{\varepsilon}\bigg(\int_0^\varepsilon D^H \tilde{u}(\mathbb{X}_{s-r|[-T,0[}, X_s)dr\bigg)ds \xrightarrow[\varepsilon\to 0^+]{\text{ucp}} \int_0^t D^H u(\mathbb{X}_s)ds.
$$

Similarly, we see that the other two terms in (25) converge ucp to zero. As a consequence, we get (22).

Regarding $I_2(\varepsilon, t)$ in (21), it can be written, by means of the following standard Taylor's expansion for a function $f \in C^2(\mathbb{R})$:

$$
\begin{aligned}
f(b) = & \ f(a) + f'(a)(b-a) + \frac{1}{2}f''(a)(b-a)^2 \\
& + \int_0^1 (1-\alpha)\big(f''(a+\alpha(b-a)) - f''(a)\big)(b-a)^2 d\alpha,
\end{aligned}
$$

as the sum of the following three terms:

$$
\begin{aligned}
I_{21}(\varepsilon, t) = & \int_0^t \partial_a \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s)\frac{X_{s+\varepsilon} - X_s}{\varepsilon}ds \\
I_{22}(\varepsilon, t) = & \ \frac{1}{2}\int_0^t \partial^2_{aa} \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s)\frac{(X_{s+\varepsilon} - X_s)^2}{\varepsilon}ds \\
I_{23}(\varepsilon, t) = & \int_0^t \bigg(\int_0^1 (1-\alpha)\big(\partial^2_{aa}\tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s + \alpha(X_{s+\varepsilon} - X_s)) \\
& - \partial^2_{aa}\tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s)\big)\frac{(X_{s+\varepsilon} - X_s)^2}{\varepsilon}d\alpha\bigg)ds.
\end{aligned}
$$

By similar arguments as in Proposition 1.2 of [39], we have

$$
I_{22}(\varepsilon, t) \xrightarrow[\varepsilon\to 0^+]{\text{ucp}} \frac{1}{2}\int_0^t \partial^2_{aa}\tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s)d[X]_s = \frac{1}{2}\int_0^t D^{VV}u(\mathbb{X}_s)d[X]_s.
$$

Regarding $I_{23}(\varepsilon, t)$, for every $\omega \in \Omega$, define $\psi_\omega\colon [0,T]\times[0,1]\times[0,1] \to \mathbb{R}$ as

$$
\psi_\omega(s, \alpha, \varepsilon) \ := \ (1-\alpha)\partial^2_{aa}\tilde{u}\big(\mathbb{X}_{s|[-T,0[}(\omega), X_s(\omega) + \alpha(X_{s+\varepsilon}(\omega) - X_s(\omega))\big),
$$

for all $(s, \alpha, \varepsilon) \in [0, T] \times [0, 1] \times [0, 1]$. Notice that ψ_ω is uniformly continuous. Denote ρ_{ψ_ω} its continuity modulus, then

$$\sup_{t\in[0,T]} |I_{23}(\varepsilon, t)| \leq \int_0^T \rho_{\psi_\omega}(\varepsilon)\frac{(X_{s+\varepsilon} - X_s)^2}{\varepsilon}ds.$$

Since X has finite quadratic variation, we deduce that $I_{23}(\varepsilon, t) \to 0$ ucp as $\varepsilon \to 0^+$. Finally, because of $I_0(\varepsilon, t)$, $I_1(\varepsilon, t)$, $I_{22}(\varepsilon, t)$, and $I_{23}(\varepsilon, t)$ converge ucp, it follows that the forward integral exists:

$$I_{21}(\varepsilon, t) \xrightarrow[\varepsilon\to 0^+]{\text{ucp}} \int_0^t \partial_a \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_s)d^- X_s = \int_0^t D^V u(\mathbb{X}_s)d^- X_s,$$

from which the claim follows.

Remark 6 Notice that, under the hypotheses of Theorem 2, the forward integral $\int_0^t D^V \mathscr{U}(\mathbb{X}_s)d^- X_s$ exists as a ucp limit, which is generally not required. □

Remark 7 The definition of horizontal derivative. Notice that our definition of horizontal derivative differs from that introduced in [17], since it is based on a limit on the left, while the definition proposed in [17] would conduct to the formula

$$D^{H,+}u(\eta) := \lim_{\varepsilon\to 0^+} \frac{\tilde{u}(\eta(\cdot + \varepsilon)1_{[-T,0[}, \eta(0)) - \tilde{u}(\eta(\cdot)1_{[-T,0[}, \eta(0))}{\varepsilon}. \tag{26}$$

To give an insight into the difference between (16) and (26), let us consider a real continuous finite quadratic variation process X with associated window process $\mathbb{X}$. Then, in the definition (26) of $D^{H,+}u(\mathbb{X}_t)$ we consider the increment $\tilde{u}(\mathbb{X}_{t|[-T,0[}(\cdot + \varepsilon), X_t) - \tilde{u}(\mathbb{X}_{t|[-T,0[}, X_t)$, comparing the present value of $u(\mathbb{X}_t) = \tilde{u}(\mathbb{X}_{t|[-T,0[}, X_t)$ with an hypothetical future value $\tilde{u}(\mathbb{X}_{t|[-T,0[}(\cdot + \varepsilon), X_t)$, obtained assuming a constant time evolution for X. On the other hand, in our definition (16) we consider the increment $\tilde{u}(\mathbb{X}_{t|[-T,0[}, X_t) - \tilde{u}(\mathbb{X}_{t-\varepsilon|[-T,0[}, X_t)$, where only the present and past values of X are taken into account, and where we also extend in a constant way the trajectory of X before time 0. In particular, unlike (26), since we do not call in the future in our formula (16), we do not have to specify a future time evolution for X, but only a past evolution before time 0. This difference between (16) and (26) is crucial for the proof of the functional Itô's formula. In particular, the adoption of (26) as definition for the horizontal derivative would require an additional regularity condition on u in order to prove an Itô formula for the process $t \mapsto u(\mathbb{X}_t)$. Indeed, as it can be seen from the proof of Theorem 2, to prove Itô's formula we are led to consider the term

$$I_1(\varepsilon, t) = \int_0^t \frac{\tilde{u}(\mathbb{X}_{s+\varepsilon|[-T,0[}, X_{s+\varepsilon}) - \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_{s+\varepsilon})}{\varepsilon}ds.$$

When adopting definition (26) it is convenient to write $I_1(\varepsilon, t)$ as the sum of the two integrals

$$I_{11}(\varepsilon, t) = \int_0^t \frac{\tilde{u}(\mathbb{X}_{s+\varepsilon|[-T,0[}, X_{s+\varepsilon}) - \tilde{u}(\mathbb{X}_{s|[-T,0[}(\cdot + \varepsilon), X_{s+\varepsilon})}{\varepsilon} ds,$$
$$I_{12}(\varepsilon, t) = \int_0^t \frac{\tilde{u}(\mathbb{X}_{s|[-T,0[}(\cdot + \varepsilon), X_{s+\varepsilon}) - \tilde{u}(\mathbb{X}_{s|[-T,0[}, X_{s+\varepsilon})}{\varepsilon} ds.$$

It can be shown quite easily that, under suitable regularity conditions on u (more precisely, if u is continuous, $D^{H,+}u$ exists everywhere on $\mathscr{C}([-T,0])$, and for every $\gamma \in \mathscr{C}([-T,0[)$ the map $(\varepsilon, a) \longmapsto D^{H,+}\tilde{u}(\gamma(\cdot + \varepsilon), a)$ is continuous on $[0,\infty) \times \mathbb{R}[$, we have

$$I_{12}(\varepsilon, t) \underset{\varepsilon \to 0^+}{\overset{\text{ucp}}{\longrightarrow}} \int_0^t D^{H,+}u(\mathbb{X}_s) ds.$$

To conclude the proof of Itô's formula along the same lines as in Theorem 2, we should prove

$$I_{11}(\varepsilon, t) \underset{\varepsilon \to 0^+}{\overset{\text{ucp}}{\longrightarrow}} 0. \tag{27}$$

In order to guarantee (27), we need to impose some additional regularity condition on $\tilde{u}$, and hence on u. As an example, (27) is satisfied if we assume the following condition on $\tilde{u}$: there exists a constant $C > 0$ such that, for every $\varepsilon > 0$,

$$|\tilde{u}(\gamma_1, a) - \tilde{u}(\gamma_2, a)| \ \le \ C\varepsilon \sup_{x \in [-\varepsilon, 0[} |\gamma_1(x) - \gamma_2(x)|,$$

for all $\gamma_1, \gamma_2 \in \mathscr{C}([-T,0[)$ and $a \in \mathbb{R}$, with $\gamma_1(x) = \gamma_2(x)$ for any $x \in [-T, -\varepsilon]$. This last condition is verified if, for example, $\tilde{u}$ is uniformly Lipschitz continuous with respect to the $L^1([-T,0])$-norm on $\mathscr{C}([-T,0[)$, namely: there exists a constant $C > 0$ such that

$$|\tilde{u}(\gamma_1, a) - \tilde{u}(\gamma_2, a)| \ \le \ C \int_{[-T,0[} |\gamma_1(x) - \gamma_2(x)| dx,$$

for all $\gamma_1, \gamma_2 \in \mathscr{C}([-T,0[)$ and $a \in \mathbb{R}$. □

We conclude this subsection providing the functional Itô's formula for a map $\mathscr{U} : [0,T] \times C([-T,0]) \to \mathbb{R}$ depending also on the time variable. Firstly, we notice that for a map $\mathscr{U} : [0,T] \times C([-T,0]) \to \mathbb{R}$ (resp. $u : [0,T] \times \mathscr{C}([-T,0]) \to \mathbb{R}$) the functional derivatives $D^H\mathscr{U}$, $D^V\mathscr{U}$, and $D^{VV}\mathscr{U}$ (resp. $D^H u$, $D^V u$, and $D^{VV}u$) are defined in an obvious way as in Definition 11 (resp. Definition 9). Moreover, given $u : [0,T] \times \mathscr{C}([-T,0]) \to \mathbb{R}$ we can define, as in Definition 8, a map $\tilde{u} : [0,T] \times \mathscr{C}([-T,0[) \times \mathbb{R} \to \mathbb{R}$. Then, we can give the following definitions.

Definition 13 Let I be $[0,T[$ or $[0,T]$. We say that $u\colon I \times \mathscr{C}([-T,0]) \to \mathbb{R}$ is of class $\mathscr{C}^{1,2}((I \times \text{past}) \times \text{present})$ if the properties below hold.

(i) u is continuous;
(ii) $\partial_t u$ exists everywhere on $I \times \mathscr{C}([-T,0])$ and is continuous;
(iii) $D^H u$ exists everywhere on $I \times \mathscr{C}([-T,0])$ and for every $\gamma \in \mathscr{C}([-T,0[)$ the map

$$(t,\varepsilon,a) \longmapsto D^H \tilde{u}(t,\gamma(\cdot-\varepsilon),a), \qquad (t,\varepsilon,a) \in I \times [0,\infty[\times\mathbb{R}$$

is continuous on $I \times [0,\infty[\times\mathbb{R}$;
(iv) $D^V u$ and $D^{VV} u$ exist everywhere on $I \times \mathscr{C}([-T,0])$ and are continuous.

Definition 14 Let I be $[0,T[$ or $[0,T]$. We say that $\mathscr{U}\colon I \times C([-T,0]) \to \mathbb{R}$ is $C^{1,2}((I \times \text{past}) \times \text{present})$ if $\mathscr{U}$ admits a (necessarily unique) extension $u\colon I \times \mathscr{C}([-T,0]) \to \mathbb{R}$ of class $\mathscr{C}^{1,2}((I \times \text{past}) \times \text{present})$.

We can now state the functional Itô's formula, whose proof is not reported, since it can be done along the same lines as Theorem 2.

Theorem 3 *Let $\mathscr{U}\colon [0,T] \times C([-T,0]) \to \mathbb{R}$ be of class $C^{1,2}(([0,T] \times \textit{past}) \times$ present) and $X = (X_t)_{t\in[0,T]}$ be a real continuous finite quadratic variation process. Then, the following* ***functional Itô's formula*** *holds, $\mathbb{P}$-a.s.,*

$$\begin{aligned}\mathscr{U}(t,\mathbb{X}_t) = {} & \mathscr{U}(0,\mathbb{X}_0) + \int_0^t \big(\partial_t \mathscr{U}(s,\mathbb{X}_s) + D^H \mathscr{U}(s,\mathbb{X}_s)\big)ds + \int_0^t D^V \mathscr{U}(s,\mathbb{X}_s)d^- X_s \\ & + \frac{1}{2}\int_0^t D^{VV}\mathscr{U}(s,\mathbb{X}_s)d[X]_s, \end{aligned} \tag{28}$$

for all $0 \le t \le T$.

Remark 8 Notice that, as a particular case, choosing $\mathscr{U}(t,\eta) = F(t,\eta(0))$, for any $(t,\eta) \in [0,T]\times C([-T,0])$, with $F \in C^{1,2}([0,T]\times\mathbb{R})$, we retrieve the classical Itô formula for finite quadratic variation processes, i.e. (4). More precisely, in this case $\mathscr{U}$ admits as unique continuous extension the map $u\colon [0,T] \times \mathscr{C}([-T,0]) \to \mathbb{R}$ given by $u(t,\eta) = F(t,\eta(0))$, for all $(t,\eta) \in [0,T] \times \mathscr{C}([-T,0])$. Moreover, we see that $D^H\mathscr{U} \equiv 0$, while $D^V\mathscr{U} = \partial_x F$ and $D^{VV}\mathscr{U} = \partial^2_{xx}F$, where $\partial_x F$ (resp. $\partial^2_{xx}F$) denotes the first-order (resp. second-order) partial derivative of F with respect to its second argument. □

2.4 Comparison with Banach Space Valued Calculus via Regularization

In the present subsection our aim is to make a link between functional Itô calculus, as derived in this paper, and Banach space valued stochastic calculus via regularization for window processes, which has been conceived in [13], see also [12, 14–16] for more recent developments. More precisely, our purpose is to identify the building blocks of our functional Itô's formula (19) with the terms appearing in the Itô formula derived in Theorem 6.3 and Sect. 7.2 in [12]. While it is expected that the vertical derivative $D^V \mathscr{U}$ can be identified with the term $D^{\delta_0}_{dx} \mathscr{U}$ of the Fréchet derivative, it is more difficult to guess to which terms the horizontal derivative $D^H \mathscr{U}$ corresponds. To clarify this latter point, in this subsection we derive two formulae which express $D^H \mathscr{U}$ in terms of Fréchet derivatives of $\mathscr{U}$.

Let us introduce some useful notations. We denote by $BV([-T, 0])$ the set of càdlàg bounded variation functions on $[-T, 0]$, which is a Banach space when equipped with the norm

$$\|\eta\|_{BV([-T,0])} := |\eta(0)| + \|\eta\|_{\mathrm{Var}([-T,0])}, \qquad \eta \in BV([-T, 0]),$$

where $\|\eta\|_{\mathrm{Var}([-T,0])} = |d\eta|([-T, 0])$ and $|d\eta|$ is the total variation measure associated to the measure $d\eta \in \mathscr{M}([-T, 0])$ generated by η: $d\eta(]-T, -t]) = \eta(-t) - \eta(-T)$, $t \in [-T, 0]$. We recall from Sect. 2.1 that we extend $\eta \in BV([-T, 0])$ to all $x \in \mathbb{R}$ setting $\eta(x) = 0$, $x < -T$, and $\eta(x) = \eta(0)$, $x \geq 0$. Let us now introduce some useful facts about tensor products of Banach spaces.

Definition 15 Let $(E, \|\cdot\|_E)$ and $(F, \|\cdot\|_F)$ be two Banach spaces.
(i) We shall denote by $E \otimes F$ the **algebraic tensor product** of E and F, defined as the set of elements of the form $v = \sum_{i=1}^n e_i \otimes f_i$, for some positive integer n, where $e \in E$ and $f \in F$. The map $\otimes\colon E \times F \to E \otimes F$ is bilinear.
(ii) We endow $E \otimes F$ with the **projective norm** π:

$$\pi(v) := \inf\left\{ \sum_{i=1}^n \|e_i\|_E \|f_i\|_F \,:\, v = \sum_{i=1}^n e_i \otimes f_i \right\}, \qquad \forall\, v \in E \otimes F.$$

(iii) We denote by $E\hat{\otimes}_\pi F$ the Banach space obtained as the completion of $E \otimes F$ for the norm π. We shall refer to $E\hat{\otimes}_\pi F$ as the **tensor product of the Banach spaces E and F**.
(iv) If E and F are Hilbert spaces, we denote $E\hat{\otimes}_h F$ the **Hilbert tensor product**, which is still a Hilbert space obtained as the completion of $E \otimes F$ for the scalar product $\langle e' \otimes f', e'' \otimes f'' \rangle := \langle e', e'' \rangle_E \langle f', f'' \rangle_F$, for any $e', e'' \in E$ and $f', f'' \in F$.
(v) The symbols $E\hat{\otimes}_\pi^2$ and $e\otimes^2$ denote, respectively, the Banach space $E\hat{\otimes}_\pi E$ and the element $e \otimes e$ of the algebraic tensor product $E \otimes E$.

Remark 9 (i) The projective norm π belongs to the class of the so-called *reasonable crossnorms* α on $E \otimes F$, verifying $\alpha(e \otimes f) = \|e\|_E \|f\|_F$.
(ii) We notice, proceeding for example as in [16] (see, in particular, formula (2.1) in [16]; for more information on this subject we refer to [41]), that the dual $(E\hat{\otimes}_\pi F)^*$ of

$E\hat{\otimes}_\pi F$ is isomorphic to the space of continuous bilinear forms $\mathscr{B}i(E, F)$, equipped with the norm $\|\cdot\|_{E,F}$ defined as

$$\|\Phi\|_{E,F} \;:=\; \sup_{\substack{e\in E,\, f\in F\\ \|e\|_E, \|f\|_F\leq 1}} |\Phi(e, f)|, \qquad \forall\, \Phi \in \mathscr{B}i(E, F).$$

Moreover, there exists a canonical isomorphism between $\mathscr{B}i(E, F)$ and $L(E, F^*)$, the space of bounded linear operators from E into F^*. Hence, we have the following chain of canonical identifications: $(E\hat{\otimes}_\pi F)^* \cong \mathscr{B}i(E, F) \cong L(E; F^*)$. □

Definition 16 Let E be a Banach space. We say that $\mathscr{U}: E \to \mathbb{R}$ is of class $C^2(E)$ if

(i) $D\mathscr{U}$, the first Fréchet derivative of $\mathscr{U}$, belongs to $C(E; E^*)$ and
(ii) $D^2\mathscr{U}$, the second Fréchet derivative of $\mathscr{U}$, belongs to $C(E; L(E; E^*))$.

Remark 10 Take $E = C([-T, 0])$ in Definition 16.
(i) *First Fréchet derivative $D\mathscr{U}$.* We have

$$D\mathscr{U}: C([-T, 0]) \;\longrightarrow\; (C([-T, 0]))^* \cong \mathscr{M}([-T, 0]).$$

For every $\eta \in C([-T, 0])$, we shall denote $D_{dx}\mathscr{U}(\eta)$ the unique measure in $\mathscr{M}([-T, 0])$ such that

$$D\mathscr{U}(\eta)\varphi \;=\; \int_{[-T,0]} \varphi(x) D_{dx}\mathscr{U}(\eta), \qquad \forall\, \varphi \in C([-T, 0]).$$

Notice that $\mathscr{M}([-T, 0])$ can be represented as the direct sum: $\mathscr{M}([-T, 0]) = \mathscr{M}_0([-T, 0]) \oplus \mathscr{D}_0$, where we recall that $\mathscr{M}_0([-T, 0])$ is the subset of $\mathscr{M}([-T, 0])$ of measures μ such that $\mu(\{0\}) = 0$, instead $\mathscr{D}_0$ (which is a shorthand for $\mathscr{D}_0([-T, 0])$) denotes the one-dimensional space of measures which are multiples of the Dirac measure δ_0. For every $\eta \in C([-T, 0])$ we denote by $(D_{dx}^\perp \mathscr{U}(\eta), D_{dx}^{\delta_0}\mathscr{U}(\eta))$ the unique pair in $\mathscr{M}_0([-T, 0]) \oplus \mathscr{D}_0$ such that

$$D_{dx}\mathscr{U}(\eta) \;=\; D_{dx}^\perp \mathscr{U}(\eta) + D_{dx}^{\delta_0}\mathscr{U}(\eta).$$

(ii) *Second Fréchet derivative $D^2\mathscr{U}$.* We have

$$\begin{aligned} D^2\mathscr{U}: C([-T, 0]) \;\longrightarrow\; L(C([-T, 0]); (C([-T, 0]))^*) &\cong \mathscr{B}i(C([-T, 0]), C([-T, 0])) \\ &\cong (C([-T, 0])\hat{\otimes}_\pi C([-T, 0]))^*, \end{aligned}$$

where we used the identifications of Remark 9(ii). Let $\eta \in C([-T, 0])$; a typical situation arises when there exists $D_{dx\,dy}\mathscr{U}(\eta) \in \mathscr{M}([-T, 0]^2)$ such that $D^2\mathscr{U}(\eta) \in L(C([-T, 0]); (C([-T, 0]))^*)$ admits the representation

$$D^2\mathscr{U}(\eta)(\varphi,\psi) = \int_{[-T,0]^2} \varphi(x)\psi(y) D_{dx\,dy}\mathscr{U}(\eta), \qquad \forall \varphi,\psi \in C([-T,0]).$$

Moreover, $D_{dx\,dy}\mathscr{U}(\eta)$ is uniquely determined. □

The definition below was given in [13].

Definition 17 Let E be a Banach space. A Banach subspace $(\chi, \|\cdot\|_\chi)$ continuously injected into $(E\hat{\otimes}_\pi^2)^*$, i.e., $\|\cdot\|_\chi \geq \|\cdot\|_{(E\hat{\otimes}_\pi^2)^*}$, will be called a **Chi-subspace** (of $(E\hat{\otimes}_\pi^2)^*$).

Remark 11 Take $E = C([-T,0])$ in Definition 17. As indicated in [13], a typical example of Chi-subspace of $C([-T,0])\hat{\otimes}_\pi^2$ is $\mathscr{M}([-T,0]^2)$ equipped with the usual total variation norm, denoted by $\|\cdot\|_{\mathrm{Var}}$. Another important Chi-subspace of $C([-T,0])\hat{\otimes}_\pi^2$ is the following, which is also a Chi-subspace of $\mathscr{M}([-T,0]^2)$:

$$\begin{aligned}\chi_0 := \big\{&\mu \in \mathscr{M}([-T,0]^2)\colon \mu(dx,dy) = g_1(x,y)dxdy + \lambda_1\delta_0(dx)\otimes\delta_0(dy)\\ &+ g_2(x)dx\otimes\lambda_2\delta_0(dy) + \lambda_3\delta_0(dx)\otimes g_3(y)dy + g_4(x)\delta_y(dx)\otimes dy,\\ &g_1 \in L^2([-T,0]^2),\ g_2,g_3 \in L^2([-T,0]),\ g_4 \in L^\infty([-T,0]),\ \lambda_1,\lambda_2,\lambda_3 \in \mathbb{R}\big\}.\end{aligned}$$

Using the notations of Example 3.4 and Remark 3.5 in [16], to which we refer for more details on this subject, we notice that χ_0 is indeed given by the direct sum $\chi_0 = L^2([-T,0]^2) \oplus \big(L^2([-T,0])\hat{\otimes}_h\mathscr{D}_0\big) \oplus \big(\mathscr{D}_0\hat{\otimes}_h L^2([-T,0])\big) \oplus \mathscr{D}_{0,0}([-T,0]^2) \oplus Diag([-T,0]^2)$. In the sequel, we shall refer to the term $g_4(x)\delta_y(dx)\otimes dy$ as the **diagonal component** and to $g_4(x)$ as the **diagonal element** of μ. □

We can now state our first representation result for $D^H\mathscr{U}$.

Proposition 7 *Let $\mathscr{U}\colon C([-T,0]) \to \mathbb{R}$ be continuously Fréchet differentiable. Suppose the following.*

(i) For any $\eta \in C([-T,0])$ there exists $D_\cdot^{ac}\mathscr{U}(\eta) \in BV([-T,0])$ such that

$$D_{dx}^\perp\mathscr{U}(\eta) = D_x^{ac}\mathscr{U}(\eta)dx.$$

(ii) There exist continuous extensions (necessarily unique)

$$u\colon \mathscr{C}([-T,0]) \to \mathbb{R}, \qquad D_\cdot^{ac}u\colon \mathscr{C}([-T,0]) \to BV([-T,0])$$

of $\mathscr{U}$ and $D_\cdot^{ac}\mathscr{U}$, respectively.

Then, for any $\eta \in C([-T,0])$,

$$D^H\mathscr{U}(\eta) = \int_{[-T,0]} D_x^{ac}\mathscr{U}(\eta)d^+\eta(x), \tag{29}$$

where we recall that the previous deterministic integral has been defined in Sect. 2.1.1. In particular, the horizontal derivative $D^H \mathscr{U}(\eta)$ and the backward integral in (29) exist.

Proof Let $\eta \in C([-T,0])$, then starting from the left-hand side of (29), using the definition of $D^H \mathscr{U}(\eta)$, we are led to consider the following increment for the function u:

$$\frac{u(\eta) - u(\eta(\cdot - \varepsilon)1_{[-T,0[} + \eta(0)1_{\{0\}})}{\varepsilon}. \tag{30}$$

We shall expand (30) using a Taylor's formula. Firstly, notice that, since $\mathscr{U}$ is C^1 Fréchet on $C([-T,0])$, for every $\eta_1 \in C([-T,0])$, with $\eta_1(0) = \eta(0)$, from the fundamental theorem of calculus we have

$$\mathscr{U}(\eta) - \mathscr{U}(\eta_1) \ = \ \int_0^1 \left(\int_{-T}^0 D_x^{\mathrm{ac}} \mathscr{U}(\eta + \lambda(\eta_1 - \eta))(\eta(x) - \eta_1(x))dx \right) d\lambda.$$

Recalling from Remark 3 the density of $C_{\eta(0)}([-T,0])$ in $\mathscr{C}_{\eta(0)}([-T,0])$ with respect to the topology of $\mathscr{C}([-T,0])$, we deduce the following Taylor's formula for u:

$$u(\eta) - u(\eta_1) \ = \ \int_0^1 \left(\int_{-T}^0 D_x^{\mathrm{ac}} u(\eta + \lambda(\eta_1 - \eta))(\eta(x) - \eta_1(x))dx \right) d\lambda, \tag{31}$$

for all $\eta_1 \in \mathscr{C}_{\eta(0)}([-T,0])$. As a matter of fact, for any $\delta \in]0, T/2]$ let (similarly to Remark 3(i))

$$\eta_{1,\delta}(x) \ := \ \begin{cases} \eta_1(x), & -T \leq x \leq -\delta, \\ \frac{1}{\delta}(\eta_1(0) - \eta_1(-\delta))x + \eta_1(0), & -\delta < x \leq 0 \end{cases}$$

and $\eta_{1,0} := \eta_1$. Then $\eta_{1,\delta} \in C([-T,0])$, for any $\delta \in]0, T/2]$, and $\eta_{1,\delta} \to \eta_1$ in $\mathscr{C}([-T,0])$, as $\delta \to 0^+$. Now, define $f\colon [-T,0] \times [0,1] \times [0,T/2] \to \mathbb{R}$ as follows

$$f(x,\lambda,\delta) \ := \ D_x^{\mathrm{ac}} u(\eta + \lambda(\eta_{1,\delta} - \eta))(\eta(x) - \eta_{1,\delta}(x)),$$

for all $(x,\lambda,\delta) \in [-T,0] \times [0,1] \times [0,T/2]$. Now $(\lambda,\delta) \mapsto \eta + \lambda(\eta_{1,\delta} - \eta)$, is continuous. Taking into account that $D_{\cdot}^{\mathrm{ac}} u\colon \mathscr{C}([-T,0]) \to BV([-T,0])$ is continuous, hence bounded on compact sets, it follows that f is bounded. Then, it follows from Lebesgue dominated convergence theorem that

$$
\begin{aligned}
&\int_0^1 \left(\int_{-T}^0 D_x^{ac} \mathscr{U}(\eta + \lambda(\eta_{1,\delta} - \eta))(\eta(x) - \eta_{1,\delta}(x)) dx \right) d\lambda \\
&= \int_0^1 \left(\int_{-T}^0 f(x, \lambda, \delta) dx \right) d\lambda \overset{\delta \to 0^+}{\longrightarrow} \int_0^1 \left(\int_{-T}^0 f(x, \lambda, 0) dx \right) d\lambda \\
&= \int_0^1 \left(\int_{-T}^0 D_x^{ac} u(\eta + \lambda(\eta_1 - \eta))(\eta(x) - \eta_1(x)) dx \right) d\lambda,
\end{aligned}
$$

from which we deduce (31), since $\mathscr{U}(\eta_{1,\delta}) \to u(\eta_1)$ as $\delta \to 0^+$. Taking $\eta_1(\cdot) = \eta(\cdot - \varepsilon)1_{[-T,0[} + \eta(0)1_{\{0\}}$, we obtain

$$
\begin{aligned}
&\frac{u(\eta) - u(\eta(\cdot - \varepsilon)1_{[-T,0[} + \eta(0)1_{\{0\}})}{\varepsilon} \\
&= \int_0^1 \left(\int_{-T}^0 D_x^{ac} u\big(\eta + \lambda\big(\eta(\cdot - \varepsilon) - \eta(\cdot)\big)1_{[-T,0[}\big) \frac{\eta(x) - \eta(x - \varepsilon)}{\varepsilon} dx \right) d\lambda \\
&= I_1(\eta, \varepsilon) + I_2(\eta, \varepsilon) + I_3(\eta, \varepsilon),
\end{aligned}
$$

where

$$
\begin{aligned}
I_1(\eta, \varepsilon) :=& \int_0^1 \left(\int_{-T}^0 \eta(x) \frac{1}{\varepsilon} \Big(D_x^{ac} u\big(\eta + \lambda\big(\eta(\cdot - \varepsilon) - \eta(\cdot)\big)1_{[-T,0[}\big) \right. \\
&\qquad \left. - D_{x+\varepsilon}^{ac} u\big(\eta + \lambda\big(\eta(\cdot - \varepsilon) - \eta(\cdot)\big)1_{[-T,0[}\big)\Big) dx \right) d\lambda, \\
I_2(\eta, \varepsilon) :=& \frac{1}{\varepsilon} \int_0^1 \left(\int_{-\varepsilon}^0 \eta(x) D_{x+\varepsilon}^{ac} u\big(\eta + \lambda\big(\eta(\cdot - \varepsilon) - \eta(\cdot)\big)1_{[-T,0[}\big) dx \right) d\lambda, \\
I_3(\eta, \varepsilon) :=& -\frac{1}{\varepsilon} \int_0^1 \left(\int_{-T-\varepsilon}^{-T} \eta(x) D_{x+\varepsilon}^{ac} u\big(\eta + \lambda\big(\eta(\cdot - \varepsilon) - \eta(\cdot)\big)1_{[-T,0[}\big) dx \right) d\lambda.
\end{aligned}
$$

Notice that, since $\eta(x) = 0$ for $x < -T$, we see that $I_3(\eta, \varepsilon) = 0$. Moreover $D_x^{ac} u(\cdot) = D_0^{ac} u(\cdot)$, for $x \geq 0$, and $\eta + \lambda(\eta(\cdot - \varepsilon) - \eta(\cdot))1_{[-T,0[} \to \eta$ in $\mathscr{C}([-T, 0])$ as $\varepsilon \to 0^+$. Since $D_x^{ac} u$ is continuous from $\mathscr{C}([-T, 0])$ into $BV([-T, 0])$, we have $D_0^{ac} u(\eta + \lambda(\eta(\cdot - \varepsilon) - \eta(\cdot))1_{[-T,0[}) \to D_0^{ac} u(\eta)$ as $\varepsilon \to 0^+$. Then

$$
\begin{aligned}
&\frac{1}{\varepsilon} \int_{-\varepsilon}^0 \eta(x) D_{x+\varepsilon}^{ac} u\big(\eta + \lambda\big(\eta(\cdot - \varepsilon) - \eta(\cdot)\big)1_{[-T,0[}\big) dx \\
&= \frac{1}{\varepsilon} \int_{-\varepsilon}^0 \eta(x) dx \, D_0^{ac} u\big(\eta + \lambda\big(\eta(\cdot - \varepsilon) - \eta(\cdot)\big)1_{[-T,0[}\big) \overset{\varepsilon \to 0^+}{\longrightarrow} \eta(0) D_0^{ac} u(\eta).
\end{aligned}
$$

So $I_2(\eta, \varepsilon) \to \eta(0) D_0^{ac} u(\eta)$. Finally, concerning $I_1(\eta, \varepsilon)$, from Fubini's theorem we obtain (denoting $\eta_{\varepsilon,\lambda} := \eta + \lambda(\eta(\cdot - \varepsilon) - \eta(\cdot))1_{[-T,0[}$)

$$\begin{aligned}
I_1(\eta,\varepsilon) &= \int_0^1\bigg(\int_{-T}^0 \eta(x)\frac{1}{\varepsilon}\Big(D_x^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda}) - D_{x+\varepsilon}^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda})\Big)dx\bigg)d\lambda \\
&= -\int_0^1\bigg(\int_{-T}^0 \eta(x)\frac{1}{\varepsilon}\bigg(\int_{]x,x+\varepsilon]} D_{dy}^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda})\bigg)dx\bigg)d\lambda \\
&= -\int_0^1\bigg(\int_{]-T,\varepsilon]}\frac{1}{\varepsilon}\bigg(\int_{(-T)\vee(y-\varepsilon)}^{0\wedge y}\eta(x)dx\bigg)D_{dy}^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda})\bigg)d\lambda \\
&= I_{11}(\eta,\varepsilon) + I_{12}(\eta,\varepsilon),
\end{aligned}$$

where

$$\begin{aligned}
I_{11}(\eta,\varepsilon) &:= -\int_0^1\bigg(\int_{]-T,\varepsilon]}\frac{1}{\varepsilon}\bigg(\int_{(-T)\vee(y-\varepsilon)}^{0\wedge y}\eta(x)dx\bigg)\Big(D_{dy}^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda}) - D_{dy}^{\mathrm{ac}}u(\eta)\Big)\bigg)d\lambda, \\
I_{12}(\eta,\varepsilon) &:= -\int_0^1\bigg(\int_{]-T,\varepsilon]}\frac{1}{\varepsilon}\bigg(\int_{(-T)\vee(y-\varepsilon)}^{0\wedge y}\eta(x)dx\bigg)D_{dy}^{\mathrm{ac}}u(\eta)\bigg)d\lambda \\
&= -\bigg(\int_{]-T,\varepsilon]}\frac{1}{\varepsilon}\bigg(\int_{(-T)\vee(y-\varepsilon)}^{0\wedge y}\eta(x)dx\bigg)D_{dy}^{\mathrm{ac}}u(\eta).
\end{aligned}$$

Recalling that $D_x^{\mathrm{ac}}u(\cdot) = D_0^{\mathrm{ac}}u(\cdot)$, for $x \geq 0$, we see that in $I_{11}(\eta,\varepsilon)$ and $I_{12}(\eta,\varepsilon)$ the integrals on $]-T,\varepsilon]$ are equal to the same integrals on $]-T,0]$, i.e.,

$$\begin{aligned}
I_{11}(\eta,\varepsilon) &= -\int_0^1\bigg(\int_{]-T,0]}\frac{1}{\varepsilon}\bigg(\int_{(-T)\vee(y-\varepsilon)}^{0\wedge y}\eta(x)dx\bigg)\Big(D_{dy}^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda}) - D_{dy}^{\mathrm{ac}}u(\eta)\Big)\bigg)d\lambda \\
&= -\int_0^1\bigg(\int_{]-T,0]}\frac{1}{\varepsilon}\bigg(\int_{y-\varepsilon}^{y}\eta(x)dx\bigg)\Big(D_{dy}^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda}) - D_{dy}^{\mathrm{ac}}u(\eta)\Big)\bigg)d\lambda, \\
I_{12}(\eta,\varepsilon) &= -\int_{]-T,0]}\frac{1}{\varepsilon}\bigg(\int_{(-T)\vee(y-\varepsilon)}^{0\wedge y}\eta(x)dx\bigg)D_{dy}^{\mathrm{ac}}u(\eta) \\
&= -\int_{]-T,0]}\frac{1}{\varepsilon}\bigg(\int_{y-\varepsilon}^{y}\eta(x)dx\bigg)D_{dy}^{\mathrm{ac}}u(\eta).
\end{aligned}$$

Now, observe that

$$|I_{11}(\eta,\varepsilon)| \;\leq\; \|\eta\|_\infty \|D_\cdot^{\mathrm{ac}}u(\eta_{\varepsilon,\lambda}) - D_\cdot^{\mathrm{ac}}u(\eta)\|_{\mathrm{Var}([-T,0])} \overset{\varepsilon\to 0^+}{\longrightarrow} 0.$$

Moreover, since η is continuous at $y \in]-T,0]$, we deduce that $\int_{y-\varepsilon}^y \eta(x)dx/\varepsilon \to \eta(y)$ as $\varepsilon \to 0^+$. Therefore, by Lebesgue's dominated convergence theorem, we get

$$I_{12}(\eta,\varepsilon) \overset{\varepsilon\to 0^+}{\longrightarrow} -\int_{]-T,0]}\eta(y)D_{dy}^{\mathrm{ac}}u(\eta).$$

So $I_1(\eta, \varepsilon) \to -\int_{]-T,0]} \eta(y) D^{ac}_{dy} u(\eta)$. In conclusion, we have

$$D^H \mathscr{U}(\eta) \; = \; \eta(0) D^{ac}_0 u(\eta) - \int_{]-T,0]} \eta(y) D^{ac}_{dy} u(\eta).$$

Notice that we can suppose, without loss of generality, $D^{ac}_{0-} \mathscr{U}(\eta) \; = \; D^{ac}_0 \mathscr{U}(\eta)$. Then, the above identity gives (29) using the integration by parts formula (12). □

For our second representation result of $D^H \mathscr{U}$ we need the following generalization of the deterministic backward integral when the integrand is a measure.

Definition 18 Let $a < b$ be two reals. Let $f : [a, b] \to \mathbb{R}$ be a càdlàg function (resp. càdlàg function with f(a) = 0) and $g \in \mathscr{M}([-T, 0])$. Suppose that the following limit

$$\int_{[a,b]} g(ds) d^+ f(s) := \lim_{\varepsilon \to 0^+} \int_{[a,b]} g(ds) \frac{f_{\overline{J}}(s) - f_{\overline{J}}(s - \varepsilon)}{\varepsilon} \tag{32}$$

$$\Big(\text{resp.} \int_{[a,b]} g(ds) d^- f(s) := \lim_{\varepsilon \to 0^+} \int_{[a,b]} g(ds) \frac{f_{\overline{J}}(s + \varepsilon) - f_{\overline{J}}(s)}{\varepsilon}\Big) \tag{33}$$

exists and it is finite. Then, the obtained quantity is denoted by $\int_{[a,b]} g d^+ f$ ($\int_{[a,b]} g d^- f$) and called (**deterministic, definite**) **backward** (**resp. forward**) **integral of** g **with respect to** f (**on** $[a, b]$).

Proposition 8 *If g is absolutely continuous with density being càdlàg (still denoted with g) then Definition 18 is compatible with the one in Definition 4.*

Proof Suppose that $g(ds) = g(s)ds$ with g càdlàg.

Identity (32). The right-hand side of (6) gives

$$\int_a^b g(s) \frac{f(s) - f(s - \varepsilon)}{\varepsilon} ds,$$

which is also the right-hand side of (32) in that case.

Identity (33). The right-hand side of (5) gives, since $f(a) = 0$,

$$\frac{1}{\varepsilon} g(a) \int_{a-\varepsilon}^{a} f(s + \varepsilon) ds + \int_a^b g(s) \frac{f_{\overline{J}}(s + \varepsilon) - f_{\overline{J}}(s)}{\varepsilon} ds$$

The first integral goes to zero. The second one equals the right-hand side of (33). □

Proposition 9 *Let $\eta \in C([-T,0])$ be such that the quadratic variation on $[-T,0]$ exists. Let $\mathscr{U} : C([-T,0]) \to \mathbb{R}$ be twice continuously Fréchet differentiable such that*

$$D^2\mathscr{U} : C([-T,0]) \longrightarrow \chi_0 \subset (C([-T,0])\hat{\otimes}_\pi C([-T,0]))^* \text{ continuously with respect to } \chi_0.$$

Let us also suppose the following.

(i) $D_x^{2,Diag}\mathscr{U}(\eta)$, *the diagonal element of the second-order derivative at η, has a set of discontinuity which has null measure with respect to $[\eta]$ (in particular, if it is countable).*

(ii) There exist continuous extensions (necessarily unique):

$$u : \mathscr{C}([-T,0]) \to \mathbb{R}, \qquad D^2_{dx\,dy}u : \mathscr{C}([-T,0]) \to \chi_0$$

of $\mathscr{U}$ and $D^2_{dx\,dy}\mathscr{U}$, respectively.

(iii) The horizontal derivative $D^H\mathscr{U}(\eta)$ exists at $\eta \in C([-T,0])$.

Then

$$D^H\mathscr{U}(\eta) \; = \; \int_{[-T,0]} D^\perp_{dx}\mathscr{U}(\eta)d^+\eta(x) - \frac{1}{2}\int_{[-T,0]} D_x^{2,Diag}\mathscr{U}(\eta)d[\eta](x). \quad (34)$$

In particular, the backward integral in (34) exists.

Proof Let $\eta \in C([-T,0])$. Using the definition of $D^H\mathscr{U}(\eta)$ we are led to consider the following increment for the function u:

$$\frac{u(\eta) - u(\eta(\cdot - \varepsilon)1_{[-T,0[} + \eta(0)1_{\{0\}})}{\varepsilon}, \quad (35)$$

with $\varepsilon > 0$. Our aim is to expand (35) using a Taylor's formula. To this end, since $\mathscr{U}$ is C^2 Fréchet, we begin noting that for every $\eta_1 \in C([-T,0])$ the following standard Taylor's expansion holds:

$$\begin{aligned}
\mathscr{U}(\eta_1) = \; & \mathscr{U}(\eta) + \int_{[-T,0]} D_{dx}\mathscr{U}(\eta)\big(\eta_1(x) - \eta(x)\big) \\
& + \frac{1}{2}\int_{[-T,0]^2} D^2_{dx\,dy}\mathscr{U}(\eta)\big(\eta_1(x) - \eta(x)\big)\big(\eta_1(y) - \eta(y)\big) \\
& + \int_0^1 (1-\lambda)\bigg(\int_{[-T,0]^2}\Big(D^2_{dx\,dy}\mathscr{U}(\eta + \lambda(\eta_1 - \eta)) \\
& - D^2_{dx\,dy}\mathscr{U}(\eta)\Big)\big(\eta_1(x) - \eta(x)\big)\big(\eta_1(y) - \eta(y)\big)\bigg)d\lambda.
\end{aligned}$$

Now, using the density of $C_{\eta(0)}([-T,0])$ into $\mathscr{C}_{\eta(0)}([-T,0])$ with respect to the topology of $\mathscr{C}([-T,0])$ and proceeding as in the proof of Proposition 7, we deduce the following Taylor's formula for u:

$$\begin{aligned}
&\frac{u(\eta) - u(\eta(\cdot-\varepsilon)1_{[-T,0[} + \eta(0)1_{\{0\}})}{\varepsilon} \\
&= \int_{[-T,0]} D^{\perp}_{dx}\mathscr{U}(\eta)\frac{\eta(x)-\eta(x-\varepsilon)}{\varepsilon} \\
&\quad - \frac{1}{2}\int_{[-T,0]^2} D^2_{dx\,dy}\mathscr{U}(\eta)\frac{(\eta(x)-\eta(x-\varepsilon))(\eta(y)-\eta(y-\varepsilon))}{\varepsilon}1_{[-T,0[\times[-T,0[}(x,y) \\
&\quad - \int_0^1 (1-\lambda)\bigg(\int_{[-T,0]^2}\Big(D^2_{dx\,dy}u(\eta+\lambda(\eta(\cdot-\varepsilon)-\eta(\cdot))1_{[-T,0[}) \\
&\quad - D^2_{dx\,dy}\mathscr{U}(\eta)\Big)\frac{(\eta(x)-\eta(x-\varepsilon))(\eta(y)-\eta(y-\varepsilon))}{\varepsilon}1_{[-T,0[\times[-T,0[}(x,y)\bigg)d\lambda.
\end{aligned} \tag{36}$$

Recalling the definition of χ_0 given in Remark 11, we notice that (due to the presence of the indicator function $1_{[-T,0[\times[-T,0[}$)

$$\begin{aligned}
&\int_{[-T,0]^2} D^2_{dx\,dy}\mathscr{U}(\eta)\frac{(\eta(x)-\eta(x-\varepsilon))(\eta(y)-\eta(y-\varepsilon))}{\varepsilon}1_{[-T,0[\times[-T,0[}(x,y) \\
&= \int_{[-T,0]^2} D^{2,L^2}_{x\,y}\mathscr{U}(\eta)\frac{(\eta(x)-\eta(x-\varepsilon))(\eta(y)-\eta(y-\varepsilon))}{\varepsilon}dxdy \\
&\quad + \int_{[-T,0]} D^{2,Diag}_{x}\mathscr{U}(\eta)\frac{(\eta(x)-\eta(x-\varepsilon))^2}{\varepsilon}dx,
\end{aligned}$$

where, by hypothesis, the maps $\eta \in \mathscr{C}([-T,0]) \mapsto D^{2,L^2}_{x\,y}u(\eta) \in L^2([-T,0]^2)$ and $\eta \in \mathscr{C}([-T,0]) \mapsto D^{2,Diag}_{x}u(\eta) \in L^\infty([-T,0])$ are continuous. In particular, (36) becomes

$$\frac{u(\eta) - u(\eta(\cdot-\varepsilon)1_{[-T,0[} + \eta(0)1_{\{0\}})}{\varepsilon} = I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon) + I_4(\varepsilon) + I_5(\varepsilon), \tag{37}$$

where

$$\begin{aligned}
I_1(\varepsilon) &:= \int_{[-T,0]} D^{\perp}_{dx}\mathscr{U}(\eta)\frac{\eta(x)-\eta(x-\varepsilon)}{\varepsilon}, \\
I_2(\varepsilon) &:= -\frac{1}{2}\int_{[-T,0]^2} D^{2,L^2}_{x\,y}\mathscr{U}(\eta)\frac{(\eta(x)-\eta(x-\varepsilon))(\eta(y)-\eta(y-\varepsilon))}{\varepsilon}dxdy, \\
I_3(\varepsilon) &:= -\frac{1}{2}\int_{[-T,0]} D^{2,Diag}_{x}\mathscr{U}(\eta)\frac{(\eta(x)-\eta(x-\varepsilon))^2}{\varepsilon}dx,
\end{aligned}$$

$$
\begin{aligned}
I_4(\varepsilon) := & -\int_0^1 (1-\lambda)\bigg(\int_{[-T,0]^2} \Big(D^{2,L^2}_{x\,y} u(\eta + \lambda(\eta(\cdot - \varepsilon) - \eta(\cdot))1_{[-T,0[})\\
& - D^{2,L^2}_{x\,y} \mathscr{U}(\eta)\Big) \frac{(\eta(x) - \eta(x-\varepsilon))(\eta(y) - \eta(y-\varepsilon))}{\varepsilon} dx\, dy\bigg) d\lambda,\\
I_5(\varepsilon) := & -\int_0^1 (1-\lambda)\bigg(\int_{[-T,0]} \Big(D^{2,Diag}_{x} u(\eta + \lambda(\eta(\cdot - \varepsilon) - \eta(\cdot))1_{[-T,0[})\\
& - D^{2,Diag}_{x} \mathscr{U}(\eta)\Big) \frac{(\eta(x) - \eta(x-\varepsilon))^2}{\varepsilon} dx\bigg) d\lambda.
\end{aligned}
$$

Firstly, we shall prove that

$$
I_2(\varepsilon) \overset{\varepsilon\to 0^+}{\longrightarrow} 0. \tag{38}
$$

To this end, for every $\varepsilon > 0$, we define the operator $T_\varepsilon \colon L^2([-T,0]^2) \to \mathbb{R}$ as follows:

$$
T_\varepsilon\, g \;=\; \int_{[-T,0]^2} g(x,y) \frac{(\eta(x) - \eta(x-\varepsilon))(\eta(y) - \eta(y-\varepsilon))}{\varepsilon} dxdy, \qquad \forall\, g \in L^2([-T,0]^2).
$$

Then $T_\varepsilon \in L^2([-T,0])^*$. Indeed, from Cauchy-Schwarz inequality,

$$
\begin{aligned}
|T_\varepsilon\, g| \le \; & \|g\|_{L^2([-T,0]^2)} \sqrt{\int_{[-T,0]^2} \frac{(\eta(x) - \eta(x-\varepsilon))^2(\eta(y) - \eta(y-\varepsilon))^2}{\varepsilon^2} dxdy}\\
= \; & \|g\|_{L^2([-T,0]^2)} \int_{[-T,0]} \frac{(\eta(x) - \eta(x-\varepsilon))^2}{\varepsilon} dx
\end{aligned}
$$

and the latter quantity is bounded with respect to ε since the quadratic variation of η on $[-T,0]$ exists. In particular, we have proved that for every $g \in L^2([-T,0]^2)$ there exists a constant $M_g \ge 0$ such that

$$
\sup_{0<\varepsilon<1} |T_\varepsilon\, g| \;\le\; M_g.
$$

It follows from Banach-Steinhaus theorem that there exists a constant $M \ge 0$ such that

$$
\sup_{0<\varepsilon<1} \|T_\varepsilon\|_{L^2([-T,0])^*} \;\le\; M. \tag{39}
$$

Now, let us consider the set $\mathscr{S} := \{g \in L^2([-T,0]^2) \colon g(x,y) = e(x)f(y)$, with $e, f \in C^1([-T,0])\}$, which is dense in $L^2([-T,0]^2)$. Let us show that

$$
T_\varepsilon\, g \overset{\varepsilon\to 0^+}{\longrightarrow} 0, \qquad \forall\, g \in \mathscr{S}. \tag{40}
$$

Fix $g \in \mathscr{S}$, with $g(x, y) = e(x) f(y)$ for any $(x, y) \in [-T, 0]$, then

$$T_\varepsilon\, g \;=\; \frac{1}{\varepsilon} \int_{[-T,0]} e(x)\big(\eta(x) - \eta(x-\varepsilon)\big)dx \int_{[-T,0]} f(y)\big(\eta(y) - \eta(y-\varepsilon)\big)dy. \quad (41)$$

We have

$$\begin{aligned} \bigg| \int_{[-T,0]} e(x)\big(\eta(x) - \eta(x - \varepsilon)\big)dx \bigg| \;&=\; \bigg| \int_{[-T,0]} \big(e(x) - e(x + \varepsilon)\big)\eta(x)dx \\ &- \int_{[-T-\varepsilon,-T]} e(x + \varepsilon)\eta(x)dx + \int_{[-\varepsilon,0]} e(x + \varepsilon)\eta(x)dx \bigg| \\ &\le\; \varepsilon \bigg(\int_{[-T,0]} |\dot{e}(x)|dx + 2\|e\|_\infty \bigg) \|\eta\|_\infty. \end{aligned}$$

Similarly,

$$\bigg| \int_{[-T,0]} f(y)\big(\eta(y) - \eta(y - \varepsilon)\big)dy \bigg| \;\le\; \varepsilon \bigg(\int_{[-T,0]} |\dot{f}(y)|dy + 2\|f\|_\infty \bigg) \|\eta\|_\infty.$$

Therefore, from (41) we find

$$|T_\varepsilon\, g| \;\le\; \varepsilon \bigg(\int_{[-T,0]} |\dot{e}(x)|dx + 2\|e\|_\infty \bigg) \bigg(\int_{[-T,0]} |\dot{f}(y)|dy + 2\|f\|_\infty \bigg) \|\eta\|_\infty^2,$$

which converges to zero as ε goes to zero and therefore (40) is established. This in turn implies that

$$T_\varepsilon\, g \overset{\varepsilon \to 0^+}{\longrightarrow} 0, \qquad \forall\, g \in L^2([-T, 0]^2). \quad (42)$$

Indeed, fix $g \in L^2([-T, 0]^2)$ and let $(g_n)_n \subset \mathscr{S}$ be such that $g_n \to g$ in $L^2([-T, 0]^2)$. Then

$$|T_\varepsilon\, g| \;\le\; |T_\varepsilon(g - g_n)| + |T_\varepsilon\, g_n| \;\le\; \|T_\varepsilon\|_{L^2([-T,0]^2)^*} \|g - g_n\|_{L^2([-T,0]^2)} + |T_\varepsilon\, g_n|.$$

From (39) it follows that

$$|T_\varepsilon\, g| \;\le\; M\|g - g_n\|_{L^2([-T,0]^2)} + |T_\varepsilon\, g_n|,$$

which implies $\limsup_{\varepsilon \to 0^+} |T_\varepsilon\, g| \le M\|g - g_n\|_{L^2([-T,0]^2)}$. Sending n to infinity, we deduce (42) and finally (38).

Let us now consider the term $I_3(\varepsilon)$ in (37). Since the quadratic variation $[\eta]$ exists, it follows from Portmanteau's theorem and hypothesis (i) that

$$I_3(\varepsilon) \;=\; \int_{[-T,0]} D_x^{2,Diag}\mathscr{U}(\eta)\frac{(\eta(x)-\eta(x-\varepsilon))^2}{\varepsilon}dx \;\underset{\varepsilon\to 0^+}{\longrightarrow}\; \int_{[-T,0]} D_x^{2,Diag}\mathscr{U}(\eta)d[\eta](x).$$

Regarding the term $I_4(\varepsilon)$ in (37), let $\phi_\eta \colon [0,1]^2 \to L^2([-T,0]^2)$ be given by

$$\phi_\eta(\varepsilon,\lambda)(\cdot,\cdot) \;=\; D^{2,L^2}_{\cdot\,\cdot} u\big(\eta + \lambda(\eta(\cdot-\varepsilon)-\eta(\cdot))1_{[-T,0[}\big).$$

By hypothesis, ϕ_η is a continuous map, and hence it is uniformly continuous, since $[0,1]^2$ is a compact set. Let ρ_{ϕ_η} denote the continuity modulus of ϕ_η, then

$$\begin{aligned}
&\big\| D^{2,L^2}_{\cdot\,\cdot} u\big(\eta+\lambda(\eta(\cdot-\varepsilon)-\eta(\cdot))1_{[-T,0[}\big) - D^{2,L^2}_{\cdot\,\cdot}\mathscr{U}(\eta)\big\|_{L^2([-T,0]^2)} \\
&= \|\phi_\eta(\varepsilon,\lambda)-\phi_\eta(0,\lambda)\|_{L^2([-T,0]^2)} \;\le\; \rho_{\phi_\eta}(\varepsilon).
\end{aligned}$$

This implies, by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\bigg|\int_0^1 (1-\lambda)\bigg(\int_{[-T,0]^2}\Big(D^{2,L^2}_{x\,y} u(\eta+\lambda(\eta(\cdot-\varepsilon)-\eta(\cdot))1_{[-T,0[}) \\
&- D^{2,L^2}_{x\,y}\mathscr{U}(\eta)\Big)\frac{(\eta(x)-\eta(x-\varepsilon))(\eta(y)-\eta(y-\varepsilon))}{\varepsilon}dxdy\bigg)d\lambda\bigg| \\
&\le \int_0^1 (1-\lambda)\big\|D^{2,L^2}_{\cdot\,\cdot} u(\eta+\lambda(\eta(\cdot-\varepsilon)-\eta(\cdot))1_{[-T,0]}) \\
&- D^{2,L^2}_{\cdot\,\cdot}\mathscr{U}(\eta)\big\|_{L^2([-T,0]^2)}\sqrt{\int_{[-T,0]^2}\frac{(\eta(x)-\eta(x-\varepsilon))^2(\eta(y)-\eta(y-\varepsilon))^2}{\varepsilon^2}dxdyd\lambda} \\
&\le \int_0^1 (1-\lambda)\rho_{\phi_\eta}(\varepsilon)\bigg(\int_{[-T,0]}\frac{(\eta(x)-\eta(x-\varepsilon))^2}{\varepsilon}dx\bigg)d\lambda \\
&= \frac{1}{2}\rho_{\phi_\eta}(\varepsilon)\int_{[-T,0]}\frac{(\eta(x)-\eta(x-\varepsilon))^2}{\varepsilon}dx \;\overset{\varepsilon\to 0^+}{\longrightarrow}\; 0.
\end{aligned}$$

Finally, we consider the term $I_5(\varepsilon)$ in (37). Define $\psi_\eta \colon [0,1]^2 \to L^\infty([-T,0])$ as follows:

$$\psi_\eta(\varepsilon,\lambda)(\cdot) \;=\; D^{2,Diag}_{\cdot} u\big(\eta+\lambda(\eta(\cdot-\varepsilon)-\eta(\cdot))1_{[-T,0[}\big).$$

We see that ψ_η is uniformly continuous. Let ρ_{ψ_η} denote the continuity modulus of ψ_η, then

$$\begin{aligned}
&\big\|D^{2,Diag}_{\cdot} u\big(\eta+\lambda(\eta(\cdot-\varepsilon)-\eta(\cdot))1_{[-T,0[}\big) - D^{2,Diag}_{\cdot}\mathscr{U}(\eta)\big\|_{L^\infty([-T,0])} \\
&= \|\psi_\eta(\varepsilon,\lambda)-\psi_\eta(0,\lambda)\|_{L^\infty([-T,0])} \;\le\; \rho_{\psi_\eta}(\varepsilon).
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\Big|\int_0^1 (1-\lambda)\bigg(\int_{[-T,0]} \Big(D_x^{2,Diag} u(\eta + \lambda(\eta(\cdot - \varepsilon) - \eta(\cdot))1_{[-T,0[})\\
&- D_x^{2,Diag}\mathscr{U}(\eta)\Big)\frac{(\eta(x) - \eta(x-\varepsilon))^2}{\varepsilon}dx\bigg)d\lambda\Big|\\
&\le \int_0^1 (1-\lambda)\bigg(\int_{[-T,0]} \rho_{\psi_\eta}(\varepsilon)\frac{(\eta(x) - \eta(x-\varepsilon))^2}{\varepsilon}dx\bigg)d\lambda\\
&= \frac{1}{2}\rho_{\psi_\eta}(\varepsilon)\int_{[-T,0]} \frac{(\eta(x) - \eta(x-\varepsilon))^2}{\varepsilon}dx \overset{\varepsilon\to 0^+}{\longrightarrow} 0.
\end{aligned}$$

In conclusion, we have proved that all the integral terms in the right-hand side of (37), unless $I_1(\varepsilon)$, admit a limit when ε goes to zero. Since the left-hand side admits a limit, namely $D^H\mathscr{U}(\eta)$, we deduce that the backward integral

$$I_1(\varepsilon) = \int_{[-T,0]} D_{dx}^{\perp}\mathscr{U}(\eta)\frac{\eta(x) - \eta(x-\varepsilon)}{\varepsilon} \overset{\varepsilon\to 0^+}{\longrightarrow} \int_{[-T,0]} D_{dx}^{\perp}\mathscr{U}(\eta)d^+\eta(x)$$

exists and it is finite, which concludes the proof. □

3 Strong-Viscosity Solutions to Path-Dependent PDEs

In the present section we study the semilinear parabolic path-dependent equation

$$\begin{cases}\partial_t\mathscr{U} + D^H\mathscr{U} + \frac{1}{2}D^{VV}\mathscr{U} + F(t,\eta,\mathscr{U},D^V\mathscr{U}) = 0, & \forall\,(t,\eta)\in[0,T[\times C([-T,0]),\\ \mathscr{U}(T,\eta) = H(\eta), & \forall\,\eta\in C([-T,0]).\end{cases} \tag{43}$$

We refer to $\mathscr{L}\mathscr{U} = \partial_t\mathscr{U} + D^H\mathscr{U} + \frac{1}{2}D^{VV}\mathscr{U}$ as the *path-dependent heat operator*. The results of this section are generalized in [9, 10], where more general path-dependent equations will be considered. Here we shall impose the following assumptions on H and F.

(A) $H\colon C([-T,0]) \to \mathbb{R}$ and $F\colon [0,T]\times C([-T,0])\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ are Borel measurable functions and satisfy, for some positive constants C and m,

$$\begin{aligned}
|F(t,\eta,y,z) - F(t,\eta,y',z')| &\le C(|y-y'| + |z-z'|),\\
|H(\eta)| + |F(t,\eta,0,0)| &\le C\big(1 + \|\eta\|_\infty^m\big),
\end{aligned}$$

for all $(t,\eta)\in[0,T]\times C([-T,0])$, $y, y'\in\mathbb{R}$, and $z, z'\in\mathbb{R}$.

3.1 Strict Solutions

In the present subsection, we provide the definition of strict solution to Eq. (43) and we state an existence and uniqueness result.

Definition 19 A function $\mathscr{U}\colon [0,T]\times C([-T,0])\to\mathbb{R}$ in $C^{1,2}(([0,T[\times\text{past})\times\text{present})\cap C([0,T]\times C([-T,0]))$, which solves Eq. (43), is called a **strict solution** to (43).

We now introduce some additional notations. Let $(\Omega,\mathscr{F},\mathbb{P})$ be a complete probability space on which a real Brownian motion $W=(W_t)_{t\geq 0}$ is defined. Let $\mathbb{F}=(\mathscr{F}_t)_{t\geq 0}$ denote the completion of the natural filtration generated by W.

- $\mathbb{S}^p(t,T)$, $p\geq 1$, $0\leq t\leq T$, the set of real càdlàg $\mathbb{F}$-adapted processes $Y=(Y_s)_{t\leq s\leq T}$ such that

$$\|Y\|^p_{\mathbb{S}^p(t,T)} := \mathbb{E}\Big[\sup_{t\leq s\leq T}|Y_s|^p\Big] < \infty.$$

- $\mathbb{H}^p(t,T)^d$, $p\geq 1$, $0\leq t\leq T$, the set of $\mathbb{R}^d$-valued predictable processes $Z=(Z_s)_{t\leq s\leq T}$ such that

$$\|Z\|^p_{\mathbb{H}^p(t,T)^d} := \mathbb{E}\bigg[\bigg(\int_t^T |Z_s|^2 ds\bigg)^{\frac{p}{2}}\bigg] < \infty.$$

  We simply write $\mathbb{H}^p(t,T)$ when $d=1$.
- $\mathbb{A}^{+,2}(t,T)$, $0\leq t\leq T$, the set of real nondecreasing predictable processes $K=(K_s)_{t\leq s\leq T}\in\mathbb{S}^2(t,T)$ with $K_t=0$, so that

$$\|K\|^2_{\mathbb{S}^2(t,T)} := \mathbb{E}\big[|K_T|^2\big].$$

- $\mathbb{L}^p(t,T;\mathbb{R}^m)$, $p\geq 1$, $0\leq t\leq T$, the set of $\mathbb{R}^m$-valued $\mathbb{F}$-predictable processes $\phi=(\phi_s)_{t\leq s\leq T}$ such that

$$\|\phi\|^p_{\mathbb{L}^p(t,T;\mathbb{R}^m)} := \mathbb{E}\bigg[\int_t^T |\phi_s|^p ds\bigg] < \infty.$$

Definition 20 Let $t\in[0,T]$ and $\eta\in C([-T,0])$. Then, we define the **stochastic flow**

$$\mathbb{W}^{t,\eta}_s(x) = \begin{cases}\eta(x+s-t), & -T\leq x\leq t-s,\\ \eta(0)+W_{x+s}-W_t, & t-s<x\leq 0,\end{cases}$$

for any $t\leq s\leq T$.

Theorem 4 *Suppose that Assumption* **(A)** *holds. Let* $\mathscr{U}\colon [0,T]\times C([-T,0])\to\mathbb{R}$ *be a strict solution to Eq.* (43)*, satisfying the polynomial growth condition*

$$|\mathscr{U}(t,\eta)| \;\leq\; C\big(1+\|\eta\|_\infty^m\big), \qquad \forall\,(t,\eta)\in[0,T]\times C([-T,0]). \tag{44}$$

for some positive constants C *and* m*. Then, we have*

$$\mathscr{U}(t,\eta) \;=\; Y_t^{t,\eta}, \qquad \forall\,(t,\eta)\in[0,T]\times C([-T,0]),$$

where $(Y_s^{t,\eta},Z_s^{t,\eta})_{s\in[t,T]} \;=\; (\mathscr{U}(s,\mathbb{W}_s^{t,\eta}),D^V\mathscr{U}(s,\mathbb{W}_s^{t,\eta})1_{[t,T[}(s))_{s\in[t,T]} \;\in\; \mathbb{S}^2(t,T)\times\mathbb{H}^2(t,T)$ *is the solution to the backward stochastic differential equation:* $\mathbb{P}$*-a.s.,*

$$Y_s^{t,\eta} \;=\; H(\mathbb{W}_T^{t,\eta})+\int_s^T F(r,\mathbb{W}_r^{t,\eta},Y_r^{t,\eta},Z_r^{t,\eta})dr-\int_s^T Z_r^{t,\eta}dW_r, \qquad t\leq s\leq T.$$

In particular, there exists at most one strict solution to Eq. (43)*.*

Proof Fix $(t,\eta)\in[0,T[\times C([-T,0])$ and set, for all $t\leq s\leq T$,

$$Y_s^{t,\eta} \;=\; \mathscr{U}(s,\mathbb{W}_s^{t,\eta}), \qquad Z_s^{t,\eta} \;=\; D^V\mathscr{U}(s,\mathbb{W}_s^{t,\eta})1_{[t,T[}(s).$$

Then, for any $T_0\in[t,T[$, applying Itô formula (28) to $\mathscr{U}(s,\mathbb{W}_s^{t,\eta})$ and using the fact that $\mathscr{U}$ solves Eq. (43), we find, $\mathbb{P}$-a.s.,

$$Y_s^{t,\eta} \;=\; Y_{T_0}^{t,\eta}+\int_s^{T_0} F(r,\mathbb{W}_r^{t,\eta},Y_r^{t,\eta},Z_r^{t,\eta})dr-\int_s^{T_0} Z_r^{t,\eta}dW_r, \qquad t\leq s\leq T_0. \tag{45}$$

The claim would follow if we could pass to the limit in (45) as $T_0\to T$. To do this, we notice that it follows from Proposition B.1 in [10] that there exists a positive constant c, depending only on T and the constants C and m appearing in the statement of the present Theorem 4, such that

$$\mathbb{E}\int_t^{T_0}|Z_s^{t,\eta}|^2ds \;\leq\; c\|Y^{t,\eta}\|^2_{\mathbb{S}^2(t,T)}+c\mathbb{E}\int_t^T|F(r,\mathbb{W}_r^{t,\eta},0,0)|^2dr, \qquad \forall\,T_0\in[t,T[.$$

We recall that, for any $q\geq 1$,

$$\mathbb{E}\Big[\sup_{t\leq s\leq T}\|\mathbb{W}_s^{t,\eta}\|_\infty^q\Big] \;<\; \infty. \tag{46}$$

Notice that from (44) and (46) we have $\|Y^{t,\eta}\|_{\mathbb{S}^2(t,T)}<\infty$, so that $Y\in\mathbb{S}^2(t,T)$. Then, from the monotone convergence theorem we find

$$\mathbb{E}\int_t^T |Z_s^{t,\eta}|^2 ds \;\le\; c\|Y^{t,\eta}\|^2_{\mathbb{S}^2(t,T)} + c\mathbb{E}\int_t^T |F(r,\mathbb{W}_r^{t,\eta},0,0)|^2 dr.$$

Therefore, it follows from the polynomial growth condition of F and (46) that $Z \in \mathbb{H}^2(t,T)$. This implies, using the Lipschitz character of F in (y,z), that $\mathbb{E}\int_t^T |F(r,\mathbb{W}_r^{t,\eta},Y_r^{t,\eta},Z_r^{t,\eta})|^2 dr < \infty$, so that we can pass to the limit in (45) and we get the claim. □

We conclude this subsection with an existence result for the *path-dependent heat equation*, namely for Eq. (43) with $F \equiv 0$, for which we provide an *ad hoc* proof. For more general cases we refer to [9].

Theorem 5 *Suppose that Assumption* **(A)** *holds. Let $F \equiv 0$ and H be given by, for all $\eta \in C([-T,0])$, (the deterministic integrals are defined according to Definition 4(i))*

$$H(\eta) \;=\; h\bigg(\int_{[-T,0]} \varphi_1(x+T)d^-\eta(x),\dots,\int_{[-T,0]} \varphi_N(x+T)d^-\eta(x)\bigg), \qquad (47)$$

where

- *h belongs $C^2(\mathbb{R}^N)$ and its second order partial derivatives satisfy a polynomial growth condition,*
- *$\varphi_1,\dots,\varphi_N \in C^2([0,T])$.*

Then, there exists a unique strict solution $\mathscr{U}$ to the path-dependent heat Eq. (43), which is given by

$$\mathscr{U}(t,\eta) \;=\; \mathbb{E}\big[H(\mathbb{W}_T^{t,\eta})\big], \qquad \forall\,(t,\eta) \in [0,T]\times C([-T,0]).$$

Proof Let us consider the function $\mathscr{U}\colon [0,T]\times C([-T,0]) \to \mathbb{R}$ given by, for all $(t,\eta) \in [0,T]\times C([-T,0])$,

$$\begin{aligned}
\mathscr{U}(t,\eta) &= \mathbb{E}\big[H(\mathbb{W}_T^{t,\eta})\big] \\
&= \mathbb{E}\bigg[h\bigg(\int_{[-t,0]} \varphi_1(x+t)d^-\eta(x) + \int_t^T \varphi_1(s)dW_s,\dots\bigg)\bigg] \\
&= \Psi\bigg(t,\int_{[-t,0]} \varphi_1(x+t)d^-\eta(x),\dots,\int_{[-t,0]} \varphi_N(x+t)d^-\eta(x)\bigg),
\end{aligned}$$

where

$$\Psi(t,x_1,\dots,x_N) \;=\; \mathbb{E}\bigg[h\bigg(x_1+\int_t^T \varphi_1(s)dW_s,\dots,x_N+\int_t^T \varphi_N(s)dW_s\bigg)\bigg],$$

for any $(t, x_1, \dots, x_N) \in [0, T] \times \mathbb{R}^N$. Notice that, for any $i, j = 1, \dots, N$,

$$D_{x_i} \Psi(t, x_1, \dots, x_N) = \mathbb{E}\Big[D_{x_i} h\Big(x_1 + \int_t^T \varphi_1(s) dW_s, \dots, x_N + \int_t^T \varphi_N(s) dW_s\Big)\Big],$$
$$D^2_{x_i x_j} \Psi(t, x_1, \dots, x_N) = \mathbb{E}\Big[D^2_{x_i x_j} h\Big(x_1 + \int_t^T \varphi_1(s) dW_s, \dots, x_N + \int_t^T \varphi_N(s) dW_s\Big)\Big],$$

so that Ψ and its first and second spatial derivatives are continuous on $[0, T] \times \mathbb{R}^N$. Let us focus on the time derivative $\partial_t \Psi$ of Ψ. We have, for any $\delta > 0$ such that $t + \delta \in [0, T]$,

$$\begin{aligned}&\frac{\Psi(t+\delta, x_1, \dots, x_N) - \Psi(t, x_1, \dots, x_N)}{\delta} \\ &= \frac{1}{\delta}\mathbb{E}\Big[h\Big(x_1 + \int_{t+\delta}^T \varphi_1(s) dW_s, \dots\Big) - h\Big(x_1 + \int_t^T \varphi_1(s) dW_s, \dots\Big)\Big].\end{aligned}$$

Then, using a standard Taylor formula, we find

$$\begin{aligned}&\frac{\Psi(t+\delta, x_1, \dots, x_N) - \Psi(t, x_1, \dots, x_N)}{\delta} \\ &= -\frac{1}{\delta}\mathbb{E}\Big[\int_0^1 \sum_{i=1}^N D_{x_i} h\Big(x_1 + \int_t^T \varphi_1(s) dW_s - \alpha \int_t^{t+\delta} \varphi_1(s) dW_s, \dots\Big) \int_t^{t+\delta} \varphi_i(s) dW_s d\alpha\Big].\end{aligned} \tag{48}$$

Now, it follows from the integration by parts formula of Malliavin calculus, see, e.g., formula (1.42) in [31] (taking into account that Itô integrals are Skorohod integrals), that, for any $i = 1, \dots, N$,

$$\begin{aligned}&\mathbb{E}\Big[D_{x_i} h\Big(x_1 + \int_t^T \varphi_1(s)\big(1 - \alpha 1_{[t,t+\delta]}(s)\big) dW_s, \dots\Big) \int_t^{t+\delta} \varphi_i(s) dW_s\Big] \\ &= (1-\alpha)\mathbb{E}\Big[\sum_{j=1}^N D^2_{x_i x_j} h\Big(x_1 + \int_t^T \varphi_1(s)\big(1 - \alpha 1_{[t,t+\delta]}(s)\big) dW_s, \dots\Big) \int_t^{t+\delta} \varphi_i(s)\varphi_j(s) ds\Big].\end{aligned} \tag{49}$$

Then, plugging (49) into (48) and letting $\delta \to 0^+$, we get (recalling that $D^2_{x_i x_j} h$ has polynomial growth, for any i, j)

$$\partial_t^+ \Psi(t, x_1, \dots, x_N) = -\frac{1}{2}\mathbb{E}\Big[\sum_{i,j=1}^N D^2_{x_i x_j} h\Big(x_1 + \int_t^T \varphi_1(s) dW_s, \dots\Big)\varphi_i(t)\varphi_j(t)\Big], \tag{50}$$

for any $(t, x_1, \dots, x_N) \in [0, T[\times \mathbb{R}^N$, where $\partial_t^+ \Psi$ denotes the right-time derivative of Ψ. Since Ψ and $\partial_t^+ \Psi$ are continuous, we deduce that $\partial_t \Psi$ exists and is continuous on $[0, T[$ (see for example Corollary 1.2, Chap. 2, in [32]). Moreover, from the representation formula (50) we see that $\partial_t \Psi$ exists and is continuous up to time T. Furthermore, from the expression of $D^2_{x_i x_j}\Psi$, we see that

$$\partial_t \Psi(t, x_1, \dots, x_N) \;=\; -\frac{1}{2}\sum_{i,j=1}^{N} \varphi_i(t)\varphi_j(t) D^2_{x_i x_j}\Psi(t, x_1, \dots, x_N).$$

Therefore, $\Psi \in C^{1,2}([0,T]\times\mathbb{R}^N)$ and is a classical solution to the Cauchy problem

$$\begin{cases} \partial_t \Psi(t,\mathbf{x}) + \frac{1}{2}\sum_{i,j=1}^N \varphi_i(t)\varphi_j(t) D^2_{x_i x_j}\Psi(t,\mathbf{x}) = 0, & \forall\,(t,\mathbf{x}) \in [0,T[\times\mathbb{R}^N, \\ \Psi(T,\mathbf{x}) = h(\mathbf{x}), & \forall\,\mathbf{x} \in \mathbb{R}^N. \end{cases} \tag{51}$$

Now we express the derivatives of $\mathscr{U}$ in terms of Ψ. We begin noting that, taking into account Proposition 4, we have

$$\int_{[-t,0]} \varphi_i(x+t) d^-\eta(x) \;=\; \eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x)\dot\varphi_i(x+t)dx, \qquad \forall\,\eta \in C([-T,0]).$$

This in turn implies that $\mathscr{U}$ is continuous with respect to the topology of $\mathscr{C}([-T,0])$. Therefore, $\mathscr{U}$ admits a unique extension $u\colon \mathscr{C}([-T,0]) \to \mathbb{R}$, which is given by

$$u(t,\eta) \;=\; \Psi\bigg(t, \int_{[-t,0]} \varphi_1(x+t)d^-\eta(x), \dots, \int_{[-t,0]} \varphi_N(x+t)d^-\eta(x)\bigg),$$

for all $(t,\eta) \in [0,T]\times\mathscr{C}([-T,0])$. We also define the map $\tilde u\colon [0,T]\times\mathscr{C}([-T,0[)\times\mathbb{R} \to \mathbb{R}$ as in (13):

$$\tilde u(t,\gamma,a) \;=\; u(t, \gamma 1_{[-T,0[} + a 1_{\{0\}}) \;=\; \Psi\bigg(t, \dots, a\varphi_i(t) - \int_{-t}^0 \gamma(x)\dot\varphi_i(x+t)dx, \dots\bigg),$$

for all $(t,\gamma,a) \in [0,T]\times\mathscr{C}([-T,0[)\times\mathbb{R}$. Let us evaluate the time derivative $\partial_t\mathscr{U}(t,\eta)$, for a given $(t,\eta) \in [0,T[\times C([-T,0])$:

$$\begin{aligned} \partial_t\mathscr{U}(t,\eta) = {}& \partial_t\Psi\bigg(t, \int_{[-t,0]} \varphi_1(x+t)d^-\eta(x), \dots, \int_{[-t,0]} \varphi_N(x+t)d^-\eta(x)\bigg) \\ & + \sum_{i=1}^N D_{x_i}\Psi\bigg(t, \dots, \int_{[-t,0]} \varphi_i(x+t)d^-\eta(x), \dots\bigg)\partial_t\bigg(\int_{[-t,0]} \varphi_i(x+t)d^-\eta(x)\bigg). \end{aligned}$$

Notice that

$$\begin{aligned} \partial_t\bigg(\int_{[-t,0]} \varphi_i(x+t)d^-\eta(x)\bigg) &= \partial_t\bigg(\eta(0)\varphi(t) - \int_{-t}^0 \eta(x)\dot\varphi_i(x+t)dx\bigg) \\ &= \eta(0)\dot\varphi(t) - \eta(-t)\dot\varphi_i(0^+) - \int_{-t}^0 \eta(x)\ddot\varphi_i(x+t)dx. \end{aligned}$$

Let us proceed with the horizontal derivative. We have

$$\begin{aligned}
D^H \mathscr{U}(t,\eta) \;&=\; D^H u(t,\eta) \;=\; D^H \tilde{u}(t, \eta_{|[-T,0[}, \eta(0)) \\
&= \lim_{\varepsilon\to 0^+} \frac{\tilde{u}(t, \eta_{|[-T,0[}(\cdot), \eta(0)) - \tilde{u}(t, \eta_{|[-T,0[}(\cdot - \varepsilon), \eta(0))}{\varepsilon} \\
&= \lim_{\varepsilon\to 0^+} \bigg(\frac{1}{\varepsilon}\Psi\bigg(t,\ldots,\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x)\dot{\varphi}_i(x+t)dx,\ldots\bigg) \\
&\quad - \frac{1}{\varepsilon}\Psi\bigg(t,\ldots,\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x-\varepsilon)\dot{\varphi}_i(x+t)dx,\ldots\bigg)\bigg).
\end{aligned}$$

From the fundamental theorem of calculus, we obtain

$$\begin{aligned}
&\frac{1}{\varepsilon}\Psi\bigg(t,\ldots,\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x)\dot{\varphi}_i(x+t)dx,\ldots\bigg) \\
&- \frac{1}{\varepsilon}\Psi\bigg(t,\ldots,\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x-\varepsilon)\dot{\varphi}_i(x+t)dx,\ldots\bigg) \\
&= \frac{1}{\varepsilon}\int_0^\varepsilon \sum_{i=1}^N D_{x_i}\Psi\bigg(t,\ldots,\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x-y)\dot{\varphi}_i(x+t)dx,\ldots\bigg)\,\partial_y\bigg(\eta(0)\varphi_i(t) \\
&\quad - \int_{-t}^0 \eta(x-y)\dot{\varphi}_i(x+t)dx\bigg)dy.
\end{aligned}$$

Notice that

$$\begin{aligned}
&\partial_y\bigg(\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x-y)\dot{\varphi}_i(x+t)dx\bigg) \;=\; -\partial_y\bigg(\int_{-t-y}^{-y} \eta(x)\dot{\varphi}_i(x+y+t)dx\bigg) \\
&\quad = -\bigg(\eta(-y)\dot{\varphi}_i(t) - \eta(-t-y)\dot{\varphi}_i(0^+) + \int_{-t-y}^{-y} \eta(x)\ddot{\varphi}_i(x+y+t)dx\bigg).
\end{aligned}$$

Therefore

$$\begin{aligned}
&D^H \mathscr{U}(t,\eta) \\
&= -\lim_{\varepsilon\to 0^+} \frac{1}{\varepsilon}\int_0^\varepsilon \sum_{i=1}^N D_{x_i}\Psi\bigg(t,\ldots,\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x-y)\dot{\varphi}_i(x+t)dx,\ldots\bigg)\bigg(\eta(-y)\dot{\varphi}_i(t) \\
&\quad - \eta(-t-y)\dot{\varphi}_i(0^+) + \int_{-t-y}^{-y} \eta(x)\ddot{\varphi}_i(x+y+t)dx\bigg)dy \\
&= -\sum_{i=1}^N D_{x_i}\Psi\bigg(t,\ldots,\eta(0)\varphi_i(t) - \int_{-t}^0 \eta(x)\dot{\varphi}_i(x+t)dx,\ldots\bigg)\bigg(\eta(0)\dot{\varphi}(t) - \eta(-t)\dot{\varphi}_i(0^+) \\
&\quad - \int_{-t}^0 \eta(x)\ddot{\varphi}_i(x+t)dx\bigg).
\end{aligned}$$

Finally, concerning the vertical derivative we have

$$\begin{aligned} D^V \mathscr{U}(t,\eta) &= D^V u(t,\eta) = \partial_a \tilde{u}(t, \eta 1_{[-T,0[} + \eta(0)1_{\{0\}}) \\ &= \sum_{i=1}^N D_{x_i}\Psi\biggl(t, \int_{[-t,0]} \varphi_1(x+t)d^-\eta(x), \ldots\biggr)\varphi_i(t) \end{aligned}$$

and

$$\begin{aligned} D^{VV} \mathscr{U}(t,\eta) &= D^{VV} u(t,\eta) = \partial^2_{aa} \tilde{u}(t, \eta 1_{[-T,0[} + \eta(0)1_{\{0\}}) \\ &= \sum_{i,j=1}^N D^2_{x_i x_j}\Psi\biggl(t, \int_{[-t,0]} \varphi_1(x+t)d^-\eta(x), \ldots\biggr)\varphi_i(t)\varphi_j(t). \end{aligned}$$

From the regularity of Ψ it follows that $\mathscr{U} \in C^{1,2}(([0,T] \times \text{past}) \times \text{present})$. Moreover, since Ψ satisfies the Cauchy problem (51), we conclude that $\partial_t \mathscr{U}(t,\eta) + D^H \mathscr{U}(t,\eta) + \frac{1}{2} D^{VV}\mathscr{U}(t,\eta) = 0$, for all $(t,\eta) \in [0,T[\times C([-T,0])$, therefore $\mathscr{U}$ is a classical solution to the path-dependent heat Eq. (43).

3.2 Towards a Weaker Notion of Solution: A Significant Hedging Example

In the present subsection, we consider Eq. (43) in the case $F \equiv 0$. This situation is particularly interesting, since it arises, for example, in hedging problems of path-dependent contingent claims. More precisely, consider a real continuous finite quadratic variation process X on $(\Omega, \mathscr{F}, \mathbb{P})$ and denote $\mathbb{X}$ the window process associated to X. Let us assume that $[X]_t = t$, for any $t \in [0,T]$. The hedging problem that we have in mind is the following: given a contingent claim's payoff $H(\mathbb{X}_T)$, is it possible to have

$$H(\mathbb{X}_T) = H_0 + \int_0^T Z_t \, d^- X_t, \tag{52}$$

for some $H_0 \in \mathbb{R}$ and some $\mathbb{F}$-adapted process $Z = (Z_t)_{t\in[0,T]}$ such that $Z_t = v(t, \mathbb{X}_t)$, with $v\colon [0,T] \times C([-T,0]) \to \mathbb{R}$? When X is a Brownian motion W and $\int_0^T |Z_t|^2 dt < \infty$, $\mathbb{P}$-a.s., the previous forward integral is an Itô integral. If H is regular enough and it is cylindrical in the sense of (47), we know from Theorem 5 that there exists a unique classical solution $\mathscr{U}\colon [0,T]\times C([-T,0]) \to \mathbb{R}$ to Eq. (43).

Then, we see from Itô's formula (28) that $\mathscr{U}$ satisfies, $\mathbb{P}$-a.s.,

$$\mathscr{U}(t, \mathbb{X}_t) = \mathscr{U}(0, \mathbb{X}_0) + \int_0^t D^V \mathscr{U}(s, \mathbb{X}_s)\, d^- X_s, \qquad 0 \le t \le T. \tag{53}$$

In particular, (52) holds with $Z_t = D^V \mathscr{U}(t, \mathbb{X}_t)$, for any $t \in [0,T]$, $H_0 = \mathscr{U}(0, \mathbb{X}_t)$.

However, a significant hedging example is the *lookback-type payoff*

$$H(\eta) = \sup_{x\in[-T,0]} \eta(x), \qquad \forall\, \eta \in C([-T,0]).$$

We look again for $\mathscr{U}\colon [0,T]\times C([-T,0]) \to \mathbb{R}$ which verifies (53), at least for X being a Brownian motion W. Since $\mathscr{U}(t,\mathbb{W}_t)$ has to be a martingale, a candidate for $\mathscr{U}$ is $\mathscr{U}(t,\eta) = \mathbb{E}[H(\mathbb{W}_T^{t,\eta})]$, for all $(t,\eta)\in[0,T]\times C([-T,0])$. However, this latter $\mathscr{U}$ can be shown not to be regular enough in order to be a classical solution to Eq. (43), even if it is "virtually" a solution to the path-dependent semilinear Kolmogorov equation (43). This will lead us to introduce a weaker notion of solution to Eq. (43). To characterize the map $\mathscr{U}$, we notice that it admits the probabilistic representation formula, for all $(t,\eta)\in[0,T]\times C([-T,0])$,

$$\begin{aligned}\mathscr{U}(t,\eta) &= \mathbb{E}\big[H(\mathbb{W}_T^{t,\eta})\big] = \mathbb{E}\Big[\sup_{-T\le x\le 0} \mathbb{W}_T^{t,\eta}(x)\Big] \\ &= \mathbb{E}\Big[\Big(\sup_{-t\le x\le 0}\eta(x)\Big)\vee\Big(\sup_{t\le x\le T}\big(W_x - W_t + \eta(0)\big)\Big)\Big] = f\Big(t, \sup_{-t\le x\le 0}\eta(x), \eta(0)\Big),\end{aligned}$$

where the function $f\colon [0,T]\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is given by

$$f(t,m,x) = \mathbb{E}\big[m\vee(S_{T-t}+x)\big], \qquad \forall\,(t,m,x)\in[0,T]\times\mathbb{R}\times\mathbb{R}, \tag{54}$$

with $S_t = \sup_{0\le s\le t} W_s$, for all $t\in[0,T]$. Recalling Remark 3, it follows from the presence of $\sup_{-t\le x\le 0}\eta(x)$ among the arguments of f, that $\mathscr{U}$ is not continuous with respect to the topology of $\mathscr{C}([-T,0])$, therefore it can not be a classical solution to Eq. (43). However, we notice that $\sup_{-t\le x\le 0}\eta(x)$ is Lipschitz on $(C([-T,0]),\|\cdot\|_\infty)$, therefore it will follow from Theorem 7 that $\mathscr{U}$ is a strong-viscosity solution to Eq. (43) in the sense of Definition 21. Nevertheless, in this particular case, even if $\mathscr{U}$ is not a classical solution, we shall prove that it is associated to the classical solution of a certain finite dimensional PDE. To this end, we begin computing an explicit form for f, for which it is useful to recall the following standard result.

Lemma 1 (Reflection principle) *For every $a>0$ and $t>0$,*

$$\mathbb{P}(S_t\ge a) = \mathbb{P}(|B_t|\ge a).$$

In particular, for each t, the random variables S_t and $|B_t|$ have the same law, whose density is given by:

$$\varphi_t(z) = \sqrt{\frac{2}{\pi t}}\, e^{-\frac{z^2}{2t}}\, 1_{[0,\infty[}(z), \qquad \forall\, z\in\mathbb{R}.$$

Proof See Proposition 3.7, Chapter III, in [35]. □

From Lemma 1 it follows that, for all $(t, m, x) \in [0, T[\times \mathbb{R} \times \mathbb{R}$,

$$f(t,m,x) \;=\; \int_0^\infty m \vee (z+x)\,\varphi_{T-t}(z)dz \;=\; \int_0^\infty m \vee (z+x)\frac{2}{\sqrt{T-t}}\varphi\Big(\frac{z}{\sqrt{T-t}}\Big)dz,$$

where $\varphi(z) = \exp(z^2/2)/\sqrt{2\pi}$, $z \in \mathbb{R}$, is the standard Gaussian density.

Lemma 2 *The function f defined in (54) is given by, for all $(t, m, x) \in [0, T[\times \mathbb{R} \times \mathbb{R}$,*

$$f(t,m,x) \;=\; 2m\Big(\Phi\Big(\frac{m-x}{\sqrt{T-t}}\Big)-\frac{1}{2}\Big)+2x\Big(1-\Phi\Big(\frac{m-x}{\sqrt{T-t}}\Big)\Big)+\sqrt{\frac{2(T-t)}{\pi}}e^{-\frac{(m-x)^2}{2(T-t)}},$$

for $x \le m$, and

$$f(t,x,m) \;=\; x+\sqrt{\frac{2(T-t)}{\pi}},$$

for $x > m$, where $\Phi(y) = \int_{-\infty}^{y} \varphi(z)dz$, $y \in \mathbb{R}$, is the standard Gaussian cumulative distribution function.

Proof *First case: $x \le m$*. We have

$$f(t,m,x) \;=\; \int_0^{m-x} m\frac{2}{\sqrt{T-t}}\varphi\Big(\frac{z}{\sqrt{T-t}}\Big)dz+\int_{m-x}^\infty (z+x)\frac{2}{\sqrt{T-t}}\varphi\Big(\frac{z}{\sqrt{T-t}}\Big)dz. \quad (55)$$

The first integral on the right-hand side of (55) becomes

$$\int_0^{m-x} m\frac{2}{\sqrt{T-t}}\varphi\Big(\frac{z}{\sqrt{T-t}}\Big)dz \;=\; 2m\int_0^{\frac{m-x}{\sqrt{T-t}}}\varphi(z)dz \;=\; 2m\Big(\Phi\Big(\frac{m-x}{\sqrt{T-t}}\Big)-\frac{1}{2}\Big),$$

where $\Phi(y) = \int_{-\infty}^{y} \varphi(z)dz$, $y \in \mathbb{R}$, is the standard Gaussian cumulative distribution function. Concerning the second integral in (55), we have

$$\begin{aligned}\int_{m-x}^\infty (z+x)\frac{2}{\sqrt{T-t}}\varphi\Big(\frac{z}{\sqrt{T-t}}\Big)dz &= 2\sqrt{T-t}\int_{\frac{m-x}{\sqrt{T-t}}}^\infty z\varphi(z)dz+2x\int_{\frac{m-x}{\sqrt{T-t}}}^\infty \varphi(z)dz\\ &= \sqrt{\frac{2(T-t)}{\pi}}e^{-\frac{(m-x)^2}{2(T-t)}}+2x\Big(1-\Phi\Big(\frac{m-x}{\sqrt{T-t}}\Big)\Big).\end{aligned}$$

Second case: $x > m$. We have

$$
\begin{aligned}
f(t,m,x) = &\int_0^\infty (z+x)\frac{2}{\sqrt{T-t}}\varphi\Big(\frac{z}{\sqrt{T-t}}\Big)dz \\
= &\ 2\sqrt{T-t}\int_0^\infty z\varphi(z)dz + 2x\int_0^\infty \varphi(z)dz \ = \ \sqrt{\frac{2(T-t)}{\pi}} + x.
\end{aligned}
$$

We also have the following regularity result regarding the function f.

Lemma 3 *The function f defined in* (54) *is continuous on* $[0,T]\times\mathbb{R}\times\mathbb{R}$*, moreover it is once* (*resp. twice*) *continuously differentiable in* (t,m) (*resp. in* x) *on* $[0,T[\times\overline{Q}$*, where* $\overline{Q}$ *is the closure of the set* $Q := \{(m,x)\in\mathbb{R}\times\mathbb{R}\colon m > x\}$*. In addition, the following Itô formula holds:*

$$
\begin{aligned}
f(t,S_t,B_t) = &\ f(0,0,0) + \int_0^t \Big(\partial_t f(s,S_s,B_s) + \frac{1}{2}\partial^2_{xx} f(s,S_s,B_s)\Big)ds \\
&+ \int_0^t \partial_m f(s,S_s,B_s)dS_s + \int_0^t \partial_x f(s,S_s,B_s)dB_s, \qquad 0 \le t \le T,\ \mathbb{P}\text{-a.s.}
\end{aligned}
\tag{56}
$$

Proof The regularity properties of f are deduced from its explicit form derived in Lemma 2, after straightforward calculations. Concerning Itô's formula (56), the proof can be done along the same lines as the standard Itô formula. We simply notice that, in the present case, only the restriction of f to $\overline{Q}$ is smooth. However, the process $((S_t,B_t))_t$ is $\overline{Q}$-valued. It is well-known that if $\overline{Q}$ would be an open set, then Itô's formula would hold. In our case, $\overline{Q}$ is the closure of its interior Q. This latter property is enough for the validity of Itô's formula. In particular, the basic tools for the proof of Itô's formula are the following Taylor expansions for the function f:

$$
\begin{aligned}
f(t',m,x) = &\ f(t,m,x) + \partial_t f(t,m,x)(t'-t) \\
&+ \int_0^1 \partial_t f(t+\lambda(t'-t),m,x)(t'-t)d\lambda, \\
f(t,m',x) = &\ f(t,m,x) + \partial_m f(t,m,x)(m'-m) \\
&+ \int_0^1 \partial_m f(t,m+\lambda(m'-m),x)(m'-m)d\lambda, \\
f(t,m,x') = &\ f(t,m,x) + \partial_x f(t,m,x)(x'-x) + \frac{1}{2}\partial^2_{xx} f(t,m,x)(x'-x)^2 \\
&+ \int_0^1 (1-\lambda)\big(\partial^2_{xx} f(t,m,x+\lambda(x'-x)) - \partial^2_{xx} f(t,m,x)\big)(x'-x)^2 d\lambda,
\end{aligned}
$$

for all $(t,m,x)\in[0,T]\times\overline{Q}$. To prove the above Taylor formulae, note that they hold on the open set Q, using the regularity of f. Then, we can extend them to the closure of Q, since f and its derivatives are continuous on $\overline{Q}$. Consequently, Itô's formula can be proved in the usual way. □

Even though, as already observed, $\mathscr{U}$ does not belong to $C^{1,2}(([0,T[\times\text{past})\times\text{present})\cap C([0,T]\times C([-T,0]))$, so that it can not be a classical solution to Eq. (43), the function f is a solution to a certain Cauchy problem, as stated in the following proposition.

Proposition 10 *The function f defined in* (54) *solves the backward heat equation:*

$$\begin{cases} \partial_t f(t,m,x) + \frac{1}{2}\partial_{xx}^2 f(t,m,x) \;=\; 0, & \forall\,(t,m,x) \in [0,T[\times\overline{Q}, \\ f(T,m,x) \;=\; m, & \forall\,(m,x) \in \overline{Q}. \end{cases}$$

Proof We provide two distinct proofs.

Direct proof. Since we know the explicit expression of f, we can derive the form of $\partial_t f$ and $\partial_{xx}^2 f$ by direct calculations:

$$\partial_t f(t,m,x) \;=\; -\frac{1}{\sqrt{T-t}}\varphi\Big(\frac{m-x}{\sqrt{T-t}}\Big), \qquad \partial_{xx}^2 f(t,m,x) \;=\; \frac{2}{\sqrt{T-t}}\varphi\Big(\frac{m-x}{\sqrt{T-t}}\Big),$$

for all $(t,m,x)\in[0,T[\times\overline{Q}$, from which the claim follows.
Probabilistic proof. By definition, the process $(f(t,S_t,B_t))_{t\in[0,T]}$ is given by:

$$f(t,S_t,B_t) \;=\; \mathbb{E}\big[S_T\big|\mathscr{F}_t\big],$$

so that it is a uniformly integrable $\mathbb{F}$-martingale. Then, it follows from Itô's formula (56) that

$$\int_0^t \Big(\partial_t f(s,S_s,B_s) + \frac{1}{2}\partial_{xx}^2 f(s,S_s,B_s)\Big)ds + \int_0^t \partial_m f(s,S_s,B_s)dS_s \;=\; 0,$$

for all $0\leq t\leq T$, $\mathbb{P}$-almost surely. As a consequence, the claim follows if we prove that

$$\int_0^t \partial_m f(s,S_s,B_s)dS_s \;=\; 0. \tag{57}$$

By direct calculation, we have

$$\partial_m f(t,m,x) \;=\; 2\Phi\Big(\frac{m-x}{\sqrt{T-t}}\Big) - 1, \qquad \forall(t,m,x)\in[0,T[\times\overline{Q}.$$

Therefore, (57) becomes

$$\int_0^t \Big(2\Phi\Big(\frac{S_s-B_s}{\sqrt{T-s}}\Big) - 1\Big)dS_s \;=\; 0. \tag{58}$$

Now we observe that the local time of $S_s - B_s$ is equal to $2S_s$, see Exercise 2.14 in [35]. It follows that the measure dS_s is carried by $\{s\colon S_s - B_s = 0\}$. This in turn implies the validity of (58), since the integrand in (58) is zero on the set $\{s\colon S_s - B_s = 0\}$.
□

3.3 Strong-Viscosity Solutions

Motivated by previous subsection and following [10], we now introduce a concept of weak (viscosity type) solution for the path-dependent Eq. (43), which we call *strong-viscosity solution* to distinguish it from the classical notion of viscosity solution.

Definition 21 A function $\mathscr{U}\colon [0,T] \times C([-T,0]) \to \mathbb{R}$ is called **strong-viscosity solution** to Eq. (43) if there exists a sequence $(\mathscr{U}_n, H_n, F_n)_n$ of Borel measurable functions $\mathscr{U}_n\colon [0,T] \times C([-T,0]) \to \mathbb{R}$, $H_n\colon C([-T,0]) \to \mathbb{R}$, $F_n\colon [0,T] \times C([-T,0]) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, satisfying the following.

(i) For all $t \in [0,T]$, the functions $\mathscr{U}_n(t,\cdot)$, $H_n(\cdot)$, $F_n(t,\cdot,\cdot,\cdot)$ are equicontinuous on compact sets and, for some positive constants C and m,

$$\begin{aligned} |F_n(t,\eta,y,z) - F_n(t,\eta,y',z')| &\leq C(|y-y'| + |z-z'|), \\ |\mathscr{U}_n(t,\eta)| + |H_n(\eta)| + |F_n(t,\eta,0,0)| &\leq C\big(1 + \|\eta\|_\infty^m\big), \end{aligned}$$

for all $(t,\eta) \in [0,T] \times C([-T,0])$, $y, y' \in \mathbb{R}$, and $z, z' \in \mathbb{R}$.

(ii) $\mathscr{U}_n$ is a strict solution to

$$\begin{cases} \partial_t \mathscr{U}_n + D^H \mathscr{U}_n + \frac{1}{2} D^{VV} \mathscr{U}_n + F_n(t,\eta,\mathscr{U}_n, D^V \mathscr{U}_n) = 0, & \forall\, (t,\eta) \in [0,T[\times C([-T,0]), \\ \mathscr{U}_n(T,\eta) = H_n(\eta), & \forall\, \eta \in C([-T,0]). \end{cases}$$

(iii) $(\mathscr{U}_n, H_n, F_n)$ converges pointwise to $(\mathscr{U}, H, F)$ as n tends to infinity.

Remark 12 (i) Notice that in [8], Definition 3.4, instead of the equicontinuity on compact sets we supposed the local equicontinuity, i.e., the equicontinuity on bounded sets (see Definition 3.3 in [8]). This latter condition is stronger when $\mathscr{U}$ (as well as the other coefficients) is defined on a non-locally compact topological space, as for example $[0,T] \times C([-T,0])$.
(ii) We observe that, for every $t \in [0,T]$, the equicontinuity on compact sets of $(\mathscr{U}_n(t,\cdot))_n$ together with its pointwise convergence to $\mathscr{U}(t,\cdot)$ is equivalent to requiring the uniform convergence on compact sets of $(\mathscr{U}_n(t,\cdot))_n$ to $\mathscr{U}(t,\cdot)$. The same remark applies to $(H_n(\cdot))_n$ and $(F_n(t,\cdot,\cdot,\cdot))_n$, $t \in [0,T]$. □

The following uniqueness result for strong-viscosity solution holds.

Theorem 6 *Suppose that Assumption* **(A)** *holds. Let* $\mathscr{U}\colon [0,T]\times C([-T,0]) \to \mathbb{R}$ *be a strong-viscosity solution to Eq.* (43)*. Then, we have*

$$\mathscr{U}(t,\eta) \;=\; Y_t^{t,\eta}, \qquad \forall\,(t,\eta) \in [0,T]\times C([-T,0]),$$

where $(Y_s^{t,\eta}, Z_s^{t,\eta})_{s\in[t,T]} \in \mathbb{S}^2(t,T)\times\mathbb{H}^2(t,T)$*, with* $Y_s^{t,\eta} = \mathscr{U}(s,\mathbb{W}_s^{t,\eta})$*, solves the backward stochastic differential equation,* $\mathbb{P}$*-a.s.,*

$$Y_s^{t,\eta} \;=\; H(\mathbb{W}_T^{t,\eta}) + \int_s^T F(r,\mathbb{W}_r^{t,\eta},Y_r^{t,\eta},Z_r^{t,\eta})dr - \int_s^T Z_r^{t,\eta}dW_r, \qquad t \le s \le T.$$

In particular, there exists at most one strong-viscosity solution to Eq. (43).

Proof Consider a sequence $(\mathscr{U}_n,H_n,F_n)_n$ satisfying conditions (i)-(iii) of Definition 21. For every $n\in\mathbb{N}$ and any $(t,\eta)\in[0,T]\times C([-T,0])$, we know from Theorem 4 that $(Y_s^{n,t,\eta},Z_s^{n,t,\eta})_{s\in[t,T]} = (\mathscr{U}_n(s,\mathbb{W}_s^{t,\eta}), D^V\mathscr{U}_n(s,\mathbb{W}_s^{t,\eta}))_{s\in[t,T]} \in \mathbb{S}^2(t,T)\times\mathbb{H}^2(t,T)$ is the solution to the backward stochastic differential equation, $\mathbb{P}$-a.s.,

$$Y_s^{n,t,\eta} \;=\; H_n(\mathbb{W}_T^{t,\eta}) + \int_s^T F_n(r,\mathbb{W}_r^{t,\eta},Y_r^{n,t,\eta},Z_r^{n,t,\eta})dr - \int_s^T Z_r^{n,t,\eta}dW_r, \qquad t \le s \le T.$$

Our aim is to pass to the limit in the above equation as $n\to\infty$, using Theorem C.1 in [10]. From the polynomial growth condition of $(\mathscr{U}_n)_n$ and estimate (46), we see that

$$\sup_n \|Y^{n,t,\eta}\|_{\mathbb{S}^p(t,T)} \;<\; \infty, \qquad \text{for any } p\ge 1.$$

This implies, using standard estimates for backward stochastic differential equations (see, e.g., Proposition B.1 in [10]) and the polynomial growth condition of $(F_n)_n$, that

$$\sup_n \|Z^{n,t,\eta}\|_{\mathbb{H}^2(t,T)} \;<\; \infty.$$

Let $Y_s^{t,\eta} = \mathscr{U}(s,\mathbb{W}_s^{t,\eta})$, for any $s\in[t,T]$. Then, we see that all the requirements of Theorem C.1 in [10] follow by assumptions and estimate (46), so the claim follows. □

We now prove an existence result for strong-viscosity solutions to the path-dependent heat equation, namely to Eq. (43) in the case $F\equiv 0$. To this end, we need the following stability result for strong-viscosity solutions.

Lemma 4 *Let* $(\mathscr{U}_{n,k},H_{n,k},F_{n,k})_{n,k}$*,* $(\mathscr{U}_n,H_n,F_n)_n$*,* $(\mathscr{U},H,F)$ *be Borel measurable functions such that the properties below hold.*

(i) For all $t\in[0,T]$*, the functions* $\mathscr{U}_{n,k}(t,\cdot)$*,* $H_{n,k}(\cdot)$*, and* $F_{n,k}(t,\cdot,\cdot,\cdot)$*,* $n,k\in\mathbb{N}$*, are equicontinuous on compact sets and, for some positive constants* C *and* m*,*

$$|F_{n,k}(t,\eta,y,z) - F_{n,k}(t,\eta,y',z')| \leq C(|y-y'| + |z-z'|),$$
$$|\mathscr{U}_{n,k}(t,\eta)| + |H_{n,k}(\eta)| + |F_{n,k}(t,\eta,0,0)| \leq C\big(1 + \|\eta\|_\infty^m\big),$$

for all $(t,\eta) \in [0,T] \times C([-T,0])$, $y, y' \in \mathbb{R}$, *and* $z, z' \in \mathbb{R}$.

(ii) $\mathscr{U}_{n,k}$ *is a strict solution to*

$$\begin{cases} \partial_t \mathscr{U}_{n,k} + D^H \mathscr{U}_{n,k} + \frac{1}{2} D^{VV} \mathscr{U}_{n,k} & \\ + F_{n,k}(t,\eta,\mathscr{U}_{n,k}, D^V \mathscr{U}_{n,k}) \; = \; 0, & \forall\, (t,\eta) \in [0,T[\times C([-T,0]), \\ \mathscr{U}_{n,k}(T,\eta) \; = \; H_{n,k}(\eta), & \forall\, \eta \in C([-T,0]). \end{cases}$$

(iii) $(\mathscr{U}_{n,k}, H_{n,k}, F_{n,k})$ *converges pointwise to* $(\mathscr{U}_n, H_n, F_n)$ *as* k *tends to infinity.*

(iv) $(\mathscr{U}_n, H_n, F_n)$ *converges pointwise to* $(\mathscr{U}, H, F)$ *as* n *tends to infinity.*

Then, there exists a subsequence $(\mathscr{U}_{n,k_n}, H_{n,k_n}, F_{n,k_n})_n$ *which converges pointwise to* $(\mathscr{U}, H, F)$ *as* n *tends to infinity. In particular,* $\mathscr{U}$ *is a strong-viscosity solution to Eq.* (43).

Proof See Lemma 3.4 in [8] or Lemma 3.1 in [10]. We remark that in [8] a slightly different definition of strong-viscosity solution was used, see Remark 12(i); however, proceeding along the same lines we can prove the present result.

Theorem 7 *Suppose that Assumption* **(A)** *holds. Let* $F \equiv 0$ *and* H *be continuous. Then, there exists a unique strong-viscosity solution* $\mathscr{U}$ *to the path-dependent heat Eq.* (43), *which is given by*

$$\mathscr{U}(t,\eta) \; = \; \mathbb{E}\big[H(\mathbb{W}_T^{t,\eta})\big], \qquad \forall\, (t,\eta) \in [0,T] \times C([-T,0]).$$

Proof Let $(e_i)_{i\geq 0}$ be the orthonormal basis of $L^2([-T,0])$ composed by the functions

$$e_0 = \frac{1}{\sqrt{T}}, \qquad e_{2i-1}(x) = \sqrt{\frac{2}{T}} \sin\left(\frac{2\pi}{T}(x+T)i\right), \qquad e_{2i}(x) = \sqrt{\frac{2}{T}} \cos\left(\frac{2\pi}{T}(x+T)i\right),$$

for all $i \in \mathbb{N}\backslash\{0\}$. Let us define the linear operator $\Lambda\colon C([-T,0]) \to C([-T,0])$ by

$$(\Lambda\eta)(x) \; = \; \frac{\eta(0) - \eta(-T)}{T} x, \qquad x \in [-T,0],\ \eta \in C([-T,0]).$$

Notice that $(\eta - \Lambda\eta)(-T) = (\eta - \Lambda\eta)(0)$, therefore $\eta - \Lambda\eta$ can be extended to the entire real line in a periodic way with period T, so that we can expand it in Fourier series. In particular, for each $n \in \mathbb{N}$ and $\eta \in C([-T,0])$, consider the Fourier partial sum

$$s_n(\eta - \Lambda\eta) \; = \; \sum_{i=0}^n (\eta_i - (\Lambda\eta)_i) e_i, \qquad \forall\, \eta \in C([-T,0]), \tag{59}$$

where (denoting $\tilde{e}_i(x) = \int_{-T}^x e_i(y)dy$, for any $x \in [-T, 0]$), by Proposition 4,

$$\begin{aligned} \eta_i &= \int_{-T}^0 \eta(x)e_i(x)dx = \eta(0)\tilde{e}_i(0) - \int_{[-T,0]} \tilde{e}_i(x)d^-\eta(x) \\ &= \int_{[-T,0]} (\tilde{e}_i(0) - \tilde{e}_i(x))d^-\eta(x), \end{aligned} \tag{60}$$

since $\eta(0) = \int_{[-T,0]} d^-\eta(x)$. Moreover we have

$$(\Lambda\eta)_i = \int_{-T}^0 (\Lambda\eta)(x)e_i(x)dx = \frac{1}{T}\int_{-T}^0 xe_i(x)dx\bigg(\int_{[-T,0]} d^-\eta(x) - \eta(-T)\bigg). \tag{61}$$

Define

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_n}{n+1}.$$

Then, by (59),

$$\sigma_n(\eta - \Lambda\eta) = \sum_{i=0}^n \frac{n+1-i}{n+1}(\eta_i - (\Lambda\eta)_i)e_i, \qquad \forall\, \eta \in C([-T, 0]).$$

We know from Fejér's theorem on Fourier series (see, e.g., Theorem 3.4, Chapter III, in [44]) that, for any $\eta \in C([-T, 0])$, $\sigma_n(\eta - \Lambda\eta) \to \eta - \Lambda\eta$ uniformly on $[-T, 0]$, as n tends to infinity, and $\|\sigma_n(\eta - \Lambda\eta)\|_\infty \le \|\eta - \Lambda\eta\|_\infty$. Let us define the linear operator $T_n\colon C([-T, 0]) \to C([-T, 0])$ by (denoting $e_{-1}(x) = x$, for any $x \in [-T, 0]$)

$$\begin{aligned} T_n\eta &= \sigma_n(\eta - \Lambda\eta) + \Lambda\eta = \sum_{i=0}^n \frac{n+1-i}{n+1}(\eta_i - (\Lambda\eta)_i)e_i + \frac{\eta(0) - \eta(-T)}{T}e_{-1} \\ &= \sum_{i=0}^n \frac{n+1-i}{n+1}x_ie_i + x_{-1}e_{-1}, \end{aligned} \tag{62}$$

where, using (60) and (61),

$$\begin{aligned} x_{-1} &= \int_{[-T,0]} \frac{1}{T}d^-\eta(x) - \frac{1}{T}\eta(-T), \\ x_i &= \int_{[-T,0]} \bigg(\tilde{e}_i(0) - \tilde{e}_i(x) - \frac{1}{T}\int_{-T}^0 xe_i(x)dx\bigg)d^-\eta(x) + \frac{1}{T}\int_{-T}^0 xe_i(x)dx\,\eta(-T), \end{aligned}$$

for $i = 0, \dots, n$. Then, for any $\eta \in C([-T, 0])$, $T_n\eta \to \eta$ uniformly on $[-T, 0]$, as n tends to infinity. Furthermore, there exists a positive constant M such that

$$\|T_n\eta\|_\infty \;\le\; M\|\eta\|_\infty, \qquad \forall\, n \in \mathbb{N},\ \forall\, \eta \in C([-T, 0]). \tag{63}$$

In particular, the family of linear operators $(T_n)_n$ is equicontinuous. Now, let us define $\bar{H}_n \colon C([-T, 0]) \to \mathbb{R}$ as follows

$$\bar{H}_n(\eta) \;=\; H(T_n\eta), \qquad \forall\, \eta \in C([-T, 0]).$$

We see from (63) that the family $(\bar{H}_n)_n$ is equicontinuous on compact sets. Moreover, from the polynomial growth condition of H and (63) we have

$$|\bar{H}_n(\eta)| \;\le\; C(1 + \|T_n\eta\|_\infty^m) \;\le\; C(1 + M^m\|\eta\|_\infty^m), \qquad \forall\, n \in \mathbb{N},\ \forall\, \eta \in C([-T, 0]).$$

Now, we observe that since $\{e_{-1}, e_0, e_1, \dots, e_n\}$ are linearly independent, then we see from (62) that $T_n\eta$ is completely characterized by the coefficients of $e_{-1}, e_0, e_1, \dots, e_n$. Therefore, the function $\bar{h}_n \colon \mathbb{R}^{n+2} \to \mathbb{R}$ given by

$$\bar{h}_n(x_{-1}, \dots, x_n) \;=\; \bar{H}_n(\eta) \;=\; H\biggl(\sum_{i=0}^{n} \frac{n+1-i}{n+1} x_i e_i + x_{-1}e_{-1}\biggr), \quad \forall\, (x_{-1}, \dots, x_n) \in \mathbb{R}^{n+2},$$

completely characterizes $\bar{H}_n$. Moreover, fix $\eta \in C([-T, 0])$ and consider the corresponding coefficients $x_{-1}, \dots, x_n$ with respect to $\{e_{-1}, \dots, e_n\}$ in the expression (62) of $T_n\eta$. Set

$$\varphi_{-1}(x) = \frac{1}{T}, \qquad \varphi_i(x) \;=\; \tilde{e}_i(0) - \tilde{e}_i(x - T) - \frac{1}{T}\int_{-T}^{0} x e_i(x)dx, \qquad x \in [0, T],$$

$$a_{-1} \;=\; -\frac{1}{T}, \qquad a_i \;=\; \frac{1}{T}\int_{-T}^{0} x e_i(x)dx.$$

Notice that $\varphi_{-1}, \dots, \varphi_n \in C^\infty([0, T])$. Then, we have

$$\bar{H}_n(\eta) = \bar{h}_n\biggl(\int_{[-T,0]} \varphi_{-1}(x+T)d^-\eta(x) + a_{-1}\eta(-T), \dots, \int_{[-T,0]} \varphi_n(x+T)d^-\eta(x) + a_n\eta(-T)\biggr).$$

Let $\phi(x) = c\exp(1/(x^2 - T^2))1_{[0,T[}(x)$, $x \ge 0$, with $c > 0$ such that $\int_0^\infty \phi(x)\,dx = 1$. Define, for any $\varepsilon > 0$, $\phi_\varepsilon(x) = \phi(x/\varepsilon)/\varepsilon$, $x \ge 0$. Notice that $\phi_\varepsilon \in C^\infty([0, \infty[)$ and (denoting $\tilde{\phi}_\varepsilon(x) = \int_0^x \phi_\varepsilon(y)dy$, for any $x \ge 0$),

$$\begin{aligned}\int_{-T}^{0} \eta(x)\phi_\varepsilon(x+T)dx &= \eta(0)\tilde{\phi}_\varepsilon(T) - \int_{[-T,0]} \tilde{\phi}_\varepsilon(x+T)d^-\eta(x)\\ &= \int_{[-T,0]} \bigl(\tilde{\phi}_\varepsilon(T) - \tilde{\phi}_\varepsilon(x+T)\bigr)d^-\eta(x).\end{aligned}$$

Therefore

$$\lim_{\varepsilon\to 0^+}\int_{[-T,0]}\big(\tilde\phi_\varepsilon(T) - \tilde\phi_\varepsilon(x+T)\big)d^-\eta(x) \;=\; \lim_{\varepsilon\to 0^+}\int_{-T}^0 \eta(x)\phi_\varepsilon(x+T)dx \;=\; \eta(-T).$$

For this reason, we introduce the function $H_n\colon C([-T,0])\to\mathbb{R}$ given by

$$H_n(\eta) \;=\; \bar h_n\bigg(\ldots,\int_{[-T,0]}\varphi_i(x+T)d^-\eta(x) + a_i\int_{[-T,0]}\big(\tilde\phi_n(T)-\tilde\phi_n(x+T)\big)d^-\eta(x),\ldots\bigg).$$

Now, for any $n\in\mathbb{N}$, let $(h_{n,k})_{k\in\mathbb{N}}$ be a locally equicontinuous sequence of $C^2(\mathbb{R}^{n+2};\mathbb{R})$ functions, uniformly polynomially bounded, such that $h_{n,k}$ converges pointwise to h_n, as k tends to infinity. Define $H_{n,k}\colon C([-T,0])\to\mathbb{R}$ as follows:

$$H_{n,k}(\eta) = h_{n,k}\bigg(\ldots,\int_{[-T,0]}\varphi_i(x+T)d^-\eta(x) + a_i\int_{[-T,0]}\big(\tilde\phi_n(T)-\tilde\phi_n(x+T)\big)d^-\eta(x),\ldots\bigg).$$

Then, we know from Theorem 5 that the function $\mathscr{U}_{n,k}\colon[0,T]\times C([-T,0])\to\mathbb{R}$ given by

$$\mathscr{U}_{n,k}(t,\eta) \;=\; \mathbb{E}\big[H_{n,k}(\mathbb{W}_T^{t,\eta})\big], \qquad \forall\,(t,\eta)\in[0,T]\times C([-T,0])$$

is a classical solution to the path-dependent heat Eq. (43). Moreover, the family $(\mathscr{U}_{n,k})_{n,\varepsilon,k}$ is equicontinuous on compact sets and uniformly polynomially bounded. Then, using the stability result Lemma 4, it follows that $\mathscr{U}$ is a strong-viscosity solution to the path-dependent heat Eq. (43). □

Acknowledgments The present work was partially supported by the ANR Project MASTERIE 2010 BLAN 0121 01. The second named author also benefited partially from the support of the "FMJH Program Gaspard Monge in optimization and operation research" (Project 2014-1607H).

