ASA 2021 Statistics and Information Systems for Policy Evaluation

BOOK OF SHORT PAPERS
of the opening conference

edited by
Bruno Bertaccini
Luigi Fabbris
Alessandra Petrucci

ASA 2021 Statistics and Information Systems for Policy Evaluation

BOOK OF SHORT PAPERS
of the opening conference

edited by
Bruno Bertaccini
Luigi Fabbris
Alessandra Petrucci

Firenze University Press

ASA 2021 Statistics and Information Systems for Policy Evaluation

BOOK OF SHORT PAPERS
of the opening conference

edited by

Bruno Bertaccini
Luigi Fabbris
Alessandra Petrucci

Cover graphic design: Lettera Meccanica SRLs

ASA 2021 Opening Conference on
STATISTICS AND INFORMATION SYSTEMS
FOR POLICY EVALUATION

University of Florence,   February 19, 2021
Bruno Bertaccini, Luigi Fabbris and Alessandra Petrucci
(Editors)

Partners

Table of contents

Preface

The Association for Applied Statistics (ASA) and the Department of Statistics, Computer Science, Applications DiSIA "*Giuseppe Parenti*" of the University of Florence, jointly with the partners AICQ (Italian Association for Quality Culture), AICQ-CN (Italian Association for Quality Culture North and Centre of Italy), AISS (Italian Academy for Six Sigma), ASSIRM (Italian Association for Marketing, Social and Opinion Research), Comune di Firenze (the Florence Municipality), SIS (the Italian Statistical Society), Regione Toscana (the Tuscany Region) and Valmon – Evaluation & Monitoring srl, have organised a scientific conference titled "*Statistics and Information Systems for Policy Evaluation*", aimed at promoting new statistical methods and applications for the evaluation of policies.

Due to the health emergency caused by the COVID-19 pandemic, the Scientific and Local Organizing Committees decided to reschedule the conference appointment in two different scientific events: an on-line Opening Conference held in February and March 2021 and a postponed on-site Conference in September 2021.

This Book includes 25 peer-reviewed short papers submitted to the Scientific Opening Conference. This event was organized in 4 on-line sessions every Friday, from February 19th, 2021 to March 12th, 2021; each session, which collected works on homogeneous issues ("Evaluation Of Educational Systems", "Innovation, Productivity and Welfare", "Health and Well-Being", "Tourism and Gastronomy"), lasted about one and half hour and was led by a Chair and one or more discussants. The papers published in this book are organized in those sessions and follow the conference program order.

On behalf of the Scientific Program Committee, we would like to thank the authors for submitting and presenting their interesting and inspiring works in the context of the evaluation of policies, the partners, the chairs, the discussants and the Local Organizing Committee. Finally, we are thankful to the members of the Scientific Committee for helping with the peer-reviewing process.

Florence (Italy), March 2021

Bruno Bertaccini, Luigi Fabbris, Alessandra Petrucci

Evaluation of educational systems

Does an entrepreneurial spirit animate fresh graduates in their work-seeking during uncertain times?

Luigi Fabbris, Manuela Scioni

1. Introduction

In this paper we examine the entrepreneurial intention of fresh graduates as a probabilistic antecedent of their propensity to create new venture, develop new business concept or behave entrepreneurially within an existing firm. The latter type of propensity, that some scholars name "intrapreneurship" (Krueger and Brazeal, 1994), refers to a proactive attitude that should drive workers' activities irrespective of workplace.

Self-employment is socially relevant because it is world-wide considered a way to improve employment at all levels and, in particular, regarding youth (Duell, 2018). In the EU, in 2017 (European Commission, 2018), self-employment without employees accounted 9.8% of total employment and another 3.9% of self-employment with employees. In Italy, the self-employed accounted 21.9% of the total employment. The problem with self-employment is that, on average, income and job satisfaction of the self-employed are lower than that of employees (Eurofound, 2015). The economic difficulties added by Covid-19 restrictions worsened even graduates' employability. Even though the full effects of pandemic on youth unemployment are yet to be detected, the graduates' transition to work remains a major concern. Besides support of public bodies–which should be addressed in particular to weaker job-seeking categories–it is claimed that graduates become more entrepreneurial (OECD / European Union, 2017). Only so, self-employment can be no longer a necessity but an ambition.

The rest of the paper is organised as follows: Section 2 presents the working hypotheses and the analytical model and Section 3 the main results of data analysis. Section 4 includes the discussion of results and final conclusions.

2. Data, models and methods

Our data refer to graduates from Padua University, the largest university of the Veneto district, Italy. People who graduated from March to September 2020 were sent an email through which they could activate an electronic questionnaire. This survey system allowed to check who responded and send them targeted reminders. The final sample, after the exclusion of medicine students, was composed of 1603 graduates.

The relational model adopted for data analysis refers to the theory of planned behaviour, as proposed in Ajzen (1991). This psychological theory plunges its roots on the hypothesis that one's behavioural intention depends both on individuals' cognitive and non-cognitive traits and their familial and social culture and norms.

A graduate's entrepreneurial intention was estimated by detecting any action related to entrepreneurial purposes he or she has put into practice while searching for a job, irrespective if he or she already had a job. With these data even a dichotomous variable was created (Y=1: at least one action; Y=0: no action). The possible predictors of graduates' entrepreneurial intention were classified in blocks, or factors, termed as follows (see also Figure 1).

a) *Human capital* (X1), including both knowledge, say the cognitive and mental structures determining how people perceive and integrate new information, and practical intelligence, say doing skills. The analysed variables were: attended

discipline, degree level, final mark, and internship and/or international experience of graduates.

b) *Social capital* (X2), namely the personal and familial relations useful to set an initiative. The analysed variables were: having a sociable personality, having attended a lyceum high school and currently attending social activities (volunteering, sports, music).

c) *Psychological capital* (PsyCap – X3), namely one's positive disposition capable of providing graduates with competitive advantage, including (Robusto et al., 2019): *self-efficacy*, that is having confidence to take on and put forth the effort necessary to succeed at challenging tasks; *optimism*, that is making a positive attribution to succeeding now and in the future; *hope*, that is persevering toward goals and, when necessary, redirecting paths to goals to succeed; and *resilience*, that is, when beset by problems and adversity, sustaining and improving to attain success. In addition, the graduates' *self-awareness*, that is the conscious understanding of their own capacity, and *extraversion*, which indicates how outgoing and social they are, were surveyed.

d) *Proactive attitudes toward labour and education* (X4A). In this work, expected income, autonomy, complexity, challenge and flexibility of the job tasks and roles were hypothesised to characterise the graduates animated by entrepreneurial disposition. The attitude toward education was indirectly measured with the availability to attend again the attended university course, he or she could go back in time.

e) *Risk propensity* (X4B), that is the extent to which graduates are willing to take a chance with respect to possible losses. The risk-propensity/courage factor was measured with five items: 1) If I feel it is relevant to me, I could compromise my relations with important persons; 2) For a valid cause, I could start a conflict in my workplace; 3) Not even an intense social pressure could refrain me from doing what I feel is the right thing to do; 4) I could expose myself to risks if I believe their outcome is important; 5) I am capable to catch sudden opportunities.

f) *Personal and social inadequacies* (X5A) *and LoC (*X5B*) that may either support or obstruct the pathways to entrepreneurship.* LoC is (Robusto et al., 2019) a generalized attitude, belief, or expectancy regarding the nature of the causal relationship between graduates' behaviour and its consequences. A person with a dominant *external* LoC tends to believe that what happens to him/her depends mainly from external forces, like fate or luck; conversely, persons with an *internal* LoC see future outcomes as being contingent on their own decisions and behaviour. Inadequacies and an external LoC may have a push effect while an internal LoC a pull (stress-reducing) effect upon entrepreneurial intention.

The hypotheses can be expressed with a system of equations: $Y=f(X_1, X_2, X_3, X_4, X_5 \mid Z)$ and $X_4=f(X_1, X_2, X_3 \mid Z)$, where: Y denotes the entrepreneurial intention; X_1 the human capital; X_2 the social capital; X_3 the psychological capital; X_4 the attitudes toward labour and education; X_5 the personal and social norms; and Z the control variables (gender: Z_1 and working at graduation: Z_2). The system of equations identifies a path analysis model, say, a model in which the relationships between sets of variables are hierarchical. In our case, the intention, Y, is influenced by one's capitals both directly and through closer-to-Y factors. Control variables are hypothesised to influence intentions only indirectly.

To process the data, we applied a PLS-SEM (Partial Least Squares–Structural Equation Modelling) model (Tenenhaus et al., 2005; Rigdon et al., 2017), a structural equation approach that fits a composite model in which more than one underlying factor is hypothesised. The data were processed with *semPLS* software (Monecke and Leisch, 2012). The software was applied both on the count of actions experienced for self-employment seeking and on a dichotomous Y. The following comments pertain to the dichotomous Y.

In PLS-SEM, let X_i ($i=1,...,k$) be a composite factor of p_i weighted factors v_{ij} ($j=1,...,p_i$), i.e., $X_i = \sum_{j=1}^{p_i} w_{ij} v_{ij} \ (i = 1, ..., k),$ where the w_{ij}'s are the weights to apply to each respective factor to obtain X_i. Each factor is a linear combination of observable variables. This implies that we are interested on the relationship between X_i and the factors and not with the observed variables.

The variance of the composite X_i is the sum of the components' variances plus twice the sum of their covariances, each adjusted by the weights:

$$\sigma^2(X_i) = \sum_{j=1}^{p_i} w_{ij}\, \sigma^2(v_{ij}) + 2 \sum_{j}^{p_i} \sum_{k>j}^{p_i} w_{ij} w_{ik} \sigma(\, v_{ij}\, v_{ik}),$$

where $\sigma^2(v_{ij})$ is the variance of v_i and $\sigma(v_{ij}\, v_{ik})$ ($i=1, ..., k$; $j \neq k=1, ..., p_i$) the covariance between indicators v_{ij} and v_{ik} of factor X_i. Random variance, being orthogonal, plays no part in the covariances.

3. Results

The responding graduates were prevalently females (61.1%) and resident in Italy (97.6%). Their activity was: studying (50.0%), realising an internship (6.5%), working (13.8%), seeking for a job (26.3%), or not doing any work- or study-related activity (3.4%). The latter category is usually confused with discouraged, or NEET, people, meaning that they do not do any activity because discouraged even to look for a job. This is not our case, since just 5.5% of these people did not receive any job offer and 87.3% was prepared to look for a job in the following twelve months. The others did not work either because of contingent reasons (disease, maternity) or because waiting to start a new job or the civil service. Definitely, the discouraged varied between 1.2 and 4.3 per thousand. In what follows, we will not analyse this uncertain category.

All disciplines were represented in our sample: hard sciences 6.5%, engineering 24.9%, life sciences 13.2%, social sciences 38.4% and the humanities 17.0%. First-level (Bachelors) numerically prevailed (60.4%) over second-level (Masters) graduates. PhDs were ignored in this research. Graduation marks ranging between 105 and 110 were 52.3% of total marks.
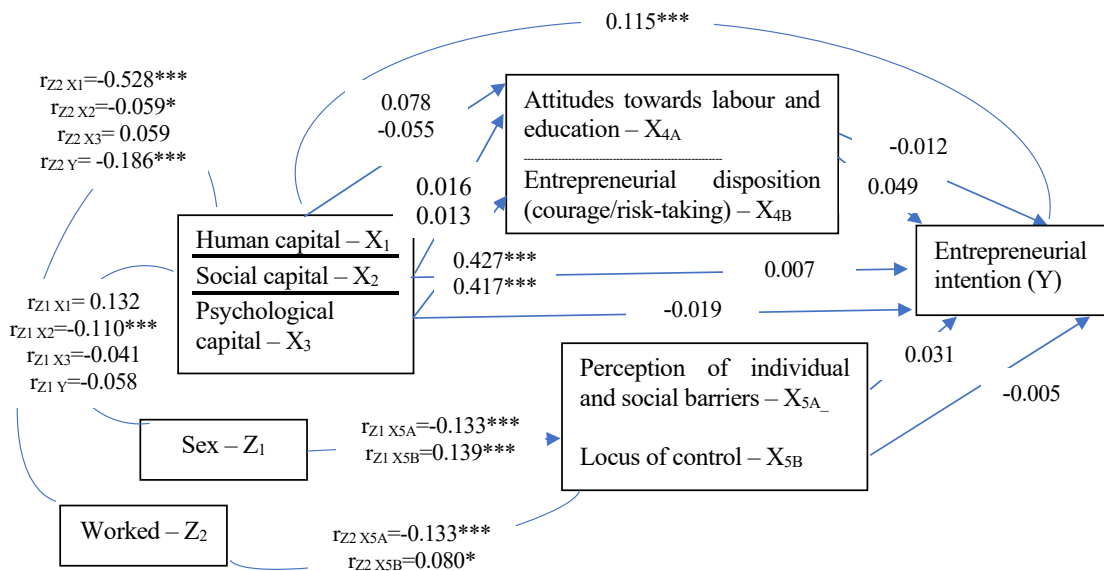
Figure 1. PLS-PM estimates of between factor relationships influencing entrepreneurial intention of fresh graduates *(Significance levels: *** <0.001; ** >0.01; * <0.05)*

Table 1. PLS-SEM estimates of the within-factor relationships (s.e. in brackets).

	% or mean	Estimates	s.e.
X_{11}: Academic discipline: Engineering	24.9	0.160	0.476
X_{11}: Academic discipline: Science (hard)	6.5	0.044	0.112
X_{11}: Academic discipline: Social science	38.4	-0.329	0.332
X_{11}: Academic discipline: Humanistic sciences	17.0	-0.001	0.236
X_{11}: Academic discipline: Life sciences	13.2	0.238***	0.077
X_{12}: High final mark	52.3	0.382*	0.184
X_{13}: Master degree	39.6	0.895***	0.111
X_{21}: Sociable personality	76.7	0.982***	0.013
X_{22}: Volunteering	17.0	0.153*	0.074
X_{23}: Music or chorus player	20.8	0.043*	0.086
X_{24}: Attended Lyceum high school	60.5	-0.248***	0.081
X_{31}: Self-efficacy scores (mean)	0.0	0.806***	0.012
X_{32}: Optimism (mean)	0.0	0.762***	0.015
X_{33}: Resilience (mean)	0.0	0.783***	0.014
X_{34}: Hope (mean)	0.0	0.526***	0.029
X_{35}: Self-awareness	67.3	0.700***	0.021
X_{36}: Extraversion	69.0	0.761***	0.014
X_{41}: Internship during studies	49.0	0.015	0.202
X_{42}: International-Erasmus mobility	20.1	0.043	0.155
X_{43}: Would attend same course	68.9	0.458***	0.086
X_{44}: Degree: less time for job finding	67.2	0.625***	0.164
X_{45}: Degree: professional career	73.6	0.744***	0.100
X_{46}: Degree: regard of peers, family	67.2	0.689***	0.052
X_{47}: Degree: strengthened self-regard	74.6	0.811***	0.051
X_{48}: Job: income relevance	49.9	-0.084	0.105
X_{49}: Job: variety of activities	20.3	0.018	0.064
X_{410}: Job: wide and complex roles	10.0	0.032	0.068
X_{411}: Job: challenging roles	24.7	0.096	0.060
X_{412}: Job: flexible schedule	9.7	-0.095	0.056
X_{413}: Job: autonomy	21.3	-0.057	0.073
X_{414}: Job: high quality outcomes	32.1	0.063	0.117
X_{415}: Courage scores (mean)	0.0	0.875***	0.071
X_{416}: Degree: chances for own business	51.8	0.548***	0.094
X_{51}: Economy unfavourable to youth	60.0	0.769***	0.022
X_{52}: Limited job offers	58.3	0.762***	0.022
X_{53}: Jobs just for friends and "wise guys"	46.4	0.622***	0.033
X_{54}: Italy is not a country for youth	47.2	0.692***	0.030
X_{55}: Youngsters high expectations	51.2	0.196***	0.066
X_{56}: Youngsters low adaptability	47.1	0.135	0.072
X_{57}: Inadequate competencies	57.2	0.540***	0.041
X_{58}: Employers just for profit	60.1	0.742***	0.026
X_{59}: Job seeking is not supported	57.3	0.681***	0.030
X_{510}: Too few insertion programs	61.6	0.709***	0.027
X_{511}: Platforms inadequate for search	49.0	0.682***	0.030
X_{512}: Internal LoC (mean, scores)	0.0	0.993***	0.157
X_{513}: External LoC (mean, scores)	0.0	-0.091	0.275
R^2 (all factors with Y)	=	*0.077*	
Average within-factor R^2	=	*0.140*	

The inclination rate for fresh graduates to start an own business is 10%. So, the entrepreneurial spirit animates a minority of highly educated people, with large differences in the number of entrepreneurial actions undertaken by those who continued studying (just 2.5%) and those who already had a job (12.1%) or were searching for it (26.4%).

We applied a PLS-SEM model including all respondents. The results of the within factor regression analyses are presented in Table 1 and outlined in Figure 1. The structural model explains 7.7% variance of fresh graduates' entrepreneurial intention. The analysis rejected most relationships hypothesised in the theoretical model; only a direct relationship of human capital and a relationship of psychological capital mediated through the risk-taking factor were confirmed. Instead, the within factor relationships are much stronger than those ascertained between the factors and the intention: indeed, the average internal-to-factor R^2 is 14%.

Regarding gender, the first-glance trend was of a significant feminine prevalence in entrepreneurial intentions: female graduates showed 11.5% intentions with respect to 7.5% of their male counterpart. The multivariate analysis, though, did not confirm this relationship either directly or through other factors.

Regarding the academic curriculum, we ascertained, among graduates who made steps toward entrepreneurship, a neat prevalence of graduates holding a Master's degree (19.3% *vs.* 4.3 of Bachelor's) in life sciences (17.5%) than in a STEM (Science, Technology, Engineering and Mathematics) discipline (science: 5.8%; engineering: 4.8%). It is puzzling that the propensity in STEM is even lower than in a social or humanistic science (respectively, 11.4 and 10.6%). Indeed, if we imagine an entrepreneur as a person who is able to put ideas into practice, this is a countertrend.

Working at graduation – that is the condition of having worked during higher education – was negatively related with human capital and even with actions undertaken to start an own business. While the former relationship was expected because working and studying at the same time generally leads to low-profile educational outcomes, the latter one may suggest that the dependent variable may not only reflect people's willingness to undertake but also availability to take into consideration any possibility in order to get a job.

We have found also a relationship between entrepreneurial intentions and final graduation mark, the intention belonging in a higher proportion to higher grades. In the extant literature (for a meta-analysis, see: Imose and Barber, 2015) this relationship is mixed. Moreover, Van Praag et al. (2009) showed that education negatively affects peoples' decisions to become an entrepreneur. Our data show that a higher graduation mark, taken alone or in conjunction with the academic discipline, seems to positively qualify people with a more determined intention to start an own business.

Finally, the way graduates retrospectively evaluated the expected effectiveness of the degree at hand – which was, as a whole, much more positive for employee-job oriented than for own-business-oriented graduates (respectively, 70.3% *versus* 56.5% positive evaluations) – is irrelevant to qualify higher levels of entrepreneurial intention if human capital factor was considered.

Concerning the psychological factors, no dimension was correlated with entrepreneurial disposition, neither PsyCap nor Loc, nor self-awareness. These results disconfirm the mainstream literature (Van Praag et al., 2009), in which both self-efficacy and being able to control own actions are psychological preconditions to develop an entrepreneurial disposition. Even the social capital showed not to influence the graduates' entrepreneurial spirit.

4. Discussion and final considerations

In this work we analysed the entrepreneurial intention of fresh graduates. We have found that just 10% of graduates is positively disposed to entrepreneurship. Bosma et al. (2020) show that a low propensity to start an own business is a worldwide phenomenon, as highlighted by the GEM - Global Entrepreneurship Monitor that yearly surveys adults of 50 countries.

Our data showed that working at graduation is negatively correlated with entrepreneurial disposition and, conversely, that good marks and the possession of a Master's degree in social and life sciences are positively correlated with graduates' entrepreneurial disposition. What

this means is unclear. Did we mix apples and oranges while defining the *Y* variable, or is this result once more the contradictory trend ascertained also in GEM that, in Italy, propensity to undertaking one's own business is low, much lower than 10%, but the proportion of people stating they possess the qualities to undertake it is high?

Our study showed that cognitive variables are much more relevant to entrepreneurial intention than non-cognitive ones. Both a positive psychological capital, an internal locus of control, positive attitudes towards labour and education, and the perception of individual and social barriers showed to be irrelevant to explain the graduates' entrepreneurial disposition. Instead, a risk-taking propensity showed a mild link with actions taken by graduates to start an own business. Therefore, an entrepreneurial intention model showed not to be fully consistent with the planned behaviour theory; moreover, the hypothesis that positively-disposed graduates are the "hive" of future entrepreneurs remains in a limbo.

The estimated R^2 is low and this may threaten the credibility of the relational model. In a future study, a more cogent definition of entrepreneurial disposition is to be tried before abandoning the hypothesis that that disposition precedes the decision to start an own business. Moreover, the study is to be circumscribed to people who effectively experienced the labour market.

References

Ajzen, I. (1991). Theory of planned behavior. *Organizational Behavior and Human Decision Processes*, **50**, pp. 179-211.

Bosma, N., Hill, S., Ionescu-Somers, A., Kelley, D., Levie, J., Tarnawa, A. (2020). *Global Entrepreneurship Monitor 2019/2020 Global Report*. Global Entrepreneurship Research Association, London.

Duell, N. (2018). *Employment and Entrepreneurship under the Youth Guarantee. Experience from the Ground*, European Commission, Directorate-General for Employment, Social Affairs and Inclusion. Publications Office of the European Union, Luxembourg.

Eurofound (2015). *Youth Entrepreneurship in Europe: Values, Attitudes, Policies*. Publications Office of the European Union, Luxembourg.

European Commission (2018). *Employment and Social Developments in Europe. Annual Review 2018*. Publications Office of the European Union, Luxembourg.

Krueger, N.F.Jr., Brazeal, D.V. (1994). Entrepreneurial potential and potential entrepreneurs. *Entrepreneurship: Theory & Practice*, Spring, pp. 91-104.

Imose, R., Barber, L.K. (2015). Using undergraduate grade point average as a selection tool: A synthesis of the literature. *The Psychologist-Manager Journal*, **18**(1), pp. 1-11.

Monecke, A., Leisch, F. (2012). semPLS: Structural Equation Modeling Using Partial Least Squares (https://www.jstatsoft.org/article/view/v048i03).

OECD/European Union (2017). *The Missing Entrepreneurs 2017: Policies for Inclusive Entrepreneurship*. OECD Publishing, Paris.

Rigdon, E.E., Sarstedt, M., Ringle, C.M. (2017). On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations. *Marketing-ZFP*, **39**(3), pp. 4–16.

Robusto, E., Maeran, R., Colledani, D., Anselmi, P., Scioni, M. (2019). Development of two scales for measuring academic psychological capital and locus of control in fresh graduates. *Statistica Applicata - Italian Journal of Applied Statistic*s, **31**, pp. 13-28.

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y.-M., Lauro, C. (2005). PLS path modeling. *Computational Statistics & Data Analysis*, **48**(1), pp. 159–205.

Van Praag, M., Van Witteloostuijn, A., Van der Sluis, J. (2009). *Returns for Entrepreneurs vs. Employees: The Effect of Education and Personal Control on the Relative Performance of Entrepreneurs vs. Wage Employees*. IZA Discussion Paper 4628. IZA, Bonn, Germany.

Nonparametric methods for stratified C-sample designs: a case study

Rosa Arboretti, Riccardo Ceccato, Luigi Salmaso

1. Introduction

The analysis of C-sample designs in the presence of stratification is a problem frequently faced by practitioners.

In the industrial field a variety of stratified analysis scenarios present themselves. Take, for example, a company that wishes to assess the performance of three different formulas for a new dishwasher detergent. Multiple dishwashers are used and multiple washes are carried out. At the end of each wash, an expert provides an evaluation of the cleaning performance of the formula. When analyzing the resulting data, the effect of using one dishwasher instead of another cannot be ignored, so each dishwasher is considered to be a separate stratum. Likewise, in the healthcare field it is quite common for multiple drugs to be tested on patients of different age groups. Each age group is again considered to be a stratum.

In this paper we focus on a scenario from the field of education. We are interested in assessing how the performance of students from different degree programs at the University of Padova changes, in terms of university credits and grades, when compared with their entrance exam results. In other words, we want to assess whether people who achieved the best results in this exam perform best during their academic career.

The entrance exam can have three possible outcomes (i.e. it is an ordinal variable). This is therefore a typical stochastic ordering problem (Basso et al., 2009; Basso and Salmaso, 2011; Bonnini et al., 2014), that is a problem in which the main interest lies in evaluating the null hypothesis $Y_1 \overset{d}{=} \ldots \overset{d}{=} Y_C$ against the alternative hypothesis $Y_1 \overset{d}{\leq} \ldots \overset{d}{\leq} Y_C$ and $\mathbb{E}[\psi(Y_1)] \leq \ldots \leq \mathbb{E}[\psi(Y_C)]$, where at least one inequality is strict, and $\psi(\cdot)$ is an increasing function (Pesarin and Salmaso, 2010). Our aim is in fact to assess whether by comparing increasing entrance exam outcomes, the $C = 3$ corresponding distributions of the student's performance measure Y are stochastically ordered.

A few nonparametric methods have been proposed in the literature to address these problems. Among them, Jonckheere's test (Jonckheere, 1954; Terpstra, 1952) is one of the first nonparametric solutions to test for ordered alternatives and is based on use of the Mann-Whitney test (Mann and Whitney, 1947) to perform all the possible $[C \times (C-1)]/2$ pairwise comparisons between C groups. Neuhäuser et al. (1998) also proposed a modification of this test that appears to be more powerful than the original test with small sample sizes (Shan et al., 2014). Additionally, permutation-based solutions involving the Non-Parametric Combination (NPC) technique (Pesarin and Salmaso, 2010; Klingenberg et al., 2009; Finos et al., 2007, 2008) were introduced.

We propose a further extension of the NPC technique to address stochastic ordering problems in the presence of stratification. Indeed, the impact of the student's choice of degree program cannot be ignored, therefore stratification must be considered in the testing procedure.

In section 2 we are going to describe the proposed permutation-based approach. In section 3 we apply it to the case study of interest related to university education. Finally, section 4

Rosa Arboretti, University of Padua, Italy, rosa.arboretti@unipd.it, 0000-0003-1263-0440
Riccardo Ceccato, University of Padua, Italy, riccardo.ceccato@unipd.it, 0000-0002-8629-8439
Luigi Salmaso, University of Padua, Italy, luigi.salmaso@unipd.it, 0000-0001-6501-1585

provides the results and conclusions.