References

1. Buckdahn, R., Ma, J., Zhang, J.: Pathwise Taylor expansions for random fields on multiple dimensional paths. Stoch. Process. Appl. **125**(7), 2820–2855 (2015)
2. Cerrai, S.: Second Order PDE's in Finite and Infinite Dimension: A Probabilistic Approach, Lecture Notes in Mathematics, vol. 1762. Springer, Berlin (2001)
3. Cerutti, M.C., Escauriaza, L., Fabes, E.B.: Uniqueness in the Dirichlet problem for some elliptic operators with discontinuous coefficients. Ann. Mat. Pura Appl. **4**(163), 161–180 (1993)
4. Chojnowska-Michalik, A.: Representation theorem for general stochastic delay equations. Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **26**(7), 635–642 (1978)
5. Cont, R., Fournié, D.-A.: Change of variable formulas for non-anticipative functionals on path space. J. Funct. Anal. **259**(4), 1043–1072 (2010)

6. Cont, R., Fournié, D.-A.: A functional extension of the Itô formula. C. R. Math. Acad. Sci. Paris **348**(1–2), 57–61 (2010)
7. Cont, R., Fournié, D.-A.: Functional Itô calculus and stochastic integral representation of martingales. Ann. Probab. **41**(1), 109–133 (2013)
8. Cosso, A., Russo, F.: A regularization approach to functional Itô calculus and strong-viscosity solutions to path-dependent PDEs. Preprint HAL, hal-00933678 version 2 (2014) (Unpublished)
9. Cosso, A., Russo, F.: Functional Itô versus Banach space stochastic calculus and strict solutions of semilinear path-dependent equations. Preprint arXiv:1505.02926v1 (2015)
10. Cosso, A., Russo, F.: Strong-viscosity solutions: semilinear parabolic PDEs and path-dependent PDEs. Preprint arXiv:1505.02927v1 (2015)
11. Crandall, M.G., Kocan, M., Soravia, P., Święch, A.: On the equivalence of various weak notions of solutions of elliptic PDEs with measurable ingredients. In: Progress in Elliptic and Parabolic Partial Differential Equations (Capri, 1994). Pitman Research Notes in Mathematics Series, vol. 350, pp. 136–162. Longman, Harlow (1996)
12. Di Girolami, C., Fabbri, G., Russo, F.: The covariation for Banach space valued processes and applications. Metrika **77**(1), 51–104 (2014)
13. Di Girolami, C., Russo, F.: Infinite dimensional stochastic calculus via regularization and applications. Preprint HAL-INRIA, inria-00473947 version 1 (2010) (Unpublished)
14. Di Girolami, C., Russo, F.: Clark-Ocone type formula for non-semimartingales with finite quadratic variation. C. R. Math. Acad. Sci. Paris **349**(3–4), 209–214 (2011)
15. Di Girolami, C., Russo, F.: Generalized covariation and extended Fukushima decomposition for Banach space-valued processes. Applications to windows of Dirichlet processes. Infinite Dimens. Anal. Quantum Probab. Relat. Top. **15**(2), 1250007, 50 (2012)
16. Di Girolami, C., Russo, F.: Generalized covariation for Banach space valued processes, Itô formula and applications. Osaka J. Math. **51**(3), 729–783 (2014)
17. Dupire, B.: Functional Itô calculus. Portfolio Research Paper, Bloomberg (2009)
18. Ekren, I., Keller, C., Touzi, N., Zhang, J.: On viscosity solutions of path dependent PDEs. Ann. Probab. **42**(1), 204–236 (2014)
19. Ekren, I., Touzi, N., Zhang, J.: Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part I. Ann. Probab, To appear in (2013)
20. Ekren, I., Touzi, N., Zhang, J.: Viscosity solutions of fully nonlinear parabolic path dependent PDEs: Part II. Ann. Probab, To appear in (2013)
21. Ekren, I., Touzi, N., Zhang, J.: Optimal stopping under nonlinear expectation. Stoch. Process. Appl. **124**(10), 3277–3311 (2014)
22. Flandoli, F., Zanco, G.: An infinite-dimensional approach to path-dependent Kolmogorov's equations. Ann. Probab, To appear in (2013)
23. Föllmer, H.: Calcul d'Itô sans probabilités. In: Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French). Lecture Notes in Mathematics, vol. 850 pp. 143–150. Springer, Berlin (1981)
24. Gozzi, F., Russo, F.: Verification theorems for stochastic optimal control problems via a time dependent Fukushima-Dirichlet decomposition. Stoch. Process. Appl. **116**(11), 1530–1562 (2006)
25. Gozzi, F., Russo, F.: Weak Dirichlet processes with a stochastic control perspective. Stoch. Process. Appl. **116**(11), 1563–1583 (2006)
26. Gradinaru, M., Nourdin, I.: Approximation at first and second order of m-order integrals of the fractional Brownian motion and of certain semimartingales. Electron. J. Probab. **8**(18), 26(electronic) (2003)
27. Jensen, R.: Uniformly elliptic PDEs with bounded, measurable coefficients. J. Fourier Anal. Appl. **2**(3), 237–259 (1996)
28. Jensen, R., Kocan, M., Święch, A.: Good and viscosity solutions of fully nonlinear elliptic equations. Proc. Amer. Math. Soc. **130**(2), 533–542(electronic) (2002)
29. Leão, D., Ohashi, A., Simas, A.B.: Weak functional Itô calculus and applications. Preprint arXiv:1408.1423v2 (2014)

30. Mohammed, S.E.A.: Stochastic Functional Differential Equations. Research Notes in Mathematics, vol. 99. Pitman (Advanced Publishing Program), Boston (1984)
31. Nualart, D.: The Malliavin Calculus and Related Topics. Probability and its Applications. Springer, New York (1995)
32. Pazy, A.: Semigroups of Linear Operators and Applications to Partial Differential Equations. Applied Mathematical Sciences, vol. 44. Springer, New York (1983)
33. Peng, S.: Note on viscosity solution of path-dependent PDE and G-martingales. Preprint arXiv:1106.1144 (2012)
34. Ren, Z., Touzi, N., Zhang, J.: Comparison of viscosity solutions of semi-linear path-dependent PDEs. Preprint arXiv:1410.7281 (2014)
35. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], third edn. Springer, Berlin (1999)
36. Rudin, W.: Real and Complex Analysis, 3rd edn. McGraw-Hill Book Co., New York (1987)
37. Russo, F., Vallois, P.: Intégrales progressive, rétrograde et symétrique de processus non adaptés. C. R. Acad. Sci. Paris Sér. I Math. **312**(8), 615–618 (1991)
38. Russo, F., Vallois, P.: Forward, backward and symmetric stochastic integration. Probab. Theory Relat. Fields **97**(3), 403–421 (1993)
39. Russo, F., Vallois, P.: The generalized covariation process and Itô formula. Stoch. Process. Appl. **59**(1), 81–104 (1995)
40. Russo, F., Vallois, P.: Elements of stochastic calculus via regularization. In: Séminaire de Probabilités XL. Lecture Notes in Mathematics, vol. 1899, pp. 147–185. Springer, Berlin (2007)
41. Ryan, R.A.: Introduction to Tensor Products of Banach Spaces. Springer Monographs in Mathematics. Springer, London (2002)
42. Tang, S., Zhang, F.: Path-dependent optimal stochastic control and associated Bellman equations: a viscosity solution approach. Preprint arXiv:1210.2078 (2013)
43. Yan, F., Mohammed, S.: A stochastic calculus for systems with memory. Stoch. Anal. Appl. **23**(3), 613–657 (2005)
44. Zygmund, A.: Trigonometric Series, vol. I, II, 3rd edn. Cambridge Mathematical Library. Cambridge University Press, Cambridge (2002) (With a foreword by Robert A. Fefferman)

Nonlinear Young Integrals via Fractional Calculus

Yaozhong Hu and Khoa N. Lê

Abstract For Hölder continuous functions $W(t, x)$ and φ_t, we define nonlinear integral $\int_a^b W(dt, \varphi_t)$ via fractional calculus. This nonlinear integral arises naturally in the Feynman-Kac formula for stochastic heat equations with random coefficients (Hu and Lê, Nonlinear Young integrals and differential systems in Hölder media. Trans. Am. Math. Soc. (in press)). We also define iterated nonlinear integrals.

Keywords Nonlinear integration · Young integral · Iterated nonlinear Young integrals

1 Introduction

Let $\{\varphi_t\,, t \geq 0\}$ be a Hölder continuous function and let $\{W(t, x), t \geq 0\,, x \in \mathbb{R}^d\}$ be another jointly Hölder continuous function of several variables (see (10) for the precise statement about the assumption on W). The aim of this paper is to define the nonlinear Young integral $\int_a^b W(dt, \varphi_t)$ by using fractional calculus.

This paper can be considered as supplementary to authors' recent paper [5], where the nonlinear Young integral is introduced to establish the Feynman-Kac formula for general stochastic partial differential equations with random coefficients, namely,

$$\partial_t u(t, x) + Lu(t, x) + u(t, x)\partial_t W(t, x) = 0\,, \tag{1}$$

Y. Hu (✉) · K.N. Lê
Department of Mathematics, University of Kansas, Lawrence, KS 66045, USA
e-mail: yhu@ku.edu

K.N. Lê
e-mail: khoale@ku.edu

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_3

where W is a Hölder continuous function of several variables (which can be a sample path of a Gaussian random field) and

$$Lu(t,x) = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(t,x,W)\partial^2_{x_i x_j} u(t,x) + \sum_{i=1}^{d} b_i(t,x,W)\partial_{x_i} u(t,x)$$

with the coefficients a_{ij} and b_i depending on W. The terminal condition for the Eq. (1) is $u(T,x) = u_T(x)$ for some given function $u_T(x)$.

To motivate our study of the nonlinear Young integral, let us recall a basic result in [5] on the Feynman-Kac formula: Let $\sigma(t,x) = (\sigma_{ij}(t,x,W))_{1\le i,j\le d}$ satisfy $a(t,x,W) = \sigma(t,x,W)\sigma(t,x,W)^T$ (we omit the explicit dependence of σ on W). Consider the following stochastic differential equation

$$dX_t^{r,x} = \sigma(t, X_t^{r,x})\delta B_t + b(t, X_t^{r,x})dt, \quad 0 \le r \le t \le T, \quad X_r^{r,x} = x, \tag{2}$$

where $(B_t, 0 \le t \le T)$ is a standard Brownian motion and δB_t denotes the Itô differential. Then it is proved in [5] that under some conditions b and σ and W (which are verified for certain Gaussian random field W), the nonlinear integral $\int_r^T W(ds, X_s^{r,x})$ is well-defined and exponentially integrable and $u(r,x) = \mathbb{E}^B\left\{u_T(X_T^{r,x})\exp\left[\int_r^T W(ds, X_s^{r,x})\right]\right\}$ is a Feynman-Kac solution to (1) with $u(T,x) = u_T(x)$. One of the main tasks in that paper is the study of the nonlinear Young integral $\int_r^T W(ds, X_s^{r,x})$. To this end we used the Riemann sum approximation and the sewing lemma of [2]. In this paper, we shall study the nonlinear Young integral $\int_a^b W(dt, \varphi_t)$ by means of fractional calculus. This approach may provide more detailed properties of the solutions to the equations (see [6, 7]).

Under certain conditions, we shall prove that the two nonlinear Young integrals, defined by Riemann sums (through sewing lemma) or by fractional calculus, are the same (see Proposition 2).

To expand the solution of a (nonlinear) differential equation with explicit remainder term we need to define (iterated) multiple integrals (see [3]). We shall also give a definition of the iterated nonlinear Young integrals. Some elementary estimates are also obtained.

The paper is organized as follows. Section 2 briefly recalls some preliminary material on fractional calculus that are needed later. Section 3 deals with the nonlinear Young integrals and Sect. 4 is concerned with iterated nonlinear Young integrals.

2 Fractional Integrals and Derivatives

In this section we recall some results from fractional calculus.

Let $-\infty < a < b < \infty$, $\alpha > 0$ and $p \ge 1$ be real numbers. Denote by $L^p(a,b)$ the space of all measurable functions on (a,b) such that

$$\|f\|_p := \left(\int_a^b |f(t)|^p dt \right)^{1/p} < \infty .$$

Denote by $C([a,b])$ the space of continuous functions on $[a,b]$. Let $f \in L^1([a,b])$. The left-sided fractional Riemann-Liouville integral $I_{a+}^{\alpha} f$ is defined as

$$I_{a+}^{\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_a^t (t-s)^{\alpha-1} f(s)\, ds, \quad t \in (a,b) \tag{3}$$

and the right-sided fractional Riemann-Liouville integral $I_{b-}^{\alpha} f$ is defined as

$$I_{b-}^{\alpha} f(t) = \frac{(-1)^{-\alpha}}{\Gamma(\alpha)} \int_t^b (s-t)^{\alpha-1} f(s)\, ds, \quad t \in (a,b) \tag{4}$$

where $(-1)^{-\alpha} = e^{-i\pi\alpha}$ and $\Gamma(\alpha) = \int_0^\infty r^{\alpha-1} e^{-r} dr$ is the Euler gamma function. Let $I_{a+}^{\alpha}(L^p)$ (resp. $I_{b-}^{\alpha}(L^p)$) be the image of $L^p(a,b)$ by the operator I_{a+}^{α} (resp. I_{b-}^{α}). If $f \in I_{a+}^{\alpha}(L^p)$ (resp. $f \in I_{b-}^{\alpha}(L^p)$) and $0 < \alpha < 1$, then the (left-sided or right-sided) Weyl derivatives are defined (respectively) as

$$D_{a+}^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \left(\frac{f(t)}{(t-a)^{\alpha}} + \alpha \int_a^t \frac{f(t)-f(s)}{(t-s)^{\alpha+1}} ds \right) \tag{5}$$

and

$$D_{b-}^{\alpha} f(t) = \frac{(-1)^{\alpha}}{\Gamma(1-\alpha)} \left(\frac{f(t)}{(b-t)^{\alpha}} + \alpha \int_t^b \frac{f(t)-f(s)}{(s-t)^{\alpha+1}} ds \right), \tag{6}$$

where $a \le t \le b$ (the convergence of the integrals at the singularity $s = t$ holds pointwise for almost all $t \in (a,b)$ if $p = 1$ and moreover in L^p-sense if $1 < p < \infty$).

It is clear that if f is Hölder continuous of order $\mu > \alpha$, then the two Weyl derivatives exist.

For any $\beta \in (0,1)$, we denote by $C^{\beta}([a,b])$ the space of β-Hölder continuous functions on the interval $[a,b]$. We will make use of the notation

$$\|f\|_{\beta;a,b} = \sup_{a<\theta<r<b} \frac{|f(r)-f(\theta)|}{|r-\theta|^{\beta}}$$

(which is a seminorm) and

$$\|f\|_{\infty;a,b} = \sup_{a\le r\le b} |f(r)|,$$

where $f : \mathbb{R} \to \mathbb{R}$ is a given continuous function.

It is well-known that $C^{\beta}([a, b])$ with the Hölder norm $\|f\|_{\beta;a,b} + \|f\|_{\infty;a,b}$ is a Banach space. However, it is *not* separable.

Using the fractional calculus, we have (see [9] and also [3])

Proposition 1 *Let* $0 < \alpha < 1$. *If* f *and* g *are continuously differentiable functions on the interval* $[a, b]$, *then*

$$\int_a^b f dg = (-1)^{\alpha} \int_a^b \left(D_{a+}^{\alpha} f(t)\right) \left(D_{b-}^{1-\alpha} g_{b-}(t)\right) dt, \tag{7}$$

where $g_{b-}(t) = g(t) - g(b)$.

In what follows κ denotes a universal generic constant depending only on λ, τ, α and independent of W, φ and a, b. The value of κ may vary from occurrence to occurrence.

For two function $f, g : [a, b] \to \mathbb{R}$, we can define the Riemann-Stieltjes integral $\int_a^b f(t)dg(t)$. Here we recall a result which is well-known (see for example [3, 9] or [6, 7]).

Lemma 1 *Let* f *and* g *be Hölder continuous functions of orders* α *and* β *respectively. Suppose that* $\alpha + \beta > 1$. *Then the Riemann-Stieltjes integral* $\int_a^b f(t)dg(t)$ *exists and for any* $\gamma \in (1 - \beta, \alpha)$, *we have*

$$\int_a^b f(t)dg(t) = (-1)^{\gamma} \int_a^b D_a^{\gamma} f(t) D_{b-}^{1-\gamma} g_{b-}(t) dt\,. \tag{8}$$

Moreover, there is a constant κ *such that*

$$\left|\int_a^b f(t)dg(t)\right| \le \kappa \|g\|_{\beta;a,b} (\|f\|_{\infty;a,b} |b-a|^{\beta} + \|f\|_{\alpha;a,b} |b-a|^{\alpha+\beta}). \tag{9}$$

Proof We refer to [9] or [3] for a proof of (8). We shall outline a proof of (9). Let γ be such that $\alpha > \gamma > 1 - \beta$. Applying fractional integration by parts formula (8), we obtain

$$\left|\int_a^b f(t)dg(t)\right| \le \int_a^b |D_{a+}^{\gamma} f(t) D_{b-}^{1-\gamma} g_{b-}(t)| dt.$$

From (5) and (6) it is easy to see that

$$|D_{b-}^{1-\gamma} g_{b-}(t)| \le \kappa \|g\|_{\beta;a,b} (b - r)^{\beta+\gamma-1}$$

and

$$|D_{a+}^{\gamma} f(t)| \le \kappa[\|f\|_{\infty;a,b}(t-a)^{-\gamma} + \|f\|_{\alpha;a,b}(t-a)^{\alpha-\gamma}].$$

Therefore

$$\left|\int_a^b f(t)dg(t)\right| \le \kappa\|g\|_{\beta;a,b}\left(\|f\|_{\infty;a,b}\int_a^b (t-a)^{-\gamma}(b-t)^{\beta+\gamma-1}dt\right.$$
$$\left. + \|f\|_{\alpha;a,b}\int_a^b (t-a)^{\alpha-\gamma}(b-t)^{\beta+\gamma-1}dt\right).$$

The integrals on the right hand side can be computed by making the substitution $t = b-(b-a)s$. Hence we derive (9).

We also need the following lemma in the proofs of our main results.

Lemma 2 *Let $f(s,t)$, $a \le s < t \le b$ be a measurable function of s and t such that*

$$\int_a^b \int_a^t \frac{|f(s,t)|}{(t-s)^{1-\alpha}} ds dt < \infty.$$

Then

$$\int_a^b I_{a+}^{\alpha,t} f(t,t')|_{t'=t} dt = (-1)^{\alpha} \int_a^b I_{b-}^{\alpha,t'} f(t,t')|_{t'=t} dt.$$

Proof An application of Fubini's theorem yields

$$\begin{aligned}
\int_a^b I_{a+}^{\alpha,t} f(t,t')|_{t'=t} dt &= \frac{1}{\Gamma(\alpha)} \int_a^b \int_a^t \frac{f(s,t)}{(t-s)^{1-\alpha}} ds dt \\
&= \frac{1}{\Gamma(\alpha)} \int_a^b \int_s^b \frac{f(s,t)}{(t-s)^{1-\alpha}} dt ds \\
&= (-1)^{\alpha} \int_a^b I_{b-}^{\alpha,t'} f(t,t')|_{t'=t} dt
\end{aligned}$$

which is the lemma.

3 Nonlinear Integral

In this section we shall use fractional calculus to define the (pathwise) nonlinear integral $\int_a^b W(dt, \varphi_t)$. This method only relies on regularity of the sample paths of W and φ. More precisely, it is applicable to stochastic processes with Hölder continuous sample paths.

Another advantage of this approach is that in the theory of stochastic processes it is usually difficult to obtain almost sure type of results. If the sample paths of the process is Hölder continuous, then one can apply this approach to each sample path and almost surely results are then automatic.

In what follows, we shall use W to denote a deterministic function $W : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$. We make the following assumption on the regularity of W:

(*W*) There are constants $\tau, \lambda \in (0, 1]$ such that for all finite $a < b$ and for all compact sets K of $\mathbb{R}^d$, the seminorm

$$\begin{aligned} &\|W\|_{\tau,\lambda;a,b,K} \\ &:= \sup_{\substack{a \le s < t \le b \\ x,y \in K; x \neq y}} \frac{|W(s,x) - W(t,x) - W(s,y) + W(t,y)|}{|t-s|^\tau |x-y|^\lambda} \\ &\quad + \sup_{\substack{a \le s < t \le b \\ x \in K}} \frac{|W(s,x) - W(t,x)|}{|t-s|^\tau} + \sup_{\substack{a \le t \le b \\ x,y \in K; x \neq y}} \frac{|W(t,y) - W(t,x)|}{|x-y|^\lambda}, \end{aligned} \tag{10}$$

is finite.

About the function φ, we assume

(ϕ) φ is locally Hölder continuous of order $\gamma \in (0, 1]$. That is, the seminorm

$$\|\varphi\|_{\gamma;a,b} = \sup_{a \le s < t \le b} \frac{|\varphi(t) - \varphi(s)|}{|t-s|^\gamma},$$

is finite for every $a < b$.

Among the three terms appearing in **(*W*)**, we will pay special attention to the first term. Thus, we denote

$$[W]_{\tau,\lambda;a,b,K} = \sup_{\substack{a \le s < t \le b \\ x,y \in K; x \neq y}} \frac{|W(s,x) - W(t,x) - W(s,y) + W(t,y)|}{|t-s|^\tau |x-y|^\lambda}.$$

If a, b is clear from the context, we frequently omit the dependence on a, b. In addition, throughout the paper, the compact set K can be chosen to be any compact set containing the image of φ on the interval of integration. Thus we omit the dependence on K as well. For instance, $\|W\|_{\tau,\lambda}$ is an abbreviation for $\|W\|_{\tau,\lambda;a,b,K}$, $\|\varphi\|_\gamma$ is an abbreviation for $\|\varphi\|_{\gamma;a,b}$ and so on. We shall assume that a and b are finite. Thus it is easy to see that for any $c \in [a,b]$

$$\sup_{a\le t\le b} |\varphi(t)| = \sup_{a\le t\le b} |\varphi(c)+\varphi(t)-\varphi(c)| \le |\varphi(c)| + \|\varphi\|_\gamma |b-a|^\gamma < \infty\,.$$

Thus assumption **($\boldsymbol{\phi}$)** also implies that

$$\|\varphi\|_{\infty;a,b} := \sup_{a\le t\le b} |\varphi(t)| < \infty\,.$$

Remark 1 Given a stochastic process indexed by (t,x), it is possible to obtain almost sure regularity of the type (10) by a multiparameter Garsia-Rodemich-Rumsey inequality. Indeed, this has been explored in [4], see also the last section of [5].

One of our main results in this section is to define $\int_a^b W(dt,\varphi_t)$ under the condition $\lambda\gamma+\tau > 1$ through a fractional integration by parts technique. The following definition is motivated from Lemma 1.

Definition 1 We define

$$\int_a^b W(dt,\varphi_t) = (-1)^\alpha \int_a^b D_{a+}^{\alpha,t'} D_{b-}^{1-\alpha,t} W_{b-}(t,\varphi_{t'})|_{t'=t} dt \tag{11}$$

whenever the right hand side makes sense.

Remark 2 Assume $d=1$. Let $W(t,x)=g(t)x$ be of the product form and let $\varphi(t)=f(t)$, where g is a Hölder continuous function of exponent τ and f is a Hölder continuous function of exponent λ. If $1-\tau<\alpha<\lambda$, then

$$\begin{aligned}\int_a^b W(dt,\varphi_t) &= (-1)^\alpha \int_a^b D_{a+}^{\alpha,t'} D_{b-}^{1-\alpha,t} W_{b-}(t,t')|_{t'=t} dt\\ &= (-1)^\alpha \int_a^b D_{b-}^{1-\alpha,t} g_{b-}(t) D_{a+}^{\alpha,t} f(t) dt\,.\end{aligned}$$

Thus from (8), $\int_a^b W(dt,\varphi_t)$ is an extension of the classical Young integral $\int_a^b f(t)dg(t)$ (see [3, 8, 9]). For general d, if $W(t,x) = \sum_{i=1}^d g_i(t)x_i$ and $\varphi_i(t) = f_i(t)$, then it is easy to see that $\int_a^b W(dt,\varphi_t) = \sum_{i=1}^d \int_a^b f_i(t)dg_i(t)$.

The following result clarifies the context in which Definition 1 is justified.

Theorem 1 *Assume the conditions* **(W)** *and* **($\boldsymbol{\phi}$)** *are satisfied. In addition, we suppose that $\lambda\gamma + \tau > 1$. Let $\alpha \in (1-\tau, \lambda\tau)$. Then the right hand side of* (11) *is finite and is independent of $\alpha \in (1-\tau, \lambda)$. As a consequence, we have*

$$
\begin{aligned}
&\int_a^b W(dt, \varphi_t)\\
&= (-1)^\alpha \int_a^b D_{a+}^{\alpha,t'} D_{b-}^{1-\alpha,t} W_{b-}(t, \varphi_{t'})|_{t'=t} dt\\
&= -\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \left\{ \int_a^b \frac{W_{b-}(t,\varphi_t)}{(b-t)^{1-\alpha}(t-a)^\alpha} dt \right.\\
&\quad + \alpha \int_a^b \int_a^t \frac{W_{b-}(t,\varphi_t) - W_{b-}(t,\varphi_r)}{(b-t)^{1-\alpha}(t-r)^{\alpha+1}} drdt\\
&\quad + (1-\alpha) \int_a^b \int_t^b \frac{W(t,\varphi_t) - W(s,\varphi_t)}{(s-t)^{2-\alpha}(t-a)^\alpha} dsdt\\
&\quad \left. + \alpha(1-\alpha) \int_a^b \int_a^t \int_t^b \frac{W(t,\varphi_t) - W(s,\varphi_t) - W(t,\varphi_r) + W(s,\varphi_r)}{(s-t)^{2-\alpha}(t-r)^{\alpha+1}} dsdrdt \right\},
\end{aligned}
\tag{12}
$$

where $W_{b-}(t,x) = W(t,x) - W(b,x)$. Moreover, there is a universal constant κ depending only on τ, λ and α, but independent W, φ and a, b such that

$$
\left| \int_a^b W(dt,\varphi_t) \right| \le \kappa \|W\|_{\tau,\lambda\,;a,b}(b-a)^\tau + \kappa \|W\|_{\tau,\lambda\,;a,b} \|\varphi\|_{\gamma\,;a,b}^\lambda (b-a)^{\tau+\lambda\gamma}, \tag{13}
$$

where $\|W\|_{\tau,\lambda\,;a,b} = \|W\|_{\tau,\lambda\,;a,b,K}$ and K is the closure of the image of $(\varphi_t, a \le t \le b)$.

Proof We denote $\|W\| = \|W\|_{\tau,\lambda;a,b}$. First by the definitions of fractional derivatives (5) and (6), we have

$$
D_{b-}^{1-\alpha,t} W_{b-}(t,\varphi_{t'}) = \frac{(-1)^{1-\alpha}}{\Gamma(\alpha)} \left(\frac{W_{b-}(t,\varphi_{t'})}{(b-t)^{1-\alpha}} + (1-\alpha) \int_t^b \frac{W(t,\varphi_{t'}) - W(s,\varphi_{t'})}{(s-t)^{2-\alpha}} ds \right).
$$

and

$$
\begin{aligned}
&D_{a+}^{\alpha,t'} D_{b-}^{1-\alpha,t} W_{b-}(t,\varphi_{t'})\\
&= \frac{(-1)^{1-\alpha}}{\Gamma(\alpha)\Gamma(1-\alpha)} \left(\frac{1}{(t'-a)^\alpha} \frac{W_{b-}(t,\varphi_{t'})}{(b-t)^{1-\alpha}} + \alpha \int_a^{t'} \frac{W_{b-}(t,\varphi_{t'}) - W_{b-}(t,\varphi_r)}{(t'-r)^{\alpha+1}(b-t)^{1-\alpha}} dr \right.\\
&\quad + \frac{1-\alpha}{(t'-a)^\alpha} \int_t^b \frac{W(t,\varphi_{t'}) - W(s,\varphi_{t'})}{(s-t)^{2-\alpha}} ds\\
&\quad \left. +(1-\alpha) \int_a^{t'} \frac{\alpha}{(t'-r)^{\alpha+1}} \int_t^b \frac{W(t,\varphi_{t'}) - W(s,\varphi_{t'}) - W(t,\varphi_r) + W(s,\varphi_r)}{(s-t)^{2-\alpha}} ds dr \right).
\end{aligned}
$$

Thus the right hand side of (11) is

$$
\begin{aligned}
&- \frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \left\{ \int_a^b \frac{W_{b-}(t,\varphi_t)}{(b-t)^{1-\alpha}(t-a)^\alpha} dt + \alpha \int_a^b \int_a^t \frac{W_{b-}(t,\varphi_t) - W_{b-}(t,\varphi_r)}{(b-t)^{1-\alpha}(t-r)^{\alpha+1}} dr dt \right.\\
&\quad + (1-\alpha) \int_a^b \int_t^b \frac{W(t,\varphi_t) - W(s,\varphi_t)}{(s-t)^{2-\alpha}(t-a)^\alpha} ds dt\\
&\quad \left. + \alpha(1-\alpha) \int_a^b \int_a^t \int_t^b \frac{W(t,\varphi_t) - W(s,\varphi_t) - W(t,\varphi_r) + W(s,\varphi_r)}{(s-t)^{2-\alpha}(t-r)^{\alpha+1}} ds dr dt \right\}\\
&=: I_1 + I_2 + I_3 + I_4 \,. \qquad (14)
\end{aligned}
$$

The condition **(*W*)** implies

$$
\begin{aligned}
I_1 &\le \kappa \|W\| \int_a^b (b-t)^{\tau+\alpha-1}(t-a)^{-\alpha} dt\\
&= \kappa \|W\| (b-a)^\tau \,. \qquad (15)
\end{aligned}
$$

Similarly, we also have

$$
\begin{aligned}
I_3 &\le \kappa \|W\| \int_a^b \int_t^b (s-t)^{\tau+\alpha-2}(t-a)^{-\alpha} ds dt\\
&\le \kappa \|W\| (b-a)^\tau \,. \qquad (16)
\end{aligned}
$$

The assumptions **(W)** and **(ϕ)** also imply

$$\begin{aligned}|W_{b-}(t,\varphi_t)-W_{b-}(t,\varphi_r)| &\le \kappa\|W\||b-t|^{\tau}|\varphi_t-\varphi_r|^{\lambda}\\ &\le \kappa\|W\|\|\varphi\|_{\gamma}^{\lambda}|b-t|^{\tau}|t-r|^{\lambda\gamma}\,.\end{aligned}$$

This implies

$$\begin{aligned}I_2 &\le \kappa\|W\|\|\varphi\|_{\gamma}^{\lambda}\int_a^b\int_a^t (b-t)^{\tau+\alpha-1}(t-r)^{\lambda\gamma-\alpha-1}drdt\\ &\le \kappa\|W\|\|\varphi\|_{\gamma}^{\lambda}(b-a)^{\tau+\lambda\gamma}\,. \end{aligned}\tag{17}$$

Using

$$|W(t,\varphi_t)-W(s,\varphi_t)-W(t,\varphi_r)+W(s,\varphi_r)|\le\kappa\|W\|\|\varphi\|_{\gamma}^{\lambda}|t-s|^{\tau}|t-r|^{\lambda\gamma}\,,$$

we can estimate I_4 as follows.

$$\begin{aligned}I_4 &\le \kappa\|W\|\|\varphi\|_{\gamma}^{\lambda}\int_a^b\int_a^t\int_t^b \frac{|t-s|^{\tau}|t-r|^{\lambda\gamma}}{(s-t)^{2-\alpha}(t-r)^{\alpha+1}}dsdrdt\\ &\le \kappa\|W\|\|\varphi\|_{\gamma}^{\lambda}(b-a)^{\tau+\lambda\gamma}\,. \end{aligned}\tag{18}$$

The inequalities (15)–(18) imply that for any $\alpha\in(1-\tau,\gamma\lambda)$, the right hand side of (11) is well-defined. The inequalities (15)–(18) also yield (13).

To show (12) is independent of α we suppose $\alpha',\alpha\in(1-\tau,\lambda\gamma)$, $\alpha'>\alpha$. Denote $\beta=\alpha'-\alpha$. Using Lemma 2, it is straightforward to see that

$$\begin{aligned}&(-1)^{\alpha}\int_a^b D_{a+}^{\alpha,t}D_{b-}^{1-\alpha,t'}W_{b-}(t,\varphi_{t'})|_{t'=t}dt\\ &=(-1)^{\alpha}\int_a^b I_{a+}^{\beta,t}D_{a+}^{\beta,t}D_{a+}^{\alpha,t}D_{b-}^{1-\alpha,t'}W_{b-}(t,\varphi_{t'})|_{t'=t}dt\\ &=(-1)^{\alpha+\beta}\int_a^b I_{b-}^{\beta,t'}D_{a+}^{\alpha+\beta,t}D_{b-}^{1-\alpha,t'}W_{b-}(t,\varphi_{t'})|_{t'=t}dt\\ &=(-1)^{\alpha'}\int_a^b D_{a+}^{\alpha',t}I_{b-}^{\beta,t'}D_{b-}^{1-\alpha,t'}W_{b-}(t,\varphi_{t'})|_{t'=t}dt\\ &=(-1)^{\alpha'}\int_a^b D_{a+}^{\alpha',t}D_{b-}^{1-\alpha',t'}W_{b-}(t,\varphi_{t'})|_{t'=t}dt\,.\end{aligned}$$

This proves the theorem.

Now we can improve the equality (13) as in the following theorem

Theorem 2 *Let the assumptions* $(\boldsymbol{W})$ *and* $(\boldsymbol{\phi})$ *be satisfied. Let* a, b, c *be real numbers such that* $a \le c \le b$*. Then there is a constant* κ *depending only on* τ, λ *and* α*, but independent* W, φ *and* a, b, c *such that*

$$\left| \int_a^b W(dt, \varphi_t) - W(b, \varphi_c) + W(a, \varphi_c) \right| \le \kappa \|W\|_{\tau,\lambda;a,b} \|\varphi\|_{\gamma;a,b}^{\lambda} (b-a)^{\tau+\lambda\gamma} . \tag{19}$$

Proof Let $a \le c < d \le b$ and let $\tilde{\varphi}(t) = \varphi(c)\chi_{[c,d)}(t)$, where $\chi_{[c,d)}$ is the indicator function on $[c, d)$. Then

$$W(t, \tilde{\varphi}(t')) = \begin{cases} W(t, \varphi(c)) & c \le t' < d \\ \\ W(t, 0) & \text{elsewhere} . \end{cases}$$

This means $W(t, \tilde{\varphi}(t')) = W(t, \varphi(c))\chi_{[c,d)}(t')$. Hence, from (8) we have

$$\begin{aligned} \int_a^b W(dt, \tilde{\varphi}(t)) &= (-1)^{\alpha} \int_a^b D_{b-}^{1-\alpha,t} W_{b-}(t, \varphi(c)) D_{a+}^{\alpha,t'} \chi_{[c,d)}(t')|_{t'=t} dt \\ &= (-1)^{\alpha} \int_a^b D_{b-}^{1-\alpha,t} W_{b-}(t, \varphi(c)) D_{a+}^{\alpha,t} \chi_{[c,d)}(t) dt \\ &= W(d, \varphi(c)) - W(c, \varphi(c)) . \end{aligned}$$

Let c be any point in $[a, b]$. Denote $\tilde{W}(t, x) = W(t, x) - W(t, \varphi_c)$. Then $\tilde{W}$ satisfies $(\boldsymbol{W})$. As in the Eq. (14), we have

$$\begin{aligned} \int_a^b W(dt, \varphi_t) - W(b, \varphi_c) + W(a, \varphi_c) &= \int_a^b \tilde{W}(dt, \varphi_t) \\ &= \tilde{I}_1 + \tilde{I}_2 + \tilde{I}_3 + \tilde{I}_4 . \end{aligned}$$

where $\tilde{I}_2 = I_2$ and $\tilde{I}_4 = I_4$ are the same as I_2 and I_4 in the proof of Theorem 1. But

$$\tilde{I}_1 = -\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_a^b \frac{W(t,\varphi_t)-W(b,\varphi_t)-W(t,\varphi_c)+W(b,\varphi_c)}{(b-t)^{1-\alpha}(t-a)^\alpha}dt$$

$$\tilde{I}_3 = -\frac{(1-\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)}\int_a^b\int_t^b \frac{W(t,\varphi_t)-W(s,\varphi_t)-W(t,\varphi_c)+W(s,\varphi_c)}{(s-t)^{2-\alpha}(t-a)^\alpha}dsdt\,.$$

From the assumptions **(*W*)** and **(*ϕ*)** we see that

$$\begin{aligned}&|W(t,\varphi_t)-W(b,\varphi_t)-W(t,\varphi_c)+W(b,\varphi_c)|\\ &\quad\le \kappa\|W\|_{\tau,\lambda;a,b}\|\varphi\|^\lambda_{\gamma;a,b}|b-t|^\tau|t-c|^{\lambda\gamma}\\ &\quad\le \kappa\|W\|_{\tau,\lambda;a,b}\|\varphi\|^\lambda_{\gamma;a,b}|b-t|^\tau|t-a|^{\lambda\gamma}\,.\end{aligned}$$

This implies that

$$\tilde{I}_1 \le \kappa\|W\|_{\tau,\lambda;a,b}\|\varphi\|^\lambda_{\gamma;a,b}(b-a)^{\tau+\lambda\gamma}\,. \tag{20}$$

Similarly, we have

$$\tilde{I}_3 \le \kappa\|W\|_{\tau,\lambda;a,b}\|\varphi\|^\lambda_{\gamma;a,b}(b-a)^{\tau+\lambda\gamma}\,. \tag{21}$$

Combining these two inequalities (20) and (21) with the inequalities (17) and (18) we have

$$\left|\int_a^b \tilde{W}(dt,\varphi_t)\right| \le \kappa\|W\|_{\tau,\lambda;a,b}\|\varphi\|^\lambda_{\gamma;a,b}(b-a)^{\tau+\lambda\gamma}\,,$$

which yields (19).

Theorem 3 *Let the assumption* **(*W*)** *be satisfied. Let* $\varphi:[a,b]\to\mathbb{R}^d$ *satisfy*

$$|\varphi(s)-\varphi(a)|\le L|s-a|^\ell \quad \forall s\in[a,b] \qquad\text{and}\qquad \sup_{a\le t<s\le b}\frac{|\varphi(s)-\varphi(t)|}{(s-t)^\gamma}\le L \tag{22}$$

for some $\ell\in(\gamma,\infty)$ *and for some constant* $L\in(0,\infty)$*. If* $\tau+\lambda\gamma>1$*, then for any* $\beta<1+\frac{\lambda\gamma+\tau-1}{\gamma}\ell$ *we have*

$$\left|\int_a^b W(dt,\varphi_t)-W(b,\varphi_a)+W(a,\varphi_a)\right|\le C(b-a)^\beta\,, \tag{23}$$

here the constant C *does not depend on* $b-a$.

Proof As in the proof of Theorem 2 we express $\int_a^b W(dt,\varphi_t)-W(b,\varphi_a)+W(a,\varphi_a)$ as the sum of the terms $\tilde{I}_j,\ j=1,2,3,4$ (we follow the notation there). First, we explain how to proceed with $\tilde{I}_4$. We shall use C to denote a generic constant independent of $b-a$. Denote

$$J := |W(t,\varphi_t) - W(s,\varphi_t) - W(t,\varphi_r) + W(s,\varphi_r)|$$

First, we know that we have

$$J \le C|t-s|^{\tau}|t-r|^{\lambda\gamma}\,. \tag{24}$$

On the other hand, we also have

$$\begin{aligned} J &\le |W(t,\varphi_t) - W(s,\varphi_t) - W(t,\varphi_a) + W(s,\varphi_a)| \\ &\qquad + |W(t,\varphi_r) - W(s,\varphi_r) - W(t,\varphi_a) + W(s,\varphi_a)| \\ &\le C|t-s|^{\tau}\left[|t-a|^{\lambda\ell} + |r-a|^{\lambda\ell}\right] \\ &\le C|t-s|^{\tau}|t-a|^{\lambda\ell} \end{aligned} \tag{25}$$

when $a \le r < t < s \le b$. Therefore, from (24) and (25) it follows that for any $\beta_1 \ge 0$ and $\beta_2 \ge 0$ with $\beta_1 + \beta_2 = 1$, we have

$$J \le C|t-s|^{\tau}|t-r|^{\beta_1\lambda\gamma}|t-a|^{\beta_2\lambda\ell}$$

If we choose α and β_1 such that

$$\tau + \alpha > 1\,, \qquad \beta_1\lambda\gamma - \alpha > 0 \tag{26}$$

then

$$\tilde{I}_4 \le C(b-a)^{\beta_1\lambda\gamma + \beta_2\lambda\ell + \tau}\,.$$

For any $\beta < 1 + \frac{\lambda\gamma+\tau-1}{\gamma}\ell$ we can choose α, β_1, and β_2 such that (26) is satisfied and

$$\tilde{I}_4 \le C(b-a)^{\beta}\,.$$

The term $\tilde{I}_2$ can be handled in a similar but easier way and a similar bound can be obtained.

Now, let us consider $\tilde{I}_3$. We have

$$|W(t,\varphi_t) - W(s,\varphi_t) - W(t,\varphi_a) + W(s,\varphi_a)| \le C|t-s|^{\tau}|t-a|^{\lambda\ell}\,.$$

This easily yields

$$\tilde{I}_3 \le C(b-a)^{\tau+\lambda\ell}\,.$$

A similar estimate holds true for $\tilde{I}_1$. However, it is easy to verify $\tau + \lambda\ell > 1 + \frac{\lambda\gamma+\tau-1}{\gamma}\ell$ if $\ell > \gamma$. The theorem is proved.

Next, we show that the nonlinear integral defined in Definition 1 coincides with the limit of Riemann sums. For this purpose, we need some preliminary set up. For every s, t in $[a, b]$, we put $\mu(s, t) = W(t, \varphi_s) - W(s, \varphi_s)$. Let $\pi = \{a = t_0 < t_1 < \cdots < t_n = b\}$ be a partition of $[a, b]$ with mesh size $|\pi| = \max_{1\le i\le n} |t_i - t_{i-1}|$. One can consider the limit of the Riemann sums

$$\lim_{|\pi|\downarrow 0} \sum_{i=1}^{n} \mu(t_{i-1}, t_i)$$

whenever it exists. A sufficient condition for convergence of the Riemann sums is provided by following two results of [2], see also [1] for a simple exposition.

Lemma 3 (The sewing map) *Let μ be a continuous function on $[0, T]^2$ with values in a Banach space B and $\varepsilon > 0$. Suppose that μ satisfies*

$$|\mu(a, b) - \mu(a, c) - \mu(c, b)| \le K|b - a|^{1+\varepsilon} \quad \forall\, 0 \le a \le c \le b \le T\,.$$

Then there exists a function $\mathscr{J}\mu(t)$ unique up to an additive constant such that

$$|\mathscr{J}\mu(b) - \mathscr{J}\mu(a) - \mu(a, b)| \le K(1 - 2^{-\varepsilon})^{-1}|b - a|^{1+\varepsilon} \quad \forall\, 0 \le a \le b \le T\,. \tag{27}$$

We adopt the notation $\mathscr{J}_a^b\mu = \mathscr{J}\mu(b) - \mathscr{J}\mu(a)$

Lemma 4 (Abstract Riemann sum) *Let $\pi = \{a = t_0 < t_1 < \cdots < t_m = b\}$ be an arbitrary partition of $[a, b]$ with $|\pi| = \sup_{i=0,\ldots,m-1} |t_{i+1} - t_i|$. Define the Riemann sum*

$$J_\pi = \sum_{i=0}^{m-1} \mu(t_i, t_{i+1})$$

then J_π converges to $\mathscr{J}_a^b\mu$ as $|\pi| \downarrow 0$.

Because $\tau + \lambda\gamma$ is strictly greater than 1, the estimate (19) together with the previous two Lemmas implies

Proposition 2 *Assume that* **(W)** *and* **(ϕ)** *hold with $\lambda\gamma + \tau > 1$. As the mesh size $|\pi|$ shrinks to 0, the Riemann sums*

$$\sum_{i=1}^{n} \left[W(t_i, \varphi_{t_{i-1}}) - W(t_{i-1}, \varphi_{t_{i-1}})\right]$$

converges to $\int_a^b W(dt, \varphi_t)$.

Remark 3 In [5], the authors define the nonlinear integral $\int W(dt, \varphi_t)$ via the sewing Lemma 3. The previous proposition shows that the approach using fractional calculus employed here produces an equivalent definition. Let us note that this is possible because of the key estimate (19) and the uniqueness part of the sewing Lemma 3.

It is easy to see from here that

$$\int_a^b W(dt,\varphi_t) = \int_a^c W(dt,\varphi_t) + \int_c^b W(dt,\varphi_t) \qquad \forall\, a < c < b\,.$$

This together with (13) imply easily the following.

Proposition 3 *Assume that* $(\boldsymbol{W})$ *and* $(\boldsymbol{\phi})$ *hold with* $\lambda\gamma + \tau > 1$. *As a function of* t, *the indefinite integral* $\left\{\int_a^t W(ds,\varphi_s)\,,\ t \le a \le b\right\}$ *is Hölder continuous of exponent* τ.

Further properties can be developed. For instance, we study the dependence of the nonlinear Young integration $\int W(ds,\varphi_s)$ with respect to the medium W and the integrand φ. We state the following two propositions whose proofs are left for readers (see, however, [5] for details).

Proposition 4 *Let* W_1 *and* W_2 *be functions on* $\mathbb{R}\times\mathbb{R}^d$ *satisfying the condition* $(\boldsymbol{W})$. *Let* φ *be a function in* $C^\gamma(\mathbb{R};\mathbb{R}^d)$ *and assume that* $\tau + \lambda\gamma > 1$. *Then*

$$\begin{aligned}|\int_a^b W_1(ds,\varphi_s) - \int_a^b W_2(ds,\varphi_s)| &\le |W_1(b,\varphi_a) - W_1(a,\varphi_a) - W_2(b,\varphi_a) + W_2(a,\varphi_a)| \\ &\quad + c(\|\varphi\|_\infty)[W_1 - W_2]_{\beta,\tau,\lambda}\|\varphi\|_\gamma |b-a|^{\tau+\lambda\gamma}\,.\end{aligned}$$

Proposition 5 *Let* W *be a function on* $\mathbb{R}\times\mathbb{R}^d$ *satisfying the condition* $(\boldsymbol{W})$. *Let* φ^1 *and* φ^2 *be two functions in* $C^\gamma(\mathbb{R};\mathbb{R}^d)$ *and assume that* $\tau + \lambda\gamma > 1$. *Let* $\theta \in (0,1)$ *such that* $\tau + \theta\lambda\gamma > 1$. *Then for any* $u < v$

$$\begin{aligned}&|\int_u^v W(ds,\varphi_s^1) - \int_u^v W(ds,\varphi_s^2)| \\ &\quad \le C_1[W]_{\tau,\lambda}\|\varphi^1 - \varphi^2\|_\infty^\lambda |v-u|^\tau \\ &\qquad + C_2[W]_{\tau,\lambda}\|\varphi^1 - \varphi^2\|_\infty^{\lambda(1-\theta)} |v-u|^{\tau+\theta\lambda\gamma}\,,\end{aligned}$$

where C_1 *is an absolute constant and* $C_2 = 2^{1-\theta}C_1(\|\varphi^1\|_\gamma^\lambda + \|\varphi^1\|_\gamma^\lambda)^\theta$.

4 Iterated Nonlinear Integral

From Remark 2 we see that if $W(t,x) = \sum_{i=1}^d g_i(t)x_i$ and $\varphi_i(t) = f_i(t)$, then $\int_a^b W(dt,\varphi_t) = \sum_{i=1}^d \int_a^b f_i(t)dg_i(t)$. We know that the multiple (iterated) integrals of the form

$$\int_{a\le s_1\le s_2\le\cdots\le s_n\le b} \varphi(s_1, s_2, \ldots, s_n)dg(s_1)dg(s_2)\cdots dg(s_n)$$

are well-defined and have applications in expanding the solutions of differential equations (see [3]). What is the extension of the above iterated integrals to the nonlinear integral? To simplify the presentation, we consider the case $d = 1$. General dimensions can be considered in a similar way with more complex notations.

We introduce the following notation. Let

$$\Delta_{n,a,b} := \{(s_1, \ldots, s_n)\,;\; a \le s_1 \le s_2 \le \cdots \le s_n \le b\}$$

be a simplex in $\mathbb{R}^n$.

Definition 2 Let $\varphi : \Delta_{n,a,b} \to \mathbb{R}$ be a continuous function. For a fixed $s_n \in [a, b]$, we can consider $\varphi(\cdot, s_n)$ as a function of $n-1$ variables. Assume we can define $\int_{\Delta_{n-1,a,s_n}} \varphi(s_1, \ldots, s_{n-1}, s_n)W(ds_1, \cdot)\cdots W(ds_{n-1}, \cdot)$, which is a function of s_n, denoted by $\phi_{n-1}(s_n)$, then we define

$$\int_{a\le s_1\le\cdots\le s_n\le b} \varphi(s_1, \ldots, s_n)W(ds_1, \cdot)\cdots W(ds_n, \cdot) = \int_a^b W(ds_n, \varphi_{n-1}(s_n))\,. \tag{28}$$

In the case $W(t, x) = f(t)x$, such iterated integrals have been studied in [3], where an important case is when $\varphi(s_1, \ldots, s_n) = \rho(s_1)$ for some function ρ of one variable. This means that $\varphi(s_1, \ldots, s_n)$ depends only on the first variable. This case appears in the remainder term when one expands the solution of a differential equation and can be dealt with in the following way.

Let $F_1, F_2, \ldots, F_n$ be jointly Hölder continuous functions on $[a, b]^2$. More precisely, for each $i = 1, \ldots, n$, F_i satisfies

$$\begin{aligned} &|F_i(s_1, t_1) - F_i(s_2, t_1) - F_i(s_1, t_2) + F_i(s_2, t_2)| \\ &\quad \le \|F_i\|_{\tau,\lambda;a,b}|s_1 - s_2|^\tau |t_1 - t_2|^\lambda\,, \quad \text{for all } s_1, s_2, t_1, t_2 \text{ in } [a, b]. \end{aligned} \tag{29}$$

We assume that $\tau + \lambda > 1$.

Suppose that F is a function satisfying (29) with $\tau + \lambda > 1$. The nonlinear integral $\int_a^b F(ds, s)$ can be defined analogously to Definition 1. Moreover, for a Hölder continuous function ρ of order λ, we set $G(s, t) = \rho(t)F(s, t)$, it is easy to see that

$$\begin{aligned}
&|G(s_1,t_1)-G(s_2,t_1)-G(s_1,t_2)+G(t_1,t_2)| \\
&\le |\rho(t_1)-\rho(t_2)||F(s_1,t_1)-F(s_2,t_1)| \\
&\qquad +|\rho(t_2)||F(s_1,t_1)-F(s_2,t_1)-F(s_1,t_2)+F_i(t_1,t_2)| \\
&\le (\|\rho\|_\tau\|+\|\rho\|_\infty)\|F\|_{\tau,\lambda}|s_1-s_2|^\tau|t_1-t_2|^\lambda.
\end{aligned}$$

Hence, the integration $\int\rho(s)F(ds,s)$ is well defined. In addition, it follows from Theorem 2 that the map $t\mapsto\int_a^t\rho(s)F(ds,s)$ is Hölder continuous of order τ.

We have then easily

Proposition 6 *Let ρ be a Hölder continuous function of order λ. Under the condition (29) and $\tau>1/2$, the iterated integral*

$$I_{a,b}(F_1,\ldots,F_n)=\int_{a\le s_1\le\cdots\le s_n\le b}\rho(s_1)F_1(ds_1,s_1)F_2(ds_2,s_2)\cdots F_n(ds_n,s_n) \tag{30}$$

is well defined.

In the simplest case when $\rho(s)=1$ and $F_i(s,t)=f(s)$ for all $i=1,\ldots,n$, the above integral becomes

$$\int_{a\le s_1\le\cdots\le s_n\le b}df(s_1)\cdots df(s_n)=\frac{(f(b)-f(a))^n}{n!}.$$

Therefore, one would expect that

$$|I_{a,b}(F_1,\ldots,F_n)|\le\kappa\frac{|b-a|^{\gamma_n}}{n!}. \tag{31}$$

This estimate turns out to be true for (30).

Theorem 4 *Let $F_1,\ldots,F_n$ satisfy (29) and ρ be Hölder continuous with exponent λ. We assume that $\rho(a)=0$. Denote $\beta=\dfrac{\lambda+\tau-1}{\lambda}$ and $\ell_n=\dfrac{\beta^{n-1}-1}{\beta-1}+\beta^{n-1}(\tau+\lambda)$. Then, for any $\gamma_n<\ell_n$, there is a constant C_n, independent of a and b (but may depend on γ_n) such that*

$$|I_{a,b}(F_1,\ldots,F_n)|\le C_n|b-a|^{\gamma_n}. \tag{32}$$

Proof Denote

$$I_{a,s}^{(k)}(F_1,\ldots,F_k)=\int\limits_{a\le s_1\le\cdots\le s_k\le s}\rho(s_1)F_1(ds_1,s_1)F_2(ds_2,s_2)\cdots F_k(ds_k,s_k).$$

Thus, we see by definition that

$$I_{a,s}^{(k+1)}(F_1,\ldots,F_{k+1})=\int_a^s F_{k+1}(dr,I_{a,r}^{(k)}(F_1,\ldots,F_k)). \tag{33}$$

We prove this theorem by induction on n. When $n=1$, the theorem follows straightforwardly from (19) with the choice $c=a$. Indeed, we have $|I_{a,t}^{(1)}|\le C|t-a|^{\lambda+\tau}$ and $|I_{a,t}^{(1)}-I_{a,s}^{(1)}|\le C|t-s|^{\tau}$.

The passage from n to $n+1$ follows from the application of (23)–(33) and this concludes the proof of the theorem.

Remark 4 The estimate of Theorem 4 also holds true for the iterated nonlinear Young integral $I_{a,b}^{(n)}(F_1,\ldots,F_n)=\int_{a\le s_1\le\cdots\le s_k\le s}F_1(ds_1,\rho(s_1))F_2(ds_2,s_2)\cdots F_n(ds_n,s_n)$, where $I_{a,b}^{(k)}(F_1,\ldots,F_k)=\int_a^b F_k(ds,I_{a,s}^{(k-1)}(F_1,\ldots,F_{k-1}))$, and $I_{a,b}^{(1)}(F_1)=\int_a^b F_1(ds,\rho(s))$.

Acknowledgments Y. Hu is partially supported by a grant from the Simons Foundation #209206 and by a General Research Fund of University of Kansas.

References

1. Feyel, D., de La Pradelle, A.: Curvilinear integrals along enriched paths. Electron. J. Probab. **11**, 860–892 (2006)
2. Gubinelli, M.: Controlling rough paths. J. Funct. Anal. **216**, 86–140 (2004)
3. Hu, Y.: Multiple integrals and expansion of solutions of differential equations driven by rough paths and by fractional Brownian motions. Stochastics **85**, 859–916 (2013)
4. Hu, Y., Le, K.: A multiparameter Garsia-Rodemich-Rumsey inequality and some applications. Stochast. Process. Appl. **123**, 3359–3377 (2013)
5. Hu, Y., Lê, N.K.: Nonlinear Young integrals and differential systems in Hölder media. Trans. Am. Math. Soc. (in press)
6. Hu, Y., Nualart, D.: Differential equations driven by Hölder continuous functions of order greater than 1/2. In: Benth, F.E. et al. (ed.) Abel Symposium on Stochastic Analysis and Applications, pp. 399–413. Springer (2007)

7. Hu, Y., Nualart, D.: Rough path analysis via fractional calculus. Trans. Am. Math. Soc. **361**, 2689–2718 (2009)
8. Young, L.C.: An inequality of the Hölder type, connected with Stieltjes integration. Acta Math. **67**, 251–282 (1936)
9. Zähle, M.: Integration with respect to fractal functions and stochastic calculus. I. Probab. Theory Relat. Fields **111**, 333–374 (1998)

A Weak Limit Theorem for Numerical Approximation of Brownian Semi-stationary Processes

Mark Podolskij and Nopporn Thamrongrat

Abstract In this paper we present a weak limit theorem for a numerical approximation of Brownian semi-stationary processes studied in [14]. In the original work of [14] the authors propose to use Fourier transformation to embed a given one dimensional (Lévy) Brownian semi-stationary process into a two-parameter stochastic field. For the latter they use a simple iteration procedure and study the strong approximation error of the resulting numerical scheme given that the volatility process is fully observed. In this work we present the corresponding weak limit theorem for the setting, where the volatility/drift process needs to be numerically simulated. In particular, weak approximation errors for smooth test functions can be obtained from our asymptotic theory.

Keywords Ambit fields · Brownian semi-stationary processes · Numerical schemes · Weak limit theorems

AMS 2010 Subject Classification: 60F05 · 65C30 · 60F17

1 Introduction

Recently, the mathematical theory of ambit fields has been intensively studied in the literature. Ambit fields is a class of spatio-temporal stochastic processes that has been originally introduced by Barndorff-Nielsen and Schmiegel in a series of papers [9–11] in the context of turbulence modelling, but which found manifold applications

M. Podolskij (✉)
Department of Mathematics, University of Aarhus, Ny Munkegade 118, 8000 Aarhus C, Denmark
e-mail: mpodolskij@creates.au.dk

N. Thamrongrat
Department of Mathematics, Heidelberg University, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany
e-mail: Thamrongrat@stud.uni-heidelberg.de

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_4

in mathematical finance and biology among other sciences; see e.g. [4, 8]. In full generality they are described via the formula

$$X_t(x) = \mu + \int_{A_t(x)} g(t,s,x,\xi)\sigma_s(\xi)L(\mathrm{d}s,\mathrm{d}\xi) + \int_{D_t(x)} q(t,s,x,\xi)a_s(\xi)\mathrm{d}s\mathrm{d}\xi \quad (1)$$

where t typically denotes time while x gives the position in space. Furthermore, $A_t(x)$ and $D_t(x)$ are ambit sets, g and q are deterministic weight functions, σ represents the volatility or intermittency field, a is a drift field and L denotes a *Lévy basis*. We recall that a Lévy basis $L = \{L(B) : \ B \in \mathscr{S}\}$, where $\mathscr{S}$ is a δ-ring of an arbitrary non-empty set S such that there exists an increasing sequence of sets $(S_n) \subset \mathscr{S}$ with $\cup_{n\in\mathbb{N}} S_n = S$, is an independently scattered random measure.

An important purely temporal subclass of ambit fields are the so called *Lévy (Brownian) semi-stationary processes*, which are defined as

$$X_t = \mu + \int_{-\infty}^{t} g(t-s)\sigma_s L(\mathrm{d}s) + \int_{-\infty}^{t} q(t-s)a_s \mathrm{d}s, \quad (2)$$

where now L is a two-sided one dimensional Lévy (Brownian) motion and the ambit sets are given via $A_t = D_t = (-\infty, t)$. The notion of a semi-stationary process refers to the fact that the process $(X_t)_{t\in\mathbb{R}}$ is stationary whenever $(a_t, \sigma_t)_{t\in\mathbb{R}}$ is stationary and independent of $(L_t)_{t\in\mathbb{R}}$. In the past years stochastic analysis, probabilistic properties and statistical inference for Lévy semi-stationary processes have been studied in numerous papers. We refer to [2, 3, 6, 7, 11, 12, 15, 17, 20, 25] for the mathematical theory as well as to [5, 26] for a recent survey on theory of ambit fields and their applications.

For practical applications in sciences numerical approximation of Lévy (Brownian) semi-stationary processes, or, more generally, of ambit fields, is an important issue. We remark that due to a moving average structure of a Lévy semi-stationary process (cf. (2)) there exists no simple iterative Euler type approximation scheme. For this reason the authors of [13, 14] have proposed two different embedding strategies to come up with a numerical simulation. The first idea is based on the embedding of a Lévy semi-stationary process into a certain two-parameter stochastic partial differential equation. The second one is based upon a Fourier method, which again interprets a given Lévy semi-stationary process as a realization of a two-parameter stochastic field. We refer to the PhD thesis of Eyjolfsson [18] for a detailed analysis of both methods and their applications to modeling energy markets. We would also like to mention a very recent work [16], which investigates numerical simulations of spatio-temporal ambit fields.

The aim of this paper is to study the weak limit theory of the numerical scheme associated with the Fourier method proposed in [14, 18]. In the original work [14] the authors have discussed the strong approximation error (in the L^2 sense) of the numerical scheme for Lévy semi-stationary processes, where the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ is assumed to be observed. We complement their study by analyzing the weak limit of the error process in the framework of Brownian semi-stationary processes, where the drift and the volatility processes need to be numerically simulated. This obviously gives a more precise assessment of the numerical error associated with the Fourier method.

The paper is organised as follows. In Sect. 2 we describe the Fourier approximation scheme for Brownian semi-stationary processes and present the main results on strong approximation error derived in [14, 18]. Section 3 is devoted to a weak limit theorem associated with a slight modification of the Fourier method.

2 Basic Assumptions and Fourier Approximation Scheme

We start with a complete filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F})_{t\in\mathbb{R}}, \mathbb{P})$, on which all processes are defined. We consider a Brownian semi-stationary process of the form

$$X_t = \mu + \int_{-\infty}^{t} g(t-s)\sigma_s W(\mathrm{d}s) + \int_{-\infty}^{t} q(t-s)a_s \mathrm{d}s, \tag{3}$$

where g and q are deterministic kernels, $(a_t)_{t\in\mathbb{R}}$ and $(\sigma_t)_{t\in\mathbb{R}}$ are adapted càdlàg processes, and W is a two sided Brownian motion. To guarantee the finiteness of the first integral appearing in (3), we assume throughout the paper that

$$\int_{-\infty}^{t} g^2(t-s)\sigma_s^2 \mathrm{d}s < \infty \qquad \text{almost surely} \tag{4}$$

for all $t \in \mathbb{R}$. When $(\sigma_t)_{t\in\mathbb{R}}$ is a square integrable stationary process, the above condition holds if $g \in L^2(\mathbb{R}_{\geq 0})$. The presence of the drift process $(a_t)_{t\in\mathbb{R}}$ will be essentially ignored in this section.

Now, we describe the Fourier approximation method introduced in [14, 18] applied to the framework of Brownian semi-stationary processes. We start with the following assumptions on kernels involved in the description (3):

Assumption (**A**):

(i) The kernel functions g and q have bounded support contained in $[0, \tau]$ for some $\tau > 0$.
(ii) $g, q \in C(\mathbb{R}_{\geq 0})$.

In some cases these conditions are rather restrictive. We will give remarks on them below. For any given $\lambda > 0$, we define

$$h(x) := g(|x|) \quad \text{and} \quad h_\lambda(x) := h(x)\exp(\lambda|x|). \tag{5}$$

Notice that $g = h$ on $[0, \tau]$. We introduce the Fourier transform of h_λ via

$$\widehat{h}_\lambda(y) := \int_{\mathbb{R}} h_\lambda(x) \exp(-ixy)\mathrm{d}x.$$

Furthermore, if we assume that $\widehat{h}_\lambda \in L^1(\mathbb{R})$, the inverse Fourier transform exists and we obtain the identity

$$h(x) = \frac{\exp(-\lambda|x|)}{2\pi} \int_{\mathbb{R}} \widehat{h}_\lambda(y) \exp(ixy)\mathrm{d}y.$$

Since the Fourier transform maps $L^1(\mathbb{R})$ functions into the space of continuous functions, we require that $h \in C(\mathbb{R})$. This fact explains the Assumption (A)(ii) for the kernel function g. Since h is an even function, for a given number $N \in \mathbb{N}$, we deduce an approximation of h via

$$h(x) \approx h_N(x) := \exp(-\lambda|x|)\left(\frac{b_0}{2} + \sum_{k=1}^{N} b_k \cos(\frac{k\pi x}{\tau})\right) \tag{6}$$

with

$$b_k = \frac{\widehat{h}_\lambda(k\pi/\tau)}{\tau}. \tag{7}$$

Obviously, the above approximation is an L^2-projection onto the linear subspace generated by orthogonal functions $\{\cos(k\pi x/\tau), \sin(k\pi x/\tau)\}_{k=0}^N$, hence we deal with a classical Fourier expansion of the function h (recall that the function h is even by definition, thus the sinus terms do not appear at (6)). Now, the basic idea of the numerical approximation method proposed in [14, 18] is based upon the following relationship:

$$\begin{aligned}\int_u^t g(t-s)\sigma_s W(\mathrm{d}s) &\approx \int_u^t h_N(t-s)\sigma_s W(\mathrm{d}s)\\ &= \int_u^t \exp(-\lambda(t-s))\left\{\frac{b_0}{2} + \sum_{k=1}^{N} b_k \cos(\frac{k\pi(t-s)}{\tau})\right\}\sigma_s W(\mathrm{d}s)\\ &= \frac{b_0}{2}\widehat{X}_{\lambda,u}(t,0) + \mathrm{Re}\sum_{k=1}^{N} b_k \widehat{X}_{\lambda,u}(t, \frac{k\pi}{\tau}),\end{aligned} \tag{8}$$

where the complex valued stochastic field $\widehat{X}_{\lambda,u}(t, y)$ is defined via

$$\widehat{X}_{\lambda,u}(t, y) := \int_u^t \exp\{(-\lambda + iy)(t-s)\}\sigma_s W(\mathrm{d}s) \tag{9}$$

and $u \in [t-\tau, t]$. In a second step, for a $\delta > 0$ small, we observe the approximation

$$\widehat{X}_{\lambda,u}(t+\delta, y) = \exp\{(-\lambda + iy)\delta\} \left(\widehat{X}_{\lambda,u}(t, y) + \int_t^{t+\delta} \exp\{(-\lambda + iy)(t-s)\}\sigma_s W(\mathrm{d}s) \right)$$
$$\approx \exp\{(-\lambda + iy)\delta\} \left(\widehat{X}_{\lambda,u}(t, y) + \sigma_t (W_{t+\delta} - W_t) \right). \qquad (10)$$

Hence, we obtain a simple iterative scheme for simulating the stochastic field $\widehat{X}_{\lambda,u}(t, y)$ in the variable t. Assume for the moment that the drift process a is zero and we wish to simulate the trajectory of $X_{t_0}, \ldots, X_{t_M}$ given the information available at time t_0. Then, the numerical simulation procedure is as follows:

(a) Simulate the independent increments $W_{t_i} - W_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})$ for $i = 1, \ldots, M$.
(b) For each $i = 1, \ldots, M$ and $k = 0, \ldots, N$, simulate $\widehat{X}_{\lambda,u}(t_i, k\pi/\tau)$ from $\widehat{X}_{\lambda,u}(t_{i-1}, k\pi/\tau)$, $W_{t_i} - W_{t_{i-1}}$ and $\sigma_{t_{i-1}}$ by using (10).
(c) Simulate X_{t_i} applying steps (a), (b) and (8) (with $u = t_0$).

Let us explain some properties of the proposed numerical scheme. First of all, there are two approximation errors, where the first one (N scale) is coming from the Fourier transformation at (6) and the second one (M scale) is coming from the discretization error obtained at (10).

It is important to understand the meaning of knowing the information about the involved processes up to time t_0. When the stochastic model for the process $(\sigma_t)_{t\in\mathbb{R}}$ is uncoupled with $(X_t)_{t\in\mathbb{R}}$, then we may use $u = t - \tau$ at (8). Indeed, in typical applications such as turbulence and finance this is the case: $(\sigma_t)_{t\in\mathbb{R}}$ is usually modeled via a jump diffusion process driven by a Lévy process, which might be correlated with the Brownian motion W. However, when the process $(X_t)_{t\in\mathbb{R}}$ is itself of a diffusion type, i.e.

$$X_t = \mu + \int_{t-\tau}^t g(t-s)\sigma(X_s)W(\mathrm{d}s) + \int_{t-\tau}^t q(t-s)a(X_s)\mathrm{d}s$$

it is in general impossible to simulate a trajectory of $(X_t)_{t\in\mathbb{R}}$, since for each value t the knowledge of the path $(X_u)_{u\in(t-\tau,t)}$ is required to compute X_t. But, in case we do know the historical path, say, $(X_u)_{u\in[-\tau,0]}$, the simulation of values X_t, $t \geq 0$, becomes possible.

The main advantage of the numerical scheme described above is that it separates the simulation of the stochastic ingredients (σ and W) and the approximation of the deterministic kernel g (or h). In other words, the stochastic field $\widehat{X}_{\lambda,u}(t, y)$ is simulated via a simple recursive scheme without using the knowledge of g, while the kernel g is approximated via the Fourier transform at (6). This is in contrast to a straightforward discretization scheme

$$\int_{t_0}^{t_j} g(t-s)\sigma_s W(\mathrm{d}s) \approx \sum_{i=1}^{j-1} g(t_j - t_i)\sigma_{t_i}(W_{t_{i+1}} - W_{t_i}).$$

This numerical property is useful when considering a whole family of kernel functions $(g_\theta)_{\theta\in\Theta}$, since for any resulting model $X_t(\theta)$ only one realization of the stochastic field $\widehat{X}_{\lambda,u}(t, y)$ needs to be simulated. This can be obviously useful for the simulation of parametric Brownian semi-stationary processes.

We may now assess the strong approximation error of the proposed numerical scheme. We start with the analysis of the error associated with the approximation of the deterministic kernel g by the function h_N. We assume for the moment that the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ is square integrable with bounded second moment. Then a straightforward computation (see e.g. [14, Eq. (4.5)]) implies that

$$\mathbb{E}\left[\left(\int_{t_0}^{t}\{g(t-s)-h_N(t-s)\}\sigma_s W(\mathrm{d}s)\right)^2\right] \le C\frac{1-\exp\{-2\lambda(t-t_0)\}}{\lambda}\left(\sum_{k=N+1}^{\infty}|b_k|\right)^2, \tag{11}$$

where C is a positive constant and the Fourier coefficients b_k have been defined at (7). We remark that $(1-\exp\{-2\lambda(t-t_0)\})/\lambda \to 2(t-t_0)$ as $\lambda\to 0$, while $(1-\exp\{-2\lambda(t-t_0)\})/\lambda \sim \lambda^{-1}$ as $\lambda\to\infty$. Thus, it is preferable to choose the parameter $\lambda > 0$ large.

Remark 1 A standard model for the kernel function g in the context of turbulence is given via

$$g(x) = x^{\alpha}\exp(-\bar{\lambda}x)$$

with $\bar{\lambda} > 0$ and $\alpha > -1/2$. Obviously, this function has unbounded support and for the values $\alpha \in (-1/2, 0)$ it is also discontinuous at 0, hence it violates the statement of the Assumption (A). However, one can easily construct an approximating function g_ε^T, which coincides with g on the interval $[\varepsilon, T]$ and satisfies the Assumption (A). Assuming again the boundedness of the second moment of the process $(\sigma_t)_{t\in\mathbb{R}}$, the approximation error is controlled via

$$\mathbb{E}\left[\left(\int_{-\infty}^{t}\{g(t-s)-g_\varepsilon^T(t-s)\}\sigma_s W(\mathrm{d}s)\right)^2\right] \le C\|g-g_\varepsilon^T\|^2_{L^2((0,\varepsilon)\cup(T,\infty))}$$

Such error can be made arbitrary small by choosing ε small and T large. Clearly, this is a rather general approach, which is not particularly related to a given class of kernel functions g. In a second step one would apply the Fourier approximation method described above to the function g_ε^T. At his stage it is important to note that the parameter $\lambda > 0$ introduced at (5) is naturally restricted through the condition $\lambda < \bar{\lambda}$; otherwise the kernel h_λ would have an explosive behaviour at ∞. Thus, the approximation error discussed at (11) cannot be made arbitrarily small in λ. □

Remark 2 The Fourier coefficients b_k can be further approximated under stronger conditions on the function h, which helps to obtain an explicit bound at (11). More specifically, when $h \in C^{2n}(\mathbb{R})$ and $h_\lambda^{(2j-1)}(\tau) = 0$ for all $j = 1, \ldots, n$, then it holds that

$$|b_k| \leq Ck^{-2n}.$$

This follows by a repeated application of integration by parts formula (see [14, Proposition 4.1] for a detailed exposition). In fact, the original work [14] defines another type of smooth interpolation functions h, rather than the mere identity $h(x) = g(|x|)$, to achieve that the relationship $h_\lambda^{(2j-1)}(\tau) = 0$ holds for all $j = 1, \ldots, n$ and some $n \in \mathbb{N}$. □

Now, let us turn our attention to the discretization error introduced at (10). We assume that $t_0 < \cdots < t_M$ is an equidistant grid with $t_i - t_{i-1} = \Delta t$. According to (10) the random variable

$$\eta_j(y) := \sum_{i=1}^{j} \exp\{(-\lambda + iy)(j + 1 - i)\Delta t\}\sigma_{t_{i-1}}(W_{t_i} - W_{t_{i-1}}) \tag{12}$$

is an approximation of $\widehat{X}_{\lambda,t_0}(t_j, y)$ for any $y \in \mathbb{R}$ whenever the drift process a is assumed to be absent. When $(\sigma_t)_{t\in\mathbb{R}}$ is a weak sense stationary process, a straightforward computation proves that

$$\mathbb{E}[|\widehat{X}_{\lambda,t_0}(t_j, y) - \eta_j(y)|^2] \leq C(t_j - t_0)\Big((\lambda^2 + y^2)(\Delta t)^2 + \mathbb{E}[|\sigma_{t_1} - \sigma_{t_0}|^2]\Big). \tag{13}$$

We refer to [14, Lemma 4.2] for a detailed proof.

Remark 3 Assume that the process $(\sigma_t)_{t\in\mathbb{R}}$ is a continuous stationary Itô semimartingale, i.e.

$$d\sigma_t = \widetilde{a}_t dt + \widetilde{\sigma}_t dB_t,$$

where B is a Brownian motion and $(\widetilde{a}_t)_{t\in\mathbb{R}}$, $(\widetilde{\sigma}_t)_{t\in\mathbb{R}}$ are stochastic processes with bounded second moment. Then the Itô isometry implies that

$$\mathbb{E}[|\sigma_{t_1} - \sigma_{t_0}|^2] \leq C\Delta t.$$

Hence, in this setting Δt becomes the dominating term in the approximation error (13). □

Combining the estimates at (11) and (13), we obtain the strong approximation error of the proposed Fourier method, which is the main result of [14] (see Propositions 4.1 and 4.3 therein).

Proposition 1 *Let $t_0 < \cdots < t_M$ be an equidistant grid with $t_i - t_{i-1} = \Delta t$. Assume that condition (A) holds and $(\sigma_t)_{t\in\mathbb{R}}$ is a weak sense stationary process. Then the L^2 approximation error associated with the Fourier type numerical scheme is given via*

$$\mathbb{E}\left[\Big|\int_{t_0}^{t_j} g(t_j-s)\sigma_s W(ds) - \Big(\frac{b_0}{2}\eta_j(0) + \sum_{k=1}^{N} b_k\eta_j(\frac{k\pi}{\tau})\Big)\Big|^2\right] \tag{14}$$
$$\le C\Bigg(\frac{1-\exp\{-2\lambda(t-t_0)\}}{\lambda}\Big(\sum_{k=N+1}^{\infty}|b_k|\Big)^2$$
$$+ (t_j-t_0)\Bigg\{\lambda^2\Big(\frac{|b_0|}{2}+\sum_{k=1}^{N}|b_k|\Big)^2(\Delta t)^2 + (\frac{\pi}{\tau})^2\Big(\sum_{k=1}^{N}k|b_k|\Big)^2(\Delta t)^2$$
$$+\Big(\frac{|b_0|}{2}+\sum_{k=1}^{N}|b_k|\Big)^2\mathbb{E}[|\sigma_{t_1}-\sigma_{t_0}|^2]\Bigg\}\Bigg)$$

for a positive constant C.

3 A Weak Limit Theorem for the Fourier Approximation Scheme

As we mentioned earlier, the Fourier approximation scheme investigated in [14, 18] basically ignored the need of simulating the volatility process $(\sigma_t)_{t\in\mathbb{R}}$ in practical applications (the same holds for the drift process $(a_t)_{t\in\mathbb{R}}$). As in the previous section we fix a time t_0 and assume the knowledge of all processes involved up to that time. Here we propose a numerical scheme for simulating the path $(X_t)_{t\in[t_0,T]}$ for a given terminal time $T > t_0$, which is a slightly modified version of the original Fourier approach. We recall the imposed condition (A), in particular, the weight functions g and q are assumed to have bounded support contained in $[0,\tau]$. First of all, we assume that we have càdlàg estimators $(a_t^M, \sigma_t^M)_{t\in[t_0,T]}$ of the stochastic process $(a_t,\sigma_t)_{t\in[t_0,T]}$ and the convergence rate $\nu_M \to \infty$ as $M \to \infty$ such that the following functional stable convergence holds:

$$\nu_M\left(a^M - a, \sigma^M - \sigma\right) \xrightarrow{d_{st}} U = (U^1, U^2) \qquad \text{on } D^2([t_0,T]), \tag{15}$$

where the convergence is on the space of bivariate càdlàg functions defined on $[t_0,T]$ equipped with the Skorohod topology $D^2([t_0,T])$. Let us briefly recall the notion of stable convergence, which is originally due to Rényi [27]. We say that a sequence of random variables Y^n with values in a Polish space $(E,\mathscr{E})$ converges stably in law to Y, where Y is defined on an extension $(\Omega',\mathscr{F}',\mathbb{P}')$ of the original probability $(\Omega,\mathscr{F},\mathbb{P})$ if and only if

$$\lim_{n\to\infty}\mathbb{E}[f(Y^n)Z] = \mathbb{E}'[f(Y)Z]$$

for any bounded and continuous function f and any bounded $\mathscr{F}$-measurable random variable Z. In this case we write ($Y^n \xrightarrow{d_{st}} Y$). In the following we will deal with the space of càdlàg processes equipped with the Skorohod topology or with the space of continuous processes equipped with the uniform topology. We refer to [1, 23] or [27] for a detailed study of stable convergence. Note that stable convergence is a stronger mode of convergence than weak convergence, but it is weaker than uniform convergence in probability.

We remark that the estimators $(a_t^M)_{t\in[t_0,T]}$ and $(\sigma_t^M)_{t\in[t_0,T]}$ might have a different *effective* convergence rate. In this case we will have either $U_1 \equiv 0$ or $U_2 \equiv 0$.

Now, we basically follow the Fourier type approach, which refers to (5) and the definition of the function $\widehat{h}_\lambda$, described in the previous section, but we replace the Fourier transform approximation proposed at (6) by a Riemann sum approximation. More specifically, we introduce the approximation

$$\begin{aligned} h(x) &= \frac{\exp(-\lambda|x|)}{2\pi} \int_{\mathbb{R}} \widehat{h}_\lambda(y) \exp(ixy) \mathrm{d}y \\ &\approx \widetilde{h}_N(x) := \frac{\exp(-\lambda|x|)}{\pi N} \sum_{k=0}^{c_N} \widehat{h}_\lambda\left(\frac{k}{N}\right) \cos\left(\frac{kx}{N}\right), \end{aligned} \tag{16}$$

where c_N is a sequence of numbers in $\mathbb{N}$ satisfying $c_N/N \to \infty$ as $N \to \infty$. In the following we will also assume that the sequence c_N additionally satisfies the condition

$$N \int_{c_N/N}^{\infty} |\widehat{h}_\lambda(y)| \mathrm{d}y \to 0 \qquad \text{as } N \to \infty. \tag{17}$$

Clearly, such a sequence exists, since $\widehat{h}_\lambda \in L^1(\mathbb{R})$. When introducing the approximation at (16), we obviously obtain two types of error: The Riemann sum approximation error and tail approximation error. Condition (17) guarantees that the Riemann sum approximation error will dominate.

Remark 4 Under some stronger conditions the tail integral at (17) can be bounded from above explicitly. Assume that $h \in C^2([-\tau, \tau])$ such that $h'(\tau) = 0$ (cf. Remark 2). Then a repeated application of integration by parts formula implies the identity

$$\widehat{h}_\lambda(y) = \int_{\mathbb{R}} h_\lambda(x) \cos(yx) \mathrm{d}x = -\frac{1}{y^2} \int_{-\tau}^{\tau} h''_\lambda(x) \cos(yx) \mathrm{d}x$$

for any $y > 0$. Thus, for any $u > 0$, we deduce the inequality

$$\int_u^\infty |\widehat{h}_\lambda(y)| \mathrm{d}y \le C\|h''\|_{L^1} \int_u^\infty y^{-2} \mathrm{d}y \le C\|h''\|_{L^1} u^{-1}.$$

Hence, condition (17) holds whenever $N^2/c_N \to 0$ as $N \to \infty$. □

Remark 5 We remark that the Fourier transform used at (6) comes from the L^2 theory. Thus, in contrast to the L^2-distance $\|h - h_N\|_{L^2}$, the limiting behaviour of a standardized version of $h(x) - h_N(x)$ is difficult to study pointwise. This is precisely the reason why we use the Riemann sum approximation instead, for which we will show the convergence of $N(h(x) - \widetilde{h}_N(x))$.

If one can freely choose the simulation rates N and M, the Fourier transform of (6) is numerically more preferable. According to the estimate (11) and the upper bound for the Fourier coefficient of Remark 2 applied for $n = 1$, we readily deduce the rate N^{-1} for the L^2-error approximation connected to (6). On the other hand, the effective sample size of the Riemann approximation at (16) is c_N. In the setting of the previous remark the overall Riemann approximation error is $\max(N^{-1}, N/c_N)$. Recalling that $c_N/N \to \infty$, the obtained rate is definitely slower than the one associated with Fourier approximation proposed at (6).

Nevertheless, as our aim is to precisely determine the asymptotics associated with the N scale, we will discuss the Riemann approximation approach in the sequel. A statement about the Fourier transform (6) will be presented in Remark 8. □

Now, we essentially proceed as in the steps (8)–(10). First of all, it holds that

$$\begin{aligned}\int_u^t g(t-s)\sigma_s W(\mathrm{d}s) &\approx \int_u^t \widetilde{h}_N(t-s)\sigma_s W(\mathrm{d}s) \\ &= \int_u^t \exp(-\lambda(t-s)) \left\{ \sum_{k=0}^{c_N} \widetilde{b}_k \cos(\frac{k(t-s)}{N}) \right\} \sigma_s W(\mathrm{d}s) \\ &= \mathrm{Re} \sum_{k=0}^{c_N} \widetilde{b}_k \widehat{X}_{\lambda,u}(t, \frac{k}{N}),\end{aligned} \tag{18}$$

where the complex valued stochastic field $\widehat{X}_{\lambda,u}(t,y)$ is defined at (9) and $\widetilde{b}_k = \widehat{h}_\lambda(k/N)/(\pi N)$. In a second step, for $\delta > 0$, we obtain the approximation

$$\begin{aligned}\widehat{X}_{\lambda,u}(t+\delta, y) &= \exp\{(-\lambda + iy)\delta\} \left(\widehat{X}_{\lambda,u}(t,y) + \int_t^{t+\delta} \exp\{(-\lambda+iy)(t-s)\}\sigma_s W(\mathrm{d}s) \right) \\ &\approx \exp\{(-\lambda + iy)\delta\} \left(\widehat{X}_{\lambda,u}(t,y) + \int_t^{t+\delta} \exp\{(-\lambda+iy)(t-s)\}\sigma_s^M W(\mathrm{d}s) \right).\end{aligned} \tag{19}$$

When the estimator σ^M is assumed to be constant on intervals $[s_{i-1}, s_i)$, $i = 1, \ldots, M$, the last integral at (19) can be easily simulated (cf. (10)). We remark that this approximation procedure slightly differs from (10) as now we leave the exponential term unchanged.

In summary, given that the information up to time t_0 is available, we arrive at the simulated value

$$X_t^{N,M} := \int_{t_0}^{t} \widetilde{h}_N(t-s)\sigma_s^M W(\mathrm{d}s) + \int_{t_0}^{t} q(t-s)a_s^M \mathrm{d}s \tag{20}$$

of the random variable

$$X_t^0 = \int_{t_0}^{t} g(t-s)\sigma_s W(\mathrm{d}s) + \int_{t_0}^{t} q(t-s)a_s \mathrm{d}s. \tag{21}$$

Note that the drift part of the Brownian semi-stationary process X is estimated in a direct manner, although other methods similar to the treatment of the Brownian part are possible. Now, we wish to study the asymptotic theory for the approximation error $X_t^{N,M} - X_t^0$. Our first result analyzes the limiting behaviour of the function $N(h(x) - \widetilde{h}_N(x))$.

Lemma 1 *Define the function* $\psi_N(x) := N(h(x) - \widetilde{h}_N(x))$. *Let us assume that the condition*

$$\widehat{yh_\lambda(y)} \in L^1(\mathbb{R}), \qquad \widehat{y^2 h_\lambda(y)} \in L^1(\mathbb{R}) \tag{22}$$

holds. Then, under Assumption (A), (17) *and* (22)*, it holds that*

$$\psi_N(x) \to \psi(x) = -\frac{\widehat{h}_\lambda(0)}{2\pi}\exp(-\lambda|x|) \qquad \textit{as } N \to \infty \tag{23}$$

for any $x \in \mathbb{R}$. *Furthermore, it holds that*

$$\sup_{N\in\mathbb{N},\ x\in[0,T]} |\psi_N(x)| \le C$$

for any $T > 0$.

Proof First, we recall a well known result from Fourier analysis (see e.g. [19, Theorem 8.22]): The condition (22) implies that

$$\widehat{h}'_\lambda \in L^1(\mathbb{R}), \qquad \widehat{h}''_\lambda \in L^1(\mathbb{R}). \tag{24}$$

Now, observe the decomposition

$$\begin{aligned}
\psi_N(x) &= \frac{N\exp(-\lambda|x|)}{\pi}\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N}\left(\kappa_x(y) - \kappa_x(\frac{k}{N})\right)\mathrm{d}y \\
&\quad + \frac{N\exp(-\lambda|x|)}{\pi}\int_{(c_N+1)/N}^{\infty}\kappa_x(y)\mathrm{d}y \\
&= \frac{N\exp(-\lambda|x|)}{\pi}\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N}\left(\kappa_x(y) - \kappa_x(\frac{k}{N})\right)\mathrm{d}y + o(1),
\end{aligned}$$

where $\kappa_x(y) = \widehat{h}_\lambda(y)\cos(yx)$ and the approximation follows by the inequality $|\kappa_x(y)| \leq |\widehat{h}_\lambda(y)|$ and condition (17). Let us denote by $\kappa'_x(y)$ the derivative of $\kappa_x(y)$ with respect to y. Since $\kappa'_x(\cdot), \kappa''_x(\cdot) \in L^1(\mathbb{R}_{\geq 0})$ because of (24), we deduce that

$$\begin{aligned}\psi_N(x) &= \frac{N\exp(-\lambda|x|)}{\pi}\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N}\kappa'_x(\frac{k}{N})\left(y-\frac{k}{N}\right)\mathrm{d}y + o(1)\\ &= \frac{\exp(-\lambda|x|)}{2\pi N}\sum_{k=0}^{c_N}\kappa'_x(\frac{k}{N}) + o(1)\\ &\to \frac{\exp(-\lambda|x|)}{2\pi}\int_0^\infty \kappa'_x(y)\mathrm{d}y \qquad \text{as } N\to\infty.\end{aligned}$$

But, since $\widehat{h}_\lambda$ vanishes at infinity, we readily obtain that

$$\int_0^\infty \kappa'_x(y)\mathrm{d}y = -\widehat{h}_\lambda(0).$$

In order to prove the second assertion of the lemma, we observe the inequality

$$|\psi_N(x)| \leq \frac{\exp(-\lambda|x|)}{\pi}\sum_{k=0}^{c_N}\int_{k/N}^{(k+1)/N}|\kappa'_x(\zeta_{k,N}(y))|\mathrm{d}y + N\int_{c_N/N}^\infty |\widehat{h}_\lambda(y)|\mathrm{d}y,$$

where $\zeta_{k,N}(y)$ is a certain value with $\zeta_{k,N}(y) \in (k/N, y)$. Clearly, the second term in the above approximation is bounded in N, since it converges to 0. On the other hand, we have that $|\kappa'_x(y)| \leq |x||\widehat{h}_\lambda(y)| + |\widehat{h}'_\lambda(y)|$, and since $\widehat{h}_\lambda, \widehat{h}'_\lambda \in L^1(\mathbb{R}_{\geq 0})$, we readily deduce that

$$\sup_{N\in\mathbb{N},\ x\in[0,T]} |\psi_N(x)| \leq C.$$

This completes the proof of the lemma. □

At this stage we need a further condition on the kernel function g to prove tightness later.

Assumption (B):

(i) The kernel function g has the form

$$g(x) = x^\alpha f(x)$$

for some $\alpha \geq 0$ and function f satisfying $f(0) \neq 0$.

(ii) $f \in C^1(\mathbb{R}_{\geq 0})$ has bounded support contained in $[0, \tau]$.

Notice that the assumption $\alpha \geq 0$ is in accordance with the condition (A)(ii). Assumption (B) implies the following approximation result:

$$\int_0^1 |g(x+\delta) - g(x)|^4 \mathrm{d}x \leq \begin{cases} C\delta^4 & \alpha = 0 \\ C\delta^{\min(4,4\alpha+1)} & \alpha > 0 \end{cases} \tag{25}$$

for $\delta \in [0, T]$. The case $\alpha = 0$ is trivial, while the other one follows along the lines of the proof of [20, Lemma 4.1]. As a matter of fact, we also require a good estimate of the left side of (25) when the kernel g is replaced by the function ψ_N defined in Lemma 1. In the following we will assume that

$$\sup_{N \in \mathbb{N}} \int_0^1 |\psi_N(x+\delta) - \psi_N(x)|^4 \mathrm{d}x \leq C\delta^{1+\varepsilon} \tag{26}$$

for some $\varepsilon > 0$ and $\delta \in [0, T]$.

Remark 6 Unfortunately, we have not been able to show the statement of (26) under the mere assumption of, say, condition (B). Obviously, as in the case of function g, condition (26) would hold if

$$\psi_N(x) = x^\alpha f_N(x),$$

where $f_N \in C^1(\mathbb{R}_{\geq 0})$ with uniformly bounded derivative in $N \in \mathbb{N}$ and x in a compact interval. We can prove condition (26) explicitly when the function g is differentiable. Assume that $y\widehat{h}_\lambda(y), y\widehat{h}'_\lambda(y) \in L^1(\mathbb{R}_{\geq 0})$ and c_N is chosen in such a way that the condition

$$N \int_{c_N/N}^{\infty} |y\widehat{h}_\lambda(y)|\mathrm{d}y \to 0 \qquad as\ N \to \infty$$

is satisfied. As in Lemma 1 we conclude that $(|\partial_x \partial_y \kappa_x(y)| \leq (|\widehat{h}_\lambda(y)| + |yx\widehat{h}_\lambda(y)| + |y\widehat{h}'_\lambda(y)|)$ and, as in the proof of Lemma 1, we deduce that

$$\begin{aligned} \sup_{x \in [0,T]} |\psi'_N(x)| \leq C \Bigg(& N \int_{c_N/N}^{\infty} |y\widehat{h}_\lambda(y)|\mathrm{d}y + N \int_{c_N/N}^{\infty} |\widehat{h}_\lambda(y)|\mathrm{d}y \\ & + \sum_{k=0}^{c_N} \int_{k/N}^{(k+1)/N} |\partial_y \kappa_x(\zeta_{k,N}(y))|\mathrm{d}y + \sum_{k=0}^{c_N} \int_{k/N}^{(k+1)/N} |\partial_x \partial_y \kappa_x(\widetilde{\zeta}_{k,N}(y))|\mathrm{d}y \Bigg) \end{aligned}$$

for certain values $\zeta_{k,N}(y), \widetilde{\zeta}_{k,N}(y)$ in the interval $(k/N, y)$. Then, due to our integrability conditions, we obtain

$$\sup_{N \in \mathbb{N},\ x \in [0,T]} |\psi'_N(x)| < \infty.$$

Moreover, condition (26) is trivially satisfied due to mean value theorem. However, showing (26) under Assumption (B) seems to be a much harder problem for $\alpha \in (0, 1)$. $\square$

The next result is the main theorem of our paper.

Theorem 1 *Assume that conditions (A), (B), (17), (22) and (26) hold, and the processes* $(\sigma_t)_{t\in[t_0,T]}$ *and* $(\sigma_t^M)_{t\in[t_0,T]}$ *has finite fourth moment with* $\sup_{t\in[t_0,T]} \mathbb{E}[\sigma_t^4] < \infty$ *and* $\sup_{t\in[t_0,T]} \sup_{M\in\mathbb{N}} \mathbb{E}[(\sigma_t^M)^4] < \infty$. *We also assume that the process* $U_t^M = \nu_M(\sigma_t^M - \sigma_t)$ *satisfies*

$$\sup_{t\in[t_0,T]} \sup_{M\in\mathbb{N}} \mathbb{E}[(U_t^M)^4] < \infty. \tag{27}$$

Then we obtain the decomposition

$$X_t^{N,M} - X_t^0 = A_t^{N,M} + B_t^M$$

such that

$$NA^{N,M} \overset{u.c.p.}{\Longrightarrow} A = \frac{\widehat{h}_\lambda(0)}{2\pi} \int_{t_0}^{\cdot} \exp(-\lambda(\cdot - s))\sigma_s W(ds) \quad \text{as } N, M \to \infty, \tag{28}$$

where ucp convergence means that $\sup_{t\in[t_0,T]} |A_t^N - A_t| \xrightarrow{\mathbb{P}} 0$, *and*

$$\nu_M B^M \xrightarrow{d_{st}} B = \int_{t_0}^{\cdot} g(\cdot - s)U_s^2 W(ds) + \int_{t_0}^{\cdot} q(\cdot - s)U_s^1 ds \quad \text{as } M \to \infty, \tag{29}$$

where the stable convergence holds on the space $C([t_0, T])$ *equipped with the uniform topology.*

Proof We start with the decomposition $X_t^{N,M} - X_t^0 = A^{N,M} + B_t^M$, where

$$A_t^{N,M} = \int_{t_0}^{t} \{\widetilde{h}_N(t-s) - g(t-s)\}\sigma_s^M W(\mathrm{d}s),$$
$$B_t^M = \int_{t_0}^{t} g(t-s)\{\sigma_s^M - \sigma_s\}W(\mathrm{d}s) + \int_{t_0}^{t} q(t-s)\{a_s^M - a_s\}\mathrm{d}s.$$

We begin by proving the stable convergence in (29). Let us first recall a classical result about weak convergence of semimartingales (see [23, Theorem VI.6.22] or [24]): Let $(Y_s^n)_{s\in[t_0,T]}$ be a sequence of càdlàg processes such that $Y^n \xrightarrow{d_{st}} Y$ on $D([t_0, T])$ equipped with the Skorohod topology. Then we obtain the weak convergence

$$\int_{t_0}^{\cdot} Y_s^n W(\mathrm{d}s) \Longrightarrow \int_{t_0}^{\cdot} Y_s W(\mathrm{d}s) \quad \text{on } C([t_0, T])$$

equipped with the uniform topology. This theorem is an easy version of the general result, since the integrator W does not depend on n and hence automatically fulfills the P-UT property. The stable nature of the aforementioned weak convergence follows by joint convergence $(\int_0^\cdot Y_s^n W(\mathrm{d}s), Y^n, W) \Longrightarrow (\int_0^\cdot Y_s W(\mathrm{d}s), Y, W)$ (cf. [24]). Hence, we deduce that

$$\int_0^\cdot Y_s^n W(\mathrm{d}s) \xrightarrow{d_{st}} \int_0^\cdot Y_s W(\mathrm{d}s) \qquad \text{on } C([t_0, T]) \tag{30}$$

equipped with the uniform topology. It is important to note that this result can not be directly applied to the process B_t^M, since this process is not a semimartingale in general. Thus, we will prove the stable convergence (29) by showing the stable convergence of finite dimensional distributions and tightness.

We fix $u_1, \ldots, u_k \in [t_0, T]$. Due to the condition (15), the finite dimensional version of (30) and continuous mapping theorem for stable convergence, we conclude the joint stable convergence

$$\begin{aligned}
&\left(\left\{\nu_M \int_{t_0}^{u_j} g(u_j - s)\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s)\right\}_{j=1,\ldots,k}, \nu_M \int_{t_0}^{\cdot} q(\cdot - s)\{a_s^M - a_s\}\mathrm{d}s\right) \\
&\xrightarrow{d_{st}} \left(\left\{\int_{t_0}^{u_j} g(u_j - s) U_s^2 W(\mathrm{d}s)\right\}_{j=1,\ldots,k}, \int_{t_0}^{\cdot} q(\cdot - s) U_s^1 \mathrm{d}s\right)
\end{aligned} \tag{31}$$

as $M \to \infty$. Here we remark that the stable convergence for the second component indeed holds, since the mapping $F : C([t_0, \tau]) \times D([t_0, T]) \to C([t_0, T])$, $F(q, a) = \int_{t_0}^{\cdot} q(\cdot - s) a_s \mathrm{d}s$ is continuous. Hence, we are left with proving tightness for the first component of the process B_t^M. We fix $u, t \in [t_0, T]$ with $t > u$ and observe the decomposition

$$\begin{aligned}
&\nu_M \left(\int_{t_0}^{t} g(t - s)\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s) - \int_{t_0}^{u} g(u - s)\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s)\right) \\
&= \nu_M \left(\int_{u}^{t} g(t - s)\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s)\right. \\
&\left. + \int_{t_0}^{u} \{g(t - s) - g(u - s)\}\{\sigma_s^M - \sigma_s\} W(\mathrm{d}s)\right) := R_M^{(1)}(t, u) + R_M^{(2)}(t, u).
\end{aligned}$$

Using Burkholder and Cauchy-Schwarz inequalities and (27), we have

$$\mathbb{E}[|R_M^{(1)}(t, u)|^4] \le C(t - u) \int_u^t |g(t - s)|^4 \mathrm{d}s.$$

Thus, we conclude that

$$\mathbb{E}[|R_M^{(1)}(t, u)|^4] \le C(t - u)^2. \tag{32}$$

Now, using the same methods we conclude that

$$\mathbb{E}[|R_M^{(2)}(t,u)|^4] \le C\int_{t_0}^{u} |g(t-s)-g(u-s)|^4 \mathrm{d}s \le C(t-u)^{\min(4,4\alpha+1)}, \quad (33)$$

where we used the inequality (25). Thus, applying (32), (33) and the Kolmogorov's tightness criteria, we deduce the tightness of the first component of the process B_t^M. This completes the proof of (29).

Now, we show the pointwise convergence at (28). Recalling the notation from (23), we need to show that

$$\int_{t_0}^{t} \{\psi_N(t-s)-\psi(t-s)\}\sigma_s^M W(\mathrm{d}s) \xrightarrow{\mathbb{P}} 0 \qquad \text{as } N, M \to \infty$$

for a fixed t. The Itô isometry immediately implies that

$$\sup_{M\in\mathbb{N}} \mathbb{E}\left[\left|\int_{t_0}^{t} \{\psi_N(t-s)-\psi(t-s)\}\sigma_s^M W(\mathrm{d}s)\right|^2\right] \le C\int_{t_0}^{t} \{\psi_N(t-s)-\psi(t-s)\}^2 \mathrm{d}s$$
$$\to 0 \qquad \text{as } N \to \infty,$$

which follows by Lemma 1 and the dominated convergence theorem. Hence, we obtain pointwise convergence at (28). Since the limiting process A is continuous, we now need to show that

$$\sup_{N,M\in\mathbb{N}} \mathbb{E}[N^4(A_t^{N,M}-A_u^{N,M})^4] \le C(t-u)^{1+\varepsilon}$$

for $t_0 < u < t$, to conclude ucp convergence from pointwise convergence in probability. Applying the same methods as in (32), (33) we deduce the inequality

$$\begin{aligned}
&\sup_{N,M\in\mathbb{N}} \mathbb{E}[N^4(A_t^{N,M}-A_u^{N,M})^4] \\
&\le C\left((t-u)\int_{u}^{t} |\psi_N(t-s)|^4 \mathrm{d}s + \int_{t_0}^{u} |\psi_N(t-s)-\psi_N(u-s)|^4 \mathrm{d}s\right) \\
&\le C(t-u)^{1+\varepsilon},
\end{aligned}$$

which follows by Lemma 1 and condition (26). This completes the proof of Theorem 1. □

Remark 7 We remark that the stronger conditions (B) and (26) are not required to prove the finite dimensional version of convergence (28) and (29). □

Theorem 1 immediately applies to the weak approximation error analysis. Assume for simplicity that $M = M(N)$ is chosen such that $\nu_M/N \to 1$, so that the Riemann sum approximation error and the simulation error from (15) are balanced. We consider a bounded test function $\varphi \in C^1(\mathbb{R})$ with bounded derivative. The mean value theorem implies the identity

$$\varphi(X_t^{N,M}) - \varphi(X_t^0) = \varphi'(\xi_{N,M})(X_t^{N,M} - X_t^0),$$

where $\xi_{N,M}$ is a random value between X_t^0 and $X_t^{N,M}$ with $\xi_{N,M} \xrightarrow{\mathbb{P}} X_t^0$ as $N \to \infty$. By properties of stable convergence we deduce that $(\xi_{N,M}, N(X_t^{N,M} - X_t^0)) \xrightarrow{d_{st}} (X_t^0, A_t + B_t)$. Hence, given the existence of the involved expectations, we conclude that

$$\mathbb{E}[\varphi(X_t^{N,M})] - \mathbb{E}[\varphi(X_t^0)] = N^{-1}\mathrm{e}'[\varphi'(X_t^0)(A_t + B_t)] + o(N^{-1}). \tag{34}$$

(Recall that the limit $A_t + B_t$ is defined on the extended probability space $(\Omega', \mathscr{F}', \mathbb{P}')$).

Remark 8 The results of Theorem 1 may also apply to the original Fourier approximated method proposed in [14, 18]. Let us keep the notation of this section and still denote the approximated value of X_t^0 by $X_t^{N,M}$. Recalling the result of (11) (see also Remark 2) and assuming that $M = M(N)$ is chosen such that $\sum_{k=N+1}^{\infty} |b_k| \ll \nu_M$, we readily deduce that

$$\nu_M(X_t^{N,M} - X_t^0) \xrightarrow{d_{st}} B_t.$$

□

Remark 9 The results of Theorem 1 might transfer to the case of Lévy semi-stationary processes

$$X_t = \mu + \int_{-\infty}^{t} g(t-s)\sigma_s L(\mathrm{d}s) + \int_{-\infty}^{t} q(t-s)a_s \mathrm{d}s$$

under suitable moment assumptions on the driving Lévy motion L (cf. [14]). However, when L is e.g. a β-stable process with $\beta \in (0, 2)$, it seems to be much harder to access the weak limit of the approximation error. □

In the following we will present some examples of convergence at (15) to highlight the most prominent results. For simplicity we assume that $a \equiv 0$ in all cases.

Example 1 Let us consider a continuous diffusion model for the volatility process σ, i.e.

$$\mathrm{d}\sigma_t = \widetilde{a}(\sigma_t)\mathrm{d}t + \widetilde{v}(\sigma_t)\mathrm{d}B_t, \qquad \sigma_{t_0} = x_0,$$

where B is a Brownian motion possibly correlated with W. We consider an equidistant partition $t_0 = s_0 < s_1 < \cdots < s_M = T$ of the interval $[t_0, T]$ and define the continuous Euler approximation of σ_t via

$$\sigma_t^M = \sigma_{s_k}^M + \widetilde{a}(\sigma_{s_k}^M)(t - s_k) + \widetilde{v}(\sigma_{s_k}^M)(B_t - B_{s_k}), \qquad t \in [s_k, s_{k+1}].$$

When the functions $\widetilde{a}$ and $\widetilde{v}$ are assumed to be globally Lipschitz and continuously differentiable, it holds that

$$\sqrt{M}(\sigma^M - \sigma) \xrightarrow{d_{st}} U^2 \qquad \text{on } C([t_0, T]),$$

where U^2 is the unique solution of the stochastic differential equation

$$\mathrm{d}U_t^2 = \widetilde{a}'(\sigma_t)U_t^2 \mathrm{d}t + \widetilde{v}'(\sigma_t)U_t^2 \mathrm{d}B_t - \frac{1}{\sqrt{2}}\widetilde{v}\widetilde{v}'(\sigma_t)\mathrm{d}W_t',$$

where W' is a new Brownian motion independent of $\mathscr{F}$. We refer to [22, Theorem 1.2] for a detailed treatment of this result. □

Example 2 Let us now consider a discontinuous diffusion model for the volatility process σ, i.e.

$$\mathrm{d}\sigma_t = \widetilde{v}(\sigma_{t-})\mathrm{d}L_t, \qquad \sigma_{t_0} = x_0,$$

where L is a purely discontinuous Lévy process. In this framework we study the discretized Euler scheme given via

$$\sigma^M_{s_{k+1}} = \widetilde{v}(\sigma^M_{s_k})(L_{s_{k+1}} - L_{s_k}), \qquad k = 0, \ldots, M-1.$$

We define the process $U_t^M = \sigma^M_{[tM]/M} - \sigma_{[tM]/M}$. In [21] several classes of Lévy processes L has been studied. For the sake of exposition we demonstrate the case of a symmetric β-stable Lévy process L with $\beta \in (0, 2)$. Let us assume that $\widetilde{v} \in C^3(\mathbb{R})$. Then, it holds that

$$(M/\log(M))^{1/\beta} U^M \xrightarrow{d_{st}} U^2 \qquad \text{on } D([t_0, T]),$$

where U^2 is the unique solution of the linear equation

$$\mathrm{d}U_t^2 = \widetilde{v}'(\sigma_{t-})U_{t-}^2 \mathrm{d}L_t - \widetilde{v}\widetilde{v}'(\sigma_{t-})\mathrm{d}L_t'$$

and L' is another symmetric β-stable Lévy process (with certain scaling parameter) independent of $\mathscr{F}$. We note that this result does not directly correspond to our condition (15) as the discretized process $\sigma_{[tM]/M}$ is used in the definition of U_t^M. □

Acknowledgments Nopporn Thamrongrat gratefully acknowledges support by Deutsche Forschungsgemeinschaft through the Research Training Group RTG 1953.

References

1. Aldous, D.J., Eagleson, G.K.: On mixing and stability of limit theorems. Ann. Probab. **6**(2), 325–331 (1978)
2. Barndorff-Nielsen, O.E., Benth, F.E., Pedersen, J., Veraart, A.: On stochastic integration for volatility modulated Lévydriven Volterra processes. Stochast. Process. Appl. **124**(1), 812–847 (2014)
3. Barndorff-Nielsen, O.E., Benth, F.E., Szozda, B.: On stochastic integration for volatility modulated Brownian-driven Volterra processes via white noise analysis. Infin. Dimens. Anal. Quantum Probab. Relat. Top. **17**(2), 1450011 (2014)
4. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.: Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. Bernoulli. **19**(3), 803–845 (2013)
5. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.: Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. To appear in Banach Center Publications (2014)
6. Barndorff-Nielsen, O.E., Corcuera, J.M., Podolskij, M.: Multipower variation for Brownian semistationary processes. Bernoulli. **17**(4), 1159–1194 (2011)
7. Barndorff-Nielsen, O.E., Corcuera, J.M., Podolskij, M.: Limit theorems for functionals of higher order differences of Brownian semi-stationary processes. In: Honor of Yuri Prokhorov, V., Shiryaev, A.N., Varadhan, S.R.S., Presman E.L. (eds.) Prokhorov and Contemporary Probability Theory, vol. 33, pp. 69–96. Springer (2013)
8. Barndorff-Nielsen, O.E., Jensen, E.B.V., Jónsdóttir, K.Y., Schmiegel, J.: Spatio-temporal modelling—with a view to biological growth. In: Finkenstdt, B., Held L., Isham, V. Statistical Methods for Spatio-Temporal Systems, pp. 47–75. Chapman and Hall/CRC, London (2007)
9. Barndorff-Nielsen, O.E., Schmiegel, J.: Ambit processes; with applications to turbulence and cancer growth. In: Benth, F.E., Nunno, G.D., Linstrøm, T., Øksendal, B., Zhang, T. (eds.). Stochastic Analysis and Applications. The Abel Symposium 2005, pp. 93–124. Springer, Heidelberg (2007)
10. Barndorff-Nielsen, O.E., Schmiegel, J.: Time change, volatility and turbulence. In: Sarychev, A., Shiryaev, A., Guerra, M., Grossinho, M.D.R. (eds.) Proceedings of the Workshop on Mathematical Control Theory and Finance, Lisbon 2007, pp. 29–53. Springer, Berlin (2008)
11. Barndorff-Nielsen, O.E., Schmiegel, J.: Brownian semistationary processes and volatility/intermittency. In: Albrecher, H., Runggaldier, W., Schachermayer, W. (eds.) Advanced Financial Modelling. Radon Series on Computational and Applied Mathematics, vol. 8, pp. 1–26 (2009)
12. Basse-O'Connor, A., Lachièze-Rey, R., Podolskij, M.: Limit theorems for stationary increments Lévy driven moving average processes. Working paper (2015)
13. Benth, F.E., Eyjolfsson, H.: Simulation of volatility modulated Volterra processes using hyperbolic stochastic partial differential equations. To appear in Bernoulli (2015)
14. Benth, F.E., Eyjolfsson, H., Veraart, A.E.D.: Approximating Lévy semistationary processes via Fourier methods in the context of power markets. SIAM J. Financ. Math. **5**(1), 71–98 (2014)
15. Benth, F.E., Süß, A.: Integration theory for infinite dimensional volatility modulated Volterra processes. Working paper. arXiv:1303.7143. To appear in Bernoulli (2013)
16. Chen, B., Chong, C., Klüppelberg, C.: Simulation of stochastic Volterra equations driven by spacetime Lévy noise. Working paper. arXiv:1501.01645 (2015)
17. Corcuera, J.M., Hedevang, E., Pakkanen, M., Podolskij, M.: Asymptotic theory for Brownian semi-stationary processes with application to turbulence. Stochastic Process. Appl. **123**, 2552–2574 (2013)
18. Eyjolfsson, H.: Approximation methods for ambit fields applied to power markets. Ph.D. thesis, Department of Mathematics, University of Oslo (2013)
19. Folland, G.B.: Real analysis—modern techniques and their applications. Wiley&Sons, J. (1984)
20. Gärtner, K., Podolskij, M.: On non-standard limits of Brownian semi-stationary processes. Stochastic Process. Appl. **125**(2), 653–677 (2015)

21. Jacod, J.: The Euler scheme for Lévy driven stochastic differential equations: limit theorems. Ann. Probab. **32**(3), 1830–1872 (2004)
22. Jacod, J., Protter, P.: Asymptotic error distributions for the Euler method for stochastic differential equations. Ann. Probab. **26**(1), 267–307 (1998)
23. Jacod, J., Shiryaev, A.N.: Limit theorems for stochastic processes, 2d edn. Springer, Berlin (2002)
24. Kurtz, T.G., Protter, P.: Weak limit theorems for stochastic integrals and stochastic differential equations. Ann. Probab. **19**(3), 1035–1070 (1993)
25. Pakkanen, M.: Limit theorems for power variations of ambit fields driven by white noise. Stochastic Process. Appl. **124**(5), 1942–1973 (2014)
26. Podolskij, M.: Ambit fields: survey and new challenges. In: Mena, R., Pardo, J.C., Rivero, V., Bravo, G.U. (eds.) To appear in Special Volume of XI Symposium of Probability and Stochastic Processes (2014)
27. Rényi, A.: On stable sequences of events. Sankhyā Ser. A. **25**, 293–302 (1963)

Non-elliptic SPDEs and Ambit Fields: Existence of Densities

Marta Sanz-Solé and André Süß

Abstract Relying on the method developed in [11], we prove the existence of a density for two different examples of random fields indexed by $(t, x) \in (0, T] \times \mathbb{R}^d$. The first example consists of SPDEs with Lipschitz continuous coefficients driven by a Gaussian noise white in time and with a stationary spatial covariance, in the setting of [9]. The density exists on the set where the nonlinearity σ of the noise does not vanish. This complements the results in [20] where σ is assumed to be bounded away from zero. The second example is an ambit field with a stochastic integral term having as integrator a Lévy basis of pure-jump, stable-like type.

Keywords Stochastic partial differential equations · Stochastic wave equation · Ambit fields · Densities

1 Introduction

Malliavin calculus has proved to be a powerful tool for the study of questions concerning the probability laws of random vectors, ranging from its very existence to the study of their properties and applications. Malliavin's probabilistic proof of Hörman-

M. Sanz-Solé: Supported by the grant MTM 2012-31192 from the *Dirección General de Investigación Científica y Técnica, Ministerio de Economía y Competitividad, Spain*.
A. Süß: Supported by the 2014/2015 SEFE project at the Centre for Advanced Study at the Norwegian Academy of Sciences and Letters.

M. Sanz-Solé (✉)
Facultat de Matemàtiques, University of Barcelona,
Gran Via de Les Corts Catalanes 585, 08007 Barcelona, Spain
e-mail: marta.sanz@ub.edu
url: http://www.ub.edu/plie/Sanz-Sole

A. Süß
Centre for Advanced Study at the Norwegian Academy of Sciences and Letters,
Drammensveien 78, 0271 Oslo, Norway
e-mail: suess.andre@web.de

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_5

der's hypoellipticity theorem for differential operators in quadratic form provided the existence of an infinitely differentiable density with respect to the Lebesgue measure on $\mathbb{R}^m$ for the law at a fixed time $t > 0$ of the solution to a stochastic differential equation (SDE) on $\mathbb{R}^m$ driven by a multi-dimension Brownian motion. The classical Malliavin's criterion for existence and regularity of densities (see, e.g. [13]) requires strong regularity of the random vector X under consideration. In fact, X should be in the space $\mathbb{D}^\infty$, meaning that it belongs to Sobolev type spaces of any degree. As a consequence, many interesting examples are out of the range of the theory, for example, SDE with Hölder continuous coefficients, and others that will be mentioned throughout this introduction.