2. Methodology

Firstly, let us further describe the stochastic ordering problem. The main interest lies in evaluating the system of hypotheses:

$$\begin{cases} H_0 : Y_1 \overset{d}{=} \ldots \overset{d}{=} Y_C \\ H_1 : Y_1 \overset{d}{\leq} \ldots \overset{d}{\leq} Y_C \text{ and at least one strict inequality } \overset{d}{<}, \end{cases}$$

where the symbol $\overset{d}{=}$ denotes equality in distribution and $\overset{d}{<}$ denotes stochastic dominance, i.e. $Y_1 \overset{d}{<} Y_2$ if and only if $F_1(z) \geq F_2(z), \forall z$ and $\exists I : F_1(z) > F_2(z), z \in I$ with $Pr(I) > 0$, where F_j is the cumulative distribution function. An alternative way to write this is:

$$\begin{cases} H_0 : F_1 = F_2 = \cdots = F_{(C-1)} = F_C \\ H_1 : F_1 \geq F_2 \geq \cdots \geq F_{(C-1)} \geq F_C \text{ and at least one strict inequality.} \end{cases} \tag{1}$$

NPC-based solutions generally consider a particular decomposition. The hypotheses are split in order to recreate the conditions of a set of two-sample problems as follows:

$$\begin{cases} H_0 : \bigcap_{i=1}^{C-1} H_{i0} = \bigcap_{i=1}^{C-1} [(F_1 = \cdots = F_i) = (F_{(i+1)} = \cdots = F_C)] \\ H_1 : \bigcup_{i=1}^{C-1} H_{i1} = \bigcup_{i=1}^{C-1} [(F_1 = \cdots = F_i) > (F_{(i+1)} = \cdots = F_C)]. \end{cases}$$

where the null hypothesis H_0 is the intersection of a number of partial hypotheses and the alternative hypothesis H_1 is the union of $C - 1$ sub-hypotheses.

For each pair of sub-hypotheses H_{i0} and H_{i1}, the first i and the last $(C - i)$ samples are pooled, so that two new samples X_1 and X_2 are achieved, with sizes N and M. The sub-problem can therefore be rewritten as:

$$\begin{cases} H_{i0} : X_1 \overset{d}{=} X_2 \\ H_{i1} : X_1 \overset{d}{<} X_2. \end{cases}$$

Each sub-hypothesis is then tested separately, using appropriate permutation tests. The adopted test statistic can differ according to the nature of the data, but a common and versatile choice is the modified version of the Anderson-Darling test statistic:

$$T = \sum_j^n [\hat{F}_1(X_j) - \hat{F}_2(X_j)] / \{\bar{F}(X_j)[1 - \bar{F}(X_j)]\}^{\frac{1}{2}} \tag{2}$$

where $X = \{X_1, X_2\}$ is the pooled sample, $\hat{F}_1(t) = \sum_j^N \mathbb{I}(X_{j1} \leq t)/N$, $\hat{F}_2(t) = \sum_j^M \mathbb{I}(X_{j2} \leq t)/M$, $\bar{F}(t) = \sum_j^n \mathbb{I}(X_j \leq t)/n$, $n = N + M$, $t \in \mathcal{R}^1$ and $\mathbb{I}(\cdot)$ is the indicator function which is 1 if $(\cdot)$ is satisfied and 0 otherwise.

According to the NPC algorithm (Pesarin and Salmaso, 2010), B permuted datasets are independently generated for each sub-problem and the related values of the test statistic $T_b^*, b = 1, \ldots, B$ are calculated to simulate the null distribution of T. Partial p-values (λ_i) and $\lambda_{ib}^*, b = 1, \ldots, B$ estimating their distributions can therefore be achieved. It is worth noting that the same permutation design is adopted for each sub-problem, to implicitly take into account the existing dependency among sub-problems.

A combination step now needs to be performed. The partial p-values $\lambda_i, i = 1, \ldots, C - 1$ related to the $C - 1$ sub-problems $\{H_{i0} \text{ vs } H_{i1}\}$ are combined using an adequate combining function, such as Fisher's combining function $T_F'' = -2 \cdot \sum_{i=1}^{C-1} \log(\lambda_i)$. The same is done for each of the B vectors $\lambda_{ib}^*, i = 1, \ldots, C - 1$. The elements of the new resulting vector represent the second-order test statistics, from which it is finally possible to achieve the global p-value λ'' to assess the system of hypotheses 1.

Given that stratification needs to be included, we propose firstly applying this procedure to each of the S strata, testing S systems of hypotheses:

$$\begin{cases} H_{0s} : F_{1s} = F_{2s} = \cdots = F_{(C-1)s} = F_{Cs} \\ H_{1s} : F_{1s} \geq F_{2s} \geq \cdots \geq F_{(C-1)s} \geq F_{Cs} \text{ and at least one strict inequality.} \end{cases} \quad (3)$$

After applying the aforementioned NPC-based approach to each stratum, the global p-values $\lambda_s'', \forall s = 1, \ldots, S$ (and the $\lambda_{sb}^*{}''$ estimating their distributions) are thus retained. Then we adopt a further combination step, using the Fisher combining function, and retrieve a final p-value λ'''. In this way, by comparing λ''' to the desired significance level α, we are able to solve the global stochastic ordering problem H_0 vs H_1.

Given that multiple systems of hypotheses H_{s0} vs $H_{s1}, \forall s = 1, \ldots, S$ are assessed, we then apply an appropriate multiplicity correction to control the false discovery rate (FDR). Our choice is the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995).

3. A case study

Let us now focus on the real stratified C-sample problem at hand. As mentioned before, we are interested in evaluating the performances of students from different degree programs at the University of Padova. In particular, we want to understand if the university credits gained at the end of the first year (Y^a), the credits gained at the end of the third year (Y^b) and the final average grade (Y^c) somehow depend on the results achieved by the student in the entrance exam. In other words, we try to indirectly assess the efficacy of this exam in evaluating and selecting future students. The analysis is performed using R (R Core Team, 2020).

Let us briefly describe the data. The total sample size is 3083 students. Firstly, the degree programs are grouped into 4 classes (identified by their Italian subject titles):

- ING_INFORMAZIONE_NON_PROFES (S1)
- ING_CIVILE_AMBIENTALE_L7 (S2)
- ING_INFORMAZIONE_L8 (S3)
- ING_INDUSTRIALE_L9 (S4).

The different classes represent different strata (i.e. $S = 4$) and have different sample sizes (see Figure 1). The variable reporting the outcome of the entrance exam has three modalities (i.e. $C = 3$), namely INSUFFICIENTE, SUFFICIENTE and PIU'_CHE_SUFFICIENTE (Insufficient, Sufficient and More Than Sufficient). For the sake of simplicity, we are going to refer to them as INS, SUF and PIU in our notation. In Figure 1, the possible outcomes are ordered from worst to best.

For each response variable $Y^j, \forall j \in \{a, b, c\}$, we want to assess if $Y_{\text{INS}}^j \overset{d}{\leq} Y_{\text{SUF}}^j \overset{d}{\leq} Y_{\text{PIU}}^j$, with at least one strict inequality, taking into account the effect of the degree program class.

Looking at credits gained at the end of the first year, a first descriptive analysis (see Figure 2) appears to support the alternative hypothesis. Indeed, in all strata, students achieving INS at the entrance exam appear to perform worse than students achieving SUF, and students achieving PIU at the entrance exam tend to perform better than students achieving SUF.

Similar conclusions can be drawn about both credits gained at the end of the third year (see Figure 3) and the average grade at the end of the academic career (see Figure 4).

Applying our testing procedure, we managed to confirm these hypotheses. We set $B = 10000$ and used the test statistic in Equation 2 and Fisher's combining function. When looking at Y^a (see Table 1), all the partial p-values and the global p-value proved to be substantially smaller than 1%. The only exceptions were ING_CIVILE_AMBIENTALE_L7 (S2) and ING_INFORMAZIONE_L8 (S3), for which the descriptive analysis shows that the order among entrance exam outcomes is less evident.
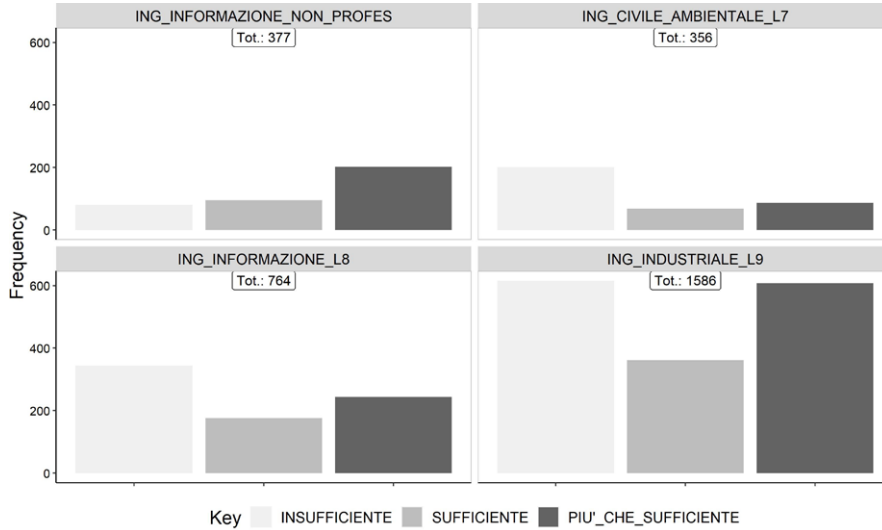
Figure 1: Description of the sample.
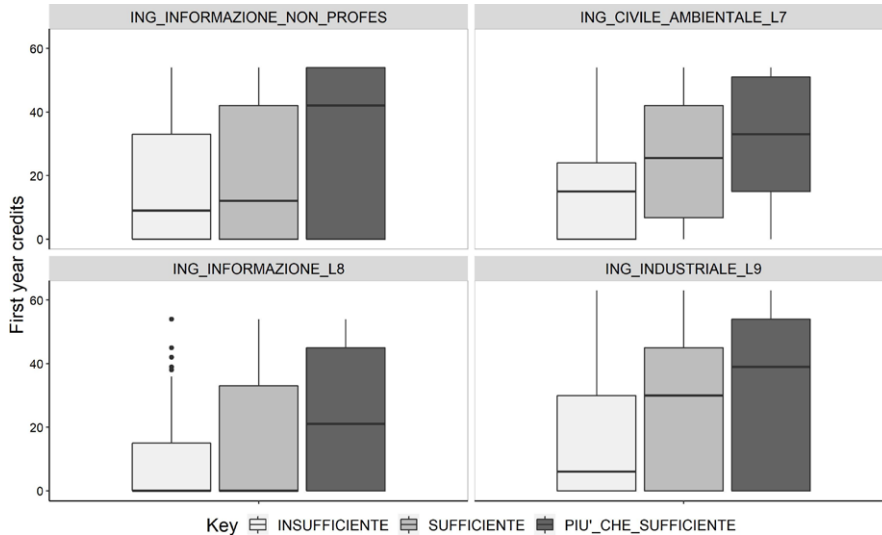
Figure 2: Credits at the end of the first year.
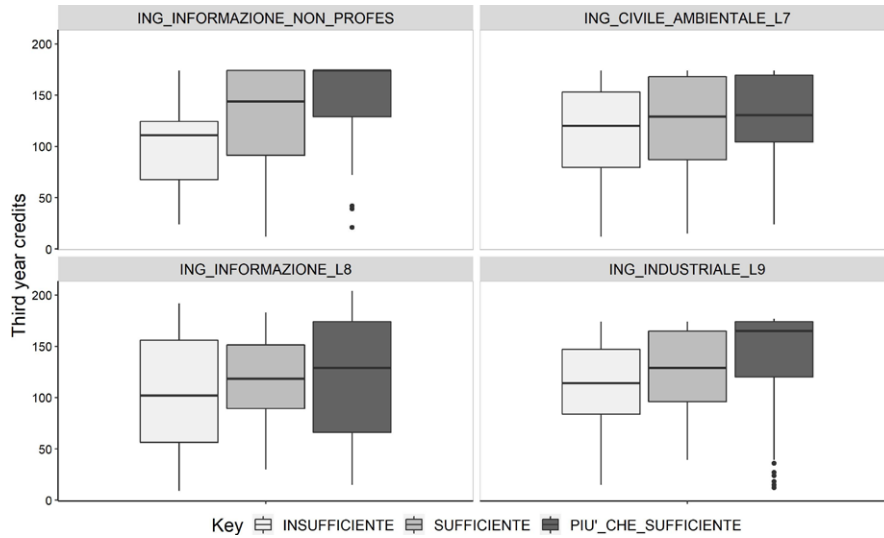
Figure 3: Credits at the end of the third year.
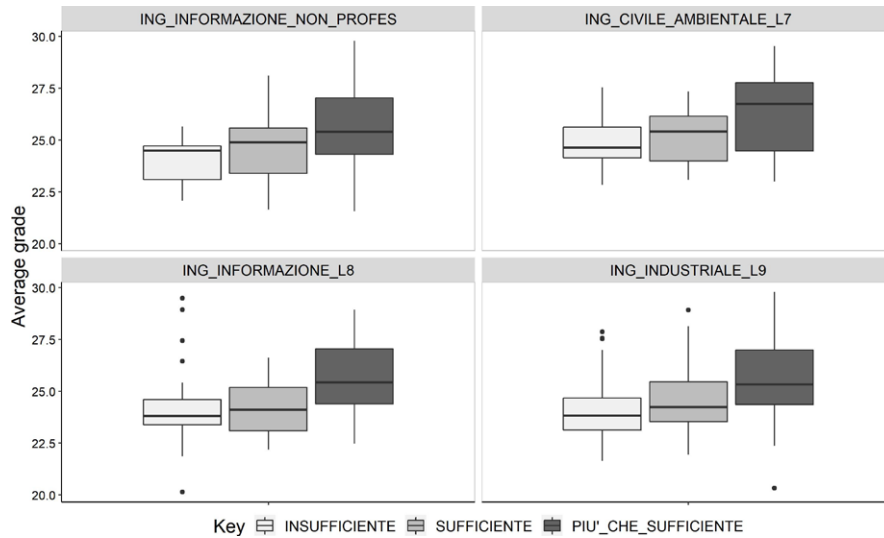
Figure 4: Average grade at the end of the academic career.

Table 1: Table of p-values for Y^a, Y^b and Y^c.

Response	Global	S1	S2	S3	S4
Y^a	1e-4	1e-4	1e-4	1e-4	1e-4
Y^b	1e-4	2e-4	0.1471	0.1185	2e-4
Y^c	1e-4	4e-4	2.9e-3	8e-4	6e-4

4. Conclusions

In this paper we presented a new solution to C-sample stochastic ordering problems in the presence of stratification, focusing on its application to a case study from the field of education.

Our proposal takes advantage of the Non-Parametric Combination (NPC) procedure (Pesarin and Salmaso, 2010), a versatile permutation-based methodology allowing us to solve several different complex problems, such as stochastic ordering. We apply this technique to evaluate the presence of stochastic ordering in each of the S existing strata and then use an appropriate combining function to assess the stochastic ordering in all the samples.

The application of this procedure allowed us to assess the efficacy of the University of Padova's entrance exams in evaluating and selecting future students. Indeed, it emerged that students with the worst results in the entrance exam tended to perform the worst during their academic career, in terms of both university credits achieved at the end of the first and third years and in terms of the final average grade, independently of the chosen degree program. The only exception was people from ING_CIVILE_AMBIENTALE_L7 and ING_INFORMAZIONE_L8. For these two strata, when the credits at the end of the third year were considered, it was not possible to find enough evidence in favor of the stochastic ordering hypothesis.

Overall, this approach appears to be significantly promising and a simulation study has been planned to further explore its performances.

References

Basso, D., Pesarin, F., Salmaso, L., Solari, A. (2009). *Permutation tests for stochastic ordering and ANOVA: theory and applications with R*. Springer Science & Business Media, New York, (NY).

Basso, D., Salmaso, L. (2011). A permutation test for umbrella alternatives. *Statistics and Computing*, **21**(1), pp. 45–54.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, **57**(1), pp. 289–300.

Bonnini, S., Prodi, N., Salmaso, L., Visentin, C. (2014). Permutation approaches for stochastic ordering. *Communications in Statistics-Theory and Methods*, **43**(10-12), pp. 2227–2235.

Finos, L., Salmaso, L., Solari, A. (2007). Conditional inference under simultaneous stochastic ordering constraints. *Journal of statistical planning and inference*, **137**(8), pp. 2633–2641.

Finos, L., Pesarin, F., Salmaso, L., Solari, A. (2008). Exact inference for multivariate ordered alternatives. *Statistical Methods and Applications*, **17**(2), pp. 195–208.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives. *Biometrika*, **41**(1/2), pp. 133–145.

Klingenberg, B., Solari, A., Salmaso, L., Pesarin, F. (2009). Testing marginal homogeneity against stochastic order in multivariate ordinal data. *Biometrics*, **65**(2), pp. 452–462.

Mann, H. B., Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, **18**(1), pp. 50–60.

Neuhäuser, M., Liu, P.-Y., Hothorn, L. A. (1998). Nonparametric tests for trend: Jonckheere's test, a modification and a maximum test. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, **40**(8), pp. 899–909.

Pesarin, F., Salmaso, L. (2010). *Permutation tests for complex data: theory, applications and software*. John Wiley & Sons, Hoboken, (NJ).

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, (AT).

Shan, G., Young, D., Kang, L. (2014). A New Powerful Nonparametric Rank Test for Ordered Alternative Problem. *PloS one*, **9**(11), pp. 1–10.

Terpstra, T. J. (1952). The asymptotic normality and consistency of Kendall's test against trend, when ties are present in one ranking. *Indagationes Mathematicae*, **14**(3), pp. 327–333.

Measuring content validity of academic psychological capital and locus of control in fresh graduates

Pasquale Anselmi, Daiana Colledani, Luigi Fabbris, Egidio Robusto, Manuela Scioni

1. Introduction

PETERE (Preferences for Employment and Training as Elected by REcent graduates) is a project of the University of Padua that investigated how fresh graduates interact with labour market to understand how to improve placement policies and support plans. One of the aims of the project was the identification of psychological patterns that could help graduates to stand the labour market in uncertain times. According to the literature, two sets of psychological variables have been identified that can be crucial to achieve academic and professional success.

The first set was developed within the framework of positive psychology (Seligman & Csikszentmihalyi, 2014) and is named "Psychological capital" (PsyCap; Luthans et al., 2007). PsyCap defines a positive psychological state characterized by feelings of self-efficacy, hope, optimism, and resilience. Self-efficacy (or confidence) describes the conviction of having all the abilities, motivation, and resources needed to successfully execute a specific task. Hope defines a positive motivational state that leads individuals to pursue their own objectives, redirecting, when it is necessary, the strategies employed to achieve them. Optimism is the subjective tendency to interpret situations and events positively. In the framework of PsyCap, this trait describes the propensity to carefully consider both positive and negative events to understand their causes and consequences (Youssef & Luthans, 2005). Optimistic individuals build positive expectancies that motivate them to persist toward their goals, dealing with difficulties, and reaching success (Chemers et al., 2001; Sharpe et al., 2011). The last trait included in PsyCap is resilience, which defines the ability to "bounce back" from adversity, failure, and uncertainty.

The second set of psychological variables is locus of control (LoC). It may be internal and external. The first defines the extent to which individuals perceive strong links between their actions and the following results. Individuals with internal LoC feel having control over their own fate. Conversely, external LoC defines the inclination to perceive a low control on ones' fate. Individuals with external LoC attribute personal outcomes to external and uncontrollable factors (Lefcourt, 2014; Rotter, 1966).

PsyCap and LoC have been extensively related to important work outcomes, including job satisfaction, job performance, and organisational commitment (e.g., Avey et al., 2010; 2011; Hansen et al., 2015; Judge & Bono, 2001). Moreover, these variables have been found to be associated with positive academic results, such as high performance and motivation, academic satisfaction, inclination to use effective and functional coping strategies, and ability to deal with stress (e.g., Clifton et al., 2004; Conti, 2000; Drago et al., 2018; Elias & Loomis, 2002; McKenzie & Schweitzer, 2001; Mohamed et al., 2018; Nunn & Nunn, 1993; Snyder et al., 2002). The attention towards these variables might be also due to the fact that they are "state-like" variables and can be modified through targeted interventions (Luthans et al., 2008; Stanton, 1982).

Scales for measuring PsyCap and LoC exist in the literature. The PsyCap Questionnaire (PCQ; Luthans et al., 2007) is meant for workers. As such, it might be inappropriate for assessing these traits among fresh graduates who were about to enter the world of work. With respect to LoC, there is a scale, called Academic Locus of Control Scale (Trice, 1985) which is intended for

students. However, this instrument is founded on a unidimensional conceptualization of LoC, which is not supported by research in this field. Levenson (1981), for instance, found that internal and external LoC are two distinct dimensions.

Recently, two brief instruments have been appositely developed for measuring PsyCap and LoC among fresh graduates: the Academic PsyCap and the LoC scales (Robusto et al., 2019). These two scales showed significant relationships with the occupational status of respondents, with their entrepreneurial disposition, and with the number of actions taken when they were looking for a job. Although the two scales showed satisfactory psychometric properties, there was room for some improvement pertaining to the content validity and the length of the six subscales (i.e., self-efficacy, hope, optimism, resilience, internal LoC, and external LoC). With respect to the former, the analysis of the content of the items included in each subscale suggested that they did not adequately cover relevant operationalizations of the different psychological variables. With respect to the latter, the length of some subscales was too small (e.g., internal LoC subscale contained only 3 items). To this purpose, in the present work new items were developed for each of the six subscales with the aim of increasing their length and improving the coverage of additional relevant operationalizations.

2. Method

Participants and procedure

To test the functioning of the new scales, a study was conducted on 1105 graduates (Males 36.7%, Mean age = 24.92, *SD* = 4.66) at the University of Padua. They were surveyed in the context of the PETERE project within one month after graduation. The survey was administered via a CAWI (Computer-Assisted Web-based Interviewing) system. Students from medicine and nursing courses were not included in the sample. To analyse the data on the Academic PsyCap and LoC scales, the total sample was randomly split into two subsamples including 550 (Males 35.9%, Mean Age = 25.11, *SD* = 4.84) and 555 (Males 37.1%, Mean Age = 24.72, *SD* = 4.47) participants, respectively.

Measures

A total of 37 items were used to measure the four facets of PsyCap: resilience (11 items, 5 of them being new), self-efficacy (9 items, 2 of them being new), optimism (9 items, 2 of them being new), and hope (8 items, 2 of them being new).

To evaluate internal and external LoC, 12 items were used, six for each subscale (3 items of internal LoC and 2 items of external LoC being new).

All items were scored on a four-point Likert scale (from 1 "Completely disagree" to 4 "Completely agree").

Analytic approach

The factor structure of Academic PsyCap and LoC scales was tested through Exploratory Structural Equation Models (ESEMs; Asparouhov & Muthén, 2009) and Confirmatory Factor Analyses (CFAs). The ESEMs were run in the first subsample (*n* = 550) whereas CFAs in the second (*n* = 555). The ESEMs were performed on all the 37 and 12 items of Academic PsyCap and LoC scales (defining four and two factors, respectively), and allowed for the identification and exclusion of poorly performing items (i.e., items with large cross-loadings or low factor loadings on the intended scale). After having removed the items with unsatisfactory performance, the factor structure of Academic PsyCap and LoC was confirmed through CFAs. ESEMs and CFAs were run using the WLSMV estimator (weighted least squares mean and variance-adjusted; Muthén & Muthén, 2012); this method is recommended for categorical observed data (e.g., Flora

& Curran, 2004; Brown, 2006). The goodness of fit of the models was evaluated by means of several fit indices: χ^2, root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR). A solution fits the data when χ^2 is non-significant ($p > .05$). Since this statistic is sensitive to sample size, the other fit measures were also taken into account in the evaluation of the models. Specifically, CFI indices close to .95 (.90 to .95 for reasonable fit), SRMR values less than .08, and RMSEA smaller than .06 (.06 to .08 for reasonable fit) are indicative of good model fit (Marsh et al., 2004).

Composite reliability was computed to measure the internal consistency of the scales. This coefficient is conceptually similar to Cronbach's α but more accurate and can be easily computed in the structural equation modeling framework (Raykov, 2001). Composite reliability ranges from 0 to 1. The closer the value to 1, the larger the internal consistency.

3. Results

Academic PsyCap

Table 1. Factor loadings (λ) from the CFA of the Academic PsyCap scale

Item		λ
	Self-efficacy	
S1	Usually, when I face a problem, I am able to identify different solutions.	0.726
S2	In difficult situations, I believe I am effective in finding a way out.	0.887
S3	I have the resources to handle even unforeseen situations.	0.729
S4	When I work hard, I can solve even the most difficult problems.	0.655
S9	I am sure I can effectively deal with unexpected events.	0.773
SNew1	I am confident in my abilities to find effective solutions to problems.	0.810
	Optimism	
O2	I always try to believe that behind every cloud there is a blue sky	0.817
O3	Thinking about my life I expect more negative than positive happenings. (R)	0.570
O5	In critical situations I usually expect them to end at best.	0.600
O8	In general, I always try to see the glass half full.	0.864
ONew1	I'm usually optimistic about the future.	0.844
ONew2	Even in difficult situations, I try to take the best opportunities and the bright side.	0.859
	Resilience	
R1	Until now, my successes have largely depended on the choices I made.	0.633
R2	The obstacles that I have overcome in my studies have certainly made me stronger and more combative.	0.638
R3	I am proud of everything I have achieved by now.	0.708
R8	Having completed my course of study or being in the process of doing so makes me proud.	0.743
RNew1	Usually, in one way or another, I try to overcome difficulties.	0.676
RNew2	I always try to give my best in all the things I do without getting discouraged in the face of obstacles.	0.715
	Hope	
H3	The goals I have achieved so far are due to my determination.	0.822
H5	In general, I plan carefully things to do to achieve my goals.	0.754
H6	I have hard times planning things to do when I have to reach a goal. (R)	0.675
H7	The goals I have achieved so far are due to my planning skills.	0.709
H9	Willpower was key to obtaining an academic degree.	0.742
HNew1	I think I will be able to achieve my current goals by counting on my determination	0.786
Correlations between factors		
	Optimism / Self-efficacy	0.632
	Resilience / Self-efficacy	0.749
	Resilience / Optimism	0.650
	Hope / Self-efficacy	0.644
	Hope / Optimism	0.449
	Hope / Resilience	0.883

Note. All factor loadings and correlation coefficients were significant $p < .001$

The four-factor ESEM run on the 37 items of the Academic PsyCap scale obtained a

successful fit. Although χ^2 was significant due to sample size ($\chi^2(524) = 1298.107$, $p < .001$), the other indices satisfied the rules of thumb (RMSEA = .052 [.048, .055]; CFI = .953; SRMR = .045).