Recently, there have been several attempts to develop techniques to prove existence of density, under weaker regularity conditions than in the Malliavin's theory, but providing much less information on the properties of the density. The idea is to avoid applying integration by parts, and use instead some approximation procedures. A pioneer work in this direction is [12], where the random vector X is compared with a *good* approximation X^ε whose law is known. The proposal of the random vector X^ε is inspired by Euler numerical approximations and the comparison is done through their respective Fourier transforms. The method is illustrated with several one-dimensional examples of stochastic equations, all of them having in common that the *diffusion coefficient* is Hölder continuous and the drift term is a measurable function: SDEs, including cases of random coefficients, a stochastic heat equation with Neumann boundary conditions, and a SDE driven by a Lévy process.

With a similar motivation, and relying also on the idea of *approximation*, A. Debussche and M. Romito prove a useful criterion for the existence of density of random vectors, see [11]. In comparison with [12], the result is formulated in an abstract form, it applies to multidimensional random vectors and provides additionally information on the space where the density lives. The precise statement is given in Lemma 1. As an illustration of the method, [11] considers finite dimensional functionals of the solutions of the stochastic Navier-Stokes equations in dimension 3, and in [10] SDEs driven by stable-like Lévy processes with Hölder continuous coefficients. A similar methodology has been applied in [1, 2, 4]. The more recent work [3] applies interpolation arguments on Orlicz spaces to obtain absolute continuity results of finite measures. Variants of the criteria provide different types of properties of the density. The results are illustrated by diffusion processes with log-Hölder coefficients and piecewise deterministic Markov processes.

Some of the methods developed in the references mentioned so far are also well-suited to the analysis of stochastic partial differential equations (SPDEs) defined by non-smooth differential operators. Indeed, consider a class of SPDEs defined by

$$Lu(t,x) = b(u(t,x)) + \sigma(u(t,x))\dot{F}(t,x),\ \ (t,x) \in (0,T] \times \mathbb{R}^d, \tag{1}$$

with constant initial conditions, where L denotes a linear differential operator, $\sigma, b : \mathbb{R} \to \mathbb{R}$, and F is a Gaussian noise, white in time with some spatial correlation (see Sect. 2 for the description of F). Under some set of assumptions, [20, Theorem 2.1] establishes the existence of density for the random field solution of (1) at any point $(t, x) \in (0, T] \times \mathbb{R}^d$, and also that the density belongs to some Besov space. The theorem applies for example to the stochastic wave equation in any spatial dimensions $d \geq 1$.

The purpose of this paper is to further illustrate the range of applications of Lemma 1 with two more examples. The first one is presented in the next Sect. 2 and complements the results of [20]. In comparison with this reference, here we are able to remove the *strong ellipticity* property on the function σ, which is crucial in most of the applications of Malliavin calculus to SPDEs (see [19]), but the class of differential operators L is more restrictive. Nevertheless, Theorem 1 below applies for example to the stochastic heat equation in any spatial dimension and to the stochastic wave equation with $d \leq 3$. For the latter example, if σ, b are smooth functions and σ is bounded away from zero, existence and regularity of the density of $u(t, x)$ has been established in [16, 17].

The second example, developed in Sect. 3, refers to ambit fields driven by a class of Lévy bases (see 14). Originally introduced in [5] in the context of modeling turbulence, ambit fields are stochastic processes indexed by time and space that are becoming popular and useful for the applications in mathematical finance among others. The expression (14) has some similarities with the mild formulation of (1) (see 3) and can also be seen as an infinite dimensional extension of SDEs driven by Lévy processes. We are not aware of previous results on densities of ambit fields.

We end this introduction by quoting the definition of the Besov spaces relevant for this article as well as the existence of density criterion by [11].

The spaces $B^s_{1,\infty}$, $s > 0$, can be defined as follows. Let $f : \mathbb{R}^d \to \mathbb{R}$. For $x, h \in \mathbb{R}^d$ set $(\Delta^1_h f)(x) = f(x + h) - f(x)$. Then, for any $n \in \mathbb{N}$, $n \geq 2$, let

$$(\Delta^n_h f)(x) = \left(\Delta^1_h (\Delta^{n-1}_h f)\right)(x) = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} f(x + jh).$$

For any $0 < s < n$, we define the norm

$$\|f\|_{B^s_{1,\infty}} = \|f\|_{L^1} + \sup_{|h| \leq 1} |h|^{-s} \|\Delta^n_h f\|_{L^1}.$$

It can be proved that for two distinct $n, n' > s$ the norms obtained using n or n' are equivalent. Then we define $B^s_{1,\infty}$ to be the set of L^1-functions with $\|f\|_{B^s_{1,\infty}} < \infty$. We refer the reader to [22] for more details.

In the following, we denote by $\mathscr{C}^\alpha_b$ the set of bounded Hölder continuous functions of degree α. The next Lemma establishes the criterion on existence of densities that we will apply in our examples.

Lemma 1 *Let κ be a finite nonnegative measure. Assume that there exist $0 < \alpha \le a < 1$, $n \in \mathbb{N}$ and a constant C_n such that for all $\phi \in \mathscr{C}_b^\alpha$, and all $h \in \mathbb{R}$ with $|h| \le 1$,*

$$\left| \int_{\mathbb{R}} \Delta_h^n \phi(y) \kappa(dy) \right| \le C_n \|\phi\|_{\mathscr{C}_b^\alpha} |h|^a. \tag{2}$$

Then κ has a density with respect to the Lebesgue measure, and this density belongs to the Besov space $B_{1,\infty}^{a-\alpha}(\mathbb{R})$.

2 Nonelliptic Diffusion Coefficients

In this section we deal with SPDEs without the classical ellipticity assumption on the coefficient σ, i.e. $\inf_{x\in\mathbb{R}^d} |\sigma(x)| \ge c > 0$. In the different context of SDEs driven by a Lévy process, this situation was considered in [10, Theorem 1.1], assuming in addition that σ is bounded. Here, we will deal with SPDEs in the setting of [9] with not necessarily bounded coefficients σ. Therefore, the results will apply in particular to Anderson's type SPDEs ($\sigma(x) = \lambda x$, $\lambda \ne 0$).

We consider the class of SPDEs defined by (1), with constant initial conditions, where L denotes a linear differential operator, and $\sigma, b{:}\mathbb{R} \to \mathbb{R}$. In the definition above, F is a Gaussian noise, white in time with some spatial correlation.

Consider the space of Schwartz functions on $\mathbb{R}^d$, denoted by $\mathscr{S}(\mathbb{R}^d)$, endowed with the following inner product

$$\langle \phi, \psi \rangle_{\mathscr{H}} := \int_{\mathbb{R}^d} dy \int_{\mathbb{R}^d} \Gamma(dx)\phi(y)\psi(y-x),$$

where Γ is a nonnegative and nonnegative definite tempered measure. Using the Fourier transform we can rewrite this inner product as

$$\langle \phi, \psi \rangle_{\mathscr{H}} = \int_{\mathbb{R}^d} \mu(d\xi) \mathscr{F}\phi(\xi) \overline{\mathscr{F}\psi(\xi)},$$

where μ is a nonnegative definite tempered measure with $\mathscr{F}\mu = \Gamma$. Let $\mathscr{H} := \overline{(\mathscr{S}, \langle\cdot,\cdot\rangle_{\mathscr{H}})}^{\langle\cdot,\cdot\rangle_{\mathscr{H}}}$, and $\mathscr{H}_T := L^2([0,T];\mathscr{H})$. It can be proved that F is an isonormal Wiener process on $\mathscr{H}_T$.

Let Λ denote the fundamental solution to $Lu = 0$ and assume that Λ is either a function or a non-negative measure of the form $\Lambda(t, dy)dt$ such that

$$\sup_{t\in[0,T]} \Lambda(t, \mathbb{R}^d) \le C_T < \infty.$$

We consider

$$u(t,x) = \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y)\sigma(u(s,y))M(ds,dy) + \int_0^t \int_{\mathbb{R}^d} \Lambda(t-s, x-y) b(u(s,y)) dy dy, \tag{3}$$

as the integral formulation of (1), where M is the martingale measure generated by F. In order for the stochastic integral in the previous equation to be well-defined, we need to assume that

$$\int_0^T ds \int_{\mathbb{R}^d} \mu(d\xi)|\mathscr{F}\Lambda(s)(\xi)|^2 < +\infty. \tag{4}$$

According to [9, Theorem 13] (see also [23]), equation (3) has a unique random field solution $\{u(t,x);\, (t,x) \in [0,T] \times \mathbb{R}^d\}$ which has a spatially stationary law (this is a consequence of the S-property in [9]), and for all $p \geq 2$

$$\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} \mathbb{E}\big[|u(t,x)|^p\big] < \infty.$$

We will prove the following result on the existence of a density.

Theorem 1 *Fix $T > 0$. Assume that for all $t \in [0,T]$, $\Lambda(t)$ is a function or a non-negative distribution such that (4) holds and $\sup_{t\in[0,T]} \Lambda(t, \mathbb{R}^d) < \infty$. Assume furthermore that σ and b are Lipschitz continuous functions. Moreover, we assume that*

$$ct^{\gamma} \leq \int_0^t ds \int_{\mathbb{R}^d} \mu(d\xi)|\mathscr{F}\Lambda(s)(\xi)|^2 \leq Ct^{\gamma_1},$$
$$\int_0^t ds |\mathscr{F}\Lambda(s)(0)|^2 \leq Ct^{\gamma_2},$$

for some $\gamma, \gamma_1, \gamma_2 > 0$ and positive constants c and C. Suppose also that there exists $\delta > 0$ such that

$$\mathbb{E}\big[|u(t,0) - u(s,0)|^2\big] \leq C|t-s|^{\delta}, \tag{5}$$

for any $s, t \in [0,T]$ and some constant $C > 0$, and that

$$\bar{\gamma} := \frac{\min\{\gamma_1, \gamma_2\} + \delta}{\gamma} > 1.$$

Fix $(t,x) \in (0,T] \times \mathbb{R}^d$. Then, the probability law of $u(t,x)$ has a density f on the set $\{y \in \mathbb{R};\, \sigma(y) \neq 0\}$. In addition, there exists $n \geq 1$ such that the function $y \mapsto |\sigma(y)|^n f(y)$ belongs to the Besov space $B^{\beta}_{1,\infty}$, with $\beta \in (0, \bar{\gamma} - 1)$.

Proof The existence, uniqueness and stationarity of the solution u is guaranteed by [9, Theorem 13]. We will apply Lemma 1 to the law of $u(t,x)$ at $x=0$. Since the solution u is stationary in space, this is enough for our purposes. Consider the measure

$$\kappa(dy) = |\sigma(y)|^n \left(P \circ u(t,0)^{-1}\right)(dy).$$

We define the following approximation of $u(t,0)$. Let for $0<\varepsilon<t$

$$u^\varepsilon(t,0) = U^\varepsilon(t,0) + \sigma(u(t-\varepsilon,0)) \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} \Lambda(t-s,-y) M(ds,dy), \tag{6}$$

where

$$\begin{aligned} U^\varepsilon(t,0) = &\int_0^{t-\varepsilon} \int_{\mathbb{R}^d} \Lambda(t-s,-y)\sigma(u(s,y)) M(ds,dy) \\ &+ \int_0^{t-\varepsilon} \int_{\mathbb{R}^d} \Lambda(t-s,-y) b(u(s,y)) dyds \\ &+ b(u(t-\varepsilon,0)) \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} \Lambda(t-s,-y) dyds. \end{aligned}$$

Applying the triangular inequality, we have

$$\begin{aligned} \left| \int_{\mathbb{R}} \Delta_h^n \phi(y) \kappa(dy) \right| &= \left| \mathbb{E}\left[|\sigma(u(t,0))|^n \Delta_h^n \phi(u(t,0)) \right] \right| \\ &\le \left| \mathbb{E}\left[(|\sigma(u(t,0))|^n - |\sigma(u(t-\varepsilon,0))|^n) \Delta_h^n \phi(u(t,0)) \right] \right| \\ &\quad + \left| \mathbb{E}\left[|\sigma(u(t-\varepsilon,0))|^n (\Delta_h^n \phi(u(t,0)) - \Delta_h^n \phi(u^\varepsilon(t,0))) \right] \right| \\ &\quad + \left| \mathbb{E}\left[|\sigma(u(t-\varepsilon,0))|^n \Delta_h^n \phi(u^\varepsilon(t,0)) \right] \right|. \end{aligned} \tag{7}$$

Remember that $\|\Delta_h^n \phi\|_{\mathscr{C}_b^\alpha} \le C_n \|\phi\|_{\mathscr{C}_b^\alpha}$. Consequently,

$$|\Delta_h^n \phi(x)| = |\Delta_h^{n-1} \phi(x) - \Delta_h^{n-1} \phi(x+h)| \le C_{n-1} \|\phi\|_{\mathscr{C}_b^\alpha} |h|^\alpha,$$

Using this fact, the first term on the right-hand side of the inequality in (7) can be bounded as follows:

$$\begin{aligned} &\left| \mathbb{E}\left[(|\sigma(u(t,0))|^n - |\sigma(u(t-\varepsilon,0))|^n) \Delta_h^n \phi(u(t,0)) \right] \right| \\ &\quad \le C_n \|\phi\|_{\mathscr{C}_b^\alpha} |h|^\alpha \mathbb{E}\left[\left| |\sigma(u(t,0))|^n - |\sigma(u(t-\varepsilon,0))|^n \right| \right]. \end{aligned} \tag{8}$$

Apply the equality $x^n - y^n = (x-y)(x^{n-1} + x^{n-2}y + \cdots + xy^{n-2} + y^{n-1})$ along with the Lipschitz continuity of σ and Hölder's inequality, to obtain

$$\begin{aligned}
&\mathbb{E}\big[\big||\sigma(u(t,0))|^n - |\sigma(u(t-\varepsilon,0))|^n\big|\big] \\
&\le \mathbb{E}\Big[\big|\sigma(u(t,0)) - \sigma(u(t-\varepsilon,0))\big| \sum_{j=0}^{n-1} |\sigma(u(t,0))|^j |\sigma(u(t,0))|^{n-1-j}\Big] \\
&\le C\Big(\mathbb{E}\big[\big|u(t,0) - u(t-\varepsilon,0)\big|^2\big]\Big)^{1/2}\Big(\mathbb{E}\Big[\Big(\sum_{j=0}^{n-1} |\sigma(u(t,0))|^j |\sigma(u(t,0))|^{n-1-j}\Big)^2\Big]\Big)^{\frac{1}{2}} \\
&\le C_n\Big(\mathbb{E}\big[\big|u(t,0) - u(t-\varepsilon,0)\big|^2\big]\Big)^{1/2} \\
&\le C_n \varepsilon^{\delta/2},
\end{aligned} \tag{9}$$

where we have used that σ has linear growth, also that $u(t,0)$ has finite moments of any order and (5). Thus,

$$\big|\mathbb{E}\big[(|\sigma(u(t,0))|^n - |\sigma(u(t-\varepsilon,0))|^n)\Delta_h^n \phi(u(t,0))\big]\big| \le C_n \|\phi\|_{\mathscr{C}_b^\alpha} |h|^\alpha \varepsilon^{\delta/2}. \tag{10}$$

With similar arguments,

$$\begin{aligned}
&\big|\mathbb{E}\big[|\sigma(u(t-\varepsilon,0))|^n (\Delta_h^n \phi(u(t,0)) - \Delta_h^n \phi(u^\varepsilon(t,0)))\big]\big| \\
&\le C_n \|\phi\|_{\mathscr{C}_b^\alpha} \mathbb{E}\big[|\sigma(u(t-\varepsilon,0))|^n |u(t,0) - u^\varepsilon(t,0)|^\alpha\big] \\
&\le C_n \|\phi\|_{\mathscr{C}_b^\alpha} \big(\mathbb{E}\big[|u(t,0) - u^\varepsilon(t,0)|^2\big]\big)^{\alpha/2} \big(\mathbb{E}\big[|\sigma(u(t-\varepsilon,0))|^{2n/(2-\alpha)}\big]\big)^{1-\alpha/2} \\
&\le C_n \|\phi\|_{\mathscr{C}_b^\alpha} \varepsilon^{\delta\alpha/2} \big(g_1(\varepsilon) + g_2(\varepsilon)\big)^{\alpha/2},
\end{aligned} \tag{11}$$

where in the last inequality we have used the upper bound stated in [20, Lemma 2.5]. It is very easy to adapt the proof of this lemma to the context of this section. Note that the constant C_n in the previous equation does not depend on α because

$$\begin{aligned}
\big(\mathbb{E}\big[|\sigma(u(t-\varepsilon,0))|^{2n/(2-\alpha)}\big]\big)^{1-\alpha/2} &\le \big(\mathbb{E}\big[(|\sigma(u(t-\varepsilon,0))| \vee 1)^{2n}\big]\big)^{1-\alpha/2} \\
&\le \mathbb{E}\big[|\sigma(u(t-\varepsilon,0))|^{2n} \vee 1\big].
\end{aligned}$$

Now we focus on the third term on the right-hand side of the inequality in (7). Let p_ε denote the density of the zero mean Gaussian random variable $\int_{t-\varepsilon}^t \int_{\mathbb{R}^d} \Lambda(t-s,-y)M(ds,dy)$, which is independent of the σ-field $\mathscr{F}_{t-\varepsilon}$ and has variance

$$g(\varepsilon) := \int_0^\varepsilon ds \int_{\mathbb{R}^d} \mu(d\xi) |\mathscr{F}\Lambda(s)(\xi)|^2 \ge C\varepsilon^\gamma.$$

In the decomposition (6), the random variable $U^\varepsilon(t,0)$ is $\mathscr{F}_{t-\varepsilon}$-measurable. Then, by conditioning with respect to $\mathscr{F}_{t-\varepsilon}$ and using a change of variables, we obtain

$$
\begin{aligned}
&\left|\mathbb{E}\left[|\sigma(u(t-\varepsilon,0))|^n \Delta_h^n \phi(u^\varepsilon(t,0))\right]\right| \\
&= \left|\mathbb{E}\big[\mathbb{E}\big[1_{\{\sigma(u(t-\varepsilon,0))\neq 0\}} |\sigma(u(t-\varepsilon,0))|^n \Delta_h^n \phi(u^\varepsilon(t,0)) \big| \mathscr{F}_{t-\varepsilon}\big]\big]\right| \\
&= \left|\mathbb{E}\left[1_{\{\sigma(u(t-\varepsilon,0))\neq 0\}} \int_{\mathbb{R}} |\sigma(u(t-\varepsilon,0))|^n \Delta_h^n \phi(U_t^\varepsilon + \sigma(u(t-\varepsilon,0))y) p_\varepsilon(y) dy\right]\right| \\
&= \Bigg|\mathbb{E}\Bigg[1_{\{\sigma(u(t-\varepsilon,0))\neq 0\}} \int_{\mathbb{R}} |\sigma(u(t-\varepsilon,0))|^n \\
&\qquad \times \phi(U_t^\varepsilon + \sigma(u(t-\varepsilon,0))y) \Delta^n_{-\sigma(u(t-\varepsilon,0))^{-1}h} p_\varepsilon(y) dy\Bigg]\Bigg| \\
&\leq \|\phi\|_\infty \mathbb{E}\left[1_{\{\sigma(u(t-\varepsilon,0))\neq 0\}} |\sigma(u(t-\varepsilon,0))|^n \int_{\mathbb{R}} \left|\Delta^n_{-\sigma(u(t-\varepsilon,0))^{-1}h} p_\varepsilon(y)\right| dy\right].
\end{aligned}
$$

On the set $\{\sigma(u(t-\varepsilon,0)) \neq 0\}$, the integral in the last term can be bounded as follows,

$$
\begin{aligned}
\int_{\mathbb{R}} \left|\Delta^n_{-\sigma(u(t-\varepsilon,0))^{-1}h} p_\varepsilon(y)\right| dy &\leq C_n |\sigma(u(t-\varepsilon,0))|^{-n} |h|^n \|p_\varepsilon^{(n)}\|_{L^1(\mathbb{R})} \\
&\leq C_n |\sigma(u(t-\varepsilon,0))|^{-n} |h|^n g(\varepsilon)^{-n/2},
\end{aligned}
$$

where we have used the property $\|\Delta_h^n f\|_{L^1(\mathbb{R})} \leq C_n |h|^n \|f^{(n)}\|_{L^1(\mathbb{R})}$, and also that $\|p_\varepsilon^{(n)}\|_{L_1} = (g(\varepsilon))^{-n/2} \leq C_n \varepsilon^{-n\gamma/2}$ (see e.g. [20, Lemma 2.3]).

Substituting this into the previous inequality yields

$$
\left|\mathbb{E}\big[|\sigma(u(t-\varepsilon,0))|^n \Delta_h^n \phi(u^\varepsilon(t,0))\big]\right| \leq C_n \|\phi\|_{\mathscr{C}_b^\alpha} |h|^n \varepsilon^{-n\gamma/2}, \tag{12}
$$

because $\|\phi\|_\infty \leq \|\phi\|_{\mathscr{C}_b^\alpha}$.

With (7), (10)–(12), we have

$$
\begin{aligned}
&\left|\int_{\mathbb{R}} \Delta_h^n \phi(y) \kappa(dy)\right| \\
&\leq C_n \|\phi\|_{\mathscr{C}_b^\alpha} \left(|h|^\alpha \varepsilon^{\delta/2} + \varepsilon^{\delta\alpha/2} \big(g_1(\varepsilon) + g_2(\varepsilon)\big)^{\alpha/2} + |h|^n \varepsilon^{-n\gamma/2}\right) \\
&\leq C_n \|\phi\|_{\mathscr{C}_b^\alpha} \left(|h|^\alpha \varepsilon^{\delta/2} + \varepsilon^{(\delta+\gamma_1)\alpha/2} + \varepsilon^{(\delta+\gamma_2)\alpha/2} + |h|^n \varepsilon^{-n\gamma/2}\right) \\
&\leq C_n \|\phi\|_{\mathscr{C}_b^\alpha} \left(|h|^\alpha \varepsilon^{\delta/2} + \varepsilon^{\gamma\bar{\gamma}\alpha/2} + |h|^n \varepsilon^{-n\gamma/2}\right)
\end{aligned} \tag{13}
$$

Let $\varepsilon = \frac{1}{2} t |h|^\rho$, with $\rho = 2n/(\gamma n + \gamma\bar{\gamma}\alpha)$. With this choice, the last term in (13) is equal to

$$
C_n \|\phi\|_{\mathscr{C}_b^\alpha} \left(|h|^{\alpha + \frac{n\delta}{\gamma(n+\bar{\gamma}\alpha)}} + |h|^{\frac{n\bar{\gamma}\alpha}{n+\bar{\gamma}\alpha}}\right).
$$

Since $\gamma_1 \leq \gamma$, by the definition of $\bar{\gamma}$, we obtain

$$\bar{\gamma} - 1 = \frac{\min\{\gamma_1, \gamma_2\}}{\gamma} + \frac{\delta}{\gamma} - 1 \leq \frac{\delta}{\gamma}.$$

Fix $\zeta \in (0, \bar{\gamma} - 1)$. We can choose $n \in \mathbb{N}$ sufficiently large and α sufficiently close to 1, such that

$$\alpha + \frac{n\delta}{\gamma(n + \bar{\gamma}\alpha)} > \zeta + \alpha \quad \text{and} \quad \frac{n\bar{\gamma}\alpha}{n + \bar{\gamma}\alpha} > \zeta + \alpha.$$

This finishes the proof of the theorem.

Remark 1 (i) Assume that σ is bounded from above but not necessary bounded away from zero. Following the lines of the proof of Theorem 1 we can also show the existence of a density without assuming the existence of moments of $u(t, x)$ of order higher than 2. This applies in particular to SPDEs whose fundamental solutions are general distributions as treated in [8], extending the result on absolute continuity given in [20, Theorem 2.1].

(ii) Unlike [20, Theorem 2.1], the conclusion on the space to which the density belongs is less precise. We do not know whether the order $\bar{\gamma} - 1$ is optimal.

3 Ambit Random Fields

In this section we prove the absolute continuity of the law of a random variable generated by an ambit field at a fixed point $(t, x) \in [0, T] \times \mathbb{R}^d$. The methodology we use is very much inspired by [10]. Ambit fields where introduced in [5] with the aim of studying turbulence flows, see also the survey papers [6, 15]. They are stochastic processes indexed by $(t, x) \in [0, T] \times \mathbb{R}^d$ of the form

$$X(t, x) = x_0 + \iint_{A_t(x)} g(t, s; x, y)\sigma(s, y)L(ds, dy) + \iint_{B_t(x)} h(t, s; x, y)b(s, y)dyds, \tag{14}$$

where $x_0 \in \mathbb{R}$, g, h are deterministic functions subject to some integrability and regularity conditions, σ, b are stochastic processes, $A_t(x), B_t(x) \subseteq [0, t] \times \mathbb{R}^d$ are measurable sets, which are called ambit sets. The stochastic process L is a *Lévy basis* on the Borel sets $\mathscr{B}([0, T] \times \mathbb{R}^d)$. More precisely, for any $B \in \mathscr{B}([0, T] \times \mathbb{R}^d)$ the random variable $L(B)$ has an infinitely divisible distribution; given $B_1, \ldots, B_k$ disjoint sets of $B \in \mathscr{B}([0, T] \times \mathbb{R}^d)$, the random variables $L(B_1), \ldots, L(B_k)$ are independent; and for any sequence of disjoint sets $(A_j)_{j \in \mathbb{N}} \subset \mathscr{B}([0, T] \times \mathbb{R}^d)$,

$$L(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} L(A_j), \quad \mathbb{P}\text{-almost surely}.$$

Throughout the section, we will consider the natural filtration generated by L, i.e. for all $t \in [0, T]$,

$$\mathscr{F}_t := \sigma(L(A); A \in [0, t] \times \mathbb{R}^d, \lambda(A) < \infty).$$

For deterministic integrands, the stochastic integral in (14) is defined as in [18]. In the more general setting of (14), one can use the theory developed in [7]. We refer the reader to these references for the specific required hypotheses on g and σ.

The class of Lévy bases considered in this section are described by infinite divisible distributions of pure-jump, stable-like type. More explicitly, as in [18, Proposition 2.4], we assume that for any $B \in \mathscr{B}([0, T] \times \mathbb{R}^d)$,

$$\log \mathbb{E}\big[\exp(i\xi L(B))\big] = \int_{[0,T]\times\mathbb{R}^d} \lambda(ds, dy) \int_{\mathbb{R}} \rho_{s,y}(dz)\big(\exp(i\xi z - 1 - i\xi z 1_{[-1,1]}(z))\big),$$

where λ is termed the *control measure* on the state space and $(\rho_{s,y})_{(s,y)\in[0,T]\times\mathbb{R}^d}$ is a family of Lévy measures satisfying

$$\int_{\mathbb{R}} \min\{1, z^2\}\rho_{s,y}(dz) = 1, \ \lambda - \text{a.s.}$$

Throughout this section, we will consider the following set of assumptions on $(\rho_{s,y})_{(s,y)\in[0,T]\times\mathbb{R}^d}$ and on λ.

Assumptions 1 Fix $(t, x) \in [0, T] \times \mathbb{R}^d$ and $\alpha \in (0, 2)$, and for any $a > 0$ let $\mathscr{O}_a := (-a, a)$. Then,

(i) for all $\beta \in [0, \alpha)$ there exists a nonnegative function $C_\beta \in L^1(\lambda)$ such that for all $a > 0$,

$$\int_{(\mathscr{O}_a)^c} |z|^\beta \rho_{s,y}(dz) \le C_\beta(s, y)a^{\beta-\alpha}, \ \lambda - \text{a.s.};$$

(ii) there exists a non-negative function $\bar{C} \in L^1(\lambda)$ such that for all $a > 0$,

$$\int_{\mathscr{O}_a} |z|^2 \rho_{s,y}(dz) \le \bar{C}(s, y)a^{2-\alpha}, \ \lambda - \text{a.s.};$$

(iii) there exists a nonnegative function $c \in L^1(\lambda)$ and $r > 0$ such that for all $\xi \in \mathbb{R}$ with $|\xi| > r$,

$$\int_{\mathbb{R}} \big(1 - \cos(\xi z)\big)\rho_{s,y}(dz) \ge c(s, y)|\xi|^\alpha, \ \lambda - \text{a.s.}$$

Example 1 Let

$$\rho_{s,y}(dz) = c_1(s,y)1_{\{z>0\}}z^{-\alpha-1}dz + c_{-1}(s,y)1_{\{z<0\}}|z|^{-\alpha-1}dz,$$

with $(s,y) \in [0,T] \times \mathbb{R}^d$, and assume that $c_1, c_{-1} \in L^1(\lambda)$. This corresponds to stable distributions (see [18, Lemma 3.7]). One can check that Assumptions 1 are satisfied with $C = \bar{C} = c_1 \vee c_{-1}$, and $c = c_1 \wedge c_{-1}$.

Assumptions 2 (**H1**) We assume that the deterministic functions $g, h\colon\{0 \le s < t \le T\} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and the stochastic processes $(\sigma(s,y); (s,y) \in [0,T] \times \mathbb{R}^d)$, $(b(s,y); (s,y) \in [0,T]\times\mathbb{R}^d)$ are such that the integrals on the right-hand side of (14) are well-defined (see the conditions in [18, Theorem 2.7] and [7, Theorem 4.1]). We also suppose that for any $y \in \mathbb{R}^d$, $p \in [2,\infty)$ we have $\sup_{s\in[0,T]} \mathbb{E}[|\sigma(s,y)|^p] < \infty$.
(**H2**) Let α be as in Assumptions 1. There exist $\delta_1, \delta_2 > 0$ such that for some $\gamma \in (\alpha, 2]$ and, if $\alpha \ge 1$, for all $\beta \in [1,\alpha)$, or for $\beta = 1$, if $\alpha < 1$,

$$\mathbb{E}\big[|\sigma(t,x) - \sigma(s,y)|^\gamma\big] \le C_\gamma(|t-s|^{\delta_1\gamma} + |x-y|^{\delta_2\gamma}), \tag{15}$$

$$\mathbb{E}\big[|b(t,x) - b(s,y)|^\beta\big] \le C_\beta(|t-s|^{\delta_1\beta} + |x-y|^{\delta_2\beta}), \tag{16}$$

for every $(t,x), (s,y) \in [0,T] \times \mathbb{R}^d$, and some $C_\gamma, C_\beta > 0$.
(**H3**) $|\sigma(t,x)| > 0$, ω-a.s.
(**H4**) Let α, $\bar{C}$, C_β and c as in Assumptions 1 and $0 < \varepsilon < t$. We suppose that

$$\int_{t-\varepsilon}^t \int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\bar{c}(s,y)|g(t,s,x,y)|^\alpha \lambda(ds,dy) < \infty, \tag{17}$$

$$c\varepsilon^{\gamma_0} \le \int_{t-\varepsilon}^t \int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)c(s,y)|g(t,s,x,y)|^\alpha \lambda(ds,dy) < \infty, \tag{18}$$

where in (17), $\bar{c}(s,y) = \bar{C}(s,y) \vee C_0(s,y)$, and (18) holds for some $\gamma_0 > 0$.
Moreover, there exist constants $C, \gamma_1, \gamma_2 > 0$ and $\gamma > \alpha$ such that

$$\int_{t-\varepsilon}^t \int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\tilde{C}_\beta(s,y)|g(t,s,x,y)|^\gamma |t-\varepsilon-s|^{\delta_1\gamma} \lambda(ds,dy) \le C\varepsilon^{\gamma\gamma_1}, \tag{19}$$

$$\int_{t-\varepsilon}^t \int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\tilde{C}_\beta(s,y)|g(t,s,x,y)|^\gamma |x-y|^{\delta_2\gamma} \lambda(ds,dy) \le C\varepsilon^{\gamma\gamma_2}. \tag{20}$$

We also assume that there exist constants $C, \gamma_3, \gamma_4 > 0$ such that for all $\beta \in [1,\alpha)$, if $\alpha \ge 1$, or for $\beta = 1$, if $\alpha < 1$,

$$\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s,x,y)|^{\beta}|t-\varepsilon-s|^{\delta_1\beta}dyds \le C\varepsilon^{\beta\gamma_3}, \tag{21}$$

$$\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s,x,y)|^{\beta}|x-y|^{\delta_2\beta}dyds \le C\varepsilon^{\beta\gamma_4}, \tag{22}$$

where $\tilde{C}_\beta$ is defined as in Lemma 3.

(**H5**) The set $A_t(x)$ "reaches t", i.e. there is no $\varepsilon > 0$ satisfying $A_t(x) \subseteq [0, t-\varepsilon] \times \mathbb{R}^d$.

Remark 2 (i) By the conditions in (**H4**), the stochastic integral in (14) with respect to the Lévy basis is well-defined as a random variable in $L^\beta(\Omega)$ for any $\beta \in (0,\alpha)$ (see Lemma 3).

(ii) One can easily derive some sufficient conditions for the assumptions in (**H4**). Indeed, suppose that

$$\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\tilde{C}_\beta(s,y)|g(t,s,x,y)|^{\gamma}\lambda(ds,dy) \le C\varepsilon^{\gamma\bar{\gamma}_1},$$

then (19) holds with $\gamma_1 = \bar{\gamma}_1 + \delta_1$. If in addition, $A_t(x)$ consists of points $(s,y) \in [0,t] \times \mathbb{R}^d$ such that $|x-y| \le |t-s|^\zeta$, for any $s \in [t-\varepsilon, t]$, and for some $\zeta > 0$, then (20) holds with $\gamma_2 = \bar{\gamma}_1 + \delta_2\zeta$. Similarly, one can derive sufficient conditions for (21), (22).

(iii) The assumption (**H5**) is used in the proof of Theorem 2, where the law of $X(t,x)$ is compared with that of an approximation $X^\varepsilon(t,x)$, which is infinitely divisible. This distribution is well-defined only if $A_t(x)$ is non-empty in the region $[t-\varepsilon, t] \times \mathbb{R}^d$.

(iv) Possibly, for particular examples of ambit sets $A_t(x)$, functions g, h, and stochastic processes σ, b, the Assumptions 2 can be relaxed. However, we prefer to keep this formulation.

We can now state the main theorem of this section.

Theorem 2 *We suppose that the Assumptions 1 and 2 are satisfied and that*

$$\frac{\min\{\gamma_1,\gamma_2,\gamma_3,\gamma_4\}}{\gamma_0} > \frac{1}{\alpha}. \tag{23}$$

Fix $(t,x) \in (0,T] \times \mathbb{R}^d$. Then the law of the random variable $X(t,x)$ defined by (14) is absolutely continuous with respect to the Lebesgue measure.

3.1 Two Auxiliary Results

In this subsection we derive two auxiliary lemmas. They play a similar role as those in [10, Sects. 5.1 and 5.2], but our formulation is more general.

Lemma 2 *Let $\rho = (\rho_{s,y})_{(s,y)\in[0,T]\times\mathbb{R}^d}$ be a family of Lévy measures and let λ be a control measure. Suppose that Assumption 1(ii) holds. Then for all $\gamma \in (\alpha, 2)$ and all $a \in (0,\infty)$*

$$\int_{|z|\le a} |z|^\gamma \rho_{s,y}(dz) \le C_{\gamma,\alpha}\bar{C}(s,y)a^{\gamma-\alpha},\ \lambda - a.s.,$$

where $C_{\gamma,\alpha} = 2^{-\gamma+2}\frac{2^{2-\alpha}}{2^{\gamma-\alpha}-1}$. Hence

$$\int_0^T \int_{\mathbb{R}^d} \int_{|z|\le a} |z|^\gamma \rho_{s,y}(dz)\lambda(ds,dy) \le Ca^{\gamma-\alpha}.$$

Proof The result is obtained by the following computations:

$$\begin{aligned}\int_{|z|\le a} |z|^\gamma \rho_{s,y}(dz) &= \sum_{n=0}^{\infty} \int_{\{a2^{-n-1}<|z|\le a2^{-n}\}} |z|^\gamma \rho_{s,y}(dz)\\ &\le \sum_{n=0}^{\infty} (a2^{-n-1})^{\gamma-2} \int_{\{|z|\le a2^{-n}\}} |z|^2 \rho_{s,y}(dz)\\ &\le \bar{C}(s,y)\sum_{n=0}^{\infty}(a2^{-n-1})^{\gamma-2}(a2^{-n})^{2-\alpha}\\ &\le C_{\gamma-\alpha}\bar{C}(s,y)a^{\gamma-\alpha}. \qquad \square\end{aligned}$$

The next lemma provides important bounds on the moments of the stochastic integrals. It plays the role of [10, Lemma 5.2] in the setting of this article.

Lemma 3 *Assume that L is a Lévy basis with characteristic exponent satisfying Assumptions 1 for some $\alpha \in (0,2)$. Let $H = (H(t,x))_{(t,x)\in[0,T]\times\mathbb{R}^d}$ be a predictable process. Then for all $0 < \beta < \alpha < \gamma \le 2$ and for all $0 \le s < t \le s+1$,*

$$\begin{aligned}&\mathbb{E}\Big[\Big|\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)g(t,r,x,y)H(r,y)L(dr,dy)\Big|^\beta\Big]\\ &\le C_{\alpha,\beta,\gamma}|t-s|^{\beta/\alpha-1}\\ &\quad\times\left(\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\tilde{C}_\beta(r,y)|g(t,r,x,y)|^\gamma\,\mathbb{E}\big[|H(u,y)|^\gamma\big]\lambda(dr,dy)\right)^{\beta/\gamma},\end{aligned} \tag{24}$$

where $\tilde{C}_\beta(r,y)$ is the maximum of $\bar{C}(r,y)$, and $(C_\beta + C_1)(r,y)$ (see Assumptions 1 for the definitions).

Proof There exists a Poisson random measure N such that for all $A \in \mathscr{B}(\mathbb{R}^d)$,

$$L([s,t]\times A) = \int_s^t \int_A \int_{|z|\le 1} z\tilde{N}(dr,dy,dz) + \int_s^t \int_A \int_{|z|>1} zN(dr,dy,dz)$$

(see e.g. [14, Theorem 4.6]), where $\tilde{N}$ stands for the compensated Poisson random measure $\tilde{N}(ds,dy,dz) = N(ds,dy,dz) - \rho_{s,y}(dz)\lambda(ds,dy)$. Then we can write

$$\mathbb{E}\Big[\Big|\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)g(t,r,x,y)H(r,y)L(dr,dy)\Big|^\beta\Big] \le C_\beta\big(I_{s,t}^1 + I_{s,t}^2 + I_{s,t}^3\big), \tag{25}$$

with

$$I_{s,t}^1 := \mathbb{E}\Big[\Big|\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|\le (t-s)^{1/\alpha}} zg(t,r,x,y)H(r,y)\tilde{N}(dr,dy,dz)\Big|^\beta\Big]$$
$$I_{s,t}^2 := \mathbb{E}\Big[\Big|\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{(t-s)^{1/\alpha}<|z|\le 1} zg(t,r,x,y)H(r,y)\tilde{N}(dr,dy,dz)\Big|^\beta\Big]$$
$$I_{s,t}^3 := \mathbb{E}\Big[\Big|\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|>1} zg(t,r,x,y)H(r,y)N(dr,dy,dz)\Big|^\beta\Big]$$

To give an upper bound for the first term, we apply first Burkholder's inequality, then the subadditivity of the function $x \mapsto x^{\gamma/2}$ (since the integral is actually a sum), Jensen's inequality, the isometry of Poisson random measures and Lemma 2. We obtain,

$$\begin{aligned}
I_{s,t}^1 &\le C_\beta \mathbb{E}\Big[\Big|\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|\le (t-s)^{1/\alpha}} \\
&\qquad \times |z|^2|g(t,r,x,y)|^2|H(r,y)|^2 N(dr,dy,dz)\Big|^{\beta/2}\Big] \\
&\le C_\beta \mathbb{E}\Big[\Big|\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|\le (t-s)^{1/\alpha}} \\
&\qquad \times |z|^\gamma|g(t,r,x,y)|^\gamma|H(r,y)|^\gamma N(dr,dy,dz)\Big|^{\beta/\gamma}\Big] \\
&\le C_\beta \Big(\mathbb{E}\Big[\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|\le (t-s)^{1/\alpha}} \\
&\qquad \times |z|^\gamma|g(t,r,x,y)|^\gamma|H(r,y)|^\gamma N(dr,dy,dz)\Big]\Big)^{\beta/\gamma} \\
&= C_\beta \Big(\mathbb{E}\Big[\int_s^t \int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\Big(\int_{|z|\le (t-s)^{1/\alpha}} |z|^\gamma \rho_{r,y}(dz)\Big)
\end{aligned}$$

$$
\begin{aligned}
&\qquad \times\, |g(t,r,x,y)|^{\gamma}|H(r,y)|^{\gamma}\lambda(dr,dy)\Big]\Big)^{\beta/\gamma}\\
&\le C_\beta\big(C_{\gamma,\alpha}(t-s)^{(\gamma-\alpha)/\alpha}\big)^{\beta/\gamma}\\
&\qquad\times\left(\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)\bar{C}(r,y)|g(t,r,x,y)|^{\gamma}\mathbb{E}\big[|H(u,y)|^{\gamma}\big]\lambda(dr,dy)\right)^{\beta/\gamma}.
\end{aligned}
$$

Notice that the exponent $(\gamma-\alpha)/\alpha$ is positive.

With similar arguments but applying now Assumption 1(i), the second term in (25) is bounded by

$$
\begin{aligned}
I_{s,t}^2 &\le C_\beta\mathbb{E}\bigg[\bigg|\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)\int_{(t-s)^{1/\alpha}<|z|\le 1}|z|^2\\
&\qquad\times\,|g(t,r,x,y)|^2|H(r,y)|^2N(dr,dy,dz)\bigg|^{\beta/2}\bigg]\\
&\le C_\beta\mathbb{E}\bigg[\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)\int_{(t-s)^{1/\alpha}<|z|\le 1}|z|^\beta\\
&\qquad\times\,|g(t,r,x,y)|^\beta|H(r,y)|^\beta N(dr,dy,dz)\bigg]\\
&= C_\beta\mathbb{E}\bigg[\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)\bigg(\int_{(t-s)^{1/\alpha}<|z|\le 1}|z|^\beta\rho_{r,y}(dz)\bigg)\\
&\qquad\times\,|g(t,r,x,y)|^\beta|H(r,y)|^\beta\lambda(dr,dy)\bigg]\\
&\le C_\beta(t-s)^{(\beta-\alpha)/\alpha}\mathbb{E}\bigg[\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)C_\beta(r,y)\\
&\qquad\times\,|g(t,r,x,y)|^\beta\big[|H(r,y)|^\beta\lambda(dr,dy)\bigg]\\
&\le C_{\beta,\gamma}(t-s)^{(\beta-\alpha)/\alpha}\bigg(\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)C_\beta(r,y)\\
&\qquad\times\,|g(t,r,x,y)|^\gamma\mathbb{E}\big[|H(r,y)|^\gamma\big]\lambda(dr,dy)\bigg)^{\beta/\gamma},
\end{aligned}
$$

where in the last step we have used Hölder's inequality with respect to the finite measure $C_\beta(r,y)\lambda(dr,dy)$.

Finally, we bound the third term in (25). Suppose first that $\beta\le 1$. Using the subadditivity of $x\mapsto x^\beta$ and Lemma 2(i) yields

$$
\begin{aligned}
I_{s,t}^3 &\le C_\beta\mathbb{E}\bigg[\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)\int_{|z|>1}|z|^\beta|g(t,r,x,y)|^\beta|H(r,y)|^\beta N(dr,dy,dz)\bigg]\\
&\le C_\beta\mathbb{E}\bigg[\int_s^t\int_{\mathbb{R}^d}1_{A_t(x)}(r,y)\bigg(\int_{|z|>1}|z|^\beta\rho_{r,y}(dz)\bigg)
\end{aligned}
$$

$$
\begin{aligned}
&\quad \times\ |g(t,r,x,y)|^{\beta}|H(r,y)|^{\beta}\lambda(dr,dy)\Big]\\
&\le C_\beta \mathbb{E}\Big[\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)C_\beta(r,y)|g(t,r,x,y)|^{\beta}|H(r,y)|^{\beta}\lambda(dr,dy)\Big]\\
&\le C_\beta\bigg(\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)C_\beta(r,y)|g(t,r,x,y)|^{\gamma}\mathbb{E}\big[|H(r,y)|^{\gamma}\big]\lambda(dr,dy)\bigg)^{\beta/\gamma},
\end{aligned}
$$

where in the last step we have used Hölder's inequality with respect to the finite measure $C_\beta(r,y)\lambda(dr,dy)$.

Suppose now that $\beta > 1$ (which implies that $\alpha > 1$). We apply Hölder's inequality with respect to the finite measure $C_1(r,y)\lambda(dr,dy)$ and Assumption 1(i)

$$
\begin{aligned}
I^3_{s,t} &\le 2^{\beta-1}\mathbb{E}\Big[\Big|\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|>1} zg(t,r,x,y)H(r,y)\tilde{N}(dr,dy,dz)\Big|^{\beta}\Big]\\
&\quad + 2^{\beta-1}\mathbb{E}\Big[\Big|\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\bigg(\int_{|z|>1}|z|\rho_{r,y}(dz)\bigg)\\
&\qquad \times\ g(t,r,x,y)H(r,y)\lambda(dr,dy)\Big|^{\beta}\Big]\\
&\le C_\beta\mathbb{E}\Big[\Big|\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|>1}|z|^2|g(t,r,x,y)|^2\\
&\qquad \times\ |H(r,y)|^2N(dr,dy,dz)\Big|^{\beta/2}\Big]\\
&\quad + C_\beta\mathbb{E}\Big[\Big|\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)C_1(r,y)|g(t,r,x,y)|H(r,y)|\lambda(dr,dy)\Big|^{\beta}\Big]\\
&\le C_\beta\mathbb{E}\Big[\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)\int_{|z|>1}|z|^{\beta}|g(t,r,x,y)|^{\beta}|H(r,y)|^{\beta}N(dr,dy,dz)\Big]\\
&\quad + C_\beta\bigg(\int_s^t\int_{\mathbb{R}^d} C_1(r,y)\lambda(dr,dy)\bigg)^{\beta-1}\\
&\qquad \times\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)C_1(r,y)|g(t,r,x,y)|^{\beta}\mathbb{E}\big[|H(r,y)|^{\beta}\big]\lambda(dr,dy)\\
&\le C_\beta\mathbb{E}\Big[\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)(C_1(r,y)+C_\beta(r,y))\\
&\qquad \times\ |g(t,r,x,y)|^{\beta}|H(u,y)|^{\beta}\lambda(dr,dy)\Big]\\
&\le C_\beta\bigg(\int_s^t\int_{\mathbb{R}^d} 1_{A_t(x)}(r,y)(C_1(r,y)+C_\beta(r,y))\\
&\qquad \times\ |g(t,r,x,y)|^{\gamma}\mathbb{E}\big[|H(u,y)|^{\gamma}\big]\lambda(dr,dy)\bigg)^{\beta/\gamma},
\end{aligned}
$$

where in the last step we have used Hölder's inequality with respect to the finite measure $(C_1(r,y)+C_\beta(r,y))\lambda(dr,dy)$. We are assuming $0<t-s\leq 1$, and $0<\beta<\alpha$. Hence, the estimates on the terms $I^i_{s,t}$, $i=1,2,3$ imply (24)

3.2 Existence of Density

With the help of the two lemmas in the previous subsection, we can now give the proof of Theorem 2. Fix $(t,x)\in(0,T]\times\mathbb{R}^d$ and let $0<\varepsilon<t$ to be determined later. We define an approximation of the ambit field $X(t,x)$ by

$$X^\varepsilon(t,x)=U^\varepsilon(t,x)+\sigma(t-\varepsilon,x)\int_{t-\varepsilon}^t\int_{\mathbb{R}^d}1_{A_t(x)}(s,y)g(t,s;x,y)L(ds,dy), \quad (26)$$

where

$$\begin{aligned} U^\varepsilon(t,x)=x_0&+\int_0^{t-\varepsilon}\int_{\mathbb{R}^d}1_{A_t(x)}(s,y)g(t,s;x,y)\sigma(s,y)L(ds,dy)\\ &+\int_0^{t-\varepsilon}\int_{\mathbb{R}^d}1_{B_t(x)}(s,y)h(t,s;x,y)b(s,y)dyds\\ &+b(t-\varepsilon,x)\int_0^{t-\varepsilon}\int_{\mathbb{R}^d}1_{B_t(x)}(s,y)h(t,s;x,y)dyds \end{aligned}$$

Note that $U^\varepsilon(t,x)$ is $\mathscr{F}_{t-\varepsilon}$-measurable.

The stochastic integral in (26) is well defined in the sense of [18] and is a random variable having an infinitely divisible distribution. Moreover, the real part of its characteristic exponent is given by

$$\Re\big(\log\mathbb{E}\big[\exp(i\xi X)\big]\big)=\int_{\mathbb{R}}\big(1-\cos(\xi z)\big)\rho_f(dz),$$

where

$$\rho_f(B)=\int_{[0,T]\times\mathbb{R}^d}\int_{\mathbb{R}}1_{\{zf(s,y)\in B\setminus\{0\}\}}\rho_{s,y}(dz)\lambda(ds,dy).$$

In the setting of this section, the next lemma plays a similar role as [20, Lemma 2.3]. It generalizes [10, Lemma 3.3] to the case of Lévy bases as integrators.

Lemma 4 *The Assumptions 1, along with (17) and (18) hold. Then, the random variable*

$$X:=\int_{t-\varepsilon}^t\int_{\mathbb{R}^d}1_{A_t(x)}(s,y)g(t,s,x,y)L(ds,dy)$$

has a $\mathscr{C}^\infty$-density $p_{t,x,\varepsilon}$, and for all $n\in\mathbb{N}$ there exists a finite constant $C_n>0$ such that $\|p^{(n)}_{t,x,\varepsilon}\|_{L^1(\mathbb{R})}\leq C_{n,t,x}(\varepsilon^{\gamma_0}\wedge 1)^{-n/\alpha}$.

Proof We follow the proof of [10, Lemma 3.3], which builds on the methods of [21]. First we show that for $|\xi|$ sufficiently large, and every $t \in (0, T]$,

$$c_{t,x,\varepsilon}|\xi|^\alpha \le \Re\Psi_X(\xi) \le C|\xi|^\alpha. \tag{27}$$

Indeed, let r be as in Assumption 1(iii). Then, for $|\xi| > r$, we have

$$\begin{aligned}
\Re\Psi_X(\xi) &= \int_{\mathbb{R}} \big(1-\cos(\xi z)\big)\rho_f(dz)\\
&= \int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} \lambda(ds,dy)\int_{\mathbb{R}} \big(1-\cos(\xi z 1_{A_t(x)}(s,y)g(t,s,x,y))\big)\rho_{s,y}(dz)\\
&\ge |\xi|^\alpha \int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)|g(t,s,x,y)|^\alpha c(s,y)\lambda(ds,dy)\\
&\ge c_{t,\varepsilon,x}\varepsilon^{\gamma_0}|\xi|^\alpha.
\end{aligned} \tag{28}$$

This proves the lower bound in (27) for $|\xi| > r$.

In order to prove the upper bound in (27), we set

$$a_{\xi,t,s,x,y} := |\xi| 1_{A_t(x)}(s,y)|g(t,s,x,y)|$$

and use the inequality $(1-\cos(x)) \le 2(x^2 \wedge 1)$ to obtain

$$\begin{aligned}
\Re\Psi_X(\xi) &= \int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} \lambda(ds,dy)\int_{\mathbb{R}} \big(1-\cos(z\xi 1_{A_t(x)}(s,y)g(t,s,x,y))\big)\rho_{s,y}(dz)\\
&\le 2\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} \lambda(ds,dy)\int_{\mathbb{R}} \big(|z|^2|\xi|^2 1_{A_t(x)}(s,y)|g(t,s,x,y)|^2 \wedge 1\big)\rho_{s,y}(dz)\\
&= 2\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} \lambda(ds,dy)\int_{|z|\le a^{-1}_{\xi,t,s,x,y}} |z|^2|\xi|^2 1_{A_t(x)}(s,y)|g(t,s,x,y)|^2\rho_{s,y}(dz)\\
&\quad + 2\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} \lambda(ds,dy)\int_{|z|\ge a^{-1}_{\xi,t,s,x,y}} \rho_{s,y}(dz).
\end{aligned} \tag{29}$$

Then, using Assumption 1(ii), the first integral in the right-hand side of the last equality in (29) can be bounded as follows:

$$\begin{aligned}
&\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} \lambda(ds,dz)|\xi|^2 1_{A_t(x)}(s,y)|g(t,s,x,y)|^2\left(\int_{|z|\le a^{-1}_{\xi,t,s,x,y}} |z|^2\rho_{s,y}(dz)\right)\\
&\le |\xi|^\alpha \int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)|g(t,s,x,y)|^\alpha \bar{C}(s,y)\lambda(ds,dy)\\
&\le C|\xi|^\alpha,
\end{aligned}$$

where in the last inequality, we have used (17).

Consider now the last integral in (29). By applying Assumption 1(i) with $\beta = 0$ and (17)

$$\begin{aligned}&\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d}\lambda(ds,dy)\left(\int_{|z|\geq a^{-1}_{\xi,t,s,x,y}}\rho_{s,y}(dz)\right)\\&\quad\leq |\xi|^{\alpha}\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d}1_{A_t(x)}(s,y)C_0(s,y)|g(t,s,x,y)|^{\alpha}\lambda(ds,dy)\\&\quad\leq C|\xi|^{\alpha}.\end{aligned}$$

Hence, we have established that

$$\Re\Psi_X(\xi)\leq C|\xi|^{\alpha},$$

for $|\xi|$ sufficiently large.

To complete the proof, we can follow the same arguments as in [10, Lemma 3.3] which rely on the result in [21, Proposition 2.3]. Note that the exponent γ_0 on the right-hand side of the gradient estimate accounts for the lower bound of the growth of the term in (18), which in the case of SDEs is equal to 1.

The next lemma shows that the error in the approximation $X^{\varepsilon}(t,x)$ in (26) and the ambit field $X(t,x)$ is bounded by a power of ε.

Lemma 5 *Assume that Assumptions 1 hold for some $\alpha \in (0,2)$ and that σ, b are Lipschitz continuous functions. Then, for any $\beta \in (0,\alpha)$, and $\varepsilon \in (0, t\wedge 1)$,*

$$\mathbb{E}\big[|X(t,x)-X^{\varepsilon}(t,x)|^{\beta}\big]\leq C_{\beta}\varepsilon^{\beta\left(\frac{1}{\alpha}+\bar{\gamma}\right)-1},$$

where $\bar{\gamma} := \min\{\gamma_1,\gamma_2,\gamma_3,\gamma_4\}$.

Proof Clearly,

$$\begin{aligned}&\mathbb{E}\big[|X(t,x)-X^{\varepsilon}(t,x)|^{\beta}\big]\\&\leq C_{\beta}\mathbb{E}\Bigg[\bigg|\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d}1_{A_t(x)}(s,y)g(t,s;x,y)(\sigma(s,y)-\sigma(t-\varepsilon,x))L(ds,dy)\bigg|^{\beta}\Bigg]\\&\quad+C_{\beta}\mathbb{E}\Bigg[\bigg|\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d}1_{B_t(x)}(s,y)h(t,s;x,y)(b(s,y)-b(t-\varepsilon,x))dyds\bigg|^{\beta}\Bigg].\end{aligned}$$

Fix $\gamma \in (\alpha,2]$ and apply Lemma 3 to the stochastic process $H(s,y) := \sigma(s,y) - \sigma(t-\varepsilon,x)$, where the arguments t,ε,x are fixed. We obtain

$$\mathbb{E}\Big[\Big|\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)g(t,s;x,y)(\sigma(s,y)-\sigma(t-\varepsilon,x))L(ds,dy)\Big|^{\beta}\Big]$$
$$\leq C_{\alpha,\beta,\gamma}\varepsilon^{\beta/\alpha-1}\Big(\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\tilde{C}_{\beta}(s,y)1_{A_t(x)}|g(t,s,x,y)|^{\gamma}$$
$$\times\ \mathbb{E}\big[|\sigma(s,y)-\sigma(t-\varepsilon,x)|^{\gamma}\big]\lambda(ds,dy)\Big)^{\beta/\gamma}$$

Owing to hypothesis (**H2**) this last expression is bounded (up to the constant $C_{\alpha,\beta,\gamma}\varepsilon^{\beta/\alpha-1}$) by

$$\Big(\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\tilde{C}_{\beta}(s,y)1_{A_t(x)}|g(t,s,x,y)|^{\gamma}$$
$$\times\ \big(|t-\varepsilon-s|^{\delta_1\gamma}+|x-y|^{\delta_2\gamma}\big)\lambda(ds,dy)\Big)^{\beta/\gamma}$$

The inequality (19) implies

$$\varepsilon^{\beta/\alpha-1}\Big(\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\tilde{C}_{\beta}(s,y)|g(t,s,x,y)|^{\gamma}|t-\varepsilon-s|^{\delta_1\gamma}\lambda(ds,dy)\Big)^{\beta/\gamma}$$
$$\leq C\varepsilon^{\beta(\frac{1}{\alpha}+\gamma_1)-1},$$

and (20) yields

$$\varepsilon^{\beta/\alpha-1}\Big(\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)\tilde{C}_{\beta}(s,y)|g(t,s,x,y)|^{\gamma}|x-y|^{\delta_2\gamma}\lambda(ds,dy)\Big)^{\beta/\gamma}$$
$$\leq C\varepsilon^{\beta(\frac{1}{\alpha}+\gamma_2)-1}.$$

Thus,

$$\mathbb{E}\Big[\Big|\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{A_t(x)}(s,y)g(t,s;x,y)(\sigma(s,y)-\sigma(t,x))L(ds,dy)\Big|^{\beta}\Big]$$
$$\leq C\varepsilon^{\beta(\frac{1}{\alpha}+[\gamma_1\wedge\gamma_2])-1}. \tag{30}$$

Assume that $\beta\geq 1$ (and therefore $\alpha>1$). Hölder's inequality with respect to the finite measure $h(t,s;x,y)dyds$, (**H2**), (21), (22), imply

$$\mathbb{E}\Big[\Big|\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)h(t,s;x,y)(b(s,y)-b(t-\varepsilon,x))dyds\Big|^{\beta}\Big]$$
$$\leq C_{\beta}\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s,x,y)|^{\beta}\mathbb{E}\big[|b(s,y)-b(t-\varepsilon,x)|^{\beta}\big]dyds$$

$$
\begin{aligned}
&\le C_\beta \int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s,x,y)|^\beta |t-\varepsilon-s|^{\delta_1\beta} dyds \\
&\quad + C_\beta \int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s,x,y)|^\beta |x-y|^{\delta_2\beta} dyds \\
&\le C\varepsilon^{\beta(\gamma_3\wedge\gamma_4)}.
\end{aligned}
$$

Suppose now that $\beta < 1$, we use Jensen's inequality and once more, (**H2**), (21), (22), to obtain

$$
\begin{aligned}
&\mathbb{E}\Big[\Big|\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)h(t,s;x,y)(b(s,y)-b(t-\varepsilon,x))dyds\Big|^\beta\Big] \\
&\le \Big(\mathbb{E}\Big[\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s;x,y)||b(s,y)-b(t-\varepsilon,x)|dyds\Big]\Big)^\beta \\
&= \Big(\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s;x,y)|\mathbb{E}\big[|b(s,y)-b(t-\varepsilon,x)|\big]dyds\Big)^\beta \\
&\le C\Big(\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d} 1_{B_t(x)}(s,y)|h(t,s;x,y)|\big[|t-s|^{\delta_1}|+|x-y|^{\delta_2}\big]dyds\Big)^\beta \\
&\le C\varepsilon^{\beta(\gamma_3\wedge\gamma_4)}.
\end{aligned}
$$

This finishes the proof.

We are now in a position to prove Theorem 2.

Proof (*Proof of Theorem* 2) We consider the inequality

$$
\begin{aligned}
|\mathbb{E}[|\sigma(t,x)|^n \Delta_h^n\phi(X(t,x))]| \le &\big|\mathbb{E}\big[\big(|\sigma(t,x)|^n - |\sigma(t-\varepsilon,x)|^n\big)\Delta_h^n\phi(X(t,x))\big]\big| \\
&+ \big|\mathbb{E}\big[|\sigma(t-\varepsilon,x)|^n\big(\Delta_h^n\phi(X(t,x)) - \Delta_h^n\phi(X^\varepsilon(t,x))\big)\big]\big| \\
&+ \big|\mathbb{E}\big[|\sigma(t-\varepsilon,x)|^n\Delta_h^n\phi(X^\varepsilon(t,x))\big]\big|. \qquad (31)
\end{aligned}
$$

Fix $\eta \in (0, \alpha \wedge 1)$. As in (8) we have

$$
\begin{aligned}
&\big|\mathbb{E}\big[\big(|\sigma(t,x)|^n - |\sigma(t-\varepsilon,x)|^n\big)\Delta_h^n\phi(X(t,x))\big]\big| \\
&\le C_n\|\phi\|_{\mathscr{C}_b^\eta}|h|^\eta \mathbb{E}\big[\big||\sigma(t,x)|^n - |\sigma(t-\varepsilon,x)|^n\big|\big]
\end{aligned}
$$

Now we proceed as in (9) using the finiteness of the moments of $\sigma(t,x)$ stated in Hypothesis (**H1**), and (**H2**). Then for all $\gamma \in (\alpha, 2]$ we have

$$
\begin{aligned}
&\mathbb{E}\big[\big||\sigma(t,x)|^n - |\sigma(t-\varepsilon,x)|^n\big|\big] \\
&= \mathbb{E}\Big[\big|\sigma(t,x)-\sigma(t-\varepsilon,x)\big|\sum_{j=0}^{n-1}|\sigma(t,x)|^j|\sigma(t-\varepsilon,x)|^{n-1-j}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\le C\Big(\mathbb{E}\big[\big|\sigma(t,x)-\sigma(t-\varepsilon,x)\big|^{\gamma}\big]\Big)^{1/\gamma}\\
&\quad\times\left(\mathbb{E}\left[\left(\sum_{j=0}^{n-1}|\sigma(t,x)|^{j}|\sigma(t-\varepsilon,x))|^{n-1-j}\right)^{\gamma/(\gamma-1)}\right]\right)^{1-1/\gamma}\\
&\le C_n\varepsilon^{\delta_1}.
\end{aligned}
$$

Therefore

$$
\big|\mathbb{E}\big[\big(|\sigma(t,x)|^n-|\sigma(t-\varepsilon,x)|^n\big)\Delta_h^n\phi(X(t,x))\big]\big|\le C_n\|\phi\|_{\mathscr{C}_b^\eta}|h|^\eta\varepsilon^{\delta_1}. \tag{32}
$$

Consider the inequality $\|\Delta_h^n\phi\|_{\mathscr{C}_b^\alpha}\le C_n\|\phi\|_{\mathscr{C}_b^\alpha}$, and apply Hölder's inequality with some $\beta\in(\eta,\alpha)$ to obtain

$$
\begin{aligned}
&\big|\mathbb{E}\big[|\sigma(t-\varepsilon,x)|^n(\Delta_h^n\phi(X(t,x))-\Delta_h^n\phi(X^\varepsilon(t,x)))\big]\big|\\
&\le C_n\|\phi\|_{\mathscr{C}_b^\eta}\mathbb{E}\big[|\sigma(t-\varepsilon,x)|^n|X(t,x)-X^\varepsilon(t,x)|^\eta\big]\\
&\le C_n\|\phi\|_{\mathscr{C}_b^\eta}\big(\mathbb{E}\big[|X(t,x)-X^\varepsilon(t,x)|^\beta\big]\big)^{\eta/\beta}\big(\mathbb{E}\big[|\sigma(u(t-\varepsilon,0))|^{n\beta/(\beta-\eta)}\big]\big)^{1-\eta/\beta}\\
&\le C_{n,\beta}\|\phi\|_{\mathscr{C}_b^\eta}\varepsilon^{\eta\left(\frac{1}{\alpha}+\bar\gamma\right)-\frac{\eta}{\beta}},
\end{aligned} \tag{33}
$$

where $\bar\gamma:=\min\{\gamma_1,\gamma_2,\gamma_3,\gamma_4\}$, and we have applied Lemma 5.

Conditionally to $\mathscr{F}_{t-\varepsilon}$, the random variable

$$
\int_{t-\varepsilon}^{t}\int_{\mathbb{R}^d}1_{A_t(x)}(s,y)g(t,s;x,y)L(ds,dy)
$$

has an infinitely divisible law and a $\mathscr{C}^\infty$-density $p_{t,x\varepsilon}$ for which a gradient estimate holds (see Lemma 4). Then, by a discrete integration by parts, and owing to (**H3**),

$$
\begin{aligned}
&\big|\mathbb{E}\big[|\sigma(t-\varepsilon,x)|^n\Delta_h^n\phi(X^\varepsilon(t,x))\big]\big|\\
&=\left|\mathbb{E}\left[\int_{\mathbb{R}}|\sigma(t-\varepsilon,x)|^n\Delta_h^n\phi(U_t^\varepsilon+\sigma(t-\varepsilon,x)y)p_{t,x,\varepsilon}(y)dy\right]\right|\\
&=\left|\mathbb{E}\left[\int_{\mathbb{R}}|\sigma(t-\varepsilon,x)|^n\phi(U_t^\varepsilon+\sigma(t-\varepsilon,x)y)\Delta^n_{-\sigma(t-\varepsilon,x)^{-1}h}p_{t,x,\varepsilon}(y)dy\right]\right|\\
&\le\|\phi\|_\infty\mathbb{E}\left[|\sigma(t-\varepsilon,x)|^n\int_{\mathbb{R}}\big|\Delta^n_{-\sigma(t-\varepsilon,x)^{-1}h}p_{t,x,\varepsilon}(y)\big|dy\right].
\end{aligned}
$$

From Lemma 4 it follows that

$$
\begin{aligned}
\int_{\mathbb{R}}\big|\Delta^n_{-\sigma(t-\varepsilon,x)^{-1}h}p_{t,x,\varepsilon}(y)\big|dy&\le C_n|\sigma(t-\varepsilon,x)|^{-n}|h|^n\|p^{(n)}_{t,x,\varepsilon}\|_{L^1(\mathbb{R})}\\
&\le C_n|\sigma(t-\varepsilon,x)|^{-n}|h|^n\varepsilon^{-n\gamma_0/\alpha},
\end{aligned}
$$

which yields

$$\left|\mathbb{E}\left[|\sigma(t-\varepsilon,x)|^n \Delta_h^n \phi(X^\varepsilon(t,x))\right]\right| \le C_n \|\phi\|_{\mathscr{C}_b^\eta} |h|^n \varepsilon^{-n\gamma_0/\alpha}, \tag{34}$$

because $\|\phi\|_\infty \le \|\phi\|_{\mathscr{C}_b^\eta}$.

The estimates (31)–(34) imply

$$|\mathbb{E}[|\sigma(t,x)|^n \Delta_h^n \phi(X(t,x))]| \le C_{n,\beta} \|\phi\|_{\mathscr{C}_b^\eta} \left(|h|^\eta \varepsilon^{\delta_1} + \varepsilon^{\eta\left(\frac{1}{\alpha}+\bar{\gamma}\right)-\frac{\eta}{\beta}} + |h|^n \varepsilon^{-n\gamma_0/\alpha}\right)$$

Set $\varepsilon = \frac{t}{2}|h|^\rho$, with $|h| \le 1$ and

$$\rho \in \left(\frac{\alpha\beta}{\beta+\alpha\beta\bar{\gamma}-\alpha}, \frac{\alpha(n-\eta)}{n\gamma_0}\right).$$

Notice that, since $\lim_{n\to\infty} \frac{\alpha(n-\eta)}{n\gamma_0} = \frac{\alpha}{\gamma_0}$, for β close to α and γ_0 as in the hypothesis, this interval is nonempty. Then, easy computations show that with the choices of ε and ρ, one has

$$|h|^\eta \varepsilon^{\delta_1} + \varepsilon^{\eta\left(\frac{1}{\alpha}+\bar{\gamma}\right)-\frac{\eta}{\beta}} + |h|^n \varepsilon^{-n\gamma_0/\alpha} \le 3|h|^\zeta,$$

with $\zeta > \eta$. Hence, with Lemma 1 we finish the proof of the theorem.

Remark 3 (i) If σ is bounded away from zero, then one does not need to assume the existence of moments of sufficiently high order. In this case one can follow the strategy in [20].

(ii) The methodology used in this section is not restricted to pure-jump stable-like noises. One can also adapt it to the case of Gaussian space-time white noises.

Acknowledgments The second author wants to thank Andreas Basse-O'Connor for a very fruitful conversation on Lévy bases.

References

1. Bally, V., Clément, E.: Integration by parts formulas and applications to equations with jumps. Probab. Theory Relat. Fields **151**, 613–657 (2011)
2. Bally, V., Clément E.: Integration by parts formulas with respect to jump times and stochastic differential equations. In: Crisan, D. (ed.) Stochastic Analysis 2010. Springer (2011)
3. Bally, V., Caramellino, L.: Convergence and regularity of probability laws by using an interpolation method. arXiv:1409.3118v1 (2014)
4. Bally, V., Fournier, N.: Regularization properties of the 2D homogeneous Boltzmann equation without cutoff. Probab. Theory Relat. Fields **151**, 659–704 (2011)

5. Barndorff-Nielsen, O., Schmiegel, J.: Lévy-based tempo-spatial modelling; with applications to turbulence. Uspekhi Mat. NAUK **59**, 65–91 (2004)
6. Barndorff-Nielsen, O., Benth, F.E., Veraart, A.: Recent advances in ambit stochastics with a view towards tempo-spatial stochastic volatility/intermittency. Banach Center Publications (2014)
7. Chong, C., Klüppelberg, C.: Integrability conditions for space-time stochastic integrals: theory and applications. Bernoulli (2015). In press
8. Conus, C., Dalang, R.C.: The non-linear stochastic wave equation in high dimensions. Electron. J. Probab. **13**, 629–670 (2008)
9. Dalang, R.C.: Extending martingale measure stochastic integral with applications to spatially homogeneous SPDEs. Electron. J. Probab. **4**, 1–29 (1999)
10. Debussche, A., Fournier, N.: Existence of densities for stable-like driven SDEs with Hölder continuous coefficients. J. Funct. Anal. **264**(8), 1757–1778 (2013)
11. Debussche, A., Romito, M.: Existence of densities for the 3D Navier-Stokes equations driven by Gaussian noise. Probab. Theory Relat. Fields **158**(3–4), 575–596 (2014)
12. Fournier, N., Printems, J.: Absolute continuity for some one-dimensional processes. Bernoulli **16**(2), 343–360 (2010)
13. Malliavin, P.: Stochastic calculus of variation and hypoelliptic operators. In: Proceedings of International Symposium on Stochastic Differential Equations, Kyoto 1976, pp. 195–263. Wiley (1978)
14. Pedersen, J.: Lévy-Itô decomposition of an independently scattered random measure. http://www.maphysto.dk/publications/MPS-RR/2003/2.pdf (2003). Accessed 27 Nov 2014
15. Podolskij M.: Ambit fields: survey and new challenges. arXiv:1405.1531 (2014)
16. Quer-Sardanyons, L., Sanz-Solé, M.: Absolute continuity of the law of the solution to the 3-dimensional stochastic wave equation. J. Funct. Anal. **206**, 1–32 (2004)
17. Quer-Sardanyons, L., Sanz-Solé, M.: A stochastic wave equation in dimension 3: smoothness of the law. Bernoulli **10**, 165–186 (2004)
18. Rajput, B., Rosinski, J.: Spectral representation of infinitely divisible distributions. Probab. Theory Relat. Fields **82**, 451–487 (1989)
19. Sanz-Solé, M.: Malliavin Calculus with Applications to Stochastic Partial Differential Equations. EPFL Press (2005)
20. Sanz-Solé, M., Süß, A.: Absolute continuity for SPDEs with irregular fundamental solution. Electron. Commun. Probab. **20**(14), 1–11 (2015)
21. Schilling, R.L., Sztonyk, P., Wang, J.: Coupling property and gradient estimates for Lévy processes via the symbol. Bernoulli **18**, 1128–1149 (2012)
22. Triebel, H.: Theory of Function Spaces. Birkhäuser (1983)
23. Walsh, J.B.: An introduction to stochastic partial differential equations. In: Ecole d'été de Probabilites de Saint Flour XIV, 1984. Lecture Notes in Mathematics, vol. 1180. Springer (1986)

Part II
Applications

Dynamic Risk Measures and Path-Dependent Second Order PDEs

Jocelyne Bion-Nadal

Abstract We propose new notions of regular solutions and viscosity solutions for path-dependent second order partial differential equations. Making use of the martingale problem approach to path-dependent diffusion processes, we explicitly construct families of time-consistent dynamic risk measures on the set of càdlàg paths $I\!R^n$ valued endowed with the Skorokhod topology. These risk measures are shown to have regularity properties. We prove then that these time-consistent dynamic risk measures provide viscosity supersolutions and viscosity subsolutions for path-dependent semi-linear second order partial differential equations.

Keywords Path-dependent PDE · Risk measures · Martingale problems

MSC: 35D40 · 35R15 · 35K55 · 60J60 · 91B30

1 Introduction

Diffusion processes are linked with parabolic second order Partial Differential Equations via the "Feynman-Kac" formula. The field of path-dependent PDEs first started in 2010 when Peng asked in [19] whether a BSDE (Backward Stochastic Differential Equation first introduced in [17]) could be considered as a solution to a path-dependent PDE. In line with the recent literature on the topic, a solution to a path-dependent second order PDE

$$H(u, \omega, \phi(u, \omega), \partial_u \phi(u, \omega), D_x \phi(u, \omega), D_x^2 \phi(u, \omega)) = 0 \quad (1)$$

is searched as a progressive function $\phi(u, \omega)$ (i.e. a path dependent function depending at time u on all the path up to time u).

J. Bion-Nadal (✉)
CMAP (UMR CNRS 7641), Ecole Polytechnique, 91128 Palaiseau Cedex, France
e-mail: jocelyne.bion-nadal@cmap.polytechnique.fr

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_6

In contrast with the classical setting, the notion of regular solution for a path-dependent PDE (1) needs to deal with càdlàg paths. Indeed to give a meaning to the partial derivatives $D_x\phi(u,\omega)$ and $D_x^2\phi(u,\omega)$ at (u_0,ω_0), one needs to assume that $\phi(u_0,\omega)$ is defined for paths ω admitting a jump at time u_0. Peng has introduced in [20] a notions of regular and viscosity solution for path-dependent second order PDEs. In [20] a regular or a viscosity solution for a path-dependent PDE is a progressive function $\phi(t,\omega)$ defined on the space of càdlàg paths endowed with the uniform norm topology and the notion of continuity and partial derivatives are those introduced by Dupire [12]. A comparison theorem is proved in this setting [20]. The motivation comes mainly from the theory of BSDE and examples of regular solutions to path-dependent PDEs can be constructed from BSDEs [18]. The main drawback for this approach based on [12] is that the uniform norm topology on the set of càdlàg paths is not separable, hence it is not a Polish space. Recently Ekren et al. proposed a notion of viscosity solution for path-dependent PDEs in the setting of continuous paths in [13, 14]. This work was motivated by the fact that a continuous function defined on the set of continuous paths does not have a unique extension into a continuous function on the set of càdlàg paths. Therefore it is suitable that the notion of viscosity solution for functions defined only on the set of continuous paths does not require to extend the function to the set of càdlàg paths. The approach developed in [13, 14] is also based on BSDE.

In the present paper we introduce a new notion of regular and viscosity solution for path-dependent second order PDEs (Sect. 2). A solution to (1) is a progressive function ϕ defined on $I\!R_+ \times \Omega$ where Ω is the set of càdlàg paths. In contrast with [20] and many works on path-dependent problems, we consider the Skorokhod topology on the set of càdlàg paths. Thus Ω is a Polish space. This property is very important. To define the continuity and regularity properties for a progressive function, we make use of the one to one correspondence between progressive functions on $I\!R_+ \times \Omega$ and strictly progressive functions on $I\!R_+ \times \Omega \times I\!R^n$ established in [3]. A function ϕ defined on $I\!R_+ \times \Omega$ is progressive if $\phi(s,\omega) = \phi(s,\omega')$ as soon as $\omega(u) = \omega'(u)$ for all $0 \le u \le s$. A function $\overline{\phi}$ defined on $I\!R_+ \times \Omega \times I\!R^n$ is strictly progressive if $\overline{\phi}(s,\omega,x) = \overline{\phi}(s,\omega',x)$ as soon as $\omega(u) = \omega'(u)$ for all $0 \le u < s$. The one to one correspondence is given by $\overline{\phi}(s,\omega,x) = \phi(s,\omega *_s x)$ where $(\omega *_s x)(u) = \omega(u)$ for all $0 \le u < s$ and $(\omega *_s x)(u) = x$ for all $u \ge s$. The continuity and regularity properties that we want for a progressive function ϕ are derived from the usual continuity and regularity properties for $\overline{\phi}$ via the above one to one correspondence. For example, $D_x\phi(u,\omega)$ is defined as $D_x\phi(u,\omega) := D_x\overline{\phi}(u,\omega,\omega(u))$ where $D_x\overline{\phi}$ is the usual partial derivative of $\overline{\phi}$ with respect to the third variable. Notice that via the above one to one correspondence, the regularity properties for a progressive function ϕ are defined in a very natural way. This is in contrast with the most commonly used regularity definitions first introduced in [12].

The notion of viscosity solution that we introduce in the present paper is motivated by our construction of a solution to semi-linear second order path-dependent PDEs based on the martingale problem approach.

Our study for viscosity solutions of path-dependent PDEs allows then to introduce a new definition of viscosity solution for path-dependent functions defined only on the set of continuous paths. As in [13, 14], this does not require to extend the function nor the coefficient functions appearing in the path-dependent PDE to the set of càdlàg paths. However our approach is very different from the one introduced in [13, 14]. In the present paper we construct then time consistent dynamic risk measures on the set Ω of càdlàg paths, to produce solutions for path-dependent semi-linear second order PDEs.

$$\begin{cases} \partial_u v(u,\omega) + \mathscr{L}^a v(u,\omega) + f(t,\omega, D_x v(u,\omega)) = 0 \text{ on } [0,t]\times\Omega \\ v(t,\omega) = h(\omega) \end{cases} \tag{2}$$

with $\mathscr{L}^a v(u,\omega) = \frac{1}{2} Tr[a(u,\omega) D_x^2 v(u,\omega)]$.
These dynamic risk measures are constructed using probability measures solution to a path-dependent martingale problem. This approach is motivated by the Feynman Kac formula and more specifically by the link between solutions of a parabolic second order PDE and probability measures solutions to a martingale problem. The martingale problem has been first introduced and studied by Stroock and Varadhan [10, 11] in the case of continuous diffusion processes. The martingale problem is linked to stochastic differential equations. However the martingale problem formulation is intrinsic and is very well suitable to construct risk measures. In [22] the martingale problem has been extended and studied to the case of jump diffusions. In [3], the study of the martingale problem is extended to the path-dependent case which means that the functions a and b (and also the jump measure) are no more defined on $I\!R_+ \times I\!R^n$ but on $I\!R_+ \times \Omega$. The question of existence and uniqueness of a solution to a path-dependent martingale problem is addressed in [3] in a general setting of diffusions with a path-dependent jump term. In the case where there is no jump term and under Lipschitz conditions on the coefficients, the existence and uniqueness of a solution has been already established in [8] from the stochastic differential equation point of view.

In Sect. 3, we recall some results from [3] on the martingale problem for path-dependent diffusion processes and study the support of a probability measure solution to the path-dependent martingale problem for $\mathscr{L}^{a,b}$.

The theory of dynamic risk measures on a filtered probability space has been developped in recent years. In the case of a Brownian filtration, dynamic risk measures coincide with g-expectations introduced by Peng [21]. An important property for dynamic risk measures is time consistency. The time consistency property for dynamic risk measures is the analogue of the Dynamic Programming Principle. For sublinear dynamic risk measures time consistency has been characterized by Delbaen [9]. For general convex dynamic risk measures two different characterizations of time consistency have been given. One by Cheridito et al. [7], the other by Bion-Nadal [5]. This last characterization of time consistency is very useful in order to construct time consistent dynamic risk measures.

Following [5], one can construct a time consistent dynamic risk measure as soon as one has a stable set of equivalent probability measures $\mathscr{Q}$ and a penalty defined on $\mathscr{Q}$ satisfying some conditions. In Sect. 4, we construct a stable set of probability measures on the set Ω of càdlàg paths. In the whole paper a is a given bounded progressively continuous function defined on $I\!R_+ \times \Omega$ such that $a(s,\omega)$ is invertible for all (s,ω). For all $r \geq 0$ and $\omega \in \Omega$ the set of probability measures $\mathcal{Q}_{r,\omega}$ is a stable set generated by probability measures $Q^{a,b}_{r,\omega}$ solution to the martingale problem for $\mathcal{L}^{a,b}$ starting from ω at time r. The functions b are assumed to satisfy some uniform *BMO* condition. In Sect. 5 we construct penalties on the stable set $\mathcal{Q}_{r,\omega}$ from a path-dependent function g. Some growth conditions are assumed on the function g to ensure integrability properties for the penalties. With such a stable set and penalties, we construct in Sect. 6 time consistent convex dynamic risk measures. More precisely for all r and ω we construct a time consistent convex dynamic risk measure $\rho^{r,\omega}_{s,t}$ on the filtered probability space $(\Omega, (\mathcal{B}_t), Q^a_{r,\omega})$ where $Q^a_{r,\omega}$ means $Q^{a,0}_{r,\omega}$ and $(\mathcal{B}_t)$ is the canonical filtration.

We prove furthermore in Sect. 7 that these time consistent dynamic risk measures satisfy the following Feller property: Let $\mathcal{C}_t$ be the set of $\mathcal{B}_t$ measurable functions h defined on Ω which can be written as $h(\omega) = k(\omega, \omega(t))$ for some continuous function k on $\Omega \times I\!R^n$ such that $k(\omega, x) = k(\omega', x)$ if $\omega(u) = \omega'(u)$ for all $u < t$. Then for all h in $\mathcal{C}_t$, there is a progressively lower semicontinuous function $R(h)$ on $[0,t] \times \Omega$ such that $R(h)(t,\omega) = h(\omega)$,

$$\rho^{r,\omega_0}_{r,t}(h) = R(h)(r,\omega_0)\ \forall 0 \leq r \leq t$$

Furthermore, for all $0 \leq r \leq s \leq t$,

$$\rho^{r,\omega_0}_{s,t}(h)(\omega') = R(h)(s,\omega')\ Q^a_{r,\omega_0}\ a.s.$$

We prove furthermore in Sect. 8 that the lower semicontinuous function $R(h)$ is a viscosity supersolution for the path-dependent semi linear second order partial differential equation (2). The function $f : I\!R^+ \times I\!R^n \times I\!R^d \to I\!R$ appearing in Eq. (2) is linked to the choice of the penalty of the risk measure. It is convex in the last variable.

We prove also that the upper semi-continuous envelope of $R(h)$ is a viscosity subsolution for (2).

When the above function h is defined only on the set of continuous paths, it is the same for the function $R(h)$. We prove then that $R(h)$ provides a viscosity supersolution and a viscosity subsolution for (2) on the set of continuous paths.

2 Solution of Path-dependent PDEs

In this section we introduce new notions for regular and viscosity solutions for second order path-dependent PDEs on the set of càdlàg paths. In contrast with [20] and all the papers using the notions of continuity and derivative introduced by Dupire [12], we work with the Skorohod topology on the set of càdlàg paths. A solution to a path-dependent PDE (1) is a progressive function $\phi(t,\omega)$ where t belongs to $I\!R_+$ and ω belongs to the set of càdlàg paths.

2.1 Topology and Regularity Properties

In the whole paper Ω denotes the set of càdlàg paths with the Skorohod topology. The set Ω is then a Polish space (i.e. is metrizable and separable). Polish spaces have nice properties which are very important in the construction of solutions for path-dependent PDEs. Among them are the existence of regular conditional probability distributions, the equivalence between relative compactness and tightness for a set of probability measures, to name a few.
To define the continuity and regularity properties for progressive functions, we use the one to one correspondence between progressive functions on $\mathbb{R}_+ \times \Omega$ and strictly progressive functions on $\mathbb{R}_+ \times \Omega \times \mathbb{R}^n$ that we have established in [3].
A function ϕ defined on $\mathbb{R}_+ \times \Omega$ is progressive if $\phi(s, \omega) = \phi(s, \omega')$ as soon as $\omega(u) = \omega'(u)$ for all $0 \le u \le s$.
A function $\overline{\phi}$ defined on $\mathbb{R}_+ \times \Omega \times \mathbb{R}^n$ is strictly progressive if $\overline{\phi}(s, \omega, x) = \overline{\phi}(s, \omega', x)$ as soon as $\omega(u) = \omega'(u)$ for all $0 \le u < s$.

The one to one correspondence $\phi \to \overline{\phi}$ is given by $\overline{\phi}(s, \omega, x) = \phi(s, \omega *_s x)$ where

$$\omega *_s x(u) = \omega(u) \; \forall\, 0 \le u < s \; \text{ and } \omega *_s x(u) = x \; \forall\, u \ge s. \tag{3}$$

Notice that $\phi(s, \omega) = \overline{\phi}(s, \omega, \omega(s))$. Accordingly a progressive function ϕ (in 2 variables (s, ω)) is said to be progressively continuous if the associated function $\overline{\phi}$ (in 3 variables (s, ω, x)) is continuous on $\mathbb{R}_+ \times \Omega \times \mathbb{R}^n$.

2.2 Regular Solution

Making use of the one to one corrrespondence between progressive functions on $\mathbb{R}_+ \times \Omega$ and strictly progressive functions on $\mathbb{R}_+ \times \Omega \times \mathbb{R}^n$, we can then give the following definition for a solution to a general path-dependent PDE.

Definition 1 Let v be a progressive function on $\mathbb{R}_+ \times \Omega$ where Ω is the set of càdlàg paths with the Skorokhod topology. v is a regular solution to the following path-dependent second order PDE

$$H(u, \omega, v(u, \omega), \partial_u v(u, \omega), D_x v(u, \omega), D_x^2 v(u, \omega)) = 0 \tag{4}$$

if the function $\overline{v}$ belongs to $\mathscr{C}^{1,0,2}(\mathbb{R}_+ \times \Omega \times \mathbb{R}^n)$ and if the usual partial derivatives of $\overline{v}$ satisfy the equation

$$H(u, \omega *_u x, \overline{v}(u, \omega, x), \partial_u \overline{v}(u, \omega, x), D_x \overline{v}(u, \omega, x), D_x^2 \overline{v}(u, \omega, x) = 0 \tag{5}$$

with $\overline{v}(u, \omega, x) = v(u, \omega *_u x)$

$(\omega *_u x)(s) = \omega(s)\ \forall s < u$, and $(\omega *_u x)(s) = x\ \forall s \geq u$. The partial derivatives of $\overline{v}$ are the usual ones, the continuity notion for $\overline{v}$ is the usual one.

2.3 Viscosity Solutions on the Set of Càdlàg Paths

The following definitions are motivated by the construction of viscosity solutions for path-dependent PDEs that we develop in the following sections. Our construction of solutions is based on the martingale problem approach for path-dependent diffusions. The support of every probability measure $Q^{a,b}_{r,\omega_0}$ solution to the martingale problem for $\mathscr{L}^{a,b}$ starting from ω_0 at time r is contained in the set of paths which coincide with ω_0 up to time r. This is a motivation for the following weak notion of continuity and also for the weak notion of local minimizer (or local maximizer) that we introduce in the definition of viscosity solution.

Definition 2 A progressively measurable function v defined on $I\!R_+ \times \Omega$ is continuous in viscosity sense at (r, ω_0) if

$$v(r,\omega_0) = \lim_{\varepsilon \to 0} \{v(s,\omega),\ \ (s,\omega) \in D_\varepsilon(r,\omega_0)\} \tag{6}$$

where

$$\begin{aligned} D_\varepsilon(r,\omega_0) = \{(s,\omega),\ \ r \leq s < r+\varepsilon,\ \ \omega(u) = \omega_0(u),\ \ \forall 0 \leq u \leq r \\ \omega(u) = \omega(s)\ \forall u \geq s,\ \text{ and } \sup_{r \leq u \leq s} ||\omega(u) - \omega_0(r)|| < \varepsilon\} \end{aligned} \tag{7}$$

v is lower (resp. upper) semi continuous in viscosity sense if Eq. (6) is satisfied replacing lim by lim inf (resp. lim sup).

Definition 3 Let v be a progressively measurable function on $(I\!R_+ \times \Omega, (\mathscr{B}_t))$ where Ω is the set of càdlàg paths with the Skorokhod topology and $(\mathscr{B}_t)$ the canonical filtration.

1. v is a viscosity supersolution of (4) if v is lower semi-continuous in viscosity sense, and if for all $(t_0, \omega_0) \in I\!R_+ \times \Omega$,
   - v is bounded from below on $D_\varepsilon(t_0, \omega_0)$ for some $\varepsilon > 0$.
   - for all strictly progressive function $\overline{\phi} \in \mathscr{C}_b^{1,0,2}(I\!R_+ \times \Omega \times I\!R^n)$ such that $v(t_0, \omega_0) = \phi(t_0, \omega_0)$, and (t_0, ω_0) is a minimizer of $v - \phi$ on $D_\varepsilon(t_0, \omega_0)$ for some $\varepsilon > 0$,

   $$H(u, \omega *_u x, \phi(u,\omega,x), \partial_u \phi(u,\omega,x), D_x \phi(u,\omega,x), D_x^2 \phi(u,\omega,x) \geq 0 \tag{8}$$

   at point $(t_0, \omega_0, \omega_0(t_0))$.

2. v is a viscosity subsolution of (4) if v is upper semi-continuous in viscosity sense, and for all (t_0, ω_0),

   - v is bounded from above on $D_\varepsilon(t_0, \omega_0)$ for some $\varepsilon > 0$
   - for all strictly progressive function $\overline{\phi} \in \mathscr{C}_b^{1,0,2}(I\!R_+ \times \Omega \times I\!R^n)$ such that $v(t_0, \omega_0) = \phi(t_0, \omega_0)$, and (t_0, ω_0) is a maximizer of $v - \phi$ on $D_\varepsilon(t_0, \omega_0)$ for some $\varepsilon > 0$,

   $$H(u, \omega *_u x, \phi(u, \omega, x), \partial_u \phi(u, \omega, x), D_x \phi(u, \omega, x), D_x^2 \phi(u, \omega, x) \leq 0 \quad (9)$$

   at point $(t_0, \omega_0, \omega_0(t_0))$.

3. v is a viscosity solution if v is both a viscosity supersolution and a viscosity subsolution.

2.4 Viscosity Solution on the Set of Continuous Paths

Recently Ekren et al. [13, 14] introduced a notion of viscosity solution of a path-dependent second order PDE for a function v defined on the set of continuous paths. One motivation for this was to define the notion of viscosity solution without extending the function v to the set of càdlàg paths.
We can notice that within our setting we can also define a notion of viscosity solution for a function v defined only on the set of continuous paths without extending v. We give the following definition which is very different from that of [13, 14] and much simpler.

Definition 4 Let v be a progressively measurable function defined on $I\!R_+ \times \mathscr{C}(I\!R_+, I\!R^n)$ with the usual uniform norm topology. v is a viscosity supersolution of (4) if v is lower semi-continuous in viscosity sense and for all $(t_0, \omega_0) \in I\!R_+ \times \mathscr{C}(I\!R_+, I\!R^n)$

- v is bounded from below on $\tilde{D}_\varepsilon(t_0, \omega_0)$ for some $\varepsilon > 0$,
- for all function strictly progressive $\overline{\phi} \in \mathscr{C}_b^{1,0,2}(I\!R_+ \times \Omega \times I\!R^n)$ such that $v(t_0, \omega_0) = \phi(t_0, \omega_0)$, and (t_0, ω_0) is a minimizer of $v - \phi$ on $\tilde{D}_\varepsilon(t_0, \omega_0)$ for some $\varepsilon > 0$,

$$H(u, \omega *_u x, \overline{\phi}(u, \omega, x), \partial_u \overline{\phi}(u, \omega, x), D_x \overline{\phi}(u, \omega, x), D_x^2 \overline{\phi}(u, \omega, x) \geq 0 \quad (10)$$

at point $(t_0, \omega_0, \omega_0(t_0))$.
Here Ω is the set of càdlàg paths with the Skorokhod topology, $\tilde{D}_\varepsilon$ is the intersection of D_ε with the set of continuous paths, and $\phi(u, \omega) = \overline{\phi}(u, \omega, \omega(u))$.

The lower semi-continuity property in viscosity sense at (t_0, ω_0) in Definition 4 means

$$v(t_0, \omega_0) = \lim_{\varepsilon \to 0} \{v(s, \omega), \ (s, \omega) \in \tilde{D}_\varepsilon(t_0, \omega_0)\}$$

We have a similar definition for a viscosity subsolution. Notice that the continuity along the sets $\tilde{D}_\varepsilon$ is also considered in [13, 14]. However the notion of viscosity solution introduced in [13, 14] is fundamentally different from ours.
We will now construct time consistent dynamic risk measures making use of probability measures solution to a path-dependent martingale problem. We will then prove that this leads to viscosity solutions to path-dependent PDEs (2).

3 Path-dependent Martingale Problem

In the classical setting, the Feynman Kac formula establishes a link between a solution of a parabolic second order PDE and probability measures solutions to a martingale problem. Assume that v is a solution of the PDE $\partial_u v(t,x) + \mathscr{L}^{a,b} v(t,x) = 0$, $v(T,.) = h$ with

$$\mathscr{L}^{a,b} v(t,x) = \frac{1}{2} Tr(a(t,x)) D_x^2(v)(t,x) + b(t,x)^* D_x v(t,x)$$

From the Feynman Kac formula, the value $v(t,x)$ can be expressed from the probability measure $Q_{t,x}^{a,b}$ solution to the martingale problem associated to the operator $\mathscr{L}^{a,b}$ starting from x at time t. $v(t,x) = E_{Q_{t,x}^{a,b}}(h(X_T))$, where (X_u) is the canonical process.
One natural way to construct soliutions for path-dependent parabolic second order partial differential equations is thus to start with probability measures solution to the path-dependent martingale problem associated to the operator $\mathscr{L}^{a,b}$ for path-dependent coefficients a and b. Let Ω be the set of càdlàg paths and $(\mathscr{B}_t)$ be the canonical filtration. Let a and b be progressively measurable functions on $\mathbb{R}_+ \times \Omega$ (a takes values in non negative invertible matrices and b in $\mathbb{R}^n$). Let $\mathscr{L}^{a,b}$ be the operator defined on $\mathscr{C}_b^2(\mathbb{R}^n)$ by

$$\mathscr{L}^{a,b}(t,\omega) = \frac{1}{2} \sum_1^n a_{ij}(t,\omega) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_1^n b_i(t,\omega) \frac{\partial}{\partial x_i} \tag{11}$$

Definition 5 Let $r \geq 0$, $\omega_0 \in \Omega$. A probability measure Q defined on $(\Omega, (\mathscr{B}_t))$ is a solution to the path-dependent martingale problem for $\mathscr{L}^{a,b}$ starting from ω_0 at time r if

$$Q(\{\omega \in \Omega \,|\omega(u) = \omega_0(u) \,\forall 0 \leq u \leq r\}) = 1$$

and if for all $f \in \mathscr{C}_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^n)$, and all t, $(Z_{r,t}^{a,b})_{r \leq t}$ given by

$$Z_{r,t}^{a,b} = f(t, X_t(\omega)) - f(r, X_r(\omega)) - \int_r^t (\frac{\partial}{\partial u} + L^{a,b}(u,\omega))(f)(u, X_u(\omega))du \tag{12}$$

is a $(Q, (\mathscr{B}_t))$ martingale.

In [3] we have studied the more general martingale problem associated with path-dependent diffusions with jumps. We have shown that the good setting to prove that the martingale problem is well posed is to deal with diffusions operators whose coefficients a and b are progressively continuous.

Recall the following result from [3].

Theorem 1 *1. Let a be a progressively continuous bounded function defined on $\mathbb{R}_+ \times \Omega$ with values in the set of non negative matrices. Assume that $a(s, \omega)$ is invertible for all (s, ω). Let b be a progressively measurable bounded function defined on $\mathbb{R}_+ \times \Omega$ with values in $\mathbb{R}^n$. For all (r, ω_0), the martingale problem for $\mathscr{L}^{a,ab}$ starting from ω_0 at time r is well posed i.e. admits a unique solution $Q^{a,ab}_{r,\omega_0}$ on the set of càdlàg paths.*

2. Assume furthermore that b is progressively continuous bounded. Consider the set of probability measures $\mathscr{M}_1(\Omega)$ equipped with the weak topology. Then the map

$$(r, \omega, x) \in \mathbb{R}_+ \times \Omega \times \mathbb{R}^n \rightarrow Q^{a,ab}_{r,\omega *_r x} \in \mathscr{M}_1(\Omega)$$

is continuous on $\{(r, \omega, x) \mid \omega = \omega *_r x\}$.

3.1 The Role of Continuous Paths

In all the following Q^a_{r,ω_0} means $Q^{a,0}_{r,\omega_0}$

We start with a result which proves that the probability measure $Q^{a,ab}_{r,\omega_0}$ is supported by paths which are continuous after time r.

Proposition 1 *Every probability measure $Q^{a,ab}_{r,\omega_0}$ solution to the martingale problem for $\mathscr{L}^{a,ab}$ starting from ω_0 at time r is supported by paths which are continuous after time r, i.e. continuous on $[r, \infty[$.*

More precisely

$$Q^{a,ab}_{r,\omega_0}(\{\omega,\ \omega(u) = \omega_0(u)\ \forall u \leq r,\ \textit{and}\ \omega_{|[r,\infty[} \in \mathscr{C}([r, \infty[, \mathbb{R}^n)\} = 1$$

Proof The probability measure $Q^{a,ab}_{r,\omega_0}$ is equivalent with Q^a_{r,ω_0}. Thus we can assume that $b = 0$.

The function a is progressively continuous. This means that the function $\overline{a}$ is continuous. Let a_n be the $\frac{1}{n}$ delayed function defined as $\overline{a}_n(u, \omega, x) = \overline{a}(u - \frac{1}{n}, \omega, x)$ for all $u \geq r + \frac{1}{n}$ and $\overline{a}_n(u, \omega, x) = \overline{a}(r, \omega, x)$ for all $0 \leq u \leq r + \frac{1}{n}$. The function a_n is also progressively continuous. Given n, let t^n_k be an increasing sequence such that $t^n_0 = r$ and $|t^n_{k+2} - t^n_k| < \frac{1}{n}$.

On a Polish space for every subsigma algebra of the Borel sigma algebra there exists a regular conditional probability distribution. It follows from [22] and the uniqueness of the solution for $\mathscr{L}^{a_n,0}$ starting from ω at time t^n_k, that for all t^n_k, $Q^{a_n}_{t^n_k,\omega}(\xi) = E_{Q^{a_n}_{r,\omega_0}}(\xi|\mathscr{B}_{t^n_k})(\omega)$ for $Q^{a_n}_{r,\omega_0}$ almost all ω. Let $a_{n,\omega}(u, x) = \overline{a}_n(u, \omega, x)$

Let $A_k = \{\omega', \ \omega'_{[t^n_k, t^n_{k+2}[} \in \mathscr{C}([t^n_k, t^n_{k+2}[)\}$. Given ω the function $a_{n,\omega}$ is not path-dependent. It follows then from [11] that $Q^{a_{n,\omega}}_{t^n_k,\omega}$ is supported by paths continuous on $[t^n_k, \infty[$. We remark that $\overline{a}_n(u, \omega', x) = a_{n,\omega}(u, x)$ for all $t^n_k \le u \le t^n_{k+2}$ and all ω' such that $\omega'(u) = \omega(u)$ for all $u \le t^n_k$. It follows that $Q^{a_n}_{t^n_k,\omega}(A_k) = 1$ for all ω. We deduce by induction that $Q^{a_n}_{r,\omega_0}(\{\omega', \ \omega'_{|[r,\infty[} \in \mathscr{C}([r, \infty[, I\!R^n)\} = 1$.
The a_n being uniformly bounded, for given r and ω_0, the set of probabilty measures $\{Q^{a_n}_{r,\omega_0}, n \in I\!N^*\}$ is weakly relatively compact. There is a subsequence weakly converging to a probability measure Q. From the continuity assumption on a, it follows that Q solves the martingale problem for $\mathscr{L}^{a,0}$ starting from ω_0 at time r. The uniqueness of the solution to this martingale problem implies that $Q = Q^a_{r,\omega_0}$. The set $\{\omega', \ \omega'_{|[r,\infty[} \in \mathscr{C}([r, \infty[, I\!R^n)\}$ is a closed subset of Ω. It follows from the Portmanteau Theorem, see e.g. [2] Theorem 2.1, that $Q^a_{r,\omega_0}(\{\omega', \omega'_{|[r,\infty[} \in \mathscr{C}([r, \infty[, I\!R^n)\} = 1$

Corollary 1 *For all continuous path ω_0 and all r, the support of the probability measure $Q^{a,ab}_{r,\omega_0}$ is contained in the set of continuous paths:*

$$Q^{a,b}_{r,\omega_0}\mathscr{C}([I\!R_+, I\!R^n)) = 1$$

Remark 1 In the simpler case where the function a is only defined on the set of continuous paths, the continuity hypothesis is just the usual continuity hypothesis for a function defined on $I\!R_+ \times \mathscr{C}([I\!R_+, I\!R^n))$ for the uniform norm topology. The associated martingale problem: probability measure solution to the martingale problem for $\mathscr{L}^{a,0}$ starting from ω_0 at time r can only be stated for initial continuous paths ω_0 (otherwise the path-dependent function $a(u, \omega)$ should be defined for paths ω which can have jumps before time u).

4 Stable Set of Probability Measures Solution to a Path-dependent Martingale Problem

In all the paper Ω denotes the set of càdlàg paths endowed with the Skorokhod topology. From now on, $a(s, \omega)$ is a given progressively continuous function on $I\!R_+ \times \Omega$ with values in non negative matrices. We assume that a is bounded and that $a(s, \omega)$ is invertible for all (s, ω). The explicit construction of dynamic risk measures developed here, making use of probability measures solutions to a martingale problem was first initiated in the unpublished preprint [4] in the Markovian case. We have introduced in [5] a general method to construct time consistent convex dynamic risk measures. This construction makes use of two tools. The first one is a set $\mathscr{Q}$ of equivalent probability measures stable by composition and stable by bifurcation (cf. [5] Definition 4.1). The second one consists in penalties $\alpha_{s,t}(Q)$, $s \le t$ defined for every probability measure Q in $\mathscr{Q}$, satisfying the local condition and the cocycle condition. The corresponding definitions are recalled in the Appendix.

4.1 Multivalued Mapping and Continuous Selector

Definition 6 X denotes the quotient of $I\!R_+ \times \Omega \times I\!R^n$ by the equivalence relation $\sim$: $(t, \omega, x) \sim (t', \omega', x')$ if $t = t'$, $x = x'$ and $\omega(u) = \omega'(u)\, \forall u < t$. The metric topology on X is induced by the one to one map from X into a subset of $I\!R_+ \times \Omega \times I\!R^n$: $(t, \omega, x) \to (t, \omega *_t x, x)$,
where $\omega *_t x$ has been defined in Sect. 2.2 Eq. (3).

The following observation is straightforward.

Remark 2 The set X is equipped with the metric topology defined above. Then every progressively continuous map on $I\!R_+ \times \Omega$ defines a unique continuous map on X. Furthermore every map continuous on the subset $\{(r, \omega, x),\ \omega = \omega *_r x\}$ of $I\!R_+ \times \Omega \times I\!R^n$ defines also a unique continuous map on X.

Recall now the definition of a multivalued mapping from X to Y. We use here the terminology chosen in [9]. Notice that the terminology used in [1] for multivalued mapping is correspondence.

Definition 7 A multivalued mapping Λ from X into $I\!R^n$ is a map Λ defined on X such that for all $(t, \omega, x) \in X$, $\Lambda(t, \omega, x)$ is a subset of $I\!R^n$. It can have additional properties:

1. Λ is convex if $\forall (t, \omega, x) \in X$, $\Lambda(t, \omega, x)$ is a convex subset of $I\!R^n$.
2. Λ is closed if for all (t, ω, x), $\Lambda(t, \omega, x)$ is closed.

Recall the following definition of a continuous selector (Definition 16.57 of [1]).

Definition 8 A selector from a multivalued mapping Λ from X into $I\!R^n$ is a function $s : X \to I\!R^n$ such that $s(t, \omega, x) \in \Lambda(t, \omega, x)$ for all $(t, \omega, x) \in X$. A continuous selector is a selector which is continuous.

Recall the following definition from [1] (Definition 16.2 and Lemma 16.5):

Definition 9 A multivalued mapping Λ from X into $I\!R^n$ is lower hemicontinuous if it satisfies the following equivalent conditions

- For every closed subset F of $I\!R^n$, $\Lambda^u(F) = \{(t, \omega, x) \in X : \Lambda(t, \omega, x) \subset F\}$ is closed
- For every open subset V of $I\!R^n$, $\Lambda^l(V) = \{(t, \omega, x) \in X : \Lambda(t, \omega, x) \cap V \neq \emptyset\}$ is open

Recall the following Michael Selection Theorem (cf. [1] Theorem 16.61)

Theorem 2 *A lower hemicontinuous mapping from a paracompact space into a Banach space with non empty closed convex values admits a continuous selector.*

Recall also that every metrizable space is paracompact (Theorem 2.86 of [1]).

4.2 Stable Set of Probability Measures Associated to a Multivalued Mapping

In all the following, Λ is a closed convex lower hemicontinuous multivalued mapping from X into $I\!R^n$. In the following (X_t) denotes the canonical process on Ω : For all càdlàg path ω, $X_t(\omega) = \omega(t)$. $(\mathscr{B}_t)$ is the canonical filtration.
Given the progressively continuous matrix valued map a, given $r \geq 0$, and $\omega \in \Omega$, we want to associate to Λ a stable set $\mathscr{Q}_{r,\omega}(\Lambda)$ of probability measures on $(\Omega, \mathscr{B})$ all equivalent with the probability measure $Q^a_{r,\omega}$ on $\mathscr{B}_t$. Furthermore we want to construct a continuous function v on X. Therefore we start with continuous selectors $\overline{\lambda}$ from Λ.

Definition 10 Let a be progressively continuous bounded defined on $I\!R_+ \times \Omega$ with values in non negative matrices, such that $a(t, \omega)$ is invertible for all (t, ω). Let Λ be a closed convex lower hemicontinuous multivalued mapping from X into $I\!R^n$.

- We define $L(\Lambda)$ to be the set of continuous bounded selectors from the multivalued mapping Λ.
- For given $r \geq 0$ and $\omega \in \Omega$, the set $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$ is the stable set of probability measures generated by the probability measures $Q^{a,a\lambda}_{r,\omega}$, $\overline{\lambda} \in L(\Lambda)$ with $\lambda(t, \omega') = \overline{\lambda}(t, \omega', X_t(\omega'))$

$$(\frac{dQ^{a,a\lambda}_{r,\omega}}{dQ^a_{r,\omega}})_{\mathscr{B}_T} = exp[\int_r^T \langle\lambda(t, \omega'), dX_t\rangle - \frac{1}{2}\int_r^T \langle\lambda(t, \omega'), a(t, \omega')\lambda(t, \omega')\rangle dt]$$

We give now a description of the set $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$.

Definition 11 We define $\tilde{L}(\Lambda)$ to be the set of processes $\overline{\mu}$ such that there is a finite subdivision $0 = s_0 < \cdot\cdot < s_i < s_{i+1} \cdot\cdot < s_k < \infty$. There is a continuous selector $\overline{\lambda}_{0,i_0}$ in $L(\Lambda)$. And for all $0 < i \leq k$ there is a finite partition $(A_{i,j})_{j\in I_i}$ of Ω into $\mathscr{B}_{s_i}$ measurable sets, and continuous selectors $\overline{\lambda}_{i,j}$ in $L(\Lambda)$ such that

$$\forall s_i < u \leq s_{i+1},\ \forall\omega' \in \Omega,\ \overline{\mu}(u, \omega', x) = \sum_{j\in I_i} \overline{\lambda}_{i,j}(u, \omega', x)1_{A_{i,j}}(\omega')$$
$$\forall s_k < u\ \forall\omega' \in \Omega,\ \overline{\mu}(u, \omega', x) = \sum_{j\in I_k} \overline{\lambda}_{k,j}(u, \omega', x)1_{A_{k,j}}(\omega')$$
$$\forall u \leq s_0,\ \forall\omega' \in \Omega,\ \overline{\mu}(u, \omega', x) = \overline{\lambda}_{0,i_0}(u, \omega', x) \qquad (13)$$

Remark 3 Every process $\overline{\mu}$ in $\tilde{L}(\Lambda)$ is bounded strictly progressive and $\mathscr{P} \times \mathscr{B}(I\!R^n)$ measurable where $\mathscr{P}$ is the predictable sigma algebra. However there is no uniform bound.

Proposition 2 *1. Let a be as above and $\overline{\mu} \in \tilde{L}(\Lambda)$. For all $r \geq 0$ and all $\omega \in \Omega$, there is a unique solution to the martingale problem for $\mathscr{L}^{a,a\mu}$ starting from ω at time r with $\mu(u,\omega) = \overline{\mu}(u,\omega,\omega(u))$. Furthermore for all $r < s$, the map $\omega' \to Q^{a,a\mu}_{s,\omega'}$ is $\mathscr{B}_s$ measurable and is a regular conditional probability distribution of $Q^{a,a\mu}_{r,\omega}$ given $\mathscr{B}_s$.*

2. Given $0 \leq r$, the set $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$ is the set of all probability measures $Q^{a,a\mu}_{r,\omega}$ for some process $\overline{\mu}$ belonging to $\tilde{L}(\Lambda)$.

Proof Let $\overline{\mu} \in \tilde{L}(\Lambda)$. There is a finite subdivision $0 = s_0 < \cdots < s_i < s_{i+1} < \cdots s_k < \infty$ such that $\overline{\mu}$ is described by Eq. (13). Let r and ω. We prove first by induction on k that there is a unique solution Q to the martingale problem for $\mathscr{L}^{a,a\mu}$ starting from ω at time r and that Q belongs to $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$.
For $k = 0$ the result is true by hypothesis.

Inductive step: Assume that $k \geq 1$ and that the result is proved for $k-1$. Let Q be a solution to the martingale problem for $\mathscr{L}^{a,a\mu}$ starting from ω at time r. Let $Q_{s_k,\omega'}$ be a regular conditional probability distribution of Q given $\mathscr{B}_{s_k}$. From [22] it follows that for Q almost all ω' in $A_{k,j}$, $Q_{s_k,\omega'}$ is a solution to the martingale problem for $\mathscr{L}^{a,a\lambda_{k,j}}$ starting from ω' at time s_k. The martingale problem for $\mathscr{L}^{a,a\lambda_{k,j}}$ is well posed. Let $Q^{a,a\lambda_{k,j}}_{s_k,\omega'}$ be the unique solution to the martingale problem for $\mathscr{L}^{a,a\lambda_{k,j}}$ starting from ω' at time s_k. It follows that $Q_{s_k,\omega'} = Q^{a,a\lambda_{k,j}}_{s_k,\omega'}$ on $A_{k,j}$ Q a.s. Thus for all ξ,

$$\begin{aligned} E_Q(\xi|\mathscr{B}_{s_k})(\omega') &= \sum_{j\in I_k} 1_{A_{k,j}}(\omega') Q^{a,a\lambda_{k,j}}_{s_k,\omega'}(\xi) \\ E_Q(\xi|\mathscr{B}_{s_k}) &= \sum_{j\in I_k} 1_{A_{k,j}} E_{Q^{a,a\lambda_{k,j}}_{r,\omega'}}(\xi|\mathscr{B}_{s_k}) \end{aligned} \tag{14}$$

On the other hand the restriction of Q to $\mathscr{B}_{s_k}$ is a solution to the martingale problem for $\mathscr{L}^{a,a\nu}$ where $\nu \in \tilde{L}(\Lambda)$ is associated to the subdivision $(s_i)_{0\leq i\leq k-1}$ and ν coincides with μ on $\mathscr{B}_{s_k}$. From the induction hypothesis it follows that the restriction of Q to $\mathscr{B}_{s_k}$ is uniquely determined, it coincides with $Q^{a,a\nu}_{r,\omega}$ and it belongs to $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$.
The end of the proof of the inductive step follows then from Eq. (14), from the $\mathscr{B}_{s_k}$ measurability of the map $\omega' \to Q^{a,a\lambda_{k,j}}_{s_k,\omega'}(\xi)$ and from the definition of $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$.
On the other hand it is easy to verify that the set $\{Q^{a,a\mu}_{r,\omega} : \overline{\mu} \in \tilde{L}(\Lambda)\}$ is stable. □

5 Construction of Penalties

In the preceding section we have constructed for all given (r,ω) a stable set of probability measures $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$ associated to a multivalued mapping Λ. In this section we construct penalties $\alpha_{st}(Q)$ for all $r \leq s \leq t$ and all $Q \in \tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$ making use of a function g.

Let $g : I\!R^+ \times \Omega \times I\!R^n \rightarrow I\!R \cup \{+\infty\}$ be a progressively measurable function. Let Λ be a closed convex multivalued Borel mapping such that for all $(t, \omega, x) \in I\!R^+ \times \Omega \times I\!R^n$, $\{0\} \subset \Lambda(t, \omega, x) \subset \{y \in I\!R^n | \ g(t, \omega *_t x, y) < \infty\}$. Define f as follows:

$$\forall z \in I\!R^d \ \ f(t, \omega, z) = \sup_{y \in \Lambda(t,\omega,X_t(\omega))} (-z.y - g(t, \omega, y)) \tag{15}$$

The following lemma is straightforward:

Lemma 1 *For all (t, ω), $f(t, \omega, .)$ is a closed convex function which is the dual transform of the function $\tilde{g}(t, \omega, .)$ where*

$$\begin{aligned} \tilde{g}(t, \omega, y) &= g(t, \omega, y) \ \ \textit{if} \ \ y \in \Lambda(t, \omega, X_t(\omega)) \\ &= +\infty \ \textit{else} \end{aligned} \tag{16}$$

For every (t, ω) $dom(\tilde{g}(t, \omega, .)) = \Lambda(t, \omega, X_t(\omega))$
If $g(t, \omega, 0) = 0 \ \forall (t, \omega)$, f takes values in $[0, \infty]$.
If g takes values in $[0, \infty]$ and satisfies $\forall (t, \omega)$, $inf_{y \in \Lambda(t,\omega,X_t(\omega))} g(t, \omega, y) = 0$ then for all (t, ω), $f(t, \omega, 0) = 0$.

Notice that, since Λ is a closed convex multivalued mapping, replacing g by $\tilde{g}$, one can always assume that for all (t, ω), $dom(g(t, \omega, .) = \{y \in I\!R^d \ |g(t, \omega, y) < \infty\}$ is closed, convex and equal to $\Lambda(t, \omega, X_t(\omega))$. We assume this in all the remainder.

Definition 12 1. g satisfies the following polynomial growth condition (GC1) if there is $K > 0$, $m \in I\!N^*$ and $\varepsilon > 0$ such that

$$\forall y \in \Lambda(u, \omega, X_u(\omega)), \ |g(u, \omega, y)| \leq K(1 + \sup_{s \leq u} ||X_s(\omega)||)^m (1 + ||y||^{2-\varepsilon}) \tag{17}$$

2. g satisfies the growth condition (GC2) if there is $K > 0$ such that

$$\forall y \in \Lambda(u, \omega, X_u(\omega)), \ |g(u, \omega, y)| \leq K(1 + ||y||^2) \tag{18}$$

Recall the following definition of BMO processes.