The inspection of factor loadings, modification indices, and item content suggested excluding 13 items from the final version of the scale. In particular, one item of the self-efficacy scale was excluded since its content was very close to that of another item of the same scale but it was characterized by a weaker factor loading. Three items of the optimism scale, one of the resilience scale and two items of the hope scale were excluded since they exhibited weak loadings on the intended factor. One item of the self-efficacy scale and three items of the resilience scale were excluded because they exhibited high cross-loadings. Finally, two items, one from the self-efficacy scale and the other from the resilience scale, were excluded according to indications of modifications indices. The new self-efficacy, optimism, resilience, and hope scales contained 6 items each, out of which: 1 item of self-efficacy, 2 items of optimism, 2 items of resilience, and 1 item of hope were new.

The results of the CFA run in the second sample, on the remaining 24 items, are reported in Table 1. The model showed an adequate fit: $\chi^2(246) = 930.574$, $p < .001$; RMSEA = .071 [.066, .076]; CFI = .941; SRMR = .066. Composite reliability was satisfactory for all scales: .89, .83, .84, and .79 for self-efficacy, optimism, resilience, and hope, respectively.

LoC

The two-factor ESEM run on the 12 items of the LoC scale obtained a successful fit. Despite χ^2 was significant due to sample size ($\chi^2(43) = 182.343$, $p < .001$), the other indices were satisfactory (RMSEA = .077 [.065, .088]; CFI = .930; SRMR = .054). The inspection of factor loadings, modification indices, and item content suggested excluding only two items, one for each subscale. In particular, the item of internal LoC was excluded because it did not load on the intended factor, whereas the item of external LoC was excluded because it had the lowest loading on the factor. The new LoC scales contained 5 items each (2 items of internal LoC and 1 item of external LoC were new).

The results of the CFA run in the second sample, on the remaining 10 items, are reported in Table 2. The model showed an adequate fit: $\chi^2(41) = 138.393$, $p < .001$; RMSEA = .075 [.062, .088]; CFI = .940; SRMR = .068. Composite reliability was satisfactory for both scales: .62 and .80 for internal and external LoC, respectively.

Table 2. Factor loadings (λ) from the CFA of the LoC scale

Item		λ
	Internal LoC	
IL1	I think that if I'm serious and prepared I will always find a good work position.	0.713
IL2	Even if it is not always true, I believe there is some relationship between the worth of individuals and their income.	0.288
IL4	I think that my academic choice will allow me to have good job opportunities.	0.610
ILNew2	I think I am directly responsible for my choices, my actions, and the results that follow.	0.515
ILNew1	I think that everything I have achieved is exclusively the result of my commitment, my skills, and my dedication.	0.326
	External LoC	
EL1	I think luck and chance are crucial for me to find the "right" job.	0.850
EL2	I think having the right contacts is more important than personal skills to find a good job.	0.540
EL3	I believe luck is crucial for me to obtain a good job position.	0.752
EL5	I think that often good job positions are achieved by completely random factors.	0.647
ELNew1	I'm convinced that against bad luck and doom there is no way out.	0.520
Correlations between factors		
	Internal LoC/ External LoC	-0.241

Note. All factor loadings and correlation coefficients were significant $p < .001$

4. Discussion

The two scales for measuring Academic PsyCap and LoC introduced by Robusto et al. (2019) have been administered to a new large sample of fresh graduates in order to develop new items and evaluate their performance. The new Academic PsyCap scale contained 24 items, 6 for each of the four subscales. The new LoC scale contained 10 items, 5 for each of the two subscales. One to two items of each subscale were new.

On the whole, the psychometric properties of the new instruments are in line with those of the original ones. However, the content validity of the new scales was improved due to the introduction of items that investigate additional relevant operationalizations of the psychological variables. Moreover, in the new version of the instruments, the subscales were balanced for item length: the four Academic PsyCap subscales contained 6 items each, while the two subscales of Academic LoC contained 5 items each. This was especially useful for the internal LoC subscale that, in the previous version, contained only three items.

References

Asparouhov, T., Muthén, B. (2009). Exploratory Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, **16**(3), pp. 397-438.

Avey, J.B., Luthans, F., Youssef, C.M. (2010). The Additive Value of Positive Psychological Capital in Predicting Work Attitudes and Behaviors. *Journal of Management*, **36**, pp. 430-452.

Avey, J.B., Reichard, R.S., Luthans, F., Mhatre, K.H. (2011). Meta-Analysis of the Impact of Positive Psychological Capital on Employee Attitudes, Behaviors, and Performance. *Human Resource Development Quarterly*, **22**, pp. 127-152.

Brown, T.A. (2006). *Confirmatory Factor Analysis for Applied Research*. Guilford Press, New York, (NY).

Chemers, M.M., Hu, L.T., Garcia, B.F. (2001). Academic Self-Efficacy and First Year College Student Performance and Adjustment. *Journal of Educational Psychology*, **93**(1), pp. 55-64.

Clifton, R.A., Perry, R.P., Stubbs, C.A., Roberts, L.W. (2004). Faculty Environments, Psychosocial Dispositions, and the Academic Achievement of College Students. *Research in Higher Education*, **45**(8), pp. 801-828.

Conti, R. (2000). College Goals: Do Self-Determined and Carefully Considered Goals Predict Intrinsic Motivation, Academic Performance, and Adjustment during the First Semester?. *Social Psychology of Education*, **4**(2), pp. 189-211.

Drago, A., Rheinheimer, D.C., Detweiler, T.N. (2018). Effects of Locus of Control, Academic Self-Efficacy, and Tutoring on Academic Performance. *Journal of College Student Retention: Research, Theory & Practice*, **19**(4), pp. 433-451.

Elias, S.M., Loomis, R.J. (2002). Utilizing Need for Cognition and Perceived Self-Efficacy to Predict Academic Performance. *Journal of Applied Social Psychology*, **32**(8), pp. 1687-1702.

Flora, D.B., Curran, P.J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis with Ordinal Data. *Psychological Methods*, **9**, pp. 466-491.

Hansen, A., Buitendach, J.H., Kanengoni, H. (2015). Psychological Capital, Subjective Well-Being, Burnout and Job Satisfaction Amongst Educators in the Umlazi Region in South Africa. *SA Journal of Human Resource Management*, **13**(1), pp. 1-9.

Judge, T.A., Bono, J.E. (2001). Relationship of Core Self-Evaluations Traits–Self-Esteem, Generalized Self-Efficacy, Locus of Control, and Emotional Stability–with Job Satisfaction and Job Performance: A Meta-Analysis. *Journal of Applied Psychology*, **86**, pp. 80-92.

Lefcourt, H.M. (2014). *Locus of Control: Current Trends in Theory & Research* (2nd ed.). Psychology Press, New York, (NY).

Levenson, H. (1981). Differentiating among Internality, Powerful Others, and Chance. *Research with the Locus of Control Construct*, **1**, pp. 15-63.

Luthans, F., Avey, J.B., Patera, J.L. (2008). Experimental Analysis of a Web-based Training Intervention to Develop Positive Psychological Capital. *Academy of Management Learning and Education*, **7**, pp. 209-221.

Luthans, F., Avolio, B.J., Avey, J.B., Norman, S.M. (2007). Positive Psychological Capital: Measurement and Relationship with Performance and Satisfaction. *Personnel Psychology*, **60**(3), pp. 541-572.

Luthans, F., Youssef, C.M., Avolio, B.J. (2007). Psychological capital: Investing and developing positive organizational behavior, in *Positive Organizational Behavior: Accentuating the Positive at Work,* eds. D. Nelson, and C.L. Cooper, Sage, Thousand Oaks, (CA), pp. 9-24.

Marsh, H.W., Hau, K.T., Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling*, **11**(3), pp. 320-341.

McKenzie, K., Schweitzer, R. (2001). Who Succeeds at University? Factors Predicting Academic Performance in First Year Australian University Students. *Higher Education Research & Development*, **20**(1), pp. 21-33.

Mohamed, A.A., Mohammed, A.M., Ahmed, H.A.E. (2018). Relation between Locus of Control and Academic Achievement of Nursing Students at Damanhour University. *IOSR Journal of Nursing and Health Science (IOSR-JNHS)*, **7**, pp. 1-13.

Muthén, B.O., Muthén, L.K. (2012). *Mplus Version 7: User's Guide*. Muthén & Muthén, Los Angeles, (CA).

Nunn, G.D., Nunn, S.J. (1993). Locus of Control and School Performance: Some Implications for Teachers. *Education*, **113**(4), pp. 636-641.

Raykov, T. (2001). Bias of Coefficient α for Fixed Congeneric Measures with Correlated Errors. *Applied Psychological Measurement*, **25**(1), pp. 69-76.

Robusto, E., Maeran, R., Colledani, D., Anselmi, P., Scioni, M. (2019). Development of two Scales for Measuring Academic Psychological Capital and Locus of Control in Fresh Graduates. *Italian Journal of Applied Statistics*, **31**(1), pp. 13-28.

Rotter, J.B. (1966). Generalized Expectancies for Internal versus External Control of Reinforcement. *Psychological Monographs: General and Applied*, **80**(1), pp. 1-28.

Seligman, M.E., Csikszentmihalyi, M. (2014). Positive psychology: An introduction, in *Flow and the Foundations of Positive Psychology. The Collected Works of Mihaly Csikszentmihalyi,* ed. M. Csikszentmihalyi, Springer, Dordrecht, (NL), pp. 279-298.

Sharpe, J.P., Martin, N.R., Roth, K.A. (2011). Optimism and the Big Five Factors of Personality: Beyond Neuroticism and Extraversion. *Personality and Individual Differences*, **51**(8), pp. 946-951.

Snyder, C.R., Shorey, H.S., Cheavens, J., Pulvers, K.M., Adams III, V.H., Wiklund, C. (2002). Hope and Academic Success in College. *Journal of Educational Psychology*, **94**(4), pp. 820-826.

Stanton, H.E. (1982). Modification of Locus of Control: Using the RSI Technique in the Schools. *Contemporary Educational Psychology*, **7**, pp. 190-194.

Trice, A.D. (1985). An academic Locus of Control Scale for College Students. *Perceptual and Motor Skills*, **61**, pp. 1043-1046.

Youssef, C.M., Luthans, F. (2005). Resiliency development of organizations, leaders and employees: Multi-level theory building for sustained performance, in *Authentic Leadership Theory and Practice: Origins, Effects and Development (Monographs in leadership and management, Vol. 3),* eds. W. Gardner, B. Avolio, and F. Walumbwa, Elsevier, Oxford, (UK), pp. 303-343.

Random effects regression trees for the analysis of INVALSI data

Giulia Vannucci, Anna Gottard, Leonardo Grilli, Carla Rampichini

1. Introduction

Multilevel data structures, where data are typically clustered in nested levels, are common in many fields. An emblematic example consists of students, that are grouped in classes and schools (individual cross-sectional data) or children growth evaluated at several time points (repeated measures). Multilevel data require specific models referred to as *multilevel*, *random effects* or *mixed* (Snijders and Bosker, 2012).

Model specification is a challenging task in mixed models. Typically, a linear model is assumed, although non-linearities and interaction effects are undeniably of interest. A worthwhile approach exploits regression trees and the CART algorithm (Breiman et al., 1984) to capture non-linearities and high-order interaction effects. In particular, regression trees are a statistical learning algorithm that shapes the regression function as piece-wise constant over a recursively found partition of the covariate space. The graphical display of the recursive partition provides an easy interpretation of this predictive algorithm. The procedure, however, assumes statistical units to be independent, which is not the case of clustered data.

Regression trees have been extended to clustered data by Hajjem et al. (2011), who proposed to model fixed effects with a decision tree while accounting for random effects via a linear mixed model in a separate, subsequent, step. In particular, they first apply the CART algorithm as if data were not clustered to estimate the fixed effects. It is shown that random effect regression trees are less sensitive to parametric assumptions and provide improved predictive power compared to linear models with random effects and regression trees without random effects. The literature has thereon grown with variants and extensions. Among others, see Sela (2012); Hajjem et al. (2014); Miller et al. (2017).

In this work, we propose a further variation of the mixed effects regression tree, where the fixed and the random part parameters are estimated jointly, using a backfitting algorithm. To ease the interpretation, our proposal incorporates a linear component additively to the regression trees. Consequently, the general trend of dependence is captured by the linear component, while the tree part captures interactions and non-linearities.

The proposed algorithm is then applied to data collected by the national institute for the evaluation of the educational system and training (INVALSI: Istituto Nazionale per la VALutazione del Sistema educativo di Istruzione e di formazione) in Italy. The study aims to compare schools' educational effectiveness impartially by measuring students' progress over their careers. We focus on test scores in Mathematics, given some characteristics of the school and the pupil. The proposed model is able to take into account the student clustering in schools and to capture interesting interactions between student-level covariates and school-level covariates.

The rest of the paper is organised as follows. Section 2 illustrates the model proposed, together with the backfitting algorithm. Section 3 describes the application of the proposal to INVALSI data. A brief section of final remarks concludes the paper.

2. A tree embedded linear mixed model

We propose a random effect model, called *Tree Embedded Linear Mixed* (TELM) *model*, able to treat both non-linear and interaction effects and cluster mean dependencies. Motivated by the application of interest, we consider in particular a two-level random effect model. Hence, we will denote as *level 1 units* the statistical units (e.g. students) and *level 2 units* the groups (e.g. schools).

The model is a piecewise-linear regression function, consisting of the sum of a tree component and a mixed effect linear component. The proposal is the mixed effect version of the semi-linear regression trees (Vannucci, 2019). It can be ideally divided into three parts: a fixed effect linear part, a fixed effect non-linear part based on a tree and a random effect part. The resulting model can be formulated as

$$Y_{ij} = \beta_0 + \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma} + T(\mathbf{X}_{ij}, \mathbf{Z}_j) + U_j + \epsilon_{ij} \tag{1}$$

where Y_{ij} is the response variable for level 1 unit i belonging to level 2 unit j, β_0 is the (fixed-effect) regression intercept, $\mathbf{X}_{ij}$ is the vector of the level 1 covariates, $\boldsymbol{\beta}$ the associated fixed effect coefficients, $\mathbf{Z}_j$ is the vector of the level 2 covariates, $\boldsymbol{\gamma}$ the associated fixed effect coefficients. Here, $T(\mathbf{X}_{ij}, \mathbf{Z}_j)$ is the tree based component depending on some or all the level 1 and the level 2 explanatory variables. Finally, $U_j \sim N(0, \sigma_u^2)$ is the random intercept for level 2 unit j and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ are the regression errors.

The model is additive in its components where the tree-component acts as a region-specific categorical variable. This can be seen in the following alternative specification

$$Y_{ij} = \beta_0 + \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma} + \sum_{m=1}^{M} \mu_m \mathbb{I}\{(\mathbf{X}_{ij}, \mathbf{Z}_j) \in R_m\} + U_j + \epsilon_{ij}, \tag{2}$$

where $R_1, \ldots, R_M$ is the partition of the predictor space corresponding to the tree-component. When the unknown regression function can be assumed to be quasi-linear (Wermuth and Cox, 1998), the number of leaf nodes M can be kept small to avoid overfitting.

To account for the contextual effects of level 1 predictors, we add the cluster mean $\overline{W}_j = (1/n_j) \sum_{i=1}^{n_j} W_{ij}$ to the set of level 2 predictors $\mathbf{Z}_j$ (Snijders and Bosker, 2012).

An iterative, backfitting-like procedure obtains model fitting. First, the tree is initialised at the mean of the response variable and the partial residuals Y^* are computed by subtracting to Y the tree prediction. Secondly, a linear random intercept effect model is fitted on Y^* and explanatory variables at the individual and group level. The corresponding partial residuals Y^{**} are obtained by subtracting to Y model predictions. These partial residuals Y^{**} are employed in the next step to fit a new tree, using the CART algorithm (Breiman et al., 1984) with a short depth. We iterate alternating the two fitting steps until convergence is reached. At the end of the procedure, model (2) is fitted by a linear random effect model using the partition associated with the tree selected at convergence. The leaf node parameters μ_m are estimated jointly with the other model parameters $\beta_0, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_u^2, \sigma_\varepsilon^2$.

The main difference of our procedure with respect to previous proposals (Hajjem et al., 2011; Sela, 2012), is the inclusion of the linear component $\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_j\boldsymbol{\gamma}$ in the random effect model (2). In the presence of quasi-linear relationships, this inclusion allows us to avoid overfitting and helps interpretation. Moreover, since the μ_m are jointly estimated in the final step, standard hypothesis tests and confidence intervals can be used for model selection and evaluation, together with the mean squared error computed on a test data set for prediction accuracy evaluation.

3. Application: Invalsi tests in Italian schools

We apply the TELM model outlined in the previous section to data on students' achievement collected by INVALSI. The Institute yearly carries out standardised tests to assess students' achievement in mathematics and reading and evaluate the overall quality of the educational offering of schools and vocational training institutes. See Arpino et al. (2019) for a discussion on this set of data.

As an illustration, we are here focusing on data on students who participated in the Maths tests at 5th and 8th grades. Specifically, the dataset is obtained by linking data on students who attended the 5th grade in 2013-2014 with data on students who attended the 8th grade in 2016-2017. The number of students who participated on both occasions of the Maths test is $409\,528$. They are grouped into $5\,773$ schools. We aim to predict the Maths test score, while understanding which of the included variables may be associate to the final score. Table 1 lists the considered explanatory variables. As shown in the table, we include both student level and school-level covariates, denoted in (1) as $\mathbf{X}_{ij}$ and $\mathbf{Z}_j$ respectively. Among the school level variables, we consider, in addition, the average of 5th grade Maths test and the average of the Socio-economic status index for each school. We are denoting these variables CM_MATH5 and CM_SES.

Table 1: Student and school level variables (INVALSI data years 2014 and 2017).

Student level variables (level one)	
MATH8	(*Response*) Test score at the 8th grade (0-100)
MATH5	Test score at the 5th grade (0-100)
SES	Socio-economic status
FEMALE	$1 = $ Yes, $0 = $ No
ENROLLED	School enrolment ($1 = $ Regualrly enrolled, $2 = $ Enrolled one year in advance, $3 = $ Enrolled one year later)
IMM	Citizenship ($0 = $Italian, $1 = 1^{st}$ generation immigrant, $2 = 2^{nd}$ generation immigrant)
School level variables (level two)	
AREA	Geographical area ($0 = $ NE, 5 categories)
TOWN	Provincial capital
CLSIZE	Average num of students per class
SCSIZE	Number of classes in the school
PUBLIC	Type of school ($0 = $ Private, $1 = $ Public)

The proposed model takes into account both linear and non-linear effects and can detect the presence of both within level and cross-level interaction effects. In particular, the tree component $T(\mathbf{X}_{ij}, \mathbf{Z}_j)$ in (1) is modelling non-linearities and interactions at once via a piece-wise linear function. Estimates for model parameters are reported in Table 2, while the tree component is also illustrated in Figure 1. The two terminal nodes without label in the plot have been automatically set in the reference category.

Individual and school level covariates not selected by the algorithm in the tree component have the usual interpretation. For example, controlling for the model covariates, females have, on average, around 1.5 points less than males in the score of math at the $8th$ grade.

Besides the usual interpretation of the coefficients of the linear components, it seems here interesting to focus on the covariates selected in the tree component of the model, namely the math score at grade 5 (MATH5) and the geographical area of the school (AREA). In particular, the tree component algorithm splits the values of MATH5 into three intervals: below 33 (2% of the observations), between 33 and 72 (55%) and above 72 (43%). Moreover, the algorithm

31

Table 2: TELM model fitted on INVALSI data: parameter estimates, standard errors and t-test.

	Estimate	Std. Error	t value
Student level			
(Intercept)	31.0733	0.9518	32.6462
MATH5	0.6263	0.0027	232.3463
SES	2.5246	0.0270	93.5909
FEMALE	-1.5021	0.0467	-32.1616
ENROLLED_2	1.8029	0.2053	8.7802
ENROLLED_3	-3.5558	0.2027	-17.5432
IMM_1	-1.1107	0.1779	-6.2434
IMM_2	-1.4869	0.1127	-13.1934
School level			
CM_MATH5	-0.2765	0.0131	-21.1820
CM_SES	1.2876	0.2260	5.6971
AREA_2 (NW)	0.4675	0.2574	1.8160
AREA_3 (Centre)	-1.9239	0.2562	-7.5080
AREA_4 (South)	8.1865	0.3559	22.9993
AREA_5 (Islands)	8.5293	0.3612	23.6133
CLSIZE	0.1629	0.0088	18.6063
SCSIZE	0.0613	0.0394	1.5562
PUBLIC	-2.1495	0.3773	-5.6971
TOWN	0.0275	0.1981	0.1388
Tree nodes			
N1: $33 \leq$ MATH5< 73 & AREA$= 4, 5$	-7.1138	0.2465	-28.8633
N2: MATH5 ≥ 73 & AREA$= 4, 5$	-11.9902	0.2806	-42.7304
N3: MATH5 ≥ 73 & AREA$= 1, 2, 3$	4.4819	0.0903	49.6540
N4: MATH5 < 33 & AREA$= 1, 2, 3$	1.9711	0.2691	7.3250
Residual variances			
School level (Intercept)		35.75	
Student level		218.35	
Number of students: 409528 Number of schools: 5773			

splits the schools into two groups depending on AREA: schools placed in North or Center Italy, and schools placed in South Italy and Islands. Thus, the algorithm suggests the presence of an interaction effect between these two variables, with the effect of AREA depending on the interval of MATH5 and vice versa. For example, for a pupil living in a region of NW of Italy, the expected difference with respect to a pupil with same characteristics living in the NE of Italy (baseline) is 2.4386 if MATH5< 33, it decreases to 0.4675 if $33 <$MATH5< 73, and it rises up to 4.9494 if MATH5 ≥ 73.

Note that the ordinary mixed effect regression model, whose parameter estimates are reported in Table 3, is nested with the TELM model. The Likelihood Ratio test comparing these two models obtains a test statistic equal to 10168, with 4 degrees of freedom, in favour of the TELM model. The variation between the estimates in the two models is due to the inclusion of the tree component, that relaxes the assumption of linearity and includes interaction effects. An interesting variation concerns the AREA coefficients estimates. Ignoring the AREA and MATH5 interaction, and the MATH5 non-linearity, completely reverse the main effect of AREA for South and Islands.
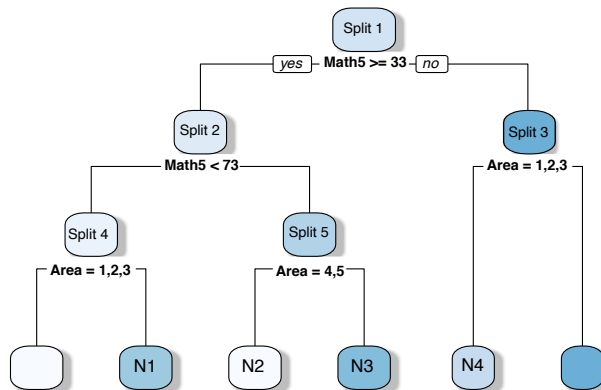
Figure 1: Graphical representation of the tree component of TELM model in Table 2 (nodes with a label correspond to a parameter in the model; the proportions of level 1 observations at each node are: left white node 0.35, N1 0.20, N2 0.17, N3 0.26, N4 0.01, right blue node 0.01)

Table 3: Random intercept model fitted on INVALSI data: parameter estimates, standard errors and t-test.

	Estimate	Std. Error	t value
Student level			
(Intercept)	36.2031	0.9438	38.3569
MATH5	0.6328	0.0015	412.2164
SES	2.5509	0.0273	93.3900
FEMALE	-1.6821	0.0473	-35.5926
ENROLLED_2	1.5694	0.2079	7.5486
ENROLLED_3	-3.5092	0.2052	-17.1025
IMM_1	-1.5243	0.1801	-8.4648
IMM_2	-1.8652	0.1140	-16.3553
School level			
CM_MATH5	-0.3240	0.0131	-24.8104
CM_SES	1.2809	0.2262	5.6620
AREA_2 (NW)	0.4945	0.2575	1.9203
AREA_3 (centre)	-1.9364	0.2564	-7.5529
AREA_4 (south)	-2.7964	0.2611	-10.7116
AREA_5 (islands)	-2.3695	0.2694	-8.7939
CLSIZE	0.1486	0.0089	16.7787
SCSIZE	0.0572	0.0394	1.4505
PUBLIC	-2.1165	0.3779	-5.6012
TOWN	0.1113	0.1982	0.5613
Residual variances			
School level (Intercept)		35.58	
Student level		223.91	
Number of students: 409528		Number of schools: 5773	

4. Conclusions

Tree Embedded Linear Mixed (TELM) models extend random effect models by including both a linear component and tree component in the regression function. The proposal increases the flexibility and the predictive ability of ordinary random effects models by handling simultaneously linear and non-linear associations and interactions.

A TELM model has the following characteristics: (1) it can handle clusters with different numbers of observations (unbalanced clusters); (2) it allows the inclusion of level 1 and level 2 covariates in the splitting process; (3) it allows observation-level covariates to have random effects. Besides, our proposal extends random effect regression trees in two directions: (i) incorporating a linear component in the final random effect model, and (ii) allowing to take into account contextual effects of level 1 covariates.

The application on INVALSI data is an illustrative example of TELM models that shows how the inclusion of a tree component helps highlight cross-level interactions.

References

Arpino, B., Bacci, S., Grilli, L., Guetto, R., Rampichini, C. (2019) Issues in prior achievement adjustment for value added analysis: an application to Invalsi tests in Italian schools. pp.17-20. In *ASA Conference 2019. Statistics for Health and Well-Being. Book of Short Papers -* ISBN:978-88-5495-135-8

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth, Belmont, (CA).

Hajjem, A., Bellavance, F., Larocque, D. (2011). Mixed Effects Regression Trees for Clustered Data. *Statistics and Probability Letters*, **81**, pp. 451–459.

Hajjem, A., Bellavance, F., Larocque, D. (2014). Mixed-effects Random Forest for Clustered Data. *Journal of Statistical Computation and Simulation*, **84**, pp. 1313–1328.

Miller, P.J., McArtor, D.B., Lubke, G.H (2017). METBOOST: Exploratory Regression Analysis with Hierarchically Clustered Data. arXiv:1702.03994v1 [stat.ML]

Sela, R.J., Simonoff, J.S. (2012). RE-EM Trees: A Data Mining Approach for Longitudinal and Clustered Data. *Machine Learning*, **86**, pp. 169–207.

Snijders, T.A.B., Bosker, R.J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd ed.)*. SAGE Publications Ltd.

Vannucci, G. (2019). Interpretable Semilinear Regression Trees. PhD Thesis, FLOrence REsearch repository.

Wermuth, N., Cox, D.R. (1998). On Association Models defined over Independence Graphs. *Bernoulli Society for Mathematical Statistics and Probability*, **4**, pp. 477–495.

Short-term and long-term international scientific mobility of Italian PhDs: An analysis by gender

Valentina Tocchioni, Alessandra Petrucci, Alessandra Minello

1. Introduction

Internationalization and globalization recently led to a large increase in high-educated and high-skilled international mobility. The increase in high-skilled mobility is also a consequence of the weakening of research and university systems of sending countries (the "brain drain" process), and the increase in skilled demand and improvements in higher education of host countries (the "brain gain" process; Boeri et al., 2012). At the micro-level, academic mobility has positive consequences on occupational prospects and careers of researchers, both in the short- and long- run (Ermini et al., 2019). For European researchers, experiencing scientific mobility is a way to advance their careers (Ackers, 2005; Mahroum, 2000; Morano-Foadi, 2005), but only a few studies focused on gender differences in opportunities for international scientific mobility (Deitch and Sanderson, 1987; Rosenfeld and Jones 1987; Mason et al., 2013; Cohen et al., 2019).

The literature suggests that women in academia tend to travel less (e.g., He et al., 2019), and especially those who are not in the humanities (Jöns, 2011). Family constraints, especially those related to childbearing and childrearing, have a stronger effect in reducing women's mobility than men's (Shauman and Xie, 1996). Due to the work-family conflict, women must be strongly determined and able to balance their professional and private lives for traveling during their academic careers (González -Ramos and Bosch, 2012). Moreover, for women in STEM (Science, Technology, Engineering and Mathematics), where the share of women is lower than in other fields of study, their performances (and hence, possibly, the chances of travelling) are much more hindered by personal events – mainly children (Ginther and Kahn, 2009). The conflict might be exacerbated in Italy, since the care responsibilities for women compared to men are higher than elsewhere in Europe: Italy is the European country (together with Romania) with the highest gender-gap in hours devoted to care during the day (Eurostat data, 2019), and it is below the European mean for the indicators of care in the European Gender Equality index (Eige, 2020). Despite it, the literature on Italy is missing on this topic.

Moving from these premises, our paper studies gender differences in short- and long-term international scientific mobility among a cohort of Italian PhDs. Moreover, we test whether these differences are diversely pronounced in female- or male- dominated fields of study, comparing the probability of moving for short- and long- periods abroad in humanities, soft- or hard- STEM (Bliglan, 1973a, 1973b).

Using Italian data on occupational conditions of PhDs collected in 2018 by Istat and modelling multinomial logistic regression analyses, we intend to deal with two research objectives. First, we aim to verify if female PhDs are associated with a lower scientific mobility irrespective their field of study. Second, we want to investigate the extent to which gender interacts differently in the humanities, soft- and hard-STEM in affecting the propensity of moving abroad after PhD qualification. We expect that women in STEM will be more penalized than women in the humanities with respect to men in the same fields. Also, the distinction between long-term and short-term mobility, which has been mainly neglected in the literature concentrating on longer stays, has taken into account across the two research objectives. In this respect, short-term mobility is a potentially high-value investment that may be pursued also by those researchers and scientists who cannot move for longer periods, such as women with caring

responsibilities (Henderson, 2019). For this reason, we expect a lower gender gap in mobility among short-term stays in comparison with long-term stays and (potential) international relocations.

In the literature, it is acknowledged that an experience abroad during early career may have positive effects on future occupational prospects. With our work, we intend to shed light on potential disparities on moving abroad that may exist among researchers in their early career by gender, and which could contribute to leave women behind in academia.

2. Data and methods

Our sample was drawn from the Istat Survey on occupational conditions of Italian PhD holders, conducted in 2018 by contacting all PhD holders who had obtained their qualification from an Italian academic institute in 2012 and 2014. After excluding foreign PhD holders (625) and those who declared to have moved abroad because of personal or familiar reasons, our final sample was formed by 15,216 observations[1]. Among them, 3,313 (21.8%) spent a period of at least three months abroad after their PhD dissertation: 799 PhDs (5.3% of PhD holders) stayed less than one year (short-term stays); 1,016 (6.7%) moved for one year or more (long-term stays); 1,498 (9.8%) were still abroad at the interview date[2] (potential international relocations); and 11,903 (78.2%) did not move.

To investigate our two research objectives, we estimated two multinomial logistic regression models, with standard errors clustered at the field of study. The response variable was a nominal variable that indicated whether the researcher remained in Italy after doctoral studies (1), or if they went, whether they moved for less than one year (2), for one year or more (3), or if they were still abroad at the interview date (4). The two key explanatory variables were student gender and the field of study, with three categories: Hard STEM; Soft STEM; and Humanities[3,4].

In our first step, in order to verify if female researchers are associated with a lower mobility irrespective their field of study, Model 1 estimated the probability of going abroad in one of the three different situations or remaining in Italy according to gender, field of study and some control covariates: parental education[5] (the highest educational level between parents, assuming the following categories: primary or lower; lower secondary; upper secondary; tertiary or post-tertiary), mother's economic activity (employed/self-employed; homemaker; retired; other condition), father's social class, classified according to EGP-class typology aggregated in a five-category classification (Goldthorpe & Erikson, 1992: higher grade professionals; lower grade professionals; routine non-manual labourers; self-employed; working class - skilled/unskilled; and a residual sixth category for those whose social class was unknown); if the researcher completed his/her PhD studies at a university outside his/her region of residence;

---

[1] PhD holders who completed the interview in 2018 were 16,057 (72.7% of all 22,099 PhD holders who defended in 2012 and 2014, which were contacted by Istat for the interview).

[2] Unfortunately, we don't know when PhDs moved during the years intercurred between the defence and the interview (this period lasted four years for those who defended in 2014, and six years for those who defended in 2012). For this reason, we opted for keeping them separated from the other two categories of short-term and long-term stays, which were defined on the basis of a specific amount of time. Moreover, we referred to this kind of mobility as "potential international relocation", because researchers could be abroad at the interview date only for a fixed amount of time.

[3] Hard STEM comprises Maths and Computer Science; Physics; Chemistry; Civil Engineering and Architecture; Industrial and Information Engineering; Soft STEM includes Earth Science; Biology; Medicine; Agricultural and Veterinary Science; Economics and Statistics; and Humanities comprises Antiquity, Philology, Literary Studies, Art History; History, Philosophy, Pedagogy, Psychology; Law; Political and Social Sciences.

[4] The percentage of women was 37.7% in the Hard-STEM, 60.1% in the Soft-STEM, and 59.9% in the Humanities.

[5] The three covariates related to the family of origin (parental education, mother's economic activity and father's social class) referred to when the researcher enrolled to the university for the first time.

the calendar year of PhD dissertation (2012 or 2014); if the researcher spent an international visiting period during PhD studies.

In the second step, Model 2 included also an interaction term between gender and field of study, for verifying if and how the field of study moderates the relationship between gender and international mobility.

3. Results

We estimated predicted probabilities of researchers' international mobility and present them graphically (full model results are available upon request to the authors). Figure 1 shows predicted probabilities of moving/not moving abroad after PhD studies according to gender and length of the stay from Model 1. Predicted probabilities show how female researchers' propensity to move is always significantly lower than their male counterparts, irrespective the length of stay. Overall, the difference between male and female researchers' propensity to mobility is about 7.8% (see Figure 1d). Looking at the three types of move, the highest propensity is of those who moved up to the interview date for both men and women, with 10.4% and 6.6%, respectively (see Figure 1c). As expected, the gender gap in the propensity towards international mobility is positively associated with the length of stay: whilst the difference in the predicted probability of moving abroad is only 1.2% between men and women for short-term stays, it raises to 2.8% for long-term stays and 3.8% for potential international relocations.

Figure 1: Results from Model 1: Predicted probabilities of moving/not moving abroad after PhD studies according to gender. CI 83%.
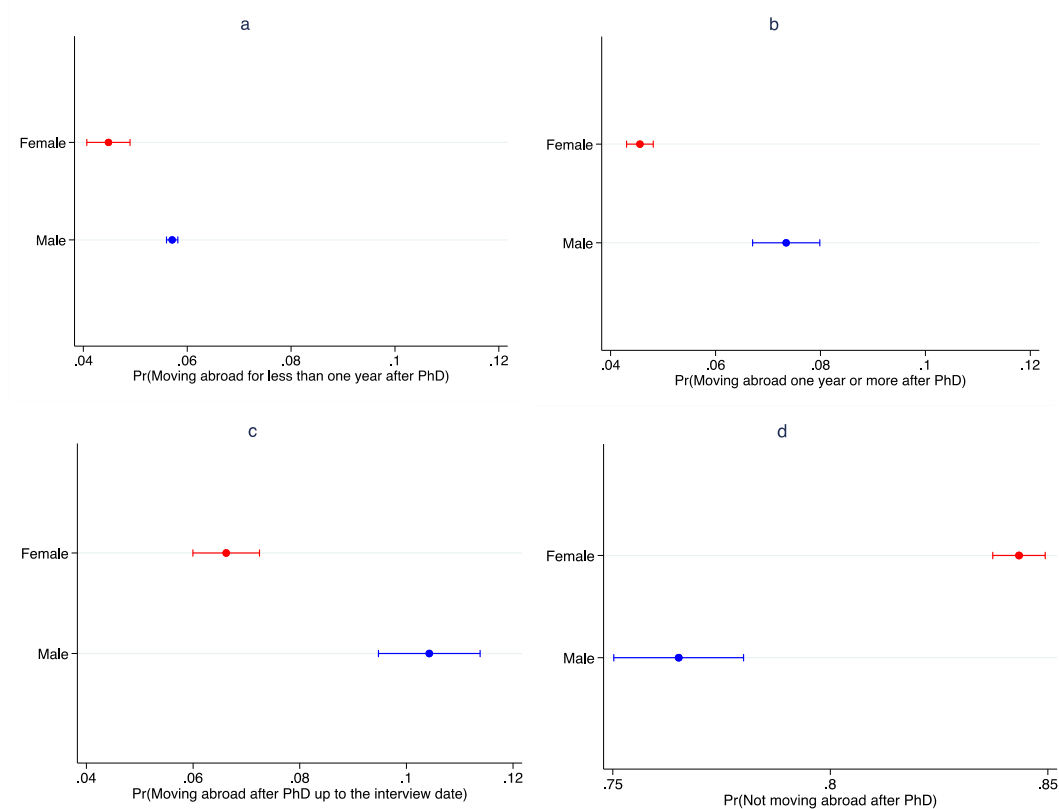
Figure 2 shows predicted probabilities of moving/not moving abroad after PhD studies according to gender and field of study from Model 2. Overall, male researchers' propensity to move is still higher than female researchers' propensity in all fields of study, and the highest gap is among those researchers who have a PhD in the Hard STEM: in this field of study, men who move abroad are 10.8% more than women, whereas this difference shrinks to 3.8% for those in the Humanities (see Figure 2d for complementary percentages). Moreover, researchers in the Hard STEM are also the ones with the highest mobility: whilst male and female researchers who moved were 31.3% and 20.3% in this field of study, respectively, these percentages decrease to 18.3% and 14.5% among researchers in the Humanities.

According to the three types of move, confidence intervals of predicted probabilities show how male researchers' propensity to move is still higher than female researchers' propensity in all combinations of field of study and length of stay, except for researchers in the Hard STEM who moved for short-term periods with the two confidence intervals overlapping (see Figure 2a). Nevertheless, researchers in the Hard STEM have a higher propensity for longer stays (both long-term stays and potential international relocations), and the gender gap in mobility is significant and the highest across all fields of study: 4.2% for and 6.3%, respectively (see Figure 2b and 2c). On the other hand, researchers in the Humanities have a higher propensity for short-term stays abroad (see Figure 2a), whereas researchers in the Soft STEM show similar percentages, with only a slightly higher propensity for potential international relocations. Differences in gender gap are very low both in the Humanities - from 0.8% to 1.9% in the different types of move - and in the Soft STEM – where the gender gap is around 1.4%-1.5% across all types of move (see Figure 2a-c).

Figure 2: Results from Model 2: Predicted probabilities of moving/not moving abroad after PhD studies according to gender and field of study. CI 83%.
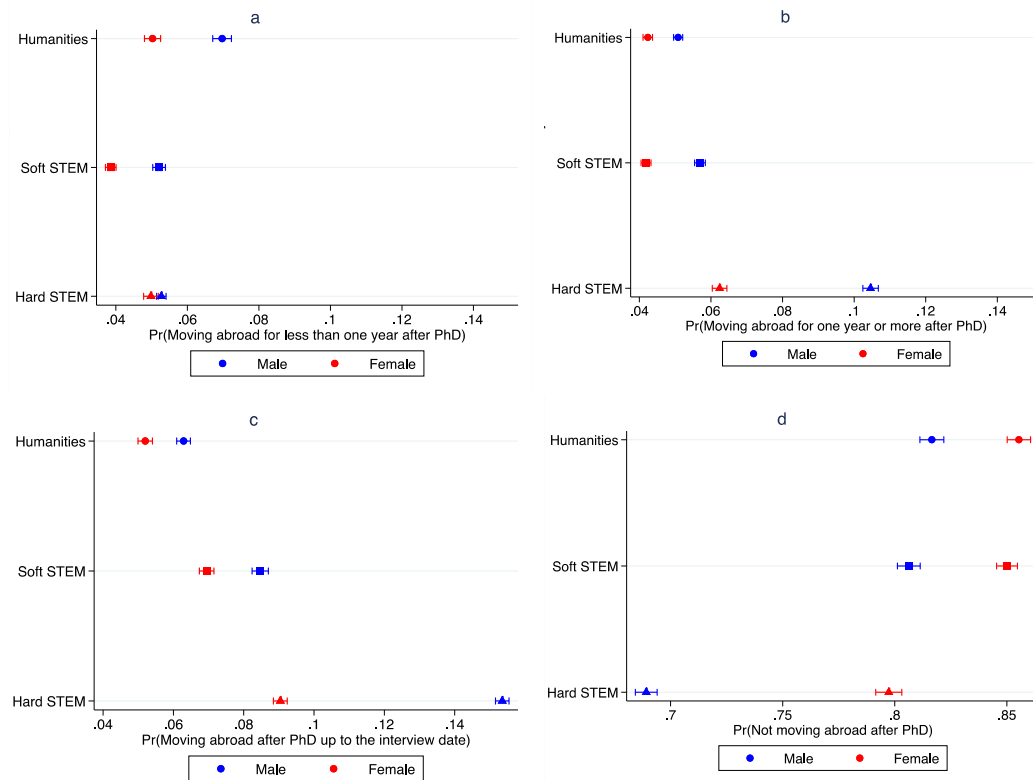
4. Conclusions

International mobility of high-educated people and researchers has positive consequences on their occupational prospects and careers, both in the short- and long- run (Ermini et al., 2019). Despite it, women in academia have a lower mobility with respect to their male counterparts, experiencing more often work-family conflicts that tend to limit their traveling during their academic careers (González -Ramos and Bosch, 2012; Jöns, 2011). In this paper, we concentrated on gender differences in short- and long-term international scientific mobility among a cohort of Italian PhDs, and the potential role of moderator played by the field of study in the relationship between gender and international mobility.

Our analyses show how women with a PhD qualification have a lower propensity to mobility compared with their male counterparts. As expected, a lower gender gap in mobility emerges among short-term stays in comparison with long-term stays and potential international relocations. In this respect, it is acknowledged that short-term mobility is presumably an investment that may be pursued also by those researchers who cannot move for longer periods, which are more often women (e.g. Henderson, 2019). Concentrating on the field of study, as expected the highest gender gap in international mobility is among women and men in hard-STEM, whereas the lowest among those researchers in the Humanities. As identified for other aspects in previous literature, to bridge the gap in hard-STEM is more difficult than in other fields of study, where the presence of women is much more pronounced (e.g., Ginther and Kahn, 2009). Nevertheless, a remark should be made. International mobility of female researchers in hard-STEM seems to be the highest among the three fields of study. Thus, a higher gender gap in international mobility in the hard-STEM could depend - at least partly - from the higher overall mobility of those researchers, and in particular that of men. In this respect, hard-STEM appears as the field of study where international mobility is more widespread, at least in Italy, and it could reveal a greater difficulty in accessing scientific research and academia positions for Italian researchers in this field of study.

Gender disparities in academia can be found in several outputs such as publications (namely, men publish more papers than women, on average: West et al. 2013), career advancement, with women having slower and more complex pattern of career (Gaiaschi and Musumeci 2020) and, as we demonstrated, in the chances to experience international short- and long- term mobility. For this final output, more than for the others there might be a direct effect of the family-work conflict. Women might be less likely to travel due to the difficulties to balance their career and their family duties. This aspect deserves further investigation. Moreover, we demonstrated that international mobility is another way to leave women behind. The direct effect of this gap on careers of women in the Italian academia should be the focus of future research.

References

Ackers, L. (2005). Moving people and knowledge: Scientific mobility in the European Union. *International migration*, **43**(5), pp. 99–131.

Biglan, A. (1973a). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, **57**, pp. 195-203.

Biglan, A. (1973b). Relationships between subject matter characteristics and the structure and output of university departments. *Journal of Applied Psychology*, **57**, pp. 204-213.

Deitch, C.H., Sanderson, S.W. (1987). Geographic constraints on married women's careers. *Work and Occupations*, **14**(4), pp. 616–634.

Boeri, T., Brücker, H., Docquier, F., Rapoport, H. (2012). *Brain drain and brain gain: The global competition to attract high-skilled migrants*. Oxford University Press, Oxford, (UK).

Cohen, S., Hanna, P., Higham, J., Hopkins, D., Orchiston, C. (2020). Gender discourses in academic mobility. *Gender, Work & Organization*, **27**(2), pp. 149-165.

Eige (2020). Gender Equality Index: Italy. Retrieved here: https://eige.europa.eu/publications/gender-equality-index-2020-italy

Ermini, B., Papi, L., Scaturro, F. (2019). Wage returns to interregional mobility among Ph.D graduates: Do occupations matter?. *Papers in Regional Science*, **98**(2), pp. 995–1025.

Eurostat (2019). How do women and men use their time. Retrieved here: https://ec.europa.eu/eurostat/statistics-explained/index.php/How_do_women_and_men_use_their_time_-_statistics

Gaiaschi, C., Musumeci, R. (2020). Just a matter of time? Women's career advancement in neo-liberal academia. An analysis of recruitment trends in Italian Universities. *Social Sciences*, **9**(9), p.163.

Ginther, D. K., Kahn, S. (2009). Does science promote women? Evidence from Academia 1973-2001, in *Science and engineering careers in the United States*, eds. R.B. Freeman, D.F. Goroff, University of Chicago Press, Chicago, (IL), pp. 163–194.

Goldthorpe, J., Erikson, R. (1992). *The constant flux: a study of class mobility in industrial societies.* Clarendon Press, Oxford (UK).

González Ramos, A.M., Bosch N.V. (2013). International mobility of women in science and technology careers: shaping plans for personal and professional purposes. *Gender, Place & Culture*, **20**(5), pp. 613-629.

He, Z., Zhen, N., Wu, C. (2019). Measuring and exploring the geographic mobility of American professors from graduating institutions: Differences across disciplines, academic ranks, and genders. *Journal of Informetrics*, **13**(3), pp. 771-784.

Henderson, E. F. (2019). A PhD In motion: Advancing a critical academic mobilities approach (CAMA) to researching short-term mobility schemes for doctoral students. *Teaching in Higher Education*, **24**(5), pp. 678–693.

Jöns, H. (2011). Transnational academic mobility and gender. *Globalisation, Societies and Education*, **9**(2), pp. 183-209.

Loeb, J.W. (2006). The status of female faculty in the U.S.: thirty-five years with equal opportunity legislation. *Management Revue*, **17**(2), pp. 157-180.

Mahroum, S. (2000). Scientists and global spaces. *Technology in Society*, **22**(4), pp. 513–523.

Mason, M.A., Wolfinger, N.H., Goulden, M. (2013). *Do babies matter? Gender and family in the ivory tower*. Rutgers University Press, New Brunswick, (NJ), and London (UK).

Morano-Foadi, S. (2005). Scientific mobility, career progression, and excellence in the european research area. *International Migration*, **43**(5), pp. 133–162.

Rosenfeld, R., Jones, J. (1987). Patterns and Effects of Geographic Mobility for Academic Women and Men. *The Journal of Higher Education*, **58**(5), pp. 493-515.

Shauman, K. A., Xie, Y. (1996). Geographic Mobility of Scientists: Sex Differences and Family Constraints. *Demography*, **33**(4), pp. 455-468.

West, J. D., Jacquet, J., King, M., Correll, S. J., Bergstrom, C. T. (2013). The Role of Gender in Scholarly Authorship. *PLoS ONE*, **8**(7), pp. e66212.

Measuring logical competences and soft skills when enrolling in a university degree course

Bruno Bertaccini, Riccardo Bruni, Federico Crescenzi, Beatrice Donati

1. Introduction

Logical abilities are a ubiquitous ingredient in all those contexts that take into account soft skills, argumentative skills or critical thinking. However, the relationship between logical models and the enhancement of these abilities is rarely explicitly considered. Two aspects of the issue are particularly critical in our opinion, namely: (i) the lack of statistically relevant data concerning these competences; (ii) the absence of reliable indices that might be used to measure and detect the possession of abilities underlying the above-mentioned soft skills. This paper addresses both aspects of this topic by presenting the results of a research that we conducted in between October and December 2020 on students enrolled in various degree courses at the University of Florence. To the best of our knowledge, this is the largest available database on the subject in the Italian University System to date[1] . It has been obtained by a three-stage initiative. We started from an "entrance" examination for assessing the students' initial abilities. This test comprised ten questions, each of which was centered on a specific reasoning construct. The results we have collected show that there is a widespread lack of understanding of basic patterns that are common in the everyday way of arguing. Students then underwent a short training course, using formal logic techniques in order to strengthen their abilities, and afterwards took an "exit" examination, replicating the structure and the questions difficulty of the entrance one in order to evaluate the effectiveness of the course. Results show that the training was beneficial.

2. Data and methods

The "entrance" test was administered to 272 students in October 2020. The short training course was scheduled in November 2020 and was not compulsory. This characteristic and the students' overall difficulties in self-organizing their study time during the health emergency due to the COVID pandemic have led to fewer "exit" exams (67). The data collected through the two exams were used to: a) estimate initial logical abilities of students engaged in a university experience; b) obtain an evaluation of the effectiveness of the short training course by comparing the abilities measured before and after attending the course itself. Both the "entrance" and "exit" exams we scheduled have the same structure in terms of type (logical constructs), number (10, one per construct), and questions difficulty. The considered logical constructs are: *Double negation* (item code N); *Disjunction negation* (item code D); *Conjunction negation* (item code C); *Hypothetical reasoning* (item code IMPL); *Sufficient and necessary conditions* (item code NEC); *Negation of the universal quantifier* (item code NU); *Negation of the existential quantifier* (item code NE); *Modus tollens* (item code MT); *Syllogism* (item code S); *Multiple steps*

---

[1]We were unable to find traces in the literature of other datasets on the topic available among other Italian universities

Bruno Bertaccini, University of Florence, Italy, bruno.bertaccini@unifi.it, 0000-0002-5816-2964
Riccardo Bruni, University of Florence, Italy, riccardo.bruni@unifi.it, 0000-0003-2695-0058
Federico Crescenzi, University of Florence, Italy, federico.crescenzi@unifi.it, 0000-0002-0701-4398
Beatrice Donati, University of Florence, Italy, beatrice.donati@unifi.it, 0000-0002-4707-8476

deduction (item code DED). These constructs correspond to what are in our experience ten of the most recurring errors made by undergraduate students. These errors have been identified in many years of teaching experience but also on the basis of the logical tradition that identifies some constructs underlying our way of reasoning. Each close-ended question (item) presents 4 answers, only one of which is true. 1 point was awarded for a correct answer, no points were assigned to missing or wrong answers. We are confident that this framework could be a good method for measuring logical abilities of students. This hypothesis is at the basis of the Item Response Theory (IRT).

Item Response Theory (IRT) is a methodology to investigate the relationship between an individuals' response to an item of a test on an overall measure of the ability that the item was intended to measure (Demars, 2010; Bartolucci et al., 2016). Knowing the item difficulty is useful when building tests to match the trait levels of a target population. For these reasons, IRT has been used proficiently either to score tests or surveys and in test development/assessment (Chen et al., 2005; Lee et al., 2008).

In presence of binary data - as those just described, that typically correspond to a set of n individuals that give wrong or correct responses to a set of items of a test/questionnaire, the main assumptions of IRT models are: unidimensionality (for each individual i who underwent the test, the responses given to the whole set of items depend on the individual ability θ_i), local independence (for each individual, the given responses are independent given the individual ability θ_i) and monotonicity (the conditional probability of responding correctly to a certain item j, known as Item Characteristic Curve, is a monotonic non-decreasing function of θ_i).

At the core of all the IRT models is the item response function (IRF). The IRF expresses the probability of getting the item j "correct" (i.e. $Y_{ij} = 1$) as a function of item characteristics and the individual's latent (i.e. unobserved) trait/ability level θ_i. In IRT literature, we distinguish between one-parameter (known also as the Rasch model), two-parameters and three-parameters logistic IRT models. Intuitively, each model extends the previous one with an additional parameter. The IRF for the three-parameters (3PL) model is:

$$P(Y_{ij} = 1|a_j, b_j, c_j, \theta_i) = c_j + (1 - c_j)\frac{e^{a_j(\theta_i - b_j)}}{1 + e^{a_j(\theta_i - b_j)}} \tag{1}$$

This function describes the probability for an individual with latent ability θ_i to endorse an item j where b denotes the item difficulty, a denotes the item discrimination and c is a parameter for guessing). Under this general configuration, higher difficulty estimates indicate that the item is harder (i.e., higher latent ability to answer correctly), and higher discriminability estimates indicate that the item has better ability to tell the difference between different levels of ability θ. Moreover, individuals with zero ability have a nonzero chance of endorsing any item, just by guessing randomly. For the sake of completeness, the guessing parameter c is not involved in the two parameters logistic (2PL) IRF function, while both the guessing parameter c and the discrimination parameter a are not involved in the one-parameter logistic (1PL, also known as Rasch) model. As usual in IRT modelling, if a parametric model for the ability distribution is not assumed, then the usual two-parameters and three-parameters logistic models present identifiability problems not encountered with the 1PL model (Haberman, 2005). These problems could be solved by imposing substantial constraints such as assuming that the ability latent trait follows a standard normal distribution. Otherwise it is possible to constrain the discriminating parameter of a reference item (usually the first one) to 1 and its threshold difficulty parameter to 0, leaving free the mean and the variance of the ability distribution (still expected normally shaped) (see Bartolucci et al., 2016). Software to estimate such class of models is available for R in the library Ltm (Rizopoulos, 2006).

All logistic IRT models were applied to our data looking for the best parameter estimations, i.e. the most reliable fitting. Results will be presented in the next section.

3. Results

The estimation of the logical abilities of students who undertook the entry test was the first objective of this work. Starting from the simplest one, we have applied all three the IRT logistic models presented above to the data collected administering the "entrance" test. Figure1 shows the Item Characteristic Curves (ICCs) and Test Information Functions respectively from the Rasch, the 2PL and the 3PL models.