Definition 13 Let $C > 0$. Let P be a probability measure. A progressively measurable process μ belongs to $BMO(P)$ and has a BMO norm less or equal to C if for all stopping times τ,

$$E_P(\int_\tau^\infty ||\mu_s||^2 ds|\mathscr{F}_\tau) \leq C$$

Recall also from [15] that the stochastic exponential $\mathscr{E}(\mu)$ of a BMO process μ is uniformly integrable, and that the BMO norms with respect to P and $P(\mathscr{E}(\mu).)$ are equivalent. Also from [15], for all $C > 0$ there is $1 < p_0 < \infty$ such that for all

$BMO(P)$ process μ with $||\mu||_{BMO(P)} \leq C$, the stochastic exponential $\mathscr{E}(\mu)$ satisfies the reverse Hölder inequality:

$$[E_P(\mathscr{E}(\mu)^{p_0}|\mathscr{B}_s)]^{\frac{1}{p_0}} \leq K_C \mathscr{E}(\mu)_s \tag{19}$$

Definition 14 Assume that g is non negative or satisfies one of the growth conditions (GC1) or (GC2). Let $0 \leq r \leq T$. For all $BMO(Q^a_{r,\omega})$ process μ Λ-valued, for all $r \leq s \leq t \leq T$, define the penalty $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})$ as follows

$$\alpha_{s,t}(Q^{a,a\mu}_{r,\omega}) = E_{Q^{a,a\mu}_{r,\omega}}(\int_s^t g(u, \omega, \mu(u, \omega))du|\mathscr{B}_s) \tag{20}$$

We need to verify that the penalties are well defined for all BMO processes and that they satisfy the local property and the cocycle condition. (Definition introduced in [5], Definition 4.3 and recalled in the Appendix).

Proposition 3 *Assume that the process μ belongs to $BMO(Q^a_{r,\omega})$. Let C such that $||\mu||_{BMO(Q^a_{r,\omega})} \leq C$.*

- *1. Assume that g satisfies the growth condition (GC1). Then Eq. (20) defines a random variable in $L_p(Q^a_{r,\omega})$ for all $1 \leq p < \infty$, and for given p, the $L_p(Q^a_{r,\omega})$ norms of $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})$ are uniformly bounded for $r \leq s \leq t \leq T$, for all μ such that $||\mu||_{BMO(Q^a_{r,\omega})} \leq C$.*
  $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})$ belongs also to $L_1(Q^{a,a\mu}_{r,\omega})$ and the $L_1(Q^{a,a\mu}_{r,\omega})$ norms of $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})$ are uniformly bounded for $||\mu||_{BMO(Q^a_{r,\omega})} \leq C$ and $r \leq s \leq t \leq T$.
  2. *Assume that g satisfies (GC2), then the random variables $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})$ belong to $L_\infty(Q^a_{r,\omega})$ and are uniformly bounded for $||\mu||_{BMO(Q^a_{r,\omega})} \leq C$ and $r \leq s \leq t \leq T$.*
  3. *In case g is non negative, Eq. (20) defines a non negative $\mathscr{B}^r_s$ random variable.*
- *Assume that g satisfies the growth condition (GC1) or (GC2). Then the penalty defined in (20) satisfies the cocycle condition for every $Q^{a,a\mu}_{r,\omega}$: Let $r \leq s \leq t \leq u$*

$$\alpha_{s,u}(Q^{a,a\mu}_{r,\omega}) = \alpha_{s,t}(Q^{a,a\mu}_{r,\omega}) + E_{Q^{a,a\mu}_{r,\omega}}(\alpha_{t,u}(Q^{a,a\mu}_{r,\omega})|\mathscr{B}_s) \tag{21}$$

- *The penalty defined in (20) is local on $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$*
- *If $g(t, \omega, 0) = 0$ $\forall(t, \omega) \in I\!R^+ \times \Omega$, The probability measure $Q^a_{r,\omega}$ has zero penalty.*

Proof 1. Assume that the function g satisfies the growth condition (GC1). Without loss of generality one can assume that $m \geq 2$. Choose $p_1 > 1$ such that $(2 - \varepsilon)p_1 = 2$. Let q be the conjugate exponent of p_1.

It follows from the conditional Hölder inequality and the equivalence of the BMO norms with respect to $Q^a_{r,\omega}$ and $Q^{a,a\mu}_{r,\omega}$ that

$$E_{Q^{a,a\mu}_{r,\omega}}(\int_s^t (1+\sup_{s\le u}||X_s(\omega)||)^m||\mu_u(\omega)||^{2-\varepsilon}du|\mathscr{B}_s)$$
$$\le K_1 C^{\frac{1}{p_1}}[E_{Q^{a,a\mu}_{r,\omega}}(\sup_{s\le s'\le u}(1+||X_{s'}||)^{mq})|\mathscr{B}_s]^{\frac{1}{q}}(t-s)^{\frac{1}{q}} \tag{22}$$

Let p_0 be such that Eq. (19) is satisfied for $P = Q^a_{r,\omega}$. Let q_0 be the conjugate exponent of p_0. It follows from conditional Hölder inequality and (19) that

$$E_{Q^{a,a\mu}_{r,\omega}}(\sup_{s'\le u}(1+||X_{s'}||)^{mq}|\mathscr{B}_s)) \le K_C E_{Q^a_{r,\omega}}(\sup_{s\le s'\le t}(1+||X_{s'}||)^{mqq_0}|\mathscr{B}_s)^{\frac{1}{q_0}}$$
$$\le K_C E_{Q^a_{r,\omega}}(\sup_{s\le s'\le t}(1+||X_{s'}||)^{mqjq_0}|\mathscr{B}_s)^{\frac{1}{q_0 j}} \tag{23}$$

for all $j \ge 1$. The first assertion of 1 of the proposition follows then from the Eqs. (22) and (23) and the inequality $E_{Q^a_{r,\omega}}(\sup_{s\le u\le t}(1+||X_t||)^k)) < \infty$ for all $k \ge 2$ ([16], Chap. 2 Sect. 5).
The second asssertion of 1. of the proposition follows from Eq. (22) and then from Eq. (23) applied with $\mathscr{B}_s$ equal to the trivial sigma algebra.

2. Assume that g satisfies the the growth condition (GC2). Thus

$$|\alpha_{st}(Q^{a,a\mu}_{r,\omega})| \le K E_{Q^{a,a\mu}_{r,\omega}}(\int_s^t (1+||\mu_u(\omega)||^2)du|\mathscr{B}_s)$$

The result follows then from the BMO condition.

3. The case g non negative is trivial.

- The cocycle condition (21) follows easily from the definition (20) and the above integrability.
- We prove now that the penalty α is local on $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$. Let $\overline{\mu}, \overline{\nu} \in \tilde{L}(\Lambda)$. The probability measures $Q^{a,a\mu}_{r,\omega}$ and $Q^{a,a\nu}_{r,\omega}$ are equivalent to $Q^a_{r,\omega}$. Let $r \le s \le t$ and A be $\mathscr{B}_s$-measurable. Assume that for all X in $L^\infty(\mathscr{B}_t)$, $E_{Q^{a,a\mu}_{r,\omega}}(X|\mathscr{B}_s)1_A = E_{Q^{a,a\nu}_{r,\omega}}(X|\mathscr{B}_s)1_A$. It follows from the equality $\frac{\mathscr{E}(a\mu)_t}{\mathscr{E}(a\mu)_s}1_A = \frac{\mathscr{E}(a\nu)_t}{\mathscr{E}(a\nu)_s}1_A$ and the $\mathscr{P} \times \mathscr{B}(\mathbb{R}^n)$ measurability of $\overline{\mu}$ and $\overline{\nu}$ that $1_{]]s,t[}1_A\mu = 1_{]]s,t[}1_A\nu$ $Q^a_{r,\omega}$ a.s. From (20) we get $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})1_A = \alpha_{s,t}(Q^{a,a\nu}_{r,\omega})1_A$. Thus the penalty α is local on $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$.
- The last point follows easily from the definition of the penalty. $\square$

6 Time Consistent Dynamic Risk Measures Associated to Path-dependent Martingale Problems

We change the sign in the classical definition of risk measures in order to avoid the minus sign which appears in the time consistency property for usual dynamic risk measures. In fact $\rho_{st}(-X)$ are "usual" dynamic risk measures.

6.1 Normalized Time-Consistent Convex Dynamic Risk Measures

Proposition 4 *Let $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$ be the stable set of probability measures defined in Definition 10. Assume that g is non negative, and that for all (u, ω'), $g(u, \omega', 0) = 0$. Let $r \le s \le t$. The formula*

$$\rho^{r,\omega}_{s,t}(Y) = \mathrm{esssup}_{\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega} \in \tilde{\mathscr{Q}}_{\mathrm{r},\omega}()}(\mathrm{E}_{\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega}}(\mathrm{Y}|\mathscr{B}_{\mathrm{s}}) - \alpha_{\mathrm{s,t}}(\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega})) \tag{24}$$

where $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})$ is given by Eq. (20) defines a normalized time consistent convex dynamic risk measure on $L^\infty(\Omega, \mathscr{B}, Q^a_{r,\omega})$.
For given $0 \le r \le t$ and Y in $L^\infty(\Omega^r, \mathscr{B}^r_t, Q^a_{r,y})$, the process $(\rho^{r,\omega}_{s,t}(Y))_{r \le s \le t}$ admits a càdlàg version.

Proof Notice that for all bounded Y,

$$-||Y||_\infty \le E_{Q^a_{r,\omega}}(Y|\mathscr{B}_s) \le \rho^{r,\omega}_{s,t}(Y) \le \mathrm{esssup}_{\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega} \in \tilde{\mathscr{Q}}_{\mathrm{r},\omega}()}(\mathrm{E}_{\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega}}(\mathrm{Y}|\mathscr{B}_{\mathrm{s}}) \le ||\mathrm{Y}||_\infty$$

Thus for all $r \le s \le t$, $||\rho^{r,\omega}_{s,t}(Y)||_\infty \le ||Y||_\infty$. The first statement follows then from Definition 10, from Propositions 2 and 3 and from Theorem 4.4 of [5].
The proof of the regularity of paths which was given in [6] Theorem 3 for normalized convex dynamic risk measures time consistent for stopping times can be extended to normalized convex dynamic risk measures time consistent for deterministic times. □

We have the following extension of the dynamic risk measure to random variables essentially bounded from below:

Corollary 2 *The definition of $\rho^{r,\omega}_{s,t}(Y)$ can be extended to random variables Y ($\mathscr{B}_t$)-measurable which are only essentially bounded from below.*
$\rho^{r,\omega}_{s,t}(Y) = \lim_{n\to\infty} \rho^{r,\omega}_{s,t}(Y \wedge n)$. For every Y essentially bounded from below, the process $(\rho^{r,\omega}_{s,t}(Y))$ is optional.

Proof Let Y be $\mathscr{B}_t$-measurable and $Q^a_{r,\omega}$-essentially bounded from below, Y is the increasing limit of $Y_n = Y \wedge n$ as n tends to ∞. Define $\rho^{r,\omega}_{s,t}(Y)$ as the increasing limit of $\rho^{r,\omega}_{s,t}(Y_n)$. As we already know that for given s and t, $\rho^{r,\omega}_{s,t}(Y)$ defined on bounded random variables by formula (24) is continuous from below, the extended definition coincides with the previous one on $Q^a_{r,\omega}$-essentially bounded random variables. From Proposition 4 for every n one can choose a càdlàg version of the process $\rho^{r,\omega}_{s,t}(Y_n)$. Thus the map $(s,\omega) \to \rho^{r,\omega}_{s,t}(Y) = \lim \rho^{r,\omega}_{s,t}(Y_n)$ is measurable for the optional sigma algebra. □

6.2 General Time-Consistent Convex Dynamic Risk Measures

In this section the function g (and thus the penalty) is not assumed to be non negative.

Definition 15 Let Q be a probability measure on $(\Omega, (\mathscr{B}_t)_{t\in\mathbb{R}_+})$. The multivalued mapping Λ is $BMO(Q)$ if there is a map $\phi \in BMO(Q)$ such that

$$\forall (u,\omega),\ \sup\{||y||,\ y \in \Lambda(u,\omega)\} \leq \phi(u,\omega)$$

In the following, p_0 is chosen such that the reverse Hölder inequality (19) is satisfied for $a\phi$.

Theorem 3 *Let (r,ω). Assume that the multivalued set Λ is $BMO(Q^a_{r,\omega})$. Let $\tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$ be the stable set of probability measures defined in Definition 10.*

Let $r \leq s \leq t$. Let

$$\rho^{r,\omega}_{s,t}(Y) = \mathrm{esssup}_{\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega} \in \tilde{\mathscr{Q}}_{\mathrm{r},\omega}()}(\mathrm{E}_{\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega}}(\mathrm{Y}|\mathscr{B}_{\mathrm{s}}) - \alpha_{\mathrm{s,t}}(\mathrm{Q}^{\mathrm{a,a}\mu}_{\mathrm{r},\omega})) \tag{25}$$

where $\alpha_{s,t}(Q^{a,a\mu}_{r,\omega})$ is given by Eq. (20)

- *Assume that g satisfies the growth condition (GC1). The above Eq. (25) defines a dynamic risk measure $(\rho^{r,\omega}_{s,t})$ on $L_p(Q^a_{r,\omega}, (\mathscr{B}_t))$ for all $q_0 \leq p < \infty$, (where q_0 is the conjugate exponent of p_0 chosen as above). These dynamic risk measures are time consistent for stopping times taking a finite number of values.*
- *Assume that g satisfies the growth condition (GC2). The above Eq. (25) defines a dynamic risk measure $(\rho^{r,\omega}_{s,t})$ on $L_\infty(Q^a_{r,\omega}, (\mathscr{B}_t))$, and also on every $L_p(Q^a_{r,\omega}, (\mathscr{B}_t))$ for $q_0 \leq p < \infty$. These dynamic risk measures are time consistent for stopping times taking a finite number of values.*

Proof There is a constant $C > 0$ such that for all $Q^{a,a\mu}_{r,\omega} \in \tilde{\mathscr{Q}}_{r,\omega}(\Lambda)$, $||a\mu||_{BMO(Q^a_{r,\omega})} \leq C$. It follows from the reverse Hölder inequality (19), that for all non negative measurable Y, $E_{Q^{a,a\mu}_{r,\omega}}(Y|\mathscr{B}_s) \leq K_C(E_{Q^a_{r,\omega}}(||Y||^{q_0}|\mathscr{B}_s)^{\frac{1}{q_0}}$. Thus $Y \to E_{Q^{a,a\mu}_{r,\omega}}(Y|\mathscr{B}_s)$ defines a linear continuous map on $L_p(Q^a_{r,\omega})$ with values $L_p(Q^a_{r,\omega})$ for all $q_0 \leq p \leq \infty$, and that for given p, the norms of these linear maps are uniformly bounded for Λ valued. From Proposition 3, it follows then that Eq. (25) defines a dynamic risk measure $(\rho^{r,\omega}_{s,t})$ on $L_p(Q^a_{r,\omega}, (\mathscr{B}_t))$ for all $q_0 \leq p < \infty$ in case (GC1). Under assumption (GC2), equation (25) defines a dynamic risk measure $(\rho^{r,\omega}_{s,t})$ on $L_p(Q^a_{r,\omega}, (\mathscr{B}_t))$ for all L_p, $q_0 \leq p \leq \infty$.

The time consistency for stopping times taking a finite number of values follows from the stability property of the set of probability measures as well as the cocycle and local property of the penalties (cf. [5] in L_∞ case). The proof is the same in L_p case.

7 Strong Feller Property

7.1 Feller Property for Continuous Parameters

We assume that the progressively measurable function g is a Caratheodory function on $\mathbb{R}_+ \times \Omega \times \mathbb{R}^n$, that is for all u, $(\omega, x) \to g(u, \omega, x)$ is continuous. The support of $Q^{a,\mu}_{r,\omega}$ is contained in the paths ω' continuous on $[r, \infty[$ (Proposition 1) and which coincide with ω on $[0, r]$. It follows that for every function λ progressively continuous bounded, and all $u > r$, the function $\omega' \to g(u, \omega', \lambda(u, \omega'))$ is continuous on the support of $Q^{a,a\lambda}_{r,\omega}$. We prove then the following Feller property for the penalty.

Proposition 5 *Let a be progressively continuous bounded such that $a(s, \omega)$ is invertible for all (s, ω). Let λ be progressively continuous bounded. Assume that g is a real valued Caratheodory function satisfying the growth condition (GC1) or (GC2).*

1. *There is a strictly progressive real valued map $L(g)$ on $[0, t] \times \Omega \times \mathbb{R}^n$ continuous on $\{(s, \omega, x),\ \omega = \omega *_s x\}$ such that*

$$E_{Q^{a,a\lambda}_{s,\omega *_s x}} \int_s^t g(u, \omega', \lambda(u, \omega')du) = L(g)(s, \omega, x) \quad \forall s \le t \ \ \omega \text{ and } x \tag{26}$$

2. *For all $0 \le r \le s \le t$, and $\omega \in \Omega$, there is a $Q^{a,\lambda}_{r,\omega}$-null set N such that for all $\omega_1 \in N^c$,*

$$E_{Q^{a,a\lambda}_{r,\omega}}\left(\int_s^t g(u, \omega', \lambda(u, \omega'))du|\mathscr{B}_s\right)(\omega_1) = L(g)(s, \omega_1, X_s(\omega_1)) \tag{27}$$

Proof We only need to prove 1.
Step 1: Assume that the function g is bounded. Then $\int_s^t g(u, \omega', \lambda(u, \omega'))du$ is a continuous bounded function of ω' on the support of $Q^{a,a\lambda}_{s,\omega}$. The map λ being progressively continuous bounded, the continuity property for $L(g)$ follows easily from Theorem 1.
Step 2: general case. Notice that λ is bounded.
Thus under assumption (GC2), $g(u, \omega', \lambda(u, \omega')$ is uniformly bounded and the continuity property of $L(g)$ follows from step 1.

Under assumption (GC1), let (s_n, ω_n, x_n), $\omega_n = \omega_n *_{s_n} x_n$ with limit (s, ω, x), $\omega = \omega *_s x$. By definition the sequence $\omega_n *_{s_n} x_n$ has limit $\omega *_s x$. It follows from [3] that the set of probability measures $\mathscr{Q} = \{Q^{a\lambda}_{s_n,\omega_n *_{s_n} x_n},\ \ n \in \mathbb{N}\} \cup \{Q^{a\lambda}_{s,\omega *_s x}\}$ is weakly relatively compact and thus tight. Thus for all $\eta > 0$, there is a compact set $\mathscr{K}$ such that $Q(\mathscr{K}^c) < \eta$ for all $Q \in \mathscr{Q}$. From the growth condition (GC1), the existence of a uniform bound for $E_Q[\int_s^t (1 + \sup_{s\leq u} ||X_u||^m)du]^k$ for $Q \in \mathscr{Q}$ and the Hölder inequality, it follows that there is a progressively continuous bounded function g_1 such that for all Q in $\mathscr{Q}$,

$$E_Q \int_s^t (||g(u, \omega, (a\lambda)(u, \omega) - g_1(u, \omega)||))du \leq \varepsilon. \tag{28}$$

The result follows then from step 1. □

We introduce now a class of $\mathscr{B}_t$ measurable functions on Ω satisfying a continuity condition derived from the progressive continuity condition that we have introduced for progressive functions and from the continuity property proved in Theorem 1.

Definition 16 Let $t > 0$. The function h defined on Ω belongs to $\mathscr{C}_t$ if there is a function $\tilde{h}$ on $\Omega \times \mathbb{R}^n$ such that

- $h(\omega) = \tilde{h}(\omega *_t X_t(\omega), X_t(\omega))$
- $\tilde{h}(\omega, x) = \tilde{h}(\omega', x)$ if $\omega(u) = \omega'(u)\ \ \forall u < t$

and such that $\tilde{h}$ is continuous bounded on $\{(\omega, x),\ \ \omega = \omega *_t x\} \subset \Omega \times \mathbb{R}^n$

Corollary 3 *Assume that a is progressively continuous bounded and that $a(s, \omega)$ is invertible for all (s, ω). Let λ be progressively continuous bounded. Let $h \in \mathscr{C}_t$. Asssume that the penalty $\alpha_{s,t}$ is given by Eq. (20) for some Caratheodory function g on $\mathbb{R}_+ \times \Omega \times \mathbb{R}^n$ satifying the growth condition (GC1) or (GC2). There is a strictly progressive map $L^{a,\lambda}(h)$ continuous on $\{(u, \omega *_u x, x),\ \ 0 \leq u \leq t\}$ such that for all $0 \leq r_0 \leq r \leq t$,*

$$\begin{aligned} L^{a,\lambda}(h)(t, \omega, x) &= \tilde{h}(\omega, x) \\ L^{a,\lambda}(h)(r, \omega, x) &= E_{Q^{a,a\lambda}_{r,\omega *_r x}}(h) - \alpha_{r,t}(Q^{a,a\lambda}_{r,\omega *_r x}) \\ &= [E_{Q^{a,a\lambda}_{r_0,\omega_0}}(h|\mathscr{B}_r) - \alpha_{r,t}(Q^{a,a\lambda}_{r_0,\omega_0})](\omega *_r x)\ \ Q^{a,a\lambda}_{r_0,\omega_0}\ a.s. \end{aligned} \tag{29}$$

Proof The result follows from Theorem 1 and from Proposition 5. □

7.2 Feller Property for the Dynamic Risk Measure

Proposition 6 *Let μ in $\tilde{L}(\Lambda)$ (Definition 11). Let $t > 0$ and $h \in \mathscr{C}_t$. Asssume that the penalty $\alpha_{s,t}$ is given by Eq. (20) for some Caratheodory function g on $\mathbb{R}_+ \times \Omega \times \mathbb{R}^n$ satifying the growth condition (GC1) or (GC2). There is a strictly progressive map*

$\Phi_\mu(h)$ continuous on $\{(u,\omega,x),\ \omega=\omega *_u x,\ u\leq t\}$ such that $\Phi_\mu(h)(t,\omega,x)=\tilde{h}(\omega,x)$ for all ω, and such that for all $r\leq s\leq t$, there is a process ν_s in $\tilde{L}(\Lambda)$ such that

$$\Phi_\mu(h)(s,\omega,x)=E_{Q^{a,a\nu_s}_{s,\omega *_s x}}(h)-\alpha_{s,t}(Q^{a,a\nu_s}_{s,\omega *_s x}) \tag{30}$$

$$E_{Q^{a,a\nu_s}_{r,\omega_0}}(h|\mathscr{B}_s)(\omega')-\alpha_{s,t}(Q^{a,a\nu_s}_{r,\omega_0})(\omega')=\Phi_\mu(h)(s,\omega',X_s(\omega'))\quad Q^a_{r,\omega_0}\ a.s. \tag{31}$$

and

$$E_{Q^{a,a\mu}_{r,\omega_0}}(h|\mathscr{B}_s)-\alpha_{s,t}(Q^{a,a\mu}_{r,\omega_0})\leq E_{Q^{a,a\nu_s}_{r,\omega_0}}(h|\mathscr{B}_s)-\alpha_{s,t}(Q^{a,a\nu_s}_{r,\omega_0})\ \ Q^a_{r,\omega_0}\ a.s. \tag{32}$$

Proof The proof is done in two steps. The first step is the construction of $\Phi_\mu(h)$ given μ. The second step is the construction of ν_s given μ and s. μ belongs to $\tilde{L}(\Lambda)$. From Definition 11, let $0=s_0<\cdot\cdot<s_i<s_{i+1}\cdot\cdot<s_k<s_{k+1}=\infty$ be a finite partition such that Eq. (13) is satisfied.

- First step: construction of $\Phi_\mu(h)$. We construct $\Phi_\mu(h)$ recursively on $[s_i,s_{i+1}[$. Let s_n be such that $s_n<t\leq s_{n+1}$. From Corollary 3, for all $j\in I_n$ there is a strictly progressive map $L^{a,\lambda_{n,j}}(h)$ continuous on $\{(u,\omega,x),\ \omega=\omega *_u x,\ u\leq t\}$ satisfying Eq. (29). Let

$$\Phi_\mu(h)(s,\omega,x)=\sup_{j\in I_n}L^{a,\lambda_{n,j}}(h)(s,\omega,x)\ \ \forall s\in[s_n,t[\tag{33}$$

  Let $h_n(\omega)=\Phi_\mu(h)(s_n,\omega,X_{s_n}(\omega))$. The function h_n belongs to $\mathscr{C}_{s_n}$ and $\tilde{h}_n(\omega,x)=\Phi_\mu(h)(s_n,\omega,x)$. Then we can proceed on $[s_{n-1},s_n[$ with h_n. We construct recursively the strictly progressive map $\Phi_\mu(h)$ continuous on $\{(u,\omega,x),\ \omega=\omega *_u x,\ u\leq t\}$. Notice that for all $s\in[s_i,s_{i+1}]$, there are $\mathscr{B}_s$ measurable sets $(C_{s,j})_{j\in I_i}$, such that

$$\Phi_\mu(h)(s,\omega,x)=\sum_{j\in I_i}1_{C_{s,j}}(\omega)L^{a,\lambda_{i,j}}(h_{i+1})(s,\omega,x) \tag{34}$$

- Second step: Given $s\in[r,t]$, construction of the process ν_s.
  There is a unique k such that $s\in]s_k,s_{k+1}]$. For $i>k$ for all $u\in]s_i,s_{i+1}]$, define

$$\nu_s(u,\omega)=\sum_{j\in I_i}1_{C_{s_i,j}}(\omega)\lambda_{i,j}(u,\omega) \tag{35}$$

  And for $u\in]s,s_{k+1}]$, define

$$\nu_s(u,\omega)=\sum_{j\in I_k}1_{C_{s,j}}(\omega)\lambda_{k,j}(u,\omega) \tag{36}$$

and $\nu_s(u,\omega)=\mu(u,\omega)$ for all $0\leq u\leq s$. It follows from the construction of ν_s that the process ν_s belongs to $\tilde{L}(\Lambda)$. It follows also recursively that for all $i>k$ and $\omega\in\Omega$,

$$E_{Q^{a,a\nu_s}_{s_i,\omega}}(h)-\alpha_{s_i,t}(Q^{a,a\nu_s}_{s_i,\omega})=\Phi_\mu(h)(s_i,\omega,X_{s_i}(\omega)) \tag{37}$$

and for all $\omega\in\Omega$,

$$E_{Q^{a,a\nu_s}_{s,\omega}}(h)-\alpha_{s,t}(Q^{a,a\nu_s}_{s,\omega})=\Phi_\mu(h)(s,\omega,X_s(\omega)) \tag{38}$$

By construction the following inequality is satisfied:

$$E_{Q^{a,a\mu}_{s,\omega}}(h)-\alpha_{s,t}(Q^{a,a\mu}_{s,\omega})\leq\Phi_\mu(h)(s,\omega,X_s(\omega)) \tag{39}$$

It follows then from Proposition 2 that

$$E_{Q^{a,a\mu}_{r,\omega_0}}(h|\mathscr{B}_s)-\alpha_{s,t}(Q^{a,a\mu}_{r,\omega_0})\leq E_{Q^{a,a\nu_s}_{r,\omega_0}}(h|\mathscr{B}_s)-\alpha_{s,t}(Q^{a,a\nu_s}_{r,\omega_0})\ \ Q^a_{r,\omega_0}\ a.s.$$

and

$$[E_{Q^{a,a\nu_s}_{r,\omega_0}}(h|\mathscr{B}_s)-\alpha_{s,t}(Q^{a,a\nu_s}_{r,\omega_0})](\omega')=\Phi_\mu(h)(s,\omega',\omega'(s))\ \ Q^a_{r,\omega_0}\ a.s. \qquad \square$$

Theorem 4 *Assume that the hypothesis of Theorem 3 are satisfied and that g is Caratheodary function. The time consistent dynamic risk measure $(\rho^{r,\omega}_{s,t})_{r\leq s\leq t}$ defined on $L^\infty(\Omega,\mathscr{B},Q^a_{r,\omega})$ by Eq. (25) satisfies the following Feller property: For every function $h\in\mathscr{C}_t$, there is a progressive map $R(h)$ on $\mathbb{R}_+\times\Omega$, $R(h)(t,\omega)=h(\omega)$, such that $\overline{R}(h)$ is lower semi continuous on $\{(u,\omega,x),\ u\leq t,\ \omega=\omega*_u x\}$ and such that the following equation is satisfied*

$$\forall s\in[r,t],\ \forall\omega'\in\Omega,\ \rho^{s,\omega'}_{s,t}(h)=R(h)(s,\omega') \tag{40}$$

$$\forall 0\leq r\leq s\leq t,\quad \rho^{r,\omega}_{s,t}(h)(\omega')=\overline{R}(h)(s,\omega',\omega'(s))\ \ Q^a_{r,\omega}\ a.s. \tag{41}$$

($\overline{R}(h)$ denotes the strictly progressive map on $\mathbb{R}_+\times\Omega\times\mathbb{R}^n$ associated to $R(h)$ in the one to one corrrespondence introduced in Sect. 2).

Proof For all $\mu\in\tilde{L}(\Lambda)$, let $\Phi_\mu(h)$ be the strictly progressive map, continuous on $\{(u,\omega,x),\ \omega=\omega*_u x,\ u\leq t\}$ constructed in Proposition 6. Let $\overline{R}(h)=\sup_{\mu\in\tilde{L}(\Lambda)}\Phi_\mu(h)$. The function $\overline{R}(h)$ is then lower semi continuous on $\{(u,\omega,x),\ \omega=\omega*_u x,\ u\leq t\}$. Let $R(h)(u,\omega)=\overline{R}(h)(u,\omega,X_u(\omega))$. Equations (40) and (41) follow easily from Proposition 6.

8 Existence of Viscosity Solutions for Path-dependent PDEs

8.1 Existence of Viscosity Supersolutions

Recall that Λ is a closed convex lower hemicontinuous multivalued mapping from X into $I\!R^n$ (Sect. 4.2). Let f be the convex conjugate of g defined as

$$f(u,\omega,z) = \sup_{y\in\Lambda(u,\omega)} (z^* y - g(u,\omega,y)) \tag{42}$$

We prove now that for all $h \in \mathscr{C}_t$ the map $R(h)$ of Theorem 4 leads to viscosity solutions for the following semi-linear second order PDE.

$$\begin{cases} -\partial_u v(u,\omega) - \mathscr{L}v(u,\omega) - f(u,\omega,a(u,\omega)D_x v(u,\omega)) = 0 \\ v(t,\omega) = f(\omega) \\ \mathscr{L}v(u,\omega) = \frac{1}{2}Tr(a(u,\omega)D_x^2(v)(u,\omega)) \end{cases} \tag{43}$$

Theorem 5 *Fix (t_0,ω_0). Assume that the mutivalued set Λ is BMO($Q^a_{t_0,\omega_0}$) (Definition 15). Assume that the function g satisfies the preceding hypothesis (g is a Caratheodory function and satisfies the growth condition (GC1) or (GC2)). Assume furthermore that g is upper semicontinuous on $\{(s,\omega,y),(s,\omega)\in X,\ y\in \Lambda(s,\omega,\omega(s)\}$. For all r and ω, let $(\rho^{r,\omega}_{s,t})$ be the dynamic risk measure given by Eq. (24) where the penalty satisfies Eq. (20).*
Let $h\in\mathscr{C}_t$. The function $R(h)$ is progressive and $\overline{R}(h)$ is lower semi continuous on $\{(u,\omega,x),\ \omega=\omega *_u x,\ u\le t\}$ (Theorem 4). $R(h)$ is a viscosity supersolution of the path-dependent second order partial differential equation (43) at each point (t_0,ω_0) such that $f(t_0,\omega_0,a(t_0,\omega_0)z)$ is finite for all z.

Proof Let $x_0=\omega_0(t_0)$. From Theorem 4, the function $R(h)$ is progressive and $\overline{R}(h)$ is lower semi continuous on $\{(u,\omega,x),\ \omega=\omega *_u x,\ u\le t\}$. We prove first that $R(h)$ is bounded from below on some $D_\varepsilon(t_0,\omega_0)$. $R(h)(u,\omega)\ge E_{Q^a_{u,\omega}}(h)-\alpha_{ut}(Q^a_{u,\omega})$. For all given $k\ge 2$, $E_{Q^a_{u,\omega}}(\sup_{s\le u\le t}(1+||X_t||)^k))$ is uniformly bounded for $(u,\omega)\in D_\varepsilon(t_0,\omega_0)$. The result follows then from either the (GC1) condition or the (GC2).
Let ϕ progressive, $\overline{\phi}$ in $\mathscr{C}_b^{1,0,2}(I\!R_+\times\Omega\times I\!R^n)$ such that $\phi(t_0,\omega_0)=R(h)(t_0,\omega_0)$ and (t_0,ω_0) is a local minimizer of $R(h)-\phi$ on $D_\eta(t_0,\omega_0)$ for some $\eta>0$.

- Step 1: Continuity properties
  By hypothesis $f(t_0,\omega_0,a(t_0,\omega_0)D_x\overline{\phi}(t_0,\omega_0,x_0))<\infty$. Thus for all $\varepsilon>0$, there is $\lambda_0\in\Lambda(t_0,\omega_0)$ such that

$$D_x\overline{\phi}(t_0,\omega_0,x_0)^t a(t_0,\omega_0)\lambda_0 - g(t_0,\omega_0,\lambda_0) > f(t_0,\omega_0,a(t_0,\omega_0)D_x\overline{\phi}(t_0,\omega_0,x_0))-\varepsilon \tag{44}$$

  The multivalued set Λ is assumed to be lower hemicontinuous. It follows that for all $K>0$, Λ_K is also lower hemicontinuous, where $\Lambda_K(u,\omega,x)=\{z\in$

$\Lambda(u, \omega, x),\ ||z|| < K\}$. Choose $K > ||\lambda_0||$. From Theorem 2, there is thus a continuous bounded selector $\overline{\lambda}(u, \omega, x)$ of Λ_K defined on X such that $\overline{\lambda}(t_0, \omega_0, x_0) = \lambda_0$. From the upper semi-continuity condition satisfied by g, and the continuity of the map $\overline{\lambda}$, it follows that for all $\varepsilon > 0$, there is $\eta_1 > 0$, such that for $t_0 \le u \le t \le t_0 + \eta_1$, $d(\omega, \omega_0) < \eta_1$ and $||\omega(t) - x_0|| < \eta_1$,

$$g(t, \omega, \overline{\lambda}(t, \omega, \omega(t))) - g(t_0, \omega_0, \lambda_0) < \varepsilon \tag{45}$$

From the continuity of the function $\overline{\lambda}$, the hypothesis $\overline{\phi} \in \mathscr{C}_b^{1,0,2}$ and the progressive continuity of a, there is η_2 such that for $t_0 \le u \le t \le t_0 + \eta_2$, $||x - x_0|| < \eta_2$ and $d(\omega *_u x, \omega_0) < \eta_2$,

$$\begin{aligned} &|\partial_u \overline{\phi}(u, \omega, x) + \frac{1}{2} Trace(D_x^2 \overline{\phi}\,\overline{a}(u, \omega, x)) + (D_x \overline{\phi}^t \overline{a}\overline{\lambda})(u, \omega, x) - \\ &\partial_u \overline{\phi}(t_0, \omega_0, x_0) + \frac{1}{2} Trace(D_x^2 \overline{\phi}\,\overline{a})(t_0, \omega_0, x_0)) + D_x \overline{\phi}^t \overline{a}(t_0, \omega_0, x_0)\lambda_0| \le \varepsilon \end{aligned} \tag{46}$$

The maps a and λ are bounded. It follows from [22] that there is $0 < \alpha < inf(\eta, \eta_1, \eta_2)$ such that

$$Q^{a,a\lambda}_{t_0,\omega_0}(A) < \varepsilon \ \ with \ \ A = \{\omega \mid \sup_{t_0 \le u \le t_0 + \alpha} ||\omega(u) - \omega_0(t_0)|| \ge inf(\eta, \eta_1, \eta_2)\ \} \tag{47}$$

Let $C = \{\omega,\ \ \omega(u) = \omega_0(u)\ \forall 0 \le u \le t_0,\ \ \ \sup_{t_0 \le u \le t_0+\alpha} ||\omega(u) - \omega_0(t_0)|| < inf(\eta, \eta_1, \eta_2)\}$

- Step 2: Time consistency
  For all $0 < \beta < \alpha$, let δ be the stopping time $\delta = \beta 1_C$. The stopping time δ takes only 2 different values. By definition of the probability measure $Q^{a,a\lambda}_{t_0,\omega_0}$, it follows from (47) that $Q^{a,a\lambda}_{t_0,\omega_0}(C) > (1 - \varepsilon)$.
  The dynamic risk measure $(\rho^{t_0,\omega_0}_{u,v})_{0 \le u \le v}$, is time consistent for stopping times taking a finite number of values, thus

$$\rho^{t_0,\omega_0}_{t_0,t}(h) = \rho^{t_0,\omega_0}_{t_0,t_0+\delta}(\rho^{t_0,\omega_0}_{t_0+\delta,t}(h)) \tag{48}$$

From Theorem 4, the lower semi continuous function $\overline{R}(h)$ satisfies: $(\rho^{t_0,\omega_0}_{t_0+\delta,t}(h)(\omega) = \overline{R}(h)(t_0 + \delta, \omega, X_{t_0+\delta}(\omega))$ $Q^a_{t_0,\omega_0}$ a.s., and $\rho^{t_0,\omega_0}_{t_0,t}(h) = R(h)(t_0, \omega_0)$. Let $\overline{\lambda}$ be the continuous function defined in step 1. By hypothesis $R(h) \ge \phi$ on $D_\eta(t_0, \omega_0)$. For all $\omega \in C$ and $u \le \beta$, $(u, \omega *_u \omega(u)) \in D_\eta(t_0, \omega_0)$. Since the functions $R(h)$ and ϕ are progressive it follows that $R(h) \ge \phi$ on $[t_0, t_0 + \beta] \times C$. It follows then from the equality $\phi(t_0, \omega_0) = R(h)(t_0, \omega_0)$ and from the definition of $\rho^{t_0,\omega_0}_{t_0,t_0+\delta}$ that

$$\phi(t_0, \omega_0) \ge E_{Q^{a,a\lambda}_{t_0,\omega_0}}[\overline{\phi}(t_0 + \delta, \omega, X_{t_0+\delta}(\omega)) - \int_{t_0}^{t_0+\delta} g(u, \omega, \lambda(u, \omega, X_u(\omega)))du] \tag{49}$$

- Step 3: Martingale problem
  The probability measure $Q^{a,a\lambda}_{t_0,\omega_0}$ is solution to the martingale problem for $\mathscr{L}^{a,a\lambda}$ starting from ω_0 at time t_0. The function $\overline{\phi}$ is strictly progressive and belongs to $\mathscr{C}^{1,0,2}_b(I\!R_+ \times \Omega \times I\!R^n)$. It follows from the martingale property proved in [3] that

$$
\begin{aligned}
0 \geq\ & E_{Q^{a,a\lambda}_{t_0,\omega_0}}\Big[\int_{t_0}^{t_0+\delta} (\partial_u\overline{\phi}(u,\omega,X_u(\omega)) + \frac{1}{2}Trace((D^2_x\overline{\phi}\overline{a})(u,\omega,X_u(\omega))))du\Big] \\
& + E_{Q^{a,a\lambda}_{t_0,\omega_0}}\Big[\int_{t_0}^{t_0+\delta} (D_x\overline{\phi}^t\overline{a}\overline{\lambda})(u,\omega,X_u(\omega)) - g(u,\omega,\overline{\lambda}(u,\omega,X_u(\omega)))du\Big] \quad (50)
\end{aligned}
$$

- Step 4: Conclusion
  Divide (50) by β and let β tend to 0. It follows from the definitions of C and δ and the Eqs. (44)–(46) proved in step 1 and $Q^{a,a\lambda}_{t_0,\omega_0}(C) > 1-\varepsilon$ that

$$
-\partial_u\overline{\phi}(t_0,\omega_0,x_0) - \mathscr{L}\overline{\phi}(t_0,\omega_0,x_0) - f(t_0,\omega_0,\sigma^t(t_0,\omega_0)D_x\overline{\phi}(t_0,\omega_0,x_0)) \geq -2\varepsilon(1-\varepsilon)
$$

  This gives the result. □

8.2 Existence of Viscosity Subsolutions

In this section we will assume that the set Λ has some uniform BMO property.

Definition 17 The multivalued mapping Λ is uniformly BMO with respect to a if there is a non negative progressively measurable map φ and $C > 0$ such that for all $0 \leq s$,

$$
\sup\{|\lambda|,\ \lambda \in \Lambda(s,\omega)\} \leq \varphi(s,\omega) \quad (51)
$$

and such that for all (s',ω'), the unique solution to the martingale problem $\mathscr{L}^a$ starting from ω' at time s' satisfies:

$$
Q^a_{s',\omega'}\Big(\int_{s'}^{\infty} \varphi(s,\omega)^2 ds\Big) \leq C \quad (52)
$$

Of course the above condition is satisfied if $\sup_\omega \int_0^\infty \varphi(s,\omega)^2 ds < \infty$.
The name "uniform BMO" property is justifed by the following result.

Lemma 2 *Assume that the multivalued mapping Λ is uniformly BMO with respect to a. Then for all (r,ω) and all process μ Λ valued such that $\overline{\mu}$ is $\mathscr{P} \times \mathscr{B}(I\!R^n)$ measurable, μ belongs to $BMO(Q^a_{r,\omega})$ and $||\mu||_{BMO(Q^a_{r,\omega})} \leq C$*

Proof Let r,ω and a stopping time $\tau \geq r$. It follows from [22] and from uniqueness of the solution to the martingale problem for $\mathscr{L}^{a,0}$ starting from ω' at time $\tau(\omega')$ that for $Q^a_{r,\omega}$ almost all ω',

$$E_{Q^a_{r,\omega}}(\int_\tau^\infty \varphi(u,\omega)^2 du|\mathscr{B}_\tau)(\omega') = E_{Q^a_{\tau(\omega'),\omega'}}(\int_{\tau(\omega')}^\infty \varphi(u,\omega)^2 du) \leq C \qquad \square$$

Let $h \in \mathscr{C}_t$. The function $\overline{R}(h)$ is lower semi continuous in viscosity sense but it is not necessarily upper semi continuous. Therefore we need to introduce the upper semi continuous envelope of $R(h)$ in the viscosity sense according to Sect. 2.3. Denote it $R(h)^*$.

$$R(h)^*(s,\omega) = \lim_{\eta \to 0} \sup\{R(h)(s',\omega'),\ (s',\omega') \in D_\eta(s,\omega)\}$$

Theorem 6 *Let (t_0,ω_0) be given. Assume that the mutivalued set Λ is uniformly BMO with respect to a. Assume that the function g is a Caratheodory function satisfying the growth condition (GC1) or (GC2) and that the Fenchel transform f of g is progressively continuous.*
Let $h \in \mathscr{C}_t$. The map $R(h)^*$ is progressive, $\overline{R(h)^*}$ is upper semicontinuous in viscosity sense. $R(h)^*$ is a viscosity subsolution of the path-dependent second order partial differential equation (43).

Proof Let $x_0 = \omega_0(t_0)$. The progressivity of $R(h)^*$ follows from the equality $D_\eta(s,\omega) = D_\eta(s,\omega *_s \omega(s))$. The upper semicontinuity property follows from the definition of $R(h)^*$. We prove first that $R(h)$ is bounded on some $D_\varepsilon(t_0,\omega_0)$. $R(h)(u,\omega) = \sup_\mu(E_{Q^{a,a\mu}_{u,\omega}}(h) - \alpha_{u,t}(Q^{a,a\mu}_{u,\omega}))$. For given $k \geq 2$, $E_{Q^a_{u,\omega}}(\sup_{s\leq u\leq t}(1 + ||X_t||)^k))$ is uniformly bounded for $(u,\omega) \in D_\varepsilon$. The result follows then from either the (GC1) condition or the (GC2), and from the uniform BMO hypothesis with similar arguments as in the proof of Proposition 3.
Let ϕ progressive, $\overline{\phi} \in \mathscr{C}_b^{1,0,2}(I\!R_+ \times \Omega \times I\!R^n)$ such that $\phi(t_0,\omega_0) = R(h)^*(t_0,\omega_0)$ and such that (t_0,ω_0) is a maximizer of $R(h)^* - \phi$ on $D_\eta(t_0,\omega_0)$.

- Step 1: Making use of the progressive continuity property of a, f and of the regularity of $\overline{\phi}$, for all $n \in I\!N^*$, there is $\eta_n > 0$, $t_0 + \eta_n < t$ such that for $t_0 \leq u \leq t_0 + \eta_n$, $d(\omega_o,\omega) < \eta_n$ and $||x_0 - \omega(u)|| < \eta_n$,

$$f(u,\omega,(\overline{a}D_x\overline{\phi})(u,\omega,\omega(u))) \leq f(t_0,\omega_0,(\overline{a}D_x\overline{\phi})(t_0,\omega_0,x_0)) + \frac{1}{n} \tag{53}$$

$$\begin{aligned}|\partial_u\overline{\phi}(u,\omega,\omega(u)) + \frac{1}{2}Trace(D^2_x\overline{\phi}\overline{a})(u,\omega,\omega(u)) - \\ \partial_u\overline{\phi}(t_0,\omega_0,x_0) + \frac{1}{2}Trace(D^2_x\overline{\phi}\overline{a})(t_0,\omega_0,x_0))| \leq \frac{1}{n}\end{aligned} \tag{54}$$

  Without loss of generality one can assume $\eta_n < \eta$.
  The matrix valued process a being bounded, it follows from [22] that there is h_n such that for all $t_0 \leq s \leq t_0 + \eta$ and $\omega \in D_\eta(t_0,\omega_0)$,

$$Q^a_{s,\omega}(\{\omega', \sup_{s\le u\le s+h_n} ||\omega'(u)-\omega(s)|| > \frac{\eta_n}{2}\}) < \varepsilon \tag{55}$$

Without loss of generality one can assume that $h_n < \frac{\eta_n}{2}$.
For all $n > 0$ choose $(t_n, \omega_n) \in D_{\frac{\eta_n}{2}}(t_0, \omega_0)$ such that $\lim_{n\to\infty} R(h)(t_n, \omega_n) = R(h)^*(t_0, \omega_0)$ and

$$\phi(t_n, \omega_n) \le R(h)(t_n, \omega_n) + \frac{h_n}{n} \tag{56}$$

Let $C_n = \{\omega', \sup_{t_n\le u\le t_n+h_n} ||\omega'(u)-\omega_n(t_n)|| \le \frac{\eta_n}{2},\ \omega'(v) = \omega_n(v),\ \forall v \le t_n\}$. It follows from Eq. (55) that $Q^a_{t_n,\omega_n}(C_n) > 1-\varepsilon$ for all n. From Lemma 2 there is a constant $C' > 0$ such that $||a\mu||_{BMO(Q^a_{t_n,\omega_n})} \le C'$ for all $\mu \in \tilde{L}(\Lambda)$ and all n. From [15] it follows that there is p_0 such that the reverse Hölder inequality (19) is satisfied for all $\mathscr{E}(a\mu)$ and thus that for all $\mu \in \tilde{L}(\Lambda)$, and all n, $Q^{a,a\mu}_{t_n,\omega_n}(C^c_n) \le K_C(\varepsilon)^{\frac{1}{q_0}}$ where q_0 is the conjugate exponent of p_0. The constants q_0 and K_c depend neither on μ nor on n. Thus ε can be chosen such that $Q^{a,a\mu}_{t_n,\omega_n}(C_n) > \frac{1}{2}$ for all n and all $\mu \in \tilde{L}(\Lambda)$.
Let δ_n be the stopping time taking only two values h_n and 0 defined by $\delta_n = h_n 1_{C_n}$.

- Step 2: Time consistency
  Making use of the time consistency of the risk measure $(\rho^{t_n,\omega_n}_{u,v})$, and of the following equations deduced from Theorem 4,

$$R(h)(t_n, \omega_n) = \rho^{t_n,\omega_n}_{t_n,t}(h)$$

$$\rho^{t_n,\omega_n}_{t_n+\delta_n,t}(h) = R(h)(t_n+\delta_n, .)\ \ Q^a_{t_n,\omega_n}\ a.s.$$

it follows that

$$R(h)(t_n, \omega_n) = \rho^{t_n,\omega_n}_{t_n,t_n+\delta_n}(R(h)(t_n+\delta_n, .))$$

It follows from the definition of the risk measure $(\rho^{t_n,\omega_n}_{t_n,t_n+\delta_n})$, that for all n there is a process μ_n in $\tilde{L}(\Lambda)$ such that

$$R(h)(t_n, \omega_n) \le E_{Q^{a,a\mu_n}_{t_n,\omega_n}}(R(h)(t_n+\delta_n, .)) - \alpha_{t_n,t_n+\delta_n}(Q^{a,a\mu_n}_{t_n,\omega_n}) + \frac{h_n}{n} \tag{57}$$

For all $\omega' \in C_n$, $(t_n+h_n, \omega' *_{t_n+h_n} \omega'(t_n+h_n)) \in D_\eta(t_0, \omega_0)$. The functions $R(h)$ and ϕ are progressive and satisfy $R(h) \le R(h)^* \le \phi$ on $D_\eta(t_0, \omega_0)$. It follows that

$$R(h)(t_n+\delta_n, \omega') \le \phi(t_n+\delta_n, \omega')\ \ \forall \omega' \in C_n \tag{58}$$

From Eqs. (56)–(58), it follows that

$$\phi(t_n, \omega_n) \le E_{Q^{a,a\mu_n}_{t_n,\omega_n}}[\phi(t_n+\delta_n, .) - \int_{t_n}^{t_n+\delta_n} g(u, \omega', \mu_n(u, \omega'))du] + 2\frac{h_n}{n} \tag{59}$$

- Step 3: Martingale problem
Given (t_n, ω_n), the probability measure $Q^{a,a\mu_n}_{t_n,\omega_n}$ is solution to the martingale problem $\mathscr{L}^{a,a\mu_n}$ starting from ω_n at time t_n. The strictly progressive function $\overline{\phi}$ belongs to $\mathscr{C}_b^{1,0,2}$. It follows from [3] and from Eq. (59) that

$$0 \le E_{Q^{a,a\mu_n}_{t_n,\omega_n}}[\int_{t_n}^{t_n+\delta_n} (\partial_u\overline{\phi}(u,\omega',(\omega'(u))) + \frac{1}{2}Trace(D_x^2\overline{\phi}\overline{a})(u,\omega',\omega'(u)))du]$$
$$+ E_{Q^{a,a\mu_n}_{t_n,\omega_n}}[\int_{t_n}^{t_n+\delta_n} (D_x\overline{\phi}^t(u,\omega',\omega'(u))a(u,\omega')\mu_n(u,\omega') - g(u,\omega',\mu_n(u,\omega')))du] + 2\frac{h_n}{n} \tag{60}$$

By definition of f it follows that

$$0 \le E_{Q^{a,a\mu_n}_{t_n,\omega_n}}[\int_{t_n}^{t_n+\delta_n} (\partial_u\overline{\phi}(u,\omega',\omega'(u))) + \frac{1}{2}Trace((D_x^2\overline{\phi}\overline{a})(u,\omega',\omega'(u)))du]$$
$$+ E_{Q^{a,a\mu_n}_{t_n,\omega_n}}[\int_{t_n}^{t_n+\delta_n} f(u,\omega',(\overline{a}D_x\overline{\phi})(u,\omega',\omega'(u)))du] + 2\frac{h_n}{n} \tag{61}$$

- Step 4: Conclusion
Divide equation (61) by h_n and let n tend to ∞. The result follows from step 1, the inequality $Q^{a,a\mu_n}_{t_n,\omega_n}(C_n) \ge \frac{1}{2}$ for all n and $\delta_n = h_n 1_{C_n}$. □

8.3 *Existence of Viscosity Solutions on the Set of Continuous Paths*

On the set of continuous paths $\mathscr{C}(I\!R_+ \times I\!R^n)$ we consider the uniform norm topology. In this section we assume that the function a is only defined on $I\!R_+ \times \mathscr{C}(I\!R_+ \times I\!R^n)$ and that it is continuous. For every continuous function h on the space of continuous paths $\mathscr{C}(I\!R_+ \times I\!R^n)$ such that $h(\omega) = h(\omega')$ if $\omega(u) = \omega'(u)$ for all $u \le t$, the corresponding function $R(h)$ is constructed as above. In this case the function $R(h)$ is defined only on the set of continuous paths (more precisely on $[0,t] \times \mathscr{C}(I\!R_+ \times I\!R^n)$). We make use of the definition of viscosity solution on the set of continuous paths introduced in Sect. 2.4 (Definition 4). To prove that $R(h)$ is a viscosity supersolution, to define $R(h)^*$ and prove that it is a viscosity subsolution, we do not need to extend the functions a nor $R(h)$. Indeed as the support of the probability measure $Q^a_{r,\omega}$ is contained in the set of continuous paths, we just need to use the restrictions of the function $\overline{\phi}$ and of its partial derivatives to the set $I\!R_+ \times \mathscr{C}(I\!R_+ \times I\!R^n) \times I\!R^n$. The proofs given for Theorems 5 and 6 can be easily adapted to prove the analog result in the setting of continuous paths.
Thus the setting of continuous paths can be considered as a "particular case" of the setting of càdlàg paths.

9 Conclusion and Perspectives

We have introduced new notions of regular solutions and viscosity solutions for path-dependent second order PDEs, Eq. (1), in the setting of càdlàg paths. In line with the recent literature on the topic, a solution of (1) must be searched among progressive functions, that is path-dependent functions depending at time t on all the path up to time t. However, the notions of solutions introduced in the present paper differ from previous notions introduced in the literature on two major points:

- In contrast with other papers, we consider on the set Ω of càdlàg paths the Skorokhod topology. Ω is thus a Polish space. This property is fundamental for the construction of solutions for path-dependent PDE that we give in the present paper.
- The notions of partial derivative for progressive functions that we introduce are defined in a very natural way by considering a progressive function of two variables as a function of three variables.

In addition we introduce also a notion of viscosity solution on the set of continuous paths.
Making use of the martingale problem approach to path-dependent diffusion processes, we then construct time consistent dynamic risk measures $\rho_{st}^{r,\omega}$. The stable set of probability measures used for the construction of $\rho_{st}^{r,\omega}$ is a set generated by probability measures solution to the path-dependent martingale problem for $\mathscr{L}^{a,a\mu}$ starting form ω at time r. The path-dependent progressively continuous bounded function a is given and takes values in the set of invertible non negative matrices. The path-dependent functions μ are progressively continuous and vary accordingly to a multivalued mapping Λ. This construction is done in a very general setting. In particular the coefficients μ are not uniformly bounded. We just assume that they satisfy some uniform BMO condition. To construct the penalties, we make use of a path-dependent function g satisfying some polynomial growth condition with respect to the path and some L_2 condition with respect to the process μ related to the BMO condition. In contrast with the usual setting of BSDE, in all this construction no Lipschitz hypothesis are assumed. Notice however that the Lipschitz setting can also be studied within our approach: instead of starting with progressively continuous maps a and μ, one could start from a subfamily of maps which, for example, satisfy some uniform continuity condition (as K Lipschitz maps).
We show that these risk measures provide explicit solutions for semi-linear path-dependent PDEs (2). First, we prove that the risk measures $\rho_{st}^{r,\omega}$ satisfy the following Feller property. For every function h $\mathscr{B}_t$ measurable having some continuity property, there is a progressively lower semi-continuous function $R(h)$ such that $\rho_{rt}^{r,\omega}(h(X_t)) = R(h)(r,\omega)$ and $R(h)(t,\omega) = h(t,\omega)$. Next, the function $R(h)$ is proved to be a viscosity supersolution for a semi-linear path-dependent PDE (2), where the function f itself is associated to g by duality on the multivalued mapping Λ. We prove also that the upper semi continuous envelope of $R(h)$ is a viscosity subsolution for the path-dependent semi linear second order PDE (2).
Here we have proved the progressive lower semi continuity for $R(h)$. To prove the progressive continuity property, additional hypothesis should be added, e.g. Lipschitz

conditions. Another way of proving the continuity is to apply a comparison Theorem. The study of comparison theorems and of continuity properties in this setting, as well as the study of solutions to fully non linear path-dependent PDE will be the subject of future work.

Acknowledgments This work was partly done while at the Center for Advanced Studies at the Norwegian Academy of Science and Letters (CAS, Oslo). It is a pleasure to thank Giulia Di Nunno and Fred Espen Benth for the invitation to contribute to the CAS program in Stochastics in Environmental and Financial Economics (SEFE), and to give a talk at the inaugural SEFE conference.

Appendix

An important way of constructing time consistent dynamic risk measures is to construct a stable set of equivalent probability measures and to define on this set a penalty which is local and satisfies the cocycle condition [5]. Recall the following definitions

Definition 18 A set $\mathcal{Q}$ of equivalent probability measures on a filtered probability space $(\Omega, \mathcal{B}, (\mathcal{B}_t))$ is stable if it satisfies the two following properties:

1. Stability by composition
   For all $s \geq 0$ for all Q and R in $\mathcal{Q}$, there is a probability measure S in $\mathcal{Q}$ such that for all X bounded $\mathcal{B}$-measurable,
   $$E_S(X) = E_Q(E_R(X|\mathcal{B}_s))$$
2. Stability by bifurcation
   For all $s \geq 0$, for all Q and R in $\mathcal{Q}$, for all $A \in \mathcal{B}_s$, there is a probability measure S in $\mathcal{Q}$ such that for all X bounded $\mathcal{B}$-measurable,
   $$E_S(X|\mathcal{F}_s) = 1_A E_Q(X|\mathcal{F}_s) + 1_{A^c} E_R(X|\mathcal{F}_s)$$

Definition 19 A penalty function α defined on a stable set $\mathcal{Q}$ of probability measures all equivalent is a family of maps $(\alpha_{s,t}),\ s \leq t$, defined on $\mathcal{Q}$ with values in the set of $\mathcal{B}_s$-measurable maps such that

(i) α is local:
For all Q, R in $\mathcal{Q}$, for all s, for all A in $\mathcal{B}_s$, the assertion $1_A E_Q(X|\mathcal{B}_s) = 1_A E_R(X|\mathcal{B}_s)$ for all X bounded $\mathcal{B}_t$ measurable implies that $1_A \alpha_{s,t}(Q) = 1_A \alpha_{s,t}(R)$.

(ii) α satisfies the cocycle condition: For all $r \leq s \leq t$, for all Q in $\mathscr{Q}$,

$$\alpha_{r,t}(Q) = \alpha_{r,s}(Q) + E_Q(\alpha_{s,t}(Q)|\mathscr{F}_r)$$

Recall the following result from [5].

Proposition 7 *Given a stable set $\mathscr{Q}$ of probability measures and a penalty $(\alpha_{s,t})$ defined on $\mathscr{Q}$ satisfying the local property and the cocycle condition,*

$$\rho_{st}(X) = \mathrm{esssup}_{\mathrm{Q}\in\mathscr{Q}}(\mathrm{E_Q}(\mathrm{X}|\mathscr{F}_\mathrm{s}) - \alpha_{\mathrm{st}}(\mathrm{Q}))$$

defines a time consistent dynamic risk measure on $L_\infty(\Omega, \mathscr{B}, (\mathscr{B}_t))$ or on $L_p(\Omega, \mathscr{B}, (\mathscr{B}_t))$ if the corresponding integrability conditions are satisfied.

References

1. Aliprantis, C., Border, K.: Infinite Dimensional Analysis, 2nd edn. Springer, Berlin (1999)
2. Billingsley, P.: Convergence of Probability Measures. Wiley series in probability and mathematical statistics. Wiley, New York (1999)
3. Bion-Nadal, J.: Martingale problem for integral-differential operator with path dependent coefficient, preprint (2015)
4. Bion-Nadal, J.: Time consistent convex Feller processes and non linear second order partial differential equations, preprint. arXiv:1207.1742 (2012)
5. Bion-Nadal, J.: Dynamic risk measures: time consistency and risk measures from BMO martingales. Finance Stochast. **12**, 219–244 (2008)
6. Bion-Nadal, J.: Time consistent dynamic risk processes. Stochast. Processes Appl. **119**, 633–654 (2008)
7. Cheredito, P., Delbaen, F., Kupper, M.: Dynamic monetary risk measures for bounded discrete time processes. Electron. J. Probab. **11**, 57–106 (2008)
8. Cont, R., Fournié, D.: Functional Ito calculus and stochastic integral representation of martingales. Ann. Probab. **41**, 109–133 (2013)
9. Delbaen, F.: The structure of m-stable sets and in particular of the set of risk neutral measures. Memoriam Paul-André Meyer, in: Lecture Notes in Mathematics **1874**, 215–258 (2006)
10. Diffusion processes with continuous coefficients II. Commun. Pure Appl. Math. **22**, 479–530 (1969)
11. Diffusion processes with continuous coefficients I. Commun. Pure Appl. Math. **22**, 345–400 (1969)
12. Dupire, B.: Functional Ito calculus, papers.ssrn.com (2009)
13. Ekren I., Touzi N., Zhang J.: Viscosity solutions of fully nonlinear parabolic path dependent PDES. arXiv:1210.0006v3 (2014)
14. Ekren, I., Keller, C., Touzi, N., Zhang, J.: On viscosity solutions of path dependent PDES. Ann. Probab. **42**(1), 204–236 (2014)
15. Kamazaki, N.: Continuous Exponential Martingales and BMO, lecture notes in mathematics 1579, Springer (1994)
16. Krylov, N.: Controlled Diffusion Processes. Springer (1980)
17. Pardoux, E., Peng, S.: Backward stochastic differential equations and quasi linear parabolic partial differential equations, lecture notes in CIS. Springer **176**, 200–217 (1992)
18. Peng, S., Wang, F.: BSDE, path dependent PDE and non linear Feynman Kac formula, preprint. arXiv:1108.4317 (2011)

19. Peng, S.: Backward stochastic differential equations nonlinear expectations and their applications, Proceedings of the International Congress of Mathematicians, Hyderabad, India 2010
20. Peng, S.: Note on viscosity solution of path dependent PDE and G Martingales, preprint. arXiv:1106.1144v2 (2012)
21. Peng, S.: Nonlinear expectations, nonlinear evaluations and risk measures. Lect. Notes Math. **1856**, 165–253 (2004)
22. Strook, D.: Diffusion processes asociated with Levy generators. Z. Wahrscheinlichkeitstheorie verw. Gebiete **32**, 209–244 (1975)

Pricing *CoCos* with a Market Trigger

José Manuel Corcuera and Arturo Valdivia

Abstract Contingent Convertible Bonds, or *CoCos*, are contingent capital instruments which are converted into shares, or may suffer a principal write-down, if certain trigger event occurs. In this paper we discuss some approaches to the problem of pricing *CoCos* when its conversion and the other relevant credit events are triggered by the issuer's share price. We introduce a new model of partial information which aims at enhancing the market trigger approach while remaining analytically tractable. We address also *CoCos* having the additional feature of being callable by the issuer at a series of pre-defined dates. These *callable CoCos* are thus exposed to a new source of risk—referred to as *extension risk*—since they have no fixed maturity, and the repayment of the principal may take place at the issuer's convenience.

Keywords Contingent convertible · Coco bond · Callable bond · Extension risk

1 Introduction

The Basel Committee on banking Supervision was created in 1974, after the collapse of the German Bank Herstatt, with the aim of establishing prudential rules of trading. During the 1980s this committee was concerned with the big moral hazard of Japanese banks that distorted the competition among countries. In 1988 it formulated a set of rules, so called Basel I, to stabilize the international banking system. Basically the main rule was that each bank should hold a minimum of 8 % of its total assets, where, for the valuation of the assets it was used some weights reflecting the credit risk of each asset. These measures produced a credit crunch and some criticism appeared, mainly related with the weights used to measure the risk of the different assets.

J.M. Corcuera (✉) · A. Valdivia
Faculty of Mathematics, University of Barcelona, Gran Via de Les Corts Catalanes, 585, 08007 Barcelona, Spain
e-mail: jmcorcuera@ub.edu

A. Valdivia
e-mail: arturo.valdivia@ub.edu

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_7

These weights only took into account the kind of institution borrowing or issuing the security and not what the spreads observed in the market. To amend the first Basel accords and take into account the market risk and interest-rate risk, it was started a process that concluded in 2004 with new rules, Basel II. These agreements are more complex and consider not only new rules for capitalization, with the introduction of VaR methodology, but also supervision and transparency rules. The regulator calculated the weights on the basis of the formula

$$K = LGD \times \Phi\left[\frac{\Phi^{-1}(PD)}{\sqrt{1-R}} + \sqrt{\frac{R}{1-R}} \times \Phi^{-1}(0.999)\right] - PD \times LGD, \qquad (1)$$

where Φ is the CDF of the standard normal distribution, LGD is the loss in case of default, PD is the probability of default, and R is the correlation between the portfolio of loans and a macroeconomic risk factor, see [24] for an explanation of its underlying model. To determine the different parameters, banks were allowed to use their own models.

In 2007 a financial crisis, originated in the U.S. home loans market, quickly spread to other markets, sectors and countries, forcing the Federal Reserve and the European Central Bank to intervene in response to the collapse of the interbank market. This gave rise, in 2010, to new regulation rules, known as Basel III, that would change the financial landscape. Some securities were not going to be allowed anymore as regulatory capital and supervisors put emphasis in the fact that capital regulatory should have a real loss absorbing capacity. This is when *Contingent Convertibles* (*CoCos*) started to play an important role.

In 2002 Flannery proposed and early form of *CoCo* that he called *Reverse Convertible Debentures*, see [22]. The idea was that whenever the bank issuing such debentures reaches a market-based capital ratio that is below a pre-specified level, a sufficient number of said debentures would convert into shares at the current market price. Later, in [23], he updated the proposal and named these assets as *Contingent Capital Certificates*. The idea behind was in agreement with what [19] wrote in *The Prudential Regulation of Banks.* In this work they formulated the *representation hypothesis*. According to this hypothesis prudential regulation should aim at replicating the corporate governance of non-financial firms, that is, acting as a representative of the debtholders of bank, regulation should play the role of creditors in nonfinancial institutions.

A Contingent Convertible is a bond issued by a financial institution where, upon the appearance of a trigger event, related with a distress of the institution, either an automatic conversion into a predetermined number of shares takes place or a partial write-down of the bond's face value is applied. It is intended to be a loss absorbing security in the sense that in case of liquidity difficulties it produces a recapitalization of the entity.

Basel III, among other regulating measures, proposed the inclusion of *CoCos* as part of Additional Tier 1 Capital, where Tier 1 is, roughly speaking, the capital or the assets that the entity have, for sure, in case of crisis, and consists of Common Equity

Tier 1 and Additional Tier 1. Chan and van Wijnbergen [5] affirm that the inclusion of Cocos in Tier 1 is a likely factor in the increase of *CoCo* issuances. In December 2013, the *CoCo* market had reached \$49 bn in size in Europe.

It is a controversial issue if *CoCos* are a stabilizing security. Koziol and Lawrenz [28] show that, under certain modelling assumptions, if *CoCos* are part of the capital structure of the company equity holders can take more risky strategies, trying to maximize the value of their shares. In their work Koziol and Lawrenz use a low level of the asset price of the company as a trigger for the conversion. Chan and van Wijnbergen [5] point out that conversion can be seen as a negative signal by the depositors of a bank and to produce bank runs. They also argue that far from lowering the risks, *CoCos* can increase even the systemic risks. On the other hand [20] defend that *CoCos* is an appropriate solution that does not lead to moral hazard provided that conversion is tied to exogenous macroeconomic shocks.

There is also disagreement about how to establish the trigger event. It is perhaps the most controversial parameter in a *CoCo*. Some advocate conversion based on book values, like the different capital ratios used in Basel III. Others defend market triggers like the market value of the equity. So far the *CoCos* issued by the private sector are based on accounting ratios.

The market for contingent convertibles started in December 2009 when the Lloyds Banking Group launched its \$13.7 bn issue of Enhanced Capital Notes. Next in line was Rabobank making its first entry in the market for contingent debt with a €1.25 bn issue early 2010. After this, things turned quiet until February 2011, when Credit Suisse launched its so-called Buffer Capital Notes (\$2 bn). This Credit Suisse issue was done on the back of the new regulatory regime in Switzerland. This was called the "Swiss Finish" and it required the larger banks such as UBS and Credit Suisse to hold loss absorbing capital up to 19 % of their risk weighted assets, see [11]. This capital had to consist of at least 10 % common equity and up to 9 % in contingent capital. In 2014 a number of banks issued *CoCos*, including Deutsche Bank and Mizuho Financial Group.

From a modelling point of view and sometimes depending on the trigger chosen for the conversion, usually a low level of a certain index related with the asset, the debt or the equity of the firm, one can follow an intensity approach or a structural approach to model the trigger. For an intensity approach for modelling the conversion time, see for instance [8, 16]. This approach is especially useful when pricing *CoCos* is the main interest, it is a kind of statistical modelling of the trigger event. In fact what one models is the law of the conversion time. In the structural approach for modelling the trigger, one models the random variable describing the conversion time and one relates it with the dynamics of the assets, debt, or equities. It is a more explanatory approach, where one can use the observed dynamics of certain economic facts to describe the conversion time.

In the structural approach one can use a market trigger based on a low level of the equity value. This approach is very appealing because the market value of equity is an observable economic variable whose dynamics can be modelled in order to fit historical data. At the same time it allows to obtain close pricing formulas, like in [13], and to define an objective trigger that can be observed immediately. Cheridito and Xu [7] also use this trigger and show that pricing and hedging problems can be treated for quite general continuous models and barriers and that solutions can always be obtained, at least numerically, using Feymann-Kac type results, translating the problem of pricing into a problem of solving a series of parabolic partial differential equations (PDE) with Dirichlet boundary conditions.

One argument against accounting triggers is that monitoring is not continuous, there is always a delay in the information. Moreover, in the recent crisis these triggers did not provide any signal of distress in troubled banks. On the contrary, when using market triggers, there exist the risk of market manipulations of the equity price trying to force the conversion or undesirable phenomena like the *death-spiral effect*. In [17] authors propose a system of multiple triggers to avoid the death spiral, whereas in [13] a system of coupon cancellations is proposed in order to alleviate this effect. Sundaresan and Wang [39] analyze this kind of trigger and find that to use low equity values as trigger is not innocuous. It can have destabilizing effects in the firm. Their reasoning is roughly speaking the following. Suppose that $(A_t)_{t\geq 0}$ represents the aggregate value of the assets of the company, $(D_t)_{t\geq 0}$ its aggregate debt, $(C_t)_{t\geq 0}$ the aggregate value of the *CoCos*, and $(L_t)_{t\geq 0}$ the aggregate liquidation value of the *CoCos* issued by the firm. Set τ for the conversion time, and assume that it happens when the (aggregate) equity value, $(E_t)_{t\geq 0}$ is lower that some level say $(H_t)_{t\geq 0}$. Since equity value is the residual value of the asset, at any time $t \leq \tau$, we will have two possibilities:

$$A_t - D_t - C_t > H_t \text{ if } t < \tau \tag{2}$$

or

$$A_t - D_t - L_t < H_t \text{ if } t = \tau. \tag{3}$$

This gives that, at any time $t \leq \tau$,

$$D_t + C_t < A_t < D_t + L_t. \tag{4}$$

So, if $C_t > L_t$ there is not any possible value for A_t and if $C_t < L_t$ there are multiple values, all of these allowing, according to [39], potential price manipulation, market uncertainty, inefficient allocation and unreliability of conversion.Obviously if $C_t = L_t$, that is, if there is no jump on the wealth of *CoCo*'s investors at the conversion time, then the equilibrium is possible, but this is considered non realistic and even problematic, since a punitive conversion for shareholders could help to maintain the *market discipline*. As a possible remedy to this situation [32] propose a trigger based not only in the value of the equities but also in the value of the *CoCos*. They also propose to include, in the *CoCo* contract, a PUT option of the issuer on the equities, in case of conversion, to avoid market price manipulations.

Another possibility, in the structural approach, is to use a low value of the asset value as a trigger. It is also quite appealing, since it allows to consider the whole capital structure of the company, to study the effect of *CoCo* debt in the equity value and to obtain the optimal conversion barrier for the shareholders. See for instance

[6, 28] or [2]. In this latter paper they also consider other triggers aiming at providing a proxy for regulatory triggers.

Nevertheless in all these papers authors consider a fixed maturity of the *CoCo* bond. However bonds often do not just have a legal maturity but can have also different *call dates*. In such cases, the bond can be called back by the issuer at these dates prior to the legal maturity. This risk of extending the life of a contract is what we call *extension risk*. This has been treated for the first time in [18] using an intensity approach and in [14] using a structural approach.

In this paper we review our work on the topic of pricing *CoCos*, and we introduce new issues like delay in the information and jumps. We always consider low values of the stock as triggers and *CoCos* that convert totally into equities in case of conversion. The paper is organized as follows. The contract features are specified in Sect. 2. In addition, a model-free formula for the *CoCo*price is presented in order to establish the general pricing problem. In Sect. 3 a model with stochastic interest rates is studied. A closed-form formula for the price is given and subsequently used in order to study the Black-Scholes model and its Greeks. In Sect. 4 two advanced models are discussed. On the one hand, the stochastic volatility Heston model is incorporated to the share price dynamics, and the correspondent prices are later on obtained by a PDE approach. On the other hand, an exponential Lévy model is proposed, and the obtainment of the correspondent prices is addressed by a Fourier method exploiting the so-called Wiener-Hopf factorization for Lévy processes. In Sect. 5 we introduce a new trigger model which aims to describe the delay of information present in accounting triggers. Finally, in Sect. 6 we show how the original pricing problem is modified when no fixed maturity is imposed on the *CoCo*. This variation leads to what we call *CoCos* with extension risk, and the pricing problem includes solving an optimal stopping time problem which, even in the Black-Scholes, turns out to be far from straightforward.

2 The Pricing Problem

The definition of a *CoCo* requires the specification of its *face value* K and *maturity* T, along with the random time τ at which the *CoCo* conversion may occur, and the prefixed price C_p at which the investor may buy the shares if conversion takes place. We refer to τ and C_p as *conversion time* and *conversion price*, respectively. The quantity $C_r := K/C_p$ is refered to as *conversion ratio*. Assuming m coupons are attached to the *CoCo*, then we further need to specify a series of credit events that may trigger a coupon cancellation. Denote by $\tau_1, \ldots, \tau_m$ the random times at which the aforementioned credit events may occur. Then the whole *coupon structure* $(c_j, T_j, \tau_j)_{j=1}^m$ of the *CoCo* is defined, in such a way that the amount c_j is paid at time T_j, provided the $\tau_j > T_j$. We establish that the last coupon is paid at maturity time, *i.e.*, $T_m := T$. The coupon cancellation feature was introduced in [13] in order to alleviate the so-called *death-spiral effect* exhibited by the traditional *CoCo*—see details in Sect. 3.1. Thus it is assumed that coupon cancellation precedes conversion according to $\tau_1 \geq \cdots \geq \tau_m \geq \tau$. Of course it suffices to set $\tau_1 = \cdots = \tau_m = \infty$ if this feature were to be excluded from modelling.

It will be assumed that the issuer of the *CoCo* pays dividends according to a deterministic function κ. From the investors side, it seems to be reasonable to assume that no dividends are paid after the conversion time τ. Thus hereafter we shall assume that the following condition holds true.

Condition (F). There are no dividends after the conversion time τ.

Remark 1 It is worth mentioning that, whereas this condition simplifies the expressions obtained for the price, it is not crucial, in the sense that the computations still can be carried on. We refer to [13] for a further discussion on this topic.

Once these features are settled, the *CoCo*'s final payoff, is given by

$$K\mathbf{1}_{\{\tau>T\}} + \frac{K}{C_p} S_\tau \mathbf{1}_{\{\tau \leq T\}} + \sum_{j=1}^{m} c_j \mathrm{e}^{\int_{T_j}^{T} r_u \mathrm{d}u} \mathbf{1}_{\{\tau_j > T_j\}}, \tag{5}$$

where $(S_t)_{t\geq 0}$ and $(r_t)_{t\geq 0}$ stand for the share price and interest rate, respectively.

2.1 A Model-Free Formula for the CoCo *Price*

In Proposition 2, a model-free formula for the *CoCo* price is given. Let us first introduce some notation required for what follows. Underlying to our market, we shall consider a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$, endowed with a filtration $\mathbb{F} := (\mathscr{F}_t)_{t\in[0,T]}$ representing the trader's information—this includes the information generated by all state variables (*e.g.*, share price, interest rates, total assets value,...) and the default-free market. All filtrations considered are assumed to satisfy the usual conditions of $\mathbb{P}$-completeness and right-continuity. We shall denote the evolution of the money in the bank account by $(B_t)_{t\in[0,T]}$, *i.e.*,

$$B_t = \exp\left\{\int_0^t r_u \mathrm{d}u\right\}, \qquad 0 \leq t \leq T.$$

Recall that two probability measures on $(\Omega, \mathscr{F})$, $\mathbb{P}_1$ and $\mathbb{P}_2$, are said to be equivalent if, for every $A \in \mathscr{F}$, $\mathbb{P}_1(A) = 0$ if and only if $\mathbb{P}_2(A) = 0$. We shall assume the existence of a *risk-neutral* probability measure $\mathbb{P}^*$, equivalent to the *real-world* probability measure $\mathbb{P}$, such that the discounted value of self-financing portfolios, $(\widetilde{V}_t := \frac{V_t}{B_t})_{t\in[0,T]}$, follows a $\mathbb{P}^*$-martingale. Hereafter the symbol tilde will be used to denote discounted prices. Now, in addition to $\mathbb{P}^*$, we shall consider other two probability measures (also equivalent to $\mathbb{P}$) which will allow us to carry on some of the computations related with the *CoCo* arbitrage-free price. First, letting $(B(t, T_j))_{t\geq 0}$ stand for the price of the default-free zero-coupon bond with maturity T_j, we define the T_j*-forward measure* $\mathbb{P}^{T_j}$ through its Radon-Nikodým derivative with respect to $\mathbb{P}^*$ as given by

$$\frac{\mathrm{d}\mathbb{P}^{T_j}}{\mathrm{d}\mathbb{P}^*} = \frac{\mathrm{e}^{-\int_0^{T_j} r_u \mathrm{d}u}}{B(0, T_j)}. \tag{6}$$

We say that $\mathbb{P}^{T_j}$ is given by taking the bond price $(B(t, T_j))_{t\geq 0}$ as *numéraire*. Similarly, but now taking the issuer's share price $(S_t)_{t\geq 0}$—without dividends—as numéraire, we obtain the *share measure* $\mathbb{P}^{(S)}$; its Radon-Nikodým derivative with respect to $\mathbb{P}^*$ is given by

$$\frac{\mathrm{d}\mathbb{P}^{(S)}}{\mathrm{d}\mathbb{P}^*} = \frac{\mathrm{e}^{-\int_0^T [r_u - \kappa(u)]\mathrm{d}u} S_T}{S_0}. \tag{7}$$

In what follows, expectation with respect to $\mathbb{P}^*$, $\mathbb{P}^{T_j}$ and $\mathbb{P}^{(S)}$ will be denoted by $\mathbb{E}^*$, $\mathbb{E}^{T_j}$ and $\mathbb{E}^{(S)}$, respectively.

Proposition 2 *The discounted* CoCo *arbitrage-free price, on the set* $\{t < \tau\}$*, equals*

$$\widetilde{\pi}_t := \mathbb{E}^*\left[\widetilde{K}\mathbf{1}_{\{\tau>T\}}\middle|\,\mathscr{F}_t\right] + C_r\mathbb{E}^*\left[\widetilde{S}_T\mathbf{1}_{\{\tau\leq T\}}\middle|\,\mathscr{F}_t\right] + \sum_{j:\,T_j>t}^{m}\mathbb{E}^*\left[\widetilde{c}_j\mathbf{1}_{\{\tau_j>T_j\}}\middle|\,\mathscr{F}_t\right] \tag{8}$$

$$= K\widetilde{B}(t,T)\mathbb{P}^T(\tau > T|\,\mathscr{F}_t) + \frac{C_r\widetilde{S}_t}{\mathrm{e}^{\int_t^T \kappa(u)\mathrm{d}u}}\mathbb{P}^{(S)}(\tau \leq T|\,\mathscr{F}_t) + \sum_{j:\,T_j>t}^{m} c_j\widetilde{B}(t,T_j)\mathbb{P}^{T_j}(\tau_j > T_j|\,\mathscr{F}_t). \tag{9}$$

Proof Due to the Condition (**F**), to receive $\frac{K}{C_p}S_\tau$ at time τ is equivalent to receive $\frac{K}{C_p}S_T$ at time T. Therefore the payoff in (5) is equivalent to

$$K\mathbf{1}_{\{\tau>T\}} + \frac{K}{C_p}S_T\mathbf{1}_{\{\tau\leq T\}} + \sum_{j=1}^{m} c_j \mathrm{e}^{\int_{T_j}^T r_u \mathrm{d}u}\mathbf{1}_{\{\tau_j>T_j\}}, \tag{10}$$

and thus expression (8) for the price follows by preconditioning, taking into account that $C_r = \frac{K}{C_p}$. As for (9), it suffices to notice that, in light of the abstract Bayes rule, for every $X \in \mathscr{L}^1(\mathscr{F}_{T_j}, \mathbb{P}^{T_j})$ we have

$$\mathbb{E}^{T_j}[X|\,\mathscr{F}_t] = \frac{\mathbb{E}^*\left[X\frac{\mathrm{d}\mathbb{P}^{T_j}}{\mathrm{d}\mathbb{P}^*}\middle|\,\mathscr{F}_t\right]}{\mathbb{E}^*\left[\frac{\mathrm{d}\mathbb{P}^{T_j}}{\mathrm{d}\mathbb{P}^*}\middle|\,\mathscr{F}_t\right]} = \frac{\mathbb{E}^*\left[X\mathrm{e}^{-\int_0^{T_j} r_u\mathrm{d}u}\middle|\,\mathscr{F}_t\right]}{\mathrm{e}^{-\int_0^t r_u\mathrm{d}u}B(t,T_j)}, \tag{11}$$

and thus

$$\mathbb{E}^*\left[X\mathrm{e}^{-\int_t^{T_j} r_u \mathrm{d}u}\middle|\,\mathscr{F}_t\right] = B(t,T_j)\mathbb{E}^{T_j}[X|\,\mathscr{F}_t]. \tag{12}$$

Similarly, for every $X \in \mathscr{L}^1(\mathscr{F}_T, \mathbb{P}^{(S)})$ we have

$$\mathbb{E}^*\left[XS_T\mathrm{e}^{-\int_t^{T} r_u \mathrm{d}u}\middle|\,\mathscr{F}_t\right] = \mathrm{e}^{-\int_t^T \kappa(u)\mathrm{d}u} S_t\mathbb{E}^{(S)}[X|\,\mathscr{F}_t]. \tag{13}$$

Combining (8) with these identities we get the expression (9).

With Proposition 2 at hand, the subsequent difference between models relies on how conversion and coupon cancellation is defined, and how the corresponding prices are evaluated. In this paper we follow a structural approach to price *CoCos*. That is to say, given a model for the share price $(S_t)_{t\geq 0}$, a series of critical time-varying barriers ℓ and ℓ^j are set in such a way that $\ell^1 \geq \cdots \geq \ell^m \geq \ell$ and the credit events are given by

$$\tau := \inf\{t > 0 : S_t \leq \ell(t)\}, \qquad \text{and} \qquad \tau_j := \inf\{t > 0 : S_t \leq \ell^j(t)\}, \quad j = 1, ..., m, \tag{14}$$

with the standard convention $\inf \emptyset := \infty$. Then our main concern is to derive analytically tractable formulas for (8)–(9), either in the form of closed formulas or efficient simulation methods. Later on in Sects. 5 and 6 we shall incorporate short-term uncertainty and extension risk to the pricing problem.

2.2 Pricing CoCos *with Write-Down*

In the case of *CoCos* with write-down, upon the appearance of the trigger event the investor does not receive a certain amount of shares. Instead, she receives only a fraction $R \in (0, 1)$ of the original face value K, provided that the issuer has not defaulted. Let δ denote the random time at which the issuer may default. Then the payoff of this *CoCo* contract with write-down equals

$$K\mathbf{1}_{\{\tau>T\}} + RK\mathbf{1}_{\{\tau\leq T\}}\mathbf{1}_{\{\delta>T\}} + \sum_{j=1}^m c_j \mathrm{e}^{\int_{T_j}^T r_u \mathrm{d}u}\mathbf{1}_{\{\tau_j>T_j\}}.$$

Similarly to Proposition 2, we can give now a model-free price formula for the *CoCo* with write-down. For this matter, we make no further assumption beyond *model consistency* in the sense that δ and τ are modelled in such a way that $\delta > \tau$ so that default may only occur after conversion.

Proposition 3 *The discounted arbitrage-free price of the* CoCo *with write-down, on the set* $\{t < \tau\}$, *equals*

$$\widetilde{\pi}_t^{wd} := (1-R)K\widetilde{B}(t,T)\mathbb{P}^T(\tau > T|\, \mathscr{F}_t) + RK\widetilde{B}(t,T)\mathbb{P}^T(\delta > T|\, \mathscr{F}_t) + \sum_{j:\, T_j > t}^{m} c_j \widetilde{B}(t,T_j)\mathbb{P}^{T_j}(\tau_j > T_j|\, \mathscr{F}_t). \tag{15}$$

Proof It suffices to notice that the payoff can be rewritten as

$$\begin{aligned} R\mathbf{1}_{\{\delta>s\}}\mathbf{1}_{\{\tau\le s\}} + \mathbf{1}_{\{\tau>s\}} &= R\mathbf{1}_{\{\delta>s\}}\mathbf{1}_{\{\tau\le s\}} + \left(R\mathbf{1}_{\{\tau>s\}} + (1-R)\mathbf{1}_{\{\tau>s\}}\right) \\ &= R\mathbf{1}_{\{\delta>s\}}\mathbf{1}_{\{\tau\le s\}} + R\mathbf{1}_{\{\tau>s\}}\mathbf{1}_{\{\delta>s\}} + (1-R)\mathbf{1}_{\{\tau>s\}} \\ &= R\mathbf{1}_{\{\delta>s\}}\left(\mathbf{1}_{\{\tau\le s\}} + \mathbf{1}_{\{\tau>s\}}\right) + (1-R)\mathbf{1}_{\{\tau>s\}}, \end{aligned}$$

where for the second equivalence we have used the identity $\mathbf{1}_{\{\tau>s\}}\mathbf{1}_{\{\delta>s\}} = \mathbf{1}_{\{\tau>s\}}$, which holds due to the consitency assumption $\tau < \delta$.

By comparing (9) with (15) we can see that the techniques used to price the *CoCo* with conversion can be readily applied to *CoCo* with write-down. Across this work we focus on the former contract.

3 A Model with Stochastic Interest Rates

In this section we assume that the price of default-free zero-coupon bonds are stochastic. More specifically, for $j \in \{1, ..., m\}$, the default-free zero-coupon bond price $(B(t,T_j))_{t\in[0,T_j]}$ is assumed to have the following $\mathbb{P}^*$-dynamics

$$\frac{\mathrm{d}B(t,T_j)}{B(t,T_j)} = r_t \mathrm{d}t + \sum_{k=1}^{d} b_k(t,T_j)\mathrm{d}W_t^k, \tag{16}$$

where each b_k is a positive deterministic càdlàg function, and $(W_t^1, ..., W_t^d)_{t\in[0,T]}$ is a d-dimensional Brownian motion with respect to the risk-neutral probability measure $\mathbb{P}^*$ and the trader's filtration $\mathbb{F}$. We shall assume as well that the share price $(S_t)_{t\in[0,T]}$ obeys

$$\frac{\mathrm{d}S_t}{S_t} = [r_t - \kappa(t)]\mathrm{d}t + \sum_{k=1}^{d} \sigma_k(t)\mathrm{d}W_t^k, \tag{17}$$

where $\sigma := (\sigma_k)_1^d$ is a positive deterministic càdlàg function such that, for all $t \in [0, T_j]$, the inequality $\left\|\sigma(t) - b(t,T_j)\right\| > 0$ is satisfied.

The conversion and coupon cancellation events in our model are linked to the asset dividends and the evolution of bond prices, in such a way that conversion is triggered as soon as $(S_t)_{t\geq 0}$ crosses

$$\ell_t := L_m B(t, T_m) \exp\left\{\int_t^{T_m} \kappa(u)\mathrm{d}u\right\}, \qquad 0 \leq t \leq T_m;$$

and similarly, for $j \in \{1, ..., m\}$, the j-th coupon cancellation is triggered as soon as $(S_t)_{t\geq 0}$ crosses

$$\ell_t^j := \begin{cases} L_j B(t, T_j) \exp\left\{\int_t^{T_j} \kappa(u)\mathrm{d}u\right\}, & 0 \leq t < T_j \\ M_{j,} & t = T_j. \end{cases} \tag{18}$$

The parameters M_j and L_j are assumed to be given non-negative constants satisfying $M_j \geq L_j$, with $L_m < C_p$ and

$$\frac{L_{j+1}}{L_j} \leq \exp\left\{-\int_{T_j}^{T_{j+1}} \kappa(u)\mathrm{d}u\right\}, \qquad j = 1, ..., m-1, \tag{19}$$

so that the required ordering $\ell_t^1 \geq \ell_t^2 \geq \cdots \geq \ell_t^m$ is fulfilled—thus ensuring that $0 \leq \tau_j \leq T_j$ implies $\tau_j < \tau_{j+1}$, for $j = 1, ..., m-1$. Clearly the *Mertonian condition* (*i.e.*, S_{T_j} must be bigger than M_j) can be removed by taking $L_j = M_j$. See Fig. 1 for an illustration of the barrier's shape and parameters.

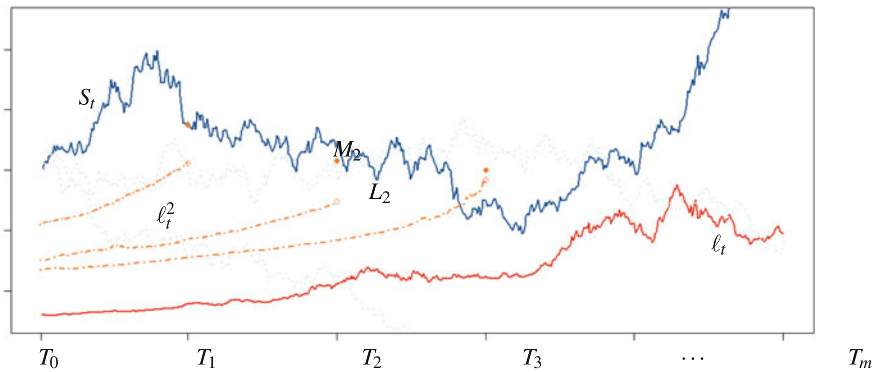

Fig. 1 The graph illustrates the share price $(S_t)_{t\geq 0}$, along with the barriers ℓ^j and its parameters L_j and M_j. The first barrier is hit at $t = T_1$, whereas the third one is hit at some $T_2 < t < T_3$. On the other hand, the second barrier is not hit since the share price stays above ℓ^2 on the whole period $[T_0, T_2]$, and the Mertonian condition is satisfied, *i.e.*, $S_{T_2} > M_2$. Conversion is not triggered either since the barrier ℓ is never hit by $(S_t)_{t\geq 0}$

In the current setting, the process $(U_t^j := \log \frac{S_t}{\ell^j(t)})_{t\geq 0}$ plays a fundamental role. Indeed, from the definition of the random times τ and τ_j (see (14)) and the barriers ℓ and ℓ^j it follows that

$$\{\tau > T_m\} = \left\{\inf_{0\leq t\leq T_m} U_t^m > 0\right\} \quad \text{and} \quad \{\tau_j > T_j\}$$
$$= \left\{\inf_{0\leq t\leq T_j} U_t^j > 0,\ U_{T_j}^j > \log\frac{M_j}{L_j}\right\}, \quad j = 1, ..., m.$$

From this observation we have that, in order to price the *CoCo* contract, we need to be able to compute the conditional joint distribution of $(\underline{U}_{T_j}^j, U_{T_j}^j)$, where we have defined $\underline{U}_{T_j}^j := \inf_{0\leq t\leq T_j} U_t^j$. To this matter, notice that an application of the Itô formula tells us that

$$\mathrm{d}U_t^j = -\frac{1}{2}\left\|\sigma(t) - b(t, T_j)\right\|^2 \mathrm{d}t + \left\|\sigma(t) - b(t, T_j)\right\| \mathrm{d}W_t^{T_j}.$$

where $(W_t^{T_j})_{t\geq 0}$ is the $\mathbb{P}^{T_j}$-Brownian motion given by the Girsanov theorem, corresponding to the probability change (6). In fact we can see that under $\mathbb{P}^{(S)}$ we have similar dynamics

$$\mathrm{d}U_t^j = \frac{1}{2}\left\|\sigma(t) - b(t, T_j)\right\|^2 \mathrm{d}t + \left\|\sigma(t) - b(t, T_j)\right\| \mathrm{d}W_t^{(S)},$$

where $(W_t^{(S)})_{t\geq 0}$ is the $\mathbb{P}^{(S)}$-Brownian motion corresponding now to the probability change (7). Consequently, a time change given by

$$a_j(t) := \int_0^t \left\|\sigma(s) - b(s, T_j)\right\|^2 \mathrm{d}s, \qquad 0 \leq t \leq T_j, \tag{20}$$

renders the fundamental process $(U_t^j := \log \frac{S_t}{\ell^j(t)})_{t\geq 0}$ a drifted Brownian motion. Then we can apply a known result (see [34]) on joint distribution of the Brownian motion and its running infimum in order to obtain the following closed-formula for the *CoCo* price. See details in [13].

Proposition 4 *In the current setting, the* CoCo *arbitrage-free price, on the set* $\{t < \tau_m\}$*, is given by*

$$\pi_t = \sum_{j:\, T_j > t}^{m} \mathbf{1}_{\{t<\tau_j\}} c_j \left(B(t, T_j)\Phi(-d_+^j - D_j) - \frac{S_t \mathrm{e}^{-\int_t^{T_j} \kappa(u)\mathrm{d}u}}{L_j} \Phi(d_-^j - D_j) \right) \tag{21}$$
$$+ KB(t, T_m) + \left(C_r - \frac{K}{L_m}\right) \left(L_m B(t, T_m)\Phi(d_+^m) + S_t \mathrm{e}^{-\int_t^{T_m} \kappa(u)\mathrm{d}u} \Phi(d_-^m)\right),$$

where

$$d_{\pm}^{j} = \frac{\log \frac{L_j B(t,T_j) \mathrm{e}^{\int_t^{T_j} \kappa(u)\mathrm{d}u}}{S_t} \pm \frac{1}{2}\int_t^{T_j} \left\|\sigma(u) - b(u,T_j)\right\|^2 \mathrm{d}u}{\sqrt{\int_t^{T_j} \left\|\sigma(u) - b(u,T_j)\right\|^2 \mathrm{d}u}}, \quad and$$

$$D_j = \frac{1}{\sqrt{\int_t^{T_j} \left\|\sigma(u) - b(u,T_j)\right\|^2 \mathrm{d}u}} \log \frac{M_j}{L_j}, \tag{22}$$

and Φ denotes the standard Gaussian cumulative distribution function.

3.1 The Black-Scholes Model and the Greeks

The Black-Scholes model is obtained as a particular case of the model above, by taking default-free bond prices with null volatilities—*i.e.*, by taking the $b(\cdot, T_j)$ in (16) to be zero. Consequently, the closed-form price formula given by Proposition 4 can be used in order to derive the *Greeks*, *Delta* Δ and *Vega* ν, which respectively describe the *CoCo*'s price sensitivity to share price and volatility. We have the following.

Proposition 5 *Let Φ and ϕ denote, respectively, the standard Gaussian cumulative distribution and density function. In the Black-Scholes model, the* CoCo*'s price sensitivity to share price $\Delta := \frac{\partial \pi_t}{\partial S_t}$ and to volatility $\nu := \frac{\partial \pi_t}{\partial \sigma}$ are respectively given by*

$$\Delta = \sum_{j:\, T_j > t}^{m} \mathbf{1}_{\{t<\tau_j\}} \frac{c_j}{L_j} \left(\frac{2}{\sigma\sqrt{T_j - t}} \phi(b_-^j) - \Phi(b_-^j) \right) + \left(\frac{K}{L_m} - C_r \right) \left(\frac{2}{\sigma\sqrt{T_m - t}} \phi(b_-^m) - \Phi(b_-^m) \right),$$

and

$$\nu = \sum_{j:\, T_j > t}^{m} \mathbf{1}_{\{t<\tau_j\}} \frac{c_j}{L_j} \frac{\log \frac{L_j}{S_t} - r(T_j - t)}{\sigma^2 \sqrt{T_j - t}} \left(\phi(-b_+^j) L_j \mathrm{e}^{-r(T_j - t)} + S_t \phi(b_-^j) \right) - \left(C_r - \frac{K}{L_m} \right) \frac{\log \frac{L_m}{S_t} - r(T_m - t)}{\sigma^2 \sqrt{T_j - t}} \left(\phi(b_+^m) L_m \mathrm{e}^{-r(T_m - t)} + S_t \phi(b_-^m) \right),$$

where

$$b_{\pm}^{j} = \frac{\log \frac{L_j}{S_t} - (r \mp \frac{1}{2}\sigma^2)(T_j - t)}{\sigma\sqrt{T_j - t}}, \qquad j = 1, ..., m. \tag{23}$$

Some remarks are in order. On the one hand, it is has been documented that by actively hedging the equity risk, investors can unintentionally force the conversion by making the share price deteriorate and eventually trigger the conversion. This situation is referred to as the *death-spiral effect*. Now, from the above expression for the Delta Δ it can be checked that it is strictly positive, and in fact one can observe that the Delta Δ increases sharply when the time to maturity T decreases. Thus one of the conclusions in [13] is that the coupon cancellation feature leads to a flatter behaviour of the Delta Δ, hence reducing the death-spiral risk. On the other hand, it can also be checked that the Vega ν is strictly negative, this tell us that an increase in the volatility translates into a decrease in the prices. Such observation is clearly in line with the intuition that a higher volatility will increase the probability of crossing the barriers ℓ and ℓ^j defining the conversion and coupon cancellation events.

4 Advanced Models

4.1 Incorporating the Heston Stochastic Volatility Model

Let us start this section by remarking the fact that the arguments preceding the obtainment of Proposition 4 hold even if the share and default-free bond price volatilities (*i.e.*, σ and $b(\cdot, T_j)$ in (17) and (16)) are no longer deterministic. However, the time change a_j in (20) would be now stochastic and, consequently, the time-changed fundamental process dynamics

$$\mathrm{d}U^j_{a_j(t)} = -\frac{1}{2}a_j(t)\mathrm{d}t + \mathrm{d}W^{T_j}_{a_j(t)}, \tag{24}$$

would no longer match those of a drifted Brownian motion. Thus one anticipates that the pricing problem, in the setting of stochastic volatility, will lead to closed-form formulas only in few cases, and require more advanced numerical tools otherwise.

As a particular model, we shall assume that the volatilities are stochastic according to the work of [26]: we consider a new stochastic factor $(V_t)_{t\geq 0}$ acting on both σ and $b(\cdot, T_j)$, in such a way that the dynamics in (17) and (16) are now replaced by

$$\frac{\mathrm{d}S_t}{S_t} = [r_t - \kappa(t)]\mathrm{d}t + \sqrt{V_t}\sum_{k=1}^{d}\sigma_k(t)\mathrm{d}W^k_t, \tag{25}$$

and

$$\frac{\mathrm{d}B(t, T_j)}{B(t, T_j)} = r_t\mathrm{d}t + \sqrt{V_t}\sum_{k=1}^{d}b_k(t, T_k)\mathrm{d}W^k_t, \tag{26}$$

respectively. The factor $(V_t)_{t\geq 0}$ is given as the solution to the following SDE

$$\mathrm{d}V_t = [\alpha - \beta V_t]\mathrm{d}t + \gamma\sqrt{V_t}\mathrm{d}Z_t,$$

where α, β and γ are constants, with $2\alpha > \gamma^2$ to ensure the positivity of the solution, and $(Z_t)_{t\geq 0}$ is a one-dimensional $\mathbb{P}^{T_j}$-Brownian motion. Similar dynamics under the share measure $\mathbb{P}^{(S)}$ are assumed to be satisfied by $(V_t)_{t\geq 0}$.

If we further assume the independence between the noises driving the prices and their volatilities, then we see that the $\mathbb{P}^{T_j}$-Brownian motion in (24) is independent of the (now stochastic) time change

$$a_j(t) = \int_0^t V_s \left\|\sigma(s) - b(s, T_j)\right\|^2 \mathrm{d}s, \qquad 0 \leq t \leq T_j. \tag{27}$$

Thus, by a preconditioning argument, we obtain the following extension of Proposition 4.

Proposition 6 *In the current setting, the* CoCo *arbitrage-free price, on the set* $\{t < \tau_m\}$, *is given by*

$$\begin{aligned}\pi_t = &\sum_{j,T_j>t}^{m} \mathbf{1}_{\{t<\tau_j\}} c_j \left(B(t,T_j)\mathbb{E}^{T_j}[\Phi(-d_+^j - D_j)|\mathscr{F}_t] - \frac{S_t\mathbb{E}^{T_j}[\Phi(d_-^j - D_j)|\mathscr{F}_t]}{L_j \mathrm{e}^{\int_t^{T_j}\kappa(u)\mathrm{d}u}} \right)\\ &+ KB(t,T_m)\mathbb{E}^{T_m}[\Phi(-d_-^m)|\,\mathscr{F}_t] - \frac{KS_t}{L_j\mathrm{e}^{\int_t^{T_j}\kappa(u)\mathrm{d}u}}\mathbb{E}^{T_m}[\Phi(d_+^m)|\,\mathscr{F}_t]\\ &+ C_r S_t \mathrm{e}^{-\int_t^{T_m}\kappa(u)\mathrm{d}u}\mathbb{E}^{(S)}[\Phi(d_-^m)|\,\mathscr{F}_t] + C_r L_m B(t,T_m)\mathbb{E}^{(S)}[\Phi(d_+^m)|\,\mathscr{F}_t],\end{aligned}$$

where

$$d_\pm^j = \frac{\log \frac{L_j B(t,T_j)\mathrm{e}^{\int_t^{T_j}\kappa(s)\mathrm{d}s}}{S_t} \pm \frac{1}{2}\int_t^{T_m} V_s\left\|\sigma(s) - b(s,T_j)\right\|^2\mathrm{d}s}{\sqrt{\int_t^{T_j} V_s\left\|\sigma(s) - b(s,T_j)\right\|^2\mathrm{d}s}}, \tag{28}$$

and

$$D_j = \frac{1}{\sqrt{\int_t^{T_j} V_s\left\|\sigma(s) - b(s,T_j)\right\|^2\mathrm{d}s}} \log\frac{M_j}{L_j}. \tag{29}$$

From the pricing formula above we can see that the *CoCo* price is related to the price of binary options. Indeed, for instance, for a binary option with maturity T_j and strike M_j we have

$$\mathbb{E}^*\left[\mathrm{e}^{-\int_t^{T_j} r_u \mathrm{d}u}\mathbf{1}_{\{S_{T_j}<M_j\}}\middle|\mathscr{F}_t\right] = B(t,T_j)\mathbb{E}^{T_j}\left[\mathbf{1}_{\{S_{T_j}<M_j\}}\middle|\mathscr{F}_t\right]$$
$$= B(t,T_j)\mathbb{P}^{T_j}\left(U^j_{T_j} \le \log\frac{M_j}{L_j}\middle|\mathscr{F}_t\right).$$

Moreover, according to our previous discussion, conditioned to $\mathscr{F}^V_{T_j} := \sigma(V_s,\ 0 \le s \le T_j)$ the random variable $U^j_{T_j}$ is normally distributed, so that we also have

$$\mathbb{P}^{T_j}\left(U^j_{T_j} \le \log\frac{M_j}{L_j}\middle|\mathscr{F}_t\right) = \mathbb{E}^{T_j}[\Phi(d^j_+ + D_j)|\mathscr{F}_t].$$

Let us now give an explicit computation of the probability above; this illustrates how the *CoCo* price can be computed. For this matter, we use the relationship between the characteristic and distribution functions (see for instance [38]) which allows us to write

$$\mathbb{P}^{T_j}\left(U^j_{T_j} \le \log\frac{M_j}{L_j}\middle|\mathscr{F}_t\right) = \frac{1}{2}+\frac{1}{2\pi}$$
$$\int_0^\infty \frac{\left(\frac{M_j}{L_j}\right)^{\mathrm{i}\xi}\varphi^{T_j}(t,U_t,V_t;-\xi) - \left(\frac{M_j}{L_j}\right)^{-\mathrm{i}\xi}\varphi^{T_j}(t,U_t,V_t;\xi)}{\mathrm{i}\xi}\mathrm{d}\xi,$$

where

$$\varphi^{T_j}(t,U_t,V_t;\xi) := \mathbb{E}^{T_j}[\exp\{i\xi(U^j_{T_j})\}|\mathscr{F}_t] = \mathbb{E}^{T_j}[\exp\{i\xi(U^j_{T_j})\}|U_t,V_t], \tag{30}$$

and the last equation holds by Markovianity. Hence the problem of computing the probability above translates into the problem of finding an expression for $\varphi^{T_j}(t,u,v;\xi)$. It follows from the Itô formula that

$$\frac{\partial\varphi^{T_j}}{\partial t} + \frac{1}{2}\frac{\partial^2\varphi^{T_j}}{\partial u^2}v\sigma^2_{T_j} + \frac{1}{2}\frac{\partial^2\varphi^{T_j}}{\partial v^2}\gamma^2 v - \frac{1}{2}\frac{\partial\varphi^{T_j}}{\partial u}v\sigma^2_{T_j} + \frac{\partial\varphi^{T_j}}{\partial v}(\alpha-\beta v) = 0, \tag{31}$$

with the boundary condition $\varphi^{T_j}(T_j,u,v;\xi) = \mathrm{e}^{\mathrm{i}\xi u}$, and where $\sigma^2_{T_j} = \|\sigma - b(\cdot,T_j)\|^2$. For an affine solution like

$$\varphi^{T_j}(t,u,v;\xi) = \mathrm{e}^{A_j(T_j-t)+B_j(T_j-t)v+\mathrm{i}\xi u}, \tag{32}$$

the PDE in (31) is reduced to the Riccati equation

$$\begin{cases} \frac{\partial B_j}{\partial t} - \frac{1}{2}\gamma^2 B_j^2 + \beta B_j = -\left(\frac{1}{2}\xi^2 - \frac{\mathrm{i}}{2}\xi\right)\sigma^2_{T_j}, \\ \frac{\partial A_j}{\partial t} = \alpha B_j, \end{cases} \tag{33}$$

with A_j and B_j vanishing at $t = T_j$. As shown in [13], for a constant function $\sigma^2_{T_j}$, such equation is explicitly solved by

$$B_j(T_j - t) = -\frac{\lambda(\xi) + \beta}{\gamma^2} \frac{\exp\left\{-\lambda(T_j - t)\right\} - 1}{\exp\left\{-\lambda(T_j - t)\right\} + \frac{\lambda(\xi)+\beta}{\lambda(\xi)-\beta}},$$

and

$$A_j(t) = -\alpha \tfrac{\lambda(\xi)-\beta}{\lambda(\xi)+\beta} \left[\tfrac{2}{\gamma^2} \log\left((\lambda(\xi) - \beta) \exp\left\{-\lambda(\xi)(T_j - t)\right\} + \tfrac{\lambda(\xi)+\beta}{\lambda(\xi)-\beta} \right) + \tfrac{\lambda(\xi)-\beta}{\gamma^2} (T_j - t) \right],$$

where $\lambda(\xi) := \sqrt{\beta^2 + \gamma^2 \sigma^2_{T_j} (\xi^2 - i\xi)}$, and $\sqrt{\cdot}$ denote the analytic extension of the real square root to $\mathbb{C} \setminus \mathbb{R}_-$.

4.2 An Exponential Lévy Model

In this section we shall consider an exponential *Lévy model* for the share price. As opposed to the previous sections, we shall now consider a numerical approach to pricing, based on exploiting the so-called *Wiener-Hopf factorization* of the driving Lévy process $(X_t)_{t \geq 0}$. This approach has been recently applied in order to price contracts with path-dependent payoffs as in [12, 31]; see more details below.

4.2.1 First-Passage Times and Wiener-Hopf Factorization

Let $(X_t)_{t\geq 0}$ be a Lévy process with characteristic triplet (μ, σ, ν), and denote its characteristic exponent by ψ_X. For details and proofs of the following arguments we refer to [1].

Recall that if $e(\lambda)$ is an exponential random variable with parameter λ, independent of $(X_t)_{t\geq 0}$, then we have the following equality in distribution

$$X_{e(\lambda)} = I + S,$$

where I and S are independent random variables, distributed as

$$\underline{X}_{e(\lambda)} := \inf_{0 \leq u \leq e(\lambda)} X_u \qquad \text{and} \qquad \overline{X}_{e(\lambda)} := \sup_{0 \leq u \leq e(\lambda)} X_u,$$

respectively. Moreover,

$$\mathbb{E}\left[\exp\left\{zX_{e(\lambda)}\right\}\right] = \mathbb{E}\left[\exp\left\{z\underline{X}_{e(\lambda)}\right\}\right] \mathbb{E}\left[\exp\left\{z\overline{X}_{e(\lambda)}\right\}\right]. \tag{34}$$

We shall refer to (34) as the *Wiener-Hopf factorization*. In fact, in general it holds that

$$\mathbb{E}\left[\exp\left\{zX_{e(\lambda)}\right\}\right] = \frac{\lambda}{\lambda - \psi_X(z)}.$$

Consequently, the knowledge of one of the factors in (34) allows us to establish the other one. A particular case of interest arises when $(X_t)_{t\geq 0}$ is a spectrally negative process since, in this case, it is known that the right factor in (34) is given by

$$\psi_\lambda^+(z) := \mathbb{E}\left[\exp\left\{z\overline{X}_{e(\lambda)}\right\}\right] = \frac{\beta_\lambda}{\beta_\lambda - z}, \tag{35}$$

where β_λ is a constant, depending on λ, defined as the solution to

$$\psi_X(\beta) = \lambda. \tag{36}$$

Therefore, once we have computed β_λ explicitly, we obtain the following expression for the left factor in (34):

$$\psi_\lambda^-(z) := \mathbb{E}\left[\exp\left\{z\underline{X}_{e(\lambda)}\right\}\right] = \frac{\lambda}{\lambda - \psi_X(z)}\frac{\beta_\lambda - z}{\beta_\lambda}. \tag{37}$$

This expression can be linked to the distribution function of $\underline{X}_t$ by partial integration. Indeed we have

$$\begin{aligned}
\psi_\lambda^-(z) &= \int_0^\infty \lambda \mathrm{e}^{-\lambda t}\mathrm{d}t \int_{-\infty}^0 z\mathrm{e}^{z\xi}\mathbb{P}(\underline{X}_t > \xi)\mathrm{d}\xi \\
&= \lambda z \int_0^\infty \int_0^\infty \mathrm{e}^{-\lambda t - z\xi} F(t,\xi)\mathrm{d}t\mathrm{d}\xi \\
&= \lambda z \widetilde{F}(\lambda, z)
\end{aligned} \tag{38}$$

where we have defined $F(t,\xi) := \mathbb{P}(\underline{X}_t > -\xi)$, and denoted its Laplace transform by $\widetilde{F}$. As argued by [33], by combining (37) and (38) we can recover F by the standard Fourier transform inversion. Further, the result can be numerically computed in an efficient way, provided the condition

$$\lim_{z\to\infty} \frac{1}{z} \int_{-\infty}^0 \left[e^{zx} - 1 - zx\mathbf{1}_{\{|x|\leq 1\}}\right] \nu(\mathrm{d}x) = 0, \tag{39}$$

holds true. This condition is imposed in order to facilitate the computation of β_λ (*i.e.*, the solution of (35)) by means of suitable integration contour change. In fact, such a change allows to take $\beta_\lambda = \lambda/\mu$. This result is summarized in the following lemma.

Lemma 7 *For fixed t and ξ, and given the parameter set $(A_1, A_2, l_1, l_2, N_1, N_2)$, define*

$$a_1 := \frac{A_1}{2tl_1}, \quad a_2 := \frac{A_2}{2\xi l_2}, \quad h_1 := \frac{\pi}{tl_1}, \quad h_2 := \frac{\pi}{\xi l_2}, \quad g(x) := \psi_X(x/\mu)$$

and, for every $N \in \mathbb{N}$,

$$s_N(t,\xi) := \frac{h_1 h_2}{4\pi^2} \sum_{n=-N}^{N} \sum_{k=-N}^{N} \frac{\mathrm{d}g}{\mathrm{d}x}(a_1 + \mathrm{i}nh_1) \\ \widetilde{F}(g(a_1 + \mathrm{i}nh_1), a_2 + \mathrm{i}kh_2)\mathrm{e}^{tg(a_1 t + \mathrm{i}nh_1) + x(a_2 + \mathrm{i}kh_2)}.$$

If the condition (39) is satisfied, then the following approximation holds true

$$\mathbb{P}(\underline{X}_t > \xi) \doteqdot \sum_{n=0}^{N_2} \binom{N_2}{n} s_{N_1+n}(t,\xi),$$

where the symbol $\doteqdot$ indicates an Euler summation.

The double sum in the lemma is used as an initial approximation of F. Here the parameters (A_1, A_2, l_1, l_2) are positive real numbers chosen large enough in order to control the aliasing error. The final Euler summation is used in order to improve the accuracy of the raw approximation s_N. It is suggested that choosing $A_1 = A_2 = 22$, $l_1 = l_2 = 1$, $N_1 = 12$ and $N_2 = 15$ gives satisfactory results. For further details see [33] and references therein.

4.2.2 The One-Sided CGMY Lévy Process

Hereafter we shall focus on a particular spectrally negative Lévy process $(X_t)_{t\geq 0}$ known as *one-sided CGMY* process, or simply *CMY* process. This process has no continuous part, and only one-sided jumps with its Lévy measure being given by

$$\nu(\mathrm{d}x) = C\exp\{-Mx\}|x|^{-1-Y}\mathbf{1}_{\{x<0\}}\mathrm{d}x, \tag{40}$$

where $C, M > 0$ and $Y < 1$ are constants. Further, its characteristic exponent can be obtained in closed-form as

$$\begin{aligned}\psi_X(z) &= \mu z + \int_{-\infty}^{0} \left[e^{zx} - 1 - zx\mathbf{1}_{\{|x|\leq 1\}}\right] \nu(\mathrm{d}x) \\ &= \mu z + C\Gamma(-Y)((M+z)^Y - M^Y),\end{aligned} \tag{41}$$

where Γ stands for the Gamma function. Thus it is apparent that the condition (39) holds true. Before we proceed, let us mention that the CGMY processes are also referred to as *Tempered stable* processes. On the other hand, by setting the parameter $Y = 0$ (resp., $Y = 1/2$) the CMY process becomes a *Gamma* process (resp., *Inverse Gaussian* process). For details on CMY processes we refer to [4, 31] and [36, Section 2.3.5].

In order to price *CoCos* we need to understand the behavior of $(X_t)_{t\geq 0}$ both under the risk-neutral measure $\mathbb{P}^*$ and the share measure $\mathbb{P}^{(S)}$. The following result shows how the Lévy characteristics of $(X_t)_{t\geq 0}$ change under Esscher transforms.