Figure 1: Entrance test: Item Characteristic Curves and Test Information Functions obtained after estimating the whole class of logistic IRT models
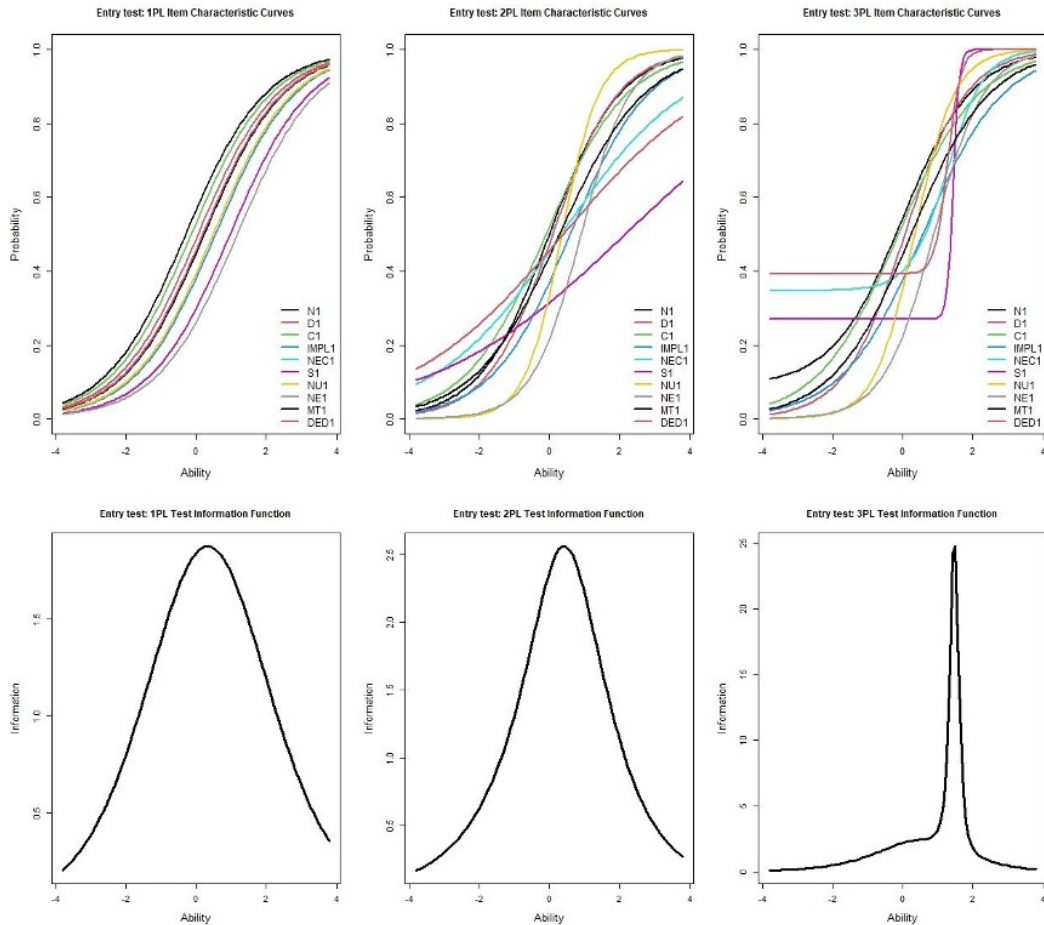

```
Likelihood ratio test
Model 1: 1PL
Model 2: constrained 2PL
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  11 -1755.9
2  18 -1743.7  7 24.364  0.0009831 ***
```

```
Likelihood ratio test
Model 1: constrained 2PL
Model 2: constrained 3PL
  #Df  LogLik Df  Chisq Pr(>Chisq)
1  18 -1743.7
2  28 -1735.4 10 16.623    0.08312 .
```

In particular, the test information functions reported in the bottom panel of Figure 1 are simply the sum of the first derivatives of the ICCs (also called Item Information Curves) in the top panel. Ideally, a good test/questionnaire should provide a good coverage of a rather wide range of latent ability levels. In this case, the information curve should be normally shaped and centred around zero. Otherwise, the test may identify a limited range of ability levels. The 1PL information curve, although centred on a value slightly greater than zero, showed a satisfactory coverage of the range of the possible abilities. Nevertheless, from the analysis of the 1PL model item-fit statistics (here not reported due to lack of space) we observed that 3 items might not fit the 1PL model so well. Also the Likelihood Ratio Test statistic (LRT, presented below Figure 1) suggests an upgrade to the 1PL model.

The 2PL model is more suitable than the Rasch one for describing our data (from the item-fit statistics only one item might still not be in line with the model). The item 2PL ICCs shows that some items provide more information about latent ability for different ability levels. In general, the higher the estimate of the item discriminability the higher the item's capability to provide information about ability levels around the point where there is a 50% chance of getting the item right (i.e. the steepest point in each ICC slope). Instead, the LRT statistic did not provide us with sufficient evidence in favour of the 3PL model (its information curve is quite far from normality), although this model is able to show how the students have tried to guess the answers of the 3 more difficult items (corresponding to NEC, S and DED logical constructs). Individual abilities were then estimated through the 2PL model.

Figure 2: Entrance test ability distribution: students taking just the "entrance" test (red) and student having taken also the "exit" test (green).
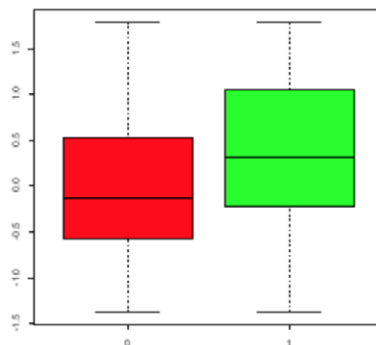
More stable results were obtained limiting the analysis of "entrance" test responses to those students who underwent the short training course and took also the "exit" test. Figure 2 shows the differences in the distribution of abilities of the entrance test respectively for those students who took only that test (red) and those students who also took the exit one (green). The application of the 2PL model to this reduced dataset (see Figure 3a for the estimated Item Information Curves) produced item-fit statistics whose p-values gave no evidence of incoherent or misfitting items. Moreover, there was no evidence against the hypothesis that we were measuring only a

single latent trait (hypothesis of unidimensionality). Interestingly, some ICCs show a different level of information in ability before and after the training course. As these are estimated by the response patterns given by all students who attended the two tests, a plausible reason of this change could lie in the fact that taking the course may have changed the attitude towards the understanding of some constructs. Of course, there may still be a source of randomness in responses because no penalty was assigned in case of incorrect response. The abilities estimated for this subset of students followed an almost perfect standard normal distribution (see Figure 4a).

To obtain an evaluation of the effectiveness of the short training course by comparing the abilities measured before and after attending the course itself, we estimated the 2PL model also on responses related to the "exit" test (see Figure 3b and Figure 4b respectively for the estimated Item Information Curves and the distribution of the estimated individual's latent ability). The comparison should be done at individual level to obtain an estimate of the course effect on students' logical abilities. Unfortunately, abilities estimated by the two models are standardized and, consequently, incomparable. The only way to solve this issue is to resort to some test equating techniques. Test equating is a statistical procedure to ensure that scores from different test forms can be compared and used interchangeably. There are several methodologies available to perform equating, some of which are based on the Classical Test Theory (CTT) framework and others are based on the Item Response Theory (IRT) framework (Gonzalez and Wiberg, 2017). Within the IRT framework, if each test form is performed independently or separately in time, their respective parameters will be on different scales and thus incomparable. Equating coefficients solves this problem by transforming the item parameters so that they are all on the same scale. In particular, in this work the abilities estimated with the "entrance" test were transformed to the scale of the "exit" form with the direct equating mean-mean method. Other popular IRT methods for equating pairs of test forms are the mean-sigma, Stocking-Lord and Haebara (Kolen and Brennan, 2014).

Figure 3: Tests underwent by students who completed the short training course: Item Information Curves of the "entrance" test (panel a) and "exit" test (panel b)
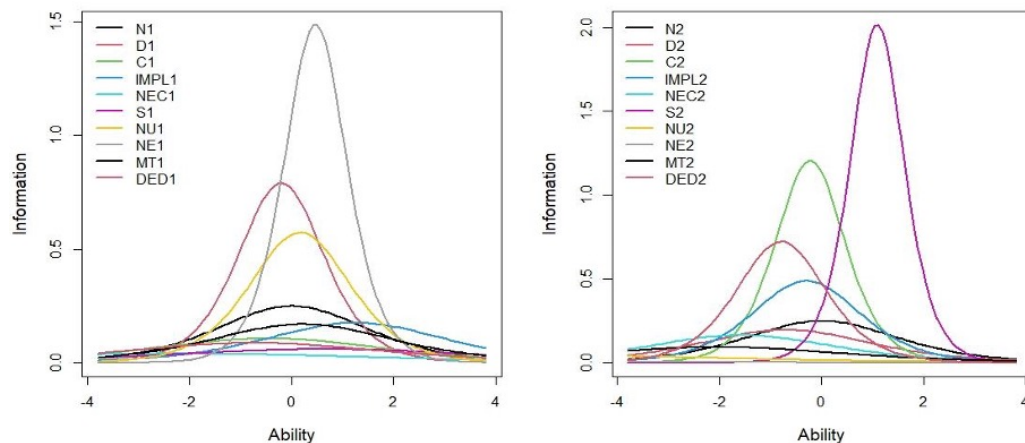
We performed the comparison using the "equateIRT" library developed for the R environment for statistical computing (Battauz, 2015). The course effect was thus estimated with a paired sample t-test for differences in abilities. The average difference of 2.07 in the ability estimated before and after taking the course confirms the validity and effectiveness of the programmed training course in incrementing logical abilities of academic students.

Figure 4: Tests underwent students who completed the short training course: distribution of the estimated individualâs latent ability for the "entrance" test (panel a) and "exit" test (panel b)
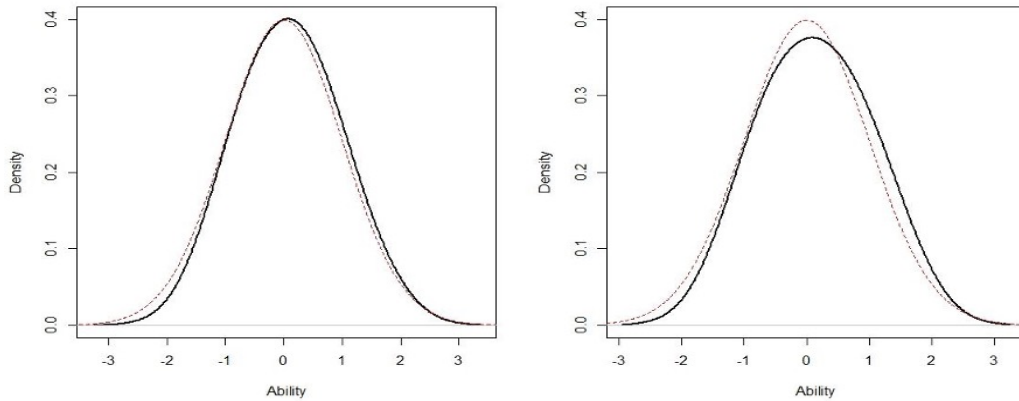
4. Conclusions

In this paper we presented the results of a research concerning the logical abilities of students enrolled in various degree courses at the University of Florence. This is the first study of this kind and this preliminary data analysis is already very promising and will help us phrasing the test items and refine the entire process. Looking at the data we can already confirm that the "entrance" test results are significant. This convinced us to strongly advise our University to design an internal policy that may become standard, testing all students and providing a mandatory logic course if their ability is below a certain threshold.

We wish to thanks Prof. Sandra Furlanetto of the University of Florence for giving life to this interesting project and providing us with the related data.

References

Bartolucci, F., Bacci, S., Gnaldi, M. (2016). *Statistical analysis of questionnaires: a unified approach based on R and Stata.* Chapman & Hall, CRC Press, Boca Raton (FL).

Battauz, M. (2015). EquateIRT: An R Package for IRT Test Equating. *Journal of Statistical Software*, **68** (7), pp 1–22.

Chen, C.M., Lee, H.M, Chen, Y.H. (2005). Personalized e-learning system using item response theory. *Computers & Education*, **44** (3), pp. 237–255.

DeMars, C. (2010). *Item response theory.* Oxford University Press, Oxford (GB).

Gonzales, J., Wiberg, M. (2017). *Applying Test Equating Methods - Using R.* Springer, Cham (CH).

Haberman, S. J. (2005). *Identifiability of Parameters in Item Response Models With Unconstrained Ability Distributions.* Educational Testing Service, Princeton, NJ. Research Report.

Kolen, M. J., Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices (3rd ed.).* Springer-Verlag, New York.

Lee, Y., Palazzo, D. J., Warnakulaso oriya, R., Pritchard, D. E. (2008). Measuring student learning with item response theory. *Physical Review Special Topics - Physics Education Research*, **4**(1)

Rizopoulos D. (2006). Ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software*, **17** (5), pp 1–25.

Innovation, productivity and welfare

A bibliometric study of global research activity in relation to the use of partial least squares for policy evaluation

Rosanna Cataldo, Laura Antonucci, Corrado Crocetta,
Maria Gabriella Grassia, Marina Marino

1. Partial Least Squares for policy evaluation

Structural Equation Modeling (SEM), especially Partial Least Squares - Path Modeling (PLS-PM) has become a mainstream method in many fields of research. Indeed, PLS-PM has been used in the social and behavioral sciences, rooted in psychometrics and in the literature on causal modeling. In the last few years it has been increasingly disseminated in a variety of disciplines, and, in particular, has been extensively used in the business and management sciences. Within these research projects, PLS-PM has been applied successfully in studies concerning the measurement of intangibles like customer and employee perceptions (e.g. satisfaction, motivation and loyalty). These kinds of model are becoming crucial to managers in order to improve their decision making processes and increase their organization's profitability. In every time and place the decision making process has always been complex. Generally, it applies evaluation principles and methods to examine the content, implementation or impact of a policy or a decision. In the last few years, researchers have been promoting statistical methods such as SEM and PLS-PM for the evaluation of policies, especially in the context of decision making. In the literature, empirical approaches which have applied PLS-PM to decision making have been identified through a systematic literature search. To better understand and characterize this trend, a bibliometric study of international papers on this subject has been developed in order to describe the use of SEM and PLS-PM approaches in policy evaluation during the last twenty years.

2. Study Methodology

A bibliometric analysis has been used to analyse the trends in the field of SEM in the context of decision making. Bibliometric analysis is a quantitative approach for the analysis of academic literature using bibliographies to provide the description, evaluation and monitoring of the published research (Garfield et al., 1964); (White and McCain, 1989). The methodological aim is to analyse publications, citations and sources of information (Rodriguez-Soler et al., 2020). Bibliographic data are processed through a workflow: study design, data collection, data analysis, data visualization and interpretation. The analysis has been performed using the Bibliometrix R-Tool (Aria and Cuccurullo, 2017), a recent R-package which facilitates a more complete bibliometric analysis employing specific tools for both bibliometric and scientometric quantitative research. The Bibliometrix R-package (http://www.bibliometrix.org) provides a set of tools for quantitative research in bibliometrics and scientometrics, supporting scholars in three key phases of analysis: 1) data importing and conversion to the R format; 2) bibliometric analysis of a publication dataset; and 3) building matrices for co-citation, coupling, collaboration and co-word analysis. The R program and the bibliometrix codes have been used to produce a descriptive bibliometric analysis and to construct the matrices. In addition, "biblioshiny" (Aria and Cuccurullo, 2017), a Bibliometrix web-interface, has been used to build a

conceptual map and network for co-citation. Matrices are the input data for the performance of network analysis, multiple correspondence analysis and certain data reduction techniques (Aria and Cuccurullo, 2017).

3. Data Collection

With the aim of understanding how the research on SEM and PLS-PM issues has evolved, the data were retrieved from two main databases commonly used by researchers: Scopus and Web of Science (WoS). Scopus and WoS are the world's most trusted independent global citation database. They are recognised as covering a broad range of relevant journals and peer-reviewed articles of high quality (Skute et al., 2019). These databases have already been used in bibliometric analysis in different disciplines, sometimes individually, in the case of WoS (Diem and Wolter, 2013); (Falagas et al. , 2006) and Scopus (Maharana, 2013); (Morandi et al., 2015), and sometimes in combination (Rodriguez-Soler et al., 2020).

We extracted articles published between 2000 and 2020 (incl.) which contained the topic "decision making" with the following keywords in the title or abstract: "PLS-PM"; "PLS Path modeling"; "PLS-Path modeling"; "SEM-PLS" ("decision making" AND "PLS-PM" OR "PLS Path modeling" OR "PLS-Path modeling" OR "SEM-PLS"). The data were downloaded on December 5, 2020. Only articles, reviews, proceedings papers and book chapters were included, with document types such as editorials, notes and corrections were excluded from the study. By merging the Scopus and WoS databases, 93 duplicate documents were removed. This process resulted in a final sample of 451 articles, which constitute the core material of this study, relating to 1,308 authors and 323 sources.

4. Analysis and Discussion

In the analysis of the data, a descriptive analysis was initially performed. Next, bibliometric techniques were developed using conceptual, intellectual or social networks.

Figure 1 shows the growth of publications from 2000 to 2020.
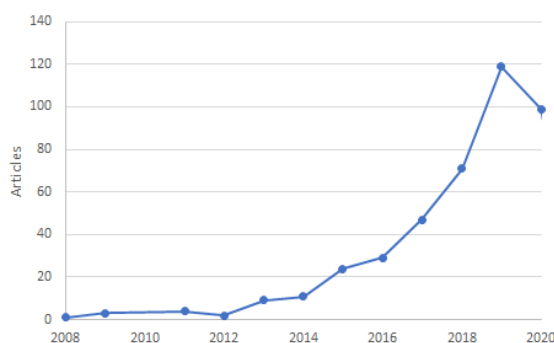
Figure 1: Growth trajectory of the literature relating to the use of PLS-PM in decision making, 2008 - 2020

As can we see, the first studies dealing with issues related to decision making were carried out in 2008. In the first years of the analysis (2008 - 2012) the number of publications is very low, emphasizing the fact that the topic was probably not very well developed and addressed by researchers. In 2019 we see a peak in the number of publications relating to the PLS-PM

approach as a statistical methods in the context of decision making.

Concerning the sources, the distribution of the articles does not present any significant concentration. The journals which included the most frequently quoted articles, containing "decision making" and "PLS-PM" as keywords, are presented in Figure 2, with the largest number of articles, namely 17, published in the *Sustainability* journal, followed by the *International journal of environmental research and public health*, a journal which deals with issues related to environmental health sciences and public health, measurement and monitoring models.

Figure 2: The most relevant sources

As regards provenance, the research activity of countries in terms of their publication output on this theme was examined. Figure 3 shows the top 20 most productive countries in terms of publication output and scientific collaborations during the period 2008-2020. In particular, the left-hand side of the Figure 3 shows the number of articles produced by the authors of different countries and the rate of cooperation of each country's authors with those of other countries, while the right-hand side of the Figure 3 shows the cooperations and networking among researchers working on and studying this subject in different countries.

Figure 3: Country production and country collaboration networks

The authors who have distinguished themselves in terms of the number of publications related to this topic come mainly from China, Malaysia, the USA and Italy. The authors from China and Malaysia have produced the same numbers of articles (35) (not shown here), but the rate of Chinese authorship with other countries is about 46% while the rate of MCP (Multiple Country Pubblications) of Malaysia is 17%. This demonstates that Chinese authors collaborate

extensively with authors from other countries. Italian authors ranked third with 30 papers, and the authorship rate for contributing articles to other authors from other countries is 27.6%. As we can see in the ranking, Italy is the European country that has most significantly increased its publication output in relation to policy evaluation in recent years, indicating that Italian researchers have been promoting statistical methods such as SEM and PLS-PM for the evaluation of policies.

The networking analysis emphasizes the strong collaboration from the Chinese researchers with those from countries such as the USA and Australia (the strength of the collaboration is indicated by the thickness of the links), while European researchers prefer to collaborate with each other. In particular, there is a strong collaboration between France, Spain and Italy. The size of the name of the country is related to the number of works published on the analyzed topic, while the different colors of the countries and of the links represent the clusters that have been formed, as determined by the program algorithm.

Figure 4 highlights some of the most frequently used topics in studies associated with "decision making" and "PLS-PM" during this period. As can be observed from Figure 4, topics relating to evaluation start to appear in 2018 ("life", "prevention", "trasportation"). The frequency increases with the passing of the years. The words most commonly used by researchers who have applied PLS-PM in their studies during the last two years have been "students", "education", "university", "perceptions", "job", "learning", "growth" and "country", topics associated with policy evaluation and decision making issues.
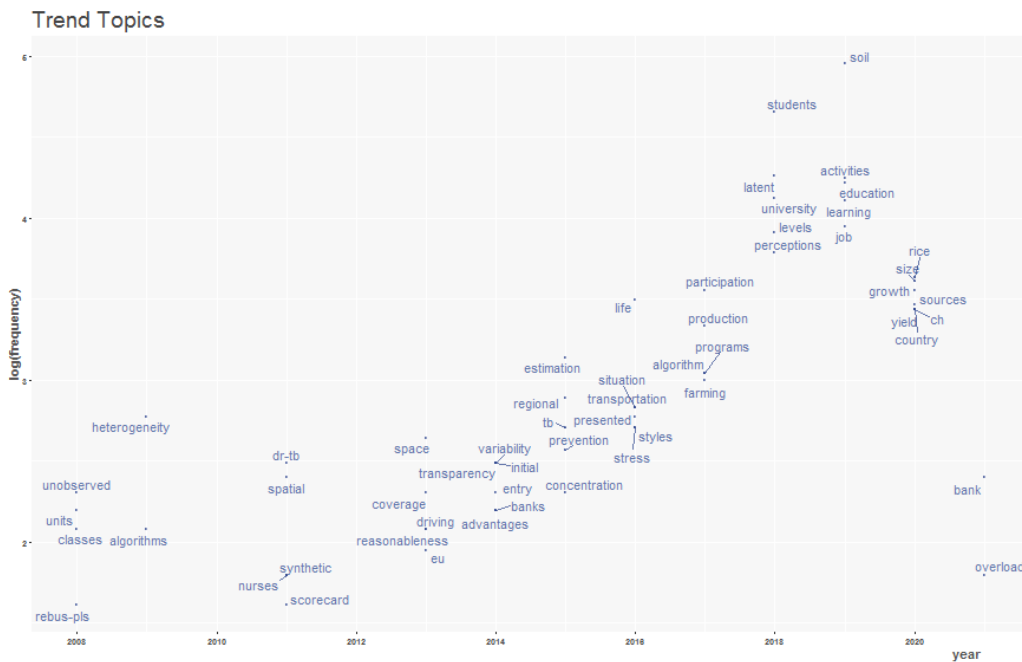
Figure 4: Trend of the topics over the time period

The final figure, Figure 5, shows the keywords considered as themes, classified by different levels of density and centrality in the network of scientific keywords. In the strategic diagram presented in Figure 5, the vertical axis measures the density, namely the strength of the internal links within a cluster represented by a theme, and the horizontal axis the centrality, namely the strength of the internal links within a cluster represented by a theme, and the horizontal axis the centrality, namely the strength of the links between the theme and other themes in the

map (Pourkhani et al., 2019). A thematic map is a very intuitive plot, enabling an analysis of themes according to the quadrant in which they are placed, namely: (1) the upper-right quadrant: motor-themes; (2) the lower-right quadrant: basic themes (3) the lower-left quadrant: emerging or disappearing themes; (4) the upper-left quadrant: very specialized/niche themes (Cataldo et al., 2019).
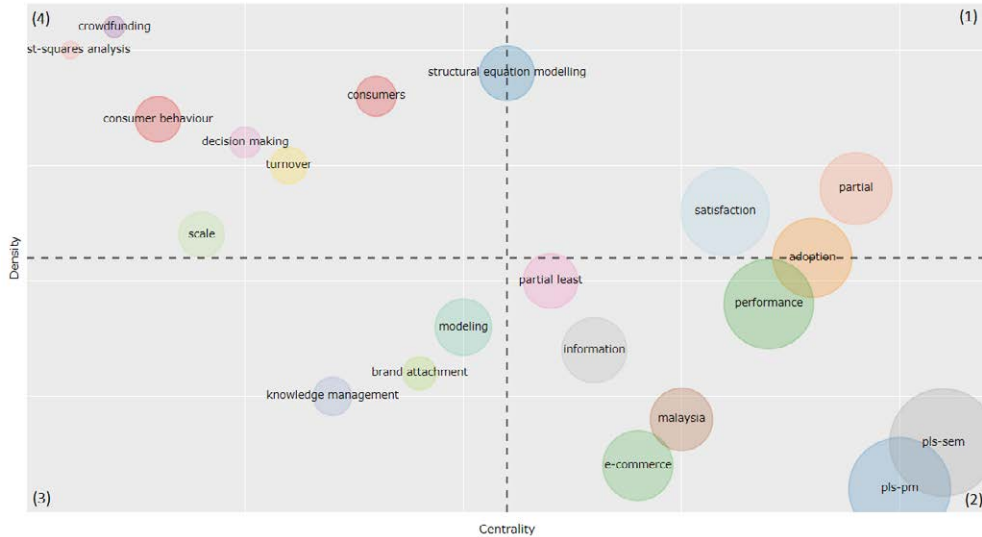
Figure 5: Thematic map

Author's keywords linked to "satisfaction" appear as a motor theme, emphasizing how in the last few years researchers have focused their attention on this theme; words like "trust", "service quality", "customer loyalty", "relationship" and "perceived value" appear in this cluster. Themes with a higher centrality include "pls-sem", "pls-pm", topics that appear ubiquitously in different scientific works and can be considered a common synthesis of the content expressed in the literature. The topic "Malaysia" appears also in this quadrant, while "China", "engagement", "finantial performance" and "risk perception" are other author keywords presented in this cluster, highlighting the predominance of Malaysian and Chinese pubblications in relation to the evaluation theme. Keywords such as "consumer behaviour", "decision-making", "consumer", "crowfunding", despite having a low centrality, have a higher frequency, showing that these themes are considered very specialized topics in these scientific works.

5. Conclusion

The decision making process has always been complex. In the last few years, researchers have been promoting the use of statistical methods such as SEM and PLS-PM for the evaluation of policies, especially in the context of decision making. To better understand and characterize the trend of the scientific pubblication relating to this theme, a bibliometric study of international papers on this subject has been developed highlighting the use of SEM and PLS-PM approaches in policy evaluation during the last twenty years. The data were retrieved from two main databases commonly used by researchers, Scopus and Web of Science, and the analysis of 451 articles was performed using bibliometrix R-Tool (Aria and Cuccurullo, 2017). The results suggest that the interest in research on this topic has increased in recent years, particularly between 2015 and 2019, indicating that this issue has become a significant topic of attention

among researchers in this period. Globally, China is ranked first in terms of production, while in Europe Italian researchers are the most prominent in the promotion of statistical methods such as SEM and PLS-PM for policy evaluation, also collaborating with scholars in Spain and France. The words most frequently used in the last two years by researchers who deal with PLS-PM in their studies have been "students", "education", "university", "perceptions", "job", "learning", "growth" and "country", topics associated with policy evaluation and decision making issues. This study has analysed scientific pubblications on databases being constantly updated. Therefore, a bibliometric analysis regarding an emergent theme may, in a few years, be subject to substantial variations. Furthemore, the present study has analysed a particular theme using two different databases. Despite them being two of the most influential databases, the global perspective could be improved with the inclusion of other databases. However, the results obtained from this analysis may assist researchers in investigation this theme and in focusing on developing the PLS-PM approach for policy evaluation and decision making in many fields of research.

References

Aria, M., Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, Elsevier, **11**(4), pp. 959- 975.

Cataldo, R., Grassia, M.G., Lauro, C.N., Marino, M., Voytsekhovska, V. (2019). A bibliometric study of the global research activity in sustainability and its dimensions. In *International Conference on Data Science and Social Research*, pp. 91-102. Springer, Cham.

Diem, A., Wolter, S.C. (2013). The use of bibliometrics to measure research performance in education sciences. *Research in Higher Education*, **54**(1), pp. 86-114.

Falagas, M. E., Karavasiou, A.I., Bliziotis, I.A. (2006). A bibliometric analysis of global trends of research productivity in tropical medicine. *Acta Tropica*, **99** (2-3), pp. 155-159.

Garfield, E., Sher, I.H., Torpie, R.J. (1964). The use of citation data in writing the history of science. Institute for Scientific Information Inc Philadelphia PA

Maharana, R.K. (2013). Bibliometric analysis of Orissa University of agricultural technology's research output as indexed in Scopus in 2008-2012. *Chinese Librarianship: An International Electronic Journal*, **36**.

Morandi, G., Guido, D., Tagliabue, A. (2015). A bibliometric study of scientific literature on the dietary therapies for epilepsy in Scopus. *Nutritional Neuroscience*, **18**(5), pp. 201-209.

Pourkhani, A., Abdipour, K., Baher, B., Moslehpour, M. (2019). The impact of social media in business growth and performance: a scientometrics analysis. *International Journal of Data and Network Science*, **3**(3), pp. 223-244.

Rodriguez-Soler, R., Uribe-Toril, J., Valenciano, J.D.P. (2020). Worldwide trends in the scientific production on rural depopulation, a bibliometric analysis using bibliometrix R-tool. *Land Use Policy*, **97**, pp. 104787.

Skute, I., Zalewska-Kurek, K., Hatak, I., de Weerd-Nederhof, P. (2019). Mapping the field: a bibliometric analysis of the literature on university-industry collaborations. *The Journal of Technology Transfer*, **44**(3), pp. 916-947.

White, H.D., McCain, K.W. (1989). Bibliometrics. *Annual Review of Information Science and Technology*, **24**, pp. 119-186.

The impact of public research expenditure on agricultural productivity: evidence from developed European countries

Alessandro Magrini

1. Introduction

Agricultural economists agree on the essential role of productivity growth to meet food demand of the rapidly increasing world population, and acknowledge the potentiality of public expenditure in agricultural research to stimulate the required productivity progress (Alston & Pardey, 2014). United States of America (USA) and developed European countries have been leaders in science-based agricultural productivity increase since the middle of the 20th century, motivating hundreds of quantitative studies aimed at assessing the impact of public research expenditure on agricultural productivity and the corresponding economic return. However, the almost totality of these studies has focused on USA (see Fuglie *et al.*, 2017; Baldos *et al.*, 2018; Andersen, 2019 for a review), with few scattered contributes on European countries (Thirtle *et al.*, 2008; Ratinger & Kristkova, 2015; Guesmi & Gil, 2017; Lemarié *et al.*, 2020).

This paper contributes to the literature by providing, for the first time, evidence on the economic return of agricultural research expenditure in developed European countries, making possible a comparison with existing studies focused on USA. We employ yearly data sourced from the United States Department of Agriculture (USDA), the Organisation for Economic Cooperation and Development (OECD), and the Food and Agriculture Organization (FAO) in the period 1970–2016. We follow the consolidated methodology based on a distributed-lag model relating a Total Factor Productivity (TFP) index to public research expenditure, with fixed effects to take into account the panel structure of the data. A Gamma lag distribution is assumed for the impact of research expenditure on productivity as in recent studies, due to its higher flexibility compared to trapezoidal and second order polynomial lag distributions (see Andersen, 2019, Section 4).

This paper is structured as follows. In Section 2, the data are described and the methodology is detailed. In Section 3, the results are reported and discussed. Section 4 contains concluding remarks and purposes for future work.

2. Data and methodology

Our analysis focused on the following countries: Austria (AT), Belgium & Luxembourg (BL), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (EL), Iceland (IS), Ireland (IE), Italy (IT), Netherlands (NL), Norway (NO), Portugal (PT), Spain (ES), Sweden (SE), Switzerland (CH) and United Kingdom (UK). We considered yearly data in the period 1970–2016, specifically agricultural TFP indices computed by USDA, and Government Budget Appropriations or Outlays for R&D (GBAORD) in agriculture made available by OECD.

USDA agricultural TFP indices are available at `https://www.ers.usda.gov` under the section *Data Products – International Agricultural Productivity*. They were computed at country-level with base year 2005 using FAO and International Labour Organization (ILO) data (see Fuglie, 2018 for details).

GBAORD data from OECD are available at `https://doi.org/10.1787/data-00194-en` and represent government budget allocations for research and development by NABS 2007

socio-economic objectives, expressed in million US dollars at 2015 prices and purchasing power parities. We selected the objective 'Agriculture' and employed these data as a proxy of public agricultural research expenditure, which is unavailable for European countries.

Data summaries by year are shown in Table 1, while Figure 1 displays quartiles and mean by year for data in level and in log return (first order difference of logarithmic values). We see that, from 1970 to 2016, the average agricultural TFP and GBAORD have increased, respectively, by 78.3% and 52.1%, with an average annual growth respectively equal to 1.3% and 0.9%.

Table 1: Data summaries by year.

Year	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Agricultural TFP (2005=100)						
1970	43.0	55.0	62.0	66.8	75.0	112.0
1985	61.0	71.0	78.0	79.1	85.0	120.0
2000	89.0	91.0	93.0	96.9	103.0	114.0
2016	104.0	107.0	114.0	119.1	124.0	160.0
GBAORD for agriculture (million 2015 US dollars)						
1970	15.0	46.2	62.1	137.6	215.3	468.1
1985	9.7	53.2	94.4	183.2	165.8	681.5
2000	18.6	61.1	237.8	185.3	274.9	567.7
2016	17.7	47.7	89.2	210.2	331.2	930.6

According to the economic theory, an increase in research expenditure involves an adoption lag, during which the effect on productivity rises from zero to a maximum, followed by a disadoption lag, during which the effect on productivity diminishes to zero. Thus, an appropriate model should weight the impact of research expenditure on productivity according to an inverted U-shaped function of the time lag. According to Fuglie *et al.* (2017), the most employed specifications for the weights of research expenditure include trapezoidal, second order polynomial and Gamma lag distributions, with this last one being increasingly popular in the last decade (see Andersen, 2019, Figure 1 for a graphical illustration).

We preliminarily checked weak stationarity of the country-level time series of agricultural TFP and GBAORD. The augmented Dickey-Fuller test (Dickey & Fuller, 1981) was unable to reject the hypothesis of unit root for all of them. Instead, the hypothesis of unit root was rejected for all the country-level time series taken in log return. In order to avoid spurious regression due to non-stationarity (Granger & Newbold, 1974), we worked on the time series in log return.

Let $j = 1, \ldots, J$ indicate the country and $t = 1971, \ldots, 2016$ denote the year. We specified the following model:

$$\Delta \log \text{TFP}_{j,t} = \alpha_j + \theta \, \text{KS}_{j,t} + \varepsilon_{j,t}$$
$$\text{KS}_{jt} = \sum_{k=0}^{\infty} w_k(\delta, \lambda) \cdot \Delta \log \text{GBAORD}_{j,t-k} \tag{1}$$

where the variable KS is interpreted as the knowledge stock deriving from past research expenditure, and $w_k(\delta, \lambda)$ are weights of a Gamma lag distribution:

$$w_k(\delta, \lambda) = \frac{(k+1)^{\frac{\delta}{1-\delta}} \lambda^k}{\sum_{l=0}^{\infty} (l+1)^{\frac{\delta}{1-\delta}} \lambda^l} \tag{2}$$

and $\varepsilon_{j,t}$ is an exogenous random error, i.e., $\text{E}(\varepsilon_{j,t}) = \text{Cov}(\varepsilon_{j,t}, \text{KS}_{j,t}) = 0$.

Several dummy variables were added to Model (1) in order to explain eventual structural breaks in the TFP series due to weather disasters and economic recessions: one dummy in 1974
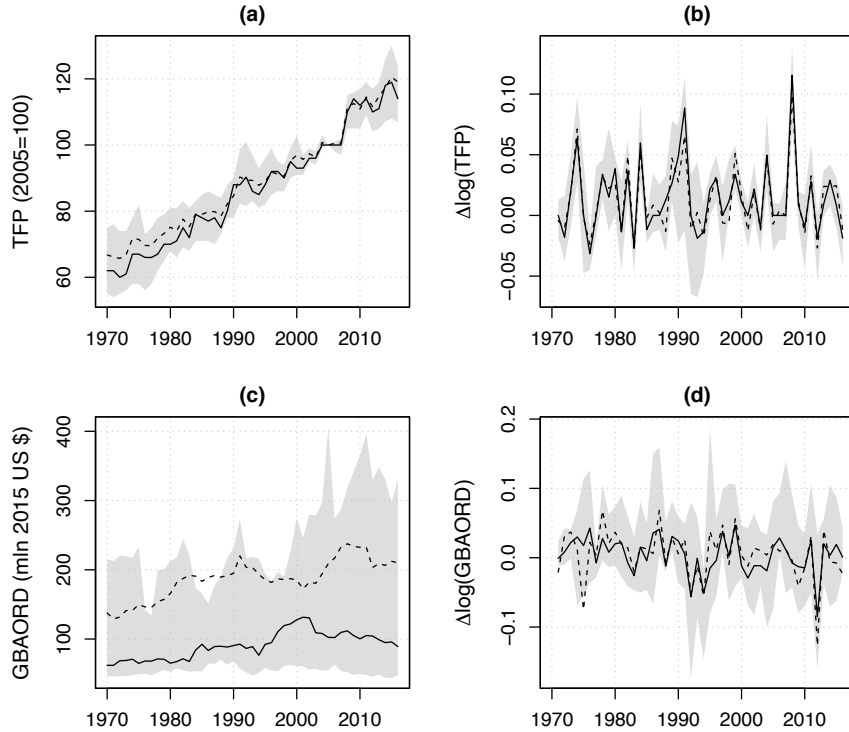
Figure 1: Time series by year. a) TFP, index 2005=100; b) TFP, log return; c) GBAORD, million 2015 US dollars; d) GBAORD, log return. Straight lines, dotted lines and shaded regions indicate, respectively, median, mean and interquartile range across the countries.

representing the European oil crisis during the 1973 Arab-Israeli war; one dummy in 2003 representing the heavy drought and heat wave which hit most European countries in that year; two dummies, one in 2008 and another one in 2012, representing the two major peaks of the European sovereign debt crisis, which was a consequence of the Great Recession in USA.

Since both TFP and GBAORD are in log return, the coefficient $\beta_k = \theta w_k$ is interpreted as the elasticity of TFP with respect to GBAORD at time lag k. Also, since the weights w_k sum to 1, parameter θ is interpreted as the long-term elasticity of TFP with respect to GBAORD.

In order to obtain maximum likelihood estimates for Model (1), we applied ordinary least squares to the models implied by several different pairs of values for δ and λ, and selected the estimates associated to the lowest residual sum of squares (see Schmidt, 1974 for details).

3. Results

We obtained the following estimates: $\hat{\delta} = 0.9$, $\hat{\lambda} = 0.6$, $\hat{\theta} = 0.172$. The standard error of $\hat{\theta}$ computed using the Heteroskedasticity and Autocorrelation Consistent (HAC) estimator (Newey & West, 1987) resulted equal to 0.084 (*p*-value: 0.040). These estimates imply the lag distribution for the impact of GBAORD on TFP shown in Figure 2, which has 99th percentile at 35 years, peak at 17 years and long-term elasticity equal to 0.172 (95% confidence interval: $[0.07, 0.337]$). All the dummy variables showed a statistically significant coefficient, with an implied structural break of positive sign for the ones in 1974 and in 2008 (estimated coefficients

57

0.061 and 0.088, respectively) and of negative sign for the ones in 2003 and in 2012 (estimated coefficients −0.024 and −0.038, respectively).

Our resulting lag distribution for the impact of public research expenditure on productivity is a bit shorter than the ones reported by recent studies on USA. For example, Baldos *et al.* (2018) found a lag distribution with 99th percentile at 51 years, peak at 24 years and long-term elasticity equal to 0.15. Since the latest studies on USA consider a period starting from the 1950s and ending no later than 2011, while our period of analysis is from 1970 to 2016, this difference may be explained by a reduction of the adoption lag in the last one or two decades.
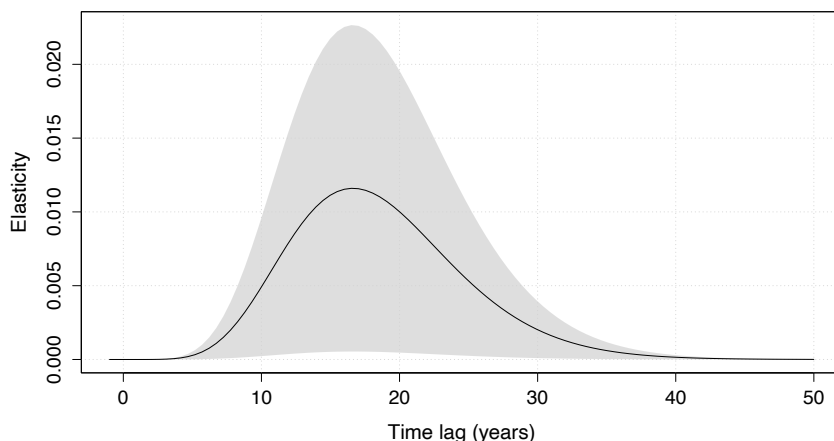
Figure 2: Estimated lag distribution for the impact of GBAORD on TFP. The shaded region represents 95% confidence bands.

Our results are only partially comparable with the ones from studies focused on specific European countries due to several reasons: a much shorter lag length is assumed (Ratinger & Kristkova, 2015; Guesmi & Gil, 2017); the lag distribution is imposed rather than estimated (Lemarié *et al.*, 2020); the considered period is outdated (Thirtle *et al.*, 2008).

After estimating Model (1), we computed the implied internal rates of return by country and compared them with the average annual change of GBAORD in recent years. To compute the internal rates of return, we employed FAO data on the real value of agricultural production in 1970–2016, available at `http://www.fao.org/faostat/en/#data` under the section *Production – Value of Agricultural Production*. Results are reported in Table 2 and displayed in Figure 3. According to our results, the countries with the highest rate of return are Germany, Spain, France and Italy (24.5–25.2%), followed by Netherlands, United Kingdom, Denmark, Greece and Belgium & Luxembourg (20.5–21.8%). However, only Germany, Denmark and Greece increased GBAORD in recent years. Norway has rate of return below the first quartile (15.8%), but it is also the country with the highest average annual change of GBAORD. Iceland, with a rate of return of 9.1%, is a negative outlier.

The estimated internal rates of return are in line with the ones reported by existing studies on USA, and they suggest that developed European countries, just like USA, could benefit from research expenditure in agriculture to a much greater extent than they currently do.

Table 2: Estimated internal rates of return and average annual change of GBAORD in different periods before 2016.

Country	Internal rate of return	Average annual % change of GBAORD		
		2001–2016	2006–2016	2011–2016
AT	18.8	−2.4	−1.8	−6.6
BL	20.5	−2.3	+2.5	−1.5
CH	16.3	−0.1	+0.8	+0.3
DE	25.2	+4.7	+6.0	+0.4
DK	21.2	−3.1	−1.4	+2.6
EL	20.7	−0.9	−2.8	+2.0
ES	25.2	+5.5	−4.4	−6.3
FI	17.4	−1.4	−4.3	−6.4
FR	25.1	−0.5	+3.6	−1.8
IE	18.8	+1.5	+1.6	−1.5
IS	9.1	−0.6	−1.2	−6.2
IT	24.5	+2.0	−4.2	−3.5
NL	21.8	−2.9	−7.6	−9.2
NO	15.8	+3.7	+3.4	+6.3
PT	17.1	−10.5	−10.8	−7.8
SE	18.2	−1.3	−2.9	−0.5
UK	21.4	+0.2	+0.9	−0.7
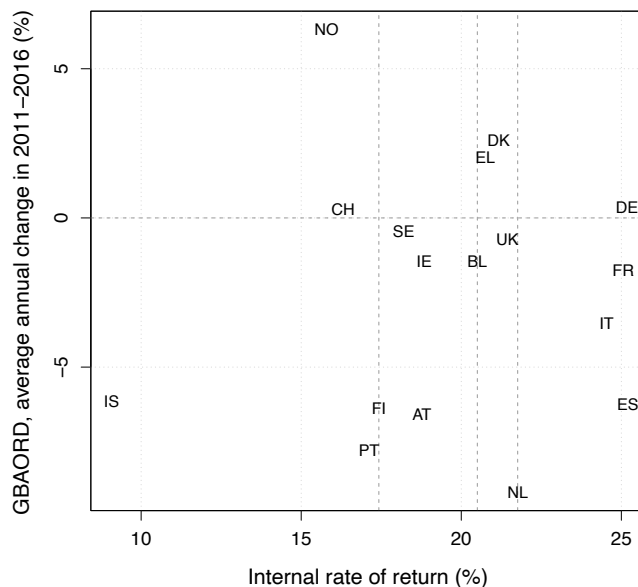
Figure 3: Internal rate of return versus average annual change of GBAORD in 2011–2016. The dotted vertical lines indicate first quartile, median and third quartile of the internal rate of return.

4. Concluding remarks

We estimated for the first time the economic return of agricultural research expenditure in developed European countries, and a comparison was made with existing studies on USA.

The main limitation of our research relies on availability and quality of data. Official statistics on actual public research expenditure in agriculture are unavailable for European countries, being available only those on government budget allocations, which have the restriction to begin in 1970, instead of in 1961 likewise USDA agricultural TFP indices. The use of budget allocations as a proxy of expenditure combined with the limited length of the time series could have significantly affected the efficiency of our estimates, as suggested by the wide confidence bands in Figure 2. Research expenditure from other countries (spillovers) and from the private sector are also expected to influence agricultural productivity, and their omission may bias the estimation of the impact of (domestic) public research expenditure. Unfortunately, data for European countries on these two further determinants of productivity are unavailable, thus they have been ignored in our analysis. In the future, we plan to estimate this missing information indirectly from available statistics. For example, spillovers could be imputed based on similarities in the budget shares for research activities across the countries (see Andersen, 2019, formula 4).

Our results highlight different rates of return across developed European countries, with Iceland being a negative outlier, suggesting the existence of unexplained heterogeneity in the relationship between research expenditure and productivity. Future work will be directed towards the identification of groups of countries with homogeneous characteristics, which could guide the specification of an opportune number of separate models.

References

Alston, J. M., Pardey, P. G. (2014). Agriculture in the global economy. *Journal of Economic Perspectives*, **28**(1), pp. 121-146.

Andersen, M. A. (2019). Knowledge productivity and the returns to agricultural research: a review. *Australian Journal of Agricultural and Resource Economics*, **63**(2), pp. 205-220.

Baldos, U. L. C., Viens, F. G., Hertel, T. W., Fuglie, K. O. (2018). R&D spending, knowledge capital, and agricultural productivity growth: a Bayesian approach. *American Journal of Agricultural Economics*, **101**(1), pp. 291-310.

Dickey, D. A., Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, **49**(4), pp. 1057-1072.

Fuglie, K. O. (2018). Is agricultural productivity slowing? *Global Food Security*, **17**, pp. 73-83.

Fuglie, K. O., Clancy, M., Heisey, P., McDonald, J. (2017). Research, productivity, and output growth in U.S. agriculture. *Journal of Agricultural and Applied Economics*, **49**(4), pp. 514-554.

Granger, C. W. J., Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, **2**(2), pp. 111-120.

Guesmi, B., Gil, J. M. (2017). Measuring the impact of agricultural research on Catalan agricultural productivity. Proceedings of the XV EAAE Congress, *Towards Sustainable Agri-food Systems: Balancing Between Markets and Society*, Aug. 29th – Sep. 1st, 2017, Parma (IT).

Lemarié, S., Orozco, V., Butault, J. P., Musolesi, A., Simioni, M., Schmitt, B. (2020). Assessing the long-term impact of agricultural research on productivity: evidence from France. Preprint. https://hal.inrae.fr/hal-02791517

Newey, W. K., West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**(3), pp. 703-708.

Ratinger, T., Kristkova, Z. (2015). R&D Investments, technology spillovers and agricultural productivity, case of the Czech Republic. *Agricultural Economics – Czech*, **61**(7), pp. 297-313.

Schmidt, P. (1974). An argument for the usefulness of the Gamma distributed lag model. *International Economic Review*, **15**(1), pp. 246-250.

Thirtle, C., Piesse, J., Schimmelpfennig, D. (2008). Modelling the length and shape of the R&D lag: an application to UK agricultural productivity. *Agricultural Economics*, **39**(1), pp. 73-85.

How to become a pastry chef:
a statistical analysis through the company requirements

Paolo Mariani, Andrea Marletta

1. Introduction

During the last years, the competitive context in which the firms are involved is definitely changed. The managers work in dynamic markets, characterised by unpredictable and complex phenomena. The companies are rapidly involved in changing processes where the available capabilities of the organisation represent a key point for success. This led to a flexible organizing model where new professional and relational competencies are emerging. In 2017, the OECD stated that sectors and nations may take advantage of better management of skills (Grundke et al., 2017).

In this study, the attention is focused on the labour market in the food & beverage sector, in particular the requirements are considered for two job profiles: pastry chef and pastry assistant. A key question is which are the requested competencies searched by the companies to hire these figures and which are the requested differences between the two roles?

The role of soft skills has increased its importance compared to the hard ones guaranteeing a competitive advantage to for the success of a company. They are transversal skills necessary to have success in the job market. This is why in this study, soft skills are considered together with the candidate's previous experience, the age and the knowledge of a foreign language. Hard skills are related to the knowledges and technical competencies useful for a specific role, while soft skills are relational and personal capacities. For this reason, they are more difficult to define and measure, and since they are not related to a learning method, it is more complicated to acquire them. According to the Excelsior informative system and by considering all the job positions, the most requested soft skills are: autonomy, flexibility, adaptability, ability to communicate, problem solving and team working (Unioncamere, 2017).

In relation to the area of interest of the study, in a pastry shop the hard skills are represented by the capability to know recipes and to use the pastry tools, while the soft skills could be intended as the goodness to satisfy the customers or the availability towards the colleagues. The pastry shop market in Italy is a key industry for the food & beverage economy because it represents an example of excellence in the world. It is a sector in expansion in which the quality of the product is going to increase thanks to education and technology. It is a growing market ready to satisfy the new requests of category of intolerant people (lactose free, gluten free, ...) and particular attention to the biological products.

In Italy, there are about $40,000$ pastry and ice-cream shops with almost $100,000$ employees. The 32.7% of these shops has $4-6$ employees, and the 47% has $2-3$ employees. Pastry shops are more spread in the North and Centre Area, while ice-cream shops are in the South Area. There is no significant difference in the distribution of employees between North and South area.

The paper is structured as follows: after the introduction, a second section is dedicated to the methodologies used to answer the research objectives. A third section will show the description of the dataset and some preliminary results. Finally, conclusions and future works will follow.

2. Methodological tools and data description

This study has several aims: first, to detect a possible relationship between the age of work experience and the possibility to hire as a pastry chef or a pastry assistant taking into account a set of possible skills; second, to detect whether the age of a candidate could represent an obstacle or an advantage in the hiring process; third, to find a classification of the analysed soft skills.

Two methodological techniques have been principally used in this study to answer these issues: logistic regression and principal component analysis. The logistic regression is a common statistical model in presence of a binary response variable. In particular, this tool could allow to answer the first two research objectives hypothesizing a model in which the binary response is the category to be hired (pastry chef or pastry assistant) and the explanatory variables are the age of previous work experience, the knowledge of a foreign language and some features of the pastry shop. To pursue the last issue, a principal component analysis has been applied on a set of soft skills to make a classification in two different groups.

Finally, the results from these two approaches are used to distinguish the figure of a pastry chef from a pastry assistant based on age, previous work experience and a set of soft skills.

Data for this analysis were collected by The AdeccoGroup in Italy in 2016 and 2017. The personal competencies needed to face the growing flexibility of the profession are the subject of specified request cross-sectional to more economic sectors. Data contain information classified on nine sectors: IT Digital, Engineering, Pharmaceutical, Finance, Tourism, Human Resource, Commercial, Food & Beverage and Production. In particular, the dataset involves about $220,000$ job position and 43 job figures. Among these figures, the selected ones are referred to the pastry shops in the Food & Beverage sector: pastry chef and pastry assistant.

The figure of pastry chef designs and creates sweets and cakes, plans and organizes candy in hotel, restaurants and pastry shops. The activities include the preparation of recipes, the estimate of costs for feed supplies, the quality monitoring of the product, the supervision and coordination of other chefs activities, the control of the equipments and the recruitment of the pastry assistants. The pastry assistant prepares ingredients, washes the dishes, cleans the common spaces and develops other tasks in support of the pastry chef. The activities contain the kitchen's cleaning, the control and the conservation of the ingredients and the preparation of simple products. It is possible to note that the two figures are hierarchically connected performing tasks at different levels, for this reason it may be expected that they have different soft skills.

The used dataset for the analysis contains information related to 76 job offers for the two professional figures and the features of the subjects that has been hired, 10 pastry chefs and 66 pastry assistants. For these subjects, information is available at the candidate level about gender, 43 women (57%) and 33 men (43%), date of birth, previous work experiences and at job offer level as company dimension, requested language and soft skills.

About the previous work experiences of the candidates, it has been expressed in terms of number of months and it involves both experiences pertinent or not to the food & beverage sector. The company dimension is categorized on three levels based on the number of employees of the company according to an internal classification of TheAdeccoGroup. The list of soft skills requested to be hired in the pastry shop is available for each job offer. To reach the purpose of this analysis, 12 dummy variables (one for each soft skill) were created with value 1 in case of presence of the competence in the job offer and 0 otherwise.