Lemma 8 *For every real number θ, consider the probability measure $\mathbb{P}^\alpha$, equivalent to $\mathbb{P}$, given by*

$$\frac{\mathrm{d}\mathbb{P}^\alpha}{\mathrm{d}\mathbb{P}} = \frac{\exp\{\alpha X_T\}}{\mathbb{E}\left[\exp\{\alpha X_T\}\right]}.$$

Assume $M_\alpha := M - \alpha > 0$. Then the Lévy exponent of $(X_t)_{t\geq 0}$ under $\mathbb{P}^\alpha$ is given by

$$\psi_X^\alpha(z) := z\mu_\alpha + \int_{-\infty}^{0} \left[e^{zx} - 1 - zx\mathbf{1}_{\{|x|\leq 1\}}\right] \nu_\alpha(\mathrm{d}x),$$

where

$$\mu_\alpha := \mu + \int_{|x|\leq 1} x(e^{\alpha x} - 1)\nu(\mathrm{d}x), \qquad \textit{and} \qquad \nu_\alpha(\mathrm{d}x) := C\exp\{-M_\alpha x\}|x|^{-1-Y}\mathbf{1}_{\{x<0\}}\mathrm{d}x.$$

Proof The first part follows from (41) and [35, Theorems 33.1 and 33.2]. Now, using the expression in (40) we have

$$\nu_\alpha(\mathrm{d}x) = C\exp\{-(M-\alpha)x\}|x|^{-1-Y}\mathbf{1}_{\{x<0\}}\mathrm{d}x = C\exp\{-M_\alpha x\}|x|^{-1-Y}\mathbf{1}_{\{x<0\}}\mathrm{d}x.$$

The assumption $M_\alpha > 0$ assures that (C, M_α, Y) is a rightful parameter set for a CMY distribution.

4.2.3 Application to *CoCos*

We shall assume that, under $\mathbb{P}^*$, there is a pure-jump (C, M, Y)-Lévy process $(X_t)_{t\geq 0}$ driving the share price $(S_t)_{t\geq 0}$ in such a way that

$$S_t := \mathrm{e}^{(r-\kappa)t}\frac{\exp\{X_t\}}{\mathbb{E}^*\left[\exp\{X_t\}\right]} = \exp\{\mu t + X_t\}, \qquad t \geq 0,$$

where the interest rate r and the dividends κ are assumed to be constants, and we define $\varpi := -\log\mathbb{E}^*[\mathrm{e}^{X_1}]$ and set $\mu := r - \kappa + \varpi$. Further, we shall assume that the parameter M is bigger than 1. We remark here that, on the one hand, this technical assumption will allow us to accommodate the *Variance Gamma* (VG) process considered in [30]. On the other hand, this assumption is consistent with the numerical experiments reported by [31].

Proposition 9 *In the current setting, the price of a* CoCo *at time* $0 \le t \le T$ *can be numerically approximated in an efficient way by means of the expression*

$$\pi_t \approx \sum_{j:\, T_j > t}^{m} \mathrm{e}^{-r(T_j - t)} P_j^0(t, T_j) + K\mathrm{e}^{-r(T_m - t)} P_m^0(t, T_m) + C_r S_t \mathrm{e}^{-\kappa(T_m - t)} P_m^1(t, T_m),$$

where

$$P_j^\alpha(t, T_j) \doteqdot \sum_{n=0}^{N_2} \binom{N_2}{n} s_{N_1+n}^\alpha \left(T_j - t, \log \frac{S_t}{\ell^j(t)} + \varpi(T_j - t)\right), \qquad j = 1, \ldots, m, \;\; \alpha = 0, 1, \tag{42}$$

with the symbol $\doteqdot$ *indicating an Euler summation, and*

$$s_N^\alpha(t, \xi) := \sum_{n=-N}^{N} \sum_{k=-N}^{N} \frac{\mu_\alpha + YC\Gamma(-Y)\left(M_\alpha + \mu_\alpha^{-1} t\right)^{Y-1}}{4\mu_\alpha t \xi l_1 l_2} \widetilde{F}_\alpha(g_\alpha(a_1 + \mathrm{i}nh_1), a_2 + \mathrm{i}kh_2)\mathrm{e}^{tg_\alpha(a_1 t + \mathrm{i}nh_1) + x(a_2 + \mathrm{i}kh_2)},$$

with the parameters $(a_1, a_2, h_1, h_2, l_1, l_2, N_1, N_2)$ *given as in Lemma 7, and*

$$\widetilde{F}_\alpha(\lambda, z) := \frac{\lambda - z\mu_\alpha}{\left(\lambda - z\mu_\alpha - C\Gamma(-Y)((M_\alpha + z)^Y - M_\alpha^Y)\right) z\lambda},$$

$$g_\alpha(x) := x + C\Gamma(-Y)\left(\left(M_\alpha + \mu_\alpha^{-1} x\right)^Y - M_\alpha^Y\right), \qquad \alpha = 0, 1,$$

where M_α *and* μ_α *are defined in Lemma 8.*

Proof Taking into account the general expression for the *CoCo* price (*c.f.* Proposition 2), computing π_t boils down to compute

$$P_j^0(t, T_j) := \mathbb{P}^*\left(\tau_j > T_j \middle| \mathscr{F}_t\right) = \mathbb{P}^*(\underline{\mathrm{X}}_{T_j - t} > -\xi)\Big|_{\xi = \log(S_t/\ell^j(t)) + \varpi(T_j - t)},$$

for $j = 1, \ldots, m$, and

$$P_m^1(t, T_m) := \mathbb{P}^{(S)}\left(\tau_m > T_m | \mathscr{F}_t\right) = \mathbb{P}^{(S)}(\underline{\mathrm{X}}_{T_m - t} > -\xi)\Big|_{\xi = \log(S_t/\ell^m(t)) + \varpi(T_m - t)}.$$

These computations can be carried out by means of Lemma 7. Indeed, under this exponential Lévy model, the share measure $\mathbb{P}^{(S)}$ (resp., risk-neutral measure $\mathbb{P}^*$)

coincides with an Esscher transform of parameter $\alpha = 1$ (resp., $\alpha = 0$). In light of Lemma 8, the driving noise $(X_t)_{t\geq 0}$ will remain a CMY process under $\mathbb{P}^{(S)}$, but now having the shifted parameter set $(C, M_1, Y) = (C, M-1, Y)$. Moreover, this implies that the Lévy measure of $(X_t)_{t\geq 0}$ also satisfies the condition (39) under $\mathbb{P}^{(S)}$. Thus, if we write $\mathbb{P}^0 = \mathbb{P}^*$ and $\mathbb{P}^1 = \mathbb{P}^{(S)}$, by the reasoning in Sect. 4.2.1 we see that the Laplace transform of $F_\alpha(t,\xi) := \mathbb{P}^\alpha(\underline{X}_t > -\xi)$ is given by $\widetilde{F}_\alpha$ as defined above. Moreover, the correspondent contour change is given by g_α.

Remark 10 [12] provides an alternative approach which exploits the Wiener-Hopf factorization in a different way: instead of computing first-time passage probabilities as done here, what is computed is the joint density of $(X_t, \underline{X}_t)$. As the noise driving share prices, the authors consider the so-called *Beta-Variance Gamma* (β-VG) process—also referred to as *Lamperti-Stable* process by [3]—which exhibits the same exponential decay as the Variance Gamma process, hence leading to a smile-conform model. For this β-VG process the distribution of the variables $\overline{X}_{e(\lambda)}$ and $\underline{X}_{e(\lambda)}$ can be specified, thus obtaining the Wiener-Hopf factors ψ_λ^+ and ψ_λ^-. Taking (8) into account, combining the knowledge of the $(X_t, \underline{X}_t)$ density with a Monte-Carlo technique due to [29], the authors provide an efficient numerical pricing of CoCos.

5 Triggering Conversion Under Short-Term Uncertainty

Linking credit events to the movements of a *fully observable* (*i.e.*, $\mathbb{F}$-adapted) process $(U_t)_{t\geq 0}$ is certainly one of the most appealing features of structural models. Indeed, this *full observability* assumption—hereafter referred to as (**A1**)—gives rise to clear and analytically tractable models as we have seen in the previous sections. When considering contingent capital contracts such as *CoCos*, however, the assumption (**A1**) seems arguable since in most cases regulatory capital depends on the balance sheets of the issuer, and those sheets are updated only at a series of predetermined dates $(t_j)_{t\in\mathbb{N}}$. Thus we are interested in considering the following *partial observability* assumption.

Assumption (**A1′**). The fundamental process $(U_t)_{t\geq 0}$ is fully observable only at predetermined dates $(t_j)_{t\in\mathbb{N}}$.

On the other hand, when the process $(U_t)_{t\geq 0}$ is related to the share price, it is also commonly assumed that the correlation between the noises driving the share price and $(U_t)_{t\geq 0}$ is equal to $\rho = 1$ (or $\rho = -1$)—hereafter this assumption will be referred to as (**A2**). Nevertheless, it would be reasonable to consider the chance that a different (possibly time-dependent) correlation parameter $\rho \in [-1, 1]$ may provide a better fit. Consequently we are also interested in considering the correlation ρ as an additional rightful model parameter by taking the following alternative to (**A2**).

Assumption (**A2′**). The correlation ρ between the noises driving the share price and $(U_t)_{t\geq 0}$ may vary.

In what follows we revisit the simple framework of Sect. 3.1 in order to illustrate these ideas and show how the pricing problem is modified under (**A1′**) and (**A2′**). For a full study of this *short-term uncertainty* model we refer to the forthcoming paper [15].

5.1 Pricing CoCos *on a Black-Scholes Model Under Short-term Uncertainty*

As shown in Sect. 3.1, in the Black-Scholes model,

$$\mathrm{d}S_t = S_t([r-\kappa]\mathrm{d}t + \sigma \mathrm{d}W_t^*),$$

the cancellation of the j-th coupon is triggered as soon as the process

$$\mathrm{d}U_t := \mathrm{d}\log\frac{S_t}{\ell_t} = -\frac{1}{2}\sigma^2\mathrm{d}t + \sigma\mathrm{d}W_t^*,$$

crosses the critical value $\log\frac{M_j}{L_j}$, $j = 1, ..., m$, whereas for conversion zero is the critical level. In this setting Assumption (**A2′**) is translated as the correlation structure between the noise driving $(S_t)_{t\geq 0}$ and that of the new process

$$\mathrm{d}U_t(\rho) := -\frac{1}{2}\sigma^2\mathrm{d}t + \sigma\mathrm{d}W_t^\rho := -\frac{1}{2}\sigma^2\mathrm{d}t + \sigma\mathrm{d}(\rho W_t^* + \sqrt{1-\rho^2}Z_t), \qquad (43)$$

where ρ is the given correlation parameter, and $(Z_t)_{t\geq 0}$ is a second Brownian motion, independent of $(W_t^*)_{t\geq 0}$. Thus, instead of the process $(U_t)_{t\geq 0}$ above, we shall now consider the parametric family $(U_t(\rho))_{t\geq 0}$ whose driving noise $(W_t^\rho)_{t\geq 0}$ is also a Brownian motion but correlated to $(W_t^*)_{t\geq 0}$, in such a way that $\mathrm{d}W_t^\rho\mathrm{d}W_t^* = \rho\mathrm{d}t$. Further, the time at which the j-th coupon may be cancelled is given by

$$\tau_j(\rho) := \inf\left\{t \geq 0 \ : \ U_t(\rho) \leq \log\frac{M_j}{L_j}\right\}.$$

As for Assumption (**A1′**), notice that the full information flow corresponds to

$$\mathscr{G}_t := \sigma(W_s^*, Z_s, \ 0 \leq s \leq t) = \mathscr{F}_t^{W^*} \vee \mathscr{F}_t^Z, \qquad t \geq 0,$$

whereas, setting $\lfloor t \rfloor := \min\{t_j \in \{0, t_1, t_2, ...\} \ : \ t_j \leq t < t_{j+1}\}$, the information available to the modeller is now given by

$$\widetilde{\mathscr{F}}_t := \mathscr{F}_t^{W^*} \vee \sigma(Z_s, \ 0 \leq s \leq \lfloor t \rfloor) = \mathscr{F}_t^{W^*} \vee \mathscr{F}_{\lfloor t \rfloor}^Z, \qquad t \geq 0.$$

Since $\mathscr{G}_t \supseteq \widetilde{\mathscr{F}}_t$, and the equality holds only at the predetermined dates $\{t_j,\ j \in \mathbb{N}\}$, we can think of $(Z_t)_{t\geq 0}$ as an extra source of noise which clears out at update times $(t_j)_{j\in\mathbb{N}}$. The fact that the extra noise is cleared out at $(t_j)_{j\in\mathbb{N}}$ motivates the notion of *short-term uncertainty*, and it has two important implications. On the one hand, our model differs from other partial or incomplete information models like [9, 10, 21], or [25] since the information structure is different. On the other hand, as opposed to other structural models, the short-term uncertainty considered here prevents the conversion time $\tau_j(\rho)$ from being a stopping time with respect to the reference filtration $\mathbb{F}$, which is generated by the relevant state variables and the risk-free market. Hence, one can investigate conditions under which $\tau_j(\rho)$ admits an intensity, as done by [9, 21], or [27].

5.2 Coupon Cancellation Probabilities Under Short-Term Uncertainty

Let us show how the coupon cancellation probabilities are modified under the assumptions (**A1**$'$) and (**A2**$'$). We begin by defining two auxiliary processes

$$\zeta_t := \sigma\sqrt{1-\rho^2}(Z_t - Z_{\lfloor t\rfloor}) \qquad \text{and} \qquad \xi_t := \rho \log \frac{S_t}{S_{\lfloor t\rfloor}} + \rho \log \frac{\ell_t}{\ell_{\lfloor t\rfloor}}, \qquad t \geq 0.$$

These processes have an important role in the computations within our short-term uncertainty model since they appear implicitly in $(U_t(\rho))_{t\geq 0}$ according to the factorization

$$U_t(\rho) = \big(U_{\lfloor t\rfloor}(\rho) + \xi_t\big) + \zeta_t. \tag{44}$$

It is apparent that the term between parentheses belongs to $\widetilde{\mathscr{F}}_t = \mathscr{F}_t^{W^*} \vee \mathscr{F}_{\lfloor t\rfloor}^{Z}$. On the other hand, ζ_t is independent of $\widetilde{\mathscr{F}}_t$, and it is normally distributed with zero mean and variance

$$\nu^2(t) := (1-\rho^2)(t - \lfloor t\rfloor)\sigma^2.$$

We note here that the variance $\nu^2(t)$ represents a key quantity within this framework. Indeed, on the one hand, it actually encodes the two new features of our model: the factor $1-\rho^2$ measures how close $(S_t)_{t\geq 0}$ and $(U_t)_{t\geq 0}$ are to being completely correlated; whereas the factor $t - \lfloor t\rfloor$ measures the elapsed time from the last information update. On the other hand, as the following result suggests, coupon cancellation probabilities and other analytical formulas obtained within our model depend explicitly on $\nu(t)$.

Proposition 11 *For every $x \in \mathbb{R}$ define the random time $\tau_x(\rho) := \inf\{s \geq 0 \;:\; U_s(\rho) \leq x\}$. Then, for every $0 \leq t \leq T$, the following equation holds true on $\{\tau > \lfloor t \rfloor\}$*

$$\mathbb{P}^*(\tau_x(\rho) > T | \widetilde{\mathscr{F}}_t) = \mathbb{E}^*\left[\Phi\left(-D_- + \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right] - \mathrm{e}^{-(U_{\lfloor t \rfloor}(\rho)-x)}\left(\frac{S_{\lfloor t \rfloor}\ell_{\lfloor t \rfloor}}{S_t \ell_t}\right)$$
$$\mathbb{E}^*\left[\mathrm{e}^{-\zeta_t}\Phi\left(D_+ - \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right],$$

where the expectations above are restricted to the values

$$D_\pm = \frac{x - U_{\lfloor t \rfloor}(\rho) - \xi_t \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}. \tag{45}$$

Moreover,

$$\mathbb{E}^*\left[\Phi\left(-D_- + \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right] = \int_{\mathbb{R}} \Phi\left(-D_- + \frac{z\nu(t)}{\sigma\sqrt{T-t}}\right)\phi(z)\mathrm{d}z \tag{46}$$

and

$$\mathbb{E}^*\left[\mathrm{e}^{-\zeta_t}\Phi\left(D_+ - \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right] = \mathrm{e}^{\frac{1}{2}\nu^2(t)}\int_{\mathbb{R}} \Phi\left(D_+ - \frac{z\nu(t) - \nu^2(t)}{\sigma\sqrt{T-t}}\right)\phi(z)\mathrm{d}z, \tag{47}$$

where Φ and ϕ stand, respectively, for the standard Gaussian cumulative distribution and density functions.

Proof Since under $\mathscr{G}_t$ the computation is known, for every $t = t_j$ we have, on $\{\tau_x(\rho) > t\}$,

$$\begin{aligned}\mathbb{P}^*(\tau_x(\rho) > T | \widetilde{\mathscr{F}}_{t_j}) &= \mathbb{P}^*(\tau_x(\rho) > T | \mathscr{G}_{t_j}) \\ &= \Phi\left(\frac{-x + U_{t_j}(\rho) + \frac{1}{2}\sigma^2(T-t_j)}{\sigma\sqrt{T-t_j}}\right) - \mathrm{e}^{-(U_{t_j}(\rho)-x)}\Phi \\ &\quad \left(\frac{x - U_{t_j}(\rho) + \frac{1}{2}\sigma^2(T-t_j)}{\sigma\sqrt{T-t_j}}\right).\end{aligned}$$

Define

$$d^1_\pm(t) = \frac{x - U_t(\rho) \pm \frac{1}{2}\sigma^2(T-t)}{\sigma\sqrt{T-t}}.$$

For $t \notin \{t_j,\ j \in \mathbb{N}\}$, by conditioning we get, on $\{\tau_x(\rho) > \lfloor t \rfloor\}$,

$$\begin{aligned}\mathbb{P}^*(\tau_x(\rho) > T|\widetilde{\mathscr{F}}_t) = \mathbb{E}^*\left[\mathbb{P}^*(\tau_x(\rho) > T|\mathscr{G}_t)\middle|\,\widetilde{\mathscr{F}}_t\right] &= \mathbb{E}^*\left[\Phi\left(-d_-^1(t)\right)\middle|\,\widetilde{\mathscr{F}}_t\right] \\ &- \mathbb{E}^*\left[\mathrm{e}^{-(U_t(\rho)-x)}\Phi\left(d_+^1(t)\right)\middle|\,\widetilde{\mathscr{F}}_t\right].\end{aligned}$$

In these terms, for $\lfloor t \rfloor < t$, the first summand above satisfies

$$\mathbb{E}^*\left[\Phi\left(-d_-(t)\right)\middle|\,\widetilde{\mathscr{F}}_t\right] = \mathbb{E}^*\left[\Phi\left(-D_- + \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right],$$

where the right-hand side of the expectation above is restricted to the value of D_- given in (45); in fact the equation above reduces to (46) since $\zeta_t \sim N(0, \nu^2(t))$ for every fixed $t \geq 0$. Similarly for the second summand, it follows from (44) that

$$\begin{aligned}\mathbb{E}^*\left[\mathrm{e}^{-(U_t(\rho)-x)}\Phi\left(d_+^1\right)\middle|\,\widetilde{\mathscr{F}}_t\right] &= \mathrm{e}^{-(U_{\lfloor t \rfloor}(\rho)-x+\xi_t)}\mathbb{E}^*\left[\mathrm{e}^{-\zeta_t}\Phi\left(D_+ - \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right] \\ &= \mathrm{e}^{-(U_{\lfloor t \rfloor}(\rho)-x)}\left(\frac{S_{\lfloor t \rfloor}\ell_{\lfloor t \rfloor}}{S_t\ell_t}\right)\mathbb{E}^*\left[\mathrm{e}^{-\zeta_t}\Phi\left(D_+ - \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right].\end{aligned}$$

In order to obtain (47), let us consider the change of measure given by

$$\begin{aligned}\frac{\mathrm{d}\mathbb{P}'}{\mathrm{d}\mathbb{P}^*} = \exp\left\{\zeta_t - \frac{1}{2}\nu^2(t)\right\} = \exp\bigg\{&\int_0^T \sigma\sqrt{1-\rho^2}\mathbf{1}_{[\lfloor t \rfloor, t]}(s)\mathrm{d}Z_s \\ &-\frac{1}{2}\int_0^T [\sigma\sqrt{1-\rho^2}\mathbf{1}_{[\lfloor t \rfloor, t]}(s)]^2\mathrm{d}s\bigg\}.\end{aligned}$$

In virtue of the Girsanov theorem, the process

$$Z_s' := Z_s - \int_0^s \sigma\sqrt{1-\rho^2}\mathbf{1}_{[\lfloor t \rfloor, t]}(u)\mathrm{d}u, \qquad s \geq 0,$$

follows a $\mathbb{P}'$-Brownian motion. Thus

$$\begin{aligned}\mathbb{E}^*\left[\mathrm{e}^{\zeta_t - \frac{1}{2}\nu^2(t)}\Phi\left(D_+ - \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right] &= \mathbb{E}_{\mathbb{P}'}\left[\Phi\left(D_+ - \frac{\zeta_t}{\sigma\sqrt{T-t}}\right)\right] \\ &= \mathbb{E}_{\mathbb{P}'}\left[\Phi\left(D_+ - \frac{\int_0^T \sigma\sqrt{1-\rho^2}\mathbf{1}_{[\lfloor t \rfloor, t]}(s)\mathrm{d}Z_s' + \nu^2(t)}{\sigma\sqrt{T-t}}\right)\right] \\ &= \int_{\mathbb{R}} \Phi\left(D_+ - \frac{x+\nu^2(t)}{\sigma\sqrt{T-t}}\right)\frac{\exp\left\{-\frac{x^2}{2\nu^2(t)}\right\}}{\sqrt{2\pi\nu^2(t)}}\mathrm{d}x \\ &= \int_{\mathbb{R}} \Phi\left(D_+ - \frac{\nu(t)z+\nu^2(t)}{\sigma\sqrt{T-t}}\right)\phi(z)\mathrm{d}z,\end{aligned}$$

where the first equivalence follows from the abstract Bayes' rule, and for the last equivalence we have simply used the standard change of variables $z = \frac{x}{\nu(t)}$.

Jeanblanc and Valchev [27] study the role of information on defaultable-bond prices within a Black-Scholes setting, with constant parameters and flat default barrier. In some sense, the short-term uncertainty model considered here can be seen as a bivariate extension of [27]. Despite this analogy, immediate differences arise. For instance, the survival probabilities obtained by [27] are constant between observation dates, *i.e.*, within each interval $[t_j, t_{j+1})$. Whereas in Proposition 11 these probabilities vary in continuous time. This difference relies on the fact that, event though within each interval $[t_j, t_{j+1})$ our knowledge of the short-term noise ζ_t is constant, we still fully observe the evolution of $(S_t)_{t\geq 0}$ and all the other $\mathbb{F}$-adapted state variables.

In order to conclude this section, let us recall that in light of our discussion in Sect. 3.1, *CoCo* prices can be obtained in our current setting once we compute expressions of the form

$$\mathbb{P}^*(\tau_j(\rho) > T_j, S_{T_j} > L_j|\widetilde{\mathscr{F}}_t), \qquad \text{and} \qquad \mathbb{P}^{(S)}(\tau_m(\rho) > T_m, S_{T_m} > L_m|\widetilde{\mathscr{F}}_t).$$

It is worth noticing that the $\widetilde{\mathscr{F}}_t$-conditional joint distribution of $(\tau_j(\rho), S_{T_j}) = (\inf_{s\leq T_j} U_s(\rho), S_{T_j})$ cannot be computed directly from Proposition 11 since the entries of this vector are driven by two different (though correlated) Brownian motions. An additional complication comes from the fact that the current information about one of them might be incomplete. The aforementioned distribution, and full details on the model, can be found in [15].

6 Extension Risk

According to the new regulatory Basel III framework, *CoCos* can be categorised as either belonging to the Additional Tier 1 or Tier 2 capital category. In order to belong to the former class, a *CoCo* is supposed to have the coupon cancellation feature and, further, no fixed maturity is to be imposed to the contract. Instead, the issuer is entitled to redeem the *CoCo* at any of the prespecified *call times* $\{T_i,\ i \in \mathbb{N}\}$. Moreover, as opposed to the common practice on callable contracts before the 2008 financial crisis, the definition of this contract does not contain any incentive (*e.g.*, a coupon step-up) for the issuer to redeem at the first call date. Investing in such a contract has the inherent risk of a financial loss due to the lengthening of the (investor's) expected maturity duration which ultimately postpones the payment of the face value K. This risk is referred to as *extension risk*. Two recent papers [14, 18] have addressed the problem of pricing *CoCos* belonging to the Additional Tier 1 capital category. As an illustration, let us revisit the Black-Scholes model in Sect. 3.1.

In order to emphasize the correspondence with call times, we add now an extra index $i \in \mathbb{N}$ to the coupon structure, in such a way that a coupon c_{ij} will be paid at T_{ij} provided $\tau_{ij} > T_{ij}$. It will be assumed that for every $i \in \mathbb{N}$ the ordering $T_{i-1} < T_{i1}, \ldots \leq T_{im} := T_i$ holds, where we set $T_0 := 0$. For the sake of clarity, let us remark that in the current setting, the barriers in (18) and their parameters become

$$\ell^{ij}(t) := \begin{cases} L_{ij}\mathrm{e}^{-(r-\kappa)(T_{ij}-t)}, & 0 \leq t < T_{ij} \\ M_{ij}, & t = T_{ij}. \end{cases}$$

From the issuer's point of view, the question of whether to postpone or not the face value K payment depends on which alternative is cheaper. Hence, similarly to the situation of Bermuda options (see for instance [37]), the discounted price of a *CoCo* belonging to Additional Tier 1 capital category equals

$$\widetilde{\Pi}_t := \begin{cases} \inf_{\theta \in \mathscr{T}_n} \mathbb{E}^*\left[\widetilde{Z}^{(n)}_\theta \middle| \mathscr{F}_{T_n}\right], & t = T_n \in \{T_i, \in \mathbb{N}\} & \text{(a)} \\ \mathbb{E}^*\left[\widetilde{\pi}_{T_{n+1}} \middle| \mathscr{F}_t\right], & t \in (T_n, T_{n+1}), & \text{(b)} \end{cases} \tag{48}$$

where $\mathscr{T}_n$ stands for the set of stopping times taking values in $\{T_i,\ i \geq n\}$ and

$$\widetilde{Z}^{(n)}_\theta = \sum_{i=n+1}^{l:T_l=\theta} \sum_{j=1}^{m} \widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij}, S_{T_{ij}}>M_{ij}\}} + K\mathrm{e}^{-r\theta}\mathbf{1}_{\{\tau>\theta\}} + \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\leq\theta\}}.$$

It is important to remark that even though the general optimal stopping theory allows us to characterize the solution to the optimization problem in (48a), this is not enough to tackle whole pricing problem. Indeed, the solution to (48a) must be obtained in a relatively explicit way in order to be able to give a reasonable expression for the price in-between call dates (48b). Here we shall address the *finite horizon* case, *i.e.*, the case where there are only finitely many call dates $\{T_1, \ldots, T_N\}$ and $T_N < \infty$.

In this case, for every fixed T_n, the solution to the optimization problem in (7.6.1) is related to the *lower Snell envelope* $(\widetilde{Y}^{(n)}_k)_{k\in\{n,\ldots,N\}}$ of the process $(\widetilde{X}^{(n)}_k)_{k\in\{n,\ldots,N\}}$ given by

$$\widetilde{X}^{(n)}_k := \widetilde{Z}^{(n)}_{T_k} = K\mathrm{e}^{-rT_k}\mathbf{1}_{\{\tau>T_k\}} + \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\leq T_k\}} + \sum_{i=n+1}^{k}\sum_{j=1}^{m}\widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij}, S_{T_{ij}}>M_{ij}\}}.$$

Having a finite horizon allows us to obtain $(\widetilde{Y}^{(n)}_k)_{k\in\{n,\ldots,N\}}$ by means of the following backwards procedure

$$\widetilde{Y}^{(n)}_k = \begin{cases} \widetilde{X}^{(n)}_N, & k = N \\ \min\left\{\widetilde{X}^{(n)}_k, \mathbb{E}^*\left[\widetilde{Y}^{(n)}_{k+1} \middle| \mathscr{F}_{T_k}\right]\right\}, & k = N-1, \ldots, n. \end{cases} \tag{49}$$

As it turns out, the lower Snell envelope $(\widetilde{Y}_k^{(n)})_{k\in\{n,\dots,N\}}$ can be obtained in a rather explicit form. Indeed, for the first iteration of (49), if $\tau > T_n$ then what we have is the raw expression

$$\begin{aligned}
\widetilde{Y}_{N-1}^{(n)} &= \min\left\{\widetilde{X}_{N-1}^{(n)}, \mathbb{E}^*\left[\widetilde{X}_N^{(n)}\middle|\mathscr{F}_{T_{N-1}}\right]\right\}\\
&= \min\Bigg\{K\mathrm{e}^{-rT_{N-1}}\mathbf{1}_{\{\tau>T_{N-1}\}} + \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\le T_{N-1}\}} + \sum_{i=n+1}^{N-1}\sum_{j=1}^{m}\widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij}\ ,\ S_{T_{ij}}>M_{ij}\}},\\
&\qquad \mathbb{E}^*\left[K\mathrm{e}^{-rT_N}\mathbf{1}_{\{\tau>T_N\}} + \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\le T_N\}} + \sum_{i=n+1}^{N}\sum_{j=1}^{m}\widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij}\ ,\ S_{T_{ij}}>M_{ij}\}}\middle|\mathscr{F}_{T_{N-1}}\right]\Bigg\}\\
&= \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\le T_{N-1}\}} + \sum_{i=n+1}^{N-1}\sum_{j=1}^{m}\widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij},S_{T_{ij}}>M_{ij}\}}\\
&\quad + \mathbf{1}_{\{\tau>T_{N-1}\}}\min\Bigg\{K\mathrm{e}^{-rT_{N-1}},\\
&\qquad \mathbb{E}^*\left[K\mathrm{e}^{-rT_N}\mathbf{1}_{\{\tau>T_N\}} + \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\le T_N\}} + \sum_{j=1}^{m}\widetilde{c}_{Nj}\mathbf{1}_{\{\tau_{Nj}>T_{Nj},S_{T_{Nj}}>M_{Nj}\}}\middle|\mathscr{F}_{T_{N-1}}\right]\Bigg\}\\
&= \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\le T_{N-1}\}} + \sum_{i=n+1}^{N-1}\sum_{j=1}^{m}\widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij},S_{T_{ij}}>M_{ij}\}} + \mathbf{1}_{\{\tau>T_{N-1}\}}\min\left\{K\mathrm{e}^{-rT_{N-1}}, \widetilde{\pi}_{T_{N-1}}\right\},
\end{aligned}\tag{50}$$

where as in the previous section, $\widetilde{\pi}_{T_{N-1}}$ denotes the price of *CoCo* here with maturity T_N and coupon structure $(c_{Nj}, T_{Nj}, \tau_{Nj})_{j=1}^m$. Due to the share price Markovianity, the price $\widetilde{\pi}_{T_{N-1}}$ can be seen as function of $S_{T_{N-1}}$; we denote this function simply by $\widetilde{\pi}_{T_{N-1}}(x)$. Now, as discussed in Sect. 3.1, the *CoCo* has a positive Delta, thus the function $\widetilde{\pi}_{T_{N-1}}(x)$ is increasing and we can find a value S_{N-1}^* such that

$$S_{N-1}^* := \inf\left\{x > 0\ :\ \widetilde{\pi}_N(x) \ge K\mathrm{e}^{-rT_{N-1}}\right\}.$$

Hence we obtain

$$\begin{aligned}
\widetilde{Y}_{N-1}^{(n)} &= \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\le T_{N-1}\}} + \sum_{i=n+1}^{N-1}\sum_{j=1}^{m}\widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij},S_{T_{ij}}>M_{ij}\}}\\
&\quad + K\mathrm{e}^{-rT_{N-1}}\mathbf{1}_{\{\tau>T_{N-1},S_{T_{N-1}}\ge S_{N-1}^*\}} + \widetilde{\pi}_{T_{N-1}}\mathbf{1}_{\{\tau>T_{N-1},S_{T_{N-1}}<S_{N-1}^*\}}\\
&= K\mathrm{e}^{-rT_{N-1}}\mathbf{1}_{\{\tau>T_{N-1},S_{T_{N-1}}\ge S_{N-1}^*\}} + \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{\tau\le T_{N-1}\}} + \sum_{i=n+1}^{N-1}\sum_{j=1}^{m}\widetilde{c}_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij},S_{T_{ij}}>M_{ij}\}}\\
&\quad + \mathbb{E}^*\Bigg[K\mathrm{e}^{-rT_N}\mathbf{1}_{\{\tau>T_N,S_{T_{N-1}}<S_{N-1}^*\}} + \frac{K}{C_p}\mathrm{e}^{\kappa\tau}\widetilde{S}_\tau\mathbf{1}_{\{S_{T_{N-1}}<S_{N-1}^*,T_{N-1}<\tau\le T_N\}}\\
&\qquad \sum_{j=1}^{m}\widetilde{c}_{Nj}\mathbf{1}_{\{\tau_{Nj}>T_{Nj},S_{T_{Nj}}>M_{Nj},S_{T_{N-1}}<S_{N-1}^*\}}\Bigg|\mathscr{F}_{T_{N-1}}\Bigg].
\end{aligned}\tag{51}$$

So far we can see that basically all indicator functions appearing originally in (50), has been augmented in (51) by an additional condition on $S_{T_{N-1}}$ (*i.e.*, $S_{T_{N-1}} < S_{N-1}^*$

or $S_{T_{N-1}} \geq S^*_{N-1}$). In [14] it has been shown that the computation of the Snell envelop $(\widetilde{Y}^{(n)}_k)_{k\in\{n,...,N\}}$ can be carried out by means of the above backward procedure. In fact, it can be seen that the payoff of a *CoCo* having the callability feature can be written in terms of

$$\begin{aligned}
&c_{ij}\mathbf{1}_{\{\tau_{ij}>T_{ij}, S_{T_0}<S^*_0,...,S_{T_{i-1}}<S^*_{i-1}, S_{T_{ij}}>M_{ij}\}} \text{ at times } T_{ij},\, i\geq 1,\\
&K\mathbf{1}_{\{\tau>T_i, S_{T_0}<S^*_0,...,S_{T_{i-1}}<S^*_{i-1}, S_{T_i}>S^*_i\}} \text{ at } T_i,\, i\geq 1,\\
&\frac{K}{C_p}\mathrm{e}^{\kappa\tau}S_\tau\mathbf{1}_{\{\lceil\tau\rceil\leq T_N, S_{T_0}<S^*_0,...,S_{\lceil\tau\rceil-1}<S^*_{\lceil\tau\rceil-1}\}} \text{ at } \tau,
\end{aligned}$$

where $\lceil\tau\rceil$ is the element of $\{T_1, ..., T_N\}$ such that $\lceil\tau\rceil - 1 < \tau \leq \lceil\tau\rceil$, and the additional variables $S^*_1, ..., S^*_{N-2}$ are defined by analogous reasoning to that behind the obtainment of S^*_{N-1}. Here $S^*_0 := \infty$ and $S^*_N := 0$ are set by convention.

Proposition 12 *If the* CoCo *with extension risk is active and conversion has not occurred, then its discounted arbitrage-free price is given by*

$$\begin{aligned}
\widetilde{\Pi}_t = &\sum_{i,j:\, T_{ij}>t} \widetilde{c}_{ij}\mathbb{P}^*\left(\tau_{ij} > T_{ij}, S_{T_0} < S^*_0, ..., S_{T_{i-1}} < S^*_{i-1}, S_{T_{ij}} > M_{ij}, \,\middle|\, \mathscr{F}_t\right)\\
&+\sum_{i:\, T_i>t} \widetilde{K}\mathbb{P}^*\left(\tau > T_i, S_{T_0} < S^*_0, ..., S_{T_{i-1}} < S^*_{i-1}, S_{T_i} > S^*_i \,\middle|\, \mathscr{F}_t\right)\\
&+\sum_{i=1}^{N} \frac{K}{C_p}\mathrm{e}^{\kappa(T_i-t)}\widetilde{S}_t\mathbb{P}^{(S)}\left(\tau \leq T_i, S_{T_0} < S^*_0, ..., S_{T_{i-1}} < S^*_{i-1}, S_{T_i} \geq S^*_i \,\middle|\, \mathscr{F}_t\right).
\end{aligned} \tag{52}$$

Proof With the explicit description of the payoff corresponding to the *CoCo* with extension risk, the result is obtained as in Proposition 2, here taking into account the following identity

$$\begin{aligned}
&\left\{\lceil\tau\rceil \leq T_N, S_{T_0} < S^*_0, ..., S_{\lceil\tau\rceil-1} < S^*_{\lceil\tau\rceil-1}\right\}\\
&\quad = \uplus_{i=1}^N \{\tau \leq T_i, S_{T_0} < S^*_0, ..., S_{T_{i-1}} < S^*_{i-1}, S_{T_i} \geq S^*_i\}.
\end{aligned}$$

In view of this proposition, the obtainment of a closed-form formula for the price *CoCo* with extension risk requires the knowledge of the conditional distribution of $(\tau, S_{T_0}, S_{T_1}, ..., S_{T_i})$ for $i = 1, ..., N$. In the Black-Scholes model, this can be achieved by means of the following general lemma obtained in [14].

Lemma 13 *Let $(B_t)_{t\geq 0}$ be a Brownian motion with drift μ and volatility σ, and denote by τ its first-passage time to level zero. Then, for arbitrary instants $T_1 < \cdots < T_n$ and arbitrary non-negative constants $a_1, ..., a_n$, on $\{\tau > t\}$ the following equation holds true*

$$
\begin{aligned}
&\mathbb{P}\left(\tau \geq T_n, B_{T_1} < a_1, \ldots, B_{T_{n-1}} < a_{n-1}, B_{T_n} > a_n \middle| \mathscr{F}_t\right) \\
&= \mathbb{P}\left(-a_1 < B_{T_1} < a_1, \ldots, -a_{n-1} < B_{T_{n-1}} < a_{n-1}, B_{T_n} > a_n \middle| \mathscr{F}_t\right) \\
&\quad - \mathrm{e}^{-2\mu\sigma^{-1}B_t}\mathbb{P}\left(-a_1 < \bar{B}_{T_1} < a_1, \ldots, -a_{n-1} < \bar{B}_{T_{n-1}} < a_{n-1}, \bar{B}_{T_n} < -a_n \middle| \mathscr{F}_t\right),
\end{aligned}
$$

where $\bar{B}_{T_j} = B_{T_j} - 2\mu(T_j - t)$, $j = 1, \ldots, n$, *and* $(\mathscr{F}_t)_{t \geq 0}$ *stands for the natural filtration generated by* $(B_t)_{t \geq 0}$.

Acknowledgments The work of J.M. Corcuera is supported by the Grant of the Spanish MCINN MTM2013-40782. This research was carried out at CAS -Centre for Advanced Study at the Norwegian Academy of Science and Letters, Research group SEFE.

References

1. Bertoin, J.: Lévy Processes. Cambridge University Press, Cambridge (1996)
2. Brigo, D., Garcia, J., Pede, N.: Coco bonds pricing with credit and equity calibrated first-passage firm value models. Int. J. Theor. Appl. Financ. **18**(3), 1550, 015 (2015). doi:10.1142/S0219024915500156
3. Caballero, M., Pardo, J., Perez, J.: On the Lamperti stable processes. Probab. Math. Stat. **30**, 1–28 (2010)
4. Carr, P., Geman, H., Madan, D., Yor, M.: The fine structure of asset returns: an empirical investigation. J. Bus. **75**, 305–332 (2002)
5. Chan, S., van Wijnbergen, S.: CoCos, Contagion and Systemic Risk (14-110/VI/DSF79) (2014). http://ideas.repec.org/p/dgr/uvatin/20140110.html
6. Chen, N., Glasserman, P., Nouri, B., Pelger, M.: CoCos, bail-in, and tail risk. In: OFR Working paper. US Department of the Treasury (2013)
7. Cheridito, P., Xu, Z.: Pricing and hedging cocos. In: Social Science Research Network Working Paper Series (2013). http://dx.doi.org/10.2139/ssrn.2201364
8. Cheridito, P., Xu, Z.: A reduced form CoCo model with deterministic conversion intensity. In: Social Science Research Network Working Paper Series (2013). http://dx.doi.org/10.2139/ssrn.2254403
9. Coculescu, D., Geman, H., Jeanblanc, M.: Valuation of default-sensitive claims under imperfect information. Financ. Stoch. **12**, 195–218 (2008)
10. Collin-Dufresne, P., Goldstein, R., Helwege, J.: P. is credit event risk priced? modeling contagion via the updating of beliefs. In: Technical Report, National Bureau of Economic Research (2010)
11. Commission of Experts: Final report of the commission of experts for limiting the economic risks posed by large companies. In: Technical Report, Swiss National Bank (2010)
12. Corcuera, J.M., De Spiegeleer, J., Ferreiro-Castilla, A., Kyprianou, A.E., Madan, D.B., Schoutens, W.: Efficient pricing of contingent convertibles under smile conform models. J. Credit Risk **9**(3), 121–140 (2013)
13. Corcuera, J.M., De Spiegeleer, J., Jönsson, H., Fajardo, J., Shoutens, W., Valdivia, A.: Close form pricing formulas for coupon cancellable cocos. J. Bank. Financ. (2014)
14. Corcuera, J.M., Fajardo, J., Schoutens, W., Valdivia, A.: CoCos with extension risk: a structural approach. In: Social Science Research Network Working Paper Series (2014). http://dx.doi.org/10.2139/ssrn.2540625

15. Corcuera, J.M., Valdivia, A.: Short-term uncertainty. Working paper, University of Barcelona (2015)
16. De Spiegeleer, J., Schoutens, W.: Pricing contingent convertibles: a derivatives approach. J. Deriv. **20**(2), 27–36 (2012)
17. De Spiegeleer, J., Schoutens, W.: Steering a bank around a death spiral: multiple trigger cocos. Wilmott **2012**(59), 62–69 (2012)
18. De Spiegeleer, J., Schoutens, W.: CoCo bonds with extension risk. Wilmott **71**, 78–91 (2014)
19. Dewatripont, M., Tirole, J.: The prudential regulation of banks. In: ULB-Universite Libre de Bruxelles, Technical Report (1994)
20. Dewatripont, M., Tirole, J.: Macroeconomic shocks and banking regulation. J. Money, Credit Bank. **44**(s2), 237–254 (2012)
21. Duffie, D., Lando, D.: Term structure of credit spreads with incomplete accounting information. Econometrica **69**, 633–664 (2001)
22. Flannery, M.: No pain, no gain? effecting market discipline via "reverse convertible debentures". In: Scott, H.S. (ed.) Capital Adequacy Beyond Basel: Banking, Securities, and Insurance, pp. 171–196 (2005)
23. Flannery, M.J.: Stabilizing large financial institutions with contingent capital certificates. In: CAREFIN Research Paper (04) (2010)
24. Gordy, M.B.: A risk-factor model foundation for ratings-based bank capital rules. J. Financ. Intermed. **12**(3), 199–232 (2003)
25. Guo, X., Jarrow, A., Zeng, Y.: Credit risk models with incomplete information. Math. Oper. Res. **34**(2), 320–332 (2009)
26. Heston, S.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. **6**(2), 327–343 (1993)
27. Jeanblanc, M., Valchev, S.: Partial information and hazard process. Int. J. Theor. Appl. Financ. **8**, 807–838 (2005)
28. Koziol, C., Lawrenz, J.: Contingent convertibles. Solving or seeding the next banking crisis? J. Bank. Financ. **36**(1), 90–104 (2012)
29. Kuznetsov, A., Kyprianou, A.E., Pardo, J.C.: A Wiener-Hopf Monte Carlo simulation technique for Lévy process. Ann. Appl. Probab. **21**(6), 2171–2190 (2011)
30. Madan, D.B., Milne, F.: Option pricing with vg martingale components. Math. Financ. **1**(4), 39–55 (1991)
31. Madan, D.B., Schoutens, W.: Break on through to the single side. J. Credit Risk **4**(3), 3–20 (2008)
32. Pennacchi, G., Vermaelen, T., Wolff, C.: Contingent capital: the case for coercs. Available at SSRN 1785380 (2011)
33. Rogers, L.: Evaluating first-passage probabilities for spectrally one-sided Lévy processes. J. Appl. Probab. **37**, 1173–1180 (2000)
34. Salminen, P., Borodin, A.: Handbook of Brownian Motion: Facts and Formulae. Birkhauser, Basel (1996)
35. Sato, K.: Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics, vol. 68. Cambridge University Press, Cambridge (2000)
36. Schoutens, W., Cariboni, J.: Lévy Processes in Credit Risk. The Wiley Finance Series. Wiley, Chichester (2009)
37. Schweizer, M.: Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann., Chapter, On Bermudan Options, pp. 257–269. Springer (2002)
38. Stuart, A., Ord, J.: Kendall's Advanced Theory of Statistics. Distribution Theory, vol. 1. Wiley, Chichester (1994)
39. Sundaresan, S., Wang, Z.: On the design of contingent capital with a market trigger. J. Financ. **70**(2), 881–920 (2015). doi:10.1111/jofi.12134

Quantification of Model Risk in Quadratic Hedging in Finance

Catherine Daveloose, Asma Khedher and Michèle Vanmaele

Abstract In this paper the effect of the choice of the model on partial hedging in incomplete markets in finance is estimated. In fact we compare the quadratic hedging strategies in a martingale setting for a claim when two models for the underlying stock price are considered. The first model is a geometric Lévy process in which the small jumps might have infinite activity. The second model is a geometric Lévy process where the small jumps are replaced by a Brownian motion which is appropriately scaled. The hedging strategies are related to solutions of backward stochastic differential equations with jumps which are driven by a Brownian motion and a Poisson random measure. We use this relation to prove that the strategies are robust towards the choice of the model for the market prices and to estimate the model risk.

Keywords Lévy models · Quadratic hedging · Model risk · Robustness · BSDEJs

MSC 2010 Codes: 60G51 · 91B30 · 91G80

1 Introduction

When jumps are present in the stock price model, the market is in general incomplete and there is no self-financing hedging strategy which allows to attain the contingent claim at maturity. In other words, one cannot eliminate the risk completely. However

C. Daveloose · M. Vanmaele (✉)
Department of Applied Mathematics, Computer Science and Statistics,
Ghent University, Krijgslaan 281-S9, 9000 Gent, Belgium
e-mail: Michele.Vanmaele@UGent.be

C. Daveloose
e-mail: Catherine.Daveloose@UGent.be

A. Khedher
Chair of Mathematical Finance, Technische Universität München, Parkring 11,
85748 Garching-Hochbrück, Germany
e-mail: Asma.Khedher@tum.de

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_8

it is possible to find 'partial' hedging strategies which minimise some risk. One way to determine these 'partial' hedging strategies is to introduce a subjective criterion according to which strategies are optimised.

In the present paper, we consider two types of quadratic hedging strategies. The first, called *risk-minimising* (RM) strategy, is replicating the option's payoff, but it is not self-financing (see, e.g., [19]). In such strategies, the hedging is considered under a *risk-neutral measure* or *equivalent martingale measure*. The aim is to minimise the risk process, which is induced by the fact that the strategy is not self-financing, under this measure. In the second approach, called *mean-variance hedging* (MVH), the strategy is self-financing and the quadratic hedging error at maturity is minimised in mean square sense (see, e.g., [19]). Again a risk-neutral setting is assumed.

The aim in this paper is to investigate whether these quadratic hedging strategies (RM and MVH) in incomplete markets are robust to the variation of the model. Thus we consider two geometric Lévy processes to model the asset price dynamics. The first model $(S_t)_{t\in[0,T]}$ is driven by a Lévy process in which the small jumps might have infinite activity. The second model $(S_t^\varepsilon)_{t\in[0,T]}$ is driven by a Lévy process in which we replace the jumps with absolute size smaller than $\varepsilon > 0$ by an appropriately scaled Brownian motion. The latter model $(S_t^\varepsilon)_{t\in[0,T]}$ converges to the first one in an L^2-sense when ε goes to 0. The aim is to study whether similar convergence properties hold for the corresponding quadratic hedging strategies.

Geometric Lévy processes describe well realistic asset price dynamics and are well established in the literature (see e.g., [5]). Moreover, the idea of shifting from a model with small jumps to another where these variations are represented by some appropriately scaled continuous component goes back to [2]. This idea is interesting from a simulation point of view. Indeed, the process $(S_t^\varepsilon)_{t\in[0,T]}$ contains a compound Poisson process and a scaled Brownian motion which are both easy to simulate. Whereas it is not easy to simulate the infinite activity of the small jumps in the process $(S_t)_{t\in[0,T]}$ (see [5] for more about simulation of Lévy processes).

The interest of this paper is the *model risk*. In other words, from a modelling point of view, we may think of two financial agents who want to price and hedge an option. One is considering $(S_t)_{t\in[0,T]}$ as a model for the price process and the other is considering $(S_t^\varepsilon)_{t\in[0,T]}$. Thus the first agent chooses to consider infinitely small variations in a discontinuous way, i.e. in the form of infinitely small jumps of an infinite activity Lévy process. The second agent observes the small variations in a continuous way, i.e. coming from a Brownian motion. Hence the difference between both market models determines a type of model risk and the question is whether the pricing and hedging formulas corresponding to $(S_t^\varepsilon)_{t\in[0,T]}$ converge to the pricing and hedging formulas corresponding to $(S_t)_{t\in[0,T]}$ when ε goes to zero. This is what we intend in the sequel by *robustness* or *stability* study of the model.

In this paper we focus mainly on the RM strategies. These strategies are considered under a martingale measure which is equivalent to the historical measure. Equivalent martingale measures are characterised by the fact that the discounted asset price processes are martingales under these measures. The problem we are facing is that the martingale measure is dependent on the choice of the model. Therefore it is clear that, in this paper, there will be different equivalent martingale measures for the two considered price models. Here we emphasise that for the robustness study, we come back to the common underlying physical measure.

Besides, since the market is incomplete, we will also have to identify which equivalent martingale measure, or measure change, to apply. In particular, we discuss some specific martingale measures which are commonly used in finance and in electricity markets: the Esscher transform, the minimal entropy martingale measure, and the minimal martingale measure. We prove some common properties for the mentioned martingale measures in the exponential Lévy setting in addition to those shown in [4, 6].

To perform the described stability study, we follow the approach in [8] and we relate the RM hedging strategies to backward stochastic differential equations with jumps (BSDEJs). See e.g. [7, 9] for an overview about BSDEs and their applications in hedging and in nonlinear pricing theory for incomplete markets.

Under some conditions on the parameters of the stock price process and of the martingale measure, we investigate the robustness to the choice of the model of the value of the portfolio, the amount of wealth, the cost and gain process in a RM strategy. The amount of wealth and the gain process in a MVH strategy coincide with those in the RM strategy and hence the convergence results will immediately follow. When we assume a fixed initial portfolio value to set up a MVH strategy we derive a convergence rate for the loss at maturity.

The BSDEJ approach does not provide a robustness result for the *optimal number* of risky assets in a RM strategy as well as in a MVH strategy. In [6] convergence rates for those optimal numbers and other quantities, such as the delta and the amount of wealth, are computed using Fourier transform techniques.

The paper is organised as follows: in Sect. 2 we introduce the notations, define the two martingale models for the stock price, and derive the corresponding BSDEJs for the value of the discounted RM hedging portfolio. In Sect. 3 we study the stability of the quadratic hedging strategies towards the choice of the model and obtain convergence rates. In Sect. 4 we conclude.

2 Quadratic Hedging Strategies in a Martingale Setting for Two Geometric Lévy Stock Price Models

Assume a finite time horizon $T > 0$. The first considered stock price process is determined by the process $L = (L_t)_{t\in[0,T]}$ which denotes a Lévy process in the filtered complete probability space $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ satisfying the usual hypotheses as defined in [18]. We work with the càdlàg version of the given Lévy process. The characteristic triplet of the Lévy process L is denoted by (a, b^2, ℓ). We consider a stock price modelled by a geometric Lévy process, i.e. the stock price is given by $S_t = S_0 e^{L_t}$, $\forall t \in [0, T]$, where $S_0 > 0$. Let $r > 0$ be the risk-free instantaneous interest rate. The value of the corresponding riskless asset equals e^{rt} for any time

$t \in [0, T]$. We denote the discounted stock price process by $\hat{S}$. Hence at any time $t \in [0, T]$ it equals

$$\hat{S}_t = \mathrm{e}^{-rt} S_t = S_0 \mathrm{e}^{-rt} \mathrm{e}^{L_t}.$$

It holds that

$$\mathrm{d}\hat{S}_t = \hat{S}_t \hat{a} \mathrm{d}t + \hat{S}_t b \mathrm{d}W_t + \hat{S}_t \int_{\mathbb{R}_0} (\mathrm{e}^z - 1) \widetilde{N}(\mathrm{d}t, \mathrm{d}z), \tag{1}$$

where W is a standard Brownian motion independent of the compensated jump measure $\widetilde{N}$ and

$$\hat{a} = a - r + \frac{1}{2} b^2 + \int_{\mathbb{R}_0} \left(\mathrm{e}^z - 1 - z 1_{\{|z|<1\}}\right) \ell(\mathrm{d}z).$$

It is assumed that $\hat{S}$ is not deterministic and arbitrage opportunities are excluded (cfr. [21]). The aim of this paper is to study the stability of quadratic hedging strategies in a martingale setting towards the choice of the model. Since the equivalent martingale measure is determined by the market model, we also have to take into account the robustness of the risk-neutral measures. Therefore we consider the case where $\mathbb{P}$ is not a risk-neutral measure, or in other words $\hat{a} \neq 0$ so that $\hat{S}$ is not a $\mathbb{P}$-martingale. Then, a change of measure, specifically determined by the market model (1), will have to be performed to obtain a martingale setting. Let us denote a martingale measure which is equivalent to the historical measure $\mathbb{P}$ by $\widetilde{\mathbb{P}}$. We consider martingale measures that belong to the class of structure preserving martingale measures, see [14]. In this case, the Lévy triplet of the driving process L under $\widetilde{\mathbb{P}}$ is denoted by $(\tilde{a}, b^2, \tilde{\ell})$. Theorem III.3.24 in [14] states conditions which are equivalent to the existence of a parameter $\Theta \in \mathbb{R}$ and a function $\rho(z; \Theta)$, $z \in \mathbb{R}$, such that

$$\int_{\{|z|<1\}} |z\,(\rho(z; \Theta) - 1)|\, \ell(\mathrm{d}z) < \infty, \tag{2}$$

and such that

$$\tilde{a} = a + b^2 \Theta + \int_{\{|z|<1\}} z\,(\rho(z; \Theta) - 1)\, \ell(\mathrm{d}z) \quad \text{and} \quad \tilde{\ell}(\mathrm{d}z) = \rho(z; \Theta) \ell(\mathrm{d}z). \tag{3}$$

For $\hat{S}$ to be a martingale under $\widetilde{\mathbb{P}}$, the parameter Θ should guarantee the following equation

$$\hat{a}_0 = \tilde{a} - r + \frac{1}{2} b^2 + \int_{\mathbb{R}_0} \left(\mathrm{e}^z - 1 - z 1_{\{|z|<1\}}\right) \tilde{\ell}(\mathrm{d}z) = 0. \tag{4}$$

From now on we denote the solution of Eq. (4)—when it exists—by Θ_0 and the equivalent martingale measure by $\widetilde{\mathbb{P}}_{\Theta_0}$. Notice that we obtain different martingale measures $\widetilde{\mathbb{P}}_{\Theta_0}$ for different choices of the function $\rho(. ; \Theta_0)$. In the next section we present some known martingale measures for specific functions $\rho(.\,;\Theta_0)$ and specific parameters Θ_0 which solve (4).

The relation between the original measure $\mathbb{P}$ and the martingale measure $\widetilde{\mathbb{P}}_{\Theta_0}$ is given by

$$\frac{\mathrm{d}\widetilde{\mathbb{P}}_{\Theta_0}}{\mathrm{d}\mathbb{P}}\Big|_{\mathscr{F}_t} = \exp\Bigg(b\Theta_0 W_t - \frac{1}{2}b^2{\Theta_0}^2 t + \int_0^t\int_{\mathbb{R}_0}\log\left(\rho(z;\Theta_0)\right)\widetilde{N}(\mathrm{d}s,\mathrm{d}z) \\ + t\int_{\mathbb{R}_0}\left(\log\left(\rho(z;\Theta_0)\right)+1-\rho(z;\Theta_0)\right)\ell(\mathrm{d}z)\Bigg).$$

From the Girsanov theorem (see e.g. Theorem 1.33 in [17]) we know that the processes W^{Θ_0} and $\widetilde{N}^{\Theta_0}$ defined by

$$\mathrm{d}W_t^{\Theta_0} = \mathrm{d}W_t - b\Theta_0\mathrm{d}t, \tag{5}$$
$$\widetilde{N}^{\Theta_0}(\mathrm{d}t,\mathrm{d}z) = N(\mathrm{d}t,\mathrm{d}z) - \rho(z;\Theta_0)\ell(\mathrm{d}z)\mathrm{d}t = \widetilde{N}(\mathrm{d}t,\mathrm{d}z) + \left(1-\rho(z;\Theta_0)\right)\ell(\mathrm{d}z)\mathrm{d}t,$$

for all $t\in[0,T]$ and $z\in\mathbb{R}_0$, are a standard Brownian motion and a compensated jump measure under $\widetilde{\mathbb{P}}_{\Theta_0}$. Moreover we can rewrite (1) as

$$\mathrm{d}\hat{S}_t = \hat{S}_t b\mathrm{d}W_t^{\Theta_0} + \hat{S}_t\int_{\mathbb{R}_0}(\mathrm{e}^z-1)\widetilde{N}^{\Theta_0}(\mathrm{d}t,\mathrm{d}z). \tag{6}$$

We consider an $\mathscr{F}_T$-measurable and square integrable random variable H_T which denotes the payoff of a contract. The discounted payoff equals $\hat{H}_T = \mathrm{e}^{-rT}H_T$. In case the discounted stock price process is a martingale, both, the mean-variance hedging (MVH) and the risk-minimising strategy (RM) are related to the Galtchouk-Kunita-Watanabe (GKW) decomposition, see [11]. In the following we recall the GKW-decomposition of the $\mathscr{F}_T$-measurable and square integrable random variable $\hat{H}_T$ under the martingale measure $\widetilde{\mathbb{P}}_{\Theta_0}$

$$\hat{H}_T = \widetilde{\mathbb{E}}^{\Theta_0}[\hat{H}_T] + \int_0^T \xi_s^{\Theta_0}\mathrm{d}\hat{S}_s + \mathscr{L}_T^{\Theta_0}, \tag{7}$$

where $\widetilde{\mathbb{E}}^{\Theta_0}$ denotes the expectation under $\widetilde{\mathbb{P}}_{\Theta_0}$, ξ^{Θ_0} is a predictable process for which we can determine the stochastic integral with respect to $\hat{S}$, and $\mathscr{L}^{\Theta_0}$ is a square integrable $\widetilde{\mathbb{P}}_{\Theta_0}$-martingale with $\mathscr{L}_0^{\Theta_0}=0$, such that $\mathscr{L}^{\Theta_0}$ is $\widetilde{\mathbb{P}}_{\Theta_0}$-orthogonal to $\hat{S}$.

The quadratic hedging strategies are determined by the process ξ^{Θ_0}. It indicates the number of discounted risky assets to hold in the portfolio. The amount invested in the riskless asset is different in both strategies and is determined by the self-financing property for the MVH strategy and by the replicating condition for the RM strategy. See [19] for more details.

We define the process

$$\hat{V}_t^{\Theta_0} = \widetilde{\mathbb{E}}^{\Theta_0}[\hat{H}_T|\mathscr{F}_t], \quad \forall t \in [0, T],$$

which equals the value of the discounted portfolio for the RM strategy. The GKW-decomposition (7) implies that

$$\hat{V}_t^{\Theta_0} = \hat{V}_0^{\Theta_0} + \int_0^t \xi_s^{\Theta_0} \mathrm{d}\hat{S}_s + \mathscr{L}_t^{\Theta_0}, \quad \forall t \in [0, T]. \tag{8}$$

Moreover since $\mathscr{L}^{\Theta_0}$ is a $\widetilde{\mathbb{P}}_{\Theta_0}$-martingale, there exist processes X^{Θ_0} and $Y^{\Theta_0}(z)$ such that

$$\mathscr{L}_t^{\Theta_0} = \int_0^t X_s^{\Theta_0} \mathrm{d}W_s^{\Theta_0} + \int_0^t \int_{\mathbb{R}_0} Y_s^{\Theta_0}(z) \widetilde{N}^{\Theta_0}(\mathrm{d}s, \mathrm{d}z), \quad \forall t \in [0, T], \tag{9}$$

and which by the $\widetilde{\mathbb{P}}_{\Theta_0}$-orthogonality of $\mathscr{L}^{\Theta_0}$ and $\hat{S}$ satisfy

$$X^{\Theta_0} b + \int_{\mathbb{R}_0} Y^{\Theta_0}(z)(\mathrm{e}^z - 1)\rho(z; \Theta_0)\ell(\mathrm{d}z) = 0. \tag{10}$$

By substituting (6) and (9) in (8), we retrieve

$$\mathrm{d}\hat{V}_t^{\Theta_0} = \left(\xi_t^{\Theta_0} \hat{S}_t b + X_t^{\Theta_0}\right) \mathrm{d}W_t^{\Theta_0} + \int_{\mathbb{R}_0} \left(\xi_t^{\Theta_0} \hat{S}_t (\mathrm{e}^z - 1) + Y_t^{\Theta_0}(z)\right) \widetilde{N}^{\Theta_0}(\mathrm{d}t, \mathrm{d}z).$$

Let $\hat{\pi}^{\Theta_0} = \xi^{\Theta_0} \hat{S}$ indicate the amount of wealth invested in the discounted risky asset in a quadratic hedging strategy. We conclude that the following BSDEJ holds for the RM strategy

$$\begin{cases} \mathrm{d}\hat{V}_t^{\Theta_0} = A_t^{\Theta_0} \mathrm{d}W_t^{\Theta_0} + \int_{\mathbb{R}_0} B_t^{\Theta_0}(z) \widetilde{N}^{\Theta_0}(\mathrm{d}t, \mathrm{d}z), \\ \hat{V}_T^{\Theta_0} = \hat{H}_T, \end{cases} \tag{11}$$

where

$$A^{\Theta_0} = \hat{\pi}^{\Theta_0} b + X^{\Theta_0} \quad \text{and} \quad B^{\Theta_0}(z) = \hat{\pi}^{\Theta_0}(\mathrm{e}^z - 1) + Y^{\Theta_0}(z). \tag{12}$$

Since the random variable $\hat{H}_T$ is square integrable and $\mathscr{F}_T$-measurable, we know by [20] that the BSDEJ (11) has a unique solution $(\hat{V}^{\Theta_0}, A^{\Theta_0}, B^{\Theta_0})$. This follows from the fact that the drift parameter of $\hat{V}^{\Theta_0}$ equals zero under $\widetilde{\mathbb{P}}_{\Theta_0}$ and thus it is Lipschitz continuous.

We introduce another Lévy process L^ε, for $0 < \varepsilon < 1$, which is obtained by truncating the jumps of L with absolute size smaller than ε and replacing them by an independent Brownian motion which is appropriately scaled. The second stock price process is denoted by $S^\varepsilon = S_0 \mathrm{e}^{L^\varepsilon}$ and the corresponding discounted stock price process $\hat{S}^\varepsilon$ is thus given by

$$\mathrm{d}\hat{S}_t^\varepsilon = \hat{S}_t^\varepsilon \hat{a}_\varepsilon \mathrm{d}t + \hat{S}_t^\varepsilon b \mathrm{d}W_t + \hat{S}_t^\varepsilon \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\widetilde{N}(\mathrm{d}t, \mathrm{d}z) + \hat{S}_t^\varepsilon G(\varepsilon)\mathrm{d}\widetilde{W}_t, \tag{13}$$

for all $t \in [0, T]$ and $\hat{S}_0^\varepsilon = S_0$. Herein $\widetilde{W}$ is a standard Brownian motion independent of W,

$$G^2(\varepsilon) = \int_{\{|z|<\varepsilon\}} (\mathrm{e}^z - 1)^2 \ell(\mathrm{d}z), \text{ and} \tag{14}$$

$$\hat{a}_\varepsilon = a - r + \frac{1}{2}\left(b^2 + G^2(\varepsilon)\right) + \int_{\{|z|\geq\varepsilon\}} \left(\mathrm{e}^z - 1 - z1_{\{|z|<1\}}\right) \ell(\mathrm{d}z).$$

From now on, we assume that the filtration $\mathbb{F}$ is enlarged with the information of the Brownian motion $\widetilde{W}$ and we denote the new filtration by $\widetilde{\mathbb{F}}$. Moreover, we also assume absence of arbitrage in this second model. It is clear that the process L^ε has the Lévy characteristic triplet $\left(a, b^2 + G^2(\varepsilon), 1_{\{|\cdot|\geq\varepsilon\}}\ell\right)$ under the measure $\mathbb{P}$.

Let $\widetilde{\mathbb{P}}_\varepsilon$ represent a structure preserving martingale measure for $\hat{S}^\varepsilon$. The characteristic triplet of the driving process L^ε w.r.t. this martingale measure is denoted by $\left(\tilde{a}_\varepsilon, b^2 + G^2(\varepsilon), \tilde{\ell}_\varepsilon\right)$. From [14, Theorem III.3.24] we know that there exist a parameter $\Theta \in \mathbb{R}$ and a function $\rho(z;\Theta)$, $z \in \mathbb{R}$, under certain conditions, such that

$$\int_{\{\varepsilon\leq|z|<1\}} |z(\rho(z;\Theta) - 1)| \, \ell(\mathrm{d}z) < \infty, \tag{15}$$

$$\tilde{a}_\varepsilon = a + \left(b^2 + G^2(\varepsilon)\right)\Theta + \int_{\{\varepsilon\leq|z|<1\}} z\left(\rho(z;\Theta) - 1\right)\ell(\mathrm{d}z), \text{ and} \tag{16}$$

$$\tilde{\ell}_\varepsilon(\mathrm{d}z) = 1_{\{|z|\geq\varepsilon\}}\rho(z;\Theta)\ell(\mathrm{d}z). \tag{17}$$

Let us assume that Θ solves the following equation

$$\tilde{a}_\varepsilon - r + \frac{1}{2}\left(b^2 + G^2(\varepsilon)\right) + \int_{\mathbb{R}_0} \left(\mathrm{e}^z - 1 - z1_{\{|z|<1\}}\right)\tilde{\ell}_\varepsilon(\mathrm{d}z) = 0, \tag{18}$$

then $\hat{S}^\varepsilon$ is a martingale under $\widetilde{\mathbb{P}}$. From now on we indicate the solution of (18)—when it exists—as Θ_ε and the martingale measure as $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$.

The relation between the original measure $\mathbb{P}$ and the martingale measure $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$ is given by

$$\frac{\mathrm{d}\widetilde{\mathbb{P}}_{\Theta_\varepsilon}}{\mathrm{d}\mathbb{P}}\Big|_{\widetilde{\mathscr{F}}_t} = \exp\Bigg(b\Theta_\varepsilon W_t - \frac{1}{2}b^2{\Theta_0}^2 t + G(\varepsilon)\Theta_\varepsilon \widetilde{W}_t - \frac{1}{2}G^2(\varepsilon){\Theta_\varepsilon}^2 t \\ + \int_0^t \int_{\{|z|\geq\varepsilon\}} \log(\rho(z;\Theta_\varepsilon))\widetilde{N}(\mathrm{d}s,\mathrm{d}z) \\ + t\int_{\{|z|\geq\varepsilon\}} \big(\log(\rho(z;\Theta_\varepsilon)) + 1 - \rho(z;\Theta_\varepsilon)\big)\ell(\mathrm{d}z)\Bigg).$$

The processes W^{Θ_ε}, $\widetilde{W}^{\Theta_\varepsilon}$, and $\widetilde{N}^{\Theta_\varepsilon}$ defined by

$$\begin{aligned} \mathrm{d}W_t^{\Theta_\varepsilon} &= \mathrm{d}W_t - b\Theta_\varepsilon \mathrm{d}t, \\ \mathrm{d}\widetilde{W}_t^{\Theta_\varepsilon} &= \mathrm{d}\widetilde{W}_t - G(\varepsilon)\Theta_\varepsilon \mathrm{d}t, \\ \widetilde{N}^{\Theta_\varepsilon}(\mathrm{d}t,\mathrm{d}z) &= N(\mathrm{d}t,\mathrm{d}z) - \rho(z;\Theta_\varepsilon)\ell(\mathrm{d}z)\mathrm{d}t \\ &= \widetilde{N}(\mathrm{d}t,\mathrm{d}z) + (1-\rho(z;\Theta_\varepsilon))\ell(\mathrm{d}z)\mathrm{d}t, \end{aligned} \tag{19}$$

for all $t \in [0,T]$ and $z \in \{z \in \mathbb{R} : |z| \geq \varepsilon\}$, are two standard Brownian motions and a compensated jump measure under $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$ (see e.g. Theorem 1.33 in [17]). Hence the process $\hat{S}^\varepsilon$ is given by

$$\mathrm{d}\hat{S}_t^\varepsilon = \hat{S}_t^\varepsilon b \mathrm{d}W_t^{\Theta_\varepsilon} + \hat{S}_t^\varepsilon \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\widetilde{N}^{\Theta_\varepsilon}(\mathrm{d}t,\mathrm{d}z) + \hat{S}_t^\varepsilon G(\varepsilon)\mathrm{d}\widetilde{W}_t^{\Theta_\varepsilon}. \tag{20}$$

We consider an $\widetilde{\mathscr{F}}_T$-measurable and square integrable random variable H_T^ε which is the payoff of a contract. The discounted payoff is denoted by $\hat{H}_T^\varepsilon = \mathrm{e}^{-rT}H_T^\varepsilon$. The GKW-decomposition of $\hat{H}_T^\varepsilon$ under the martingale measure $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$ equals

$$\hat{H}_T^\varepsilon = \widetilde{\mathbb{E}}^{\Theta_\varepsilon}[\hat{H}_T^\varepsilon] + \int_0^T \xi_s^{\Theta_\varepsilon} \mathrm{d}\hat{S}_s^\varepsilon + \mathscr{L}_T^{\Theta_\varepsilon}, \tag{21}$$

where $\widetilde{\mathbb{E}}^{\Theta_\varepsilon}$ is the expectation under $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$, ξ^{Θ_ε} is a predictable process for which we can determine the stochastic integral with respect to $\hat{S}^\varepsilon$, and $\mathscr{L}^{\Theta_\varepsilon}$ is a square integrable $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$-martingale with $\mathscr{L}_0^{\Theta_\varepsilon} = 0$, such that $\mathscr{L}^{\Theta_\varepsilon}$ is $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$-orthogonal to $\hat{S}^\varepsilon$.

The value of the discounted portfolio for the RM strategy is defined by

$$\hat{V}_t^{\Theta_\varepsilon} = \widetilde{\mathbb{E}}^{\Theta_\varepsilon}[\hat{H}_T^\varepsilon | \widetilde{\mathscr{F}}_t], \quad \forall t \in [0,T].$$

From the GKW-decomposition (21) we have

$$\hat{V}_t^{\Theta_\varepsilon} = \hat{V}_0^{\Theta_\varepsilon} + \int_0^t \xi_s^{\Theta_\varepsilon} \mathrm{d}\hat{S}_s^\varepsilon + \mathscr{L}_t^{\Theta_\varepsilon}, \quad \forall t \in [0,T]. \tag{22}$$

Moreover since $\mathscr{L}^{\Theta_\varepsilon}$ is a $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$-martingale, there exist processes X^{Θ_ε}, $Y^{\Theta_\varepsilon}(z)$, and Z^{Θ_ε} such that

$$\mathscr{L}_t^{\Theta_\varepsilon} = \int_0^t X_s^{\Theta_\varepsilon} \mathrm{d}W_s^{\Theta_\varepsilon} + \int_0^t \int_{\{|z|\geq\varepsilon\}} Y_s^{\Theta_\varepsilon}(z) \widetilde{N}^{\Theta_\varepsilon}(\mathrm{d}s, \mathrm{d}z) + \int_0^t Z_s^{\Theta_\varepsilon} \mathrm{d}\widetilde{W}_s^{\Theta_\varepsilon}, \quad \forall t \in [0, T]. \tag{23}$$

The $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$-orthogonality of $\mathscr{L}^{\Theta_\varepsilon}$ and $\hat{S}^\varepsilon$ implies that

$$X^{\Theta_\varepsilon} b + \int_{\{|z|\geq\varepsilon\}} Y^{\Theta_\varepsilon}(z)(\mathrm{e}^z - 1)\rho(z; \Theta_\varepsilon)\ell(\mathrm{d}z) + Z^{\Theta_\varepsilon} G(\varepsilon) = 0. \tag{24}$$

Combining (20) and (23) in (22), we get

$$\begin{aligned}\mathrm{d}\hat{V}_t^{\Theta_\varepsilon} &= \left(\xi_t^{\Theta_\varepsilon} \hat{S}_t^\varepsilon b + X_t^{\Theta_\varepsilon}\right) \mathrm{d}W_t^{\Theta_\varepsilon} + \int_{\{|z|\geq\varepsilon\}} \left(\xi_t^{\Theta_\varepsilon} \hat{S}_t^\varepsilon (\mathrm{e}^z - 1) + Y_t^{\Theta_\varepsilon}(z)\right) \widetilde{N}^{\Theta_\varepsilon}(\mathrm{d}t, \mathrm{d}z) \\ &\quad + \left(\xi_t^{\Theta_\varepsilon} \hat{S}_t^\varepsilon G(\varepsilon) + Z_t^{\Theta_\varepsilon}\right) \mathrm{d}\widetilde{W}_t^{\Theta_\varepsilon}.\end{aligned}$$

Let $\hat{\pi}^{\Theta_\varepsilon} = \xi^{\Theta_\varepsilon} \hat{S}^\varepsilon$ denote the amount of wealth invested in the discounted risky asset in the quadratic hedging strategy. We conclude that the following BSDEJ holds for the RM strategy

$$\begin{cases} \mathrm{d}\hat{V}_t^{\Theta_\varepsilon} = A_t^{\Theta_\varepsilon} \mathrm{d}W_t^{\Theta_\varepsilon} + \int_{\{|z|\geq\varepsilon\}} B_t^{\Theta_\varepsilon}(z) \widetilde{N}^{\Theta_\varepsilon}(\mathrm{d}t, \mathrm{d}z) + C_t^{\Theta_\varepsilon} \mathrm{d}\widetilde{W}_t^{\Theta_\varepsilon}, \\ \hat{V}_T^{\Theta_\varepsilon} = \hat{H}_T^\varepsilon, \end{cases} \tag{25}$$

where

$$\begin{aligned} &A^{\Theta_\varepsilon} = \hat{\pi}^{\Theta_\varepsilon} b + X^{\Theta_\varepsilon}, \quad B^{\Theta_\varepsilon}(z) = \hat{\pi}^{\Theta_\varepsilon}(\mathrm{e}^z - 1) + Y^{\Theta_\varepsilon}(z), \quad \text{and} \\ &C^{\Theta_\varepsilon} = \hat{\pi}^{\Theta_\varepsilon} G(\varepsilon) + Z^{\Theta_\varepsilon}. \end{aligned} \tag{26}$$

Since the random variable $\hat{H}_T^\varepsilon$ is square integrable and $\widetilde{\mathscr{F}}_T$-measurable we know by [20] that the BSDEJ (25) has a unique solution $\left(\hat{V}^{\Theta_\varepsilon}, A^{\Theta_\varepsilon}, B^{\Theta_\varepsilon}, C^{\Theta_\varepsilon}\right)$. This results from the fact that the drift parameter of $\hat{V}^{\Theta_\varepsilon}$ equals zero under $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$ and thus is Lipschitz continuous.

3 Robustness of the Quadratic Hedging Strategies

The aim of this section is to study the stability of the quadratic hedging strategies to the variation of the model, where we consider the two stock price models defined in (1) and (13). We study the stability of the RM strategy extensively and at the end of this section we come back to the MVH strategy. Since we work in the martingale setting, we first present some specific martingale measures which are commonly used in finance and in electricity markets. Then we discuss some common properties which are fulfilled by these measures. This is the topic of the next subsection.

3.1 Robustness of the Martingale Measures

Recall from the previous section that the martingale measures $\widetilde{\mathbb{P}}_{\Theta_0}$ and $\widetilde{\mathbb{P}}_{\Theta_\varepsilon}$ are determined via the functions $\rho(.\,;\Theta_0)$, $\rho(.\,;\Theta_\varepsilon)$ and the parameters Θ_0, Θ_ε, respectively. We present the following assumptions on these characteristics.

Assumptions 1 For Θ_0, Θ_ε, $\rho(.\,;\Theta_0)$, and $\rho(.\,;\Theta_\varepsilon)$ satisfying Eqs. (2)–(4), and Eqs. (15)–(18) we assume the following, where C denotes a positive constant and $\Theta \in \{\Theta_0, \Theta_\varepsilon\}$.

(i) Θ_0 and Θ_ε exist and are unique.

(ii) It holds that

$$|\Theta_0 - \Theta_\varepsilon| \le C\widetilde{G}^2(\varepsilon),$$

where $\widetilde{G}(\varepsilon) = \max(G(\varepsilon), \sigma(\varepsilon))$. Herein $\sigma(\varepsilon)$ equals the standard deviation of the jumps of L with size smaller than ε, i.e.

$$\sigma^2(\varepsilon) = \int_{\{|z|<\varepsilon\}} z^2 \ell(\mathrm{d}z).$$

(iii) On the other hand, Θ_ε is uniformly bounded in ε, i.e.

$$|\Theta_\varepsilon| \le C.$$

(iv) For all z in $\{|z| < 1\}$ it holds that

$$|\rho(z;\Theta)| \le C.$$

(v) We have

$$\int_{\{|z|\ge 1\}} \rho^4(z;\Theta)\ell(\mathrm{d}z) \le C.$$

(vi) It is guaranteed that

$$\int_{\mathbb{R}_0} \big(1 - \rho(z;\Theta)\big)^2 \ell(\mathrm{d}z) \le C.$$

(vii) It holds for $k \in \{2, 4\}$ that

$$\int_{\mathbb{R}_0} \big(\rho(z; \Theta_0) - \rho(z; \Theta_\varepsilon)\big)^k \ell(\mathrm{d}z) \le C\widetilde{G}^{2k}(\varepsilon).$$

Widely used martingale measures in the exponential Lévy setting are the Esscher transform (ET), minimal entropy martingale measure (MEMM), and minimal martingale measure (MMM), which are specified as follows.

- In order to define the ET we assume that

$$\int_{\{|z|\ge 1\}} \mathrm{e}^{\theta z} \ell(\mathrm{d}z) < \infty, \qquad \forall \theta \in \mathbb{R}. \tag{27}$$

  The Lévy measures under the ET are given in (3) and (17) where $\rho(z; \Theta) = \mathrm{e}^{\Theta z}$. The ET for the first model is then determined by the parameter Θ_0 satisfying (4). For the second model the ET corresponds to the solution Θ_ε of (18). See [13] for more details.
- Let us impose that

$$\int_{\{|z|\ge 1\}} \mathrm{e}^{\theta (\mathrm{e}^z - 1)} \ell(\mathrm{d}z) < \infty, \qquad \forall \theta \in \mathbb{R}, \tag{28}$$

  and that $\rho(z; \Theta) = \mathrm{e}^{\Theta(\mathrm{e}^z - 1)}$ in the Lévy measures. Then the solution Θ_0 of Eq. (4) determines the MEMM for the first model, and Θ_ε being the solution of (18) characterises the MEMM for the second model. The MEMM is studied in [12].
- Let us consider the assumption

$$\int_{\{z\ge 1\}} \mathrm{e}^{4z} \ell(\mathrm{d}z) < \infty. \tag{29}$$

  The MMM implies that $\rho(z; \Theta) = \Theta(\mathrm{e}^z - 1) - 1$ in the Lévy measures and the parameters Θ_0 and Θ_ε are the solutions of (4) and (18). More information about the MMM can be found in [1, 10].

In [4, 6] it was shown that the ET, the MEMM, and the MMM fulfill statements (i), (ii), (iii), and (iv) of Assumptions 1 in the exponential Lévy setting. The following proposition shows that items (v), (vi), and (vii) of Assumptions 1 also hold for these martingale measures.

Proposition 1 *The Lévy measures given in* (3) *and* (17) *and corresponding to the ET, MEMM, and MMM, satisfy (v), (vi), and (vii) of Assumptions* 1.

Proof Recall that the Lévy measure satisfies the following integrability conditions

$$\int_{\{|z|<1\}} z^2 \ell(\mathrm{d}z) < \infty \qquad \text{and} \qquad \int_{\{|z|\geq 1\}} \ell(\mathrm{d}z) < \infty. \tag{30}$$

We show that the statement holds for the considered martingale measures.

- Under the ET it holds for $\Theta \in \{\Theta_0, \Theta_\varepsilon\}$ that

$$\rho^4(z;\Theta) = \mathrm{e}^{4\Theta z} \leq \mathrm{e}^{4C|z|},$$

  because of (iii) in Assumptions 1. By the mean value theorem (MVT), there exists a number Θ' between 0 and Θ such that

$$\big(1-\rho(z;\Theta)\big)^2 = z^2 \mathrm{e}^{2\Theta' z}\Theta^2 \leq \left(1_{\{|z|<1\}}\mathrm{e}^{2C} z^2 + 1_{\{|z|\geq 1\}}\mathrm{e}^{(2C+2)z}\right) C,$$

  where we used again Assumptions 1 (iii). For $k \in \{2,4\}$, we derive via the MVT that

$$\big(\rho(z;\Theta_0)-\rho(z;\Theta_\varepsilon)\big)^k = \mathrm{e}^{k\Theta_0 z}\left(1-\mathrm{e}^{(\Theta_\varepsilon-\Theta_0)z}\right)^k = \mathrm{e}^{k\Theta_0 z} z^k \mathrm{e}^{k\Theta'' z}(\Theta_0-\Theta_\varepsilon)^k,$$

  where Θ'' is a number between 0 and $\Theta_\varepsilon - \Theta_0$. Assumptions 1 (ii) imply that

$$\big(\rho(z;\Theta_0)-\rho(z;\Theta_\varepsilon)\big)^k \leq \left(1_{\{|z|<1\}}\mathrm{e}^{k(|\Theta_0|+C)} z^2 + 1_{\{|z|\geq 1\}}\mathrm{e}^{k(\Theta_0+1+C)z}\right) C\widetilde{G}^{2k}(\varepsilon).$$

  The obtained inequalities and integrability conditions (27) and (30) prove the statement.
- Consider the MEMM and $\Theta \in \{\Theta_0, \Theta_\varepsilon\}$. We have

$$\rho^4(z;\Theta) = \mathrm{e}^{4\Theta(\mathrm{e}^z-1)} \leq \mathrm{e}^{4C|\mathrm{e}^z-1|},$$

  because of (iii) in Assumptions 1. The latter assumption and the MVT imply that

$$\begin{aligned}\big(1-\rho(z;\Theta)\big)^2 &= (\mathrm{e}^z-1)^2\mathrm{e}^{2\Theta'(\mathrm{e}^z-1)}\Theta^2 \\ &\leq \left(1_{\{|z|<1\}}\mathrm{e}^{2C(\mathrm{e}+1)+2} z^2 + 1_{\{|z|\geq 1\}}\mathrm{e}^{(2C+2)(\mathrm{e}^z-1)}\right) C.\end{aligned}$$

  We determine via the MVT and properties (ii) and (iii) in Assumptions 1 for $k \in \{2,4\}$ that

$$\begin{aligned}&\big(\rho(z;\Theta_0)-\rho(z;\Theta_\varepsilon)\big)^k \\ &= \mathrm{e}^{k\Theta_0(\mathrm{e}^z-1)}\left(1-\mathrm{e}^{(\Theta_\varepsilon-\Theta_0)(\mathrm{e}^z-1)}\right)^k \\ &= \mathrm{e}^{k\Theta_0(\mathrm{e}^z-1)}(\mathrm{e}^z-1)^k\mathrm{e}^{k\Theta''(\mathrm{e}^z-1)}(\Theta_0-\Theta_\varepsilon)^k \\ &\leq \left(1_{\{|z|<1\}}\mathrm{e}^{k(|\Theta_0|(\mathrm{e}+1)+1+C(\mathrm{e}+1))} z^2 + 1_{\{|z|\geq 1\}}\mathrm{e}^{k(\Theta_0+1+C)(\mathrm{e}^z-1)}\right) C\widetilde{G}^{2k}(\varepsilon).\end{aligned}$$

From (28) and (30) we conclude that (v), (vi), and (vii) in Assumptions 1 are in force.

- For the MMM we have

$$\rho^4(z;\Theta) = \big(\Theta(\mathrm{e}^z-1)-1\big)^4 \leq C(\mathrm{e}^{4z}+1).$$

Moreover it holds that

$$\big(1-\rho(z;\Theta)\big)^2 = (\mathrm{e}^z-1)^2\Theta^2 \leq \Big(1_{\{|z|<1\}}\mathrm{e}^2z^2 + 1_{\{|z|\geq 1\}}(\mathrm{e}^{2z}+1)\Big)\,C.$$

We get through (ii) and (iii) in Assumptions 1 that

$$\begin{aligned}\big(\rho(z;\Theta_0)-\rho(z;\Theta_\varepsilon)\big)^k &= (\mathrm{e}^z-1)^k(\Theta_0-\Theta_\varepsilon)^k\\ &\leq \Big(1_{\{|z|<1\}}\mathrm{e}^k z^2 + 1_{\{|z|\geq 1\}}(\mathrm{e}^{kz}+1)\Big)\,C\widetilde{G}^{2k}(\varepsilon),\end{aligned}$$

for $k\in\{2,4\}$. The proof is completed by involving conditions (29) and (30). □

3.2 Robustness of the BSDEJ

The aim of this subsection is to study the robustness of the BSDEJs (11) and (25). First, we prove the L^2-boundedness of the solution of the BSDEJ (11) in the following lemma.

Lemma 1 *Assume point (vi) from Assumptions 1. Let $(\hat{V}^{\Theta_0}, A^{\Theta_0}, B^{\Theta_0})$ be the solution of (11). Then we have for all $t\in[0,T]$*

$$\mathbb{E}\Big[\int_t^T(\hat{V}_s^{\Theta_0})^2\mathrm{d}s\Big]+\mathbb{E}\Big[\int_t^T(A_s^{\Theta_0})^2\mathrm{d}s\Big]+\mathbb{E}\Big[\int_t^T\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2\ell(\mathrm{d}z)\mathrm{d}s\Big]\leq C\mathbb{E}[\hat{H}_T^2],$$

where C represents a positive constant.

Proof Via (5) we rewrite the BSDEJ (11) as follows

$$\begin{aligned}\mathrm{d}\hat{V}_t^{\Theta_0} &= \Big(-b\Theta_0 A_t^{\Theta_0} + \int_{\mathbb{R}_0} B_t^{\Theta_0}(z)(1-\rho(z;\Theta_0))\ell(\mathrm{d}z)\Big)\mathrm{d}t\\ &\quad + A_t^{\Theta_0}\mathrm{d}W_t + \int_{\mathbb{R}_0} B_t^{\Theta_0}(z)\widetilde{N}(\mathrm{d}t,\mathrm{d}z).\end{aligned}$$

We apply the Itô formula to $\mathrm{e}^{\beta t}(\hat{V}_t^{\Theta_0})^2$ and find that

$$
\begin{aligned}
&\mathrm{d}\left(\mathrm{e}^{\beta t}(\hat{V}_t^{\Theta_0})^2\right)\\
&= \beta \mathrm{e}^{\beta t}(\hat{V}_t^{\Theta_0})^2 \mathrm{d}t + 2\mathrm{e}^{\beta t}\hat{V}_t^{\Theta_0}\left(-b\Theta_0 A_t^{\Theta_0} + \int_{\mathbb{R}_0} B_t^{\Theta_0}(z)(1-\rho(z;\Theta_0))\ell(\mathrm{d}z)\right)\mathrm{d}t\\
&\quad + 2\mathrm{e}^{\beta t}\hat{V}_t^{\Theta_0}A_t^{\Theta_0}\mathrm{d}W_t + \mathrm{e}^{\beta t}(A_t^{\Theta_0})^2\mathrm{d}t\\
&\quad + \int_{\mathbb{R}_0}\mathrm{e}^{\beta t}\left(\left(\hat{V}_{t-}^{\Theta_0} + B_t^{\Theta_0}(z)\right)^2 - (\hat{V}_{t-}^{\Theta_0})^2\right)\widetilde{N}(\mathrm{d}t,\mathrm{d}z) + \int_{\mathbb{R}_0}\mathrm{e}^{\beta t}(B_t^{\Theta_0}(z))^2\ell(\mathrm{d}z)\mathrm{d}t.
\end{aligned}
$$

By integration and taking the expectation we recover that

$$
\begin{aligned}
&\mathbb{E}\left[\mathrm{e}^{\beta t}(\hat{V}_t^{\Theta_0})^2\right]\\
&= \mathbb{E}\left[\mathrm{e}^{\beta T}(\hat{V}_T^{\Theta_0})^2\right] - \beta\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(\hat{V}_s^{\Theta_0})^2\mathrm{d}s\right]\\
&\quad - 2\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}\hat{V}_s^{\Theta_0}\left(-b\Theta_0 A_s^{\Theta_0} + \int_{\mathbb{R}_0} B_s^{\Theta_0}(z)(1-\rho(z;\Theta_0))\ell(\mathrm{d}z)\right)\mathrm{d}s\right] \qquad (31)\\
&\quad - \mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(A_s^{\Theta_0})^2\mathrm{d}s\right] - \mathbb{E}\left[\int_t^T\int_{\mathbb{R}_0}\mathrm{e}^{\beta s}(B_s^{\Theta_0}(z))^2\ell(\mathrm{d}z)\mathrm{d}s\right].
\end{aligned}
$$

Because of the properties

$$
\text{for all } a,b\in\mathbb{R} \text{ and } k\in\mathbb{R}_0^+ \text{ it holds that } \pm 2ab \le ka^2 + \frac{1}{k}b^2 \qquad (32)
$$

and

$$
\text{for all } n\in\mathbb{N} \text{ and for all } a_i\in\mathbb{R},\, i=1,\dots,n \text{ we have that } \left(\sum_{i=1}^n a_i\right)^2 \le n\sum_{i=1}^n a_i^2, \qquad (33)
$$

the third term in the right hand side of (31) is estimated by

$$
\begin{aligned}
&- 2\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}\hat{V}_s^{\Theta_0}\left(-b\Theta_0 A_s^{\Theta_0} + \int_{\mathbb{R}_0} B_s^{\Theta_0}(z)(1-\rho(z;\Theta_0))\ell(\mathrm{d}z)\right)\mathrm{d}s\right]\\
&\le \mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}\left\{k(\hat{V}_s^{\Theta_0})^2 + \frac{1}{k}\left(-b\Theta_0 A_s^{\Theta_0} + \int_{\mathbb{R}_0} B_s^{\Theta_0}(z)(1-\rho(z;\Theta_0))\ell(\mathrm{d}z)\right)^2\right\}\mathrm{d}s\right]\\
&\le k\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(\hat{V}_s^{\Theta_0})^2\mathrm{d}s\right] + \frac{2}{k}b^2{\Theta_0}^2\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(A_s^{\Theta_0})^2\mathrm{d}s\right]\\
&\quad + \frac{2}{k}\int_{\mathbb{R}_0}(1-\rho(z;\Theta_0))^2\,\ell(\mathrm{d}z)\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2\ell(\mathrm{d}z)\mathrm{d}s\right].
\end{aligned}
$$

Substituting the latter inequality in (31) leads to

$$\mathbb{E}\left[e^{\beta t}(\hat{V}_t^{\Theta_0})^2\right] + (\beta - k)\mathbb{E}\left[\int_t^T e^{\beta s}(\hat{V}_s^{\Theta_0})^2 ds\right] + \left(1 - \frac{2}{k}b^2{\Theta_0}^2\right)\mathbb{E}\left[\int_t^T e^{\beta s}(A_s^{\Theta_0})^2 ds\right]$$
$$+ \left(1 - \frac{2}{k}\int_{\mathbb{R}_0}(1 - \rho(z;\Theta_0))^2\,\ell(dz)\right)\mathbb{E}\left[\int_t^T e^{\beta s}\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2\ell(dz)ds\right]$$
$$\leq \mathbb{E}\left[e^{\beta T}(\hat{V}_T^{\Theta_0})^2\right]. \tag{34}$$

Let k guarantee that

$$1 - \frac{2}{k}b^2{\Theta_0}^2 \geq \frac{1}{2} \quad \text{and} \quad 1 - \frac{2}{k}\int_{\mathbb{R}_0}(1 - \rho(z;\Theta_0))^2\,\ell(dz) \geq \frac{1}{2}.$$

Hence we choose

$$k \geq 4\max\left(b^2{\Theta_0}^2, \int_{\mathbb{R}_0}(1 - \rho(z;\Theta_0))^2\,\ell(dz)\right) > 0,$$

which exists because of (vi) from Assumptions 1. Besides we assume that $\beta \geq k + \frac{1}{2} > 0$. Then for $s \in [0, T]$ it follows that $1 \leq e^{\beta s} \leq e^{\beta T}$ and from (34) we achieve

$$\mathbb{E}\left[\int_t^T(\hat{V}_s^{\Theta_0})^2 ds\right] + \mathbb{E}\left[\int_t^T(A_s^{\Theta_0})^2 ds\right] + \mathbb{E}\left[\int_t^T\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2\ell(dz)ds\right] \leq C\mathbb{E}[(\hat{V}_T^{\Theta_0})^2],$$

which proves the claim. □

In order to study the robustness of the BSDEJs (11) and (25), we consider both models under the enlarged filtration $\widetilde{\mathbb{F}}$ since we have for all $t \in [0, T]$ that $\mathscr{F}_t \subset \widetilde{\mathscr{F}}_t$. Let us define

$$\bar{V}^\varepsilon = \hat{V}^{\Theta_0} - \hat{V}^{\Theta_\varepsilon}, \quad \bar{A}^\varepsilon = A^{\Theta_0} - A^{\Theta_\varepsilon}, \quad \bar{B}^\varepsilon(z) = B^{\Theta_0}(z) - 1_{\{|z|\geq\varepsilon\}}B^{\Theta_\varepsilon}(z).$$

We derive from (5), (11), (19), and (25) that

$$d\bar{V}_t^\varepsilon = \alpha_t^\varepsilon dt + \bar{A}_t^\varepsilon dW_t + \int_{\mathbb{R}_0}\bar{B}_t^\varepsilon(z)\widetilde{N}(dt, dz) - C_t^{\Theta_\varepsilon}d\widetilde{W}_t, \tag{35}$$

where

$$\alpha^\varepsilon = -b(\Theta_0 A^{\Theta_0} - \Theta_\varepsilon A^{\Theta_\varepsilon}) + G(\varepsilon)\Theta_\varepsilon C^{\Theta_\varepsilon} \tag{36}$$
$$+ \int_{\mathbb{R}_0}\left(B^{\Theta_0}(z)\,(1 - \rho(z;\Theta_0)) - 1_{\{|z|\geq\varepsilon\}}B^{\Theta_\varepsilon}(z)\,(1 - \rho(z;\Theta_\varepsilon))\right)\ell(dz).$$

The process α^ε (36) plays a crucial role in the study of the robustness of the BSDEJ. In the following lemma we state an upper bound for this process in terms of the solutions of the BSDEJs.

Lemma 2 *Let Assumptions 1 hold true. Consider α^ε as defined in (36). For any $t \in [0, T]$ and $\beta \in \mathbb{R}$ it holds that*

$$\begin{aligned}
&\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(\alpha_s^\varepsilon)^2 \mathrm{d}s\right] \\
&\leq C\left(\widetilde{G}^4(\varepsilon)\left\{\mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(A_s^{\Theta_0})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right\}\right. \\
&\qquad + \mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(\bar{A}_s^\varepsilon)^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}\int_{\mathbb{R}_0}(\bar{B}_s^\varepsilon(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] \\
&\qquad \left. + \mathbb{E}\left[\int_t^T \mathrm{e}^{\beta s}(C_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right]\right),
\end{aligned}$$

where C is a positive constant.

Proof Parts (ii) and (iii) of Assumptions 1 imply that

$$\begin{aligned}
\left|-b\left(\Theta_0 A_s^{\Theta_0} - \Theta_\varepsilon A_s^{\Theta_\varepsilon}\right)\right| &\leq |b||\Theta_0 - \Theta_\varepsilon||A_s^{\Theta_0}| + |b||\Theta_\varepsilon||A_s^{\Theta_0} - A_s^{\Theta_\varepsilon}| \\
&\leq C\widetilde{G}^2(\varepsilon)|A_s^{\Theta_0}| + C|\bar{A}_s^\varepsilon|
\end{aligned}$$

and

$$|G(\varepsilon)\Theta_\varepsilon C_s^{\Theta_\varepsilon}| \leq C|C_s^{\Theta_\varepsilon}|.$$

From Hölder's inequality and Assumptions 1 (vi) and (vii) it follows that

$$\begin{aligned}
&\left|\int_{\mathbb{R}_0}\left(B_s^{\Theta_0}(z)\,(1-\rho(z;\Theta_0)) - 1_{\{|z|\geq\varepsilon\}}B_s^{\Theta_\varepsilon}(z)\,(1-\rho(z;\Theta_\varepsilon))\right)\ell(\mathrm{d}z)\right| \\
&\leq \left|\int_{\mathbb{R}_0} B_s^{\Theta_0}(z)\,(\rho(z;\Theta_0)-\rho(z;\Theta_\varepsilon))\,\ell(\mathrm{d}z)\right| + \left|\int_{\mathbb{R}_0}\bar{B}_s^\varepsilon(z)\,(1-\rho(z;\Theta_\varepsilon))\,\ell(\mathrm{d}z)\right| \\
&\leq \left(\int_{\mathbb{R}_0}\left(\rho(z;\Theta_0)-\rho(z;\Theta_\varepsilon)\right)^2\ell(\mathrm{d}z)\right)^{1/2}\left(\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2\ell(\mathrm{d}z)\right)^{1/2} \\
&\quad + \left(\int_{\mathbb{R}_0}\left(1-\rho(z;\Theta_\varepsilon)\right)^2\ell(\mathrm{d}z)\right)^{1/2}\left(\int_{\mathbb{R}_0}(\bar{B}_s^\varepsilon(z))^2\ell(\mathrm{d}z)\right)^{1/2} \\
&\leq C\widetilde{G}^2(\varepsilon)\left(\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2\ell(\mathrm{d}z)\right)^{1/2} + C\left(\int_{\mathbb{R}_0}(\bar{B}_s^\varepsilon(z))^2\ell(\mathrm{d}z)\right)^{1/2}.
\end{aligned}$$

We conclude that

$$(\alpha_s^\varepsilon)^2 \leq C\left(\widetilde{G}^4(\varepsilon)\left\{(A_s^{\Theta_0})^2 + \int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2\ell(\mathrm{d}z)\right\} + (\bar{A}_s^\varepsilon)^2 + \int_{\mathbb{R}_0}(\bar{B}_s^\varepsilon(z))^2\ell(\mathrm{d}z) + (C_s^{\Theta_\varepsilon})^2\right).$$

The statement is easily deduced from this inequality. □

With these two lemmas ready for use, we state and prove the main result of this subsection which is the robustness of the BSDEJs for the discounted portfolio value process of the RM strategy.

Theorem 1 *Assumptions 1 are in force. Let* $(\hat{V}^{\Theta_0}, A^{\Theta_0}, B^{\Theta_0})$ *be the solution of* (11) *and* $(\hat{V}^{\Theta_\varepsilon}, A^{\Theta_\varepsilon}, B^{\Theta_\varepsilon}, C^{\Theta_\varepsilon})$ *be the solution of* (25)*. For some positive constant C and any* $t \in [0, T]$ *we have*

$$\begin{aligned}
&\mathbb{E}\left[\int_t^T(\hat{V}_s^{\Theta_0} - \hat{V}_s^{\Theta_\varepsilon})^2\mathrm{d}s\right] + \mathbb{E}\left[\int_t^T(A_s^{\Theta_0} - A_s^{\Theta_\varepsilon})^2\mathrm{d}s\right]\\
&\quad + \mathbb{E}\left[\int_t^T\int_{\mathbb{R}_0}\left(B_s^{\Theta_0}(z) - 1_{\{|z|\geq\varepsilon\}}B_s^{\Theta_\varepsilon}(z)\right)^2\ell(\mathrm{d}z)\mathrm{d}s\right] + \mathbb{E}\left[\int_t^T(C_s^{\Theta_\varepsilon})^2\mathrm{d}s\right]\\
&\leq C\left(\mathbb{E}\left[(\hat{H}_T - \hat{H}_T^\varepsilon)^2\right] + \widetilde{G}^4(\varepsilon)\mathbb{E}[\hat{H}_T^2]\right).
\end{aligned}$$

Proof We apply the Itô formula to $\mathrm{e}^{\beta t}(\bar{V}_t^\varepsilon)^2$

$$\begin{aligned}
\mathrm{d}\left(\mathrm{e}^{\beta t}(\bar{V}_t^\varepsilon)^2\right) &= \beta\mathrm{e}^{\beta t}(\bar{V}_t^\varepsilon)^2\mathrm{d}t + 2\mathrm{e}^{\beta t}\bar{V}_t^\varepsilon\alpha_t^\varepsilon\mathrm{d}t + 2\mathrm{e}^{\beta t}\bar{V}_t^\varepsilon\bar{A}_t^\varepsilon\mathrm{d}W_t - 2\mathrm{e}^{\beta t}\bar{V}_t^\varepsilon C_t^{\Theta_\varepsilon}\mathrm{d}\widetilde{W}_t\\
&\quad + \mathrm{e}^{\beta t}(\bar{A}_t^\varepsilon)^2\mathrm{d}t + \mathrm{e}^{\beta t}(C_t^{\Theta_\varepsilon})^2\mathrm{d}t + \int_{\mathbb{R}_0}\mathrm{e}^{\beta t}(\bar{B}_t^\varepsilon(z))^2\ell(\mathrm{d}z)\mathrm{d}t\\
&\quad + \int_{\mathbb{R}_0}\mathrm{e}^{\beta t}\left((\bar{V}_{t-}^\varepsilon + \bar{B}_t^\varepsilon(z))^2 - (\bar{V}_{t-}^\varepsilon)^2\right)\widetilde{N}(\mathrm{d}t, \mathrm{d}z).
\end{aligned}$$

Integration over the interval $[t, T]$ and taking the expectation under $\mathbb{P}$ results into

$$\begin{aligned}
\mathbb{E}\left[\mathrm{e}^{\beta t}(\bar{V}_t^\varepsilon)^2\right] &= \mathbb{E}\left[\mathrm{e}^{\beta T}(\bar{V}_T^\varepsilon)^2\right] - \beta\mathbb{E}\left[\int_t^T\mathrm{e}^{\beta s}(\bar{V}_s^\varepsilon)^2\mathrm{d}s\right] - 2\mathbb{E}\left[\int_t^T\mathrm{e}^{\beta s}\bar{V}_s^\varepsilon\alpha_s^\varepsilon\mathrm{d}s\right]\\
&\quad - \mathbb{E}\left[\int_t^T\mathrm{e}^{\beta s}(\bar{A}_s^\varepsilon)^2\mathrm{d}s\right] - \mathbb{E}\left[\int_t^T\int_{\mathbb{R}_0}\mathrm{e}^{\beta s}(\bar{B}_s^\varepsilon(z))^2\ell(\mathrm{d}z)\mathrm{d}s\right]\\
&\quad - \mathbb{E}\left[\int_t^T\mathrm{e}^{\beta s}(C_s^{\Theta_\varepsilon})^2\mathrm{d}s\right],
\end{aligned}$$

or equivalently

$$
\begin{aligned}
&\mathbb{E}\left[e^{\beta t}(\bar{V}_t^{\varepsilon})^2\right] + \mathbb{E}\left[\int_t^T e^{\beta s}(\bar{A}_s^{\varepsilon})^2 \mathrm{d}s\right] \\
&\quad + \mathbb{E}\left[\int_t^T \int_{\mathbb{R}_0} e^{\beta s}(\bar{B}_s^{\varepsilon}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] + \mathbb{E}\left[\int_t^T e^{\beta s}(C_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \\
&= \mathbb{E}\left[e^{\beta T}(\bar{V}_T^{\varepsilon})^2\right] - \beta \mathbb{E}\left[\int_t^T e^{\beta s}(\bar{V}_s^{\varepsilon})^2 \mathrm{d}s\right] - 2\mathbb{E}\left[\int_t^T e^{\beta s}\bar{V}_s^{\varepsilon}\alpha_s^{\varepsilon} \mathrm{d}s\right] \\
&\leq \mathbb{E}\left[e^{\beta T}(\bar{V}_T^{\varepsilon})^2\right] + (k-\beta)\mathbb{E}\left[\int_t^T e^{\beta s}(\bar{V}_s^{\varepsilon})^2 \mathrm{d}s\right] + \frac{1}{k}\mathbb{E}\left[\int_t^T e^{\beta s}(\alpha_s^{\varepsilon})^2 \mathrm{d}s\right], \quad (37)
\end{aligned}
$$

where we used property (32). The combination of (37) with Lemma 2 provides

$$
\begin{aligned}
&\mathbb{E}\left[e^{\beta t}(\bar{V}_t^{\varepsilon})^2\right] + (\beta-k)\mathbb{E}\left[\int_t^T e^{\beta s}(\bar{V}_s^{\varepsilon})^2 \mathrm{d}s\right] + \left(1-\frac{C}{k}\right)\mathbb{E}\left[\int_t^T e^{\beta s}(\bar{A}_s^{\varepsilon})^2 \mathrm{d}s\right] \\
&+ \left(1-\frac{C}{k}\right)\mathbb{E}\left[\int_t^T e^{\beta s}\int_{\mathbb{R}_0}(\bar{B}_s^{\Theta_\varepsilon}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] + \left(1-\frac{C}{k}\right)\mathbb{E}\left[\int_t^T e^{\beta s}(C_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \\
&\leq \mathbb{E}\left[e^{\beta T}(\bar{V}_T^{\varepsilon})^2\right] + \frac{C}{k}\widetilde{G}^4(\varepsilon)\left\{\mathbb{E}\left[\int_t^T e^{\beta s}(A_s^{\Theta_0})^2 \mathrm{d}s\right]\right. \quad (38) \\
&\quad \left. + \mathbb{E}\left[\int_t^T e^{\beta s}\int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right\}.
\end{aligned}
$$

Let us choose k and β such that $1-\frac{C}{k} \geq \frac{1}{2}$ and $\beta - k \geq \frac{1}{2}$. This means we choose $k \geq 2C > 0$ and $\beta \geq \frac{1}{2} + k > 0$. Thus for any $s \in [t, T]$ it holds that $1 < e^{\beta s} \leq e^{\beta T}$. We derive from (38) that

$$
\begin{aligned}
&\mathbb{E}\left[\int_t^T (\bar{V}_s^{\varepsilon})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_t^T (\bar{A}_s^{\varepsilon})^2 \mathrm{d}s\right] \\
&\quad + \mathbb{E}\left[\int_t^T \int_{\mathbb{R}_0}(\bar{B}_s^{\Theta_\varepsilon}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] + \mathbb{E}\left[\int_t^T (C_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \\
&\leq C\left(\mathbb{E}\left[(\bar{V}_T^{\varepsilon})^2\right] + \widetilde{G}^4(\varepsilon)\left\{\mathbb{E}\left[\int_t^T (A_s^{\Theta_0})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_t^T \int_{\mathbb{R}_0}(B_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right\}\right).
\end{aligned}
$$

By Lemma 1 we conclude the proof. □

This main result leads to the following theorem concerning the robustness of the discounted portfolio value process of the RM strategy.

Theorem 2 *Assume Assumptions 1. Let $\hat{V}^{\Theta_0}$, $\hat{V}^{\Theta_\varepsilon}$ be part of the solution of* (11), (25) *respectively. Then we have*

$$\mathbb{E}\left[\sup_{0\le t\le T}(\hat{V}_s^{\Theta_0}-\hat{V}_s^{\Theta_\varepsilon})^2\right]\le C\left(\mathbb{E}[(\hat{H}_T-\hat{H}_T^\varepsilon)^2]+\widetilde{G}^4(\varepsilon)\mathbb{E}[\hat{H}_T^2]\right),$$

for a positive constant C.

Proof Integration of the BSDEJ (35) results into

$$\bar{V}_t^\varepsilon=\bar{V}_T^\varepsilon-\int_t^T\alpha_s^\varepsilon \mathrm{d}s-\int_t^T\bar{A}_s^\varepsilon \mathrm{d}W_s-\int_t^T\int_{\mathbb{R}_0}\bar{B}_s^\varepsilon(z)\widetilde{N}(\mathrm{d}s,\mathrm{d}z)+\int_t^T C_s^{\Theta_\varepsilon}\mathrm{d}\widetilde{W}_s.$$

By property (33) we arrive at

$$\begin{aligned}&\mathbb{E}\left[\sup_{0\le t\le T}(\bar{V}_t^\varepsilon)^2\right]\\&\le 5\left(\mathbb{E}\left[(\bar{V}_T^\varepsilon)^2\right]+\mathbb{E}\left[\int_0^T(\alpha_s^\varepsilon)^2\mathrm{d}s\right]+\mathbb{E}\left[\sup_{0\le t\le T}\left(\int_t^T\bar{A}_s^\varepsilon\mathrm{d}W_s\right)^2\right]\right.\\&\qquad\left.+\mathbb{E}\left[\sup_{0\le t\le T}\left(\int_t^T\int_{\mathbb{R}_0}\bar{B}_s^\varepsilon(z)\widetilde{N}(\mathrm{d}s,\mathrm{d}z)\right)^2\right]+\mathbb{E}\left[\sup_{0\le t\le T}\left(\int_t^T C_s^{\Theta_\varepsilon}\mathrm{d}\widetilde{W}_s\right)^2\right]\right).\end{aligned}$$

Burkholder's inequality (see e.g., Theorem 3.28 in [15]) guarantees the existence of a positive constant C such that

$$\begin{aligned}&\mathbb{E}\left[\sup_{0\le t\le T}\left(\int_t^T\bar{A}_s^\varepsilon\mathrm{d}W_s\right)^2\right]\le C\mathbb{E}\left[\int_0^T(\bar{A}_s^\varepsilon)^2\mathrm{d}s\right],\\&\mathbb{E}\left[\sup_{0\le t\le T}\left(\int_t^T\int_{\mathbb{R}_0}\bar{B}_s^\varepsilon(z)\widetilde{N}(\mathrm{d}s,\mathrm{d}z)\right)^2\right]\le C\mathbb{E}\left[\int_0^T\int_{\mathbb{R}_0}(\bar{B}_s^\varepsilon(z))^2\ell(\mathrm{d}z)\mathrm{d}s\right],\\&\mathbb{E}\left[\sup_{0\le t\le T}\left(\int_t^T C_s^{\Theta_\varepsilon}\mathrm{d}\widetilde{W}_s\right)^2\right]\le C\mathbb{E}\left[\int_0^T(C_s^{\Theta_\varepsilon})^2\mathrm{d}s\right].\end{aligned}$$

Application of Lemma 2 for $t=0$, $\beta=0$, Lemma 1, and Theorem 1 completes the proof. □

3.3 Robustness of the Risk-Minimising Strategy

Theorem 2 in the previous subsection concerns the robustness result of the value process of the discounted portfolio in the RM strategy. Before we present the stability of the amount of wealth in the RM strategy, we study the relation between $\hat{\pi}^{\Theta_0}$ (resp. $\hat{\pi}^{\Theta_\varepsilon}$) and the solution of the BSDEJ of type (11) (resp. (25)) in the first (resp. second) model. Consider the processes A^{Θ_0} and $B^{\Theta_0}(z)$ defined in (12), then it holds that

$$\begin{aligned}
&A^{\Theta_0} b + \int_{\mathbb{R}_0} B^{\Theta_0}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_0)\ell(\mathrm{d}z) \\
&= \hat{\pi}^{\Theta_0} b^2 + X^{\Theta_0} b + \int_{\mathbb{R}_0} \left(\hat{\pi}^{\Theta_0}(\mathrm{e}^z - 1)^2 \rho(z;\Theta_0) + Y^{\Theta_0}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_0)\right)\ell(\mathrm{d}z) \\
&= \hat{\pi}^{\Theta_0}\left\{b^2 + \int_{\mathbb{R}_0}(\mathrm{e}^z - 1)^2\rho(z;\Theta_0)\ell(\mathrm{d}z)\right\} \\
&\quad + X^{\Theta_0} b + \int_{\mathbb{R}_0} Y^{\Theta_0}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_0)\ell(\mathrm{d}z).
\end{aligned}$$

From property (10) we attain that

$$\hat{\pi}^{\Theta_0} = \frac{1}{\kappa_0}\left(A^{\Theta_0} b + \int_{\mathbb{R}_0} B^{\Theta_0}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_0)\ell(\mathrm{d}z)\right), \tag{39}$$

where $\kappa_0 = b^2 + \int_{\mathbb{R}_0}(\mathrm{e}^z - 1)^2\rho(z;\Theta_0)\ell(\mathrm{d}z)$. Similarly for the second setting we have for the processes A^{Θ_ε}, $B^{\Theta_\varepsilon}(z)$, and C^{Θ_ε} defined in (26) that

$$\begin{aligned}
&A^{\Theta_\varepsilon} b + \int_{\{|z|\geq\varepsilon\}} B^{\Theta_\varepsilon}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_\varepsilon)\ell(\mathrm{d}z) + C^{\Theta_\varepsilon} G(\varepsilon) \\
&= \hat{\pi}^{\Theta_\varepsilon}\left\{b^2 + \int_{\{|z|\geq\varepsilon\}}(\mathrm{e}^z - 1)^2\rho(z;\Theta_\varepsilon)\ell(\mathrm{d}z) + G^2(\varepsilon)\right\} \\
&\quad + X^{\Theta_\varepsilon} b + \int_{\{|z|\geq\varepsilon\}} Y^{\Theta_\varepsilon}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_\varepsilon)\ell(\mathrm{d}z) + Z^{\Theta_\varepsilon} G(\varepsilon).
\end{aligned}$$

Property (24) leads to

$$\hat{\pi}^{\Theta_\varepsilon} = \frac{1}{\kappa_\varepsilon}\left(A^{\Theta_\varepsilon} b + \int_{\{|z|\geq\varepsilon\}} B^{\Theta_\varepsilon}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_\varepsilon)\ell(\mathrm{d}z) + C^{\Theta_\varepsilon} G(\varepsilon)\right), \tag{40}$$

where $\kappa_\varepsilon = b^2 + \int_{\{|z|\geq\varepsilon\}}(\mathrm{e}^z - 1)^2\rho(z;\Theta_\varepsilon)\ell(\mathrm{d}z) + G^2(\varepsilon)$.

We introduce the following additional assumption on the Lévy measure which we need for the robustness results studied later.

Assumption 2 For the Lévy measure ℓ the following integrability condition holds

$$\int_{\{z\geq 1\}} \mathrm{e}^{4z}\ell(\mathrm{d}z) < \infty.$$

Note that the latter assumption, combined with (30), implies for $k \in \{2, 4\}$ that

$$\int_{\mathbb{R}_0} (e^z - 1)^k \ell(dz) \le C \left(\int_{\{|z|<1\}} z^2 \ell(dz) + \int_{\{|z|\ge 1\}} \ell(dz) + \int_{\{z \ge 1\}} e^{4z} \ell(dz) \right) < \infty. \tag{41}$$

Moreover Assumption 2 is fulfilled for the considered martingale measures described in Sect. 3.1. Indeed, consider the ET, applying (27) for $\theta = 4$ and restricting the integral over $\{z \ge 1\}$ implies Assumption 2. On the set $\{z \ge 1\}$ it holds that $z \le e^z - 1$ and therefore Assumption 2 follows from (28) by choosing $\theta = 4$. For the MMM, condition (29) corresponds exactly to Assumption 2.

Theorem 3 *Impose Assumptions 1 and 2. Let the processes $\hat{\pi}^{\Theta_0}$ and $\hat{\pi}^{\Theta_\varepsilon}$ denote the amounts of wealth in a RM strategy. There is a positive constant C such that for any $t \in [0, T]$*

$$\mathbb{E}\left[\int_t^T (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 ds \right] \le C \left(\mathbb{E}[(\hat{H}_T - \hat{H}_T^\varepsilon)^2] + \widetilde{G}^4(\varepsilon) \mathbb{E}[\hat{H}_T^2] \right).$$

Proof Consider the amounts of wealth in (39) and (40). Let us denote $\hat{\pi}^{\Theta_0} = \frac{1}{\kappa_0} \Upsilon^0$ and $\hat{\pi}^{\Theta_\varepsilon} = \frac{1}{\kappa_\varepsilon} \Upsilon^\varepsilon$. Then it holds that

$$\left(\hat{\pi}^{\Theta_0} - \hat{\pi}^{\Theta_\varepsilon} \right)^2 \le 2 \left(\left(\frac{\kappa_0 - \kappa_\varepsilon}{\kappa_0 \kappa_\varepsilon} \right)^2 (\Upsilon^0)^2 + \frac{1}{\kappa_\varepsilon^2} (\Upsilon^0 - \Upsilon^\varepsilon)^2 \right).$$

Herein we have because of the Hölder's inequality, (14), (41), and properties (iv) and (vii) in Assumptions 1 that

$$\begin{aligned}
\left(\frac{\kappa_0 - \kappa_\varepsilon}{\kappa_0 \kappa_\varepsilon} \right)^2 &\le \frac{3}{b^8} \left(\left(\int_{\{|z|<\varepsilon\}} (e^z - 1)^2 \rho(z; \Theta_0) \ell(dz) \right)^2 \right. \\
&\quad \left. + \left(\int_{\{|z|\ge\varepsilon\}} (e^z - 1)^2 (\rho(z; \Theta_0) - \rho(z; \Theta_\varepsilon)) \ell(dz) \right)^2 + G^4(\varepsilon) \right) \\
&\le \frac{3}{b^8} \left(C \left(\int_{\{|z|<\varepsilon\}} (e^z - 1)^2 \ell(dz) \right)^2 \right. \\
&\quad \left. + \int_{\mathbb{R}_0} (e^z - 1)^4 \ell(dz) \int_{\mathbb{R}_0} (\rho(z; \Theta_0) - \rho(z; \Theta_\varepsilon))^2 \ell(dz) + G^4(\varepsilon) \right) \\
&\le C \widetilde{G}^4(\varepsilon).
\end{aligned}$$

On the other hand it is clear from (39) and (40) that

$$
\begin{aligned}
&(\Upsilon^0 - \Upsilon^\varepsilon)^2 \\
&\le 3 \big((\bar{A}^\varepsilon)^2 b^2 + (C^{\Theta_\varepsilon})^2 G^2(\varepsilon) \\
&\qquad + \left(\int_{\mathbb{R}_0} (B^{\Theta_0}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_0) - 1_{\{|z|\ge\varepsilon\}} B^{\Theta_\varepsilon}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_\varepsilon))\ell(\mathrm{d}z) \right)^2 \Bigg).
\end{aligned}
$$

Herein we derive via Hölder's inequality, (30), (41), and points (iv), (v), and (vii) in Assumptions 1 that

$$
\begin{aligned}
&\left(\int_{\mathbb{R}_0} (B^{\Theta_0}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_0) - 1_{\{|z|\ge\varepsilon\}} B^{\Theta_\varepsilon}(z)(\mathrm{e}^z - 1)\rho(z;\Theta_\varepsilon))\ell(\mathrm{d}z) \right)^2 \\
&= \left(\int_{\mathbb{R}_0} (B^{\Theta_0}(z)(\rho(z;\Theta_0) - \rho(z;\Theta_\varepsilon))(\mathrm{e}^z - 1) + \bar{B}^\varepsilon(z)\rho(z;\Theta_\varepsilon)(\mathrm{e}^z - 1))\ell(\mathrm{d}z) \right)^2 \\
&\le \int_{\mathbb{R}_0} (B^{\Theta_0}(z))^2 \ell(\mathrm{d}z) \int_{\mathbb{R}_0} (\rho(z;\Theta_0) - \rho(z;\Theta_\varepsilon))^2 (\mathrm{e}^z - 1)^2 \ell(\mathrm{d}z) \\
&\quad + \int_{\mathbb{R}_0} (\bar{B}^\varepsilon(z))^2 \ell(\mathrm{d}z) \int_{\mathbb{R}_0} \rho^2(z;\Theta_\varepsilon)(\mathrm{e}^z - 1)^2 \ell(\mathrm{d}z) \\
&\le \int_{\mathbb{R}_0} (B^{\Theta_0}(z))^2 \ell(\mathrm{d}z) \left(\int_{\mathbb{R}_0} (\rho(z;\Theta_0) - \rho(z;\Theta_\varepsilon))^4 \ell(\mathrm{d}z) \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}_0} (\mathrm{e}^z - 1)^4 \ell(\mathrm{d}z) \right)^{\frac{1}{2}} \\
&\quad + \int_{\mathbb{R}_0} (\bar{B}^\varepsilon(z))^2 \ell(\mathrm{d}z) \left(\int_{\{|z|\ge 1\}} \rho^4(z;\Theta_\varepsilon)\ell(\mathrm{d}z) \int_{\{|z|\ge 1\}} (\mathrm{e}^z - 1)^4 \ell(\mathrm{d}z) \right)^{\frac{1}{2}} \\
&\quad + C \int_{\mathbb{R}_0} (\bar{B}^\varepsilon(z))^2 \ell(\mathrm{d}z) \int_{\{|z|<1\}} z^2 \ell(\mathrm{d}z) \\
&\le C \widetilde{G}^4(\varepsilon) \int_{\mathbb{R}_0} (B^{\Theta_0}(z))^2 \ell(\mathrm{d}z) + C \int_{\mathbb{R}_0} (\bar{B}^\varepsilon(z))^2 \ell(\mathrm{d}z).
\end{aligned}
$$

The results above show that

$$
\begin{aligned}
\left(\hat{\pi}_t^{\Theta_0} - \hat{\pi}_t^{\Theta_\varepsilon} \right)^2 \le C \Bigg(&(\bar{A}_t^\varepsilon)^2 + \int_{\mathbb{R}_0} (\bar{B}_t^\varepsilon(z))^2 \ell(\mathrm{d}z) + (C_t^{\Theta_\varepsilon})^2 \\
&+ \widetilde{G}^4(\varepsilon) \left\{ (A_t^{\Theta_0})^2 + \int_{\mathbb{R}_0} (B_t^{\Theta_0}(z))^2 \ell(\mathrm{d}z) \right\} \Bigg).
\end{aligned}
$$

Therefore

$$
\begin{aligned}
&\mathbb{E}\left[\int_t^T (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \\
&\leq C \Bigg(\mathbb{E}\left[\int_t^T (\bar{A}_s^\varepsilon)^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_t^T \int_{\mathbb{R}_0} (\bar{B}_s^\varepsilon(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] + \mathbb{E}\left[\int_t^T (C_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \\
&\qquad + \widetilde{G}^4(\varepsilon)\left\{\mathbb{E}\left[\int_t^T (A_s^{\Theta_0})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_t^T \int_{\mathbb{R}_0} (B_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right\}\Bigg).
\end{aligned}
$$

By Lemma 1 and Theorem 1 we conclude the proof. $\square$

The trading in the risky assets is gathered in the gain processes defined by $\hat{G}_t^{\Theta_0} = \int_0^t \xi_s^{\Theta_0} \mathrm{d}\hat{S}_s$ and $\hat{G}_t^{\Theta_\varepsilon} = \int_0^t \xi_s^{\Theta_\varepsilon} \mathrm{d}\hat{S}_s^\varepsilon$. The following theorem shows the robustness of this gain process.

Theorem 4 *Under Assumptions 1 and 2, there exists a positive constant C such that for any $t \in [0, T]$*

$$
\mathbb{E}\left[\left(\hat{G}_t^{\Theta_0} - \hat{G}_t^{\Theta_\varepsilon}\right)^2\right] \leq C \left(\mathbb{E}[(\hat{H}_T - \hat{H}_T^\varepsilon)^2] + \widetilde{G}^2(\varepsilon)\mathbb{E}[\hat{H}_T^2]\right).
$$

Proof From (5) and (6) we know that

$$
\begin{aligned}
\xi_s^{\Theta_0} \mathrm{d}\hat{S}_s &= \xi_s^{\Theta_0} \hat{S}_s b \mathrm{d}W_s^{\Theta_0} + \xi_s^{\Theta_0} \hat{S}_s \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)\widetilde{N}^{\Theta_0}(\mathrm{d}s, \mathrm{d}z) \\
&= \hat{\pi}_s^{\Theta_0} \Bigg(\left(-b^2 \Theta_0 + \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)\,(1 - \rho(z; \Theta_0))\,\ell(\mathrm{d}z)\right) \mathrm{d}s \\
&\qquad + b\mathrm{d}W_s + \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)\widetilde{N}(\mathrm{d}s, \mathrm{d}z)\Bigg).
\end{aligned}
$$

In the other setting we have from (19) and (20) that

$$
\begin{aligned}
\xi_s^{\Theta_\varepsilon} \mathrm{d}\hat{S}_s^\varepsilon &= \xi_s^{\Theta_\varepsilon} \hat{S}_s^\varepsilon b \mathrm{d}W_s^{\Theta_\varepsilon} + \xi_s^{\Theta_\varepsilon} \hat{S}_s^\varepsilon \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\widetilde{N}^{\Theta_\varepsilon}(\mathrm{d}s, \mathrm{d}z) + \xi_s^{\Theta_\varepsilon} \hat{S}_s^\varepsilon G(\varepsilon)\mathrm{d}\widetilde{W}_s^{\Theta_\varepsilon} \\
&= \hat{\pi}_s^{\Theta_\varepsilon} \Bigg(\left(-b^2 \Theta_\varepsilon + \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\,(1 - \rho(z; \Theta_\varepsilon))\,\ell(\mathrm{d}z) - G^2(\varepsilon)\Theta_\varepsilon\right) \mathrm{d}s \\
&\qquad + b\mathrm{d}W_s + \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\widetilde{N}(\mathrm{d}s, \mathrm{d}z) + G(\varepsilon)\mathrm{d}\widetilde{W}_s\Bigg).
\end{aligned}
$$

We derive from the previous SDEs that

$$
\begin{aligned}
&\hat{G}_t^{\Theta_0} - \hat{G}_t^{\Theta_\varepsilon} = \int_0^t \xi_s^{\Theta_0} \mathrm{d}\hat{S}_s - \int_0^t \xi_s^{\Theta_\varepsilon} \mathrm{d}\hat{S}_s^\varepsilon \\
&= \left(-b^2\Theta_0 + \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)\,(1 - \rho(z;\Theta_0))\,\ell(\mathrm{d}z)\right) \int_0^t \hat{\pi}_s^{\Theta_0} \mathrm{d}s \\
&\quad - \left(-b^2\Theta_\varepsilon + \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\,(1 - \rho(z;\Theta_\varepsilon))\,\ell(\mathrm{d}z) - G^2(\varepsilon)\Theta_\varepsilon\right) \int_0^t \hat{\pi}_s^{\Theta_\varepsilon} \mathrm{d}s \\
&\quad + b\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon}) \mathrm{d}W_s + \int_0^t \int_{\mathbb{R}_0} \left(\hat{\pi}_s^{\Theta_0}(\mathrm{e}^z - 1) - \hat{\pi}_s^{\Theta_\varepsilon} 1_{\{|z|\geq\varepsilon\}}(\mathrm{e}^z - 1)\right) \widetilde{N}(\mathrm{d}s, \mathrm{d}z) \\
&\quad - G(\varepsilon) \int_0^t \hat{\pi}_s^{\Theta_\varepsilon} \mathrm{d}\widetilde{W}_s.
\end{aligned}
$$

Via the Cauchy-Schwartz inequality and the Itô isometry we obtain that

$$
\begin{aligned}
&\mathbb{E}\left[\left(\hat{G}_t^{\Theta_0} - \hat{G}_t^{\Theta_\varepsilon}\right)^2\right] \\
&\leq C\left(\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0})^2 \mathrm{d}s\right] \left\{\left(-b^2\Theta_0 + \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)\,(1 - \rho(z;\Theta_0))\,\ell(\mathrm{d}z)\right)\right.\right. \\
&\quad \left. - \left(-b^2\Theta_\varepsilon + \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\,(1 - \rho(z;\Theta_\varepsilon))\,\ell(\mathrm{d}z) - G^2(\varepsilon)\Theta_\varepsilon\right)\right\}^2 \\
&\quad + \mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \\
&\quad \times \left(-b^2\Theta_\varepsilon + \int_{\{|z|\geq\varepsilon\}} (\mathrm{e}^z - 1)\,(1 - \rho(z;\Theta_\varepsilon))\,\ell(\mathrm{d}z) - G^2(\varepsilon)\Theta_\varepsilon\right)^2 \\
&\quad + b^2\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] + G^2(\varepsilon)\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \\
&\quad \left. + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \left(\hat{\pi}_s^{\Theta_0}(\mathrm{e}^z - 1) - \hat{\pi}_s^{\Theta_\varepsilon} 1_{\{|z|\geq\varepsilon\}}(\mathrm{e}^z - 1)\right)^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right),
\end{aligned}
$$

wherein

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \left(\hat{\pi}_s^{\Theta_0}(\mathrm{e}^z - 1) - \hat{\pi}_s^{\Theta_\varepsilon} 1_{\{|z|\geq\varepsilon\}}(\mathrm{e}^z - 1)\right)^2 \ell(\mathrm{d}z)\mathrm{d}s\right] \\
&\leq 2\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} \left((\hat{\pi}_s^{\Theta_0})^2(\mathrm{e}^z - 1)^2 1_{\{|z|<\varepsilon\}} + (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2(\mathrm{e}^z - 1)^2 1_{\{|z|\geq\varepsilon\}}\right) \ell(\mathrm{d}z)\mathrm{d}s\right] \\
&\leq 2\left(\int_{\{|z|<\varepsilon\}} (\mathrm{e}^z - 1)^2 \ell(\mathrm{d}z)\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0})^2 \mathrm{d}s\right]\right. \\
&\qquad \left. + \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)^2 \ell(\mathrm{d}z)\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right]\right),
\end{aligned}
$$

and

$$\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] \leq 2\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] + 2\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0})^2 \mathrm{d}s\right].$$

By relation (14), Assumptions 1, (39), (41), Lemma 1, and Theorem 3 we prove the statement. □

The following result shows the robustness of the process $\mathscr{L}^{\Theta}$ appearing in the GKW-decomposition. This plays an important role in the stability of the cost process of the RM strategy.

Theorem 5 *Let Assumptions 1 and 2 hold true. Let the processes $\mathscr{L}^{\Theta_0}$ and $\mathscr{L}^{\Theta_\varepsilon}$ be as in (9) and (23), respectively. For any $t \in [0, T]$ it holds that*

$$\mathbb{E}[(\mathscr{L}_t^{\Theta_0} - \mathscr{L}_t^{\Theta_\varepsilon})^2] \leq C\left(\mathbb{E}[(\hat{H}_T - \hat{H}_T^\varepsilon)^2] + \widetilde{G}^2(\varepsilon)\mathbb{E}[\hat{H}_T^2]\right),$$

for a positive constant C.

Proof By (5) we can rewrite (9) as

$$\begin{aligned}\mathrm{d}\mathscr{L}_t^{\Theta_0} &= \left(-b\Theta_0 X_t^{\Theta_0} + \int_{\mathbb{R}_0} Y_t^{\Theta_0}(z)(1-\rho(z;\Theta_0))\ell(\mathrm{d}z)\right)\mathrm{d}t \\ &\quad + X_t^{\Theta_0}\mathrm{d}W_t + \int_{\mathbb{R}_0} Y_t^{\Theta_0}(z)\widetilde{N}(\mathrm{d}t,\mathrm{d}z).\end{aligned}$$

and similarly by (19) we obtain for (23)

$$\begin{aligned}\mathrm{d}\mathscr{L}_t^{\Theta_\varepsilon} &= \left(-b\Theta_\varepsilon X_t^{\Theta_\varepsilon} + \int_{\{|z|\geq\varepsilon\}} Y_t^{\Theta_\varepsilon}(z)(1-\rho(z;\Theta_\varepsilon))\ell(\mathrm{d}z) - G(\varepsilon)\Theta_\varepsilon Z_t^{\Theta_\varepsilon}\right)\mathrm{d}t \\ &\quad + X_t^{\Theta_\varepsilon}\mathrm{d}W_t + \int_{\{|z|\geq\varepsilon\}} Y_t^{\Theta_\varepsilon}(z)\widetilde{N}(\mathrm{d}t,\mathrm{d}z) + Z_t^{\Theta_\varepsilon}\mathrm{d}\widetilde{W}_t.\end{aligned}$$

Hence we recover that

$$\mathrm{d}(\mathscr{L}_t^{\Theta_0} - \mathscr{L}_t^{\Theta_\varepsilon}) = \gamma_t^\varepsilon \mathrm{d}t + \bar{X}_t^\varepsilon \mathrm{d}W_t + \int_{\mathbb{R}_0} \bar{Y}_t^\varepsilon(z)\widetilde{N}(\mathrm{d}t,\mathrm{d}z) - Z_t^{\Theta_\varepsilon}\mathrm{d}\widetilde{W}_t,$$

where

$$\begin{aligned}\gamma^\varepsilon &= -b(\Theta_0 X^{\Theta_0} - \Theta_\varepsilon X^{\Theta_\varepsilon}) + G(\varepsilon)\Theta_\varepsilon Z^{\Theta_\varepsilon} \\ &\quad + \int_{\mathbb{R}_0} \left(Y^{\Theta_0}(z)(1-\rho(z;\Theta_0)) - 1_{\{|z|\geq\varepsilon\}}Y^{\Theta_\varepsilon}(z)(1-\rho(z;\Theta_\varepsilon))\right)\ell(\mathrm{d}z), \\ \bar{X}^\varepsilon &= X^{\Theta_0} - X^{\Theta_\varepsilon}, \\ \bar{Y}^\varepsilon(z) &= Y^{\Theta_0}(z) - 1_{\{|z|\geq\varepsilon\}}Y^{\Theta_\varepsilon}(z).\end{aligned}$$

By integration over $[0, t]$ and taking the square we retrieve using (33) that

$$(\mathscr{L}_t^{\Theta_0} - \mathscr{L}_t^{\Theta_\varepsilon})^2 \le C\Bigg(\bigg(\int_0^t \gamma_s^\varepsilon \mathrm{d}s\bigg)^2 + \bigg(\int_0^t \bar{X}_s^\varepsilon \mathrm{d}W_s\bigg)^2 + \bigg(\int_0^t \int_{\mathbb{R}_0} \bar{Y}_s^\varepsilon(z)\widetilde{N}(\mathrm{d}s, \mathrm{d}z)\bigg)^2 + \bigg(\int_0^t Z_s^{\Theta_\varepsilon} \mathrm{d}\widetilde{W}_s\bigg)^2\Bigg).$$

Via the Cauchy-Schwartz inequality and the Itô isometry it follows that

$$\mathbb{E}[(\mathscr{L}_t^{\Theta_0} - \mathscr{L}_t^{\Theta_\varepsilon})^2] \le C\Bigg(\mathbb{E}\bigg[\int_0^t (\gamma_s^\varepsilon)^2 \mathrm{d}s\bigg] + \mathbb{E}\bigg[\int_0^t (\bar{X}_s^\varepsilon)^2 \mathrm{d}s\bigg] + \mathbb{E}\bigg[\int_0^t \int_{\mathbb{R}_0} (\bar{Y}_s^\varepsilon(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\bigg] + \mathbb{E}\bigg[\int_0^t (Z_s^{\Theta_\varepsilon})^2 \mathrm{d}s\bigg]\Bigg).$$

Concerning the term $\mathbb{E}\left[\int_0^t (\gamma_s^\varepsilon)^2 \mathrm{d}s\right]$ we derive through (ii) and (iii) in Assumptions 1 that

$$\begin{aligned}
&\mathbb{E}\bigg[\int_0^t \left(\Theta_0 X_s^{\Theta_0} - \Theta_\varepsilon X_s^{\Theta_\varepsilon}\right)^2 \mathrm{d}s\bigg] \\
&\le 2\Bigg(\mathbb{E}\bigg[\int_0^t (\Theta_0 - \Theta_\varepsilon)^2 (X_s^{\Theta_0})^2 \mathrm{d}s\bigg] + \mathbb{E}\bigg[\int_0^t {\Theta_\varepsilon}^2 (X_s^{\Theta_0} - X_s^{\Theta_\varepsilon})^2 \mathrm{d}s\bigg]\Bigg) \\
&\le C\Bigg(\widetilde{G}^4(\varepsilon)\mathbb{E}\bigg[\int_0^t (X_s^{\Theta_0})^2 \mathrm{d}s\bigg] + \mathbb{E}\bigg[\int_0^t (\bar{X}_s^\varepsilon)^2 \mathrm{d}s\bigg]\Bigg)
\end{aligned}$$

and via (vi) and (vii) in Assumptions 1 it follows that

$$\begin{aligned}
&\mathbb{E}\Bigg[\int_0^t \bigg\{\int_{\mathbb{R}_0} \left(Y_s^{\Theta_0}(z)\,(1 - \rho(z;\Theta_0)) - 1_{\{|z|\ge\varepsilon\}} Y_s^{\Theta_\varepsilon}(z)\,(1 - \rho(z;\Theta_\varepsilon))\right) \ell(\mathrm{d}z)\bigg\}^2 \mathrm{d}s\Bigg] \\
&\le \int_{\mathbb{R}_0} (\rho(z;\Theta_0) - \rho(z;\Theta_\varepsilon))^2\, \ell(\mathrm{d}z)\mathbb{E}\bigg[\int_0^t \int_{\mathbb{R}_0} (Y_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\bigg] \\
&\quad + \int_{\mathbb{R}_0} (1 - \rho(z;\Theta_\varepsilon))^2\, \ell(\mathrm{d}z)\mathbb{E}\bigg[\int_0^t \int_{\mathbb{R}_0} (\bar{Y}_s^\varepsilon(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\bigg] \\
&\le C\Bigg(\widetilde{G}^4(\varepsilon)\mathbb{E}\bigg[\int_0^t \int_{\mathbb{R}_0} (Y_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\bigg] + \mathbb{E}\bigg[\int_0^t \int_{\mathbb{R}_0} (\bar{Y}_s^\varepsilon(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\bigg]\Bigg).
\end{aligned}$$

Thus we obtain that

$$
\begin{aligned}
&\mathbb{E}[(\mathscr{L}_t^{\Theta_0} - \mathscr{L}_t^{\Theta_\varepsilon})^2] \\
&\leq C\left(\widetilde{G}^4(\varepsilon)\left\{\mathbb{E}\left[\int_0^t (X_s^{\Theta_0})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (Y_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right\}\right. \\
&\qquad \left. + \mathbb{E}\left[\int_0^t (\bar{X}_s^{\varepsilon})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (\bar{Y}_s^{\varepsilon}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] + \mathbb{E}\left[\int_0^t (Z_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right]\right).
\end{aligned}
\tag{42}
$$

Let us consider the terms appearing in the latter expression separately.

- Definition (12) implies that

$$
\mathbb{E}\left[\int_0^t (X_s^{\Theta_0})^2 \mathrm{d}s\right] \leq 2\left(\mathbb{E}\left[\int_0^t (A_s^{\Theta_0})^2 \mathrm{d}s\right] + b^2 \mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0})^2 \mathrm{d}s\right]\right)
$$

  and

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (Y_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] \\
&\leq 2\left(\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (B_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] + \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)^2 \ell(\mathrm{d}z) \mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0})^2 \mathrm{d}s\right]\right).
\end{aligned}
$$

- Combining (12) and (26) in

$$
\bar{X}_t^{\varepsilon} = X_t^{\Theta_0} - X_t^{\Theta_\varepsilon} = \bar{A}_t^{\varepsilon} - (\hat{\pi}_t^{\Theta_0} - \hat{\pi}_t^{\Theta_\varepsilon})b,
$$

  it easily follows that

$$
\mathbb{E}\left[\int_0^t (\bar{X}_s^{\varepsilon})^2 \mathrm{d}s\right] \leq C\left(\mathbb{E}\left[\int_0^t (\bar{A}_s^{\varepsilon})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right]\right).
$$

- Similarly, from (12) and (26) we find

$$
\bar{Y}_t^{\varepsilon}(z) = Y_t^{\Theta_0}(z) - Y_t^{\Theta_\varepsilon}(z) = \bar{B}_t^{\varepsilon}(z) - (\hat{\pi}_t^{\Theta_0} - \hat{\pi}_t^{\Theta_\varepsilon})(\mathrm{e}^z - 1).
$$

  Hence

$$
\begin{aligned}
&\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (\bar{Y}_s^{\varepsilon}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] \\
&\leq 2\left(\mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (\bar{B}_s^{\varepsilon}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right] + \int_{\mathbb{R}_0} (\mathrm{e}^z - 1)^2 \ell(\mathrm{d}z) \mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right]\right).
\end{aligned}
$$

- From (26), the estimate

$$(Z_t^{\Theta_\varepsilon}(z))^2 \le C\left((C_t^{\Theta_\varepsilon})^2 + (\hat{\pi}_t^{\Theta_0} - \hat{\pi}_t^{\Theta_\varepsilon})^2 G^2(\varepsilon) + (\hat{\pi}_t^{\Theta_0})^2 G^2(\varepsilon)\right)$$

  leads to

$$\mathbb{E}\left[\int_0^t (Z_s^{\Theta_\varepsilon}(z))^2 \mathrm{d}s\right] \le C\left(\mathbb{E}\left[\int_0^t (C_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] + G^2(\varepsilon)\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right]\right.$$
$$\left. + G^2(\varepsilon)\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0})^2 \mathrm{d}s\right]\right).$$

- Because of (39) and (vi) in Assumptions 1 we notice that

$$\mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0})^2 \mathrm{d}s\right] \le C\left(\mathbb{E}\left[\int_0^t (A_s^{\Theta_0})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (B_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right).$$

Using (41) and the combination of the above inequalities in (42) show that

$$\mathbb{E}[(\mathscr{L}_t^{\Theta_0} - \mathscr{L}_t^{\Theta_\varepsilon})^2] \le C\left(\widetilde{G}^2(\varepsilon)\left\{\mathbb{E}\left[\int_0^t (A_s^{\Theta_0})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (B_s^{\Theta_0}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]\right\}\right.$$
$$+ \mathbb{E}\left[\int_0^t (\bar{A}_s^{\varepsilon})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^t \int_{\mathbb{R}_0} (\bar{B}_s^{\varepsilon}(z))^2 \ell(\mathrm{d}z)\mathrm{d}s\right]$$
$$\left. + \mathbb{E}\left[\int_0^t (C_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right] + \mathbb{E}\left[\int_0^t (\hat{\pi}_s^{\Theta_0} - \hat{\pi}_s^{\Theta_\varepsilon})^2 \mathrm{d}s\right]\right).$$

Finally by Lemma 1 and Theorems 1 and 3 we conclude the proof. □

The cost processes of the quadratic hedging strategy for $\hat{H}_T$, $\hat{H}_T^{\varepsilon}$ are defined by $K^{\Theta_0} = \mathscr{L}^{\Theta_0} + \hat{V}_0^{\Theta_0}$ and $K^{\Theta_\varepsilon} = \mathscr{L}^{\Theta_\varepsilon} + \hat{V}_0^{\Theta_\varepsilon}$. The upcoming result concerns the robustness of the cost process and follows directly from the previous theorem.

Corollary 1 *Under Assumptions 1 and 2, there exists a positive constant C such that it holds for all $t \in [0, T]$ that*

$$\mathbb{E}[(K_t^{\Theta_0} - K_t^{\Theta_\varepsilon})^2] \le C\left(\mathbb{E}[(\hat{H}_T - \hat{H}_T^{\varepsilon})^2] + \widetilde{G}^2(\varepsilon)\mathbb{E}[\hat{H}_T^2]\right).$$

Proof Notice that

$$\mathbb{E}[(K_t^{\Theta_0} - K_t^{\Theta_\varepsilon})^2] \le 2\left(\mathbb{E}[(\mathscr{L}_t^{\Theta_0} - \mathscr{L}_t^{\Theta_\varepsilon})^2] + \mathbb{E}[(\hat{V}_0^{\Theta_0} - \hat{V}_0^{\Theta_\varepsilon})^2]\right),$$

wherein

$$\mathbb{E}[(\hat{V}_0^{\Theta_0} - \hat{V}_0^{\Theta_\varepsilon})^2] \leq \mathbb{E}\left[\sup_{0\leq t\leq T} (\hat{V}_t^{\Theta_0} - \hat{V}_t^{\Theta_\varepsilon})^2\right].$$

Theorems 2 and 5 complete the proof. □

3.4 Robustness Results for the Mean-Variance Hedging

Since the optimal numbers ξ^{Θ_0} and ξ^{Θ_ε} of risky assets are the same in the RM and the MVH strategy, the amounts of wealth $\hat{\pi}^{\Theta_0}$ and $\hat{\pi}^{\Theta_\varepsilon}$ and the gain processes $\hat{G}^{\Theta_0}$ and $\hat{G}^{\Theta_\varepsilon}$ also coincide for both strategies. Therefore we conclude that the robustness results of the amount of wealth and gain process also hold true for the MVH strategy, see Theorems 3 and 4.

The cost for a MVH strategy is not the same as for the RM strategy. However, under the assumption that a fixed starting amount $\widetilde{V}_0$ is available to set up a MVH strategy, we derive a robustness result for the loss at time of maturity. For the models (1) and (13), it holds that the losses at time of maturity T are given by

$$L^{\Theta_0} = \hat{H}_T - \widetilde{V}_0 - \int_0^T \xi_s^{\Theta_0} \mathrm{d}\hat{S}_s,$$
$$L^{\Theta_\varepsilon} = \hat{H}_T^\varepsilon - \widetilde{V}_0 - \int_0^T \xi_s^{\Theta_\varepsilon} \mathrm{d}\hat{S}_s^\varepsilon.$$

When Assumptions 1 and 2 are imposed, we derive via Theorem 4 that

$$\mathbb{E}[(L^{\Theta_0} - L^{\Theta_\varepsilon})^2] \leq C\left(\mathbb{E}[(\hat{H}_T - \hat{H}_T^\varepsilon)^2] + \widetilde{G}^2(\varepsilon)\mathbb{E}[\hat{H}_T^2]\right),$$

for a positive constant C.

Note that we cannot draw any conclusions from the results above about the robustness of the value of the discounted portfolio for the MVH strategy, since the portfolios are strictly different for both strategies.

4 Conclusion

Two different geometric Lévy stock price models were considered in this paper. We proved that the RM and the MVH strategies in a martingale setting are stable against the choice of the model. To this end the two models were considered under different risk-neutral measures that are dependent on the specific price models. The robustness results are derived through the use of BSDEJs and the obtained L^2-convergence rates are expressed in terms of estimates of the form $\mathbb{E}[(\hat{H}_T - \hat{H}_T^{\varepsilon})^2]$. The latter estimate is a well studied quantity, see [3, 16]. In the current paper, we considered two possible models for the price process. Starting from the initial model (1) other models could be constructed by truncating the small jumps and possibly rescaling the original Brownian motion (cfr. [8]). Similar robustness results hold for quadratic hedging strategies in a martingale setting in these other models.

In [8] a semimartingale setting was considered and conditions had to be imposed to guarantee the existence of the solutions to the BSDEJs. In this paper however, we considered a martingale setting and, since there is no driver in the BSDEJs, the existence of the solution to the BSDEJs was immediately guaranteed. On the other hand, since the two models were considered under two different martingale measures, we had to fall back on the common historical measure for the robustness study. Therefore, a robustness study of the martingale measures had to be performed and additional terms made some computations more involved compared to the semimartingale setting studied in [8].

In this approach based on BSDEJs we could not find explicit robustness results for the optimal number of risky assets. Therefore we refer to [6], where a robustness study is performed in a martingale and semimartingale setting based on Fourier transforms. Note that in [6] robustness was mainly studied in the L^1-sense and the authors noted that their results can be extended into L^2-convergence, whereas L^2-robustness results are explicitly derived in the current paper.

Acknowledgments The financial support from the Agency for Innovation by Science and Technology in Flanders (IWT) is gratefully acknowledged by Catherine Daveloose. Asma Khedher thanks the KPMG Center of Excellence in Risk Management for the financial support. The research by Asma Khedher and Michèle Vanmaele was carried out with support of CAS - Centre of Advanced Study, at the Norwegian Academy of Science and Letter, research program SEFE. Michèle Vanmaele acknowledges the Research Foundation Flanders (FWO) and the Special Research Fund (BOF) of the Ghent University for providing the possibility to go on sabbatical leave to CAS.

References

1. Arai, T.: Minimal martingale measures for jump diffusion processes. J. Appl. Probab. **41**, 263–270 (2004)
2. Asmussen, S., Rosinski, J.: Approximations of small jump Lévy processes with a view towards simulation. J. Appl. Probab. **38**, 482–493 (2001)
3. Benth, F.E., Di Nunno, G., Khedher, A.: Robustness of option prices and their deltas in markets modelled by jump-diffusions. Comm. Stoch. Anal. **5**(2), 285–307 (2011)
4. Benth, F.E., Di Nunno, G., Khedher, A.: A note on convergence of option prices and their Greeks for Lévy models. Stochastics **85**(6), 1015–1039 (2013)
5. Cont, R., Tankov, P.: Financial Modelling with Jump Processes. Chapman & Hall (2004)

6. Daveloose, C., Khedher, A., Vanmaele, M.: Robustness of quadratic hedging strategies in finance via Fourier transforms. J. Comput. Appl. Math. (2015). doi:10.1016/j.cam.2015.09.005
7. Delong, L.: Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications: BSDEs with Jumps. Springer, New York (2013)
8. Di Nunno, G., Khedher, A., Vanmaele, M.: Robustness of quadratic hedging strategies in finance via backward stochastic differential equations. Appl. Math. Optim. (2015). doi:10.1007/s00245-014-9283-z
9. El Karoui, N., Peng, S., Quenez, M.C.: Backward stochastic differential equations in finance. Math. Finance **7**(1), 1–71 (1997)
10. Föllmer, H., Schweizer, M.: Hedging of contingent claims under incomplete information. In: Davis, M.H.A., Elliot, R.J. (eds.) Applied Stochastic Analysis, Stochastic Monographs, vol. 5, pp. 389–414. Gordon and Breach, London/New York (1991)
11. Föllmer, H., Sondermann, D.: Hedging of non-redundant contingent claims. In: Hildenbrand, W., Mas-Colell, A. (eds.) Contributions to Mathematical Economics, pp. 205–223. North-Holland, Elsevier (1986)
12. Fujiwara, T., Miyahara, Y.: The mimimal entropy martingale measures for geometric Lévy processes. Fin. Stoch. **7**, 509–531 (2003)
13. Gerber, H.U., Shiu, E.S.W.: Actuarial bridges to dynamic hedging and option pricing. Insur. Math. Econom. **18**, 183–218 (1996)
14. Jacod, J., Shiryaev, A.: Limit Theorems for Stochastic Processes. Springer, Heidelberg (2002)
15. Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Graduate Texts in Mathematics, 2nd edn. Springer (1991)
16. Kohatsu-Higa, A., Tankov, P.: Jump-adapted discretization schemes for Lévy-driven SDEs. Stochastic Process. Appl. **120**(11), 2258–2285 (2010)
17. Øksendal, B., Sulem, A.: Applied Stochastic Control of Jump Diffusions. Springer, Heidelberg (2009)
18. Protter, P.: Stochastic integration and differential equations, 2nd edn. Version 2.1. Stochastic Modelling and Applied Probability, vol. 21. Springer, Berlin (2005)
19. Schweizer, M.: A guided tour through quadratic hedging approaches. In: Jouini, E., Cvitanić, J., Musiela, M. (eds.) Option Pricing, Interest Rates and Risk Management, pp. 538–574. Cambridge University Press (2001)
20. Tang, S., Li, X.: Necessary conditions for optimal control of stochastic systems with random jumps. SIAM J. Control Optim. **32**(5), 1447–1475 (1994)
21. Tankov, P.: Pricing and hedging in exponential Lévy models: review of recent results. Paris-Princeton Lecture Notes in Mathematical Finance (2010)

Risk-Sensitive Mean-Field Type Control Under Partial Observation

Boualem Djehiche and Hamidou Tembine

Abstract We establish a stochastic maximum principle (SMP) for control problems of partially observed diffusions of mean-field type with risk-sensitive performance functionals.

Keywords Time inconsistent control · Maximum principle · Mean-field SDE · Risk-sensitive control · Partial observation

AMS subject classification: 93E20 · 60H30 · 60H10 · 91B28.

1 Introduction

In optimal control problems for diffusions of mean-field type the performance functional, drift and diffusion coefficient depend not only on the state and the control but also on the probability distribution of the state-control pair. The mean-field coupling makes the control problem time-inconsistent in the sense that the Bellman Principle is no longer valid, which motivates the use of the stochastic maximum (SMP) approach to solve this type of optimal control problems instead of trying extensions of the dynamic programming principle (DPP). This class of control problems has been studied by many authors including [1, 2, 5, 7, 15, 20]. The performance functionals

B. Djehiche (✉)
Department of Mathematics, KTH Royal Institute of Technology,
100 44 Stockholm, Sweden
e-mail: boualem@math.kth.se

H. Tembine
New York University, Abu Dhabi, United Arab Emirates
e-mail: tembine@nyu.edu

H. Tembine
Learning and Game Theory Lab, Division of Engineering, NYU Abu Dhabi,
Saadiyat Island, Abu Dhabi, United Arab Emirates

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_9

considered in these papers have been of risk-neutral type i.e. the running cost/profit terms are expected values of stage-additive payoff functions. Not all behavior, however, can be captured by risk-neutral performance. One way of capturing risk-averse and risk-seeking behaviors is by exponentiating the performance functional before expectation (see [17]).

The first paper that we are aware of and which deals with risk-sensitive optimal control in a mean field context is [24]. Using a matching argument, the authors derive a verification theorem for a risk-sensitive mean-field game whose underlying dynamics is a Markov diffusion. This matching argument freezes the mean-field coupling in the dynamics, which yields a standard risk-sensitive HJB equation for the value-function. The mean-field coupling is then retrieved through the Fokker-Planck equation satisfied by the marginal law of the optimal state.

In a recent paper [11], the authors have established a risk-sensitive SMP for mean-field type control. The risk-sensitive control problem was first reformulated in terms of an augmented state process and terminal payoff problem. An intermediate stochastic maximum principle was then obtained by applying the SMP of ([5], Theorem 2.1.) for loss functionals without running cost but with augmented state in higher dimension and complete observation of the state. Then, the intermediate first- and second-order adjoint processes are transformed into a simpler form using a logarithmic transformation derived in [12].

Optimal control of partially observed diffusions (without mean-field coupling) has been studied by many authors including the non-exhaustive references [3, 4, 8–10, 13, 14, 16, 19, 21, 23, 26, 27], using both the DPP and SMP approaches. Reference [23] derives an SMP for the most general model of optimal control of partially observed diffusions under risk-neutral performance functionals. Recently, Wang et al. [25], extended the SMP for partially observable optimal control of diffusions for risk-neutral performance functionals of mean-field type.

The purpose of this paper is to establish a stochastic maximum principle for a class of risk-sensitive mean-field type control problems under partial observation. Following the above mentioned papers of optimal control under partial observation, in particular [23], our strategy is to transform the partially observable control problem into a completely observable one and then apply the approach suggested in [11] to derive a suitable risk-sensitive SMP. To the best to our knowledge, the risk-sensitive maximum principle under partial observation without passing through the DPP, and in particular, for mean-field type controls was not established in earlier works.

The paper is organized as follows. In Sect. 2, we present the model and state the partially observable risk-sensitive SMP which constitutes the main result, whose proof is given in Sect. 3. Finally, in Sect. 4, we apply the risk-sensitive SMP to the linear-exponential-quadratic setup under partial observation. To streamline the presentation, we only consider the one-dimensional case. The extension to the multidimensional case is by now straightforward. Furthermore, we consider diffusion models where the control enters only the drift coefficient, which leads to an SMP with only one pair of adjoint processes. The general Peng-type SMP can be derived following e.g. [11, 23].

2 Statement of the Problem

Let $T > 0$ be a fixed time horizon and $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ be a given filtered probability space on which there are defined two independent standard one-dimensional Brownian motions $W = \{W_s\}_{s\geq 0}$ and $Y = \{Y_s\}_{s\geq 0}$. Let $\mathscr{F}_t^W$ and $\mathscr{F}_t^Y$ be the $\mathbb{P}$-completed natural filtrations generated by W and Y, respectively. Set $\mathbb{F}^Y := \{\mathscr{F}_t^Y,\ 0 \leq s \leq T\}$ and $\mathbb{F} := \{\mathscr{F}_s,\ 0 \leq s \leq T\}$, where, $\mathscr{F}_t = \mathscr{F}_t^W \vee \mathscr{F}_t^Y$.

We consider a mean-field type version the stochastic controlled system with partial observation considered in [23] which is an extension of the model considered by [4, 14] to which we refer for further details.

The model is defined as follows.

(i) An admissible control u is an $\mathbb{F}^Y$-adapted process with values in a non-empty subset (not necessarily convex) U of $\mathbb{R}$ and satisfies $E[\int_0^T |u(t)|^2 dt] < \infty$. We denote the set of all admissible controls by $\mathscr{U}$. The control u is called partially observable.

(ii) Given a control process $u \in \mathscr{U}$, we consider the signal-observation pair (x^u, Y) which satisfies the following SDE of mean-field type

$$\begin{cases} dx^u(t) = b(t, x^u(t), E[x^u(t)], u(t))dt + \sigma(t, x^u(t), E[x^u(t)])dW_t \\ \qquad\qquad + \alpha(t, x^u(t), E[x^u(t)])d\widetilde{W}_t^u, \quad x^u(0) = x_0, \\ dY_t \qquad = \beta(t, x^u(t))dt + d\widetilde{W}_t^u, \quad Y_0 = 0, \end{cases} \tag{1}$$

where,

$$b(t, x, m, u) : \ [0, T] \times \mathbb{R} \times \mathbb{R} \times U \longrightarrow \mathbb{R},$$

$$\alpha(t, x, m),\ \sigma(t, x, m) : \ [0, T] \times \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$$

and $\beta(t, x) : [0, T] \times \mathbb{R} \longrightarrow \mathbb{R}$ are Borel measurable function.

In this model, the observation process Y, which carries out the controls u, is assumed to be a given Brownian motion independent of W and is supposed to admit a decomposition as a trend $\int_0^{\cdot} \beta(t, x^u(t))dt$ (a functional of the state process x^u) corrupted by a process $\widetilde{W}^u$ which are a priori not observable. The case $\alpha = 0$ corresponds to the model considered in [4, 14]. A more general model of the function β would be to let it depend on the control u and be of mean-field type. To keep the presentation simpler, we skip this case in this paper. But, the main results do extend to this case.

Before we formulate the control problem, we show that the system (1) has a weak solution. Introduce the density process defined on $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ by

$$\rho^u(t) := \exp\left\{\int_0^t \beta(s, x^u(s))dY_s - \frac{1}{2}\int_0^t |\beta(s, x^u(s))|^2 ds\right\},$$

which solves the linear SDE

$$d\rho^u(t) = \rho^u(t)\beta(t, x^u(t))dY_t, \quad \rho^u(0) = 1.$$

Assuming the function β bounded (see Assumption 1, below), ρ is a uniformly integrable martingale such that, for every $p \geq 2$,

$$E[\sup_{0\leq t\leq T} \rho_t^p] \leq C, \tag{2}$$

where, C is a constant which depends only on the bound of β, p and T. Define $d\mathbb{P}^u = \rho^u(T)d\mathbb{P}$. By Girsanov's Theorem, $\mathbb{P}^u$ is a probability measure. Moreover, $\widetilde{W}^u$ is a $\mathbb{P}^u$-standard Brownian motion independent of W. This in turn entails that $(\mathbb{P}^u, x^u, Y, W, \widetilde{W}^u)$ is a weak solution of (1).

The objective is to characterize admissible controls which minimize the risk-sensitive cost functional given by

$$J^\theta(u(\cdot)) = E^u\left[\exp\left(\theta\left[\int_0^T f(t, x^u(t), E^u[x^u(t)], u(t))\,dt + h(x^u(T), E^u[x^u(T)])\right]\right)\right],$$

where, θ is the risk-sensitivity index,

$$\begin{aligned} &f(t, x, m, u): \ [0, T] \times \mathbb{R} \times \mathbb{R} \times U \longrightarrow \mathbb{R}, \\ &h(x, m): \ \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}. \end{aligned}$$

Any $\bar{u}(\cdot) \in \mathscr{U}$ which satisfies

$$J^\theta(\bar{u}(\cdot)) = \inf_{u(\cdot)\in\mathscr{U}} J^\theta(u(\cdot)) \tag{3}$$

is called a risk-sensitive optimal control under partial observation.

Let $\Psi_T = \int_0^T f(t, x(t), E^u[x(t)], u(t))dt + h(x(T), E^u[x(T)])$ and consider the payoff functional given by

$$\widetilde{\Psi}_\theta := \frac{1}{\theta}\log E^u e^{\theta\Psi_T}.$$

When the risk-sensitive index θ is small, the loss functional $\widetilde{\Psi}_\theta$ can be expanded as

$$E^u[\Psi_T] + \frac{\theta}{2}\text{var}_u(\Psi_T) + O(\theta^2),$$

where, $\mathrm{var}_u(\Psi_T)$ denotes the variance of Ψ_T w.r.t. $\mathbb{P}^u$. If $\theta < 0$, the variance of Ψ_T, as a measure of risk, improves the performance $\widetilde{\Psi}_\theta$, in which case the optimizer is called *risk seeker*. But, when $\theta > 0$, the variance of Ψ_T worsens the performance $\widetilde{\Psi}_\theta$, in which case the optimizer is called *risk averse*. The risk-neutral loss functional $E^u[\Psi_T]$ can be seen as a limit of risk-sensitive functional $\widetilde{\Psi}_\theta$ when $\theta \to 0$.

Since $d\mathbb{P}^u = \rho^u(T)d\mathbb{P}$, the associated risk-sensitive cost functional becomes

$$J^\theta(u(\cdot)) = E\left[\rho^u(T)e^{\theta\left[\int_0^T f(t,x^u(t),E[\rho^u(t)x^u(t)],u(t))\,dt+h(x^u(T),E[\rho^u(T)x^u(T)])\right]}\right], \tag{4}$$

where, on $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$, the process (ρ^u, x^u) satisfies the following dynamics:

$$\begin{cases} d\rho^u(t) = \rho^u(t)\beta(s, x^u(s))dY_t, \\ dx^u(t) = \{b(t, x^u(t), E[x^u(t)], u(t)) - \alpha(t, x^u(t), E[x^u(t)])\beta(t, x^u(t))\}\,dt \\ \qquad\qquad + \sigma(t, x^u(t), E[x^u(t)])dW_t + \alpha(t, x^u(t), E[x^u(t)])dY_t, \\ \rho^u(0) \ = 1,\ \ x^u(0) = x_0. \end{cases} \tag{5}$$

We have recast the partially observable control problem into the following completely observable control problem: Minimize $J^\theta(u(\cdot))$ defined by (4) subject to (5).

The main result of this paper is a stochastic maximum principle (SMP) in terms of necessary optimality conditions for the problem (3) subject to (5).

We will only consider the case where the risk-sensitive parameter is positive, $\theta > 0$. The case $\theta < 0$ can be treated in a similar fashion by considering $\theta = -\bar{\theta}, \bar{\theta} > 0$, and $\bar{f} = -f, \bar{h} = -h$ in the performance functional (4).

We will make the following assumption.

Assumption 1 The functions $b, \sigma, \alpha, \beta, f, h$ are twice continuously differentiable with respect to (x, m). Moreover, these functions and their first derivatives with respect to (x, m) are bounded and continuous in (x, m, u).

To keep the presentation less technical, we impose these assumptions although they are restrictive and can be made weaker.
Under these assumptions, in view of Girsanov's theorem and [18], Proposition 1.2., for each $u \in \mathscr{U}$, the SDE (5) admits a unique weak solution (ρ^u, x^u).
We now state an SMP to characterize optimal controls $\bar{u}(\cdot) \in \mathscr{U}$ which minimize (4), subject to (5). Let $(\bar{\rho}, \bar{x}) := (\rho^{\bar{u}}, x^{\bar{u}})$ denote the corresponding state process, defined as the solution of (5).

We introduce the following notation.

$$
\begin{array}{l}
X := \begin{pmatrix} \rho \\ x \end{pmatrix}, \quad \bar{X} := \begin{pmatrix} \bar{\rho} \\ \bar{x} \end{pmatrix}, \quad X_0 = \bar{X}_0 := \begin{pmatrix} 1 \\ x_0 \end{pmatrix}, \quad B_t := \begin{pmatrix} Y_t \\ W_t \end{pmatrix}, \\
F(t, X, m, u) := \begin{pmatrix} 0 \\ c(t, x, m, u) \end{pmatrix}, \quad G(t, X, m) := \begin{pmatrix} \rho\beta(t, x) & 0 \\ \alpha(t, x, m) & \sigma(t, x, m) \end{pmatrix}, \\
c(t, x, m, u) := b(t, x, m, u) - \alpha(t, x, m)\beta(t, x), \ \phi(X) := x, \ \tilde{\phi}(X) := \rho x, \\
\phi(\bar{X}) := \bar{x}, \ \tilde{\phi}(\bar{X}) := \bar{\rho}\bar{x}.
\end{array}
\tag{6}
$$

We define the risk-neutral Hamiltonian as follows. For $(p, q) \in \mathbb{R}^2 \times \mathbb{R}^{2\times 2}$,

$$
H(t, X, m, p, q, u) := \langle F(t, X, m, u), p\rangle + \mathrm{tr}(G^*(t, X, m)q) - f(t, x, m, u), \tag{7}
$$

where, $'*'$ denotes the transposition operation of a matrix or a vector.

We also introduce the risk-sensitive Hamiltonian: $(p, q, \ell) \in \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^2$,

$$
\begin{aligned}
H^\theta(t, X, m, u, p, q, \ell) := \langle F(t, X, m, u), p\rangle &- f(t, x, m, u) \\
&+ \mathrm{tr}(G^*(t, X, m)(q + \theta\ell p^*)).
\end{aligned}
\tag{8}
$$

We have $H = H^0$.

Setting

$$
\ell := \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix}, \quad p := \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \quad q := \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix},
$$

the explicit form of the Hamiltonian (8) reads

$$
\begin{aligned}
H^\theta(t, X, m, u, p, q, \ell) := \ & c(t, x, m, u)p_2 - f(t, x, m, u) + \rho\beta(t, x)(q_{11} + \theta\ell_1 p_1) \\
& + \ \alpha(t, x, m)(q_{21} + \theta\ell_2 p_1) + \sigma(t, x, m)(q_{22} + \theta\ell_2 p_2).
\end{aligned}
$$

Setting $\theta = 0$, we obtain the explicit form of the Hamiltonian (7):

$$
\begin{aligned}
H(t, X, m, u, p, q) := \ & c(t, x, m, u)p_2 - f(t, x, m, u) + \rho\beta(t, x)q_{11} \\
& + \ \alpha(t, x, m)q_{21} + \sigma(t, x, m)q_{22}.
\end{aligned}
$$

With the obvious notation for the derivatives of the functions $b, \alpha, \beta, \sigma, f, h$, w.r.t. the arguments x and m, we further set

$$
\begin{cases}
H_x^\theta(t, X, m, u, p, q) := c_x(t, x, m, u)p_2 - f_x(t, x, m, u) + \rho\beta_x(t, x)(q_{11} + \theta\ell_1 p_1) \\
\qquad\qquad\qquad\qquad + \ \alpha_x(t, x, m)(q_{21} + \theta\ell_2 p_1) + \sigma_x(t, x, m)(q_{22} + \theta\ell_2 p_2), \\
\check{H}_m^\theta(t, X, m, u, p, q) := c_m(t, x, m, u)p_2 + \alpha_m(t, x, m)(q_{21} + \theta\ell_2 p_1) \\
\qquad\qquad\qquad\qquad + \ \sigma_m(t, x, m)(q_{22} + \theta\ell_2 p_2), \\
H_\rho^\theta(t, X, m, u, p, q) := \beta(t, x)(q_{11} + \theta\ell_1 p_1).
\end{cases}
$$

With this notation, the system (5) can be rewritten in the following compact form

$$\begin{cases} dX(t) = F(t, X(t), E[\phi(X(t))], u(t))dt + G(t, X(t), E[\phi(X(t))])dB_t, \\ X(0) \;\; = X_0. \end{cases} \tag{9}$$

We define the risk-neutral Hamiltonian associated with random variables X such that $\phi(X)$ and $\tilde{\phi}(X)$ are $L^1(\Omega, \mathscr{F}, \mathbb{P})$ as follows (with the obvious abuse of notation): For $(p, q) \in \mathbb{R}^2 \times \mathbb{R}^{2\times 2}$,

$$\begin{aligned} H(t, X, p, q, u) := \langle F(t, X, E[\phi(X)], u), p\rangle - f(t, x, E[\tilde{\phi}(X)], u) \\ + \mathrm{tr}(G^*(t, X, E[\phi(X)])q). \end{aligned} \tag{10}$$

We also introduce the risk-sensitive Hamiltonian: $(p, q, \ell) \in \mathbb{R}^2 \times \mathbb{R}^{2\times 2} \times \mathbb{R}^2$,

$$\begin{aligned} H^\theta(t, X, u, p, q, \ell) := \langle F(t, X, E[\phi(X)], u), p\rangle - f(t, x, E[\tilde{\phi}(X)], u) \\ + \mathrm{tr}(G^*(t, X, E[\phi(X)])(q + \theta \ell p^*)). \end{aligned} \tag{11}$$

For $g \in \{b, c, \sigma, \alpha, \beta\}$ and $u \in U$, we set

$$g_x(t) := g_x(t, \bar{x}(t), E[\bar{x}(t)], \bar{u}(t)), \quad g_m(t) := g_m(t, \bar{x}(t), E[\bar{x}(t)], \bar{u}(t)) \tag{12}$$

and

$$\begin{cases} f_x(t) := f_x(t, \bar{x}(t), E[\bar{\rho}(t)\bar{x}(t)], \bar{u}(t)), \quad f_m(t) := f_m(t, \bar{x}(t), E[\bar{\rho}(t)\bar{x}(t)], \bar{u}(t)), \\ h_x(t) := h_x(\bar{x}(t), E[\bar{\rho}(t)\bar{x}(t)]), \quad h_m(t) := h_m(\bar{x}(t), E[\bar{\rho}(t)\bar{x}(t)]). \end{cases} \tag{13}$$

Let

$$\psi_T^\theta := \bar{\rho}(T) \exp \theta \left[\int_0^T f(t, \bar{x}(t), E[\bar{\rho}(t)\bar{x}(t)], \bar{u}(t))dt + h(\bar{x}(T), E[\bar{\rho}(T)\bar{x}(T)]) \right].$$

We introduce the adjoint equations involved in the risk-sensitive SMP for our control problem.

$$\begin{cases} d\hat{p}(t) \;\; = -\begin{pmatrix} H_\rho^\theta(t) + \frac{1}{v^\theta(t)} E[v^\theta(t) \check{H}_m^\theta(t)] - \frac{\bar{x}(t)}{v^\theta(t)} E[v^\theta(t) f_m(t)] \\ H_x^\theta(t) + \frac{1}{v^\theta(t)} E[v^\theta(t) \check{H}_m^\theta(t)] - \frac{\bar{\rho}(t)}{v^\theta(t)} E[v^\theta(t) f_m(t)]] \end{pmatrix} dt \\ \qquad\qquad + \hat{q}(t)(-\theta \ell(t)dt + dB_t), \\ dv^\theta(t) = \theta v^\theta(t) \langle \ell(t), dB_t \rangle, \\ \hat{p}(T) \;\; = -\begin{pmatrix} (\theta \bar{\rho}(T))^{-1} \\ h_x(T) \end{pmatrix} - \begin{pmatrix} \bar{x}(T) \\ \bar{\rho}(T) \end{pmatrix} \frac{1}{\psi_T^\theta} E[\psi_T^\theta h_m(T)], \\ v^\theta(T) \;\; = \psi_T^\theta, \end{cases} \tag{14}$$

where, in view of (2) and (13), for $k \in \{\rho, x\}$,

$$
\begin{aligned}
H_k^{\theta}(t) := \langle F_k(t, \bar{X}(t), E[\phi(\bar{X}(t))], \bar{u}(t)), \hat{p}(t)\rangle - f_k(t, \bar{x}(t), E[\tilde{\phi}(\bar{X}(t))], \bar{u}(t)) \\
+ \operatorname{tr}(G_k^*(t, \bar{X}(t), E[\phi(\bar{X}(t))])(\hat{q}(t) + \theta\ell\hat{p}^*(t))
\end{aligned}
$$

and

$$
\begin{aligned}
\check{H}_m^{\theta}(t) := \langle F_m(t, \bar{X}(t), E[\phi(\bar{X}(t))], \bar{u}(t)), \hat{p}(t)\rangle \\
+ \operatorname{tr}(G_m^*(t, \bar{X}(t), E[\phi(\bar{X}(t))])(\hat{q}(t) + \theta\ell\hat{p}^*(t)).
\end{aligned}
$$

We note that the processes $(\hat{p}, \hat{q}, \ell)$ may depend on the sensitivity index θ. To ease notation, we omit to make this dependence explicit.

Below, we will show that, under Assumption 1, (14) admits a unique $\mathbb{F}$-adapted solution $(\hat{p}, \hat{q}, \hat{v}^{\theta}, \ell)$ such that

$$
E\left[\sup_{t\in[0,T]} |\hat{p}(t)|^2 + \sup_{t\in[0,T]} |v^{\theta}(t)|^2 + \int_0^T \left(|\hat{q}(t)|^2 + |\ell(t)|^2\right) dt\right] < \infty. \tag{15}
$$

Moreover,

Lemma 1 *The process defined on* $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P})$ *by*

$$
L_t^{\theta} := \frac{v^{\theta}(t)}{v^{\theta}(0)} = \exp\left(\int_0^t \theta\langle\ell(s), dB_s\rangle - \frac{\theta^2}{2}\int_0^t |\ell(s)|^2 ds\right), \quad 0 \le t \le T, \tag{16}
$$

is a uniformly integrable $\mathbb{F}$*-martingale.*

The process L^{θ} defines a new probability measure $\mathbb{P}^{\theta}$ equivalent to $\mathbb{P}$ by setting $L_t^{\theta} := \frac{d\mathbb{P}^{\theta}}{d\mathbb{P}}|_{\mathscr{F}_t}$. By Girsanov's theorem, the process $B_t^{\theta} := B_t - \theta\int_0^t \ell(s)ds, \ 0 \le t \le T$ is a $\mathbb{P}^{\theta}$-Brownian motion.

The following theorem is the main result of the paper. Let $E^{\theta}[\,\cdot\,]$ denote the expectation w.r.t. $\mathbb{P}^{\theta}$.

Theorem 1 (**Risk-sensitive maximum principle**) *Let Assumption 1 hold. If the process* $(\bar{\rho}(\cdot), \bar{x}(\cdot), \bar{u}(\cdot))$ *is an optimal solution of the risk-sensitive control problem (3)–(5), then there are two pairs of* $\mathbb{F}$*-adapted processes* (v^{θ}, ℓ) *and* $(\hat{p}, \hat{q})$ *which satisfy (14)–(15), such that*

$$
E^{\theta}[H^{\theta}(t, \bar{\rho}(t), \bar{x}(t), \hat{p}(t), \hat{q}(t), \ell(t), u) - H^{\theta}(t, \bar{\rho}(t), \bar{x}(t), \hat{p}(t), \hat{q}(t), \ell(t), \bar{u}(t))|\mathscr{F}_t^Y] \le 0,
$$

for all $u \in U$, *almost every* t *and* $\mathbb{P}^{\theta}$*−almost surely.*

Remark 1 The boundedness assumption of the involved coefficients and their derivatives imposed in Assumption 1, in Theorem 1, guarantees the solvability of the system of forward-backward SDEs (5) and (14). In fact Theorem 1 applies provided we can solve system of forward-backward SDEs (5) and (14). A typical example of such a situation is the classical Linear-Quadratic (LQ) control problem (see Sect. 4 below), in which the involved coefficients are at most quadratic, but not necessarily bounded.

3 Proof of the Main Result

In this section we give a proof of Theorem 1 here presented in several steps.

3.1 An Intermediate SMP for Mean-Field Type Control

In this subsection we first reformulate the risk-sensitive control problem associated with (4)–(5) in terms of an augmented state process and terminal payoff problem. An intermediate stochastic maximum principle is then obtained by applying the SMP obtained in ([1], Theorem 3.1 or [5], Theorem 2.1) for loss functionals without running cost. Then, we transform the intermediate first-order adjoint processes to a simpler form. The mean-field type control problem for the cost functional (4) under the dynamics (5) is equivalent to

$$\inf_{u(\cdot)\in\mathscr{U}} E\left[\rho(T)e^{\theta[h(x(T),E[\rho(T)x(T)])+\xi(T)]}\right], \tag{17}$$

subject to

$$\begin{cases} d\rho(t) = \rho(t)\beta(t,x(t))dY_t, \\ dx(t) = \{b(t,x(t),E[x(t)],u(t)) - \alpha(t,x(t),E[x(t)])\beta(t,x(t))\}\,dt \\ \qquad\quad + \sigma(t,x(t),E[x(t)])dW_t + \alpha(t,x(t),E[x(t)])dY_t, \\ d\xi \quad\; = f(t,x(t),E[\rho(t)x(t)],u(t))dt, \\ \rho(0) \;\; = 1,\; x(0)=x_0,\; \xi(0)=0. \end{cases} \tag{18}$$

We introduce the following notation.

$$R := \begin{pmatrix}\rho\\ x\\ \xi\end{pmatrix} = \begin{pmatrix}X\\ \xi\end{pmatrix},\;\; \bar{R} := \begin{pmatrix}\bar{\rho}\\ \bar{x}\\ \bar{\xi}\end{pmatrix} = \begin{pmatrix}\bar{X}\\ \bar{\xi}\end{pmatrix},\;\; R_0 = \bar{R}_0 := \begin{pmatrix}X_0,\\ 0\end{pmatrix},$$

$$\Gamma(t,R,m) := \begin{pmatrix}G(t,X,m,u)\\ 0\end{pmatrix},$$

$$\phi(R) = \phi(X) = x,\;\; \tilde{\phi}(R) = \tilde{\phi}(X) = \rho x,\;\; \phi(\bar{R}) = \phi(\bar{X}) = \bar{x},\;\; \tilde{\phi}(\bar{R}) = \tilde{\phi}(\bar{X}) = \bar{\rho}\bar{x}.$$

With this notation, the system (18) can be rewritten in the following compact form

$$\begin{cases} dR(t) = \begin{pmatrix} F(t, R(t), E[\phi(R(t))], u(t)) \\ f(t, R(t), E[\tilde{\phi}(R(t))], u(t)) \end{pmatrix} dt + \Gamma(t, R(t), E[\phi(R(t))]) dB_t, \\ R(0) \ = R_0 \end{cases}$$

and the risk-sensitive cost functional (4) becomes

$$J^{\theta}(u(\cdot)) := E[\Phi\left(R(T), E[\tilde{\phi}(R(T))]\right)],$$

where,

$$\Phi\left(R(T), E[\phi(R(T))]\right) := \rho(T) \exp\left(\theta h(x(T), E[\rho(T)x(T)]) + \theta\xi(T)\right).$$

We define the Hamiltonian associated with random variables R such that $\phi(R) \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ as follows. For $(p, q) \in \mathbb{R}^3 \times \mathbb{R}^{3\times 3}$,

$$\begin{aligned} H^e(t, R, p, q, u) := \langle \begin{pmatrix} F(t, R(t), E[\phi(R(t))], u(t)) \\ f(t, R(t), E[\tilde{\phi}(R(t))], u(t)) \end{pmatrix}, p \rangle \\ + \mathrm{tr}(\Gamma^*(t, R, E[\phi(R)])q), \end{aligned} \tag{19}$$

where, Γ^* denotes the transpose of the matrix Γ.

Setting

$$p := \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}, \quad q := \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \\ q_{31} & q_{32} \end{pmatrix}, \tag{20}$$

the explicit form of the Hamiltonian (19) reads

$$\begin{aligned} H^e(t, \rho, x, \xi, p, q, u) := H^e(t, R, p, q, u) = c(t, x, E[x], u)p_2 + f(t, x, E[\rho x], u)p_3 \\ + \sigma(t, x, E[x])q_{22} + \rho\beta(t, x)q_{11} + \alpha(t, x, E[x])q_{21}. \end{aligned}$$

In view of (12), we set

$$\begin{cases} H_x^e(t) \ := c_x(t)p_2(t) + f_x(t)p_3(t) + \sigma_x(t)q_{22}(t) + \bar{\rho}(t)\beta_x(t)q_{11}(t) + \alpha_x(t)q_{21}(t), \\ \check{H}_m^e(t) := c_m(t)p_2(t) + f_m(t)p_3(t) + \sigma_m(t)q_{22}(t) + \alpha_m(t)q_{21}(t), \\ H_\rho^e(t) \ = \ \beta(t, \bar{x}(t))q_{11}(t). \end{cases}$$

We may apply the SMP for risk-neutral mean-field type control (cf. [1], Theorem 3.1 or [5], Theorem 2.1) to the augmented state dynamics (ρ, x, ξ) to derive the first order adjoint equation:

$$\begin{cases} dp(t) = -\begin{pmatrix} H^e_\rho(t) + E[\check{H}^e_m(t)] + \bar{x}(t)E[f_m(t)p_3(t)] \\ H^e_x(t) + E[\check{H}^e_m(t)] + \bar{\rho}(t)E[f_m(t)p_3(t)] \\ 0 \end{pmatrix} dt + q(t)dB_t, \\ p(T) = -\theta\psi^\theta_T \begin{pmatrix} (\theta\bar{\rho}(T))^{-1} \\ h_x(T) \\ 1 \end{pmatrix} - \theta \begin{pmatrix} \bar{x}(T) \\ \bar{\rho}(T) \\ 0 \end{pmatrix} E[\psi^\theta_T h_m(T)]. \end{cases} \tag{21}$$

This is a system of linear backward SDEs of mean-field type which, in view of ([6], Theorem 3.1), under Assumption 1, admits a unique $\mathbb{F}$-adapted solution (p, q) satisfying

$$E\left[\sup_{t\in[0,T]} |p(t)|^2 + \int_0^T |q(t)|^2 dt\right] < \infty, \tag{22}$$

where, $|\cdot|$ denotes the usual Euclidean norm with appropriate dimension.

We may apply the SMP for SDEs of mean-field type control from ([1], Theorem 3.1 or [5], Theorem 2.1) together with the SMP for risk-neutral partially observable SDEs derived in ([23], Theorem 2.1) to obtain the following SMP.

Proposition 1 *Let Assumption 1 hold. If $(\bar{R}(\cdot), \bar{u}(\cdot))$ is an optimal solution of the risk-neutral control problem (17) subject to the dynamics (18), then there is a unique pair of $\mathbb{F}$-adapted processes (p, q) which satisfies (21)–(22) such that*

$$E[H^e(t, \bar{R}(t), p(t), q(t), u) - H^e(t, \bar{R}(t), p(t), q(t), \bar{u}(t))|\mathscr{F}^Y_t] \le 0,$$

for all $u \in U$, almost every t and $\mathbb{P}-$almost surely.

3.2 Transformation of the First Order Adjoint Process

Although the result of Proposition 1 is a good SMP for the risk-sensitive mean-field type control with partial observations, augmenting the state process with the third component ξ yields a system of three adjoint equations that appears complicated to solve in concrete situations. In this section we apply the transformation of the adjoint processes (p, q) introduced in [11] in such a way to get rid of the third component (p_3, q_{31}, q_{32}) in (21) and express the SMP in terms of only two adjoint process that we denote $(\hat{p}, \hat{q})$, where

$$\hat{p} := \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix}, \quad \hat{q} := \begin{pmatrix} \hat{q}_1 \\ \hat{q}_2 \end{pmatrix}, \quad \hat{q}_i := (\hat{q}_{i1}, \hat{q}_{i2}), \quad i = 1, 2. \tag{23}$$

Indeed, noting that from (21), we have $dp_3(t) = \langle q_3(t), dB_t \rangle$ and $p_3(T) = -\theta \psi_T^\theta$, the explicit solution of this backward SDE is

$$p_3(t) = -\theta E[\psi_T^\theta \mid \mathscr{F}_t] = -\theta v^\theta(t), \tag{24}$$

where,

$$v^\theta(t) := E[\psi_T^\theta \mid \mathscr{F}_t], \qquad 0 \le t \le T.$$

In particular, we have $v^\theta(0) = E[\psi_T^\theta]$. Therefore, in view of (24), it would be natural to choose a transformation of (p, q) into an adjoint process $(\hat{p}, \hat{q})$, where,

$$\hat{p} := \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \\ \hat{p}_3 \end{pmatrix}, \quad \hat{q} := \begin{pmatrix} \hat{q}_{11} & \hat{q}_{12} \\ \hat{q}_{21} & \hat{q}_{22} \\ \hat{q}_{31} & \hat{q}_{32} \end{pmatrix},$$

such that

$$\hat{p}_3(t) = \frac{p_3(t)}{\theta v^\theta(t)} = -1, \qquad 0 \le t \le T. \tag{25}$$

This would imply that, for almost every $0 \le t \le T$,

$$\hat{q}_3(t) = (\hat{q}_{31}(t), \hat{q}_{32}(t)) = 0, \quad \mathbb{P} - \text{a.s.}, \tag{26}$$

which in turn would reduce the number of adjoint processes to those of the form given by (23).

We consider the following transform:

$$\hat{p}(t) := \frac{1}{\theta v^\theta(t)} p(t), \qquad 0 \le t \le T. \tag{27}$$

In view of (21), we have

$$\hat{p}(T) = -\begin{pmatrix} (\theta \bar{\rho}(T))^{-1} \\ h_x(T) \\ 1 \end{pmatrix} - \begin{pmatrix} \bar{x}(T) \\ \bar{\rho}(T) \\ 0 \end{pmatrix} \frac{1}{\psi_T^\theta} E[\psi_T^\theta h_m(T)]. \tag{28}$$

We should identify the processes $\hat{\alpha}$ and $\hat{q}$ such that

$$d\hat{p}(t) = -\hat{\alpha}(t)dt + \hat{q}(t)dB_t, \tag{29}$$

for which (25) and (26) are satisfied.

In order to investigate the properties of these new processes $(\hat{p}, \hat{q})$, the following properties of the generic martingale v^θ, used in [11], are essential. We reproduce them here for the sake of completeness. Since, by Assumption 1, f and h are bounded by some constant $C > 0$, we have

$$0 < e^{-(1+T)C\theta}\rho(T) \le \psi_T^\theta \le e^{(1+T)C\theta}\rho(T).$$

Therefore, v^θ is a uniformly integrable $\mathbb{F}$-martingale satisfying

$$0 < e^{-(1+T)C\theta}\rho(t) \le v^\theta(t) \le e^{(1+T)C\theta}\rho(t), \qquad 0 \le t \le T.$$

Hence, in view of (2), we have

$$E[\sup_{0\le t\le T} |v^\theta(t)|^2] \le C. \tag{30}$$

Furthermore, the martingale v^θ enjoys the following useful *logarithmic transform* established in ([12], Proposition 3.1)

$$v^\theta(t) = \exp\left(\theta Z_t + \theta \int_0^t f(s, \bar{x}(s), E[\bar{\rho}(s)\bar{x}(s)], \bar{u}(s))ds\right), \quad 0 \le t \le T, \tag{31}$$

and

$$v^\theta(0) = E[\psi_T^\theta] = \exp(\theta Z_0).$$

Moreover, the process Z is the first component of the $\mathbb{F}$-adapted pair of processes (Z, ℓ) which is the unique solution to the following quadratic BSDE:

$$\begin{cases} dZ_t = -\{f(t, \bar{x}(t), E[\bar{\rho}(s)\bar{x}(s)], \bar{u}(t)) + \frac{\theta}{2}|\ell(t)|^2\}dt + \langle \ell(t), dB_t\rangle, \\ Z_T = \frac{1}{\theta}\ln\bar{\rho}(T) + h(\bar{x}_T, E[\bar{\rho}(T)\bar{x}(T)]). \end{cases} \tag{32}$$

where, $\ell(t) = (\ell_1(t), \ell_2(t))$ satisfies

$$E\left[\int_0^T |\ell(t)|^2 dt\right] < \infty. \tag{33}$$

In particular, v^θ solves the following linear backward SDE

$$dv^\theta(t) = \theta v^\theta(t)\langle \ell(t), dB_t\rangle, \quad v^\theta(T) = \psi_T^\theta. \tag{34}$$

Hence,

Proof of Lemma 1. In view of (30),

$$\frac{v^{\theta}(t)}{v^{\theta}(0)} = \exp\left(\int_0^t \theta\langle \ell(s), dB_s\rangle - \frac{\theta^2}{2}\int_0^t |\ell(s)|^2 ds\right) := L_t^{\theta}, \quad 0 \le t \le T, \qquad (35)$$

is a uniformly integrable $\mathbb{F}$-martingale. □

To identify the processes $\tilde{\alpha}$ and $\tilde{q}$ such that

$$d\hat{p}(t) = -\hat{\alpha}(t)dt + \hat{q}(t)dB_t,$$

we may apply Itô's formula to the process $p(t) = \theta v^{\theta}\tilde{p}(t)$, use (21) and (34) and identify the coefficients. We obtain

$$\begin{cases} \hat{\alpha}(t) = \frac{1}{\theta v^{\theta}(t)}\begin{pmatrix} H_{\rho}^{e}(t) + E[\check{H}_m^{e}(t)] + \bar{x}(t)E[f_m(t)p_3(t)] \\ H_x^{e}(t) + E[\check{H}_m^{e}(t)] + \bar{\rho}(t)E[f_m(t)p_3(t)] \\ 0 \end{pmatrix} + \theta\hat{q}(t)\ell(t), \\ \hat{q}(t) = \frac{1}{\theta v^{\theta}(t)}q(t) - \theta\hat{p}(t)\ell(t). \end{cases} \qquad (36)$$

Therefore,

$$\begin{cases} d\hat{p}(t) = -\frac{1}{\theta v^{\theta}(t)}\begin{pmatrix} H_{\rho}^{e}(t) + E[\check{H}_m^{e}(t)] + \bar{x}(t)E[f_m(t)p_3(t)] \\ H_x^{e}(t) + E[\check{H}_m^{e}(t)] + \bar{\rho}(t)E[f_m(t)p_3(t)] \\ 0 \end{pmatrix} dt + \hat{q}(t)dB_t^{\theta}, \\ \hat{q}(t) = \frac{1}{\theta v^{\theta}(t)}q(t) - \theta\hat{p}(t)\ell(t), \\ dv^{\theta}(t) = \theta v^{\theta}(t)\langle \ell(t), dB_t\rangle, \\ \hat{p}(T) = -\begin{pmatrix} (\theta\bar{\rho}(T))^{-1} \\ h_x(T) \\ 1 \end{pmatrix} - \begin{pmatrix} \bar{x}(T) \\ \bar{\rho}(T) \\ 0 \end{pmatrix}\frac{1}{\psi_T^{\theta}}E[\psi_T^{\theta}h_m(T)], \\ v^{\theta}(T) = \psi_T^{\theta}, \end{cases} \qquad (37)$$

where, $B_t^{\theta} := B_t - \theta\int_0^t \ell(s)ds, \ 0 \le t \le T$, which is, in view of (35) and Girsanov's Theorem, a $\mathbb{P}^{\theta}$-Brownian motion, where $\left.\frac{d\mathbb{P}^{\theta}}{d\mathbb{P}}\right|_{\mathscr{F}_t} := L_t^{\theta}$.

In particular,

$$d\hat{p}_3(t) = \langle \hat{q}_3(t), -\theta\ell(t)dt + dB_t\rangle, \quad \hat{p}_3(T) = -1.$$

Therefore, noting that $\hat{p}_3(t) := [\theta v^\theta(t)]^{-1} p_3(t)$ is square-integrable, we obtain

$$\hat{p}_3(t) = E^{\mathbb{P}^\theta}[\hat{p}_3(T)|\mathscr{F}_t] = -1.$$

Thus, its quadratic variation becomes $\int_0^T |\hat{q}_3(t)|^2 dt = 0,\ \mathbb{P}^\theta -$ a.s. This implies that, for almost every $0 \le t \le T$, $\hat{q}_3(t) = 0,\ \mathbb{P}^\theta$ and $\mathbb{P} -$ a.s.

Hence, we can drop the last components from the adjoint processes $(\hat{p}, \hat{q})$ and only consider (keeping the same notation)

$$\hat{p} := \begin{pmatrix} \hat{p}_1 \\ \hat{p}_2 \end{pmatrix}, \quad \hat{q} := \begin{pmatrix} \hat{q}_{11} & \hat{q}_{12} \\ \hat{q}_{21} & \hat{q}_{22} \end{pmatrix},$$

for which (37) reduces to the risk-sensitive adjoint equation:

$$\begin{cases} d\hat{p}(t) &= -\frac{1}{\theta v^\theta(t)} \begin{pmatrix} H^e_\rho(t) + E[\check{H}^e_m(t)] - \bar{x}(t)E[f_m(t)] \\ H^e_x(t) + E[\check{H}^e_m(t)] - \bar{\rho}(t)E[f_m(t)] \end{pmatrix} dt + \hat{q}(t)dB^\theta_t, \\ \hat{q}(t) &= \frac{1}{\theta v^\theta(t)} q(t) - \theta\hat{p}(t)\ell(t), \\ dv^\theta(t) &= \theta v^\theta(t)\langle \ell(t), dB_t\rangle, \\ \hat{p}(T) &= -\begin{pmatrix} (\theta\bar{\rho}(T))^{-1} \\ h_x(T) \end{pmatrix} - \begin{pmatrix} \bar{x}(T) \\ \bar{\rho}(T) \end{pmatrix} \frac{1}{\psi^\theta_T} E[\psi^\theta_T h_m(T)], \\ v^\theta(T) &= \psi^\theta_T. \end{cases} \tag{38}$$

In view of the uniqueness of $\mathbb{F}$-adapted pairs (p, q), solution of (21), and the pair (v^θ, ℓ) obtained satisfying (33) and (34), the solution of the system of backward SDEs (38) is unique and satisfies (15).

3.3 Risk-Sensitive Stochastic Maximum Principle

We may use the transform (27) and (36) to obtain the explicit form (11) of the risk-sensitive Hamiltonian H^θ defined by

$$H^\theta(t, \bar{X}(t), \hat{p}(t), \hat{q}(t), \ell(t), u) := \frac{1}{\theta v^\theta(t)} H^e(t, \bar{R}(t), p(t), q(t), u), \tag{39}$$

where, H^e is defined by (19).
Let

$$\delta H^e(t) := H^e(t, \bar{R}(t), p(t), q(t), u) - H^e(t, \bar{R}(t), p(t), q(t), \bar{u}(t))$$

and

$$\delta H^{\theta}(t) = H^{\theta}(t, \bar{X}(t), \hat{p}(t), \hat{q}(t), \ell(t), u) - H^{\theta}(t, \bar{X}(t), \hat{p}(t), \hat{q}(t), \ell(t), \bar{u}(t)).$$

We have

$$E[\delta H^{e}(t)|\mathscr{F}_t^Y] = \theta E[v^{\theta}(t)\delta H^{\theta}(t)|\mathscr{F}_t^Y] = \theta v^{\theta}(0) E^{\theta}[\delta H^{\theta}(t)|\mathscr{F}_t^Y],$$

where, we recall that $v^{\theta}(t)/v^{\theta}(0) = L_t^{\theta} = d\mathbb{P}^{\theta}/d\mathbb{P}|_{\mathscr{F}_t}$.

Now, since $\theta > 0$ and $v^{\theta}(0) = E[\psi_T^{\theta}] > 0$, the variational inequality (1) translates into

$$E^{\theta}[H^{\theta}(t, \bar{\rho}(t), \bar{x}(t), \hat{p}(t), \hat{q}(t), \ell(t), u) - H^{\theta}(t, \bar{\rho}(t), \bar{x}(t), \hat{p}(t), \hat{q}(t), \ell(t), \bar{u}(t))|\mathscr{F}_t^Y] \leq 0.$$

for all $u \in U$, almost every t and $\mathbb{P}^{\theta}-$almost surely. This finishes the proof of Theorem 1.

4 Illustrative Example: Linear-Quadratic Risk-Sensitive Model Under Partial Observation

To illustrate our approach, we consider a one-dimensional linear diffusion with exponential quadratic cost functional. Perhaps, the easiest example of a linear-quadratic (LQ) risk-sensitive control problem with mean-field coupling is

$$\begin{cases} \inf_{u(\cdot)\in\mathscr{U}} E^u e^{\theta\left[\frac{1}{2}\int_0^T u^2(t)dt + \frac{1}{2}x^2(T) + \mu E^u[x(T)]\right]}, \\ \text{subject to} \\ dx(t) = (ax(t) + bu(t))\, dt + \sigma dW_t + \alpha d\widetilde{W}_t^u, \\ dY_t = \beta x(t)dt + d\widetilde{W}_t^u, \\ x(0) = x_0,\ Y_0 = 0, \end{cases}$$

where, a, b, α, β, μ and σ are real constants.

In this section we will illustrate our approach by only considering the LQ risk-sensitive control under partial observation without the mean-field coupling i.e. ($\mu = 0$) so that our result can be compared with [8] where a similar example (in many dimensions) is studied using the Dynamic Programming Principle. The case $\mu \neq 0$ can treated in a similar fashion (cf. [11]).

We consider the linear-quadratic risk-sensitive control problem:

$$\begin{cases} \inf_{u(\cdot)\in\mathscr{U}} E^u e^{\theta\left[\frac{1}{2}\int_0^T u^2(t)dt+\frac{1}{2}x^2(T)\right]}, \\ \text{ subject to} \\ dx(t) = (ax(t)+bu(t))\,dt+\sigma dW_t+\alpha d\widetilde{W}_t^u, \\ dY_t = \beta x(t)dt + d\widetilde{W}_t^u, \\ x(0) = x_0,\ Y_0 = 0, \end{cases} \tag{40}$$

where, a, b, α, β and σ are real constants.

An admissible process $(\bar{\rho}(\cdot), \bar{x}(\cdot), \bar{u}(\cdot))$ satisfying the necessary optimality conditions of Theorem 1 is obtained by solving the following system of forward-backward SDEs (cf. (5) and (14)) (see Remark 1, above).

$$\begin{cases} d\bar{\rho}(t) &= \beta\bar{\rho}(t)\bar{x}(t)dY_t, \\ d\bar{x}(t) &= \{c\bar{x}(t)+b\bar{u}(t)\}\,dt+\sigma dW_t+\alpha dY_t, \\ dp(t) &= -\begin{pmatrix} H_\rho^\theta(t) \\ H_x^\theta(t) \end{pmatrix} dt + q(t)(-\theta\ell(t)dt+dB_t), \\ dv^\theta(t) &= \theta v^\theta(t)\langle \ell(t), dB_t\rangle, \\ p(T) &= -\begin{pmatrix} (\theta\bar{\rho}(T))^{-1} \\ \bar{x}(T) \end{pmatrix}, \\ v^\theta(T) &= \psi_T^\theta, \\ \bar{\rho}(0) &= 1,\ \ \bar{x}(0) = x_0, \end{cases} \tag{41}$$

where,

$$c := a-\alpha\beta,\ \ B_t := \begin{pmatrix} Y_t \\ W_t \end{pmatrix},\ \ \ell := \begin{pmatrix} \ell_1 \\ \ell_2 \end{pmatrix},\ \ p := \begin{pmatrix} p_1 \\ p_2 \end{pmatrix},\ \ q := \begin{pmatrix} q_{11}\ q_{12} \\ q_{21}\ q_{22} \end{pmatrix},$$

$$\psi_T^\theta := \bar{\rho}(T)e^{\theta\left[\frac{1}{2}\int_0^T \bar{u}^2(t)dt+\frac{1}{2}\bar{x}^2(T)\right]},$$

and the associated risk-sensitive Hamiltonian is

$$\begin{aligned} H^\theta(t,\rho,x,u,p,q,\ell) := (cx+bu)p_2 - \tfrac{1}{2}u^2 + \rho\beta x(q_{11}+\theta\ell_1 p_1) \\ +\alpha(q_{21}+\theta\ell_2 p_1)+\sigma(q_{22}+\theta\ell_2 p_2). \end{aligned} \tag{42}$$

In general the solution (v^θ, ℓ) primarily gives the correct form of the process ℓ which may be a function of the optimal control $\bar{u}$. Inserting ℓ in the BSDE satisfied by (p, q) in the system (41) and solving for (p, q), we arrive at the characterization the optimal control of our problem.

For the LQ-control problem it turns out that by considering the BSDE satisfied by (v^θ, ℓ), we will find an explicit form of the optimal control $\bar{u}$. Indeed, by (31), this is equivalent to consider the BSDE satisfied by (Z, ℓ):

$$\begin{cases} dZ_t &= -\{\frac{1}{2}\bar{u}^2(t) + \frac{\theta}{2}|\ell(t)|^2\}dt + \langle \ell(t), dB_t\rangle, \\ Z_T &= \frac{1}{\theta}\ln \bar{\rho}(T) + \frac{1}{2}\bar{x}_T^2. \end{cases}$$

Since $\bar{u}$ is $\mathscr{F}_t^Y$, the form of Z_T suggests that we characterize $\bar{u}$ and ℓ such that

$$E^\theta[Z_t|\mathscr{F}_t^Y] = E^\theta[\frac{\gamma(t)}{2}\bar{x}^2(t) + \frac{1}{\theta}\ln \bar{\rho}(t) + \eta(t)|\mathscr{F}_t^Y], \quad 0 \le t \le T,$$

where, γ and η are deterministic functions such that $\gamma(T) = 1$ and $\eta(T) = 0$. In view of the SDEs satisfied by $(\bar{\rho}, \bar{x})$ in (41), applying Itô's formula and identifying the coefficients, we get

$$\ell_1(t) = (\alpha\gamma(t) + \beta/\theta)\bar{x}(t), \quad \ell_2(t) = \sigma\gamma(t)\bar{x}(t) \tag{43}$$

and

$$\begin{aligned} E^\theta[\tfrac{1}{2}\left(\dot{\gamma}(t) + 2(c+\alpha\beta)\gamma(t) + (\theta(\sigma^2+\alpha^2) - b^2)\gamma^2(t)\right)\bar{x}^2(t)|\mathscr{F}_t^Y] \\ + E^\theta[\dot{\eta}(t) + \tfrac{1}{2}(\sigma^2+\alpha^2)\gamma(t) + (\bar{u}(t) + b\gamma(t)\bar{x}(t))^2|\mathscr{F}_t^Y] = 0. \end{aligned}$$

Hence,

$$\begin{cases} \dot{\gamma}(t) + 2(c+\alpha\beta)\gamma(t) + (\theta(\sigma^2+\alpha^2) - b^2)\gamma^2(t) = 0, \quad \gamma(T) = 1, \\ \dot{\eta}(t) + \frac{1}{2}(\sigma^2+\alpha^2)\gamma(t) = 0, \quad \eta(T) = 0, \end{cases} \tag{44}$$

where, the first equation is the risk-sensitive Riccati equation, and

$$E^\theta[(\bar{u}(t) + b\gamma(t)\bar{x}(t))^2|\mathscr{F}_t^Y] = 0.$$

By the conditional Jensen's inequality, we have

$$\left|E^\theta[\bar{u}(t) + b\gamma(t)\bar{x}(t)|\mathscr{F}_t^Y]\right|^2 \le E^\theta[(\bar{u}(t) + b\gamma(t)\bar{x}(t))^2|\mathscr{F}_t^Y].$$

Therefore, the optimal control is

$$\bar{u}(t) = -b\gamma(t)E^\theta[\bar{x}(t)|\mathscr{F}_t^Y], \tag{45}$$

and the optimal dynamics solves the linear SDE

$$d\bar{x}(t) = \left(c\bar{x}(t) - b^2\gamma(t)E^\theta[\bar{x}(t)|\mathscr{F}_t^Y]\right)dt + \sigma dW_t + \alpha dY_t, \quad \bar{x}(0) = x_0, \tag{46}$$

where, by the filter equation of Theorem 8.1 in [22], $\pi_t(\bar{x}) := E^\theta[\bar{x}(t)|\mathscr{F}_t^Y]$ is the solution of the SDE on $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P}^\theta)$:

$$\pi_t(\bar{x}) = x_0 + \int_0^t (c - b^2\gamma(s))\pi_s(\bar{x})ds + \int_0^t \left(\alpha + (\theta\alpha\gamma(t) + \beta)\left[\pi_s(\bar{x}^2) - \pi_s^2(\bar{x})\right]\right) d\bar{Y}_s^\theta,$$

where, $\bar{Y}_t^\theta = Y_t - \int_0^t (\theta\alpha\gamma(s) + \beta)\pi_s(\bar{x})ds$ is an $(\Omega, \mathscr{F}, \mathbb{F}^Y, \mathbb{P}^\theta)$-Brownian motion. Inserting the form (43) of ℓ in the BSDE satisfied by (p, q) in the system (41) and solving for (p, q), we arrive at the same characterization the optimal control of our problem, obtained as a maximizer of the associated H^θ given by (42). We sketch the main steps and omit the details.

We have

$$H_u^\theta = bp_2 - u, \quad H_\rho^\theta = \beta x(q_{11} + \theta\ell_1 p_1), \quad H_x^\theta = cp_2 + \beta\rho(q_{11} + \theta\ell_1 p_1).$$

The BSDE satisfied by (p, q) then reads

$$\begin{cases} dp_1(t) = -\{q_{11}(t)(\beta\bar{x}(t) + \theta\ell_1(t)) + \theta(\ell_1(t)p_1(t)\bar{x}(t) + q_{12}(t)\ell_2(t))\}\,dt \\ \qquad\qquad + q_{11}(t)dY_t + q_{12}(t)dW_t, \\ dp_2(t) = -\{cp_2(t) + \beta\rho(t)(q_{11}(t) + \theta\ell_1(t)p_1(t))\}\,dt \\ \qquad\qquad + \theta(q_{21}\ell_1(t) + q_{22}\ell_2(t))dt + q_{21}(t)dY_t + q_{22}(t)dW_t, \\ p_1(T) = -\frac{1}{\theta\bar{\rho}(T)}, \quad p_2(T) = -\bar{x}(T). \end{cases} \tag{47}$$

In view of Theorem 1, if $\bar{u}$ is an optimal control of the system (40), it is necessary that

$$E^\theta[bp_2(t) - \bar{u}(t)|\mathscr{F}_t^Y] = 0.$$

This yields

$$\bar{u}(t) = bE^\theta[p_2(t)|\mathscr{F}_t^Y].$$

The associated state dynamics $\bar{x}$ solves then the SDE

$$d\bar{x}(t) = \left\{c\bar{x}(t) + b^2 E^\theta[p_2(t)|\mathscr{F}_t^Y]\right\} dt + \sigma dW_t + \alpha dY_t.$$

It remains to compute $E^\theta[p_2(t)|\mathscr{F}_t^Y]$. Indeed, inserting the form (43) of ℓ in the BSDE satisfied by (p, q) in the system (47), by Itô's formula and identifying the coefficients, it is easy to check that $(p_1(t), q_{11}(t), q_{12}(t))$ given by

$$p_1(t) := -\frac{1}{\theta\bar{\rho}(t)}, \quad q_{11}(t) := \frac{\beta}{\theta}\frac{\bar{x}(t)}{\bar{\rho}(t)}, \quad q_{12}(t) := 0$$

solves the first adjoint equation in (47). Furthermore, since $p_2(T) = -\bar{x}(T)$, setting

$$E^\theta[p_2(t)|\mathscr{F}_t^Y] = -\lambda(t)E^\theta[\bar{x}(t)|\mathscr{F}_t^Y],$$

where, λ is a deterministic function such that $\lambda(T) = 1$, and identifying the coefficients, we find that λ satisfies the risk-sensitive Riccati equation in (44). Moreover,

$$q_{21}(t) = -\sigma\lambda(t), \quad q_{22}(t) = -\alpha\lambda(t).$$

By uniqueness of the solution of the risk-sensitive Riccati equation in (44), it follows that $\lambda = \gamma$. Therefore,

$$E^\theta[p_2(t)|\mathscr{F}_t^Y] = -\gamma(t)E^\theta[\bar{x}(t)|\mathscr{F}_t^Y], \quad q_{21}(t) = -\sigma\gamma(t), \quad q_{22}(t) = -\alpha\gamma(t).$$

Summing up: the optimal control of the LQ-problem (41) is

$$\bar{u}(t) = -b\gamma(t)E^\theta[\bar{x}(t)|\mathscr{F}_t^Y], \tag{48}$$

where, γ solves the risk-sensitive Riccati equation

$$\dot{\gamma}(t) + 2(c+\alpha\beta)\gamma(t) + (\theta(\sigma^2+\alpha^2) - b^2)\gamma^2(t) = 0, \quad \gamma(T) = 1. \tag{49}$$

The optimal dynamics solves the linear SDE

$$d\bar{x}(t) = \left(c\bar{x}(t) - b^2\gamma(t)E^\theta[\bar{x}(t)|\mathscr{F}_t^Y]\right)dt + \sigma dW_t + \alpha dY_t, \quad \bar{x}(0) = x_0, \tag{50}$$

and the filter $\pi_t(\bar{x}) := E^\theta[\bar{x}(t)|\mathscr{F}_t^Y]$ is solution of the SDE on $(\Omega, \mathscr{F}, \mathbb{F}, \mathbb{P}^\theta)$:

$$\pi_t(\bar{x}) = x_0 + \int_0^t (c - b^2\gamma(s))\pi_s(\bar{x})ds + \int_0^t \left(\alpha + (\theta\alpha\gamma(t)+\beta)\left[\pi_s(\bar{x}^2) - \pi_s^2(\bar{x})\right]\right) d\bar{Y}_s^\theta,$$

where, $\bar{Y}_t^\theta = Y_t - \int_0^t(\theta\alpha\gamma(s)+\beta)\pi_s(\bar{x})ds$ is an $(\Omega, \mathscr{F}, \mathbb{F}^Y, \mathbb{P}^\theta)$-Brownian motion.

References

1. Andersson, D., Djehiche, B.: A maximum principle for SDE's of mean-field type. Appl. Math. Optim. **63**(3), 341–356 (2010)
2. Bensoussan, A., Sung, K.C.J., Yam, S.C.P., Yung, S.P.: Linear-quadratic mean field games, Preprint. arXiv:1404.5741, (2014)
3. Baras, J.S., Elliott, R.J., Kohlmann, M.: The partially observed stochastic minimum principle. SIAM J. Control Optim. **27**(6), 1279–1292 (1989)

4. Bensoussan, A.: Maximum principle and dynamic programming approaches of the optimal control of partially observed diffusions. Stochastics **9**, 169–222 (1983)
5. Buckdahn, R., Djehiche, B., Li, J.: A general stochastic maximum principle for SDEs of mean-field type. Appl. Math. Optim. **64**(2), 197-216 (2011)
6. Buckdahn, R., Li, J., Peng, S.: Mean-field backward stochastic differential equations and related partial differential equations. Stoch. Process. Appl. **119**(10), 3133–3154 (2009)
7. Carmona, R., Delarue, F.: Forward-backward stochastic differential equations and controlled mckean vlasov dynamics. arXiv:1303.5835, (2013)
8. Charalambous, C.D.: Partially observable nonlinear risk-sensitive control problems: dynamic programming and verification theorem. IEEE Trans. Autom. Control **42**(8), 1130–1138 (1997)
9. Charalambous, C.D., Hibey, J.: Minimum principle for partially observable nonlinear risk-sensitive control problems using measure-valued decompositions. Stochast. Stochast. Rep. **57**, 247–288 (1996)
10. Davis, M.H.A., Varaiya, P.: Dynamic programming conditions for partially observable stochastic systems. SIAM J. Control Optim. **11**(2), 226–261 (1973)
11. Djehiche, B., Tembine, H., Tempone, R.: A stochastic maximum principle for risk-sensitive mean-field type control. IEEE Trans. Autom. Control (2014). doi:10.1109/TAC.2015.2406973
12. El-Karoui, N., Hamadène, S.: BSDEs and risk-sensitive control, zero-sum and nonzero-sum game problems of stochastic functional differential equations. Stoch. Process. Appl. **107**, 145–169 (2003)
13. Fleming, W.H.: Optimal control of partially observable diffusions. SIAM J. Control Optim. **6**, 194–214 (1968)
14. Hausmann, U.G.: The maximum principle for optimal control of diffusions with partial information. SIAM J. Control Optim. **25**, 341–361 (1987)
15. Hosking, J.: A stochastic maximum principle for a stochastic differential game of a mean-field type. Appl. Math. Optim. **66**, 415–454 (2012)
16. Huang, J., Wang, G., Xiong, J.: A maximum principle for partial information backward stochastic control problems with applications. SIAM J. Control Optim. **48**, 2106–2117 (2009)
17. Jacobson, D.H.: Optimal stochastic linear systems with exponential criteria and their relation to differential games. Trans. Autom. Control AC-**18**, 124-131 (1973)
18. Jourdain, B., Méléard, S., Woyczynski, W.: Nonlinear SDEs driven by Lévy processes and related PDEs. Alea **4**, 1–29 (2008)
19. Kwakernaak, H.: A minimum principle for stochastic control problems wth output feedback. Syst. Control Lett. **1**, 74–77 (1981)
20. Li, J.: Stochastic maximum principle in the mean-field controls. Automatica **48**, 366–373 (2012)
21. Li, X., Tang, S.: General necessary conditions for partially observed optimal stochastic controls. J. Appl. Probab. **32**, 1118–1137 (1995)
22. Liptser, R.S., Shiryayev, A.N.: Statistics of random process, vol. 1. Springer, New York (1977)
23. Tang, S.: The Maximum Principle for partially observed optimal control of stochastic differential equations. SIAM J. Control Optim. **36**(5), 1596–1617 (1998)
24. Tembine, H., Zhu, Q., Basar, T.: Risk-sensitive mean-field games. IEEE Trans. Autom. Control **59**(4), 835–850 (2014)
25. Wang, G., Zhang, C., Zhang, W.: Stochastic maximum principle for mean-field type optimal control under partial information. IEEE Trans. Autom. Control **59**(2), 522–528 (2014)
26. Whittle, P.: A risk-sensitive maximum principle: the case of imperfect state observations. IEEE Trans. Autom. Control **36**, 793–801 (1991)
27. Zhou, X.Y.: On the necessary conditions of optimal control for stochastic partial differential equations. SIAM J. Control Optim. **31**, 1462–1478 (1993)

Risk Aversion in Modeling of Cap-and-Trade Mechanism and Optimal Design of Emission Markets

Paolo Falbo and Juri Hinz

Abstract According to theoretical arguments, a properly designed emission trading system should help reaching pollution reduction at low social burden based on the theoretical work of environmental economists, cap-and-trade systems are put into operations all over the world. However, the practice from emissions trading yields a real stress test for the underlying theory and reveals a number of its weak points. This paper aims to fill the gap between general welfare concepts underlying understanding of liberalized market and specific issues of real-world emission market operation. In our work, we present a novel technique to analyze emission market equilibrium in order to address diverse questions in the setting of *risk-averse* market players. Our contribution significantly upgrades all existing models in this field, which neglect risk-aversion aspects at the cost of having a wide range of singularities in their conclusions, now resolved in our approach. Furthermore, we show both how the architecture of an environmental market can be optimized under the realistic assumption of risk-aversion.

Keywords Emission markets · Social burden · Environmental policy · Risk aversion · General equilibrium · Risk-neutral measure

---

This research was partially supported under Australian Research Council's Discovery Projects funding scheme (project number: DP13010335)

P. Falbo
Department of Economics and Management, University of Brescia,
Contrada S. Chiara, 50, 25122 Brescia, BS, Italy
e-mail: falbo@eco.unibs.it

J. Hinz (✉)
Sydney - School of Mathematics, University of Technology,
P.O. Box 123, Broadway, Sydney, NSW 2007, Australia
e-mail: juri.hinz@uts.edu.au

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_10

1 Practice of the EU ETS

A properly designed emission trading system should help reducing pollution reduction with low social burden.

In this paper we understand it as a burden to the society, caused by energy production. We assume that it can be measured in monetary units including both, the overall production costs and an appropriately quantified environmental impact of energy production.

Originated from this idea, and based on the theoretical work of environmental economists, cap-and-trade systems have been put into operations all over the world.

The problem of design optimization for emission trading schemes has been addressed in [4]. This work shows that, in general, a traditional architecture of environmental markets is far from being optimal, meaning that appropriate alterations may provide significant improvements in emission reduction performances at lower social burden. Such improvements can be achieved by extending a regulatory framework, which we address below as *extended scheme*.

Let us explain this.

In the *traditional scheme*, it is assumed that the administrator allocates a predetermined allowance number to the market and sets a compliance date at which a penalty must be paid for each unit of pollutant not covered by allowances. Hence, the policy maker can exercise merely two controls, the so-called cap (total amount of allowances allocated to the market) and the penalty size. In theory, a desired pollution reduction can be reached at some costs for the society by an appropriate choice of these parameters. However, in practice, there is not much flexibility, since the cap is motivated politically and the penalty is determined to provide enough incentives for the required pollution reduction. As a result, the performance of the traditional scheme could be very poor in terms of social burden for the achieved reduction.

In an *extended scheme*, the policy maker has much more influence. The regulator can tax or subsidize the production in terms of monetary units or in terms of emission certificates. These *additional controls* can be implemented in a technology-sensitive way. Doing so, the merit order of technologies can be changed significantly. On this account, emission savings, triggered by certificate prices, also become controllable. The work [4] illustrates that, by an appropriate choice of *additional controls*, the market can reach a targeted pollution reduction at much lower social burden.

Although these theoretical findings are sound, intuitive, and practically important, the optimization of environmental market architectures could not be brought to a level suitable for practical implementation. There are two reasons for this.

(1) The existing approach [4] is based on the unrealistic assumption that each of the market players is non-risk-averse in the sense that it realizes a *linear utility function*. This assumption is not conform with the modern view and creates a number of singularities in the model. A priori, it is not even clear which conclusions of this work do hold under risk-aversion.

(2) Although the practical advantage of such market design optimization is obvious, policy makers hardly can use the theoretical findings of [4], because their quantitative assessment requires optimal control techniques whose numerics is difficult.

In this work, we address both issues, namely:

(1′) We assume a non-linear utility function for market agents and show several properties of the market equilibrium which make market design optimization possible. With this, our model is brought in line with standard economic theory and is appropriate for further developments. We also emphasize that, to capture risk-aversion, a completely new argumentation has been developed.
(2′) We provide our study in a one-period setting. Being accessible without optimal control techniques, the results become evident and potentially usable for a broad audience, including practitioners and decision makers.

The paper is organized as follows. Section 2 discusses the literature developed concerning markets of emission certificates. In Sect. 3 we introduce our equilibrium model. Section 4 deepens the analysis of the equilibrium. Section 5 studies the social optimality of the equilibrium and proves that it corresponds to the overall minimum-cost policy under a risk-neutral probability distribution. Section 6 discusses some perspectives of optimal market design. The final Sect. 7 provides conclusions.

2 Theory of Marketable Pollution Rights

The efficiency properties of environmental markets have been first addressed in [6, 10], who first advocated the principle that the "environment" is a good that can not be "consumed" for free. In particular, Montgomery describes a system of tradable certificates issued by a public authority coupled with fixing a cap to the total emissions, and, doing so, to force polluting companies paying proportionally to the environmental damage generated by their production activity. An emission certificate is representative of the permission to emit a given quantity of pollutant without being penalized. Companies with low environmental impact can sell excess certificates and the resulting revenue represents a general incentive to reduce pollution. Montgomery shows that the equilibrium price for a certificate must be driven by the cost of the most virtuous company to abate its marginal unit of pollutant. The key result of his analysis is that such a system guarantees that the reduction of pollution is distributed among the companies efficiently, that is minimizing their total costs.

After the seminal analysis of Montgomery, which is based on a deterministic and static model, the following research has taken the direction to the stochastic and multi-period settings. A literature review on the research which has developed after Montgomery's work can be found in [14]. A common result shared by all the analyses developed so far is that cap-and-trade systems indeed represent the most efficient way to reduce and control the environmental damage generated by the industrial activity.

Let us mention the contributions which are directly related to our analysis. A majority of relatively recent papers [1–5, 11, 13] are related to equilibrium models, where risk-neutral individuals optimize the expected value of their profit or cost function. The hypothesis of risk-neutrality of the agents is gracefully assumed in those contributions, since it significantly simplifies the proof that environmental markets are efficient. Some papers have considered explicitly risk-averse decision makers. One of them, [9], develops a pricing model for the spot and derivative pricing of environmental certificates in a single-period economy. In [7], the authors also develop a (multi-period) equilibrium pricing model for contingent claims depending on environmental certificates, where risk-averse agents maximize the expected utility of their profit function.

3 One-Period Equilibrium of Emission Market

To explain the emission price mechanism, we present a market model where a finite number of agents, indexed by the set I, is confronted with abatement of pollution. The key assumptions are:

- We consider a trading scheme in isolation, within a time horizon $[0, T]$, without credit transfer from and to other markets. That is, unused emission allowances expire worthless.
- There is no production strategy adjustment within the compliance period $[0, T]$. This means that the agents schedule their production plans for the entire period $[0, T]$ at the beginning. Allowances can be traded twice: at time $t = 0$ at the beginning and at time $t = T$ immediately before emission reports are surrendered to the regulator.
- For the sake of simplicity, we set the interest rate to zero.
- Each agent decides how much energy to produce and how many allowances to trade.

Note that this one-period model is best suited for our needs to explain the core mechanism of market operation and to discuss its properties. A generalization to a multi-period framework is possible, but it gives no additional insights related to the goal of this work.

The ith agent is specified by the set Ξ^i of feasible production plans for the generation of energy (electricity) within one time period from $t = 0$ to $t = T$. Further, we consider the following mappings, defined on Ξ^i, for each agent $i \in I$:

$$\xi_0^i \mapsto V_0^i(\xi_0^i), C_0^i(\xi_0^i), E_T^i(\xi_0^i),$$

with the interpretation that for production plan $\xi_0^i \in \Xi^i$, the values $V_0^i(\xi_0^i)$, $C_0^i(\xi_0^i)$, and $E_0^i(\xi_0^i)$ stand for the total production volume, the total production costs, and the total carbon dioxide emission, respectively.

Production: At time $t = 0$, each agent $i \in I$ faces the energy demand $D_0 \in \mathbb{R}_+$ of the entire market, the realized electricity price $P_0 \in \mathbb{R}_+$, and the emission allowance price $A_0 \in \mathbb{R}_+$. Based on this information, each agent decides on its production plan $\xi_0^i \in \Xi^i$, where Ξ^i is the set of feasible production plans. Given $\xi_0^i \in \Xi^i$, at time T, agent realizes the total production costs,

$$C_0^i(\xi_0^i) \in \mathbb{R}, \tag{1}$$

the production volume

$$V_0^i(\xi_0^i) \in \mathbb{R}, \tag{2}$$

and the total revenue, $P_0 V_0^i(\xi_0^i)$, from the electricity sold.

Allowance allocation: We assume that the administrator allocates a pre-determined number $\gamma_0^i \in [0, \infty[$ of allowances to each agent i.

So far, we have introduced deterministic quantities. Let us now turn to uncertainties modeled by random variables on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Emission from production: Following the production plan ξ_0^i, the total pollution of agent i is expressed as $E_T^i(\xi_0^i)$.

Remark (Randomness in demand and production) The question of randomness in energy demand and production deserves a careful argumentation. The reader may be confused by the assumption that in our one-period modeling, the time unit may correspond to the entire compliance period (which suggests a rather long time), such that our assumption on deterministic demand and unflexible production schedule appears unrealistic. To ease understanding, one shall imagine an artificial emission market model for short time period, say one day until compliance. The point of our proposal is that the elements, the arguments and the techniques required to define the optimal production plan on a daily basis are the same of those required to identify the plan ξ_0^i over a generalized period $[0, T]$. The value of this toy model is that it allows a straight-forward generalization to the multi-period situation. In our one-period modeling, we assume that the nominal energy demand D_0 is non-random and is observed at the time $t = 0$ when production decisions are made. We also suppose that the production plan ξ_0^i of each agent is deterministically scheduled at time $t = 0$. This view is in line with the current practice in energy business, where a nominal energy production volume along with a detailed schedule of production units is planed non-randomly in advance. Of course, the realized energy consumption deviates from what has been predicted. However, based on our experience in energy markets, it does not make sense to include this random factor into equilibrium modeling, since all decisions are made on the basis of a non-random demand anticipation and non-random customer's requests for energy delivery. To maintain energy consumption fluctuations in real-time, diverse auxiliary mechanisms are used. They can be considered as purely technical measures (security of supply by reserve margins). For this reason, we believe that it is natural to assume that, although the energy demand D_0 is known and production plan ξ_0^i is deterministically scheduled at time $t = 0$, the total emission, associated with this production can not be predicted with certainty at time $t = 0$ when the production and trading decisions are made. In fact, in practice the producers have to manage diverse source of randomness while following production, (demand fluctuation, outages of generators) which yields usually small but unpredictable deviations N^i from the nominal emission $E_0^i(\xi_0^i)$ associated with production plan ξ_0^i. Thus, let us agree that $E_T^i(\xi_0^i)$ is modeled as a random variable given as a sum

$$E_T^i(\xi_0^i) = E_0^i(\xi_0^i) + N^i, \qquad \xi_0^i \in \Xi^i, \; i \in I \tag{3}$$

with deterministic function

$$\xi_0^i : \Xi^i \to \mathbb{R}, \quad \xi_0^i \mapsto E_0^i(\xi_0^i)$$

describing the dependence of the nominal emission on the production plan and a random variable N^i standing for the deviation from the nominal emission. Note that the random emission $E_T^i(\xi_0^i)$ will be the only source of uncertainty in our model.

To ease our analysis, let us agree on the natural assumption that for production schedules $\xi_0^i \in \Xi_i$ $i \in I$ the total market emission $\sum_{i \in I} E_T^i(\xi_0^i)$ possesses no point masses:

$$\mathbb{P}(\sum_{i \in I} E_T^i(\xi_0^i) = z) = 0 \qquad \text{for all } z \in \mathbb{R}. \tag{4}$$

Allowance trading: At times $t = 0, T$ the allowance permits can be exchanged between agents by trading at the prices A_0 and A_T, respectively. Denote by ϑ_0^i, ϑ_T^i the change at times $t = 0, T$ of the allowance number held by agent $i \in I$. Such trading yields a revenue, which is

$$-\vartheta_0^i A_0 - \vartheta_T^i A_T. \tag{5}$$

Note that ϑ_0^i and A_0 are deterministic, whereas ϑ_T^i and A_T are modeled as random variables. Observe that sales are described by negative values of ϑ_0^i, ϑ_T^i, therefore (5) is non-negative random variable, if permits are sold.

Penalty payment: As mentioned above, the penalty $\pi \in [0, +\infty[$ must be paid at maturity T for each unit of pollutant not covered by allowances. Given the changes at times $t = 0, T$ due to allowance trading, i.e. ϑ_0^i and ϑ_T^i, the production ξ_0^i, and the total number γ_0^i of allowances allocated to agent $i \in I$, the loss of agent i due to potential penalty payment is given by

$$\pi(E_0^i(\xi_T^i) - \vartheta_0^i - \vartheta_T^i - \gamma_0^i)^+. \tag{6}$$

Individual profit: In view of (1)–(6), the profit of agent $i \in I$ following trading and production strategy $(\vartheta^i, \xi^i) = (\vartheta_0^i, \vartheta_T^i, \xi_0^i)$ depends on the market prices $(A, P) =$

(A_0, A_T, P_0) for allowances and energy and is given by

$$L^{A,P,i}(\vartheta^i, \xi^i) = -\vartheta_0^i A_0 - \vartheta_T^i A_T - C_0^i(\xi_0^i) + P_0 V_0^i(\xi_0^i) \\ -\pi(E_T^i(\xi_0^i) - \vartheta_0^i - \vartheta_T^i - \gamma_0^i)^+.$$

Note that the individual profit could be negative.

Risk-aversion and rational behavior: Suppose that the risk attitudes of each agent $i \in I$ are described by a pre-specified strictly increasing utility function $U^i : \mathbb{R} \to \mathbb{R}$. With this, the rational behavior of the agent i is targeted on the maximization of the functional

$$(\vartheta^i, \xi^i) \mapsto \mathbb{E}(U^i(L^{A,P,i}(\vartheta^i, \xi^i)))$$

over all the possible trading and production strategies $(\vartheta^i, \xi^i) = (\vartheta_0^i, \vartheta_T^i, \xi_0^i)$.

Energy demand: Suppose that at time $t = 0$ all agents observe the total energy demand, which is described by $D_0 \in \mathbb{R}_+$. Let us agree that the demand must be covered.

Market equilibrium: Following standard apprehension, a realistic market state is described by the so-called equilibrium—a situation where all allowance prices, all allowance positions, and all production decisions are such that each agent is satisfied by its own policy and, at the same time, natural restrictions are fulfilled.

Definition 1 Given energy demand $D_0 \in \mathbb{R}_+$, the prices $(A^*, P^*) = (A_0^*, A_T^*, P_0^*)$ are called equilibrium prices, if, for each agent $i \in I$, there exists a strategy $(\vartheta^{i*}, \xi^{i*}) = (\vartheta_0^{i*}, \vartheta_T^{i*}, \xi_0^{i*})$ such that:

(i) the energy demand is covered

$$\sum_{i \in I} V_0^i(\xi_0^{i*}) = D_0,$$

(ii) the emission certificates are in zero net supply

$$\sum_{i \in I} \vartheta_t^{i*} = 0 \quad \text{almost surely for}\, t = 0 \,\text{and}\, t = T, \tag{7}$$

(iii) each agent $i \in I$ is satisfied by its own policy in the sense that

$$\mathbb{E}(U^i(L^{A^*,P^*,i}(\vartheta^{i*}, \xi^{i*}))) \geq \mathbb{E}(U^i(L^{A^*,P^*,i}(\vartheta^i, \xi^i))) \tag{8}$$

holds for any alternative strategy (ϑ^i, ξ^i).

The main objective of this section is to prove that in the present model the electricity price formation is determined by the usual merit order arguments, where the effect of emission regulation causes emission allowance prices to enter the costs of production at the specific emission rate. This issue can be considered as the core mechanism of any cap-and-trade system, since including pollution costs into final product prices causes a change of the merit order of production technologies towards a cleaner production. To formulate this result, let us elaborate on the opportunity costs and introduce additional definitions.

Opportunity costs: In the economic literature, they stand for the forgone benefit from using a certain strategy compared to the next best alternative. For example, the opportunity costs of farming own land is the amount which could be obtained by renting the land to someone else. Let us explain how the opportunity costs necessarily lead to windfall profits.

When facing energy (electricity) generation, producers consider a profit, which could be potentially realized when, instead of production, unused emission allowances were sold to the market. For instance, if the price of the emission certificates is 12 € per tonne of CO2 and the production of one Megawatt-hour (MWh) emits two tonnes of CO2 (say, using a coal-fired steam turbine), then the producer must decide between two strategies which are equivalent in terms of their emission certificate balance:

- produce and sell one MWh to the market,
- do not produce this MWh and sell allowances covering two tonnes of CO2.

In this situation, the opportunity costs of producing one MWh are $2 \times 12 = 24$ €. Obviously, the agent produces energy only if the first strategy is at least as profitable as the second one. Thereby, both the production and the opportunity costs must be considered in the formation of the electricity market price. Clearly, if the production costs of electricity are 30 € per MWh, then the energy will be produced only if its price covers both the production and the opportunity costs. Thus electricity can only be delivered at a price exceeding $30 + 2 \times 12 = 54$ €. That is, in order to trigger the electricity production, the opportunity costs must be added to the production costs.