On 76 job offers, the 12 soft skills are:

1. Self-control (36.8%)
2. Customer orientation (15.8%)
3. Learning and innovation (13.2%)
4. Autonomy and initiative (13.2%)
5. Quality orientation (11.8%)
6. Team Working (11.8%)
7. Communication (10.5%)
8. Adaptability (7.9%)
9. Participation and responsibility (6.6%)
10. Impact and influence (6.6%)
11. Motivation (2.6%)
12. Planning and organisation (1.3%)

where the value in brackets is the percentage of presence of the competence in the job offers, these means the self-control is the most requested competence desired in 28 of the 76 cases. Customer orientation is requested in 12 job offers, at the third place, learning and innovation and autonomy and initiative in 10 job offers

If a sum of the soft skills is provided for job offer, as expected a number of higher skills is requested for the pastry chef, since it is a role of superior degree. On the other hand, the number of requested skills seems to be independent by the company dimension.

3. Principal results

A logistic regression gives the possibility to detect a relation between a set of explanatory variables as the length of the previous work experience (expressed in month), the age of the candidate and the job position:

- 0 for the professional figure pastry assistant
- 1 for the professional figure pastry chef

Considering a 95% confidence interval, the only variabile with a β coefficient different from 0 is the age of the candidate. $\beta = 0.132$ and $exp(\beta) = 1.141$, this means that for each year of age the probability to be a pastry chef is 1.14 times respect to a pastry assistant. The number of requested languages, the dimension of the company and the length of the previous work experience (expressed in month) did not seem to have an impact on the hiring between the two professional figures.

The second technique presented here is the Principal Component Analysis (PCA), usually applied to reduce the space of dimensions (Jolliffe, 2002). In this case it is used to group the set of 10 soft skills, Motivation and Planning and organisation have been deleted because of low frequencies. The number of components here chosen is 3 explaining 75% of the variance. Once established the number of components, a Varimax rotation was applied to the components matrix, in order to improve the interpretation of the three groups. In table 1, the classification of the soft skills is presented.

Group 1: Efficiency	Group 2: Outward	Group 3: Synergy
Quality orientation	Customer orientation	Team Working
Learning and innovation	Communication	Self-control (-)
Participation and responsibility	Adaptability	
Autonomy and initiative	Impact and influence	

Table 1: Classification of soft skills using PCA, 2016-2017, Italy

The first group classifies soft skills more tangible and related to the product as the quality orientation, so it may be named Efficiency. The second one relates to competencies towards others like customer orientation and communication, for this reason this group may be named Outward. The last group is positively correlated with team working and negatively correlated with self-control and it is called Synergy.

4. Conclusions

The present study aimed to underline the importance of soft skills as a requested requirement in the job market for a successful job career. Using the AdeccoGroup dataset on job figures, a detailed study was conducted to investigate the food & beverage sector and in particular the pastry shop field. The application classified some desired features for candidates hired as pastry chef and pastry assistant in Italy in 2016 and 2017.

The most interesting result is about the presence of different soft skills and the influence of age and the previous work experience in the two job figures. As it is possible to expect from a hierarchical point of view, the oldest candidates with more experience are addressed to the figure of pastry chef, while the youngest ones or the least experienced candidates are more desired as a pastry assistant. Another important expected result is that the pastry chef requests a higher number of soft skills than to the pastry assistant. For the pastry chef figure, the most requested soft skills are: autonomy and initiative, quality orientation and learning and innovation. On the other hand, for the pastry assistant figure, the desired soft skills are: self control, team working and customer orientation. Finally, an additional classification allowed to divide the set of soft skills in Efficiency soft skills, Outward and Synergy soft skills.

Future works could considering other explanatory variables in the logistic regression model to detect the influence of other factor in the choice of a new hiring between the two figures. Moreover, a similar analysis could be conducted on other job figures to confirm the soft skills classification in the three groups detected.

References

Grundke, R., Jamet, S., Kalamova, M., Keslair, F., Squicciarini, M. (2017). *Skills and global value chains*. OECD, Paris.

Jolliffe, I. T. (2002). *Principal Component Analysis, 2nd edn. Series: Springer Series in Statistics*, XXIX, 487. illus. Springer, NY, 28.

Unioncamere. (2017). Sistema informativo excelsior. In Progetto Excelsior. Sintesi dei principali risultati. http://excelsior.unioncamere.net

Measuring the movement between employment and self-employment: a survey proposal

Luigi Fabbris, Paolo Feltrin

1. Introduction

The *Oxford English Dictionary* defines self-employment as 'the state of working for oneself as a freelancer or the owner of a business rather than for an employer'. This definition highlights that a self-employed person works for themselves 'rather than for an employer'. However, some people work for an employer but are fiscally independent. These include, among others, the so called internal independent workers—individuals who are registered as self-employed but work at a firm where they are subject to organisational rules as are employees. Should these people be considered to be dependent or independent workers? Even Eurostat (European Union, 2018) sees as peculiar this professional status called *dependent self-employment*—itself a linguistic paradox.[1] Other ambivalent professional statuses occur worldwide. These definitions based on dichotomies are inadequate to define borderline types of employment.

Following the hypothesis that a multidimensional definition is needed to classify workers according to professional status, we analyse a plurality of viewpoints and propose a survey to statistically measure both the disputed and the undisputed categories of self-employment. All viewpoints pertain to people's a priori conditions and not to their outcomes.

The rest of the paper is organised as follows. Section 2 highlights the critical factors in the movement between employment and self-employment to help researchers build categories, or blocks, of workers who—possibly in a future rebuilding of professional classification—could be considered to be self-employed. Section 3 presents a scheme to analyse Italian self-employment in relation to the European literature. Section 4 concludes with a nation-wide survey proposal.

2. Trends of self-employment

According to Eurostat, a self-employed person is the sole or joint owner of the unincorporated (i.e., formed into a legal corporation) enterprise in which they work, unless they are also in paid employment that is their main activity (they then are considered to be an employee). The self-employed also include unpaid family workers; outworkers, or those who work outside their usual workplace, for instance, at home; and workers engaged in production done entirely for their own final use or capital formation, either individually or collectively. The self-employed without employees are called *own-account self-employed*, and the self-employed with their own employees are called *employers* (https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Self-employed).

OECD/European Union (2017) data show that only a scant minority of Italian workforce seek self-employment. Only 1.6% of Italian job seekers are oriented towards independent employment, while 1.2% of the unemployed seek self-employment, and 1.8% of those previously or currently employed in dependent positions are willing to change status. Indeed, a large proportion of job seekers (23.4%) show indifference towards the professional status of

---

[1] Dependent self-employment is a phenomenon of some importance in countries where few professions are regulated, such as the Netherlands (5.3% of the self-employed workforce), the Czech Republic (5.8%), the United Kingdom (6.7%), Cyprus (7.3%) and Slovakia (9.9%). In the EU-28, this category amounted 3.5% of self-employment and 0.5% of overall employment in 2017. In Italy, dependent self-employment amounted to 4.3% of self-employment and 0.9% of overall employment (European Union, 2018).

Luigi Fabbris, University of Padua, Italy, luigi.fabbris@unipd.it, 0000-0001-8657-8361
Paolo Feltrin, Tolomeo Studi e Ricerche, Italy, paolo.feltrin@gmail.com, 0000-0003-2801-5151

the sought job—and perhaps some might enter independent employment. Nevertheless, the data highlight that few Italians endorse the aim of starting one's own business, but many more express the same level of liking for this professional status as employment.

Moreover, the propensity to exit self-employment is much higher than that to enter it. In 2019, 61.5% of self-employed workers who lost their jobs stated they were looking for employee positions, while 7.7% insisted on staying in self-employment, and another 30.8% were indifferent to both statuses. These results are somewhat expected because those who lose their job face psychological pressure to not repeat the negative experience. Instead, the propensity to stay in their dependent professional status is higher for former employees (83.9%), of whom only 1.8% would accept self-employment, while 14.3% are indifferent to both statuses.

In contrast, comparing the movement between professional statuses over time (Table 1), we see that 99.5% of Italian employees remain in a dependent status, and 98.8% of the self-employed remain in an independent status after one year. The largest movement is shown in the para-subordinate category,[2] in which only 86.8% remain after one year. The movement between dependent and independent work is the largest in absolute terms: in 2019, about 72,000 workers transitioned from a dependent status to an independent status, and 61,000 moved the other way. If we add the movement to and from a para-subordinate status to self-employment, the new entrants in an independent status were about 76,000, while about 62,000 left it. In relative terms, the data show an almost equal balance: those newly in an independent status numbered only 0.1 percentage points (23.2 versus 23.1) more than those who left it. The movement balance is null if we consider para-subordinates to be self-employed.

Among the self-employed categories, employers are the most stable (99.2% staying in the same category for one year), followed by those self-employed in craft, commerce and agriculture (98.7%), and freelancers (98.3%). Even cooperative workers dominantly stay (96.1%), though given their double status as associated owners and employees of their cooperatives, they might classify themselves as dependent or independent at different times. Family workers (97.3% stay) and para-subordinates (86.8%) are similarly hybrid but differ in commitment with respect to self-employment.

Table 1. Transition matrix (%) between current and one-year-earlier professional status of Italian workers, Italy, year 2019 (*Authors' analysis of Italian labour force survey data*).

Condition 1 year earlier	Current condition							
	Dependent	Para-subordin.	Employer	Freelance	Craft, commerce	Family worker	Cooperat. Worker	Row total (%)
Employee	99.5	0.1	…	0.1	0.2	…	…	100.0 (76.1)
Parasubordinate	9.9	86.8	…	2.0	1.2	0.1	0.0	100.0 (0.7)
Employer	0.4	0.0	99.2	0.1	0.3	…	0.0	100.0 (1.2)
Freelance	1.4	0.1	…	98.3	0.2	…	0.0	100.0 (6.3)
Craft, commerce	1.1	0.1	…	0.1	98.7	…	…	100.0 (13.5)
Family work	0.9	…	0.2	…	1.5	97.3	…	100.0 (1.5)
Cooper. worker	3.3	0.0	0.0	…	0.5	…	96.1	100.0 (0.6)
Self-employment	1.2	…	5.3	26.7	57.3	6.4	2.6	100.0 (23.1)
(Total)	(76.0)	(0.6)	(1.2)	(6.3)	(13.6)	(1.5)	(0.6)	(100.0)

[2] In Italy, the para-subordinate, or pseudo self-employment, category includes temporary or ad-hoc contracts of collaboration with a company for which the company requires that collaborators register themselves at the Chamber of Commerce as self-employed and pay directly their social security fees.

3. Model of the movement between employment and self-employment

The definition of self-employment has legal, organizational and economical aspects.[3] The legal aspect refers to who the employer is—a company or self-employed individuals themselves. This dimension leaves in a limbo some categories of workers who are recognized in Italy as self-employed but, strictly speaking, do not employ themselves (e.g., unpaid family workers and cooperatives' working partners). In another peculiar category are the dependent self-employed, who are legally self-employed but possess dependent traits of employees.

Organizational and economic aspects are also added to the definition of self-employment. A main economic characteristic is a worker's dependence on their income source—whether the source is unique or nearly so, or a worker can work for as many clients they want. Even organizational dependency, which refers to the work time and schedule, task order and content and ownership of work equipment (e.g., tools, space and premises), could distinguish employment from self-employment. The rationale is that these economic and organizational aspects reduce the ability to classify workers as self-employed to the degree that they depend on others' will. Unfortunately, there is no clear-cut rule. Suppose, for instance, that a worker is registered as self-employed but uses equipment at a company's workplace, and all organizational aspects of the job are ruled by this company that is this worker's only client. Should this worker be classified as an anomalous self-employed or an anomalous employee? Alternatively, suppose a worker autonomously organizes work tasks and can adopt a flexible working schedule. Is this worker more self-employed rather the previous anomalous worker? It is difficult to say; we can only state that other people's discretion concerning the workday and organization of the work environment makes self-employment classification a dubious task and requires further research.

From a social viewpoint, it is relevant to understand the reasons why workers became self-employed. Indeed, there is a clear distinction in classification based on whether a worker's decision to start their own business depends on their consolidated will or familial traditions or, instead, on contingent external pressures (e.g., the need for a flexible schedule or an explicit request from a former employer) or nearly random situations (e.g., a sudden opportunity a worker was prompted to take or a lack of opportunities to find a job as an employee). In the 2017 European labour force survey on self-employment (European Union, 2018), smoothly entering self-employment, as if it were written in one's destiny, was by far the most frequent answer when the self-employed were asked about their reasons for entering their current job (38.7%, to which can be added another 7.2% who stated that self-employment was a common practice in their field). This reason was followed by continuation of a family business (24%), environmental pressures (12.6%, of whom 10.3% could not find a job as an employee and 2.3% received a request from a former employer) and, finally, contingent reasons (7.5% for the need for flexible hours and 8.2% for other reasons).

It can be concluded that in the EU, the large majority of workers entered self-employment either instinctively or as a consequence of opportunistic reasoning. Consequently, only 1 of 8 self-employed workers would switch to employment if they could. In any case, even if motivation might influence the stability of self-employed workers' willingness to maintain this status, it does not affect their professional status.

To start and maintain a self-employed business ("pull factors") is a positive attitude, particularly the feeling of being involved in and satisfied with work tasks. Involvement is a multidimensional condition that includes the opportunity to work with partners or family members who share responsibilities and efforts; the possible number of subcontractors and/or employees involved and the subcontractors and/or employees recently recruited; a high level

---

[3] We could also consider fiscal and social security aspects. Concerning social security, Italy's National Social Security Institute includes the following among independent workers: small business owners in agriculture, craft and commerce; partners in ventures; home sellers; voucher-based workers; and freelancers.

of dedicated financial resources; and, in general, positive business prospects.

Another important attitude is workers' general disposition towards self-employment. This disposition is a direct consequence of workers' personal beliefs and system of values influenced by culture and life experience. A positive (or negative) disposition towards self-employment might strengthen or weaken at any turning point in life. Whatever happens to people may influence their attitudes towards professional decisions. Positive examples given by parents, relatives and peers certainly can push a young worker to start in the family business. Even a professional decision that seems determined from birth—as it may look for a younger generation continuing a family business—is effectively influenced throughout their lives. Indeed, some choose not to maintain a family business.

Let us now examine the barriers to self-employment ("push factors"). A first distinction is between individual and external sources. The main individual barriers to starting one's own business include a social disposition negatively oriented toward self-employment, the possible effects of impairments or chronic diseases on work activities and potential difficulties accessing and managing credit. The external barriers include workers' reflections on the social barriers, such as the time and guarantees required to get financing, level of administrative burdens, level of social protections and retirement fund coverage, and economic difficulties that restrain clients from asking for work or that delay or reduce payments.

The relevance of both individual and social barriers as perceived by workers should be observed with specific survey questions because social barriers affect individuals' business decisions depending on their coping ability. Social barriers influence the movement between employment and self-employment both before the initial choice and at all turning points during one's work life. The survey, therefore, should investigate the attitudes and perceptions of social barriers among all the adults, not only workers.

Figure 1 represents the pull and push factors that may influence the movement between employment and self-employment. The model can be considered an adaptation of Ajzen's (1991) theory of planned behaviour whose dependent variable is behavioural intention. A more comprehensive model to predict the movement between types of employment should include human, social and psychological capital and a set of control variables. This paper, though, does not include types of capital or control variables (dashed lines in Figure 1) because of space limits.

Figure 1. Factors influencing the movement between employment and self-employment.
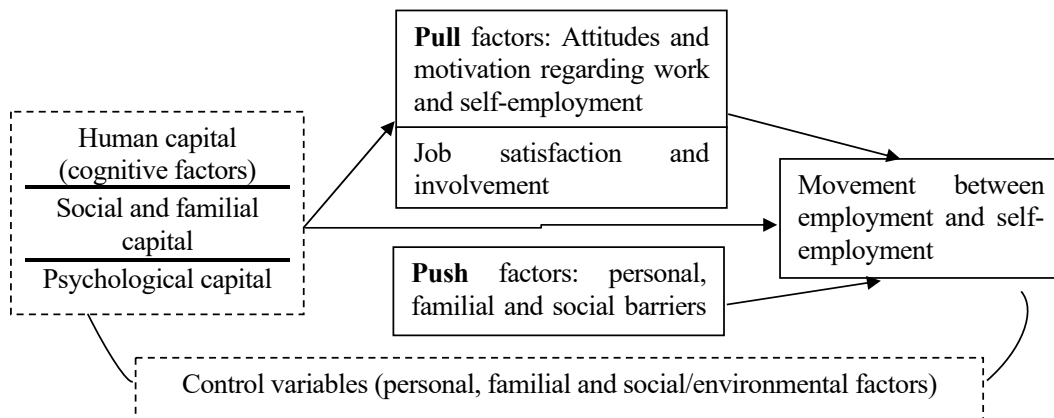
4. A survey proposal

Our analysis is aimed at defining a set of survey questions on the movement between

employment and self-employment. Our main conclusion is that the questionnaire should include a set of common questions for the entire labour force, not only workers, and the movement should encompass both employees and self-employed persons at any occupational turning point.

Our rationale is that no professional status is permanent, and when combined appropriately, certain variables are able to predict individuals' decision to start their own business. In this way, workers can be classified into statistical blocks with varying levels of self-employment scale. Our proposal is to pinpoint the crucial elements of this matter also to suggest official statistics institutions how to measure this changing-nature phenomenon.

The initial question to be submitted to all the adult population is 'Are you waiting to start a new job?'. This question is needed as some respondents who have completed their education and are simply waiting to start a new business might be confused with those not in education, employment or training, as shown in Fabbris and Scioni (2020). The answer options will be: 'Yes, waiting to become an employee'; 'Yes, waiting to become self-employed, possibly in a family business'; 'Yes, waiting to work as a cooperative member'; and 'No'.

The disposition towards self-employment and, conversely, employment will be highlighted with a question about the type of employment sought and why. This question will be posed to people looking for a job in a slightly different manner than those not looking for a job. For job seekers (including those who want to change jobs or are seeking a second job), the question will be: 'Generally, do you prefer to work as an employee or as self-employed?' The answer options will be: 'Prefer to work as an employee', 'Prefer to work as a self-employed or in a family business', 'Prefer to work as a cooperative member' and 'No preference'. For those not seeking a job, the question, with the same answer options, can be rephrased: 'In case you aim to seek a (new) job, would you prefer to work as an employee or self-employed?' The comparison between current and preferred status allows statistical evaluation of the strength of the respondents' preference for the professional status they are in, and the proportion of the respondents staying in or moving from the employment and the self-employment categories.

The reasons for the self-employment choice (pull factors) are described in Section 3. In addition, the reader is pointed to the questionnaire discussed in European Union (2018) and the Italian labour force survey questionnaire (Istat, 2018). In summary, we suggest that all workers should describe their job, either current or expected, in reference to the following dimensions:

- *Organisational independence*—the flexibility of work time, hours and schedule; work-life balance; ownership of work equipment, tools and machinery; working in one's own or others' offices and premises; possibilities for smart working; and having one or more than one client

- *Economic and social independence*—the quantity and stability of income, social security and retirement plans; strength of ties with employers and clients giving work; career prospects; and social relevance of job outcomes

- *Task autonomy and challenges*—personal influence over task content and order in the main job; level of responsibility for task complexity management; confidence in one's own professional means; quantity and variety of used competencies; risk-taking capability; proactiveness in seeking work opportunities; and learning in the workplace; may be considered to be a proxy for work quality

- *Business network width*—the number of partners, subcontractors, employees and mentors who usually work with the respondent; and engagement level in control of the budget and business volume; may be considered to be a proxy for work quantity.

Some of these questions could appear to pertain to only self-employment, but they are also

applicable to company managers and, in general, so-called *intrapreneurs*, who are described by Krueger and Brazeal (1994) and Douglas and Fitzsimmons (2013) as having a proactive attitude that drives their work activities irrespective of their workplace. These questions, therefore, will be asked of all workers in forms that depend on the communication channel with the respondents (i.e., telephone, face-to-face or www systems). In addition, surveying positive attitudes towards labour will require questioning all workers about their job satisfaction (e.g., overall job satisfaction, then for income, autonomy, complexity, challenge, career/business prospects and flexibility of job tasks).

Finally, a survey should include the socio-psychological barriers to self-employment (push factors). Limitations on succeeding at work can come from physical and social problems; previous failures; family expectations; care work for children and relatives; educational inadequacies; gender, age and other characteristics of the respondents that could interfere with work tasks; time and guarantees required to get financing; level of administrative burdens; coverage of social protection; difficulties recovering credit and finding work orders; and the like.

These varied questions can be organised in a battery so that they can be administered with the same scale to all the respondents. The umbrella question will ask how much the respondents perceive that current economic and social difficulties interfere with their decision to start their own business, say: 'How much do you agree with the following statements?', and the answer options will be 'People with disabilities are discriminated against in the workplace'; 'In private companies, women are discriminated against'; 'It is no use seeking a job in such poor economic conditions'; 'Only sly guys or those with friends in the right place get jobs in Italy'; 'In Italy, it is difficult to access financial credit to start a business on your own'; 'Once you fail, there is no chance for you to start your own business'; 'It is only possible to balance private life and work if you work for a public organisation'; 'A woman cannot have a job if she has children'; and so on.

It is worth saying that our model ignores the economic sector in which the worker acts and the harmonisation with European rules as possible interaction factors for self-employment classification purposes. These factors should be considered after the analysis of the survey results.

References

Ajzen, I. (1991). Theory of planned behavior. *Organizational Behavior and Human Decision Processes*, **50**, pp. 179-211.

Douglas, E.J., Fitzsimmons, J.R. (2013). Intrapreneurial intentions versus entrepreneurial intentions: Distinct constructs with different antecedents. *Small Business Economics*, **41**(1), pp. 115-132.

European Union (2018). *Labour Force Survey (LFS) Ad-hoc Module 2017 on the Self-Employed Persons. Assessment Report 2018 Edition.* Publications Office of the European Union, Luxembourg.

Fabbris, L., Scioni, M. (2020). Does an entrepreneurial spirit animate fresh graduates in their work-seeking during uncertain times?. (in this volume).

ISTAT (2018). *Rilevazione sulle forze di lavoro. Questionario unico. 1° trimestre 2018.* (https://www.istat.it/it/files/2011/01/Facsimile-Questionario-in-italiano.pdf?title=La+rilevazione+ sulle+forze+di+lavoro++-+10%2Fgen%2F2018+-+Facsimile+questionario+in+italiano.pdf).

Krueger, N.F.Jr., Brazeal, D.V. (1994). Entrepreneurial potential and potential entrepreneurs. *Entrepreneurship: Theory & Practice*, **18**(3), pp. 91-104.

OECD/European Union (2017). *The Missing Entrepreneurs 2017. Policies for Inclusive Entrepreneurship*. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264283602-en.

Innovation and sustainability: the Italian scenario

Rosanna Cataldo, Maria Gabriella Grassia, Paolo Mazzocchi, Claudio Quintano, Antonella Rocca

1. Introduction

Recent public and governmental concerns regarding sustainability have increased attention on the possibility of improving firms' efficiency in terms of the emerging topic of sustainable innovation. The perspective of what represents innovation has changed significantly in the pioneering and the wide usage of patent statistics. In fact, a large number of research papers have suggested significant advancements in the usage of indicators connected to measuring innovation (see, among others, Rothwel, 1992; Hagedoorn and Cloodt, 2003; Smith, 2005; Gössling and Rutten, 2007; Makkonen and Van der Have, 2013). One of the most frequently used set of indicators to assess the innovation level of European countries is the European Innovation Scoreboard (EIS; European Commission, 2020), while the Regional Innovation Scoreboard (RIS; European Commission, 2019) represents a regional extension of the EIS. Compared to EIS, the RIS assesses the innovation performance using a limited number of indicators. The fourth edition of the Oslo manual (OECD, 2018) proposed a detailed updated guideline focused on measuring innovation in the business sector, and Dziallas and Blind (2019) contributed to the literature review of innovation measurements by carrying out an extensive analysis. Nevertheless, there still remains a broad discussion on these issues.

Sustainable innovation combines the innovation topic and the characteristics connected to sustainable development, which in turn involve three dimensions of sustainability: economic, social and environmental (or ecological) features (Sood and Tellis, 2005). These subjects can also be investigated considering several goals of sustainable development. Among others, Carrillo-Hermosilla et al. (2009; 2010) presented an overview of connections among innovation, ecological sustainability, eco-innovation and sustainable innovation.

Since the research question connected to the impact of the innovation on sustainability is still open, the present work attempts to shed light upon this relationship, considering the Italian Regions. As for the theoretical model, the present article considers a higher order construct (Wetzels et al., 2009), also known as a hierarchical (component) model (HCM), which is based on the Structural Equation Model (SEM) Partial Least Squares (PLS) Path Modelling (PM). In the authors' opinion, from a policy maker's and managerial point of view, the possibility of improving firms' efficiency in terms of several dimensions of sustainable innovation represents a relevant topic that must be investigated.

2. Sustainable innovation in the business sector

As mentioned above, the OECD (2018) manual focuses on measuring innovation in the business sector following the SNA 2008 recommendations. It suggests a framework for measuring innovation using a common definition, and it recommends—for international comparisons—several specifications to avoid weaknesses in empirical analysis. According to a similar perspective, to provide homogenous and comparable indicators—and to avoid the exclusion of relevant dimensions— specific economic activity boundaries and spatial perimeters of the firms investigated

Rosanna Cataldo, University of Naples Federico II, Italy, rosanna.cataldo2@unina.it, 0000-0002-6324-8252
Maria Gabriella Grassia, University of Naples Federico II, Italy, mariagabriella.grassia@unina.it, 0000-0002-7128-7323
Paolo Mazzocchi, University of Naples Parthenope, Italy, paolo.mazzocchi@uniparthenope.it, 0000-0002-6632-314X
Claudio Quintano, University of Naples Suor Orsola Benincasa, Italy, claudio.quintano@unisob.na.it, 0000-0001-8315-8476
Antonella Rocca, University of Naples Parthenope, Italy, antonella.rocca@uniparthenope.it, 0000-0001-8171-3149

can be fixed, considering small-sized and medium-sized enterprises (SMEs). Durst and Edvardsson (2012) highlighted that SMEs are the drivers of most nations all over the world; the present research dedicates special prominence to them, also considering potentially innovative SMEs that could become innovative but cannot because they do not yet have all the requirements. In addition, since some economic sectors are more interested in innovation than others, and since international comparisons of innovation features require the specification of a homogeneous structure to perform the analysis, consequently specific NACE codes can be considered for each Italian Region.