In the scientific community, this phenomenon is well-known under the name of *cost-pass-through*. An empirical analysis, see [12] confirms that the strategy of *cost-pass-through* is currently followed by the European energy producers. Furthermore, the detailed investigation of mathematical market models shows that the *cost-pass-through* is the only possible strategy in the so-called equilibrium state of the market. This can be interpreted as follows: when behaving optimally, the energy producers must pass the allowance price on to the consumers. Note that the producer obtains a windfall profit of 24 € in anycase: if electricity price is higher than 54, by passing-through the price of the certificates (that he has received for free); if the price is less than 54, by selling 2 certificates at 12 € on the emission market.

More importantly, it turns out that the *cost-pass-through* is nothing but the core mechanism responsible for the emission savings. Namely, due to the opportunity costs, clean technologies appear cheaper than emission-intense production strategies. For instance, the alternative generation technology represented by a gas turbine, which yields energy at the price of 40 € and emits only one tonne of CO2, hardly competes with a coal-fired steam turbine under generic regime (without emissions regulation). Namely, if there is no regulatory framework, then the coal-fired steam turbine is scheduled first and the gas turbine has to wait until the energy demand can not be covered by coal-fired steam technologies. However, given an emission regulation, the opposite is true: say, if the allowance price is equal to 12 € per tonne of CO2 as above, then the gas technology appears cheaper, operating at full costs of $40 + 1 \times 12 = 52$ €. Thus, the gas turbine is scheduled first, followed by the coal-fired steam turbine, which runs only if the installed gas turbine capacity does not cover the energy demand.

In the next section, we will show that the only rational behavior in equilibrium is to pass the opportunity costs on to the consumers. For this, we require additional notions.

Definition 2 Consider a given energy amount $d \in \mathbb{R}_+$ and a given allowance price $a \in \mathbb{R}_+$.
(i) Introduce the *individual opportunity merit order* costs of agent $i \in I$ as

$$\mathscr{C}^i(d, a) = \inf\{C_0^i(\xi_0^i) + aE_0^i(\xi_0^i) \ : \ \xi_0^i \in \Xi^i, \ \ V_0^i(\xi_0^i) \geq d\}.$$

An individual production plan $\xi_0^i \in \Xi^i$ is called conform with opportunity costs at emission price $a \in \mathbb{R}_+$ if

$$\mathscr{C}^i(V_0^i(\xi_0^i), a) = C_0^i(\xi_0^i) + aE_0^i(\xi_0^i).$$

that is to say ξ_0^i is confirm if it minimizes the production and emission costs among all the alternative plans offering the same generation and given emission price a.
(ii) Introduce the *cumulative opportunity merit order* costs as

$$\mathscr{C}(d, a) = \inf\{\sum_{i \in I}(C_0^i(\xi_0^i) + aE_0^i(\xi_0^i)) \ : \ \xi_0^i \in \Xi^i, \ \ i \in I, \ \ \sum_{i \in I} V_0^i(\xi_0^i) \geq d\}.$$

The production plans $\xi_0^i \in \Xi^i$, $i \in I$, are called conform with opportunity costs at emission price $a \in \mathbb{R}_+$ if

$$\mathscr{C}(\sum_{i \in I} V_0^i(\xi_0^i), a) = \sum_{i \in I} \left(C_0^i(\xi_0^i) + aE_0^i(\xi_0^i)\right).$$

(iii) Any price $p \in \mathbb{R}_+$ with the property that

$$-\mathscr{C}(\tilde{d}, a) + p\tilde{d} \leq -\mathscr{C}(d, a) + pd \text{ for all } \tilde{d} \in \mathbb{R}_+ \tag{9}$$

is referred to as an opportunity merit order electricity price at (d, a) and it can be understand as the marginal cost for the entire generation sector given the level of demand d and emission price a.

4 Properties of Equilibrium

With these definitions, we now show that, within any equilibrium, the production plans are always conform with opportunity costs. Furthermore, the equilibrium electricity price is always an opportunity merit order price.

Proposition 1 *Given energy demand D_0, let $(A^*, P^*) = (A_0^*, A_T^*, P_0^*)$ be the equilibrium prices with the corresponding strategies $(\vartheta^{i*}, \xi^{i*})$, $i \in I$, then the following points hold:*
(i) For each agent $i \in I$, the individual production plan ξ_0^{i*} is conform with opportunity costs at emission price A_0^*:

$$\mathscr{C}^i(V_0^i(\xi_0^{i*}), A_0^*) = C_0^i(\xi_0^{i*}) + A_0^* E_0^i(\xi_0^{i*}). \tag{10}$$

(ii) The market production schedule ξ_0^{i*}, $i \in I$, is conform with opportunity costs at emission price A_0^*:

$$\mathscr{C}(\sum_{i \in I} V_0^i(\xi_0^{i*}), A_0^*) = \sum_{i \in I}(C_0^i(\xi_0^{i*}) + A_0^* E_0^i(\xi_0^{i*})). \tag{11}$$

(iii) P_0^* is an opportunity merit order price at $(\sum_{i \in I} V_0^i(\xi_0^{i*}), A_0^*)$.

The direct economic consequence of this mathematical result is that each individual will organize its own production strategy by scheduling power production units in an increasing price order. Thereby their variable costs (which include the opportunity costs of using the emission certificates) are considered. Within such a schedule, a demand d is satisfied by gradually turning on the most economic plants, until a generation level matching d is reached. Furthermore, the above proposition states that such schedule is reached not only on the individual level, but also for the entire market. Namely, an overall demand d is satisfied by gradually turning on the most economic plants until reaching a production which covers the demand d. Such aggregate ordering is usually called merit order, for this reason we call $\mathscr{C}^i(d, a)$ and $\mathscr{C}(d, a)$ (agent's i) opportunity merit order costs and cumulative opportunity merit order costs, respectively.

It is worth noticing that the opportunity merit order electricity price as defined in (9) is equal to the marginal cost of generating electricity when the level of demand is d given certificate price a. Coupling this property with the merit order production in the electricity sector, implies that the most expensive production in the plan ξ_0^i will determine the marginal cost at demand level d and emission price a. The opportunity merit order electricity price at (d, a) is defined as the lowest price, which is able to trigger the required production level.

Proof (i) Consider the equilibrium strategy $(\vartheta_0^{i*}, \xi_0^{i*})$ of agent $i \in I$. Assume that the agent deviates from this strategy following an alternative production plan $\xi_0^i \in \Xi^i$. However, to keep the same emission credit balance, the difference $E_0^i(\xi_0^i) - E_0^i(\xi_0^{i*})$

is traded at the market in addition to ϑ_0^{i*}. That is, we change the equilibrium trading strategy $(\vartheta_0^{i*}, \vartheta_T^{i*})$ to an alternative trading strategy $(\vartheta_0^i, \vartheta_T^i)$ given by

$$\vartheta_0^i = \vartheta_0^{i*} + E_0^i(\xi_0^i) - E_0^i(\xi_0^{i*}), \qquad \vartheta_T^i = \vartheta_T^{i*}.$$

Note that we have changed only the initial position, from ϑ_0^{i*} to ϑ_0^i, whereas the final position is the same $\vartheta_T^i = \vartheta_T^{i*}$. A direct calculation shows that the profit of this alternative strategy $(\vartheta^i, \xi^i) = (\vartheta_0^i, \vartheta_T^i, \xi_0^i)$ can be written as

$$L^{A^*,P^*,i}(\vartheta^i, \xi^i) = L^{A^*,P^*,i}(\vartheta^{i*}, \xi^{i*}) + R(\xi_0^i, \xi_0^{i*}),$$

i.e. it differs form the original profit $L^{A^*,P^*,i}(\vartheta^{i*}, \xi^{i*})$ by the amount

$$R(\xi_0^i, \xi_0^{i*}) = P_0^*(V_0^i(\xi_0^i) - V_0^i(\xi_0^{i*})) + (C_0^i(\xi_0^{i*}) - C_0^i(\xi_0^i)) + A_0^*(E_0^i(\xi_0^{i*}) - E_0^i(\xi_0^i)).$$

Note that $R(\xi_0^i, \xi_0^{i*})$ can not be positive, since otherwise

$$L^{A^*,P^*,i}(\vartheta^i, \xi^i) > L^{A^*,P^*,i}(\vartheta^{i*}, \xi^{i*})$$

would yield

$$\mathbb{E}(U^i(L^{A^*,P^*,i}(\vartheta^i,\ \xi^i))) > \mathbb{E}(U^i(L^{A^*,P^*,i}(\vartheta^{i*}, \xi^{i*}))),$$

thus contradicting the optimality of the equilibrium strategy $(\vartheta^{i*}, \xi^{i*})$ (see (8)). Now, from $R(\xi_0^i, \xi_0^{i*}) \leq 0$ we conclude that

$$-C_0^i(\xi_0^{i*}) - A_0^* E_0^i(\xi_0^{i*}) + P_0^* V_0^i(\xi_0^{i*}) \geq -C_0^i(\xi_0^i) - A_0^* E_0^i(\xi_0^i) + P_0^* V_0^i(\xi_0^i) \tag{12}$$

for each $\xi_0^i \in \Xi^i$. With this, we conclude the desired assertion (10) as follows: any alternative production plan ξ_0^i which produces an energy amount $V_0^i(\xi_0^i)$ at least equal to $V_0^i(\xi_0^{i*})$ must satisfy

$$C_0^i(\xi_0^{i*}) + A_0^* E_0^i(\xi_0^{i*}) \leq C_0^i(\xi_0^i) + A_0^* E_0^i(\xi_0^i).$$

Thus,

$$C_0^i(\xi_0^{i*}) + A_0^* E_0^i(\xi_0^{i*}) = \inf\{C_0^i(\xi_0^i) + A_0^* E_0^i(\xi_0^i) : \xi_0^i \in \Xi^i,\ \ V_0^i(\xi_0^i) \geq V_0^i(\xi_0^{i*})\}.$$

(ii) Summing up (12) over $i \in I$, yields, for arbitrary choices of $\xi_0^i \in \Xi^i$, $i \in I$,

$$-\sum_{i\in I}\Big(C_0^i(\xi_0^{i*})+A_0^*E_0^i(\xi_0^{i*})\Big)+P_0^*\sum_{i\in I}V_0^i(\xi_0^{i*})$$
$$\geq -\sum_{i\in I}\Big(C_0^i(\xi_0^{i})+A_0^*E_0^i(\xi_0^{i})\Big)+P_0^*\sum_{i\in I}V_0^i(\xi_0^{i}). \tag{13}$$

From this, we conclude that, for any choice $\xi_0^i\in\Xi^i$, $i\in I$, of production plans satisfying

$$\sum_{i\in I}V_0^i(\xi_0^i)\geq\sum_{i\in I}V_0^i(\xi_0^{i*}),$$

it holds

$$\sum_{i\in I}\Big(C_0^i(\xi_0^{i*})+A_0^*E_0^i(\xi_0^{i*})\Big)\leq\sum_{i\in I}\Big(C_0^i(\xi_0^{i})+A_0^*E_0^i(\xi_0^{i})\Big),$$

implying the desired assertion (11).
(iii) We need to prove that, for any $\tilde{d}\in\mathbb{R}_+$,

$$-\mathscr{C}(\tilde{d},A_0^*)+P_0^*\tilde{d}\leq-\mathscr{C}(\sum_{i\in I}V_0^i(\xi_0^{i*}),A_0^*)+P_0^*\sum_{i\in I}V_0^i(\xi_0^{i*}).$$

For each choice of production strategies $\xi_0^i\in\Xi^i$, $i\in I$, estimate (13), combined with (11), gives

$$-\mathscr{C}(\sum_{i\in I}V_0^i(\xi_0^{i*}),A_0^*)+P_0^*\sum_{i\in I}V_0^i(\xi_0^{i*})$$
$$\geq-\sum_{i\in I}\Big(C_0^i(\xi_0^{i})+A_0^*E_0^i(\xi_0^{i})\Big)+P_0^*\sum_{i\in I}V_0^i(\xi_0^{i}).$$

In particular, if the strategies are chosen from

$$\{(\xi_0^i)_{i\in I}\ :\ \xi_0^i\in\Xi^i,\quad i\in I,\qquad\sum_{i\in I}V_0^i(\xi_0^i)\geq\tilde{d}\},$$

then it holds that

$$-\mathscr{C}(\sum_{i\in I}V_0^i(\xi_0^{i*}),A_0^*)+P_0^*\sum_{i\in I}V_0^i(\xi_0^{i*})\geq-\sum_{i\in I}\Big(C_0^i(\xi_0^{i})+A_0^*E_0^i(\xi_0^{i})\Big)+P_0^*\tilde{d}.$$

Passing on the right-hand side of this inequality to

$$\mathscr{C}(\tilde{d},A_0^*):=\inf\{\sum_{i\in I}(C_0^i(\xi_0^i)+A_0^*E_0^i(\xi_0^i))\ :\ \xi_0^i\in\Xi^i,\ i\in I,\ \sum_{i\in I}V_0^i(\xi_0^i)\geq\tilde{d}\},$$

yields the desired assertion

$$-\mathscr{C}(\sum_{i\in I} V_0^i(\xi_0^{i*}), A_0^*) + P_0^* \sum_{i\in I} V_0^i(\xi_0^{i*}) \geq -\mathscr{C}(\tilde{d}, A_0^*) + P_0^* \tilde{d}.$$

Remark The statement (ii) of the above proposition characterizes equilibrium in terms of aggregated quantities. Once the equilibrium is reached, the production schedule represents the cheapest way to satisfy the demand. From this perspective, the reader may conclude that the equilibrium production schedule can be obtained as a production plan which minimizes the overall costs among those which cover a given demand, indicating that only aggregated quantities do influence the equilibrium. However, we shall emphasize that the equilibrium still heavily depends on individual ingredients (such as initial endowments and risk aversion), which enter through the initial allowance price.

Now, we show another natural property of the equilibrium allowance prices. It turns out that there is no arbitrage allowance trading and that the terminal allowance price is digital.

Proposition 2 *Given energy demand D_0, let $(A^*, P^*) = (A_0^*, A_T^*, P_0^*)$ be the equilibrium prices with the corresponding strategies $(\vartheta^{i*}, \xi^{i*})$, $i \in I$,. It holds:*
(i) There exists a risk-neutral measure $\mathbb{Q}^* \sim \mathbb{P}$ such that $A^* = (A_0^*, A_T^*)$ follows a martingale with respect to $\mathbb{Q}^*$.
(ii) The terminal allowance price in equilibrium is digital

$$A_T^* = \pi 1_{\{\sum_{i\in I} E_T^i(\xi_0^{i*}) - \gamma_0 \geq 0\}}. \tag{14}$$

Proof (i) According to the first fundamental theorem of asset pricing, see [8] in discrete-time setting, the existence of the so-called equivalent martingale measure satisfying $A_0^* = \mathbb{E}^{\mathbb{Q}^*}(A_T)$ is ensured by the absence of arbitrage. Fortunately, in our framework, the absence of arbitrage follows from the equilibrium notion, as we show next. We thus conclude (i) of the above theorem and it remains to verify that the equilibrium rules out all arbitrage opportunities for allowance trading. Let us follow an indirect proof, assuming that ν_0 is an arbitrage allowance trading, meaning that

$$\mathbb{P}(\nu_0(A_T^* - A_0^*) \geq 0) = 1, \quad \mathbb{P}(\nu_0(A_T^* - A_0^*) > 0) > 0. \tag{15}$$

Based on this we obtain a contradiction by showing that each agent i can change its own policy $(\vartheta^{i*}, \xi^{i*})$ to an improved strategy $(\tilde{\vartheta}^i, \xi^{i*})$ satisfying

$$\mathbb{E}\left(U^i\left(L^{A^*,i}(\vartheta^{i*}, \xi^{i*})\right)\right) < \mathbb{E}\left(U^i(L^{A^*,i}(\tilde{\vartheta}^i, \xi^{i*}))\right). \tag{16}$$

The improvement is achieved by incorporating arbitrage ν_0 into the allowance trading of each agent i as follows:

$$\tilde{\vartheta}_0^i := \vartheta_0^{i*} + \nu_0, \qquad \tilde{\vartheta}_T^i := \vartheta_T^{i*} - \nu_0.$$

Indeed, the revenue improvement from allowance trading is

$$-\tilde{\vartheta}_0^i A_0^* - \tilde{\vartheta}_T^i A_T^* = -\vartheta_0^i A_0^* - \vartheta_T^i A_T^* + \nu_0(A_T^* - A_0^*),$$

which we combine with (15) to see that

$$\mathbb{P}\left(L^{A,i}(\vartheta^{i*}, \xi^{i*}) \leq L^{A,i}(\tilde{\vartheta}^i, \xi^{i*})\right) = 1, \qquad \mathbb{P}\left(L^{A,i}(\vartheta^{i*}, \xi^i) < L^{A,i}(\tilde{\vartheta}^i, \xi^{i*})\right) > 0,$$

which implies (16), therefore contradicting the optimality of $(\vartheta^{i*}, \xi^{i*})$.
(ii) From equilibrium property (8), it follows that for almost each $\omega \in \Omega$ the terminal allowance position adjustment $\vartheta_T(\omega)$ is a maximizer on $\mathbb{R}$ to

$$z \mapsto -zA_T^*(\omega) - \pi(E_T^i(\xi_0^{i*})(\omega) - \vartheta_0^{i*} - \gamma_0^i - z)^+. \tag{17}$$

Note that a maximizer of this mapping exists only if $0 \leq A_T^*(\omega) \leq \pi$. That is, the terminal allowance price in equilibrium must be within the interval $A_T^* \in [0, \pi]$ almost surely. Let us show now that the price actually attains only boundary values almost surely, i.e.

$$A_T^* \in \{0, \pi\} \quad \text{almost surely.} \tag{18}$$

Suppose that an intermediate value $A_T^*(\omega) \in]0, \pi[$ is taken, then the unique maximizer of function (17) is attained on $E_T^i(\xi_T^{i*})(\omega) - \vartheta_0^{i*} - \gamma_0^i$. This implies that $\vartheta_T^{i*}(\omega) = E_T^i(\xi_0^{i*})(\omega) - \vartheta_0^{i*} - \gamma_0^i$ holds for each $i \in I$, and a summation over i yields

$$\sum_{i \in I} \vartheta_T^{i*}(\omega) = \sum_{i \in I} \left(E_T^i(\xi_0^{i*})(\omega) - \vartheta_0^{i*} - \gamma_0^i\right) = \sum_{i \in I} E_T^i(\xi_0^{i*})(\omega) - \gamma_0.$$

Note that equilibrium property (7) ensures that the random variable on the left-hand side of the above equality is zero almost surely. Thus, the inclusion

$$\{\omega : A_T^* \in]0, \pi[\} \subseteq \{\omega : \sum_{i \in I} E_T^i(\xi_0^{i*}) - \gamma_0 = 0\} \tag{19}$$

holds almost surely. Because of (4), the probability of the event on the right-hand side of the above inclusion is zero, which shows (18).

If $A_T^*(\omega) = 0$, then a maximizer $\vartheta_T^{i*}(\omega)$ to the function (17) is attained on $[E_T^i(\xi_0^{i*})(\omega) - \vartheta_0^{i*} - \gamma_0^i, +\infty[$. Hence

$$\{\omega : A_T^* = 0\} \subseteq \{\omega : E_T^i(\xi_0^{i*}) - \vartheta_0^{i*} - \gamma_0^i \leq \vartheta_T^{i*}\}$$

holds almost surely for each $i \in I$, which implies that

$$\{\omega : A_T^* = 0\} \subseteq \{\omega : \sum_{i \in I} E_T^i(\xi_0^{i*}) - \gamma_0 \le \sum_{i \in I} \vartheta_T^{i*}\}$$

holds almost surely. Now, because of the equilibrium property (7), we obtain the almost sure inclusion

$$\{\omega : A_T^* = 0\} \subseteq \{\omega : \sum_{i \in I} E_T^i(\xi_0^{i*}) - \gamma_0 \le 0\}.$$

Since the probability of $A_T^* \in]0, \pi[$ is zero (19), we conclude for the complementary event that

$$\{\omega : A_T^* = \pi\} \supseteq \{\omega : \sum_{i \in I} E_T^i(\xi_0^{i*}) - \gamma_0 \ge 0\} \tag{20}$$

holds almost surely. Let us show the opposite inclusion. If $A_T^*(\omega) = \pi$, then a maximizer $\vartheta_T^{i*}(\omega)$ to function (17) is attained on $] - \infty, E_T^i(\xi_0^{i*})(\omega) - \vartheta_0^{i*} - \gamma_0^i]$. Hence,

$$\{\omega : A_T^* = \pi\} \subseteq \{E_T^i(\xi_0^{i*}) - \vartheta_0^{i*} - \gamma_0^i \ge \vartheta_T^{i*}\}$$

holds almost surely for each $i \in I$, which implies that

$$\{\omega : A_T^* = \pi\} \subseteq \{\omega : \sum_{i \in I} E_T^i(\xi_0^{i*}) - \gamma_0 \ge \sum_{i \in I} \vartheta_T^{i*}\}$$

holds almost surely. Now, because of the equilibrium property (7), we obtain

$$\{\omega : A_T^* = \pi\} \subseteq \{\omega : \sum_{i \in I} E_T^i(\xi_0^{i*}) - \gamma_0 \ge 0\}. \tag{21}$$

Finally, combine inclusions (20) and (21) to obtain assertion (14).

5 Social Optimality

To formulate social optimality, we require additional notations. Given production strategies $\xi_0^i \in \Xi^i$, $i \in I$, we denote the overall market production schedule by $\xi_0 = (\xi_0^i)_{i \in I}$ and introduce the total production costs C_0, the total production volume V_0, and the total carbon dioxide emission E_T and the total nominal carbon dioxide emission, defined by

$$C_0(\xi_0) = \sum_{i \in I} C_0^i(\xi_0^i), \quad V(\xi_0) = \sum_{i \in I} V_0^i(\xi_0^i), \quad E_T(\xi_0) = \sum_{i \in I} E_T^i(\xi_0^i), \quad E_0(\xi_0) = \sum_{i \in I} E_0^i(\xi_0^i).$$

Having in mind that $C_0(\xi_0)$ stands for the overall costs of the production and interpreting $\pi(E_T(\xi_0) - \gamma_0)^+$ as a proxy of the environmental impact of the production schedule ξ_0, let us agree that

$$B(\xi_0) = C_0(\xi_0) + \pi(E_T(\xi_0) - \gamma_0)^+$$

expresses the social burden caused by the overall production plan $\xi_0 \in \times_{i \in I} \Xi^i$.

It turns out that the equilibrium strategy minimizes the social burden among all production strategies which cover a given demand.

Proposition 3 *Given energy demand D_0, let $(A^*, P^*) = (A_0^*, A_T^*, P_0^*)$ be the equilibrium prices with the corresponding strategies $(\vartheta^{i*}, \xi^{i*})$, $i \in I$. Let $\mathbb{Q}^*$ be a risk-neutral measure whose existence is shown in Proposition 2. Then*

$$E_0^{\mathbb{Q}^*}(B(\xi_0^*)) \leq E_0^{\mathbb{Q}^*}(B(\xi_0)) \tag{22}$$

holds for each production schedule $\xi_0 = (\xi_0^i)_{i \in I} \in \times_{i \in I} \Xi^i$ which yields at least the same production volume, $V_0(\xi_0) \geq V_0(\xi_0^*) = D_0$.

Proof For each convex function $f : \mathbb{R} \to \mathbb{R}, x \mapsto f(x)$, it holds $f(x) + \nabla f(x) h \leq f(x+h)$, $h \in \mathbb{R}$, where $\nabla f(x)$ stands for one of the sub-gradients of f at the point x. In particular, for convex function $f : \mathbb{R} \to \mathbb{R}_+$, $x \mapsto x^+$, we obtain $x^+ + 1_{\{x \geq 0\}} h \leq (x+h)^+$ for all $x, h \in \mathbb{R}$. With the equilibrium production strategy $\xi_0^* = (\xi_0^{i*})_{i \in I}$, we conclude that

$$(E_T(\xi_0^*) - \gamma_0)^+ + 1_{\{E_T(\xi_0^*) - \gamma_0 \geq 0\}}(E_T(\xi_0) - E_T(\xi_0^*)) \leq (E_T(\xi_0) - \gamma_0)^+$$

holds almost surely for any production strategy $\xi_0 \in \times_{i \in I} \Xi^i$. Using our model assumption (3) we deduce $E_T(\xi_0) - E_T(\xi_0^*) = E_0(\xi_0) - E_0(\xi_0^*)$ which gives

$$(E_T(\xi_0^*) - \gamma_0)^+ + 1_{\{E_T(\xi_0^*) - \gamma_0 \geq 0\}}(E_0(\xi_0) - E_0(\xi_0^*)) \leq (E_T(\xi_0) - \gamma_0)^+.$$

Calculating the expectations with respect to $\mathbb{Q}^*$ on both sides and multiplying both sides by π, we obtain

$$\begin{aligned} \pi \mathbb{E}^{\mathbb{Q}^*}\left((E_T(\xi_0^*) - \gamma_0)^+\right) + \pi \mathbb{E}^{\mathbb{Q}^*}\left(1_{\{E_T(\xi_0^*) - \gamma_0 \geq 0\}}\right)(E_0(\xi_0) - E_0(\xi_0^*)) \leq \\ \leq \pi \mathbb{E}^{\mathbb{Q}^*}\left((E_T(\xi_0) - \gamma_0)^+\right). \end{aligned}$$

Using the martingale property and the digital terminal value of the equilibrium allowance prices shown in Proposition 2, we finally obtain

$$\pi \mathbb{E}^{\mathbb{Q}^*}\left((E_T(\xi_0^*) - \gamma_0)^+\right) + A_0^*(E_0(\xi_0) - E_0(\xi_0^*)) \leq \pi \mathbb{E}^{\mathbb{Q}^*}\left((E_T(\xi_0) - \gamma_0)^+\right). \tag{23}$$

For the case that the strategy ξ_0 yields at least the total production volume of the equilibrium strategy, $V_0(\xi_0) \geq V_0(\xi_0^*)$, assertion (11) in Proposition 1 yields the estimate

$$C_0(\xi_0^*) + A_0^* E_0(\xi_0^*) \leq C_0(\xi_0) + A_0^* E_0(\xi_0),$$

which is equivalent to

$$C_0(\xi_0^*) - C_0(\xi_0) \leq A_0^*(E_0(\xi_0) - E_0(\xi_0^*)).$$

Now, combining the last inequality with (23), we obtain

$$C_0(\xi_0^*) + \pi \mathbb{E}^{\mathbb{Q}^*}\left((E_T(\xi_0^*) - \gamma_0)^+\right) \leq C_0(\xi_0) + \pi \mathbb{E}^{\mathbb{Q}^*}\left((E_T(\xi_0) - \gamma_0)^+\right),$$

which proves our claim (22).

In Proposition 3, the equilibrium production schedule ξ_0^* was characterized as a solution to the minimization problem

$$\min\{\mathbb{E}^{\mathbb{Q}^*}(B(\xi_0)) \;:\; \xi_0 \in \times_{i \in I} \Xi^i, \quad V_0(\xi_0) \geq D_0\}. \tag{24}$$

Although this fact is about minimization of social burden, it should not be interpreted as one of the classical welfare results, which typically follow from equilibrium considerations.

An interesting point here is that this type of cost-optimality needs to be taken with great care: due to the opportunity cost-pass-through, the consumers can not expect that an (inappropriately designed) cap-and-trade mechanism indeed implements the cheapest way of emission reduction, from their perspective.

To see this point, remember that the price per unit of electricity under the merit order system includes the opportunity costs of consuming the emission certificates. Therefore, given emission price A_0 and an overall production schedule ξ_0, the consumers pay the costs $\sum_{i \in I} A_0 E^i(\xi_0^i)$ to switch in the merit order and to reduce emissions. From a global perspective, this costs stands for a wealth re-distribution. From the consumer's perspective, it is associated with a burden.

6 Equilibrium-Like Risk-Neutral Modeling

Another interesting observation from Proposition 3 is that the expectation $\mathbb{E}^{\mathbb{Q}^*}(B(\xi_0))$ of the social burden $B(\xi_0)$ is minimized with respect to a risk-neutral measure $\mathbb{Q}^*$ which differs from the objective measure $\mathbb{P}$. The measure $\mathbb{Q}^*$ is an outcome of the equilibrium, and, as such, it heavily depends on the many model components, for instance on the risk-aversions, on the certificate endowments, and on the production technologies of the agents. However, it is surprising that, once the measure $\mathbb{Q}^*$ is known, other important equilibrium outcomes can be deduced from aggregated quantities only.

In particular, given $\mathbb{Q}^*$, the equilibrium production schedule ξ_0^* can be obtained as the solution of optimization problem (24). Such solution is determined by aggregated quantities, since the social burden is by definition $B(\xi_0) = C_0(\xi_0) + \pi(E_T(\xi_0) - \gamma_0)^+$ and, apart the quantities γ_0 and π decided by the authority, it depends only on technologies present in the market. Having obtained the equilibrium production schedule ξ_0^* as the solution of optimization problem (24), the equilibrium allowance price A_0^* is calculated applying martingale pricing:

$$A_0^* = \pi \mathbb{E}^{\mathbb{Q}^*}(1_{\{E_T(\xi_0)-\gamma_0 \geq 0\}}).$$

Finally, given the production schedule ξ_0^* and the allowance price A_0^*, also the electricity price P_0^* is determined as the marginal price of the most expensive technology, which is active in the schedule ξ_0^*. Note that the opportunity costs must be included when identifying the most expensive active technology.

Summarizing, we conclude that given $\mathbb{Q}^*$, merely aggregated market parameters are needed to obtain ξ_0^*, A_0^*, and P_0^*. This observation can be used to establish and to analyze realistic equilibrium-like emission market models. Such models are needed, since in real emission trading it is nearly impossible to estimate the equilibrium from a market model, because the individual parameters are highly undetermined. For instance, within the EU ETS, there are more than 25, 000 agents, each with a specific production, its own certificate endowment, and a completely unknown risk-aversion. On the contrary, the aggregated quantities are well-known, since high-quality market data on total allowance allocation and electricity production, including capacities, costs, and emission rates, are available.

In view of this, we suggest an alternative way to estimate the market equilibrium based on aggregated quantities and using an exogenously specified proxy for risk-neutral measure $\mathbb{Q}^*$. This general approach follows the standard methodology of financial mathematics, which successfully describes the stochastic evolution of equilibrium prices on financial markets under an appropriately chosen risk-neutral measure.

6.1 Market Equilibrium Under a Risk-Neutral Measure

We sketch the following program for equilibrium-type modeling of emission markets:
(1) Determine a risk-neutral measure $\mathbb{Q}^*$, which corresponds to an equilibrium situation of the emission market in the sense of (i) of Proposition 2.
(2) Observe that, because of Proposition 3, the equilibrium production schedule ξ_0^* must be a solution to the *deterministic optimization problem*

$$\begin{aligned} &\text{minimize } C_0(\xi_0) + \pi \mathbb{E}^{\mathbb{Q}^*}((E_T(\xi_0) - \gamma_0)^+) \\ &\text{subject to } V_0(\xi_0) \geq D_0, \text{ over } \xi_0 \in \times_{i \in I} \Xi^i. \end{aligned} \tag{25}$$

To address the problem further, a specification of the space $\times_{i\in I} \Xi^i$ of market production strategies along with the functions C_0, V_0 and E_T is required.
(3) Given the equilibrium production schedule, calculate the total production costs $C_0(\xi_0^*)$, the total carbon dioxide emission $E_T(\xi_0^*)$, and the energy price P^* to assess the performance in emission reduction of the current market architecture.

Remark Note that a risk neutral measure is not unique. Clearly, finding a realistic candidate for the risk-neutral measure $\mathbb{Q}^*$ can be difficult. However, notice that one merely needs to specify the fluctuations of the non-predictable emissions under a risk-neutral measure. This distribution can be described in a parameter-dependent way, which adds flexibility to the model. For instance, having assumed a Gaussian framework under objective measure and modeling the density of the risk-neutral measure in terms of a Girsanov kernel. Given theoretical initial emission price depending on the parameters of the Girsanov kernel, these parameters shall be adjusted to match the observed emission prices. Similar techniques have been applied in financial modeling under the framework of *implicit model calibration*. Being one of the central questions in quantitative finance, the connection between risk-neutral and objective measures has been successfully addressed over the recent decades. In view of this development, modeling from a risk-neutral measure perspective can be based on a variety of different methods, ranging from benchmark approach, estimation of risk premia, state price density from portfolio optimization theory, to several econometric methods for the estimation of the so-called market price of risk.

Finally, the performance of the cap-and-trade mechanism can be examined leveraging on the dependence of the major economic indicators, i.e. total consumers' costs $P_0^* D_0$, total (producers') production costs $C_0(\xi_0^{i*})$, and total carbon dioxide emission $E_0(\xi_0^{i*})$, on the controls available to the regulator.

Note that in the standard scheme the regulator controls two key parameters: the total allowance allocation γ_0 and the penalty size π. The performance of regulation could be assessed in terms of relation between the increase of consumers' costs versus the achieved emission reduction. Such analysis may uncover and visualize inappropriate market architectures, where unlucky choices of γ_0 and π cause consumers to pay too much, compared to emission savings. Complementary or supplementary policies can be evaluated at this point next to the cap-and-trade system. In particular, different forms of subsidies and carbon tax mechanisms can have a strong impact on the merit order of different technologies.

7 Conclusions

In this paper, we show how equilibrium analysis and optimization of an environmental market can be carried out under the realistic assumption of risk-averse market players.

This generalization is based on a novel approach. Thereby, we obtain a number of interesting observations, which allow studying equilibrium market situations in terms of aggregated market quantities under a risk-neutral measure. Our findings show how market design optimization can be achieved incorporating risk-aversion. The choice to develop our approach in one-period setting, yields explicit results which constructively contribute to better understand the working principles of financial instruments and to improve both effectiveness and efficiency of environmental policy.

References

1. Barrieu, P., Fehr, M.: Integrated EUA and CER price modeling and application for spread option pricing. Centre for Climate Change Economics and Policy Working Papers 50, Centre for Climate Change Economics and Policy, London, UK (2011)
2. Carmona, R., Fehr, M.: The clean development mechanism and joint price formation for allowances and CERs. In: Dalang, R., Dozzi, M., Russo, F. (eds.) Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability, vol. 63, pp. 341–383. Springer, Basel (2011)
3. Carmona, R., Fehr, M., Hinz, J.: Optimal stochastic control and carbon price formation. SIAM J. Control Optim. **48**(4), 2168–2190 (2009)
4. Carmona, R., Fehr, M., Hinz, J., Porchet, A.: Market design for emission trading schemes. SIAM Rev. **52**(3), 403–452 (2010)
5. Chesney, M., Taschini, L.: The endogenous price dynamics of the emission allowances and an application to CO_2 option pricing. Swiss Finance Institute Research Papers 08-02, Swiss Finance Institute, Zurich, Switzerland (2008)
6. Dales, J.H.: Land, water, and ownership. Can. J. Econ. **1**(4), 791–804 (1968)
7. Hinz, J., Novikov, A.: On fair pricing of emission-related derivatives. Bernoulli **16**(4), 1240–1261 (2010)
8. Kabanov, Y., Sticker, C.: A teachers' note on no-arbitrage criteria. In: Azéma, J., Émery, M., Ledoux, M., Yor, M. (eds.) Séminaire de Probabilités XXXV. Lecture Notes in Mathematics, vol. 1755, pp. 149–152. Springer, Berlin (2001)
9. Kijima, M., Maeda, A., Nishide, K.: Equilibrium pricing of contingent claims in tradable permit markets. J. Futures Mark. **30**(6), 559–589 (2010)
10. Montgomery, W.D.: Markets in licenses and efficient pollution control programs. J. Econ. Theory **5**(3), 395–418 (1972)
11. Seifert, J., Uhrig-Homburg, M., Wagner, M.: Dynamic behavior of CO_2 spot prices. J. Environ. Econ. Manage. **56**(2), 180–194 (2008)
12. Sijm, J., Neuhoff, K., Chen, Y.: CO2 cost pass-through and windfall profits in the power sector. Clim. Policy **6**, 49–72 (2006)
13. Stevens, B., Rose, A.: A dynamic analysis of the marketable permits approach to global warming policy: a comparison of spatial and temporal flexibility. J. Environ. Econ. Manage. **44**(1), 45–69 (2002)
14. Taschini, L.: Environmental economics and modeling marketable permits. Asia-Pac. Finan. Mark. **17**(4), 325–343 (2010)

Exponential Ergodicity of the Jump-Diffusion CIR Process

Peng Jin, Barbara Rüdiger and Chiraz Trabelsi

Abstract In this paper we study the jump-diffusion CIR process (shorted as JCIR), which is an extension of the classical CIR model. The jumps of the JCIR are introduced with the help of a pure-jump Lévy process $(J_t, t \geq 0)$. Under some suitable conditions on the Lévy measure of $(J_t, t \geq 0)$, we derive a lower bound for the transition densities of the JCIR process. We also find some sufficient conditions under which the function $V(x) = x$, $x \geq 0$, is a Forster-Lyapunov function for the JCIR process. This allows us to prove that the JCIR process is exponentially ergodic.

Keywords CIR model with jumps · Exponential ergodicity · Forster-Lyapunov functions · Stochastic differential equations

MSC: 60H10 · 60J60

1 Introduction

The Cox-Ingersoll-Ross model (or CIR model) was introduced in [1] by Cox et al. in order to describe the random evolution of interest rates. The CIR model captures many features of the real world interest rates. In particular, the interest rate in the CIR model is non-negative and mean-reverting. Because of its vast applications in

---

P. Jin (✉) · B. Rüdiger
Fachbereich C, Bergische Universität Wuppertal, Gaußstraße 20,
42119 Wuppertal, Germany
e-mail: jin@uni-wuppertal.de

B. Rüdiger
e-mail: ruediger@uni-wuppertal.de

C. Trabelsi
Department of Mathematics, Faculty of Sciences of Tunis, University of Tunis El Manar,
Campus Universitaire Farhat Hached, 2092 Tunis, Tunisia
e-mail: trabelsichiraz@hotmail.fr

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_11

mathematical finance, some extensions of the CIR model have been introduced and studied, see e.g. [2, 5, 15].

In this paper we study an extension of the CIR model including jumps, the so-called jump-diffusion CIR process (shorted as JCIR). The JCIR process is defined as the unique strong solution $X := (X_t, t \geq 0)$ to the following stochastic differential equation

$$dX_t = a(\theta - X_t)dt + \sigma\sqrt{X_t}dW_t + dJ_t, \quad X_0 \geq 0, \tag{1}$$

where $a, \sigma > 0$, $\theta \geq 0$ are constants, $(W_t, t \geq 0)$ is a 1-dimensional Brownian motion and $(J_t, t \geq 0)$ is a pure-jump Lévy process with its Lévy measure ν concentrated on $(0, \infty)$ and satisfying

$$\int_{(0,\infty)} (\xi \wedge 1)\nu(d\xi) < \infty, \tag{2}$$

independent of the Brownian motion $(W_t, t \geq 0)$. The initial value X_0 is assumed to be independent of $(W_t, t \geq 0)$ and $(J_t, t \geq 0)$. We assume that all the above processes are defined on some filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F})_{t\geq 0}, P)$. We remark that the existence and uniqueness of strong solutions to (1) are guaranteed by [7, Theorem 5.1].

The term $a(\theta - X_t)$ in (1) defines a mean reverting drift pulling the process towards its long-term value θ with a speed of adjustment equal to a. Since the diffusion coefficient in the SDE (1) is degenerate at 0 and only positive jumps are allowed, the JCIR process $(X_t, t \geq 0)$ stays non-negative if $X_0 \geq 0$. This fact can be shown rigorously with the help of comparison theorems for SDEs, for more details we refer the readers to [7].

The JCIR defined in (1) includes the basic affine jump-diffusion (or BAJD) as a special case, in which the Lévy process $(J_t, t \geq 0)$ takes the form of a compound Poisson process with exponentially distributed jumps. The BAJD was introduced by Duffie and Gârleanu [2] to describe the dynamics of default intensity. It was also used in [5, 12] as a short-rate model. Motivated by some applications in finance, the long-time behavior of the BAJD has been well studied. According to [12, Theorem 3.16] and [10, Proposition 3.1], the BAJD possesses a unique invariant probability measure, whose distributional properties were later investigated in [9, 11]. We remark that the results in [10, 11] are very general and hold for a large class of affine process with state space $\mathbf{R}_+$, where $\mathbf{R}_+$ denotes the set of all non-negative real numbers. The existence and some approximations of the transition densities of the BAJD can be found in [6]. A closed formula of the transition densities of the BAJD was recently derived in [9].

In this paper we are interested in two problems concerning the JCIR defined in (1). The first one is to study the transition density estimates of the JCIR. Our first main result of this paper is a lower bound on the transition densities of the JCIR. Our idea to establish the lower bound of the transition densities is as follows. Like the BAJD, the JCIR is also an affine processes in $\mathbf{R}_+$. Roughly speaking, affine processes are

Markov processes for which the logarithm of the characteristic function of the process is affine with respect to the initial state. Affine processes on the canonical state space $\mathbf{R}_+^m \times \mathbf{R}^n$ have been investigated in [3, 5, 13, 14]. Based on the exponential-affine structure of the JCIR, we are able to compute its characteristic function explicitly. Moreover, this enables us to represent the distribution of the JCIR as the convolution of two distributions. The first distribution is known and coincides with the distribution of the CIR model. However, the second distribution is more complicated. We will give a sufficient condition such that the second distribution is singular at the point 0. In this way we derive a lower bound estimate of the transition densities of the JCIR.

The other problem we consider in this paper is the exponential ergodicity of the JCIR. According to the main results of [10] (see also [12]), the JCIR has a unique invariant probability measure π, given that some integrability condition on the Lévy measure of $(J_t, t \geq 0)$ is satisfied. Under some sharper assumptions we show in this paper that the convergence of the law of the JCIR process to its invariant probability measure under the total variation norm is exponentially fast, which is called the exponential ergodicity. Our method is the same as in [9], namely we show the existence of a Forster-Lyapunov function and then apply the general framework of [16–18] to get the exponential ergodicity.

The remainder of this paper is organized as follows. In Sect. 2 we collect some key facts on the JCIR and in particular derive its characteristic function. In Sect. 3 we study the characteristic function of the JCIR and prove a lower bound of its transition densities. In Sect. 4 we show the existence of a Forster-Lyapunov function and the exponential ergodicity for the JCIR.

2 Preliminaries

In this section we use the exponential-affine structure of the JCIR process to derive its characteristic functions.

We recall that the JCIR process $(X_t, t \geq 0)$ is defined to be the solution to (1) and it depends obviously on its initial value X_0. From now on we denote by $(X_t^x, t \geq 0)$ the JCIR process started from an initial point $x \geq 0$, namely

$$dX_t^x = a(\theta - X_t^x)dt + \sigma\sqrt{X_t^x}dW_t + dJ_t, \quad X_0^x = x. \tag{3}$$

Since the JCIR process is an affine process, the corresponding characteristic functions of $(X_t^x, t \geq 0)$ is of exponential-affine form:

$$E\big[e^{uX_t^x}\big] = e^{\phi(t,u)+x\psi(t,u)}, \quad u \in \mathscr{U} := \{u \in \mathbf{C} : \Re u \leq 0\}, \tag{4}$$

where $\Re u$ denotes the real part of u and the functions $\phi(t,u)$ and $\psi(t,u)$ in turn are given as solutions of the generalized Riccati equations

$$\begin{cases} \partial_t \phi(t,u) = F\big(\psi(t,u)\big), & \phi(0,u) = 0, \\ \partial_t \psi(t,u) = R\big(\psi(t,u)\big), & \psi(0,u) = u \in \mathscr{U}, \end{cases} \tag{5}$$

with the functions F and R given by

$$F(u) = a\theta u + \int_{(0,\infty)} (e^{u\xi} - 1)\nu(d\xi),$$
$$R(u) = \frac{\sigma^2 u^2}{2} - au.$$

Solving the system (5) gives $\phi(t,u)$ and $\psi(t,u)$ in their explicit forms:

$$\psi(t,u) = \frac{ue^{-at}}{1 - \frac{\sigma^2}{2a}u(1-e^{-at})} \tag{6}$$

and

$$\phi(t,u) = -\frac{2a\theta}{\sigma^2}\log\big(1 - \frac{\sigma^2}{2a}u(1-e^{-at})\big) + \int_0^t \int_{(0,\infty)} \Big(e^{\xi\psi(s,u)} - 1\Big)\nu(d\xi)ds. \tag{7}$$

Here the complex-valued logarithmic function $\log(\cdot)$ is understood to be its main branch defined on $\mathbf{C}\setminus\{0\}$. For $t \geq 0$ and $u \in \mathscr{U}$ we define

$$\varphi_1(t,u,x) := \big(1 - \frac{\sigma^2}{2a}u(1-e^{-at})\big)^{-\frac{2a\theta}{\sigma^2}} \exp\Big(\frac{xue^{-at}}{1 - \frac{\sigma^2}{2a}u(1-e^{-at})}\Big),$$
$$\varphi_2(t,u) := \exp\Big(\int_0^t \int_0^\infty \Big(e^{\xi\psi(s,u)} - 1\Big)\nu(d\xi)ds\Big), \tag{8}$$

where the complex-valued power function $z^{-2a\theta/\sigma^2} := \exp\big(-(2a\theta/\sigma^2)\log z\big)$ is also understood to be its main branch defined on $\mathbf{C}\setminus\{0\}$. One can notice that $\varphi_2(t,u)$ resembles the characteristic function of a compound Poisson distribution.

It follows from (4), (6) and (7) that the characteristic functions of $(X_t^x, t \geq 0)$ is given by

$$E[e^{uX_t^x}] = \varphi_1(t,u,x)\varphi_2(t,u), \qquad u \in \mathscr{U}, \tag{9}$$

where $\varphi_1(t,u,x)$ and $\varphi_2(t,u)$ are defined by (8).

According to the parameters of the JCIR process we look at two special cases:

2.1 Special Case (i): $\nu = 0$, No Jumps

Notice that the case $\nu = 0$ corresponds to the classical CIR model $(Y_t, t \geq 0)$ satisfying the following stochastic differential equation

$$dY_t^x = a(\theta - Y_t^x)dt + \sigma\sqrt{Y_t^x}dW_t, \quad Y_0^x = x \geq 0. \tag{10}$$

It follows from (9) that the characteristic function of $(Y_t^x, t \geq 0)$ coincides with $\varphi_1(t, u, x)$, namely for $u \in \mathscr{U}$

$$E[e^{uY_t^x}] = \varphi_1(t, u, x). \tag{11}$$

It is well known that the classical CIR model $(Y_t^x, t \geq 0)$ has transition density functions $f(t, x, y)$ given by

$$f(t, x, y) = \kappa e^{-u-v}\left(\frac{v}{u}\right)^{\frac{q}{2}} I_q\left(2(uv)^{\frac{1}{2}}\right) \tag{12}$$

for $t > 0$, $x > 0$ and $y \geq 0$, where

$$\kappa \equiv \frac{2a}{\sigma^2\left(1 - e^{-at}\right)}, \qquad u \equiv \kappa x e^{-at},$$

$$v \equiv \kappa y, \qquad q \equiv \frac{2a\theta}{\sigma^2} - 1,$$

and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q. For $x = 0$ the formula of the density function $f(t, x, y)$ is given by

$$f(t, 0, y) = \frac{c}{\Gamma(q+1)} v^q e^{-v} \tag{13}$$

for $t > 0$ and $y \geq 0$.

2.2 Special Case (ii): $\theta = 0$ and $x = 0$

We denote by $(Z_t, t \geq 0)$ the JCIR process given by

$$dZ_t = -aZ_t dt + \sigma\sqrt{Z_t}dW_t + dJ_t, \quad Z_0 = 0. \tag{14}$$

In this particular case the characteristic functions of $(Z_t, t \geq 0)$ is equal to $\varphi_2(t, u)$, namely for $u \in \mathscr{U}$

$$E[e^{uZ_t}] = \varphi_2(t, u). \tag{15}$$

3 A Lower Bound for the Transition Densities of JCIR

In this section we will find some conditions on the Lévy measure ν of $(J_t, t \geq 0)$ such that an explicit lower bound for the transition densities of the JCIR process given in (3) can be derived. As a first step we show that the law of X_t^x, $t > 0$, in (3) is absolutely continuous with respect to the Lebesgue measure and thus possesses a density function.

Lemma 1 *Consider the JCIR process* $(X_t^x, t \geq 0)$ *(started from* $x \geq 0$*) that is defined in* (3)*. Then for any* $t > 0$ *and* $x \geq 0$ *the law of* X_t^x *is absolutely continuous with respect to the Lebesgue measure and thus possesses a density function* $p(t, x, y)$, $y \geq 0$.

Proof As shown in the previous section, it holds

$$E[e^{uX_t^x}] = \varphi_1(t, u, x)\varphi_2(t, u) = E[e^{uY_t^x}]E[e^{uZ_t}],$$

therefore the law of X_t^x, denoted by $\mu_{X_t^x}$, is the convolution of the laws of Y_t^x and Z_t. Since $(Y_t^x, t \geq 0)$ is the well-known CIR process and has transition density functions $f(t, x, y)$, $t > 0, x, y \geq 0$ with respect to the Lebesgue measure, thus $\mu_{X_t^x}$ is also absolutely continuous with respect to the Lebesgue measure and possesses a density function.

In order to get a lower bound for the transition densities of the JCIR process we need the following lemma.

Lemma 2 *Suppose that* $\int_{(0,1)} \xi \ln(1/\xi)\nu(d\xi) < \infty$*. Then* φ_2 *defined by* (8) *is the characteristic function of a compound Poisson distribution. In particular,* $P(Z_t = 0) > 0$ *for all* $t > 0$*, where* $(Z_t, t \geq 0)$ *is defined by* (14).

Proof It follows from (6), (8) and (15) that

$$E[e^{uZ_t}] = \varphi_2(t, u) = \exp\left(\int_0^t \int_{(0,\infty)} \left(\exp\left(\tfrac{\xi u e^{-as}}{1-(\sigma^2/2a)(1-e^{-as})u}\right) - 1\right)\nu(d\xi)ds\right),$$

where $u \in \mathscr{U}$. Define

$$\Delta := \int_0^t \int_{(0,\infty)} \left(\exp\left(\frac{\xi u e^{-as}}{1-(\sigma^2/2a)(1-e^{-as})u}\right) - 1\right)\nu(d\xi)ds.$$

If we rewrite

$$\exp\left(\frac{\xi e^{-as} u}{1-(\sigma^2/2a)(1-e^{-as})u}\right) = \exp\left(\frac{\alpha u}{\beta - u}\right), \tag{16}$$

where

$$\begin{cases} \alpha := \dfrac{2a\xi}{\sigma^2(e^{as}-1)} > 0, \\ \beta := \dfrac{2ae^{as}}{\sigma^2(e^{as}-1)} > 0, \end{cases} \tag{17}$$

then we recognize that the right-hand side of (16) is the characteristic function of a Bessel distribution with parameters α and β. Recall that a probability measure $\mu_{\alpha,\beta}$ on $\big(\mathbf{R}_+, \mathscr{B}(\mathbf{R}_+)\big)$ is called a Bessel distribution with parameters α and β if

$$\mu_{\alpha,\beta}(dx) = e^{-\alpha}\delta_0(dx) + \beta e^{-\alpha-\beta x}\sqrt{\frac{\alpha}{\beta x}}I_1(2\sqrt{\alpha\beta x})dx, \tag{18}$$

where δ_0 is the Dirac measure at the origin and I_1 is the modified Bessel function of the first kind, namely

$$I_1(r) = \frac{r}{2}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{4}r^2\right)^k}{k!(k+1)!}, \qquad r \in \mathbf{R}.$$

For more properties of Bessel distributions we refer the readers to [8, Sect. 3] (see also [4, p. 438] and [9, Sect. 3]). Let $\hat{\mu}_{\alpha,\beta}$ denote the characteristic function of the Bessel distribution $\mu_{\alpha,\beta}$ with parameters α and β which are defined in (17). It follows from [9, Lemma 3.1] that

$$\hat{\mu}_{\alpha,\beta}(u) = \exp\Big(\frac{\alpha u}{\beta - u}\Big) = \exp\Big(\frac{\xi e^{-as}u}{1-(\sigma^2/2a)(1-e^{-as})u}\Big).$$

Therefore

$$\begin{aligned} \Delta &= \int_0^t \int_{(0,\infty)} \big(\hat{\mu}_{\alpha,\beta}(u) - 1\big)\nu(d\xi)ds \\ &= \int_0^t \int_{(0,\infty)} \Big(e^{\frac{\alpha u}{\beta-u}} - e^{-\alpha} + e^{-\alpha} - 1\Big)\nu(d\xi)ds. \end{aligned}$$

Set

$$\begin{aligned} \lambda &:= \int_0^t \int_{(0,\infty)} (1-e^{-\alpha})\nu(d\xi)ds \\ &= \int_0^t \int_{(0,\infty)} \Big(1 - e^{-\frac{2a\xi}{\sigma^2(e^{as}-1)}}\Big)\nu(d\xi)ds. \end{aligned} \tag{19}$$

If $\lambda < \infty$, then

$$\begin{aligned}\Delta &= \int_0^t \int_{(0,\infty)} \left(e^{\frac{\alpha u}{\beta - u}} - e^{-\alpha}\right)\nu(d\xi)ds - \lambda \\ &= \lambda\Big(\frac{1}{\lambda}\int_0^t \int_{(0,\infty)} \big(e^{\frac{\alpha u}{\beta - u}} - e^{-\alpha}\big)\nu(d\xi)ds - 1\Big).\end{aligned}$$

The fact that $\lambda < \infty$ will be shown later in this proof.

Next we show that the term $(1/\lambda)\int_0^t \int_{(0,\infty)} \big(\exp\big(\alpha u/(\beta - u)\big) - \exp(-\alpha)\big)\nu(d\xi)ds$ can be viewed as the characteristic function of a probability measure ρ. To define ρ, we first construct the following measures

$$m_{\alpha,\beta}(dx) := \beta e^{-\alpha-\beta x}\sqrt{\frac{\alpha}{\beta x}} I_1(2\sqrt{\alpha\beta x})dx, \quad x \geq 0,$$

where I_1 is the modified Bessel function of the first kind. Noticing that the measure $m_{\alpha,\beta}$ is the absolute continuous component of the measure $\mu_{\alpha,\beta}$ in (18), we easily get

$$\hat{m}_{\alpha,\beta}(u) = \hat{\mu}_{\alpha,\beta}(u) - e^{-\alpha} = e^{\frac{\alpha u}{\beta - u}} - e^{-\alpha},$$

where $\hat{m}_{\alpha,\beta}(u) := \int_0^\infty e^{ux} m_{\alpha,\beta}(dx)$ for $u \in \mathscr{U}$. Recall that the parameters α and β defined by (17) depend on the variables ξ and s. We can define a measure ρ on $\mathbf{R}_+$ as follows:

$$\rho := \frac{1}{\lambda}\int_0^t \int_{(0,\infty)} m_{\alpha,\beta}\ \nu(d\xi)ds.$$

By the definition of the constant λ in (19) we get

$$\begin{aligned}\rho(\mathbf{R}_+) &= \frac{1}{\lambda}\int_0^t \int_{(0,\infty)} m_{\alpha,\beta}(\mathbf{R}_+)\nu(d\xi)ds \\ &= \frac{1}{\lambda}\int_0^t \int_{(0,\infty)} (1 - e^{-\alpha})\nu(d\xi)ds \\ &= 1,\end{aligned}$$

i.e. ρ is a probability measure on $\mathbf{R}_+$, and for $u \in \mathscr{U}$

$$\begin{aligned}\hat{\rho}(u) &= \int_{(0,\infty)} e^{ux}\rho(dx) \\ &= \frac{1}{\lambda}\int_0^t \int_{(0,\infty)} (e^{\frac{\alpha u}{\beta - u}} - e^{-\alpha})\nu(d\xi)ds.\end{aligned}$$

Thus $\Delta = \lambda(\hat{\rho}(u) - 1)$ and $E[e^{uZ_t}] = e^{\lambda(\hat{\rho}(u)-1)}$ is the characteristic function of a compound Poisson distribution.

Now we verify that $\lambda < \infty$. Noticing that

$$
\begin{aligned}
\lambda &= \int_0^t \int_{(0,\infty)} \left(1 - e^{-\alpha}\right) \nu(d\xi) ds \\
&= \int_0^t \int_{(0,\infty)} \left(1 - e^{-\frac{2a\xi}{\sigma^2(e^{as}-1)}}\right) \nu(d\xi) ds \\
&= \int_{(0,\infty)} \int_0^t \left(1 - e^{-\frac{2a\xi}{\sigma^2(e^{as}-1)}}\right) ds \nu(d\xi),
\end{aligned}
$$

we introduce the change of variables $\dfrac{2a\xi}{\sigma^2(e^{as}-1)} := y$ and then get

$$
\begin{aligned}
dy &= -\frac{2a\xi}{\sigma^2(e^{as}-1)^2} a e^{as} ds \\
&= -y^2 \frac{\sigma^2}{2\xi} \left(\frac{2a\xi}{\sigma^2 y} + 1\right) ds.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\lambda &= \int_{(0,\infty)} \nu(d\xi) \int_{\infty}^{\frac{2a\xi}{\sigma^2(e^{at}-1)}} (1 - e^{-y}) \frac{-2\xi}{2a\xi y + \sigma^2 y^2} dy \\
&= \int_{(0,\infty)} \nu(d\xi) \int_{\frac{2a\xi}{\sigma^2(e^{at}-1)}}^{\infty} (1 - e^{-y}) \frac{2\xi}{2a\xi y + \sigma^2 y^2} dy \\
&= \int_{(0,\infty)} \nu(d\xi) \int_{\frac{\xi}{\delta}}^{\infty} (1 - e^{-y}) \frac{2\xi}{2a\xi y + \sigma^2 y^2} dy,
\end{aligned}
$$

where $\delta := \dfrac{\sigma^2(e^{at}-1)}{2a}$. Define

$$
M(\xi) := \int_{\frac{\xi}{\delta}}^{\infty} (1 - e^{-y}) \frac{2\xi}{2a\xi y + \sigma^2 y^2} dy.
$$

Then $M(\xi)$ is continuous on $(0, \infty)$. As $\xi \to 0$ we get

$$
\begin{aligned}
M(\xi) &= \int_{\frac{\xi}{\delta}}^{1} (1 - e^{-y}) \frac{2\xi}{2a\xi y + \sigma^2 y^2} dy + 2\xi \int_1^{\infty} (1 - e^{-y}) \frac{dy}{2a\xi y + \sigma^2 y^2} \\
&\le 2\xi \int_{\frac{\xi}{\delta}}^{1} \frac{y}{2a\xi y + \sigma^2 y^2} dy + 2\xi \int_1^{\infty} \frac{1}{2a\xi y + \sigma^2 y^2} dy \\
&\le 2\xi \int_{\frac{\xi}{\delta}}^{1} \frac{1}{2a\xi + \sigma^2 y} dy + 2\xi \int_1^{\infty} \frac{1}{\sigma^2 y^2} dy.
\end{aligned}
$$

Since

$$2\xi \int_{\frac{\xi}{\delta}}^{1} \frac{1}{2a\xi + \sigma^2 y} dy = \frac{2\xi}{\sigma^2}\Big[\ln(2a\xi + \sigma^2 y)\Big]_{\frac{\xi}{\delta}}^{1}$$
$$= \frac{2\xi}{\sigma^2}\ln(2a\xi + \sigma^2) - \frac{2\xi}{\sigma^2}\ln(2a\xi + \frac{\sigma^2\xi}{\delta})$$
$$\leq c_1\xi + c_2\xi\ln(\frac{1}{\xi}) \leq c_3\xi\ln(\frac{1}{\xi})$$

for sufficiently small ξ, we conclude that

$$M(\xi) \leq c_4\xi\ln(\frac{1}{\xi}), \quad \text{as } \xi \to 0.$$

If $\xi \to \infty$, then

$$M(\xi) \leq \int_{\frac{\xi}{\delta}}^{\infty}(1 - e^{-y})\frac{2\xi}{2a\xi y + \sigma^2 y^2}dy$$
$$\leq \int_{\frac{\xi}{\delta}}^{\infty}\frac{2\xi}{2a\xi y + \sigma^2 y^2}dy \leq 2\xi\int_{\frac{\xi}{\delta}}^{\infty}\frac{1}{\sigma^2 y^2}dy$$
$$= \frac{2\xi}{\sigma^2}\int_{\frac{\xi}{\delta}}^{\infty} d(-\frac{1}{y}) = \frac{2\xi}{\sigma^2}\Big[-\frac{1}{y}\Big]_{\frac{\xi}{\delta}}^{\infty}$$
$$= \frac{2\xi}{\sigma^2}\frac{\delta}{\xi} = \frac{2\delta}{\sigma^2} := c_5 < \infty.$$

Therefore,

$$\lambda \leq c_4\int_0^1 \xi\ln(\frac{1}{\xi})\nu(d\xi) + c_5\int_1^{\infty} 1\nu(d\xi) < \infty.$$

With the help of the Lemma 2 we can easily prove the following proposition.

Proposition 1 *Let $p(t, x, y)$, $t > 0, x, y \geq 0$ denote the transition density of the JCIR process $(X_t^x, t \geq 0)$ defined in (3). Suppose that $\int_{(0,1)} \xi\ln(\frac{1}{\xi})\nu(d\xi) < \infty$. Then for all $t > 0, x, y \geq 0$ we have*

$$p(t, x, y) \geq P(Z_t = 0)f(t, x, y),$$

where $P(Z_t = 0) > 0$ for all $t > 0$ and $f(t, x, y)$ are transition densities of the CIR process (without jumps).

Proof According to Lemma 2, we have $P(Z_t = 0) > 0$. Since

$$E[e^{uX_t^x}] = \varphi_1(t, u, x)\varphi_2(t, u) = E[e^{uY_t^x}]E[e^{uZ_t}],$$

the law of X_t^x, denoted by $\mu_{X_t^x}$, is the convolution of the laws of Y_t^x and Z_t. Thus for all $A \in \mathscr{B}(\mathbf{R}_+)$

$$\begin{aligned}\mu_{X_t^x}(A) &= \int_{\mathbf{R}_+} \mu_{Y_t^x}(A-y)\mu_{Z_t}(dy)\\ &\geq \int_{\{0\}} \mu_{Y_t^x}(A-y)\mu_{Z_t}(dy)\\ &\geq \mu_{Y_t^x}(A)\mu_{Z_t}(\{0\})\\ &\geq P(Z_t=0)\mu_{Y_t^x}(A)\\ &\geq P(Z_t=0)\int_A f(t,x,y)dy,\end{aligned}$$

where $f(t,x,y)$ are the transition densities of the classical CIR process given in (12). Since $A \in \mathscr{B}(\mathbf{R}_+)$ is arbitrary, we get

$$p(t,x,y) \geq P(Z_t=0)f(t,x,y)$$

for all $t > 0$, $x, y \geq 0$.

4 Exponential Ergodicity of JCIR

In this section we find some sufficient conditions such that the JCIR process is exponentially ergodic. We have derived a lower bound for the transition densities of the JCIR process in the previous section. Next we show that the function $V(x) = x$, $x \geq 0$, is a Forster-Lyapunov function for the JCIR process.

Lemma 3 *Suppose that* $\int_{(1,\infty)} \xi\nu(d\xi) < \infty$. *Then the function* $V(x) = x$, $x \geq 0$, *is a Forster-Lyapunov function for the JCIR process defined in* (3)*, in the sense that for all* $t > 0$, $x \geq 0$,

$$E[V(X_t^x)] \leq e^{-at}V(x) + M,$$

where $0 < M < \infty$ *is a constant.*

Proof We know that $\mu_{X_t^x} = \mu_{Y_t^x} * \mu_{Z_t}$, therefore

$$E[X_t^x] = E[Y_t^x] + E[Z_t].$$

Since $(Y_t^x, t \geq 0)$ is the CIR process starting from x, it is known that $\mu_{Y_t^x}$ is a non-central Chi-squared distribution and thus $E[Y_t^x] < \infty$. Next we show that $E[Z_t] < \infty$.

Let $u \in (-\infty, 0)$. By using Fatou's Lemma we get

$$
\begin{aligned}
E[Z_t] &= E\Big[\lim_{u\to 0}\frac{e^{uZ_t}-1}{u}\Big] \\
&\le \liminf_{u\to 0} E\Big[\frac{e^{uZ_t}-1}{u}\Big] = \liminf_{u\to 0}\frac{E[e^{uZ_t}]-1}{u}.
\end{aligned}
$$

Recall that

$$
E[e^{uZ_t}] = \varphi_2(t,u) = \exp\Big(\int_0^t\int_{(0,\infty)}\Big(e^{\frac{\xi u e^{-as}}{1-(\sigma^2/2a)(1-e^{-as})u}}-1\Big)\nu(d\xi)ds\Big) = e^{\Delta(u)}.
$$

Then we have for all $u \le 0$

$$
\begin{aligned}
&\frac{\partial}{\partial u}\Big(\exp\Big(\frac{\xi u e^{-as}}{1-(\sigma^2/2a)(1-e^{-as})u}\Big)-1\Big) \\
&= \frac{\xi e^{-as}}{\big(1-(\sigma^2/2a)(1-e^{-as})u\big)^2}\exp\Big(\frac{\xi u e^{-as}}{1-(\sigma^2/2a)(1-e^{-as})u}\Big) \\
&\le \frac{\xi e^{-as}}{\big(1-(\sigma^2/2a)(1-e^{-as})u\big)^2} \le \xi e^{-as}
\end{aligned}
$$

and further

$$
\int_0^t\int_{(0,\infty)}\xi e^{-as}\nu(d\xi)ds < \infty.
$$

Thus $\Delta(u)$ is differentiable in u and

$$
\Delta'(0) = \int_0^t\int_{(0,\infty)}\xi e^{-as}\nu(d\xi)ds = \frac{1-e^{-at}}{a}\int_{(0,\infty)}\xi\nu(d\xi).
$$

It follows that

$$
\begin{aligned}
E[Z_t] &\le \liminf_{u\to 0}\frac{\varphi_2(t,u)-\varphi_2(t,0)}{u} \\
&= \frac{\partial\varphi_2(t,u)}{\partial u}\Big|_{u=0} = e^{\Delta(0)}\Delta'(0) \\
&= \frac{1-e^{-at}}{a}\int_{(0,\infty)}\xi\nu(d\xi).
\end{aligned}
$$

Therefore under the assumption $\int_{(0,\infty)}\xi\nu(d\xi) < \infty$ we have proved that $E[Z_t] < \infty$. Furthermore,

$$
E[Z_t] = \frac{\partial}{\partial u}\Big(E[e^{uZ_t}]\Big)\Big|_{u=0} = \frac{1-e^{-at}}{a}\int_{(0,\infty)}\xi\nu(d\xi).
$$

On the other hand,

$$E[e^{uY_t^x}] = \big(1 - (\sigma^2/2a)u(1 - e^{-at})\big)^{-2a\theta/\sigma^2} \exp\Big(\frac{xue^{-at}}{1 - (\sigma^2/2a)u(1 - e^{-at})}\Big).$$

With a similar argument as above we get

$$E[Y_t^x] = \frac{\partial}{\partial u}\Big(E[e^{uY_t^x}]\Big)\Big|_{u=0} = \theta(1 - e^{-at}) + xe^{-at}.$$

Altogether we get

$$\begin{aligned} E[X_t^x] &= E[Y_t^x] + E[Z_t] \\ &= (1 - e^{-at})\big(\theta + \frac{1 - e^{-at}}{a}\big) + xe^{-at} \\ &\le \theta + \frac{1}{a} + xe^{-at}, \end{aligned}$$

namely

$$E[V(X_t^x)] \le \theta + \frac{1}{a} + e^{-at}V(x).$$

Remark 1 If $\int_{(1,\infty)} \xi\nu(d\xi) < \infty$, then there exists a unique invariant probability measure for the JCIR process. This fact follows from [12, Theorem 3.16] and [10, Proposition 3.1].

Let $\|\cdot\|_{TV}$ denote the total-variation norm for signed measures on $\mathbf{R}_+$, namely

$$\|\mu\|_{TV} = \sup_{A \in \mathscr{B}(\mathbf{R}_+)} \{|\mu(A)|\}.$$

Let $P^t(x, \cdot) := P(X_t^x \in \cdot)$ be the distribution of the JCIR process at time t started from the initial point $x \ge 0$. Now we prove the main result of this paper.

Theorem 1 *Assume that*

$$\int_{(1,\infty)} \xi\ \nu(d\xi) < \infty \quad \textit{and} \quad \int_{(0,1)} \xi \ln(\frac{1}{\xi})\nu(d\xi) < \infty.$$

Let π be the unique invariant probability measure for the JCIR process. Then the JCIR process is exponentially ergodic, namely there exist constants $0 < \beta < 1$ and $0 < B < \infty$ such that

$$\|P^t(x, \cdot) - \pi\|_{TV} \le B\big(x + 1\big)\beta^t, \quad t \ge 0, \quad x \in \mathbf{R}_+. \tag{20}$$

Proof Basically, we follow the proof of [18, Theorem 6.1]. For any $\delta > 0$ we consider the δ-skeleton chain $\eta_n^x := X_{n\delta}^x$, $n \in \mathbf{Z}_+$, where $\mathbf{Z}_+$ denotes the set of all non-negative integers. Then $(\eta_n^x)_{n\in\mathbf{Z}_+}$ is a Markov chain on the state space $\mathbf{R}_+$ with transition kernel $P^\delta(x,\cdot)$ and starting point $\eta_0^x = x$. It is easy to see that the measure π is also an invariant probability measure for the chain $(\eta_n^x)_{n\in\mathbf{Z}_+}$, $x \geq 0$.

Let $V(x) = x$, $x \geq 0$. It follows from the Markov property and Lemma 3 that

$$E[V(\eta_{n+1}^x)|\eta_0^x, \eta_1^x, \ldots, \eta_n^x] = \int_{\mathbf{R}_+} V(y)P^\delta(\eta_n^x, dy) \leq e^{-a\delta}V(\eta_n^x) + M,$$

where M is a positive constant. If we set $V_0 := V$ and $V_n := V(\eta_n^x)$, $n \in \mathbf{N}$, then

$$E[V_1] \leq e^{-a\delta}V_0(x) + M$$

and

$$E[V_{n+1}|\eta_0^x, \eta_1^x, \ldots, \eta_n^x] \leq e^{-a\delta}V_n + M, \quad n \in \mathbf{N}.$$

Now we proceed to show that the chain $(\eta_n^x)_{n\in\mathbf{Z}_+}$, $x \geq 0$, is λ-irreducible, strong aperiodic, and all compact subsets of $\mathbf{R}_+$ are petite for the chain $(\eta_n^x)_{n\in\mathbf{Z}_+}$.

"*λ-irreducibility*": We show that the Lebesgue measure λ on $\mathbf{R}_+$ is an irreducibility measure for $(\eta_n^x)_{n\in\mathbf{Z}_+}$. Let $A \in \mathscr{B}(\mathbf{R}_+)$ and $\lambda(A) > 0$. Then it follows from Proposition 1 that

$$P[\eta_1^x \in A|\eta_0^x = x] = P(X_\delta^x \in A) \geq P(Z_\delta = 0)\int_A f(\delta, x, y)dy > 0,$$

since $f(\delta, x, y) > 0$ for any $x \in \mathbf{R}_+$ and $y > 0$. This shows that the chain $(\eta_n^x)_{n\in\mathbf{Z}_+}$ is irreducible with λ being an irreducibility measure.

"*Strong aperiodicity*"(see [16, p. 561] for a definition): To show the strong aperiodicity of $(\eta_n^x)_{n\in\mathbf{Z}_0}$, we need to find a set $B \in \mathscr{B}(\mathbf{R}_+)$, a probability measure m with $m(B) = 1$, and $\varepsilon > 0$ such that

$$L(x, B) > 0, \qquad x \in \mathbf{R}_+, \tag{21}$$

and

$$P(\eta_1^x \in A) \geq \varepsilon m(A), \quad x \in C, \quad A \in \mathscr{B}(\mathbf{R}_+), \tag{22}$$

where $L(x, B) := P(\eta_n^x \in B$ for some $n \in \mathbf{N})$. To this end set $B := [0, 1]$ and $g(y) := \inf_{x\in[0,1]} f(\delta, x, y)$, $y > 0$. Since for fixed $y > 0$ the function $f(\delta, x, y)$ is strictly positive and continuous in $x \in [0, 1]$, thus we have $g(y) > 0$ and $0 < \int_{(0,1]} g(y)dy \leq 1$. Define

$$m(A) := \frac{1}{\int_{(0,1]} g(y)dy}\int_{A\cap(0,1]} g(y)dy, \qquad A \in \mathscr{B}(\mathbf{R}_+).$$

Then for any $x \in [0, 1]$ and $A \in \mathscr{B}(\mathbf{R}_+)$ we get

$$\begin{aligned} P(\eta_1^x \in A) &= P(X_\delta^x \in A) \\ &\geq P(Z_\delta = 0) \int_A f(\delta, x, y) dy \\ &\geq P(Z_\delta = 0) \int_{A \cap (0,1]} g(y) dy \\ &\geq P(Z_\delta = 0) m(A) \int_{(0,1]} g(y) dy, \end{aligned}$$

so (22) holds with $\varepsilon := P(Z_\delta = 0) \int_{(0,1]} g(y) dy$.

Obviously

$$L(x, [0, 1]) \geq P(\eta_1^x \in [0, 1]) = P(X_\delta^x \in [0, 1]) \geq P(Z_\delta = 0) \int_{[0,1]} f(\delta, x, y) dy > 0$$

for all $x \in \mathbf{R}_+$, which verifies (21).

"*Compact subsets are petite*": We have shown that λ is an irreducibility measure for $(\eta_n^x)_{n \in \mathbf{Z}_+}$. According to [16, Theorem 3.4(ii)], to show that all compact sets are petite, it suffices to prove the Feller property of $(\eta_n^x)_{n \in \mathbf{Z}_+}$, $x \geq 0$. But this follows from the fact that $(\eta_n^x)_{n \in \mathbf{Z}_+}$ is a skeleton chain of the JCIR process, which is an affine process and possess the Feller property.

According to [16, Theorem 6.3] (see also the proof of [16, Theorem 6.1]), the probability measure π is the only invariant probability measure of the chain $(\eta_n^x)_{n \in \mathbf{Z}_+}$, $x \geq 0$, and there exist constants $\beta \in (0, 1)$ and $C \in (0, \infty)$ such that

$$\| P^{\delta n}(x, \cdot) - \pi \|_{TV} \leq C\big(x + 1\big)\beta^n, \quad n \in \mathbf{Z}_+, \quad x \in \mathbf{R}_+.$$

Then for the rest of the proof we can proceed as in [18, p. 536] and get the inequality (20).

Acknowledgments We gratefully acknowledge the suggestions of the editors and the anonymous referee. Their valuable comments have led to a significant improvement of this manuscript. The research was supported by the Research Program "DAAD—Transformation: Kurzmaßnahmen 2012/13 Program". This research was also carried out with the support of CAS—Centre for Advanced Study, at the Norwegian Academy of Science and Letter, research program SEFE.

References

1. Cox, J.C., Ingersoll Jr, J.E., Ross, S.A.: A theory of the term structure of interest rates. Econometrica **53**(2), 385–407 (1985)
2. Duffie, D., Gârleanu, N.: Risk and valuation of collateralized debt obligations. Financ. Anal. J. **57**(1), 41–59 (2001)
3. Duffie, D., Filipović, D., Schachermayer, W.: Affine processes and applications in finance. Ann. Appl. Probab. **13**(3), 984–1053 (2003)
4. Feller, W.: An Introduction to Probability Theory and Its Applications, vol. II, 2nd edn. Wiley, New York (1971)
5. Filipović, D.: A general characterization of one factor affine term structure models. Finance Stoch. **5**(3), 389–412 (2001)
6. Filipović, D., Mayerhofer, E., Schneider, P.: Density approximations for multivariate affine jump-diffusion processes. J. Econometrics **176**(2), 93–111 (2013)
7. Fu, Z., Li, Z.: Stochastic equations of non-negative processes with jumps. Stochast. Process. Appl. **120**(3), 306–330 (2010)
8. Grigelionis, B.: Thorin classes of Lévy processes and their transforms. Lith. Math. J. **48**(3), 294–315 (2008)
9. Jin, P., Rüdiger, B., Trabelsi, C.: Positive Harris recurrence and exponential ergodicity of the basic affine jump-diffusion. ArXiv e-prints. http://arxiv.org/abs/1501.03638 (2015). Accessed 5 May 2015
10. Keller-Ressel, M.: Moment explosions and long-term behavior of affine stochastic volatility models. Math. Finance **21**(1), 73–98 (2011)
11. Keller-Ressel, M., Mijatović, A.: On the limit distributions of continuous-state branching processes with immigration. Stochast. Process. Appl. **122**(6), 2329–2345 (2012)
12. Keller-Ressel, M., Steiner, T.: Yield curve shapes and the asymptotic short rate distribution in affine one-factor models. Finance Stoch. **12**(2), 149–172 (2008)
13. Keller-Ressel, M., Schachermayer, W., Teichmann, J.: Affine processes are regular. Probab. Theor. Relat. Fields **151**(3–4), 591–611 (2011)
14. Keller-Ressel, M., Steiner, T.: Yield curve shapes and the asymptotic short rate distribution in affine one-factor models. Finance Stoch. **12**(2), 149–172 (2008)
15. Li, Z., Ma, C.: Asymptotic properties of estimators in a stable Cox-Ingersoll-Ross model. Stochast. Process. Appl. **125**(8), 3196–3233 (2015)
16. Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes. I. Criteria for discrete-time chains. Adv. Appl. Probab. **24**(3), 542–574 (1992)
17. Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes. II. Continuous-time processes and sampled chains. Adv. Appl. Probab. **25**(3), 487–517 (1993)
18. Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. Adv. Appl. Probab. **25**(3), 518–548 (1993)

Optimal Control of Predictive Mean-Field Equations and Applications to Finance

Bernt Øksendal and Agnès Sulem

Abstract We study a coupled system of controlled stochastic differential equations (SDEs) driven by a Brownian motion and a compensated Poisson random measure, consisting of a forward SDE in the unknown process $X(t)$ and a *predictive mean-field* backward SDE (BSDE) in the unknowns $Y(t)$, $Z(t)$, $K(t, \cdot)$. The driver of the BSDE at time t may depend not just upon the unknown processes $Y(t)$, $Z(t)$, $K(t, \cdot)$, but also on the predicted future value $Y(t + \delta)$, defined by the conditional expectation $A(t) := E[Y(t+\delta)|\mathscr{F}_t]$. We give a sufficient and a necessary maximum principle for the optimal control of such systems, and then we apply these results to the following two problems: (i) Optimal portfolio in a financial market with an *insider influenced asset price process.* (ii) Optimal consumption rate from a cash flow modeled as a geometric Itô-Lévy SDE, with respect to *predictive recursive utility*.

Keywords Predictive (time-advanced) mean-field BSDE · Coupled FBSDE system · Optimal control · Maximum principles · Optimal portfolio · Insider influenced financial market · Predictive recurrent utility · Utility maximizing consumption rate

MSC (2010): 60HXX · 60J65 · 60J75 · 93E20 · 91G80

B. Øksendal (✉)
Department of Mathematics, University of Oslo,
P.O. Box 1053 Blindern, N-0316 Oslo, Norway
e-mail: oksendal@math.uio.no

B. Øksendal
Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway

A. Sulem
INRIA Paris-Rocquencourt, Domaine de Voluceau, Rocquencourt,
BP 105, 78153 Le Chesnay Cedex, France
e-mail: agnes.sulem@inria.fr

A. Sulem
Université Paris-Est - Marne la Vallée, Champs-sur-Marne, France

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_12

1 Introduction

The purpose of this paper is to introduce and study a pricing model where beliefs about the future development of the price process influence its current dynamics. We think this can be a realistic assumption in price dynamics where human psychology is involved, for example in electricity prices, oil prices and energy markets in general. It can also be a natural model of the risky asset price in an insider influenced market. See Sect. 5.1.

We model such price processes as backward stochastic differential equations (BSDEs) driven by Brownian motion and a compensated Poisson random measure, where the coefficients depend not only of the current values of the unknown processes, but also on their predicted future values. These predicted values are expressed mathematically in terms of conditional expectation, and we therefore name such equations *predictive mean-field equations*. To the best of our knowledge such systems have never been studied before.

In applications to portfolio optimization in a financial market where the price process is modeled by a predictive mean-field equation, we are led to consider coupled systems of forward-backward stochastic differential equations (FBSEDs), where the BSDE is of predictive mean-field type. In this paper we study solution methods for the optimal control of such systems in terms of maximum principles. Then we apply these methods to study

(i) optimal portfolio in a financial market with an *insider influenced asset price process.* (Sect. 5.1), and
(ii) optimal consumption rate from a cash flow modeled as a geometric Itô-Lévy SDE, with respect to *predictive recursive utility* (Sect. 5.2).

2 Formulation of the Problem

We now present our model in detail. We refer to [5] for information about stochastic control of jump diffusions.

Let $B(t) = B(t, \omega)$; $(t, \omega) \in [0, \infty) \times \Omega$ and $\tilde{N}(dt, d\zeta) = N(dt, d\zeta) - \nu(d\zeta)dt$ be a Brownian motion and an independent compensated Poisson random measure, respectively, on a filtered probability space $\big(\Omega, \mathbb{E}, \mathbb{F} = \{\mathscr{F}_t\}_{t\geq 0}, P\big)$ satisfying the usual conditions. We consider a controlled system of *predictive* (time-advanced) coupled mean-field forward-backward stochastic differential equations (FBSDEs) of the form ($T > 0$ and $\delta > 0$ are given constants)

- Forward SDE in $X(t)$:

$$\begin{cases} dX(t) &= dX^u(t) = b(t, X(t), Y(t), A(t), Z(t), K(t, \cdot), u(t), \omega)dt \\ &\quad +\sigma(t, X(t), Y(t), A(t), Z(t), K(t, \cdot), u(t), \omega)dB(t) \\ &\quad + \int_{\mathbb{R}} \gamma(t, X(t), Y(t), A(t), Z(t), K(t, \cdot), u(t), \zeta, \omega)\tilde{N}(dt, d\zeta)\ ;\ t \in [0, T] \\ X(0) &= x \in \mathbb{R} \end{cases}$$

- Predictive BSDE in $Y(t)$, $Z(t)$, $K(t)$:

$$\begin{cases} dY(t) & = -g(t, X(t), Y(t), A(t), Z(t), K(t,\cdot), u(t), \omega)dt + Z(t)dB(t) \\ & \quad + \int_{\mathbb{R}} K(t,\zeta)\tilde{N}(dt, d\zeta)\ ;\ t \in [0,T) \\ Y(T) & = h(X(T), \omega). \end{cases} \tag{1}$$

We set

$$Y(t) := L\ ;\ t \in (T, T+\delta], \tag{2}$$

where L is a given bounded $\mathscr{F}$-measurable random variable, representing a "cemetery" state of the process Y after time T. The process $A(t)$ represents our *predictive mean-field* term. It is defined by

$$A(t) := E[Y(t+\delta) \mid \mathscr{F}_t]\ ;\ t \in [0,T]. \tag{3}$$

Here $\mathscr{R}$ is the set of functions from $\mathbb{R}_0 := \mathbb{R}\backslash\{0\}$ into $\mathbb{R}$, $h(x,\omega)$ is a C^1 function (with respect to x) from $\mathbb{R} \times \Omega$ into $\mathbb{R}$ such that $h(x,\cdot)$ is $\mathscr{F}_T$-measurable for all x, and

$$g : [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathscr{R} \times \mathbb{U} \times \Omega \to \mathbb{R}$$

is a given function (driver) such that $g(t, x, y, a, z, k, u, \cdot)$ is an $\mathbb{F}$-adapted process for all $x, y, a, z \in \mathbb{R}$, $k \in \mathscr{R}$ and $u \in \mathbb{U}$, which is the set of admissible control values. The process $u(t)$ is our control process, assumed to be in a given family $\mathscr{A} = \mathscr{A}_{\mathbb{G}}$ of *admissible* processes, assumed to be càdlàg and adapted to a given subfiltration $\mathbb{G} = \{\mathscr{G}_t\}_{t\geq 0}$ of the filtration $\mathbb{F}$, i.e. $\mathscr{G}_t \subseteq \mathscr{F}_t$ for all t. The sigma-algebra $\mathscr{G}_t$ represents the information available to the controller at time t.

We assume that for all $u \in \mathscr{A}$ the coupled system (1)–(3) has a unique solution $X(t) = X^u(t) \in L^2(m \times P), Y(t) = Y^u(t) \in L^2(m \times P), A(t) = A^u(t) \in L^2(m \times P), Z(t) = Z^u(t) \in L^2(m \times P), K(t,\zeta) = K^u(t,\zeta) \in L^2(m \times \nu \times P)$, with $X(t), Y(t), A(t)$ being càdlàg and $Z(t), K(t,\zeta)$ being predictable. Here and later m denotes Lebesgue measure on $[0,T]$.

To the best of our knowledge this system, (1)–(3), of *predictive mean-field FBSDEs* has not been studied before. However, the predictive BSDE (1)–(3) is related to the time-advanced BSDE which appears as an adjoint equation for stochastic control problems of a stochastic differential *delay* equation. See [7] and the references therein.

The process $A(t)$ models the predicted future value of the state Y at time $t+\delta$. Therefore (1)–(3) represent a system where the dynamics of the state is influenced by beliefs about the future. This is a natural model for situations where human behavior is involved, for example in pricing issues in financial or energy markets.

The performance functional associated to $u \in \mathscr{A}$ is defined by

$$J(u) = E\left[\int_0^T f(t, X(t), Y(t), A(t), u(t), \omega)dt + \varphi(X(T), \omega) + \psi(Y(0))\right] \tag{4}$$

where $f : [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{U} \times \Omega \to \mathbb{R}$, $\varphi : \mathbb{R} \times \Omega \to \mathbb{R}$ and $\psi : \mathbb{R} \to \mathbb{R}$ are given C^1 functions, with $f(t, x, y, a, u, \cdot)$ being $\mathbb{F}$-adapted for all $x, y, a \in \mathbb{R}$, $u \in \mathbb{U}$. We assume that $\varphi(x, \cdot)$ is $\mathscr{F}_T$-measurable for all x.

We study the following predictive mean-field stochastic control problem:
Find $u^* \in \mathscr{A}$ such that

$$\sup_{u \in \mathscr{A}} J(u) = J(u^*). \tag{5}$$

In Sect. 3 we give a sufficient and a necessary maximum principle for the optimal control of forward-backward predictive mean-field systems of the type above.

An existence and uniqueness result for predictive mean-field BSDEs is given in Sect. 4.
Then in Sect. 5 we apply the results to the following problems:

- Portfolio optimization in a market where the stock price is modeled by a predictive mean-field BSDE,
- Optimization of consumption with respect to *predictive recursive utility*.

3 Solution Methods for the Stochastic Control Problem

3.1 A Sufficient Maximum Principle

For notational simplicity we suppress the dependence of ω in f, g, h, φ and ψ in the sequel. We first give sufficient conditions for optimality of the control u by modifying the stochastic maximum principle given in, for example, [6], to our new situation:

We define the *Hamiltonian* $H : [0, T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathscr{R} \times \mathbb{U} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}) \to \mathbb{R}$ associated to the problem (5) by

$$\begin{aligned} H(t, x, y, a, z, k, u, p, q, r, \lambda) = {} & f(t, x, y, a, u) + b(t, x, y, a, z, k, u)p + \sigma(t, x, y, a, z, k, u)q \\ & + \int_{\mathbb{R}} \gamma(t, x, y, a, z, k, u, \zeta)\tilde{N}(dt, d\zeta) + g(t, x, y, a, z, k, u)\lambda. \end{aligned} \tag{6}$$

We assume that f, b, σ, γ and g, and hence H, are Fréchet differentiable (C^1) in the variables x, y, a, z, k, u and that the Fréchet derivative $\nabla_k H$ of H with respect to $k \in \mathscr{R}$ as a random measure is absolutely continuous with respect to ν, with

Radon-Nikodym derivative $\frac{d\nabla_k H}{d\nu}$. Thus, if $\langle \nabla_k H, h \rangle$ denotes the action of the linear operator $\nabla_k H$ on the function $h \in \mathscr{R}$ we have

$$\langle \nabla_k H, h \rangle = \int_{\mathbb{R}} h(\zeta) d\nabla_k H(\zeta) = \int_{\mathbb{R}} h(\zeta) \frac{d\nabla_k H(\zeta)}{d\nu(\zeta)} d\nu(\zeta). \tag{7}$$

The associated backward-forward system of equations in the *adjoint processes* $p(t), q(t), r(t), \lambda(t)$ is defined by

- BSDE in $p(t), q(t), r(t)$:

$$\begin{cases} dp(t) &= -\frac{\partial H}{\partial x}(t)dt + q(t)dB(t) + \int_{\mathbb{R}} r(t,\zeta)\tilde{N}(dt,d\zeta)\ ;\ 0 \le t \le T \\ p(T) &= \varphi'(X(T)) + \lambda(T)h'(X(T)). \end{cases} \tag{8}$$

- SDE in $\lambda(t)$:

$$\begin{cases} d\lambda(t) &= \left\{ \frac{\partial H}{\partial y}(t) + \frac{\partial H}{\partial a}(t-\delta)\chi_{[\delta,T]}(t) \right\} dt + \frac{\partial H}{\partial z}(t)dB(t) \\ &\quad + \int_{\mathbb{R}} \frac{d\nabla_k H}{d\nu}(t,\zeta)\tilde{N}(dt,d\zeta)\ ;\ 0 \le t \le T \\ \lambda(0) &= \psi'(Y(0)), \end{cases} \tag{9}$$

where we have used the abbreviated notation

$$H(t) = H(t, X(t), Y(t), A(t), Z(t), K(t,\cdot), u(t), p(t), q(t), r(t), \lambda(t)).$$

Note that, in contrast to the *time advanced* BSDE (1)–(3), (9) is a (forward) stochastic differential equation with *delay*.

Theorem 1 (Sufficient maximum principle) *Let $\hat{u} \in \mathscr{A}$ with corresponding solution $\hat{X}(t), \hat{Y}(t), \hat{A}(t), \hat{Z}(t), \hat{K}(t,\cdot), \hat{p}(t), \hat{q}(t), \hat{r}(t), \hat{\lambda}(t)$ of (1)–(3), (8) and (9). Assume the following:*

- $$\hat{\lambda}(T) \ge 0 \tag{10}$$

- *For all t, the functions*

$$x \to h(x), x \to \varphi(x), x \to \psi(x) \text{ and}$$
$$(x,y,a,z,k,u) \to H(t,x,y,a,z,k,u,\hat{p}(t),\hat{q}(t),\hat{r}(t),\hat{\lambda}(t))$$

are concave (11)

- *For all t the following holds,*

(The conditional maximum principle)

$$\underset{v\in\mathbb{U}}{ess\ sup}\ E[H(t,\hat{X}(t),\hat{Y}(t),\hat{A}(t),\hat{Z}(t),\hat{K}(t,\cdot),v,\hat{\lambda}(t),\hat{p}(t),\hat{q}(t),\hat{r}(t,\cdot))\mid\mathscr{G}_t]$$
$$=E[H(t,\hat{X}(t),\hat{Y}(t),\hat{A}(t),\hat{Z}(t),\hat{K}(t,\cdot),\hat{u}(t),\hat{\lambda}(t),\hat{p}(t),\hat{q}(t),\hat{r}(t,\cdot))\mid\mathscr{G}_t]\ ;\ t\in[0,T] \tag{12}$$

- $$\left\| \frac{d\nabla_k \hat{H}(t,.)}{d\nu} \right\| < \infty \textit{ for all } t\in[0,T]. \tag{13}$$

Then $\hat{u}$ is an optimal control for the problem (5).

Proof By replacing the terminal time T by an increasing sequence of stopping times τ_n converging to T as n goes to infinity, and arguing as in [6] we see that we may assume that all the local martingales appearing in the calculations below are martingales.

Much of the proof is similar to the proof of Theorem 3.1 in [6], but due to the predictive mean-field feature of the BSDE (1)–(3), there are also essential differences. Therefore, for the convenience of the reader, we sketch the whole proof:

Choose $u\in\mathscr{A}$ and consider

$$J(u)-J(\hat{u})=I_1+I_2+I_3, \tag{14}$$

with

$$I_1:=E\left[\int_0^T\{f(t)-\hat{f}(t)\}\right]dt,\quad I_2:=E[\varphi(X(T))-\varphi(\hat{X}(T))],\quad I_3:=\psi(Y(0))-\psi(\hat{Y}(0)), \tag{15}$$

where $\hat{f}(t)=f(t,\hat{Y}(t),\hat{A}(t),\hat{u}(t))$ etc., and $\hat{Y}(t)=Y^{\hat{u}}(t)$ is the solution of (1)–(3) when $u=\hat{u}$, and $\hat{A}(t)=E[\hat{Y}(t)\mid\mathscr{F}_t]$.

By the definition of H we have

$$\begin{aligned}I_1=E\Bigg[\int_0^T&\{H(t)-\hat{H}(t)-\hat{p}(t)\tilde{b}(t)-\hat{q}(t)\tilde{\sigma}(t)\\&-\int_{\mathbb{R}}\hat{r}(t,\zeta)\tilde{\gamma}(t,\zeta)\nu(d\zeta)-\hat{\lambda}(t)\tilde{g}(t)\Bigg],\end{aligned} \tag{16}$$

where we from now on use the abbreviated notation

$$H(t)=H(t,X(t),Y(t),A(t),Z(t),K(t,\cdot),u(t),\hat{\lambda}(t))$$
$$\hat{H}(t)=H(t,\hat{X}(t),\hat{Y}(t),\hat{A}(t),\hat{Z}(t),\hat{K}(t,\cdot),\hat{u}(t),\hat{\lambda}(t))$$

and we put

$$\tilde{b}(t) := b(t) - \hat{b}(t),$$

and similarly with $\tilde{X}(t) := X(t) - \hat{X}(t)$, $\tilde{Y}(t) := Y(t) - \hat{Y}(t)$, $\tilde{A}(t) := A(t) - \hat{A}(t)$, etc.

By concavity of φ, (9) and the Itô formula,

$$\begin{aligned}
I_2 &\leq E[\varphi'(\hat{X}(T))\tilde{X}(T)] \\
&= E[\hat{p}(T)\tilde{X}(T)] - E[\hat{\lambda}(T)h'(\hat{X}(T))\tilde{X}(T)] \\
&= \Bigg(E\Bigg[\int_0^T \hat{p}(t^-)d\tilde{X}(t) + \int_0^T \tilde{X}(t^-)d\hat{p}(t) + \int_0^T \hat{q}(t)\tilde{\sigma}(t)dt \\
&\qquad + \int_0^T \int_{\mathbb{R}} \hat{r}(t,\zeta)\tilde{\gamma}(t,\zeta)\nu(d\zeta)dt\Bigg]\Bigg) - E[\hat{\lambda}(T)h'(\hat{X}(T))\tilde{X}(T)] \\
&= E\Bigg[\int_0^T \hat{p}(t)\tilde{b}(t)dt + \int_0^T \tilde{X}(t)\Bigg(-\frac{\partial \hat{H}}{\partial x}(t)\Bigg)dt \\
&\qquad + \int_0^T \hat{q}(t)\tilde{\sigma}(t)dt + \int_0^T \int_{\mathbb{R}} \hat{r}(t,\zeta)\tilde{\gamma}(t,\zeta)\nu(d\zeta)dt\Bigg] \\
&\quad - E[\hat{\lambda}(T)h'(\hat{X}(T))\tilde{X}(T)].
\end{aligned} \tag{17}$$

By concavity of ψ and h, (10) and the Itô formula we have

$$\begin{aligned}
I_3 &\leq E\left[\psi'(Y(0))\tilde{Y}(0)\right] = E[\hat{\lambda}(0)\tilde{Y}(0)] \\
&= E[\hat{\lambda}(T)\tilde{Y}(T)] - E\left[\int_0^T \hat{\lambda}(t)d\tilde{Y}(t) + \int_0^T \tilde{Y}(t)d\hat{\lambda}(t) + \int_0^T d[\tilde{Y},\hat{\lambda}](t)\right] \\
&= E[\hat{\lambda}(T)(h(X(T)) - h(\hat{X}(T)))] \\
&\quad - E\left[\int_0^T \hat{\lambda}(t)d\tilde{Y}(t) + \int_0^T \tilde{Y}(t)d\hat{\lambda}(t) + \int_0^T d[\tilde{Y},\hat{\lambda}](t)\right] \\
&\leq E[\hat{\lambda}(T)h'(\hat{X}(T))\tilde{X}(T)] \\
&\quad + E\Bigg[\int_0^T \hat{\lambda}(t)\tilde{g}(t)dt + \int_0^T \tilde{Y}(t)\left[-\frac{\partial \hat{H}}{\partial y}(t) - \frac{\partial \hat{H}}{\partial a}(t-\delta)\chi_{[\delta,T]}(t)\right]dt \\
&\qquad + \int_0^T \frac{\partial \hat{H}}{\partial z}(t)\tilde{Z}(t)dt + \int_0^T \int_{\mathbb{R}} \frac{d\nabla_k \hat{H}}{d\nu}(t,\zeta)\tilde{K}(t,\zeta)\nu(d\zeta)dt\Bigg]
\end{aligned} \tag{18}$$

Adding (16), (17) and (18) we get, by (9),

$$
\begin{aligned}
J(u) - J(\hat{u}) &= I_1 + I_2 + I_3 \\
&\leq E\Bigg[\int_0^T \Bigg\{ H(t) - \hat{H}(t) - \frac{\partial \hat{H}}{\partial x}\tilde{X}(t) - \frac{\partial \hat{H}}{\partial y}\tilde{Y}(t) \\
&\qquad - \frac{\partial H}{\partial a}(t-\delta)\chi_{[\delta,T]}(t)\tilde{Y}(t) - \frac{\partial H}{\partial z}(t)\tilde{Z}(t) - \langle \nabla_k \hat{H}(t,\cdot), \tilde{K}(t,\cdot)\rangle \Bigg\} dt\Bigg].
\end{aligned}
\tag{19}
$$

Note that, since $Y(s) = \hat{Y}(s) = L$ for $s \in (T, T+\delta]$ by (1), we get

$$
\begin{aligned}
&E\left[\int_0^T \frac{\partial \hat{H}}{\partial a}(t-\delta)\tilde{Y}(t)\chi_{[\delta,T]}(t)dt\right] = E\left[\int_0^{T-\delta} \frac{\partial \hat{H}}{\partial a}(s)\tilde{Y}(s+\delta)ds\right] \\
&= E\left[\int_0^{T-\delta} E\left[\frac{\partial \hat{H}}{\partial a}(s)\tilde{Y}(s+\delta) \mid \mathscr{F}_s\right] dt\right] \\
&= E\left[\int_0^{T-\delta} \frac{\partial \hat{H}}{\partial a}(s) E\left[\tilde{Y}(s+\delta) \mid \mathscr{F}_s\right] ds\right] = E\left[\int_0^T \frac{\partial \hat{H}}{\partial a}(s)\tilde{A}(s)ds\right].
\end{aligned}
\tag{20}
$$

Substituted into (19) this gives, by concavity of H,

$$
\begin{aligned}
J(u) - J(\hat{u}) &= I_1 + I_2 + I_3 \\
&\leq E\Bigg[\int_0^T \Bigg\{ H(t) - \hat{H}(t) - \frac{\partial \hat{H}}{\partial x}(X(t) - \hat{X}(t)) - \frac{\partial \hat{H}}{\partial y}(Y(t) - \hat{Y}(t)) \\
&\qquad - \frac{\partial H}{\partial a}(t)(A(t) - \hat{A}(t)) - \frac{\partial H}{\partial z}(t)(Z(t) - \hat{Z}(t)) \\
&\qquad - \langle \nabla_k \hat{H}(t,\cdot), (K(t,\cdot) - \hat{K}(t,\cdot)\rangle \Bigg\} dt\Bigg] \\
&\leq E\left[\int_0^T \frac{\partial \hat{H}}{\partial u}(t)(u(t) - \hat{u}(t))dt\right] \\
&= E\left[\int_0^T E[\frac{\partial \hat{H}}{\partial u}(t)|\mathscr{G}_t](u(t) - \hat{u}(t))dt\right] \leq 0,
\end{aligned}
\tag{21}
$$

since $u = \hat{u}(t)$ maximizes $E[\hat{H}(t)|\mathscr{G}_t]$. □

3.2 A Necessary Maximum Principle

We proceed to prove a partial converse of Theorem 1, in the sense that we give *necessary* conditions for a control $\hat{u}$ to be optimal. In this case we can only conclude that $\hat{u}(t)$ is a critical point for the Hamiltonian, not necessarily a maximum point. On the other hand, we do not need any concavity assumptions, but instead we need some properties of the set $\mathscr{A}$ of admissible controls, as described below.

Theorem 2 (Necessary maximum principle) *Suppose $\hat{u} \in \mathscr{A}$ with associated solutions $\hat{X}, \hat{Y}, \hat{Z}, \hat{K}, \hat{p}, \hat{q}, \hat{r}, \hat{\lambda}$ of (1)–(3) and (8) and (9). Suppose that for all processes $\beta(t)$ of the form*

$$\beta(t) := \chi_{[t_0,T]}(t)\alpha, \tag{22}$$

where $t_0 \in [0, T)$ and $\alpha = \alpha(\omega)$ is a bounded $\mathscr{G}_{t_0}$-measurable random variable, there exists $\delta > 0$ such that the process

$$\hat{u}(t) + r\beta(t) \in \mathscr{A} \text{ for all } r \in [-\delta, \delta].$$

We assume that the derivative processes defined by

$$x(t) = x^{\beta}(t) = \frac{d}{dr}X^{\hat{u}+r\beta}(t) \mid_{r=0}, \tag{23}$$

$$y(t) = y^{\beta}(t) = \frac{d}{dr}Y^{\hat{u}+r\beta}(t) \mid_{r=0}, \tag{24}$$

$$a(t) = a^{\beta}(t) = \frac{d}{dr}A^{\hat{u}+r\beta}(t) \mid_{r=0}, \tag{25}$$

$$z(t) = z^{\beta}(t) = \frac{d}{dr}Z^{\hat{u}+r\beta}(t) \mid_{r=0}, \tag{26}$$

$$k(t) = k^{\beta}(t) = \frac{d}{dr}K^{\hat{u}+r\beta}(t) \mid_{r=0}, \tag{27}$$

exist and belong to $L^2(m \times P)$, $L^2(m \times P)$, $L^2(m \times P)$, and $L^2(m \times P \times \nu)$, respectively.

Moreover, we assume that $x(t)$ satisfies the equation

$$
\begin{cases}
dx(t) = \Big\{ \frac{\partial b}{\partial x}(t)x(t) + \frac{\partial b}{\partial y}(t)y(t) + \frac{\partial b}{\partial a}(t)a(t) + \frac{\partial b}{\partial z}(t)z(t) + \langle \nabla_k b, k(t,\cdot)\rangle \\
\qquad + \frac{\partial b}{\partial u}(t)\beta(t) \Big\} dt \\
\qquad + \Big\{ \frac{\partial \sigma}{\partial x}(t)x(t) + \frac{\partial \sigma}{\partial y}(t)y(t) + \frac{\partial \sigma}{\partial a}(t)a(t) + \frac{\partial \sigma}{\partial z}(t)z(t) + \langle \nabla_k \sigma, k(t,\cdot)\rangle \\
\qquad + \frac{\partial \sigma}{\partial u}(t)\beta(t) \Big\} dB(t) \\
\qquad + \int_{\mathbb{R}} \Big\{ \frac{\partial \gamma}{\partial x}(t,\zeta)x(t) + \frac{\partial \gamma}{\partial y}(t,\zeta)y(t) + \frac{\partial \gamma}{\partial a}(t,\zeta)a(t) + \frac{\partial \gamma}{\partial z}(t,\zeta)z(t) \\
\qquad\qquad + \langle \nabla_k \gamma(t,\zeta), k(t,\cdot)\rangle + \frac{\partial \gamma}{\partial u}(t,\zeta)\beta(t) \Big\} \tilde{N}(dt,d\zeta)\ ;\ t \in [0,T] \\
x(0) = 0
\end{cases}
\tag{28}
$$

and that $y(t)$ satisfies the equation

$$
\begin{cases}
dy(t) &= -\Big\{ \frac{\partial g}{\partial x}(t)x(t) + \frac{\partial g}{\partial y}(t)y(t) + \frac{\partial g}{\partial a}(t)a(t) + \frac{\partial g}{\partial z}(t)z(t) \\
&\qquad + \langle \nabla_k g(t), k(t,\cdot)\rangle + \frac{\partial g}{\partial u}(t)\beta(t) \Big\} dt \\
&\qquad + z(t)dB(t) + \int_{\mathbb{R}} k(t,\zeta)\tilde{N}(dt,d\zeta)\ ;\ 0 \le t < T \\
y(T) &= h'(X(T))x(T) \\
y(t) &= 0\ ;\ T < t \le T+\delta,
\end{cases}
\tag{29}
$$

where we have used the abbreviated notation

$$
\frac{\partial g}{\partial x}(t) = \frac{\partial}{\partial x} g(t,x,y,a,z,k,u)_{x=X(t),y=Y(t),a=A(t),z=Z(t),k=K(t),u=u(t)} \text{ etc.}
$$

Then the following, (i) and (ii), are equivalent:

(i) $\frac{d}{dr}J(\hat{u}+r\beta)_{r=0} = 0$ *for all β of the form (22)*

(ii) $\frac{d}{du}E[H(t,\hat{Y}(t),\hat{A}(t),\hat{Z}(t),\hat{K}(t),u,\hat{\lambda}(t))_{u=\hat{u}(t)}|\mathcal{G}_t] = 0,$
where $(\hat{Y},\hat{A},\hat{Z},\hat{K},\hat{\lambda})$ is the solution of (1), (3) and (9) corresponding to $u=\hat{u}$.

Proof As in Theorem 1, by replacing the terminal time T by an increasing sequence of stopping times τ_n converging to T as n goes to infinity, we obtain as in [6] that we may assume that all the local martingales appearing in the calculations below are martingales. The proof has many similarities with the proof of Theorem 3.2 in [6], but since there are some essential differences due to the predictive mean-field term, we sketch the whole proof. For simplicity of notation we drop the hats in the sequel, i.e. we write u instead of $\hat{u}$ etc.

(i) $\Rightarrow$ (ii): We can write $\frac{d}{dr}J(u+r\beta)\mid_{r=0}= I_1 + I_2 + I_3$, where

$$I_1 = \frac{d}{dr}E\left[\int_0^T f(t, Y^{u+r\beta}(t), A^{u+r\beta}(t), Z^{u+r\beta}(t), K^{u+r\beta}(t), u(t)+r\beta(t))dt\right]_{r=0}$$
$$I_2 = \frac{d}{dr}[\varphi(X^{u+r\beta}(T))]_{r=0}$$
$$I_3 = \frac{d}{dr}[\psi(Y^{u+r\beta}(0))]_{r=0}.$$

By our assumptions on f and ψ we have

$$I_1 = \left[\int_0^T \left\{\frac{\partial f}{\partial x}(t)x(t) + \frac{\partial f}{\partial y}(t)y(t) + \frac{\partial f}{\partial a}(t)a(t) + \frac{\partial f}{\partial z}(t)z(t) + \langle \nabla_k f(t,\cdot), k(t,\cdot)\rangle + \frac{\partial f}{\partial u}(t)\beta(t)\right\}dt\right] \tag{30}$$

$$I_2 = E[\varphi'(X(T)x(T)] = E[p(T)x(T)] \tag{31}$$
$$I_3 = \psi'(Y(0))y(0) = \lambda(0)y(0). \tag{32}$$

By the Itô formula and (28)

$$\begin{aligned}
I_2 &= E[p(T)x(T)] = E\left[\int_0^T p(t)dx(t) + \int_0^T x(t)dp(t) + \int_0^T d[p,x](t)\right]\\
&= E\Bigg[\int_0^T p(t)\left\{\frac{\partial b}{\partial x}(t)x(t) + \frac{\partial b}{\partial y}(t)y(t) + \frac{\partial b}{\partial a}(t)a(t) + \frac{\partial b}{\partial z}(t)z(t)\right.\\
&\quad \left. + \langle \nabla_k b(t), k(t,\cdot)\rangle + \frac{\partial b}{\partial u}(t)\beta(t)\right\}dt + \int_0^{\tau_n} x(t)\left(-\frac{\partial H}{\partial x}(t)\right)dt\\
&\quad + \int_0^{\tau_n} q(t)\left\{\frac{\partial \sigma}{\partial x}(t)x(t) + \frac{\partial \sigma}{\partial y}(t)y(t) + \frac{\partial \sigma}{\partial a}(t)a(t) + \frac{\partial \sigma}{\partial z}(t)z(t)\right.\\
&\quad \left. + \langle \nabla_k \sigma(t), k(t,\cdot)\rangle + \frac{\partial \sigma}{\partial u}(t)\beta(t)\right\}dt\\
&\quad + \int_0^T \int_{\mathbb{R}} r(t,\zeta)\left\{\frac{\partial \gamma}{\partial x}(t,\zeta)x(t) + \frac{\partial \gamma}{\partial y}(t,\zeta)y(t) + \frac{\partial \gamma}{\partial a}(t,\zeta)a(t) + \frac{\partial \gamma}{\partial z}(t,\zeta)z(t)\right.\\
&\quad \left. + < \nabla_k \gamma(t,\zeta), k(t,\cdot) > + \frac{\partial \gamma}{\partial u}(t,\zeta)\beta(t)\right\}\nu(d\zeta)dt\Bigg]\\
&= E\Bigg[\int_0^T x(t)\left\{\frac{\partial b}{\partial x}(t)p(t) + \frac{\partial \sigma}{\partial x}(t)q(t) + \int_{\mathbb{R}}\frac{\partial \gamma}{\partial x}(t,\zeta)r(t,\zeta)\nu(d\zeta) - \frac{\partial H}{\partial x}(t)\right\}dt\\
&\quad + \int_0^T y(t)\left\{\frac{\partial b}{\partial y}(t)p(t) + \frac{\partial \sigma}{\partial y}(t)q(t) + \int_{\mathbb{R}}\frac{\partial \gamma}{\partial y}(t,\zeta)r(t,\zeta)\nu(d\zeta)\right\}dt\\
&\quad + \int_0^T a(t)\left\{\frac{\partial b}{\partial a}(t)p(t) + \frac{\partial \sigma}{\partial a}(t)q(t) + \int_{\mathbb{R}}\frac{\partial \gamma}{\partial a}(t,\zeta)r(t,\zeta)\nu(d\zeta)\right\}dt
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^T z(t)\left\{\frac{\partial b}{\partial z}(t)p(t)+\frac{\partial \sigma}{\partial z}(t)q(t)+\int_{\mathbb{R}}\frac{\partial \gamma}{\partial z}(t,\zeta)r(t,\zeta)\nu(d\zeta)\right\}dt\\
&+\int_0^T\int_{\mathbb{R}}\langle k(t,\cdot),\nabla_k b(t)p(t)+\nabla_k\sigma(t)q(t)\\
&+\int_{\mathbb{R}}\nabla_k\gamma(t,\zeta)r(t,\zeta)\nu(d\zeta)\rangle\nu(d\zeta)dt\Bigg]\\
=&\,E\left[\int_0^T x(t)\left\{-\frac{\partial f}{\partial x}(t)-\lambda(t)\frac{\partial g}{\partial x}(t)\right\}dt\right.\\
&+\int_0^T y(t)\left\{\frac{\partial H}{\partial y}(t)-\frac{\partial f}{\partial y}(t)-\lambda(t)\frac{\partial g}{\partial y}(t)\right\}dt\\
&+\int_0^T a(t)\left\{\frac{\partial H}{\partial a}(t)-\frac{\partial f}{\partial a}(t)-\lambda(t)\frac{\partial g}{\partial a}(t)\right\}dt\\
&+\int_0^T z(t)\left\{\frac{\partial H}{\partial z}(t)-\frac{\partial f}{\partial z}(t)-\lambda(t)\frac{\partial g}{\partial z}(t)\right\}dt\\
&+\int_0^T\int_{\mathbb{R}}k(t,\zeta)\{\nabla_k H(t)-\nabla_k f(t)-\lambda(t)\nabla_k g(t)\}\nu(d\zeta)dt\\
&\left.+\int_0^T\beta(t)\left\{\frac{\partial H}{\partial u}(t)-\frac{\partial f}{\partial u}(t)-\lambda(t)\frac{\partial g}{\partial u}(t)\right\}dt\right]\\
=&-I_1-E\left[\int_0^T\lambda(t)\left\{\frac{\partial g}{\partial x}(t)x(t)+\frac{\partial g}{\partial y}(t)y(t)+\frac{\partial g}{\partial z}(t)z(t)\right.\right.\\
&\left.\left.+\langle\nabla_k g(t),k(t,\cdot)\rangle+\frac{\partial g}{\partial u}(t)\beta(t)\right\}dt\right]\\
&+E\left[\int_0^T\left\{\frac{\partial H}{\partial y}(t)y(t)+\frac{\partial H}{\partial z}(t)z(t)+\langle\nabla_k H(t),k(t,\cdot)\rangle+\frac{\partial H}{\partial u}(t)\beta(t)\right\}dt\right] \qquad (33)
\end{aligned}
$$

By the Itô formula and (29),

$$
\begin{aligned}
I_3=\lambda(0)y(0)&=E\left[\lambda(T)y(T)-\left(\int_0^T\lambda(t)dy(t)+\int_0^T y(t)d\lambda(t)+\int_0^T d[\lambda,y](t)\right)\right]\\
&=E[\lambda(T)y(T)]\\
&\quad-\left(E\left[\int_0^T\lambda(t)\left\{-\frac{\partial g}{\partial y}(t)y(t)-\frac{\partial g}{\partial a}(t)a(t)-\frac{\partial g}{\partial z}(t)z(t)\right.\right.\right.\\
&\qquad\qquad\left.-\langle\nabla_k g(t),k(t,\cdot)\rangle-\frac{\partial g}{\partial u}(t)\beta(t)\right\}dt\\
&\qquad+\int_0^T y(t)\frac{\partial H}{\partial y}(t)dt+y(t)\frac{\partial H}{\partial a}(t-\delta)\chi_{[\delta,T]}(t)dt+\int_0^T z(t)\frac{\partial H}{\partial z}(t)dt\\
&\qquad\left.\left.+\int_0^T\int_{\mathbb{R}}k(t,\zeta)\nabla_k H(t,\zeta)\nu(d\zeta)dt\right]\right). \qquad (34)
\end{aligned}
$$

Adding (30), (33) and (34) and using that

$$E[\int_0^T y(t)\frac{\partial H}{\partial a}(t-\delta)\chi_{[\delta,T]}dt] = E[\int_0^{T-\delta} y(s+\delta)\frac{\partial H}{\partial a}(s)ds]$$
$$= E[\int_0^T \frac{\partial H}{\partial a}(s)E[y(s+\delta)|\mathscr{F}_s]ds] = E[\int_0^T y(t)\frac{\partial H}{\partial a}(s)a(s)ds], \tag{35}$$

we get

$$\frac{d}{dr}J(u+r\beta)\mid_{r=0} = I_1 + I_2 = E\left[\int_0^T \frac{\partial H}{\partial u}(t)\beta(t)dt\right].$$

We conclude that

$$\frac{d}{dr}J(\hat{u}+r\beta)\mid_{r=0} = 0$$

if and only if

$$E\left[\int_0^T \frac{\partial \hat{H}}{\partial u}(t)\beta(t)dt\right] = 0 \quad \text{for all bounded } \beta \in \mathscr{A}_{\mathbb{G}} \text{ of the form (12.3.17).}$$

Since this holds for all such β, we obtain that if (i) holds, then

$$\int_{t_0}^T E\left[\frac{\partial \hat{H}}{\partial u}(t) \mid \mathscr{G}_{t_0}\right]dt = 0 \text{ for all } t_0 \in [0,T). \tag{36}$$

Differentiating with respect to t_0 and using continuity of $\frac{\partial \hat{H}}{\partial u}(t)$, we conclude that (ii) holds.

(ii) $\Rightarrow$ (i): This is proved by reversing the above argument. We omit the details. $\square$

4 Existence and Uniqueness of Predictive Mean-Field Equations

In this section we study the existence and uniqueness of predictive mean-field BSDEs in the unknowns $Y(t)$, $Z(t)$, $K(t,\zeta)$ of the form

$$\begin{cases} dY(t) &= -g(t,Y(t),A(t),Z(t),K(t,\cdot),\omega)dt + Z(t)dB(t) \\ &\quad + \int_{\mathbb{R}} K(t,\zeta)\tilde{N}(dt,d\zeta)\ ;\ t\in[0,T) \\ Y(t) &= L\ ;\ t\in[T,T+\delta]\ ;\ \delta>0 \text{ fixed,} \end{cases} \tag{37}$$

where $L \in L^2(P)$ is a given $\mathcal{F}_T$-measurable random variable, and the process $A(t)$ as before is defined by

$$A(t) = E[Y(t+\delta) \mid \mathcal{F}_t]\ ;\ t \in [0, T]. \tag{38}$$

To this end, we can use the same argument which was used to handle a similar, but different, time-advanced BSDE in [7]. For completeness we give the details:

Theorem 3 *Suppose the following holds*

$$E[\int_0^T g^2(t, 0, 0, 0, 0)dt] < \infty \tag{39}$$

There exists a constant C *such that*

$$|g(t, y_1, a_1, z_1, k_1) - g(t, y_2, a_2, z_2, k_2)| \le C(|y_1 - y_2| + |z_1 - z_2| + \left(\int_{\mathbb{R}} |k_1(\zeta) - k_2(\zeta)|^2 \nu(d\zeta)\right)^{\frac{1}{2}}) \tag{40}$$

for all $t \in [0, T]$, *a.s. Then there exists a unique solution triple* $(Y(t), Z(t), K(t, \zeta))$ *of (37) such that the following holds:*

$$\begin{cases} Y \text{ is cadlag and } E[\sup_{t\in[0,T]} Y^2(t)] < \infty, \\ Z, K \text{ are predictable and } E[\int_0^T \{Z^2(t) + \int_{\mathbb{R}} K^2(t,\zeta)\nu(d\zeta)\}dt] < \infty. \end{cases}$$

Proof We argue backwards, starting with the interval $[T - \delta, T]$:
Step 1. In this interval we have $A(t) = E[L|\mathcal{F}_t]$ and hence we know from the theory of classical BSDEs (see e.g. [8, 9] and the references therein), that there exists a unique solution triple $(Y(t), Z(t), K(t, \zeta))$ such that the following holds:

$$\begin{cases} Y \text{ is cadlag and } E[\sup_{t\in[T-\delta,T]} Y^2(t)] < \infty, \\ Z, K \text{ are predictable and } E[\int_{T-\delta}^T \{Z^2(t) + \int_{\mathbb{R}} K^2(t,\zeta)\nu(d\zeta)\}dt] < \infty. \end{cases}$$

Step 2. Next, we continue with the interval $[T - 2\delta, T - \delta]$. For t in this interval, the value of $Y(t+\delta)$ is known from the previous step and hence $A(t) = E[Y(t+\delta)|\mathcal{F}_t]$ is known. Moreover, by Step1 the terminal value for this interval, $Y(T-\delta)$, is known and in $L^2(P)$. Hence we can again refer to the theory of classical BSDEs and get a unique solution in this interval.
Step n. We continue this iteration until we have reached the interval $[0, T - n\delta]$, where n is a natural number such that

$$T - (n + 1)\delta \le 0 < T - n\delta.$$

Combining the solutions from each of the subintervals, we get a solution for the whole interval. □

5 Applications

In this section we illustrate the results of the previous sections by looking at two examples.

5.1 Optimal Portfolio in an Insider Influenced Market

In the seminal papers by Kyle [4] and Back [2] it is proved that in a financial market consisting of

- *noise traders* (where noise is modeled by Brownian motion),
- *an insider* who knows the value L of the price of the risky asset at the terminal time $t = T$ and
- *a market maker* who at any time t clears the market and sets the market price,

the corresponding equilibrium price process (resulting from the insider's portfolio which maximizes her expected profit), will be a *Brownian bridge terminating at the value L at time* $t = T$. In view of this we see that a predictive mean-field equation can be a natural model of the *risky asset price in an insider influenced market.*

Accordingly, suppose we have a market with the following two investment possibilities:

- A risk free asset, with unit price $S_0(t) := 1$ for all t
- A risky asset with unit price $S(t) := Y(t)$ at time t, given by the predictive mean-field equation

$$\begin{cases} dY(t) &= -A(t)\mu(t)dt + Z(t)dB(t);\ \ t \in [0, T) \\ Y(t) &= L(\omega);\quad t \in [T, T + \delta], \end{cases} \tag{41}$$

where $\mu(t) = \mu(t, \omega)$ is a given bounded adapted process and L is a given bounded $\mathscr{F}_T$-measurable random variable, being the terminal state of the process Y at time T.

Let $u(t)$ be a portfolio, representing the number of risky assets held at time t. We assume that $\mathbb{G} = \mathbb{F}$. If we assume that the portfolio is self-financing, the corresponding wealth process $X(t) = X^u(t)$ is given by

$$\begin{cases} dX(t) = u(t)dY(t) = u(t)A(t)\mu(t)dt + u(t)Z(t)dB(t);\ \ t \in [0, T) \\ X(0) = x > 0. \end{cases} \tag{42}$$

Let $U : [0, \infty) \mapsto [-\infty, \infty)$ be a given utility function, assumed to be increasing, concave and C^1 on $(0, \infty)$. We study the following portfolio optimization problem:

Problem 1 Find $u^* \in \mathscr{A}$ such that

$$\sup_{u \in \mathscr{A}} E[U(X^u(T))] = E[U(X^{u^*}(T))]. \tag{43}$$

This is a problem of the type discussed in the previous sections, with $f = \psi = N = 0, \varphi = U$ and $h(x, \omega) = L(\omega)$, and we can apply the maximum principles from Sect. 3 to study it.

By (6) the Hamiltonian gets the form

$$H(t, x, y, a, z, k, u, p, q, r, \lambda) = ua\mu(t)p + uzq + a\mu(t)\lambda. \tag{44}$$

The associated backward-forward system of equations in the adjoint processes $p(t), q(t), \lambda(t)$ becomes

- BSDE in $p(t), q(t)$:

$$\begin{cases} dp(t) & = q(t)dB(t)\ ;\ 0 \le t \le T \\ p(T) & = U'(X(T)), \end{cases} \tag{45}$$

- SDE in $\lambda(t)$:

$$\begin{cases} d\lambda(t) = \mu(t-\delta)[u(t-\delta)p(t-\delta) + \lambda(t-\delta)]\chi_{[\delta,T]}(t)dt \\ \qquad\qquad + u(t)q(t)dB(t)\ ;\ 0 \le t \le T \\ \lambda(0) = 0. \end{cases} \tag{46}$$

The Hamiltonian can only have a maximum with respect u if

$$A(t)\mu(t)p(t) + Z(t)q(t) = 0. \tag{47}$$

Substituting this into (45) we get

$$\begin{cases} dp(t) & = -\theta(t)p(t)dB(t); 0 \le t \le T \\ p(T) & = U'(X(T)), \end{cases} \tag{48}$$

where

$$\theta(t) := \frac{A(t)\mu(t)}{Z(t)}. \tag{49}$$

From this we get

$$p(t) = c \exp(-\int_0^t \theta(s)dB(s) - \frac{1}{2}\int_0^t (\theta(s))^2 ds); 0 \le t \le T \tag{50}$$

where the constant

$$c = p(0) = E[U'(X(T)] \tag{51}$$

remains to be determined.

In particular, putting $t = T$ in (50) we get

$$U'(X(T)) = p(T) = c\exp(-\int_0^T \theta(s)dB(s) - \frac{1}{2}\int_0^T (\theta(s))^2 ds) \tag{52}$$

or

$$X(T) = (U')^{-1}(c\exp(-\int_0^T \theta(s)dB(s) - \frac{1}{2}\int_0^T (\theta(s))^2 ds)). \tag{53}$$

Define

$$\Gamma(T) = \exp(\int_0^T \theta(s)dB(s) - \frac{1}{2}\int_0^T (\theta(s))^2 ds). \tag{54}$$

Then by the Girsanov theorem the measure Q defined on $\mathscr{F}_T$ by

$$dQ(\omega) = \Gamma(T)dP(\omega) \tag{55}$$

is an equivalent martingale measure for the market (41). Therefore, by (53),

$$x = E_Q[X(T)] = E[(U')^{-1}(c\exp(-\int_0^T \theta(s)dB(s) - \frac{1}{2}\int_0^T (\theta(s))^2 ds))\Gamma(T)]. \tag{56}$$

This equation determines implicitly the value of the constant c and hence by (53) the optimal terminal wealth $X(T) = X^{u^*}(T)$. To find the corresponding optimal portfolio u^* we proceed as follows:

Define

$$Z_0(t) := u^*(t)Z(t). \tag{57}$$

Then $(X^{u^*}(t), Z_0(t))$ is found by solving the linear BSDE

$$\begin{cases} dX^{u^*}(t) = \frac{A(t)\mu(t)Z_0(t)}{Z(t)}dt + Z_0(t)dB(t);\ 0 \leq t \leq T \\ X^{u^*}(T) = E[(U')^{-1}(c\exp(-\int_0^T \theta(s)dB(s) - \frac{1}{2}\int_0^T (\theta(s))^2 ds))\Gamma(T)]. \end{cases} \tag{58}$$

We have proved:

Theorem 4 (Optimal portfolio in an insider influenced market) *The optimal portfolio u^* for the problem* (43) *is given by*

$$u^*(t) = \frac{Z_0(t)}{Z(t)}, \tag{59}$$

where $Z_0(t)$, $Z(t)$ are the solutions of the BSDEs (41), (58)*, respectively, and c and θ are given by* (56) *and* (49)*, respectively.*

5.2 Predictive Recursive Utility Maximization

Consider a cash flow $X(t) = X^c(t)$ given by

$$\begin{cases} dX(t) = X(t)[\mu(t)dt + \sigma(t)dB(t) \\ \qquad\qquad + \int_R \gamma(t,\zeta)\tilde{N}(dt,d\zeta)] - c(t)X(t)dt; \;\; t \in [0,T) \\ X(0) = x > 0. \end{cases} \tag{60}$$

Here $\mu(t), \sigma(t), \gamma(t,\zeta)$ are given bounded adapted processes, while $u(t) := c(t)$ is our control, interpreted as our *relative consumption rate* from the cash flow. We say that c is *admissible* if c is $\mathbb{F}$-adapted, $c(t) > 0$ and $X^c(t) > 0$ for all $t \in [0,T)$. We put $\mathbb{G} = \mathbb{F}$.

Let $Y(t) = Y^c(t), Z(t) = Z^c(t), K(t,\zeta) = K^c(t,\zeta)$ be the solution of the predictive mean-field BSDE defined by

$$\begin{cases} dY(t) = -\{\alpha(t)A(t) + \ln(c(t)X(t))\}dt + Z(t)dB(t) \\ \qquad\qquad + \int_R K(t,\zeta)\tilde{N}(dt,d\zeta); \;\; t \in [0,T) \\ Y(T) = 0, \end{cases} \tag{61}$$

where $\alpha(t) > 0$ is a given bounded $\mathbb{F}$-adapted process. Then, inspired by classical definition of recursive utility in [3], we define $Y^c(0)$ to be the *predictive recursive utility* of the relative consumption rate c.

We now study the following predictive recursive utility maximization problem:

Problem 2 Find $c^* \in \mathscr{A}$ such that

$$\sup_{c \in \mathscr{A}} Y^c(0) = Y^{c*}(0). \tag{62}$$

We apply the maximum principle to study this problem. In this case we have $f = \varphi = h = 0$, $\psi(x) = x$, and the Hamiltonian becomes

$$\begin{aligned} H(t,x,y,a,z,k,u,p,q,r,\lambda) = x[(\mu(t) - c)p + \sigma(t)q + \int_{\mathbb{R}} \gamma(t,\zeta)r(\zeta)\nu(dt,d\zeta)] \\ + [a\alpha(t) + \ln c + \ln x]\lambda. \end{aligned} \tag{63}$$

The associated backward-forward system of equations in the adjoint processes $p(t), q(t), \lambda(t)$ becomes

- BSDE in $p(t), q(t)$:

$$\begin{cases} dp(t) &= -[(\mu(t) - c(t))p(t) + \sigma(t)q(t) + \int_{\mathbb{R}} \gamma(t,\zeta)\nu(dt,d\zeta) + \frac{\lambda(t)}{X(t)}]dt \\ & \quad +q(t)dB(t) + \int_{\mathbb{R}} r(t,\zeta)\tilde{N}(dt,d\zeta) \ ; \ 0 \le t \le T \\ p(T) &= 0, \end{cases} \tag{64}$$

- SDE in $\lambda(t)$:

$$\begin{cases} d\lambda(t) = \alpha(t-\delta)\lambda(t-\delta)]\chi_{[\delta,T]}(t)dt \ ; \ 0 \le t \le T \\ \lambda(0) = 1. \end{cases} \tag{65}$$

The delay SDE (65) does not contain any unknown parameters, and it is easily seen that it has a unique continuous solution $\lambda(t) > 1$, which we may consider known.

We can now proceed along the same lines as in Sect. 5.2 of [1]: Maximizing H with respect to c gives the first order condition

$$c(t) = \frac{\lambda(t)}{X(t)p(t)}. \tag{66}$$

The solution of the linear BSDE (64) is given by

$$\Gamma(t)p(t) = E[\int_t^T \frac{\lambda(s)\Gamma(s)}{X(s)} ds | \mathscr{F}_t], \tag{67}$$

where

$$\begin{cases} d\Gamma(t) = \Gamma(t^-)[(\mu(t) - c(t))dt + \sigma(t)dB(t) + \int_{\mathbb{R}} \gamma(t,\zeta)\tilde{N}(dt,d\zeta)] \ ; \ 0 \le t \le T \\ \Gamma(0) = 1. \end{cases} \tag{68}$$

Comparing with (60) we see that

$$X(t) = x\Gamma(t) \ ; \ 0 \le t \le T. \tag{69}$$

Substituting this into (67) we obtain

$$p(t)X(t) = E[\int_t^T \lambda(s)ds | \mathscr{F}_t] \ ; \ 0 \le t \le T. \tag{70}$$

Substituting this into (66) we get the following conclusion:

Theorem 5 *The optimal relative consumption rate $c^*(t)$ for the predictive recursive utility consumption problem (62) is given by*

$$c^*(t) = \frac{\lambda(t)}{E[\int_t^T \lambda(s)ds|\mathscr{F}_t]} \; ; \; 0 \leq t < T, \tag{71}$$

where $\lambda(t)$ is the solution of the delay SDE (65).

Acknowledgments This research was carried out with support of CAS—Centre for Advanced Study, at the Norwegian Academy of Science and Letters, within the research program SEFE.

References

1. Agram, N., Øksendal, B.: Infinite horizon optimal control of forward-backward stochastic differential equations with delay. J. Comput. Appl. Math. **259**, 336–349 (2014)
2. Back, K.: Insider trading in continuous time. Rev. Financ. Stud. **5**(3), 387–409 (1992)
3. Duffie, D., Epstein, L.: Stochastic differential utility. Econometrica **60**, 353–394 (1992)
4. Kyle, A.S.: Continuous auctions and insider trading. Econometrica **53**(6), 1315–1336 (1985)
5. Øksendal, B., Sulem, A.: Applied Stochastic Control of Jump Diffusions, 2nd edn. Springer (2007)
6. Øksendal, B., Sulem, A.: Risk minimization in financial markets modeled by Itô -Lévy processes. Afrika Matematika (2014). doi:10.1007/s13370-014-02489-9
7. Øksendal, B., Sulem, A., Zhang, T.: Optimal control of stochastic delay equations and time-advanced backward stochastic differential equations. Adv. Appl. Probab. **43**, 572–596 (2011)
8. Quenez, M.-C.: Backward stochastic differential equations. In: Encyclopedia of Quantitative Finance, pp. 134–145. Wiley (2010)
9. Quenez, M.-C., Sulem, A.: BSDEs with jumps, optimization and applications to dynamic risk measures. Stoch. Proc. Appl. **123**, 3328–3357 (2013)

Modelling the Impact of Wind Power Production on Electricity Prices by Regime-Switching Lévy Semistationary Processes

Almut E.D. Veraart

Abstract This paper studies the impact of wind power production on electricity prices in the European energy market. We propose a new modelling framework based on so-called regime-switching Lévy semistationary processes to account for forward-looking information consisting of predicted wind power generation. We show that our new regime-switching model, where the regime switch depends on the so-called wind penetration index, can describe recent electricity price data well.

1 Introduction

Renewable sources of energy are of increasing importance in modern energy markets. For instance, the European Union has set the target of increasing the share of energy from renewable sources by 2020 to 20 %. In the German and Austrian energy market, which will be the focus of this paper, the most important source of renewable energy is wind, followed by biogas and solar. Since many renewable sources are highly dependent on weather conditions, they tend to increase the volatility of the corresponding energy prices. It is hence urgent and important to find reliable models which can describe electricity prices in these changing market conditions which can be used for risk assessment and management in energy markets.

The recent literature has presented a variety of both discrete-time and continuous-time time series models which promise to describe the stylised facts of energy markets, see e.g. [3, 17] for reviews. However, reliable models which incorporate information on renewable sources have only recently emerged and have currently been restricted to discrete-time models, see e.g. [9–12, 18].

This paper contributes to the continuous-time literature by introducing for the first time a modelling framework which takes the forward-looking information available to market participants through wind production forecasts into account when

A.E.D. Veraart (✉)
Department of Mathematics, Imperial College London,
180 Queen's Gate, SW7 2AZ London, UK
e-mail: a.veraart@imperial.ac.uk

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_13

modelling electricity day-ahead prices. In doing so, it extends recent work by [1] who proposed to model electricity spot prices by so-called Lévy semistationary processes. Their model consists of a reduced form approach, which only considers electricity prices directly and does not take any price information from other fuels or commodities into account. Note that other types of forward-looking information, such as capacity constraints, have been incorporated in models for electricity prices by [8]. Also, [2] have developed a framework for incorporating forward-looking information in electricity or weather markets through an enlargement of filtrations approach.

With the increasing power generation through wind farms, we have observed that electricity prices at the European Energy Exchange (EEX) started to become negative, which happened for the first time in October 2008, and partially even exhibited rather extreme negative price spikes, see e.g. [16] and the references therein. Also, various articles have argued that increasing wind power production seems to decrease the overall price level, but tends to increase the observed volatility in the market, see e.g. [12]. These are important findings, which need to be incorporated into a modelling framework, one of which will be presented in this paper.

The outline for the remaining part of this article is as follows. In Sect. 2 we give a detailed description of the data from the EEX which will be used in our empirical analysis and we carry out an exploratory data analysis to motivate the new model we are going to introduce in this paper. Section 3 contains the main contribution, where we introduce the new class of regime-switching Lévy semistationary processes and show how they can be calibrated to our empirical data. Finally, Sect. 4 concludes.

2 Exploratory Data Analysis

This section presents the results of an exploratory data analysis of electricity price and wind production data from the European Energy Exchange, which motivates the new modelling framework which we will introduce in Sect. 3.

2.1 Description of the Data

Our empirical data analysis focuses on electricity prices and wind data for a time period from 01.01.2011 to 31.07.2014, i.e. consisting of 1308 days.

We consider three sets of data: electricity prices, their corresponding volumes (sometimes called loads) and wind production data. More precisely, daily EEX Phelix baseload and EEX Phelix baseload volume data (for Germany and Austria) were downloaded from Datastream and the EPEX spot website. In addition, we downloaded the *forecasted wind production data* for the four German Transmission System Operator (TSOs) (50 Hz Transmission, Amprion, Tennet TSO, EnBW Transportnetze (Transnet)) and one Austrian TSO (Verbund (APG)). These data

Table 1 Summary statistics of the EEX Phelix baseload from 01.01.2011 to 31.07.2014

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
–56.87	33.87	43.10	41.96	50.69	98.98

have been aggregated to obtain *daily* forecasts for the wind production for each of the five TSOs.

2.2 EEX Phelix Baseload Prices

First of all, we want to explore the specific properties of the time series of the electricity price data. We focus on the day-ahead electricity prices determined by a daily auction at 12:00 pm, 7 days a week all year (including statutory holidays). The underlying quantity to be traded is the electricity for delivery the following day in 24 h intervals. The prices are bounded (currently between [–500, 3000] EUR/MWh). The EEX Phelix baseload is obtained as the daily averages of the 24 h day-ahead prices for Germany and Austria.

From the summary statistics in Table 1, we notice that the times series does not feature any truly extremes spikes, which have occurred in older data sets from the EEX market. In addition, we notice that there are negative prices even in the baseload prices which consists of the average of the 24 h prices, see also Fig. 1. The plot of the autocorrelation function reveals a clear weekly pattern which is one of the well-known stylised facts of such data.

When we study the distributional properties of the price data, see Fig. 2, we clearly observe that the empirical distribution is not well described by a Gaussian distribution, but appears to be asymmetric and features heavier tails, the latter is particularly pronounced when we focus on the left tail of the distribution.

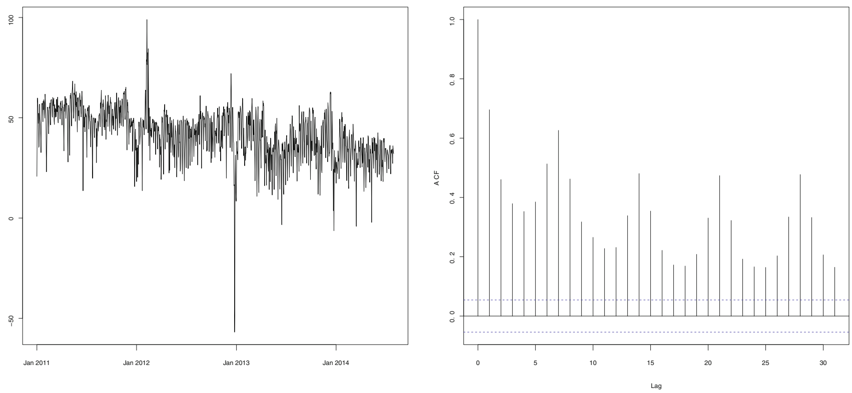

Fig. 1 Time series plot of the baseload prices (in EUR/MWh) and autocorrelation plot

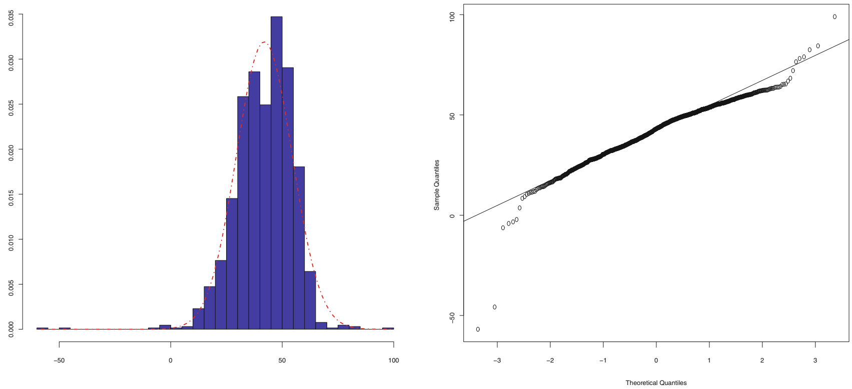

Fig. 2 Distributional properties of the prices: The standardised histogram of the empirical distribution and estimated kernel density function of the prices are depicted in the first plot. The second plot compares the empirical distribution to a Gaussian distribution via a quantile-quantile plot

2.3 Predicted Wind Energy Feed-In

Next we investigate the data from the five Transmission System Operator (TSOs) in Germany and Austria. Note that we are studying the *one-day ahead predicted wind feed-in* since we assume that this is the quantity which impacts the one-day ahead electricity prices determined in the daily auction.

Note that for each TSO, the data is available in 15 min intervals, where the unit of measurement is Megawatt (MW). In order to get the hourly forecasts, we aggregated the data as follows. Let $V_{t,q(i)}^{(q)}$ denote the 15 min wind power forecast for quarter i within hour t. We then obtain the hourly forecasts (recorded in Megawatt hours (MWh)), denoted by $V_{t(j)}^{(h)}$, where $t(j)$ denotes the jth hour on the tth day, from

$$V_{t(j)}^{(h)}[\mathrm{MW}\cdot \mathrm{h}] = \sum_{i=1}^{4} \frac{1}{4}[\mathrm{h}] V_{t(j),q(i)}^{(q)}[\mathrm{MW}].$$

Further, we obtain the daily forecasts, denoted by $V_t^{(d)}$ from

$$V_t^{(d)}[\mathrm{MW}\cdot \mathrm{h}] = \sum_{j=1}^{24} V_{t(j)}^{(h)}[\mathrm{MW}\cdot \mathrm{h}].$$

The data are summarised by TSO in Table 2.

In the following study, we focus on the aggregated time series which accumulates the predicted wind feed-in of all five TSOs, see Table 3 for the summary statistics and Fig. 3 for the time series plot. We clearly observe a yearly cycle in the wind data showing that the predicted wind feed-in is always highest during the winter time period.

Table 2 Forecasted wind feed-in in Gigawatt hours (GWh) (rounded)

	50 Hz	Amprion	APG	Tennet	Transnet	Total
2011	19490	6687	18	18484	416	45095
2012	20203	7253	25	20464	246	48191
2013	19129	7742	31	21259	425	48585
2014 (Until July)	11648	4799	21	13412	399	30279

Table 3 Summary statistics of the forecasted aggregated wind production data (in GWh) from 01.01.2011 to 31.07.2014

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
12510	53870	98940	131600	176000	572600

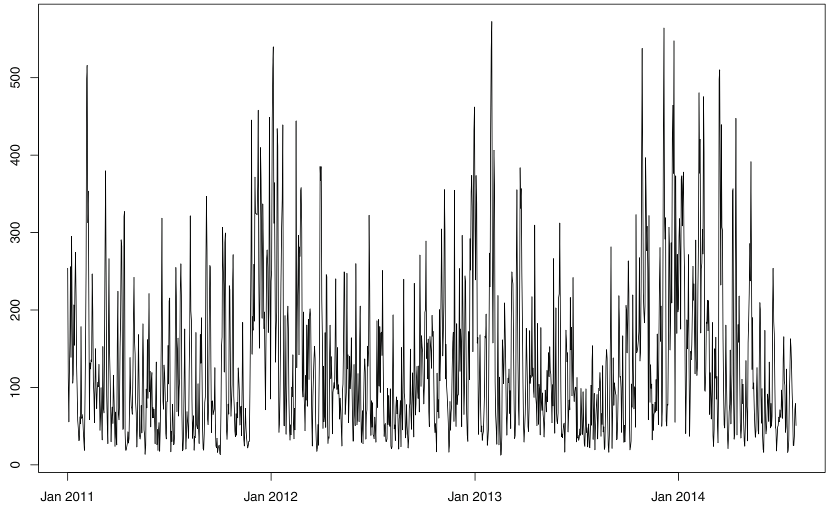

Fig. 3 Time series of the forecasted aggregated wind production data from 01.01.2011 to 31.07.2014 reported in GWh

2.4 *Wind Penetration Index*

Jónsson et al. [10] pointed out that there is a non-linear and time dependent relationship between wind power forecasts and spot prices. Moreover, they found that "it is in fact the ratio between the forecasted wind power generation and the forecasted load that has the strongest association with the spot prices", see [10, p. 314]. Hence, in the following analysis, we will not use the predicted wind feed-in data directly, but rather focus on the so-called *wind penetration index*, which describes the percentage of the wind feed-in compared to the total energy production.

Table 4 Summary statistics of the wind penetration index from 01.01.2011 to 31.07.2014

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.02104	0.08796	0.15480	0.19050	0.26450	0.67540

In order to compute this index, we follow the approach outlined in [10]: They argued that according to recent work by [15], state-of-the-art forecasting models lead to load forecasts where the predicted load equals the actual load plus an error term, i.e. let AL_t denote the actual load on day t and let L_t denote the corresponding predicted load. Then

$$AL_t = L_t + \epsilon_t, \quad \text{where } \epsilon_t \sim N(0, \sigma^2).$$

Typically, the standard deviation σ is chosen as 2 % of the average load for the period considered.

Following this methodology, we downloaded the actual load data from Datastream and the EPEX website and computed the predicted load data by adding Gaussian perturbations to the actual load data. Clearly, this is not exactly the same as working with the predicted load information from each TSO directly, see [10] for a discussion, but the practical impact of this approximation has been found to be marginal.

We can now define the so-called *wind penetration index* on day t as

$$WP_t := \frac{V_t^{(d)}}{L_t}.$$

This is in fact the prediction of the wind penetration on day t, which is available on day $t - 1$ and can hence be considered as forward-looking information. We provide the summary information of the wind penetration index for our sample in Table 4.

Moreover, a time series plot and the corresponding histogram of the wind penetration index is depicted in Fig. 4. We observe that the time series plot of the wind penetration index resembles the one for the original wind data—including a yearly seasonal pattern. Also, the wind penetration index is overall rather low, which is indicated by the histogram and the quantile information contained in Table 4. This is not surprising since the conventional fuels still account for the majority of the electricity production in the European energy market.

2.5 The Relation Between Prices and Wind Data

Finally, we carry out an exploratory study of the relation between the electricity prices and the wind penetration index. In Fig. 5, we plot the electricity prices versus the wind penetration index to check whether we can spot any association between the

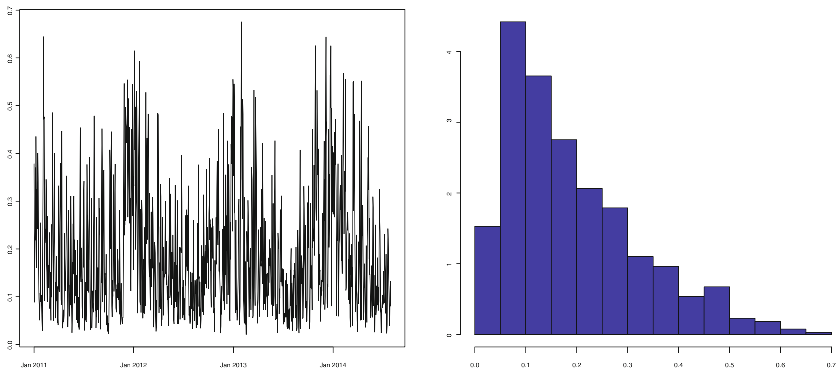

Fig. 4 Time series plot and histogram of the wind penetration index

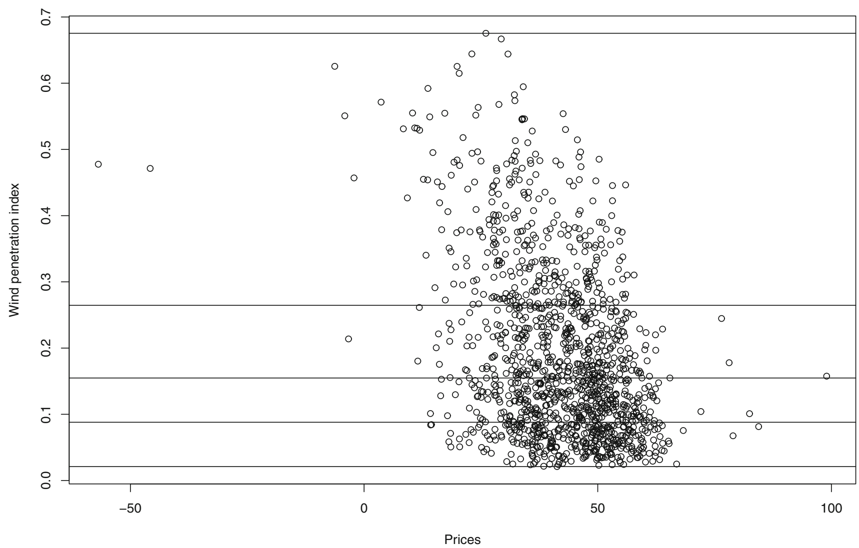

Fig. 5 The electricity prices are plotted versus the wind penetration index. The five *horizontal lines* correspond to the minimum, the 25 %, 50 %, 75 % quantiles and the maximum of the wind penetration index, respectively

two variables. We observe that the two lowest electricity prices are associated with a rather high wind penetration index. Also, for a very high wind penetration index, the prices seem to be below their mean value. This is in line with earlier studies which found that a high wind production typically results in lower electricity prices. However, we need to keep in mind that by comparing the wind and the electricity prices, we can only obtain a partial picture, since clearly other fuels, such as coal, gas and nuclear, play a key role in determining the corresponding electricity price and are for the purpose of this study excluded from the analysis.

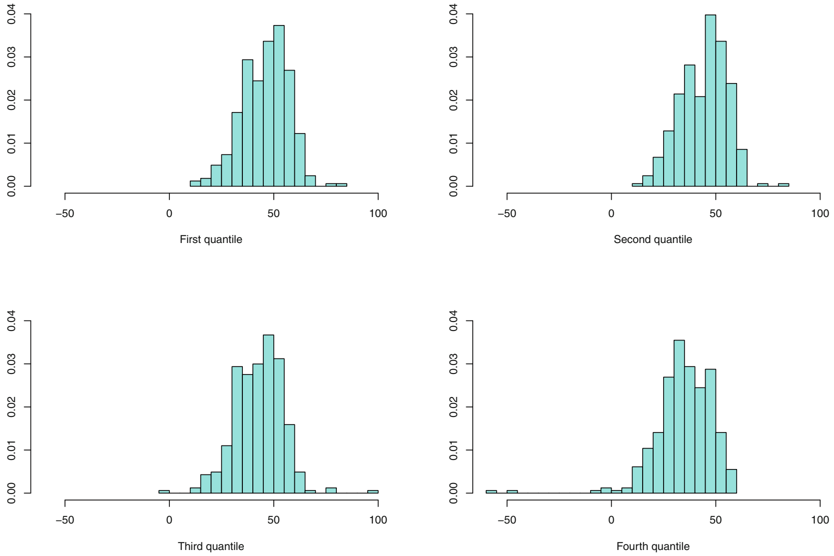

Fig. 6 Distribution of the electricity prices for different quartiles of the wind penetration index. E.g. the first plot corresponds to a low wind penetration index (the first quartile) and the last one to a rather high wind penetration index (the fourth quartile)

We also compare the distribution of the electricity prices associated with different quartiles of the wind penetration index. That is, we have divided our price data into four groups corresponding to the 1st, 2nd, 3rd and 4th quartile of the wind penetration index. When comparing the corresponding marginal distributions, we observe again that smaller price data are associated with a higher wind penetration index, see Fig. 6.

The finding from this exploratory study motivates the new modelling framework which we are going to introduce in the next section.

3 Model Building

Recent work by [1] suggests that the class of so-called *Lévy semistationary* (LSS) processes is very suitable for modelling electricity day-ahead prices. In their work, the class of LSS processes was used in a truly reduced form modelling set-up, meaning that the (deseasonalised) electricity prices were modelled directly by LSS processes and no other external variables were included in the analysis.

Here we will go one step further and explore the possibility of including *forward-looking information* in form of the wind penetration index into a new modelling framework which is based on LSS processes.

In a first step, we are going to review the basic traits of LSS processes and then we will discuss how an LSS-based model can be extended to account for forward-looking information.

An LSS process $Y = \{Y(t)\}_{t \in \mathbb{R}}$ on $\mathbb{R}$ without drift is defined as

$$Y(t) = \int_{-\infty}^{t} g(t-s)\sigma(s-)dL(s), \tag{1}$$

where L denotes a two-sided Lévy process, $g : \mathbb{R} \to \mathbb{R}$ denotes a deterministic weight function satisfying $g(s) = 0$ whenever $s < 0$ and σ denotes a càdlàg, adapted stochastic volatility process, which is assumed to be independent of L. In order to ensure the existence of the stochastic integral, we need suitable integrability conditions on the kernel function g, see [1] for details.

Note that the name Lévy semistationary process indicates that the process Y is stationary as soon as the stochastic volatility process σ is a stationary process. The reason for choosing a stationary process for modelling deseasonalised electricity prices is that commodity prices typically exhibit strong mean reversion. A stationary process can in fact mimic such a behaviour since it ensures that the process cannot move away from its long term mean indefinitely, but will need to return to it since otherwise the stationarity would not be preserved.

Many well-known stochastic processes belong to the LSS class, including volatility modulated Ornstein-Uhlenbeck processes, continuous-time autoregressive moving average (CARMA) processes and fractionally integrated CARMA processes.

It is important to note that LSS processes are in general not semimartingales, which has been discussed in detail in [1]. However, in this paper, we will in fact be staying within the traditional semimartingale framework since we will be working with volatility modulated CARMA processes as the main building blocks for our new model. To this end, let us briefly recall their definition.

Suppose we have nonnegative integers $p > q$ and we wish to define a CARMA (p, q) process. We introduce the autoregressive (AR) and moving average (MA) polynomials:

$$\begin{aligned} P^{\mathrm{AR}(p)}(z) &= z^p + a_1 z^{p-1} + \cdots + a_p, \\ P^{\mathrm{MA}(q)}(z) &= b_0 + b_1 z + \cdots + b_{p-1} z^{p-1}, \end{aligned}$$

where $b_q = 1$ and $b_j = 0$ for $q < j < p$. Moreover, we assume that the polynomials have no common roots and then write formally

$$P^{\mathrm{AR}(p)}(D)Y(t) = P^{\mathrm{MA}(q)}(D)DL(t),$$

where $D = \frac{d}{dt}$. We can make sense of the “derivative” of the Lévy process through a state space representation, where we write

$$Y(t) = \mathbf{b}^\top \mathbf{V}(t)\,, \qquad \text{for } d\mathbf{V}(t) = \mathbf{A}\mathbf{V}(t)dt + \mathbf{e}dL(t), \quad \text{where}$$

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -a_p & -a_{p-1} & \cdots & \cdots & -a_1 \end{pmatrix}, \qquad \mathbf{e} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{p-1} \end{pmatrix}.$$

Assuming that all eigenvalues of $\mathbf{A}$ have negative real parts, we know that

$$\mathbf{V}(t) = \int_{-\infty}^{t} e^{\mathbf{A}(t-s)}\mathbf{e}\, dL(s)$$

is the (strictly) stationary solution of the stochastic differential equation above, see [5]. That is, in our LSS specification we can choose $g(x) = \mathbf{b}^{\top} e^{\mathbf{A}x}\mathbf{e}$ and $\sigma \equiv 1$ to obtain a CARMA(p, q) process. As soon as a stochastic volatility process is added, we would call the corresponding LSS process a *volatility modulated CARMA process*.

3.1 Deseasonalising the Data

We argued before that stationary processes can easily accommodate key stylised facts of commodity prices. However, at the same time, we cannot ignore that strong seasonal effects are typically present in such markets and need to be accounted for. We proceed by introducing an *arithmetic* model for the electricity day-ahead price, denoted by $S = (S(t))_{t\geq 0}$, where

$$S(t) = \Lambda(t) + Y(t).$$

Here, Λ denotes a deterministic seasonality and trend function and Y denotes a stochastic process. In the original framework proposed by [1], the process Y was chosen to be an LSS process. In the following, however, we will introduce a modification of that modelling framework.

The seasonality and trend function is chosen to be

$$\Lambda(t) = c_0 + c_1 t + c_2 \cos\left(\frac{\tau_1 + 2\pi t}{365}\right) + c_3 \cos\left(\frac{\tau_2 + 2\pi t}{7}\right),$$

which accounts for a linear trend and weekly and yearly seasonal cycles. We used a robust least squares estimation procedure to estimate the parameters (by truncating the spikes in the estimation procedure) and obtained the following estimates and standard errors, see Table 5, all of which were highly significant. The estimated parameters in the seasonality function confirm the existence of both weekly and yearly seasonality as well as the presence of a negative trend.

Table 5 Estimated parameters in the seasonality and trend function Λ

	c_0	c_1	c_2	c_3	τ_1	τ_2
Estimate	53.47	–0.01724	2.377	–6.815	349.25	33.59
Standard error	0.4693	0.0006203	0.3356	0.3302	50.46	0.3394

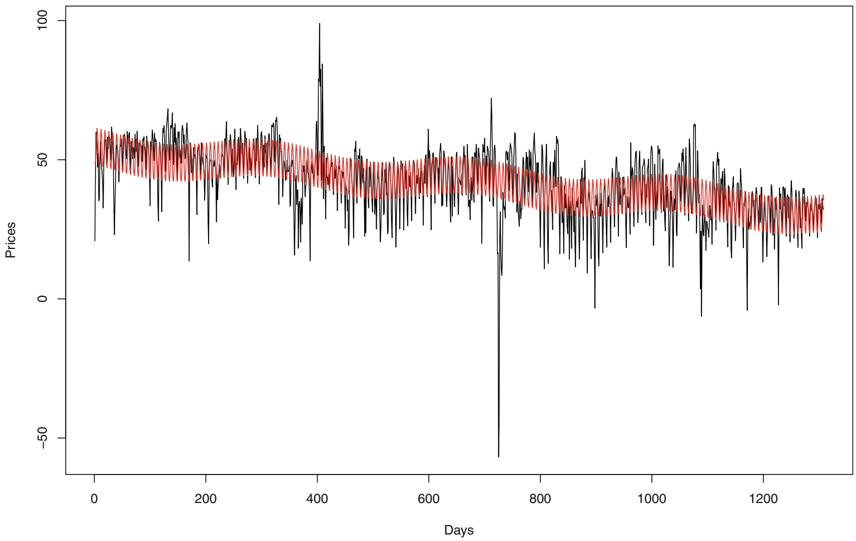

Fig. 7 Electricity day-ahead data and fitted seasonality and trend function

Figure 7 shows a graph of the original data with the fitted seasonality function super-imposed. Also, Fig. 8 depicts the deseasonalised price data and their autocorrelation function.

It should be noted that a variety of alternative procedures could be followed to deal with the problem of seasonality in the electricity prices. Here we are dealing with a rather simple deterministic parametric function to mimic the trend and the seasonal cycles. Using e.g. weekly and yearly dummy variables could refine this approach even further, but would result in a less parsimonious model. A more interesting alternative to the approach pursued here is to acknowledge the fact that the seasonality cannot only be determined by historical data, but also through other market data available to market participants. Example, it has been observed that gas and coal prices, given that they are important fuels used to produce electricity, play a key role in determining trend and also seasonal cycles of electricity prices. This suggests that e.g. forward curve data for gas and/or coal could be used to model the trend, see e.g. [8] for research along those lines. Also [14] give a detailed account on robust estimation procedures of the long-term seasonal component of electricity prices.

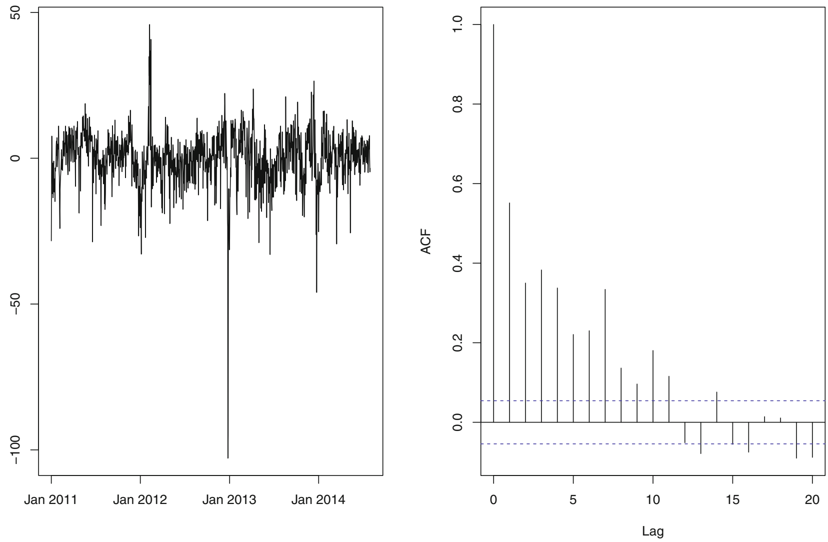

Fig. 8 Time series and autocorrelation plot of the deseasonalised electricity day-ahead data

Table 6 Estimated parameters of the associated ARMA(2,1) process

	AR1	AR2	MA
Estimate	1.1480	–0.2324	–0.6962
Standard error	0.0597	0.0447	0.0501

Here AR1 and AR2 corresponds to the first and second autoregressive parameter, respectively, and MA corresponds to the moving average parameter

3.2 Fitting a CARMA Process

After the seasonality has been removed, we need to find a suitable model for the stochastic process Y. Following the success of the CARMA processes in describing electricity prices, we choose a Lévy semistationary process where the kernel function is given by a kernel associated with a CARMA(p,q) process. More specifically we choose $p = 2$ and $q = 1$. Note that when choosing the order of the CARMA(p,q) process, we need to consider pairs (p, q) such that $p > q$ so that the CARMA process is well defined. We choose a CARMA(2,1) process due to reasons of analytical tractability and increased flexibility compared to a simple Ornstein-Uhlenbeck model. In our goodness-of-fit study, we indeed find that such a model choice is suitable here.

Note that [6] have discussed in detail how a discretely sampled CARMA process can be represented as a weak ARMA process. Using this representation we have first estimated the corresponding ARMA(2,1) parameters by a quasi-maximum likelihood method. The corresponding parameters estimates and standard errors are provided in Table 6.

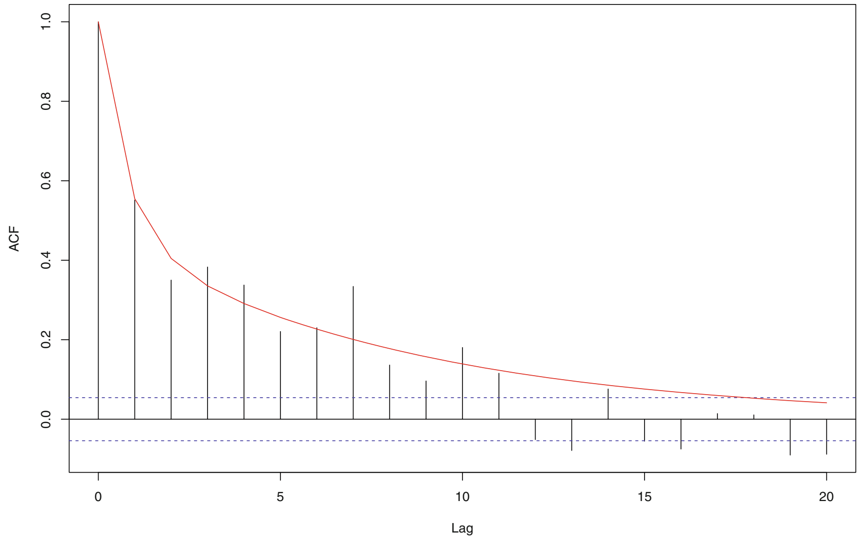

Fig. 9 Empirical (*bars*) and estimated (*solid line*) autocorrelation function of the estimated CARMA(2,1) process

Following the procedure outlined in [6] we can then recover the corresponding continuous-time parameters. In our case, we have $a_1 = 1.459, a_2 = 0.162, b_0 = 0.383$. Note that one can easily verify that the estimates satisfy the condition that the eigenvalues of $\mathbf{A}$ have negative real parts, which implies a stationary model. We compare the empirical and estimated autocorrelation function in Fig. 9, where we observe a good fit.

Under the assumption that the CARMA process is driven by a subordinator, [6] have shown how the corresponding increments of the driving Lévy process can be recovered from a discretely observed CARMA process, see also [7] for the multivariate case. Here we have implemented their algorithm for the case of a CARMA(2,1) process and have recovered the driving process. Note that the original algorithm was designed for driving Lévy processes, but can in fact be adapted to the case of volatility modulated Lévy processes as well, the case which will be relevant in the next section.

Let us briefly recap our estimation procedure until now: We started off with a spot price model $S(t) = \Lambda(t) + Y(t)$, where we have estimated the seasonality function Λ and have removed it from the data. In the next step, we have assumed that Y is an LSS process of the form

$$Y(t) = \int_{-\infty}^{t} g(t-s)dM(s),$$

where g corresponds to the kernel function associated with a CARMA(2,1) process and initially M was assumed to be a Lévy process, which can be recovered from the observations of the CARMA process. We will now leave this traditional framework behind and will introduce a new regime-switching model based on LSS processes.

3.3 The New Model Based on a Regime-Switching LSS Process

The predicted wind penetration index can be viewed as forward-looking information since the information is available before the prices for the next day are determined in the auction market. Hence it is reasonable to try to incorporate this information in the model.

Previous studies have included such information in discrete-time models such as e.g. ARMAX-GARCHX models, see [12], where the wind is treated as an exogenous variable. However, we are interested in a continuous-time modelling framework. E.g. one could consider CARMA-X models or regime-switching models. Here we will follow the latter approach which is motivated by the work by [8] who incorporated forward-looking capacity constraints into a jump-diffusion model for electricity prices.

We introduce an exogenous regime-switching variable based on the forward-looking variable given by the predicted wind penetration index ρ, where

$$\rho(s) = \begin{cases} 1, & \text{if the predicted wind penetration at time } s \text{ is ``high''} \\ 0, & \text{if the predicted wind penetration at time } s \text{ is ``low''}. \end{cases} \tag{2}$$

The new spot price model is then given by $S(t) = \Lambda(t) + Y(t)$, where

$$Y(t) = \int_{-\infty}^{t} g(t-s)dM(s), \qquad t \geq 0.$$

Here

$$dM(s) = \rho(s)dM^{(1)}(s) + (1-\rho(s))dM^{(2)}(s),$$

where

$$dM^{(i)}(s) = a^{(i)}(s)ds + \sigma^{(i)}(s-)dL^{(i)}(s),$$

for independent Lévy processes $L^{(1)}$ and $L^{(2)}$. Also, $a^{(i)}$ denote suitable drift and $\sigma^{(i)}$ stochastic volatility processes, for $i \in \{1, 2\}$.

The key question which remains to be addressed is how exactly the regime-switching variable ρ should be chosen, given that the expression in (2) appears

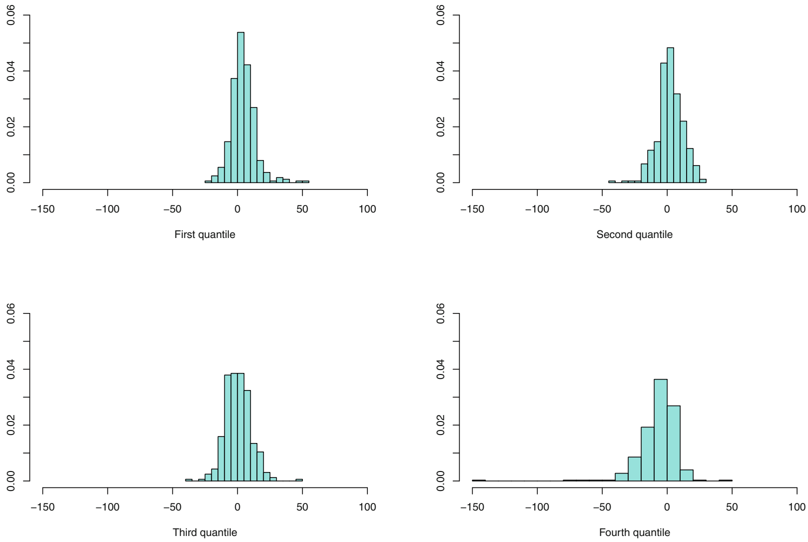

Fig. 10 Standardised histograms of the increments of M for different levels (associated with the four quartiles) of the wind penetration index

rather casual. In order to answer this question, we study the empirical properties of the recovered driving process M. More precisely, we investigate how the marginal distribution of the driving process of the CARMA process changes in relation to different levels of the wind penetration index. We split the sample of the recovered increments of M into four parts corresponding to the four quartiles of the wind penetration index, which are given in Table 4.

Their empirical distributions are described in form of standardised histograms, which describe the empirical probability density functions, in Fig. 10. Similarly to the finding in our exploratory data analysis, we observe that also the distribution of the increments of the driving process M changes quite remarkably for different levels of the wind penetration index. In particular, we observe that rather extreme negative increments are associated with a relatively high wind penetration index.

One can imagine a variety of rather sophisticated methods for choosing the cut-off point for our regime-switching variable. Here we are interested in a rather simple rule, which at the same time allows for a reasonable amount of observations in the high regime so that inference is still feasible and does not just rely on a very small number of observations. Hence, we choose the cut-off point to be 26.4 % as a hard threshold, meaning that all increments of M associated with a wind penetration index in the fourth quartile belong to the high regime.

3.4 Model for M Based on the Generalised Hyperbolic Distribution

Motivated by the empirical study in [1], we will fit the class of *generalised hyperbolic* (GH) distributions to the increments of M in the two regimes.

Our notation for the GH distribution follows the one used in [13]. See also [4] for more details on the implementation of the corresponding estimation procedures in `R` available through the `ghyp` package.

Let us denote by $d, k \in \mathbb{N}$ some constants and let $\mathbf{X}$ denote a k–dimensional random vector. Recall that we say that the law of $\mathbf{X}$ is given by the *multivariate generalised hyperbolic* (GH) distribution if

$$\mathbf{X} \overset{\text{law}}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}\mathbf{C}\mathbf{Z},$$

where $\mathbf{Z} \sim N(\mathbf{0}, I_k)$, $\mathbf{C} \in \mathbb{R}^{d\times k}$, $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^d$. Here $W \geq 0$ denotes a one-dimensional random variable, independent of $\mathbf{Z}$ and with Generalised Inverse Gaussian (GIG) distribution, i.e. $W \sim GIG(\lambda, \chi, \psi)$. The density of the GIG distribution with parameters (λ, χ, ψ) has the following functional form:

$$f_{GIG}(x) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{x^{\lambda-1}}{2K_\lambda(\sqrt{\chi\psi})} \exp\left(-\frac{1}{2}\left(\frac{\chi}{x} + \psi x\right)\right),$$

where K_λ denotes the modified Bessel function of the third kind, and the parameters have to satisfy one of the following three restrictions

$$\chi > 0, \psi \geq 0, \lambda < 0, \quad \text{or} \quad \chi > 0, \psi > 0, \lambda = 0, \quad \text{or} \quad \chi \geq 0, \psi > 0, \lambda > 0.$$

The parameter $\boldsymbol{\mu}$ is called the location parameter, $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}'$ is the dispersion matrix and $\boldsymbol{\gamma}$ is the symmetry or skewness parameter. The three (scalar) parameters λ, χ, ψ of the GIG distribution determine the shape of the GH distribution. The parametrisation described above is referred to as the so-called $(\lambda, \chi, \psi, \mu, \boldsymbol{\Sigma}, \gamma)$–parametrisation of the GH distribution. Since this parametrisation causes an identifiability problem when one tries to estimate the parameters, we will work with the so-called $(\lambda, \overline{\alpha}, \mu, \boldsymbol{\Sigma}, \gamma)$–parametrisation in our empirical study. One can show that the $(\lambda, \chi, \psi, \mu, \boldsymbol{\Sigma}, \gamma)$–parametrisation can be obtained from the $(\lambda, \overline{\alpha}, \mu, \boldsymbol{\Sigma}, \gamma)$–parametrisation by setting

$$\psi = \overline{\alpha}\frac{K_{\lambda+1}(\overline{\alpha})}{K_\lambda(\overline{\alpha})}, \qquad \chi = \frac{\overline{\alpha}^2}{\psi} = \overline{\alpha}\frac{K_\lambda(\overline{\alpha})}{K_{\lambda+1}(\overline{\alpha})},$$

and $\lambda, \boldsymbol{\Sigma}, \gamma$ remain the same, cf. [4].

We estimated 11 distributions within the GH class—consisting of the asymmetric and symmetric versions of the GH, hyperbolic, Student's t, variance gamma, normal inverse Gaussian distribution and the Gaussian distribution and compared them

Table 7 Parameter estimates for the Student's t-distribution in the low regime (symmetric case) and high regime (asymmetric case)

	$\widehat{\nu}$	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\gamma}$
Low regime	6.70	2.26	9.77	0
High regime	4.76	0.57	11.31	–7.68

Note that the parameter $\overline{\alpha} = 0$ in the case of the Student's t-distribution

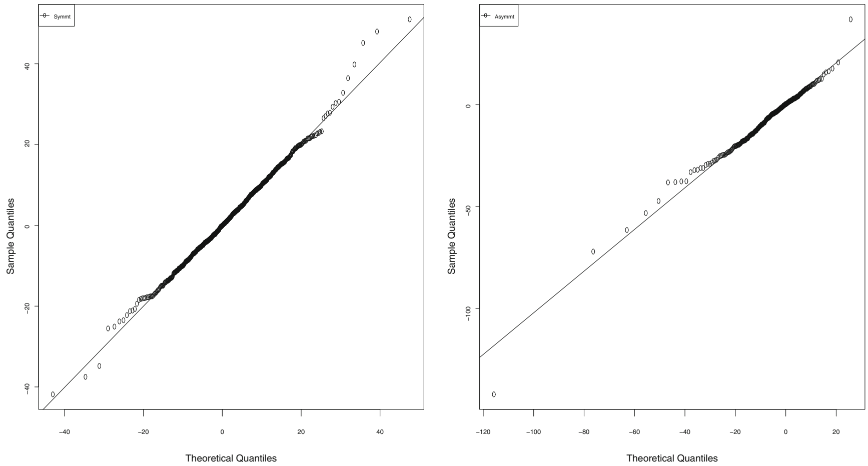

Fig. 11 Diagnostic quantile-quantile plots: The first picture compares the empirical quantiles of the data in the *low* regime with the estimated symmetric Student's t-quantiles, and the second picture compares the empirical quantiles of the data in the *high* regime with the estimated asymmetric Student's t-quantiles

according to the Akaike information criterion. We found that the best model for the low regime is given by the symmetric Student's t-distribution and for the high regime by the asymmetric Student's t-distribution, see Table 7 for the corresponding parameter estimates.

When comparing the parameter estimates for the low and the high regime provided in Table 7, we observe that the skewness, fatness of the tails and the volatility increases for the high regime and that the mean parameter decreases compared with the low regime. This is in line with previous findings in the literature, which suggest that the price level typically decreases with increasing wind energy production and that the volatility and the risk for negative spikes (represented through negative skewness and fatter tails) typically increases.

Also, we have provided quantile-quantile plots to assess the goodness-of-fit of the Student's t-distribution in Fig. 11, which overall look reasonable.

The estimation results suggest that a good model for the driving process M in the regime-switching LSS specification is given by

$$dM(s) = \rho(s)dM^{(1)}(s) + (1-\rho(s))dM^{(2)}(s),$$

where

$$dM^{(1)}(s) = \left(\mu^{(1)} + \gamma\left(\sigma^{(1)}(s)\right)^2\right)ds + \sigma^{(1)}(s)dW^{(1)}(s),$$
$$dM^{(2)}(s) = \mu^{(2)}ds + \sigma^{(2)}(s)dW^{(2)}(s),$$

for independent Brownian motions $W^{(1)}$ and $W^{(2)}$. Here the stochastic volatility processes $\sigma^{(i)}$ are chosen as Ornstein-Uhlenbeck processes with inverse Gamma marginal distribution, since a mean-variance mixture with the inverse Gamma distribution results in the Student's t-distribution.

Note that the reason for choosing volatility modulated Brownian motions rather than Lévy processes with Student's t-distribution is that we found a significant short term (2 lags) autocorrelation in the increments of the recovered process M suggesting that a stochastic volatility model is more suitable than a pure jump model. This finding reveals that stochastic volatility is a key feature in energy markets, but it typically only exhibits short memory. Stochastic volatility is naturally incorporated into the LSS framework making it a convincing modelling tool for energy markets.

4 Conclusion

This paper has presented an extension of the modelling framework based on Lévy semistationary (LSS) processes introduced by [1]. Since forward-looking information in terms of weather forecasts is available to market participants, the corresponding predictions for the day-ahead wind production can be derived and used when determining day-ahead electricity spot prices. We incorporated this information through the so-called predicted wind penetration index in a regime-switching model based on LSS processes. We have observed that the flexibility offered through the regime switching component allows to model electricity prices in a more refined way than it was possible in the original (reduced form) LSS modelling framework. In particular, we have found that a relatively high wind penetration index leads to a lower mean level, higher skewness, fatter tails and increased volatility in the distribution of the electricity prices. This confirms earlier findings in the literature and for the first time links them to a flexible continuous-time stochastic modelling framework.

Given the increasing importance of renewable sources of energy, it will be interesting to extend the current investigation to include a wider range of renewables, including solar and biogas which up to now do not play as big of a role as wind power generation in the European energy market.

Another area for future research would be to develop a stochastic model for the wind penetration index, which could either result in a regime-switching model with a stochastic switching parameter or in a joint model for electricity prices and the wind generation index. A preliminary analysis along those lines has revealed that a reasonable model for ρ needs to take both yearly seasonality and clusters into account. This could be seen as a first step for constructing models which can be used for mid-term forecasts of electricity prices influenced by renewables and would help to find modelling and inference tools for reliable risk management in energy markets.

Acknowledgments A.E.D. Veraart acknowledges financial support by a Marie Curie FP7 Integration Grant within the 7th European Union Framework Programme.

References

1. Barndorff-Nielsen, O.E., Benth, F.E., Veraart, A.E.D.: Modelling energy spot prices by volatility modulated Lévy-driven Volterra processes. Bernoulli **19**(3), 803–845 (2013)
2. Benth, F.E., Meyer-Brandis, T.: The information premium for non-storable commodities. J. Energy Markets **2**(3), 111–140 (2009)
3. Benth, F.E., Šaltytė Benth, J., Koekebakker, S.: Stochastic modelling of electricity and related markets. In: Advanced Series on Statistical Science and Applied Probability, vol. 11. World Scientific (2008)
4. Breymann, W., Lüthi, D.: ghyp: a package on generalized hyperbolic distributions. Manual for the R package $ghyp$ (2010)
5. Brockwell, P.: Representations of continuous-time ARMA processes. J. Appl. Probab. **41**(A), 375–382 (2004)
6. Brockwell, P.J., Davis, R.A., Yang, Y.: Estimation for non-negative Lévy-driven CARMA processes. J. Bus. Econ. Statist. **29**(2), 250–259 (2011)
7. Brockwell, P.J., Schlemm, E.: Parametric estimation of the driving Lévy process of multivariate CARMA processes from discrete observations. J. Multivar. Anal. **115**, 217–251 (2013)
8. Cartea, Á., Figueroa, M.G., Geman, H.: Modelling electricity prices with forward looking capacity constraints. Appl. Math. Finance **16**(1–2), 103–122 (2009)
9. Forrest, S., MacGill, I.: Assessing the impact of wind generation on wholesale prices and generator dispatch in the Australian national electricity market. Energy Policy **59**, 120–132 (2013)
10. Jónsson, T., Pinson, P., Madsen, H.: On the market impact of wind energy forecasts. Energy Econ. **32**(2), 313–320 (2010)
11. Jónsson, T., Pinson, P., Nielsen, H., Madsen, H., Nielsen, T.: Forecasting electricity spot prices accounting for wind power predictions. Sustain. Energy, IEEE Trans. **4**(1), 210–218 (2013). Jan
12. Ketterer, J.C.: The impact of wind power generation on the electricity price in Germany. Energy Econ. **44**, 270–280 (2014)
13. McNeil, A.J., Frey, R., Embrechts, P.: Quantitative risk management: concepts, techniques and tools. Princeton Series in Finance. Princeton University Press, Princeton, NJ (2005)
14. Nowotarski, J., Tomczyk, J., Weron, R.: Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. Energy Econ. **39**, 13–27 (2013)
15. Taylor, J.W., McSharry, P.E.: Short-term load forecasting methods: an evaluation based on European data. IEEE Trans. Power Syst. **22**(4), 2213–2219 (2007)

16. Veraart, A.E.D., Veraart, L.A.M.: Modelling electricity day-ahead prices by multivariate Lévy semi-stationary processes. In: Benth, F.E., Kholodnyi, V., Laurence, P. (eds.) Wolfgang Pauli Proceedings, pp. 157–188. Springer (2014)
17. Weron. R.: Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. Wiley, Chichester (2006). With CD-ROM with MFE Toolbox
18. Würzburg, K., Labandeira, X., Linares, P.: Renewable generation and electricity prices: taking stock and new evidence for Germany and Austria. Energy Econ. **40**, S159–S171 (2013). Supplement Issue: Fifth Atlantic Workshop in Energy and Environmental Economics

Pricing Options on EU ETS Certificates with a Time-Varying Market Price of Risk Model

Ya Wen and Rüdiger Kiesel

Abstract To price options on emission certificates reduced-form models have proved to be useful. We empirically analyse the performance of the model proposed in Carmona and Hinz [2] and Hinz [8]. As we find evidence for a time-varying market price of risk, we extend the Carmona-Hinz framework by introducing a bivariate pricing model. We show that the extended model is able to extract information on the market price of risk and evaluate its impact on the EUA options.

Keywords Carbon market · EUA futures · Risk-neutral valuation · Market price of risk · Option pricing

1 Introduction

The European Union Emissions Trading Scheme (EU ETS) was launched in 2005 and constitutes still the world's largest carbon market. The EU ETS was set up as a cap-and-trade scheme and split up into three phases, namely Phase I (2005–07) without the possibility to bank unused permits; Phase II (2008–12) in which banking was allowed; and the current Phase III (2013–20) which, compared to the two previous periods, introduced significant changes, such as the abolishment of national allocation plans and the auctioning of permits.

Besides permits, futures and options on permits are being traded. Various authors have discussed the design of the market and the pricing of the permits and the derivatives traded. The fundamental concepts for emission trading and the market mechanism have been reviewed in the paper of Taschini [15], which also provides a literature overview. Equilibrium models for allowance permit markets have been

Y. Wen · R. Kiesel (✉)
Faculty of Business Administration and Economics, University of Duisburg-Essen,
Universitätsstraße 12, 45145 Essen, Germany
e-mail: ruediger.kiesel@uni-due.de

R. Kiesel
e-mail: ya.wen@uni-due.de

F.E. Benth and G. Di Nunno (eds.), *Stochastics of Environmental and Financial Economics*, Springer Proceedings in Mathematics and Statistics 138,
DOI 10.1007/978-3-319-23425-0_14

widely used to capture the theoretical properties of emission trading schemes. Examples are the dynamic but deterministic model proposed by Rubin [13] and stochastic equilibrium approaches such as Seifert et al. [14], Wagner [16] and Carmona et al. [3]. These models use optimal stochastic control to investigate the dynamic emission trading in the risk-neutral framework. Carmona et al. [4] derive the permit price formula which can be described as the discounted penalty multiplied by the probability of the excess demand event. Its historical model fit has been evaluated by Grüll and Taschini [6]. Grüll and Kiesel [5] used the formula to analyse the emission permit price drop during the first compliance period. Carmona and Hinz [2] and Hinz [8] propose a reduced-form model which is particular feasible for the calibration of EUA futures and options as it directly models the underlying price process. Both Paolella and Taschini [12] and Benz and Trück [1] provide an econometric analysis for the short-term spot price behavior and the heteroscedastic dynamics of the price returns. For the option pricing, Carmona and Hinz [2] derive a option pricing formula from their reduced-form model for a single trading period. They also discuss the extension of the formula to two trading periods. Hitzemann and Uhrig-Homburg [9, 10] develop an option pricing model for multi-compliance periods by considering a remaining value component in the pricing formula capturing the expected value after a finite number of trading periods.

As emission certificates are traded assets their price paths carry information on the market participant expectations on the development of the fundamental price drivers of the certificates including the regulatory framework. In particular, prices of futures and options of certificates carry forward-looking information which can be extracted by using appropriate valuation models. In this paper we derive such a model by extending the reduced-form pricing model of Carmona and Hinz [2]. Using an extensive data set we extract a time series for the implied market price of risk, which relates to the risk premium the investors attach to the certificates. This requires a calibration of the model to historical price data during varying time periods and with different maturities of futures and options. A crucial step in the calibration procedure is a price transformation of normalized futures prices of permits from a pricing measure to the historical measure. We find that the implied market price of risk possesses stochastic characteristics. Therefore, we extend the existing reduced-form model by modeling the dynamics of the market price of risk as an Ornstein-Uhlenbeck process and show that the extended model captures the appropriate properties of the market. The market price of risk is an implied value related to the permit prices, this requires an extension of the univariate permit pricing model to a bivariate one. In this context, the standard Kalman filter algorithm is considered to be an effective way to calibrate to the historical prices. We apply this methodology and estimate the implied risk premia. Once the risk premia have been determined, EUA option prices can be derived to fit the bivariate model setting, which helps to improve the accuracy of the pricing.

This paper is organized as follows. In Sect. 2 we introduce the basic reduced-form model based on a risk-neutral framework. We calibrate the model with an extended data series and compare the calibration results. In Sect. 3 an extended bivariate pricing model will be introduced in order to capture the market information of the risk premia.

We demonstrate how to calibrate the extended model by applying the standard Kalman filter algorithm, we present the estimation results of this procedure and discuss the model fit. In Sect. 4 we evaluate the option pricing performance by taking into account the calibration results of the bivariate model. Section 5 concludes.

2 Univariate EUA Pricing Model and Parameter Estimation

The basic reduced-form model based on a risk-neutral framework was introduced by Carmona and Hinz [2]. Here, the aggregated normalised emissions are modelled directly and it can be shown that the emission certificate futures process solves a stochastic differential equation. In this section we give a brief introduction to the model and discuss the quality of the model calibration.

2.1 Univariate Model

We consider an emission trading scheme with a single trading phase with horizon $[0, T]$. The price evolution of emission permits is assumed to be adapted stochastic processes on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,T]}, \mathbb{P})$ with an equivalent risk-neutral measure $\mathbb{Q} \sim \mathbb{P}$. Based on the assumption of a market compliance at time T the price has only two possible outcomes, namely zero or the penalty level. The argument is as follows. If there are sufficient emission allowances in the market to cover the total emissions at compliance time, surplus allowances will become worthless. Otherwise, for undersupplied permits the price will increase to the penalty level.

We introduce the reduced-form model of Carmona and Hinz [2]. Here the normalized futures price process is a martingale under $\mathbb{Q}$ given by

$$a_t = \mathbb{E}^{\mathbb{Q}}[1_{\{\Gamma_T \geq 1\}} | \mathscr{F}_t], \quad t \in [0, T]. \tag{1}$$

The process $(\Gamma_t)_{t\in[0,T]}$ denotes the aggregated normalized emission, and is assumed to follow a lognormal process given by

$$\Gamma_t = \Gamma_0 e^{\int_0^t \sigma_s d\widetilde{W}_s - \frac{1}{2}\int_0^t \sigma_s^2 ds}, \quad \Gamma_0 \in (0, \infty),$$

where σ_t stands for the volatility of the emission pollution rate. $t \hookrightarrow \sigma_t$ is a deterministic function which is continuous and square-integrable. $(\widetilde{W}_t)_{t\in[0,T]}$ is a Brownian motion with respect to $\mathbb{Q}$. Carmona and Hinz [2] prove that a_t solves the stochastic differential equation

$$da_t = \Phi'(\Phi^{-1}(a_t))\sqrt{z_t} d\widetilde{W}_t, \tag{2}$$

with the function $t \hookrightarrow z_t$, $t \in (0, T)$, given by

$$z_t = \frac{\sigma_t^2}{\int_t^T \sigma_u^2 du}. \tag{3}$$

In order to calibrate the model, Carmona and Hinz [2] suggest to use the function

$$z_t = \beta(T-t)^{-\alpha},$$

with $\alpha \in \mathbb{R}$ and $\beta \in (0, \infty)$, so one has

$$da_t = \Phi'(\Phi(a_t))\sqrt{\beta(T-t)^{-\alpha}}d\widetilde{W}_t. \tag{4}$$

To estimate the parameters one has to determine the distribution of the price variable. For this purpose one considers the price transformation process ξ_t defined by $a_t = \Phi(\xi_t)$, where Φ denotes the cumulative distribution function of the standard normal distribution. By applying Itô's formula one has

$$d\xi_t = \left(\frac{1}{2}z_t\xi_t + \sqrt{z_t}h\right)dt + \sqrt{z_t}dW_t, \tag{5}$$

where W_t denotes the Brownian motion under the objective measure $\mathbb{P}$ and h is the market price of risk which is assumed to be constant. Moreover, it can be shown that ξ_t is conditional Gaussian with mean μ_t and variance σ_t^2 so that the log-likelihood can be calculated and Maximum-likelihood estimation can be applied to find the model parameters.

2.2 Estimation

We calibrate the model to different emission trading periods during the first and second EU ETS trading phases. We consider the daily prices of the EUA futures with maturities in December from 2005 to 2012. Their historical price series are shown in Fig. 1.

The price transformation ξ_t is conditional Gaussian with its mean μ_t and variance σ_t^2. We consider the daily historical observations of the EUA futures at time $t_1, t_2, \ldots, t_n$. Their corresponding price transformations can be determined using the definition $a_t = \Phi(\xi_t)$. Thus the parameters α, β, h can be estimated by maximizing the log-likelihood function given by

$$L_{\xi_{t_i},\ldots,\xi_{t_n}}(h,\alpha,\beta) = \sum_{i=1}^{n}\left(-\frac{(\xi_{t_i} - \mu_i(h,\alpha,\beta))^2}{2\sigma_i^2(\alpha,\beta)} - \ln\left(\sqrt{2\pi\sigma_i^2(\alpha,\beta)}\right)\right). \tag{6}$$

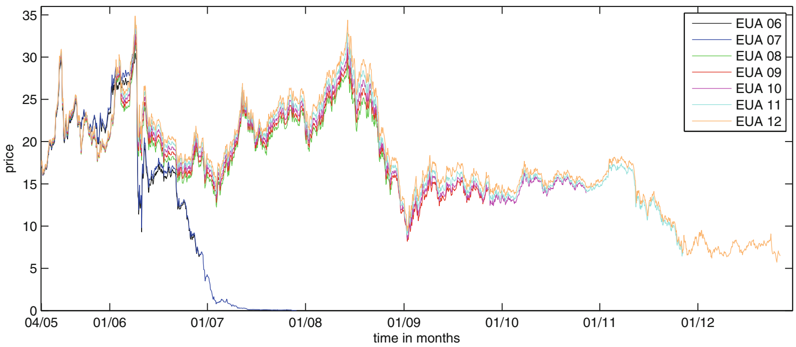

Fig. 1 Historical prices of the EUA futures with maturities in December from 2005 to 2012

Under the model assumptions the residuals

$$w_i = \frac{\xi_{t_i} - \mu_i(h, \alpha, \beta)}{\sqrt{\sigma_i^2(\alpha, \beta)}}, \quad i = 1, \ldots, n, \tag{7}$$

must be a series of independent standard normal random variables. So standard statistical analysis can be applied to test the quality of the model fit. We show our estimation results in Table 1. The horizons of the price data are two years, starting from the first trading day in January of the previous trading year to the last trading day in December of the next year.

Comparing the estimation values in Table 1, the instability of the parameter values in each cell for different time periods can be observed. Note the value for the market price of risk changes its sign during the first and second trading phase. This implies the inappropriateness of the assumption for a constant market price of risk. The fourth value in each cell is the negative of LLF. Note the -LLF are much lower after the first trading phase because the price collapse during 2006 to 2007 affects the data partially.

From Figs. 2, 3, 4, 5 and 6 we display the time series of the residuals w_i, their empirical auto-correlations, empirical partial auto-correlations and quantile-quantile-plots. We choose EUA futures with maturity in December 2007 (EUA 07) and EUA futures with maturity in December 2012 (EUA 12) as examples.

The time series w_i show an effect of volatility clustering. This is confirmed by significant values to high lags in the sample autocorrelation and sample partial autocorrelation. Also the Q-Q plots, especially for the first trading phase, indicate heavy tails and a non-Gaussian behavior. A formal analysis with an application of Jarque-Bera test rejects the hypothesis that the data set is generated from normally distributed random variables. In order to improve the model fit we extend the model by introduction of a dynamic market price of risk in Sect. 3.

Table 1 Parameter estimate results

	Time period						
Maturity	Apr.05–Dec.06	Jan.06–Dec.07	Jan.07–Dec.08	Jan.08–Dec.09	Jan.09–Dec.10	Jan.10–Dec.11	Jan.11–Dec.12
Dec.06	−0.3501						
	0.7818						
	0.7470						
	−644.5836						
Dec.07	−0.4303	0.1326					
	0.5556	0.7599					
	0.6002	1.3959					
	−646.0240	−739.0301					
Dec.08	−0.8022	−1.9960	0.0870				
	0.0632	0.0260	0.0751				
	−0.1473	−0.3918	0.0551				
	−989.4653	−1276.0276	−1324.2164				
Dec.09	−0.8995	−3.2116	0.0191	−0.2602			
	0.0438	0.0028	0.0748	0.0811			
	−0.2260	−0.4051	0.0899	0.4049			
	−984.3740	−1279.0503	−1358.3749	−1312.4067			
Dec.10	−1.1852	−4.3385	0.0104	−0.3174	−0.7881		
	0.0221	0.0002	0.0750	0.0645	0.0544		
	−0.2480	−0.4157	0.0896	0.6054	0.2404		
	−986.8374	−1281.3926	−1357.4328	−1331.5990	−1455.5480		

(continued)

Table 1 (continued)

	Time period						
Maturity	Apr.05–Dec.06	Jan.06–Dec.07	Jan.07–Dec.08	Jan.08–Dec.09	Jan.09–Dec.10	Jan.10–Dec.11	Jan.11–Dec.12
Dec.11	−1.3765	−5.7102	0.0579	−0.4098	−2.2939	0.4460	
	0.0129	0.0001	0.0791	0.0506	0.0092	0.0331	
	−0.2729	−0.4358	0.0834	0.5952	0.1349	0.4657	
	−978.9760	−1279.6562	−1361.8570	−1332.8319	−1497.0551	−1528.8444	
Dec.12	−1.6486	−6.8341	0.0277	−0.5419	−3.5805	1.4572	0.1279
	0.0062	0.0000	0.0782	0.0373	0.0008	0.1090	0.0680
	−0.2951	−0.4138	0.0611	0.5816	0.1823	0.4183	0.6948
	−976.2664	−1273.2228	−1355.5116	−1332.0014	−1500.6628	−1524.7310	−1355.8639

In each cell the first value stands for α, second for β, third for the market price of risk (MPR) h, the last one for the negative of LLF. Note that from 2005 to 2007 is the first trading phase, from 2008 to 2012 is the second trading phase

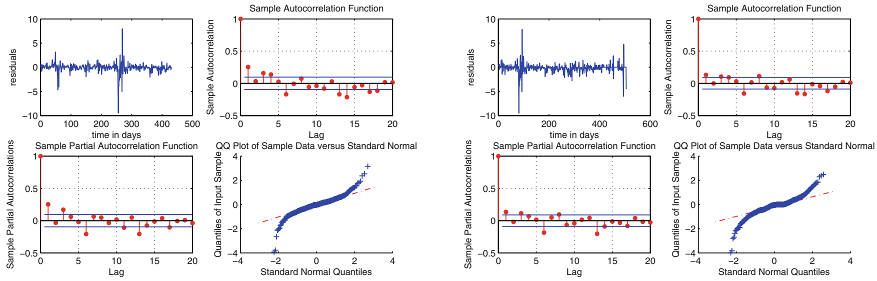

Fig. 2 Statistical analysis of EUA 07, time period 05–06 and 06–07

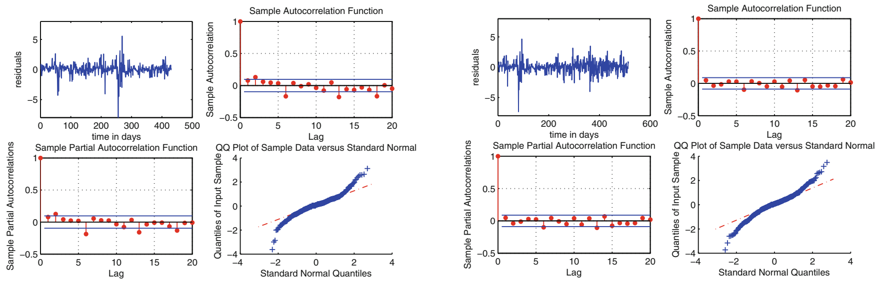

Fig. 3 Statistical analysis of EUA 12, time period 05–06 and 06–07

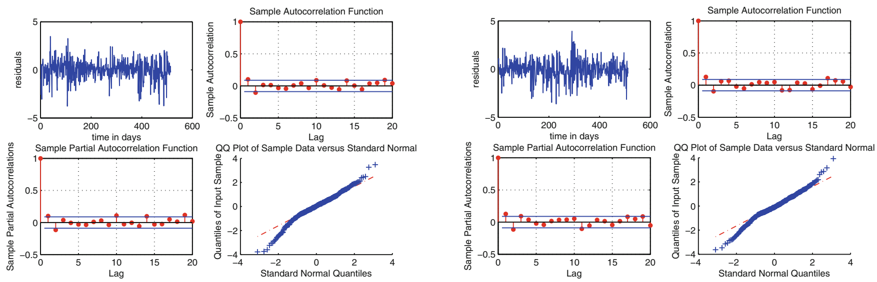

Fig. 4 Statistical analysis of EUA 12, time period 07–08 and 08–09

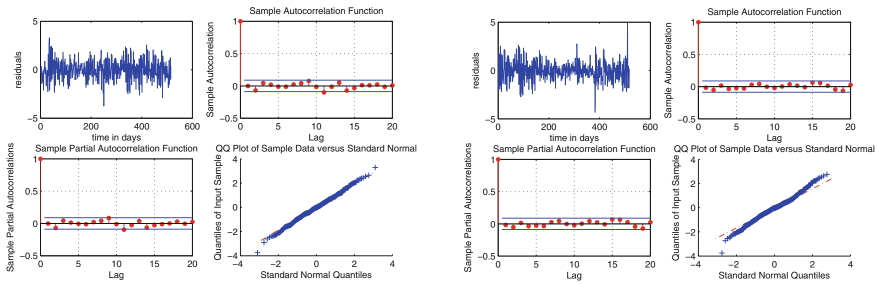

Fig. 5 Statistical analysis of EUA 12, time period 09–10 and 10–11

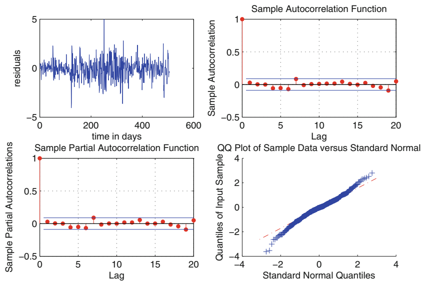

Fig. 6 Statistical analysis of EUA 12, time period 11–12

3 Bivariate Pricing Model for EUA

The evidence in the previous section shows that the market price of risk is actually time varying rather than constant. In order to illustrate the dynamic property of the market price of risk we consider a bivariate permit pricing model in this section.

3.1 Model Description

We model the market price of risk as an Ornstein-Uhlenbeck process as its value can be either positive or negative and denote it by λ_t. Recall the equation for the normalized price process under the risk-neutral measure $\mathbb{Q}$ given by (2). According to Girsanov's theorem, the bivariate pricing model under the objective measure $\mathbb{P}$ is given by

$$\begin{aligned} da_t &= \Phi'(\Phi^{-1}(a_t))\sqrt{z_t}(\lambda_t dt + dW_t^1), \\ d\lambda_t &= \theta(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda dW_t^2, \\ dW_t^1 dW_t^2 &= \rho dt, \end{aligned}$$

where W_t^1 and W_t^2 are two one-dimensional Brownian motions with correlation coefficient ρ. Note that under the model assumptions, the filtration $(\mathscr{F}_t)$ in the probability space must be assumed to be generated by the bivariate Brownian motion.

The use of Girsanov's theorem in the bivariate model requires the condition that the process Z_t given by

$$Z_t = \exp\left(\int_0^t \lambda_s dW_s - \frac{1}{2}\int_0^t \lambda_s^2 ds\right) \tag{8}$$

is a martingale. A sufficient condition for (8) is Novikov's condition:

$$\mathbb{E}\Big[\exp\left(\frac{1}{2}\int_0^T \lambda_s^2 ds\right)\Big] < \infty. \tag{9}$$

Under our model assumptions, this condition is always satisfied.[1]

To calibrate the model we use the transformed price process to avoid complex numerical calculations in the calibration procedure. The bivariate model can be reformulated as

$$d\xi_t = \left(\frac{1}{2}z_t\xi_t + \sqrt{z_t}\lambda_t\right)dt + \sqrt{z_t}dW_t^1, \tag{10}$$

$$d\lambda_t = \theta(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda dW_t^2, \tag{11}$$

$$dW_t^1 dW_t^2 = \rho dt. \tag{12}$$

In (11), $\bar{\lambda}$ represents the long-term mean value. θ denotes the rate with which the shocks dissipate and the variable reverts towards the mean. σ_λ is the volatility of the market price of risk. According to Carmona and Hinz [2], the price transformation is conditional Gaussian and its SDE can be solved explicitly.

3.2 Calibration to Historical Data

We consider the discretization of the model (10)–(12). By assuming the constant volatility terms in the time interval $[t_{k-1}, t_k]$, the model equations can be discretized under Euler's scheme given by

$$\xi_{t_k} = \sqrt{z_{t_{k-1}}}\triangle t\lambda_{t_{k-1}} + \left(1 + \frac{1}{2}z_{t_{k-1}}\right)\xi_{t_{k-1}} + \sqrt{z_{t_{k-1}}\triangle t}\mathscr{E}_{t_k}^1, \tag{13}$$

$$\lambda_{t_k} = (1 - \theta\triangle t)\lambda_{t_{k-1}} + \theta\bar{\lambda}\triangle t + \sigma_\lambda\sqrt{\triangle t}\mathscr{E}_{t_k}^2, \tag{14}$$

$$Cov(\mathscr{E}_{t_k}^1, \mathscr{E}_{t_k}^2) = \rho, \tag{15}$$

where $\triangle t = t_k - (t_{k-1})$, namely the time interval, and $\mathscr{E}_{t_k}^1, \mathscr{E}_{t_k}^2 \sim \mathscr{N}(0, 1)$. z_{t_k} can be modeled by using the function $\beta(T - t_k)^{-\alpha}$. The model parameter-set is therefore $\psi = [\theta, \bar{\lambda}, \sigma_\lambda, \rho, \alpha, \beta]$.

As λ_{t_k} is a hidden state variable related to the price transformation, and only values of ξ_{t_k} at time points $t_1, t_2, \ldots, t_n$ can be determined from the market observations, the market price of risk series can be estimated by applying the Kalman filter algorithm. We have chosen to use the transformation process instead of the normalized price a_{t_k}. Because of the linear form of Eqs. (13) and (14) the standard Kalman filter algorithm is considered to be an efficient method for the model calibration. A detailed procedure to apply the standard Kalman filter can be found in [7]. To apply the Kalman filter model Eqs. (13)–(15) must be put into the state space representation to fit the model framework. The measurement equation links the unobservable state to observations. It can be derived from (13) and (14). After some manipulations, the equations of the

[1] A proof can be found in Appendix 1.

state space form for the model can be rewritten[2] as

$$S_{t_k} = \sqrt{\beta(T-t_k)^{-\alpha}}\Delta t \lambda_{t_k} + \left(1 + \frac{1}{2}(\beta(T-t_k)^{-\alpha})\right)\xi_{t_k} + \sqrt{\beta(T-t_k)^{-\alpha}\Delta t}\bar{\mathscr{E}}^1_{t_k}, \tag{16}$$

and

$$\begin{aligned}\lambda_{t_k} = &\left[\theta\bar{\lambda}\Delta t - \frac{\sigma_\lambda \rho}{\sqrt{\beta(T-t_k)^{-\alpha}}}\left(\left(1+\frac{1}{2}(\beta(T-t_k)^{-\alpha})\right)\xi_{t_{k-1}} - \xi_{t_k}\right)\right]\\ &+(1-\theta\Delta t - \sigma_\lambda\rho\Delta t)\lambda_{t_{k-1}} + \sigma_\lambda\sqrt{(1-\rho^2)\Delta t}\bar{\mathscr{E}}^2_{t_k},\end{aligned} \tag{17}$$

where $\bar{\mathscr{E}}^1_{t_k}$ and $\bar{\mathscr{E}}^2_{t_k}$ are independent, standard normally distributed random variables.

For the estimation of the parameter vector $\psi = [\theta, \bar{\lambda}, \sigma_\lambda, \rho, \alpha, \beta]$ consider the variable ξ_{t_k}. In each iteration of the filtering procedure, the conditional mean $\mathbb{E}[\xi_{t_k}|\xi_{t_1}, \ldots, \xi_{t_{k-1}}]$ and the conditional variance $Var(\xi_{t_k}|\xi_{t_1}, \ldots, \xi_{t_{k-1}})$ can be calculated. We denote mean and variance by $\mu_{t_k}(\psi)$ and $\Sigma_{t_k}(\psi)$, respectively. The joint probability density function of the observations is denoted by $f(\xi_{t_{1:n}}|\psi)$ and is given by

$$f(\xi_{t_{1:n}}|\psi) = \prod_{k=1}^{n}\frac{1}{\sqrt{2\pi\Sigma_{t_k}(\psi)}}\exp\left(-\frac{(\xi_{t_k}-\mu_{t_k}(\psi))^2}{2\Sigma_{t_k}(\psi)}\right),$$

where $\xi_{t_{1:n}}$ summarize the observations from ξ_{t_1} to ξ_{t_n}. Its corresponding log-likelihood function is given by

$$\mathscr{L}_{obs}(\psi|\xi_{t_{1:n}}) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\sum_{k=1}^{n}\log\Sigma_{t_k}(\psi) - \frac{1}{2}\sum_{k=1}^{n}\frac{(\xi_{t_k}-\mu_{t_k}(\psi))^2}{\Sigma_{t_k}(\psi)}. \tag{18}$$

The estimation results, their standard errors, t-tests and p-values can be found in Table 2. Figure 7 shows the estimation results of the market price of risk, compared with the price transformation and the historical futures price. In Fig. 8, a negative correlation between the price transformation and the market price of risk can be seen. The market price of risk is the return in excess of the risk-free rate that the market wants as compensation for taking the risk.[3] It is a measure of the extra required rate of return, or say, a risk premium, that investors need for taking the risk. The more risky an investment is, the higher the additional expected rate of return should be. So in order to achieve a higher required rate of return, the asset must be discounted and thus will be sold at a lower price. Figure 8 reveals this inverse relationship.

Moreover, we use the mean pricing errors (MPE) and the root mean squared errors (RMSE) given by

---

[2]For a derivation of the state equation see Appendix 2.

[3]For an economical explanation see [11], Chap. 27. or [17], Chap. 30.

Table 2 Test of model parameters at significance level 5 %, sample size 1536

Parameter	Coeff	Std Err	t-test	p-value
θ	1.5130	0.3195	5.7601	0.0000
$\bar{\lambda}$	0.4091	0.6117	4.0641	0.0001
σ_λ	0.2913	0.0193	17.6365	0.0000
ρ	0.0017	0.0016	9.0910	0.0000
α	-1.5772	0.0256	61.5603	0.0000
β	0.0172	0.0005	35.6312	0.0000

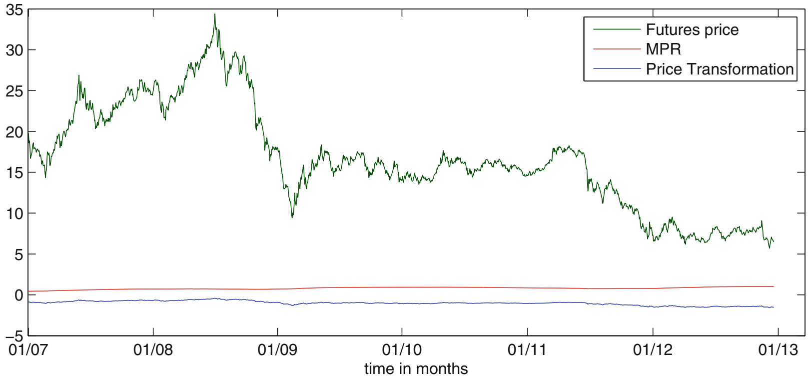

Fig. 7 MPR, futures price and price transformation from Jan. 2007 to Dec. 2012

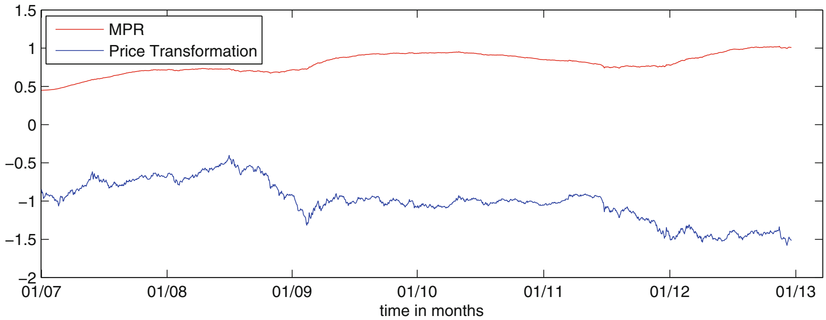

Fig. 8 Negative correlation of MPR and price transformation

$$MPE = \frac{1}{N}\sum_{t_i=1}^{N}(\bar{\xi}_{t_i,\tau} - \xi_{t_i,\tau}),$$

$$RMSE = \left(\frac{1}{N}\sum_{t_i=1}^{N}(\bar{\xi}_{t_i,\tau} - \xi_{t_i,\tau})^2\right)^{\frac{1}{2}},$$

Table 3 Performance of MPE and RMSE with 2000 observations

Maturity	MPE	RMSE
1 month	−0.0153	0.0182
3 months	−0.0208	0.0234
6 months	−0.0366	0.0397
9 months	−0.1273	0.1302

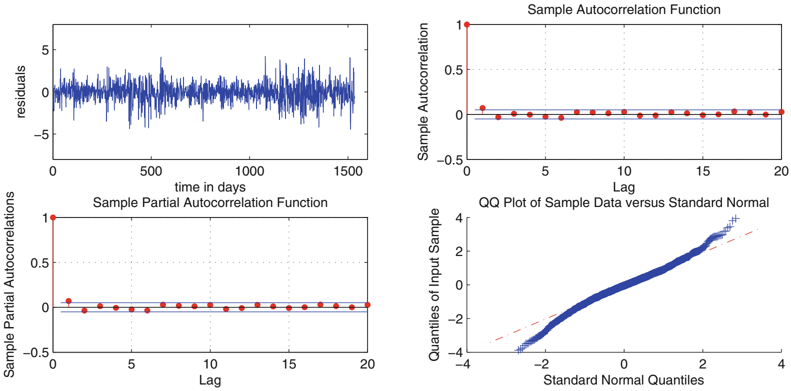

Fig. 9 Statistical tests for the residuals in trading phase 2

respectively, to assess the quality of model fit. Here N denotes the number of observations, $\bar{\xi}_{t_i,\tau}$ is the estimated price to maturity τ, and $\xi_{t_i,\tau}$ is the observed price. Their values can be seen in Table 3. The absolute values of MPE and RMSE increase with time but still remain very low even 9 months before the maturity. Therefore, the conclusion is that the model is able to reproduce the price dynamics.

In Fig. 9 we show the standard statistical test results of the residuals by taking into account the dynamic market price of risk. Comparing with the results from Figs. 2, 3, 4, 5 and 6, the time series of the residuals is relative stable with smaller variance. The sample auto-correlations and sample partial auto-correlations reveal very weak linear dependence of the variables at different time points. Also, the Q-Q plot indicates a better fit of a Gaussian distribution.

4 Option Pricing and Market Forward Looking Information

A general pricing formula of a European call is given by

$$C_t = e^{-\int_t^\tau r_s ds}\mathbb{E}^{\mathbb{Q}}[(A_\tau - K)^+|\mathscr{F}_t],$$

where $\{r_s\}_{s\in[0,T]}$ stands for a deterministic rate, A_t denotes the futures price, $K \geq 0$ is the strike price, and $\tau \in [0, T]$ is the maturity. The normalized price process a_t is given by $a_t = A_t/P$, where P denotes the penalty for each ton of exceeding emissions, and therefore we have $A_t = P\Phi(\xi_t)$. A call option price formula written on EUA has been derived by Carmona and Hinz [2] under the assumption of a constant market price of risk. Under the assumption of a dynamic market price of risk, the option price formula is coherent with the formula in [2] given by

$$C_t = e^{-\int_t^\tau r_s ds} \int_{\mathbb{R}} (P\Phi(x) - K)^+ \varphi(\mu_{t,\tau}, \sigma_{t,\tau}^2) dx,$$

where φ stands for the density function of a standard normal distribution. Here $\mu_{t,\tau}$ and $\sigma_{t,\tau}^2$ are the parameters of the distribution of ξ_t, which is conditional Gaussian. Under the risk neutral measure $\mathbb{Q}$, $\mu_{t,\tau}$ and $\sigma_{t,\tau}^2$ are given by

$$\mu_{t,\tau} = e^{\frac{1}{2}\int_t^\tau z_s ds} \xi_t, \qquad \sigma_{t,\tau}^2 = \int_t^\tau z_s e^{\int_s^\tau z_u du} ds.$$

In the following example, the penalty level is $P = 100$, the initial time $t = 0$ starts in April 2005. EUA futures has maturity T on the last trading day in 2012. The European calls written on EUA futures with a strike at $K = 15$ and maturity T will be considered under a constant interest rate at $r = 0.05$. Figure 10 shows the call option prices and the futures prices. The red curve stands for the option prices under dynamic MPR while the green curve stands for the option prices under constant MPR.

To measure the impact of the dynamic market price of risk on the EUA option for different strikes we calculate the option price in the univariate and bivariate model setting respectively. Durations from 1, 3, 6 and 12 months to maturity are chosen for calls written on EUA 2012. The results are plotted in Fig. 11. The red curve stands for the option prices evaluated by the bivariate model and the blue curve by the univariate

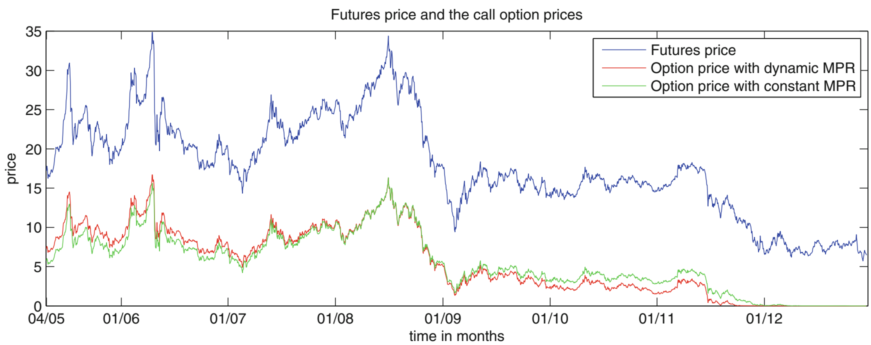

Fig. 10 Futures price and call option prices with $K = 15$ from 2005 to 2012

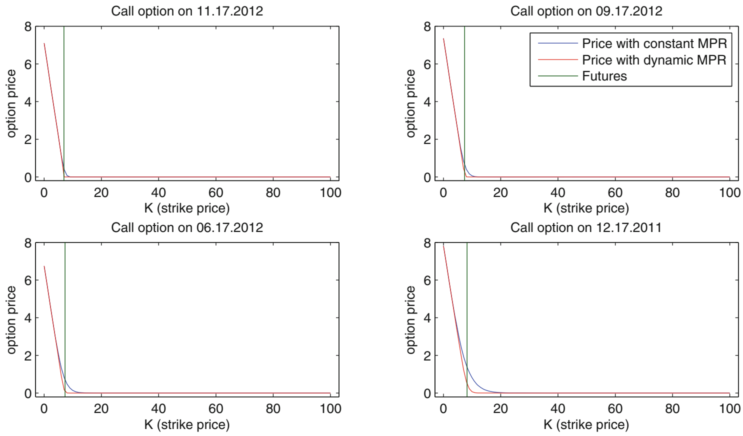

Fig. 11 Call option prices comparison for durations of 1, 3, 6, 12 months on EUA 2012 for different strikes

model. The green line is the corresponding futures price at the given time. In most cases, one is interested in the option prices near the underlying price. According to the figure, the option prices in different model settings coincide except for a interval around the corresponding futures. In a short time before the maturity of EUA 2012, Fig. 11 shows a price overestimation by the constant MPR. This result is consistent with the result shown in Fig. 10, where we take $K = 15$ as a sample path.

Moreover, one notes that the call price process with constant MPR develops below the call process with dynamic MPR in the first trading phase before 2008 and then increases slowly and moves to the upside of the call process with dynamic MPR during the second trading phase, before both processes vanish to the maturity because of lower underlying prices. The reason for the price underestimation before 2008 and overestimation thereafter can be explained as the assumption of a constant MPR in the whole trading periods and thus causes a neglect on the information of the market participants. Due to the regulatory framework of the carbon market, certificates carry information on the market participant expectations on the development of the fundamental price drivers. Since the implied risk premia increase with time and exceed their 'average' level in 2008, asset price must be discounted to compensate the higher risk. By using appropriate valuation models, this risk premia and the forward-looking information carried by prices of futures and options of certificates can be extracted.

5 Conclusion

We extract forward-looking information in the EU ETS by applying an extended pricing model of EUA futures and analyzing its impact on option prices. We find that the implied risk premium is time-varying and has to be modeled by a stochastic process. Using the information given by the risk premium we show that the option prices during the first and second trading phases are underestimated and overestimated, respectively. The reason for the pricing deviation is caused by the assumption of a constant market price of risk which rigidifies the market participant expectations on the development of price drivers. The over- and underestimated prices are mostly concentrated in the interval including the corresponding futures, which is the area where the price most likely will evolve in the future. Although there is not a closed form for the option pricing formula, a simple numerical approach can be used to determine the price.

Appendix 1

In order to show the condition in (8), it is sufficient to prove the Novikov's condition given by

$$\mathbb{E}\Big[\exp\Big(\frac{1}{2}\int_0^T \lambda_s^2 ds\Big)\Big] < \infty.$$

In the bivariate EUA pricing model, where λ_t follows a Ornstein-Uhlenbeck-Process given by

$$\lambda_t = \theta(\bar{\lambda} - \lambda_t)dt + \sigma_\lambda dWt,$$

this condition is always satisfied.

Proof We first show that there exists a constant $\varepsilon > 0$ such that for any $S \in [0, T]$, we have

$$\mathbb{E}\Big[\exp\Big(\frac{1}{2}\int_S^{S+\varepsilon} \lambda_t^2 dt\Big)\Big] < \infty. \tag{19}$$

To show (19) we consider the term in the expectation notation. By applying Jensen's inequality we have

$$\exp\Big(\frac{1}{2}\int_S^{S+\varepsilon} \lambda_t^2 dt\Big) = \exp\Big(\int_S^{S+\varepsilon} \frac{1}{\varepsilon}\frac{\varepsilon}{2}\lambda_t^2 dt\Big)$$
$$= \exp\Big(\frac{1}{\varepsilon}\int_S^{S+\varepsilon} \frac{\varepsilon}{2}\lambda_t^2 dt\Big) \le \frac{1}{\varepsilon}\int_S^{S+\varepsilon} \exp\Big(\frac{\varepsilon}{2}\lambda_t^2\Big)dt.$$

By applying Fubini's theorem (19) becomes

$$\frac{1}{\varepsilon}\int_S^{S+\varepsilon} \mathbb{E}\Big[\exp\Big(\frac{\varepsilon}{2}\lambda_t^2\Big)\Big]dt. \tag{20}$$

The process λ_t is a Gaussian process with mean and variance given by

$$\mathbb{E}[\lambda_t] = \mu_t = \lambda_0 e^{-\theta t} + \bar{\lambda}(1 - e^{-\theta t}),$$
$$\mathrm{Var}(\lambda_t) = \sigma_t^2 = \frac{\sigma_\lambda^2}{2\theta}(1 - e^{-2\theta t}).$$

We have $\lambda_t \sim \mathcal{N}(\mu_t, \sigma_t^2)$. Now let Z be a standard normal-distributed random variable $Z \sim \mathcal{N}(0, 1)$. So in (20) we have

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\frac{\varepsilon}{2}\lambda_t^2\Big)\Big] &= \mathbb{E}\Big[\exp\Big(\frac{\varepsilon}{2}(\mu_t + \sigma_t Z)^2\Big)\Big] \\
&= \mathbb{E}\Big[\exp\Big(\frac{\varepsilon\mu_t^2}{2} + \varepsilon\mu_t\sigma_t Z + \frac{\varepsilon\sigma_t^2 Z^2}{2}\Big)\Big] \\
&= \int_{\mathbb{R}} \exp\Big(\frac{\varepsilon\mu_t^2}{2} + \varepsilon\mu_t\sigma_t x + \frac{\varepsilon\sigma_t^2 x^2}{2}\Big)\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{x^2}{2}\Big)dx \\
&= \exp\Big(\frac{\varepsilon\mu_t^2}{2}\Big)\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1-\varepsilon\sigma_t^2}{2}x^2 + \varepsilon\mu_t\sigma_t x\Big)dx.
\end{aligned}$$

To calculate the integration term above, let $a_t = 1 - \varepsilon\sigma_t^2$ and $b_t = \varepsilon\mu_t\sigma_t$, and make the integral-substitution. Then we have

$$\begin{aligned}
&\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1-\varepsilon\sigma_t^2}{2}x^2 + \varepsilon\mu_t\sigma_t x\Big)dx \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}(a_t x^2 - 2b_t x)\Big)dx \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\Big(y^2 - 2b_t\frac{1}{\sqrt{a_t}}y\Big)\Big)\frac{1}{\sqrt{a_t}}dy \\
&= \frac{1}{\sqrt{a_t}}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{1}{2}\Big(y^2 - \frac{2b_t}{\sqrt{a_t}}y + \Big(\frac{b_t}{\sqrt{a_t}}\Big)^2 - \Big(\frac{b_t}{\sqrt{a_t}}\Big)^2\Big)\Big)dy \\
&= \frac{1}{\sqrt{a_t}}\exp\Big(\frac{b_t^2}{2a_t}\Big)\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{(y - \frac{2b_t}{\sqrt{a_t}})^2}{2}\Big)dy \\
&= \frac{1}{\sqrt{a_t}}\exp\Big(\frac{b_t^2}{2a_t}\Big).
\end{aligned}$$

According to the assumptions $a_t = 1 - \varepsilon\sigma_t^2$ is positive and the expectation is convergent for a small ε and its value is

$$\begin{aligned}
\mathbb{E}\Big[\exp\Big(\frac{\varepsilon}{2}\lambda_t^2\Big)\Big] &= \frac{1}{\sqrt{a_t}}\exp\Big(\frac{\varepsilon\mu_t^2}{2}\Big)\exp\Big(\frac{b_t^2}{2a_t}\Big) \\
&= \frac{1}{\sqrt{1-\varepsilon\sigma_t^2}}\exp\Big(\frac{\varepsilon\mu_t^2}{2}\Big)\exp\Big(\frac{\varepsilon^2\mu_t^2\sigma_t^2}{2-2\varepsilon\sigma_t^2}\Big).
\end{aligned}$$

Thus the integral in (20) is finite and the exponential term in (19) is integrable.

In order to show Z_t is a martingale we first consider that Z_t is a local martingale, hence it is a supermartingale. Therefore, Z_t is a martingale if and only if the condition $\mathbb{E}[Z_t] = 1$, $\forall t \in [0, T]$, is satisfied. This martingale property can be shown by induction. Suppose $\mathbb{E}[Z_0] = 1$ which is trivial and $\mathbb{E}[Z_t] = 1$ for $t \in [0, S]$ for $S < T$. Let now $t \in [S, S + \varepsilon]$ and set

$$Z_S^t = \exp\Big(\int_S^t \lambda_s dW_s - \frac{1}{2}\int_S^t \lambda_s^2 ds\Big).$$

According to Novikov condition and (19), Z_S^t is a martingale. Then we have

$$\mathbb{E}[Z_t] = \mathbb{E}[Z_S Z_S^t] = \mathbb{E}[\mathbb{E}[Z_S Z_S^t]|\mathscr{F}_S] = \mathbb{E}[Z_S \mathbb{E}[Z_S^t|\mathscr{F}_S]] = \mathbb{E}[Z_S Z_S^S] = \mathbb{E}[Z_S],$$

since

$$Z_S^S = \exp\Big(\int_S^S \lambda_s dW_s - \frac{1}{2}\int_S^S \lambda_s^2 ds\Big) = \exp(0) = 1.$$

It follows

$$\mathbb{E}[Z_t] = \mathbb{E}[Z_S] = 1 \quad \text{for } t \in [S, S + \varepsilon].$$

Then we have $\mathbb{E}[Z_t] = 1$ for $t \in [0, S + \varepsilon]$. Repeat this induction for $\frac{T-S}{\varepsilon}$ times we have $\mathbb{E}[Z_t] = 1$ for $t \in [0, T]$, which implies Z_t defined in (8) is a martingale. □

Appendix 2

The bivariate EUA pricing model can be described as follows:

$$\begin{aligned}
\xi_{t_k} &= \sqrt{z_{t_{k-1}}}\triangle t \lambda_{t_{k-1}} + \left(1 + \frac{1}{2} z_{t_{k-1}}\right)\xi_{t_{k-1}} + \sqrt{z_{t_{k-1}}\triangle t}\mathscr{E}_{t_k}^1,\\
\lambda_{t_k} &= (1 - \theta\triangle t)\lambda_{t_{k-1}} + \theta\bar{\lambda}\triangle t + \sigma_\lambda\sqrt{\triangle t}\mathscr{E}_{t_k}^2,\\
Cov(\mathscr{E}_{t_k}^1, \mathscr{E}_{t_k}^2) &= \rho,
\end{aligned}$$

where $\mathscr{E}_{t_k}^1$ and $\mathscr{E}_{t_k}^2$ are both random variables of the standard normal distribution. We want to put the model into the state space form. Price transformation depends on the current level of the market price of risk, which is an unobservable variable and therefore must be modeled in the equation of λ_{t_k}. We first let

$$\mathscr{E}_{t_k}^1 = \bar{\mathscr{E}}_{t_k}^1, \qquad \mathscr{E}_{t_k}^2 = \sqrt{1-\rho^2}\bar{\mathscr{E}}_{t_k}^2 + \rho\bar{\mathscr{E}}_{t_k}^1,$$

where $\bar{\mathscr{E}}^1_{t_k}$ and $\bar{\mathscr{E}}^2_{t_k}$ are both random variables of the standard normal distribution as well. This fact can be easily seen since we have

$$Cov(\bar{\mathscr{E}}^1_{t_k}, \bar{\mathscr{E}}^2_{t_k}) = Cov\left(\mathscr{E}^1_{t_k}, \frac{\mathscr{E}^2_{t_k} - \rho\mathscr{E}^1_{t_k}}{\sqrt{1-\rho^2}}\right) = Cov\left(\mathscr{E}^1_{t_k}, \frac{\mathscr{E}^2_{t-k}}{\sqrt{1-\rho^2}}\right) + Cov\left(\mathscr{E}^1_{t_k}, -\frac{\rho\mathscr{E}^1_{t_k}}{\sqrt{1-\rho^2}}\right) = 0.$$

Note that

$$\sqrt{z_{t_{k-1}}}\Delta t\lambda_{t_{k-1}} + \left(1 + \frac{1}{2}z_{t_{k-1}}\right)\xi_{t_{k-1}} + \sqrt{z_{t_{k-1}}\Delta t}\mathscr{E}^1_{t_k} - \xi_{t_k} = 0.$$

Multiplying $-\sigma_\lambda \rho (z_{t_{k-1}})^{-\frac{1}{2}}$ at the both sides of the equation and sum it to the equation of λ_{t_k}, it follows that

$$\begin{aligned}
\lambda_{t_k} &= (1 - \theta\Delta t)\lambda_{t_{k-1}} - \sigma_\lambda\rho\Delta t\lambda_{t_{k-1}} + \theta\bar{\lambda}\Delta t, \\
&\quad -\frac{\sigma_\lambda\rho}{\sqrt{z_{t_{k-1}}}}\left(\left(1 + \frac{1}{2}z_{t_{k-1}}\right)\xi_{t_{k-1}} - \xi_{t_k}\right) + \sigma_\lambda\sqrt{\Delta t}\mathscr{E}^2_{t_k} - \sigma_\lambda\sqrt{\Delta t}\rho\mathscr{E}^1_{t_k} \\
&= (1 - \theta\Delta t - \sigma_\lambda\rho\Delta t)\lambda_{t_{k-1}} + \left[\theta\bar{\lambda}\Delta t - \frac{\sigma_\lambda\rho}{\sqrt{z_{t_{k-1}}}}\left(\left(1 + \frac{1}{2}z_{t_{k-1}}\right)\xi_{t_{k-1}} - \xi_{t_k}\right)\right] \\
&\quad + \sigma_\lambda\sqrt{\Delta t}\sqrt{1-\rho^2}\mathscr{E}^2_{t_k}.
\end{aligned}$$

This is the transition equation in the state space form, and the measurement equation would be

$$S_{t_k} = \xi_{t_{k+1}} = \sqrt{z_{t_k}}\Delta t\lambda_{t_k} + \left(1 + \frac{1}{2}z_{t_k}\right)\xi_{t_k} + \sqrt{z_{t_k}\Delta t}\mathscr{E}^1_{t_k}.$$

References

1. Benz, E., Trück, S.: Modeling the price dynamics of CO_2 emission allowances. Energy Econ. **31**, 4–15 (2009)
2. Carmona, R., Hinz, J.: Risk-neutral models for emission allowance prices and option valuation. Manag. Sci. **57**, 1453–1468 (2011)
3. Carmona, R., Fehr, M., Hinz, J.: Optimal stochastic control and carbon price formation. SIAM J. Control Optim. **48**, 2168–2190 (2009)
4. Carmona, R., Fehr, M., Hinz, J., Porchet, A.: Market design for emission trading schemes. SIAM Rev. **52**, 403–452 (2010)
5. Grüll, G., Kiesel, R.: Quantifying the CO_2 permit price sensitivity. Z. Energiewirtsch **36**, 101–111 (2012)
6. Grüll, G., Taschini, L.: A Comparison of reduced-form permit price models and their empirical performances. LSE Research Online (2010). http://eprints.lse.ac.uk/37603/ of subordinate document. Accessed 30 May 2015
7. Harvey, A.C.: Forecasting. Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge (1989)

8. Hinz, J.: Quantitative modeling of emission markets. Jahresber. Dtsch. Mathematiker-Ver. **112**, 195–216 (2010)
9. Hitzemann, S., Uhrig-Homburg M.: Empirical performance of reduced-form models for emission permit prices. Social Science Research Network (2013). http://ssrn.com/abstract=2297121 of subordinate document. Accessed 30 May 2015
10. Hitzemann, S., Uhrig-Homburg, M.: Equilibrium price dynamics of emission permits. Social Science Research Network (2014). http://ssrn.com/abstract=1763182 of subordinate document. Accessed 30 May 2015
11. Hull, J.C.: Options, Futures, and Other Derivatives. Prentice Hall, New Jersey (2009)
12. Paolella, M.S., Taschini, L.: An econometric analysis of emission trading allowances. J. Bank. Finance **32**, 2022–2032 (2008)
13. Rubin, J.: A model of intertemporal emission trading, banking and borrowing. J. Environ. Econ. Manage. **31**(3), 269–286 (1996)
14. Seifert, J., Uhrig-Homburg, M., Wagner, M.: Dynamic behavior of CO_2 spot prices. J. Environ. Econ. Manage. **56**, 180–194 (2008)
15. Taschini, L.: Environmental economics and modeling marketable permits: a survey. Asian Pac. Finan. Markets **17**(4) (2010)
16. Wagner, M.: Emissionszertifikate, Preismodellierung und Derivatebewertung. Dissertation, Universität Karlsruhe (2006)
17. Wilmott, P.: Paul Wilmott on Quantitative Finance. Wiley, Chichester (2006)