As stated earlier, this empirical work is performed to address the research question aimed at verifying the impact on the sustainable innovation via HCM, and the authors postulate that the sustainable innovation is the only higher-order latent variable in their model. To investigate the statistical significance of the relationship, one endogenous variable (*Sustainable Innovation*) is estimated using four (exogenous) latent constructs: *Business Standard Innovation (BSI), SMEs Innovation (SmI), Economic Sustainability (EcS)* and *Social Sustainability (SoS)*. Given this definition, the authors express the following general form of the *Sustainable Innovation* equation:

$$Sustainable\ innovation = f(BSI, SmI, EcS, SoS) \qquad (1)$$

The model proposed is based on a well-defined path diagram shown in Figure 1 to describe the relationships between the different dimensions. More details about codes and variable definitions are provided in Table 1.
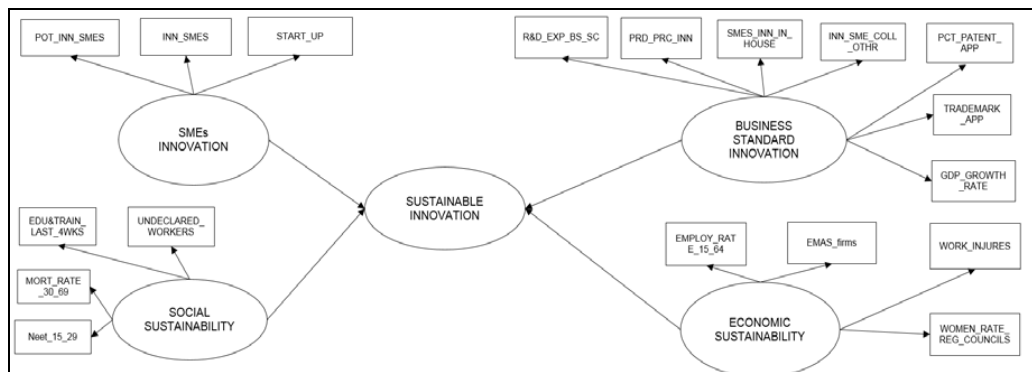
Figure 1: Path diagram.

3. Preliminary results

Since the higher order construct has no manifest variables connected to it, among the methods proposed in the literature, this contribution considers the Two-Stage (step) Approach (TSA) (Ringle et al. 2012; Wetzels et al. 2009) to state this limitation. TSA refers to the scores obtained through a principal component analysis (PCA) applied to the lower order components. All the manifest variables of the lower order construct have been treated in a reflective way (each manifest variable reflects-and it is an effect of- the corresponding latent variable), while the higher order construct involves a formative mode (see Hair et al., 2017 for an extensive evaluation of these issues). Concerning the outer model assessment, since the model is supposed to be reflective, all the blocks of manifest variables must be one-dimensional and homogeneous, and Table 2 checks the homogeneity and the one-dimensionality of the constructs. This table shows three main indices for checking the block homogeneity and one-dimensionality - Cronbach's α, Dillon-Goldstein's ρ (or Jöreskog's ρ) and the PCA eigenvalues – which confirm that the model assumptions seem to be appropriate. To prevent these indices from appearing inadequate in the estimations, several variables required a transformation since these indicators had their original scale inverted.

Table 1: Latent dependent variables, manifest variables, codes and sources.

Latent variable names	Manifest variable names	Codes	Sources
SMEs innovation	Potential innovative SMEs	POT_INN_SMES	I
	Innovative SMEs	INN_SMES	I
	Innovative start-up	START_UP	I
Business Standard Innovation	R&D expenditure business sector	R&D_EXP_BS_SC	II
	Product or process innovators	PRD_PRC_INN	II
	SMEs innovating in-house	SMES_INN_IN_HOUSE	II
	Innovative SMEs collaborating with others	INN_SME_COLL_OTHR	II
	PCT patent applications	PCT_PATENT_APP	II
	Trademark applications	TRADEMARK_APP	II
	Annual GDP growth rate	GDP_GROWTH_RATE	III
Social Sustainability	Neet (15-29)	NEET_15_29	III
	Mortality rate (leading causes of death) [30-69]	MORT_RATE_30_69	III
	Education and training activities during the last 4 weeks [percentage participation rate]	EDU&TRAIN_LAST_4WKS	III
	Undeclared workers	UNDECLARED_WORKERS	III
	Employment rate (15-64)	EMPLOY_RATE_15_64	III
Economic Sustainability	EMAS firms for 1,000 employees of local units	EMAS_FIRMS	III
	Work injuries	WORK_INJURES	III
	Women rate nominated to regional councils	WOMEN_RATE_REG_COUNCILS	III

Sources: I—Bureau van Dijk (Amadeus) database (https://amadeus.bvdinfo.com); II—Regional Innovation Index 2019 (https://interactivetool.eu/RIS/RIS_2.html); III—ASviS (https://asvis.it/database-sugli-sdgs/). Full description of each variable and more details about the sources are available on request.

Table 2: Main indices for checking the block homogeneity and one-dimensionality.

Latent variable names	Dimensions	Cronbach's α	Dillon-Goldstein's ρ	PCA eigenvalues
SMEs innovation	3	0.800	0.883	2.144; 0.510
Business Standard Innovation	7	0.874	0.909	4.243; 1.078
Social Sustainability	5	0.962	0.971	4.354; 0.423
Economic Sustainability	3	0.658	0.815	1.788; 0.738
Sustainable innovation	4			2.641; 0.664

Different variables, originally involved in the model, have been removed from the analysis due to the fact that they presented several weaknesses which require further investigation (for instance: the percentage of renewable energy consumption expressed as a percentage of final energy consumption; the energy produced by using renewable sources; the number of the spin off for regions; the usage of public transport by employees and students; etc.).

Since the SEM–PLS literature indicates several measurements to assess the quality of the outer, the inner and the global models, Figure 2 and Table 3 present the corresponding results. In more detail, Figure 2 shows the loading between (1) the manifest variables and their own latent variables and (2) the manifest variables and the remaining latent variables. This figure visually verifies that the shared variance between a construct and its indicators is larger than the variance with other constructs. Table 3 summarises the weights, the loadings, the communalities, the R^2 and the GOF.

Figure 2: Cross-loadings.

When the TSA SEM-PLS approach is performed, analysing the path coefficients, it appears that *Sustainable innovation* depends on its latent variables expressing the equation in the following form:

$$Sustainable\ innovation = 0.302BSI + 0.303SmI + 0.304EcS + 0.306\ SoS$$

The equation above indicates that all the latent constructs appear to be positively (and significantly) correlated with sustainable innovation. The coefficients are significant at the 0.05 level, and the non-parametric bootstrap procedure has been used to statistically validate the model. Supplementary findings that can derive from the latent variable scores, and more details concerning the analysis, are available on request.

Table 3: SEM-PLS assessment: indices.

Latent variables	Manifest indicators	Weighs	Loadings	Communalities	R^2	GOF
SMEs Innovation	POT_INN_SMES	0.457	0.856	0.734		
	INN_SMES	0.331	0.854	0.729		
	START_UP	0.397	0.821	0.674		
Business standard innovation	R.D_EXP_BS_SC	0.207	0.787	0.619		
	PRD_PRC_INN	0.195	0.913	0.833		
	SMES_INN_IN_HOUSE	0.185	0.914	0.835		
	INN_SME_COLL_OTHR	0.084	0.466	0.217		
	PCT_PATENT_APP	0.263	0.906	0.820		
	TRADEMARK_APP	0.215	0.876	0.767		
	GDP_GROWTH_RATE	0.068	0.370	0.137		
	Neet_15_29	0.231	0.985	0.971		0.829
Social sustainability	MORT_RATE_30_69	0.171	0.813	0.662		
	EDU.TRAIN_LAST_4WKS	0.199	0.907	0.823		
	UNDECLARED_WORKERS	0.227	0.973	0.947		
	EMPLOY_RATE_15_64	0.238	0.975	0.951		
Economic sustainability	EMAS_firms	0.371	0.677	0.458		
	WORK_INJURES	0.477	0.800	0.641		
	WOMEN_RATE_REG_COU NCILS	0.443	0.829	0.687		
Sustainable innovation					0.999	

4. Future work

The preliminary significant and positive relationships presented in this work require a certain caution in analysing the interaction among the *Sustainable Innovation* and its latent constructs. Potential awareness might be relevant from a policy point of view considering that the topic of the study is the analysis of the effects that may affect *Sustainable Innovation*. Prospective research endeavours could consider several model modifications to strengthen sustainable strategies and, in future investigations, the number of indicators and the contextual factors may also be extended. Supplementary considerations can originate from the possible causal relationships between the manifest variables and/or different constructs, which can also have an impact on *Sustainable Innovation.* Since the path coefficients represent the direct effects, it is important to evaluate the indirect effects. In addition, the interaction effects – which refer to the influence that an additional variable might have on the relationship between an independent and a dependent variable – can be investigated as well. According to a similar perspective, the analysis of moderating effects - which imply the involvement of a variable as a moderator indicator and which could change the strength and the direction of a relationship between the constructs in the model - cannot be omitted.

References

Carrillo-Hermosilla, J., del Río, P., Könnölä, T. (2009). *Eco-innovation. When Sustainability and Competitiveness Shake Hands*. Palgrave. London (UK).

Carrillo-Hermosilla, J., del Río, P., Könnölä, T. (2010). Diversity of Eco-Innovations: Reflections from Selected Case Studies. *Journal of Cleaner Production,* **18** (10–11), pp. 1073–1083.

Durst, S., Edvardsson, I. R. (2012). Knowledge Management in SMEs: A Literature Review. *Journal of Knowledge Management*, **16** (6), pp. 879–903.

Dziallas, M., Blind, K. (2019). Innovation Indicators Throughout the Innovation Process: An Extensive Literature Analysis. *Technovation*, **80-81**, pp. 3–29.

European Commission (2019). Regional Innovation Scoreboard, Luxembourg, available at https://ec.europa.eu/growth/sites/growth/files/ris2019.pdf

European Commission (2020). European Innovation Scoreboard, Luxembourg, available at https://ec.europa.eu/docsroom/documents/42981/attachments/1/translations/en/renditions/native

Gössling, T., Rutten, R. (2007). Innovation in Regions. *European Planning Studies*, **15,** pp. 253–270.

Hagedoorn, J., Cloodt, M. (2003). Measuring Innovative Performance: Is there an Advantage in using Multiple Indicators? *Research Policy*, **32**(8), pp. 1365–1379.

Hair, J.F., Sarstedt, M., Ringle, C.M., Gudergan, S.P. (2017), *Advanced Issues in Partial Least Squares Structural Equation Modeling*, Sage, Los Angeles.

Makkonen, T, van der Have, R. P. (2013). Benchmarking Regional Innovative Performance: Composite Measures and Direct Innovation Counts. *Scientometrics,* **94**(1), pp. 247–262.

Organisation for Economic Co-operation and Development - OECD (2018). Oslo Manual 2018: *The Measurement of Scientific and Technological Activities: Guidelines for Collecting, reporting and using data on Innovation*. OECD, Paris.

Ringle, C.M., Sarstedt, M. and Straub, D.W. (2012). Editor's comments: A critical look at the use of pls-sem in "mis quarterly". In *MIS quarterly*, iii–xiv.

Rothwell, R. (1992). Successful industrial innovation: Critical factors for the 1990s. *R&D Management*, **22**(3), pp. 221–238.

Smith, K. (2005). Measuring Innovation. In Fagerberg, J., Mowery, D., Nelson, R. (Eds.), *The Oxford handbook of innovation*. Oxford University Press, New York, pp. 148–177.

Sood, A., Tellis, G. J. (2005). Technological Evolution and Radical Innovation. *Journal of Marketing*, **69**(3), pp. 152–168.

Wetzels, M., Odekerken-Schröder, G., Van Oppen, C. (2009). Using PLS Path Modeling for Assessing Hierarchial Construct Models: Guidelines and Empirical Illustration. *MIS Quarterly*, **33**(1), pp. 177–195.

The Financial Wellbeing Index: "Donne al quadrato" and the relevant impact measurement

Claudia Segre, Serena Spagnolo, Valentina Gabella, Valentina Langella

1. Introduction

Financial well-being describes the condition in which a person can fully meet current financial obligations, feel secure in their financial future and is able to make autonomous choices. The expression itself, "financial well-being", underlines how the economic and financial aspects are inextricably linked to our individual and social well-being. Helping people to improve their financial well-being, in a broad sense, is, therefore, the first impact indicator that financial education professionals must ask themselves.

For this reason, the Global Thinking Foundation has decided to measure the impact of the Donne al quadrato project through collaboration with ALTIS - Università Cattolica, analyzing activities' progress to identify the impact of the project, of its strengths and weaknesses and possible paths for improvement and enhancement. The intervention developed along two main axes:

- Scientific review, validation and expansion of stakeholder engagement (by increasing the number of courses monitored), methodology and measurement process in place;
- Definition of a synthetic indicator of Financial Wellbeing, to assess the overall effects generated within the Donne al quadrato project. This indicator intends to measure the level of security and freedom of people regarding their economic situation and their financial capabilities, considering micro indicators, specific to the sample examined, and macro variables, belonging to the territorial context, to purify the changes measured by the macroeconomic effects affecting the entire population.

The conceptualisation of the theoretical reference framework for measuring the impact generated started from analysing the literature on the mechanisms that regulate people's financial behaviours. These aspects can be modified by didactic-training activity.

The analysis of the literature shows that can place three concepts at the basis of the definition of people's financial behaviour: financial well-being ((Delafrooz & Paim, 2011; Gerrans, Speelman & Campitelli, 2014), financial literacy (Atkinson e Messy (2011), (2012)) and financial capability (Holzmann et al. 2013; Kempson et al, 2013a; Kempson et al. 2013b).

After examining various theories, in this study, it was decided to use the analysis of the links between literacy and financial capacity and of the main components of financial well-being described in the following guide "Financial Well-Being" (Kempson et al. 2017).

2. The Financial Wellbeing Index

Methodology

Using the framework theorized by Kempson, for measuring of the components of financial well-being decided to set up a synthetic indicator.

The Financial Wellbeing Index (FWI) was designed to provide an accurate, consistent and comparable measure over time of how much participation in the Donne al quadrato course has influenced people's perception of security and freedom, about their economic situation and their financial capabilities. The FWI, conceived based on of a methodology used by the University of Bristol (Hayes, D., Evans, J., & Finney, A., 2016), aims to provide a complete, concise and easily communicable image, which describes the impact picture and its evolution over time (trend).

The index is measured on a scale of 0 to 180, where higher scores represent greater financial well-being. As shown in figure 1, 83% of the overall score of the index is based on a micro index,

Claudia Segre, Global Thinking Foundation, Milan, Italy, claudia.segre@gltfoundation.com
Serena Spagnolo, Global Thinking Foundation, Milan, Italy serenaspagnolo2407@libero.it
Valentina Gabella, Catholic University of Sacro Cuore, Italy, valentina.gabella@unicatt.it
Valentina Langella, Catholic University of Sacro Cuore, Italy, valentina.langella@unicatt.it

calculated on the basis of the results collected thanks to the administration of a questionnaire to participants and participants in courses provided by the Foundation. The remaining 17% is made up of the index's macro component, which is constructed using three nationally recognized economic indicators of a territorial context (Istat). The overall score of the index is calculated by adding the macro index to the results of the individual micro indexes of the respondents.
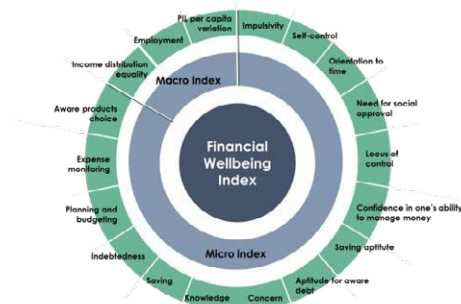
Figure 1 - Graphical representation of the FWI.

Index creation and composition

The overall index comprises the sum of the values of fifteen individual micro aspects and three macro aspects, as explained below.

Macro component

The index's macro component is based on three macroeconomic indicators chosen to provide a global overview of the economy on a national and regional basis. This result in measure the level of employment, the equality of income distribution (GINI coefficient) and the variability of per capita GDP.

For the calculation of this component, the recent historical values provided by Istat were used and rescaled so that they provide a score on a scale of 1-10, where a higher score always corresponds to a scenario of greater socioeconomic well-being, or levels of falling unemployment, and high variability of per capita GDP values and a higher level of equality. Note that since the macro component provides a snapshot of the macroeconomic context, it is invariant for all individuals for which the index was calculated.

The methods for calculating the three elements are detailed below.

Occupation

To calculate the employment score, the historical series, provided by Istat, of the annual unemployment rate for the last nine years, at a national level and for each region, was reworked. At the regional level, the choice not to limit ourselves to using the precise data of the latest survey was made to contextualize the data in their recent historical evolution. The range of variability of the time frame under consideration was, therefore calculated as follows. The minimum and maximum of the historical series of all the Italian regions were considered, and a buffer was applied to them to take into account the uncertainty of these limits (10% of the average point between the minimum and maximum). The minimum and maximum used, for regional scores, are therefore unique and the same for all regions.

As regards to the national data, on the other hand, the minimum and maximum of the relative historical series were used, without reference to the regional curves. For this reason, it may happen that the national score is not directly comparable with the regional ones (for example it does not represent the weighted average of the regional scores); in fact, the two types of data represent slightly different concepts: the employment rate is quantified and re-proportioned, at a national level, concerning to its evolution over time, while at the regional level with respect to time and connections with other regions.

The score was then calculated by reporting the most recent data of the historical series from the variability interval to the scale of 1-10 and then considering the reciprocal, so that higher

unemployment levels correspond to lower scores and vice versa. In other words, if T_d is the most recent unemployment rate and U and L the maximum and minimum limits of the range of variation, calculated as described above, the employment score is determined using the formula:

$$P_o = 10 - \frac{T_d - L}{U - L} 10$$

Equality of income distribution

A similar procedure was applied to the historical series of the Gini coefficients (source Istat) to calculate the income distribution equality score.

The Gini coefficient is, in fact, an internationally used measure of the inhomogeneity in the distribution of net household income within a country. It is calculated by comparing the effective distribution of income with the theoretically entirely fair one: a Gini coefficient of 0 represents a perfect income distribution equality (in which 10% of the population receives 10% of the national income, 50 per cent receives 50%, etc.), while a coefficient of 100 represents perfect inequality (in which only one person receives 100% of the income).

Similarly, to what is described for the occupancy score, to determine the variability interval of the regional data, the lower and upper thresholds of the variability interval of all the regional time series of the Gini coefficients starting from 2010 were calculated. For the national interval, on the other hand, and correspond to the minimum and maximum of the national time series. Also, in this case, the national score may not be directly comparable with the regional ones since the two types of data represent slightly different concepts: the Gini coefficient is quantified and re-proportioned, at a national level, in relation to its evolution over time, while at a regional scale for time and relations with other regions.

The income distribution equality score was then obtained by reporting T_g, the Gini coefficient for the available last year, from the variability interval L - U on the scale of 1 - 10 and considering the reciprocal, as follows:

$$P_g = 10 - \frac{T_g - L}{U - L} 10$$

Change in GDP

The methods for calculating the score relating to the change in per capita GDP were completely similar to those described for the unemployment and income equality scores. The only peculiarity of this indicator lies in the fact that the historical series of GDP values tout-court were not used, which are by no means representative to define the FWI. Still, those of the percentage changes compared to the previous year. Indeed, long-term GDP per capita, in absolute terms, tends to increase. At the same time, its annual variation gives a more accurate measure of the real improvement or deterioration of the local macroeconomic context. In other words, the proposed GDP change score observes the pace of growth (or decline) of the Gross Domestic Product indicator and a static economy.

In other words, the proposed GDP change score observes the growth rate (or decline) of the Gross Domestic Product indicator and a stagnant economy, without annual growth or decline, would therefore obtain 5 points out of 10.

Once the limits of the L and U variability interval have been determined, with the two regional-national methods, as described for the other indicators, the score relating to the variation in GDP is calculated using the following formula:

$$P_P = \frac{D_P - L}{U - L} 10$$

Note that in this case it is not necessary to consider the reciprocal (10-) since high changes in GDP will have to correspond to high GDP scores. The regional and national scores are shown below, calculated in the manner described above, in the two updates T0 (2019) and T1 (2020). For the first update, the 2009-2017 time series was considered, while for the second, 2010-2018.

Micro component

For the structuring of the theoretical reference framework and the data collection questionnaire relating to the micro component of the FWI, the Financial Wellbeing Conceptual Model proposed by prof. Elaine Kempson (University of Bristol).

This model starts from the definition of financial well-being, on the basis of the three elements that make it up and attempts to describe it by taking into consideration the relationships between four key issues that influence it. The elements that make up financial wellbeing are:

1. The ability to meet financial commitments (e.g. rents, utility bills, and loan payments)

2. The extent to which people have felt comfortable with their financial situation over the past year, how comfortable they imagine they will be in the near future, and how their finances have allowed them to enjoy life

3. Resilience for the future - the ability to cope with a significant unexpected expense or decline in revenue.

While the four key themes are:

- *Social and economic environment*;
- *Financial knowledge and experience acquired;*
- *Attitudes, motivation and biases*, or psychological factors such as attitudes, motivations and cognitive biases;
- *Financial capable behaviour*, or financially aware behaviour.

The micro component of the index was formulated on the basis of this theoretical model. The corresponding themes selected, to which the fifteen aspects that make up the component refer, are: Personality, Knowledge, Attitudes and Behaviours (Table1). Each of these aspects is evaluated on a scale from 1 to 10 and is added to the others with equal weight, thus originating a maximum score of 150 points.

Theme	Topic
Personality	Time orientation
	Impulsiveness
	Need for social approval
	Self-control
	Locus of control
Knowledge	Financial knowledge
Attitudes	Saving attitude
	Attitude for aware debt
	Confidence in one's ability to manage money
	Concern about one's future financial situation
Behavior	Saving ability and awareness
	Recourse to debt
	Planning and budgeting
	Expense monitoring
	Informed choice of products

Table 1 – Structure of the micro component of the Financial Wellbeing Index

Below is a brief description of each of the aspects considered:

- *Time orientation*: propensity of the individual to think in perspective, to plan and to focus on the long term;
- *Impulsiveness*: propensity of the individual to ponder and evaluate situations in detail before acting;
- *Need for social approval*: propensity of the individual to seek acceptance and esteem for the social context to which he or she belongs;
- *Self control*: propensity of the individual to control their impulses;
- *Locus of control*: propensity of the individual to believe that the events of his existence are caused by internal causes (his behaviour and his actions) or external (chance, actions or will of others) independent of his will;

- *Financial knowledge*: survey of some basic economic and financial knowledge, discussed during the course;
- *Savings attitude*: aptitude of the individual to spend and save;
- *Attitude to aware debt*: aptitude of the individual to be aware of the debts he intends to contract (e.g. purchases in instalments);
- *Confidence in one selves ability to manage money*: aptitude of the individual to think that he has the knowledge and skills necessary to understand the economic and financial choices presented to him;
- *Fear and concern about the financial situation of the following year:* frequency with which the individual tends to be in a state of apprehension about his own economic possibilities, relative to the next twelve months (not having the capacity to save, not meeting the debts, being unemployed or in a job that is not profitable enough);
- *Savings ability and awareness:* what percentage of income is saved and in what instruments it is invested in;
- *Recourse to debt:* frequency with which the individual incurs debts to meet daily or unexpected needs;
- *Planning and budgeting:* the extent to which the individual plans his future expenses and allocates resources to different areas;
- *Expense tracking:* extent to which the individual monitors his past spending and savings;
- *Informed choice of products*: extent to which the individual tends to inquire about possible products to buy, both financial and non-financial.

The micro component of the FWI was assessed thanks to the processing of the data obtained from the questionnaires administered to the students, direct beneficiaries of the action of the GLT Foundation.

3. The questionnaire

The questionnaire, designed to assess the micro-component of the financial well-being index, includes 2-4 questions for each aspect described in the previous section.
The survey is structured and consists of 60 closed questions, 6 of which from the personal data, which capture a specific aspect of financial well-being, relative to the element in question.

The scores of the items, on a scale of 1-10, have been attributed in such a way that a higher score always corresponds to a higher level of financial well-being.

The questions were all similarly weighted within each domain, i.e. their score contributes equally to the score of the area evaluated.

4. The experimentation

The case: the Donne al Quadrato project

The Donne al Quadrato project conceived, implemented and promoted by the Global Thinking Foundation allowed the FWI experimentation to start, with the aim of measuring the social impact on the participants of the financial education courses provided within the project.

Results

This report presented research results regarding the assessment of the impacts of the Donne al quadrato financial training course in the year 2019/2020.

To show the impacts, reference was made to financial well-being described by different dimensions, subjective and objective, which make up people's financial behaviour.

The construction of a synthetic index, based on the studies of the World Bank and the University of Bristol, has made it possible to analyse a series of objective and subjective financial characteristics and statistically describe the way in which various components relate to the financial well-being of a group of people. The experimentation was then carried out on samples from different geographic regions and at different times. Therefore, the index provides a holistic method for measuring the financial well-being of individuals over time and space. The results of the trial showed that financial education could generate a range of changes not only in knowledge but also in financial skills and behaviour, as well as the financial well-being of participants. The findings help us understand the role of "what people know and do" for their financial well-being.

Financial education can help individuals improve, their financial situations and ultimately their financial well-being by helping them improve their economic attitudes and behaviour.

The results show that the aspect that recorded the most consistent growth concerns the individual's aptitude to be aware of the debts he intends to contract (e.g. instalment purchases) (Aptitude to aware debt +21%), also followed by personality aspects, such as the individual's propensity to ponder and evaluate situations in detail before acting (Impulsivity + 10%) and the individual's propensity to believe that the events of his existence are caused by internal causes (his behaviour and his actions) (Locus of control + 9%). Also, with regard to behaviours, the results show a significant improvement, for example, in the extent to which the individual monitors his expenses and savings (Monitoring of expenses + 12%). In particular, the study suggests the need to increase plans and projects for the development of financial skills and attitudes which, through the generation of virtuous behaviour, reduce fears about one's economic possibilities (not having the capacity to save, not meeting the debts contracted, being unemployed or in a job that is not profitable enough, etc...) increasing self-efficacy and financial well-being.

References

Atkinson, A., Messy, F. A. (2011). Assessing financial literacy in 12 countries: an OECD/INFE international pilot exercise. *Journal of Pension Economics & Finance*, **10**(4), pp. 657-665.

Atkinson, A. and F. Messy (2012), Measuring Financial Literacy: Results of the OECD / International Network on Financial Education (INFE) Pilot Study, *OECD Working Papers on Finance, Insurance and Private Pensions*, No. **15**, OECD Publishing, Paris, (FR)

Delafrooz, N., Paim, L. H. (2011). Determinants of financial wellness among Malaysia workers. *African Journal of Business Management*, **5**(24), 10092-10100.

Gerrans, P., Speelman, C., Campitelli, G. (2014). The relationship between personal financial wellness and financial wellbeing: A structural equation modelling approach. *Journal of Family and Economic Issues*, **35**(2), pp. 145-160.

Hayes, D., Evans, J., Finney, A. (2016). *Momentum UK Household Financial Wellness Index: Wave one*. Personal Finance Research Centre (PFRC), University of Bristol for Momentum UK.

Holzmann, R., Mulaj, F., & Perotti, V. (2013). *Financial capability in low-and middle-income countries: measurement and evaluation*, World Bank, Washington, DC, (USA).

Kempson, E., Perotti, V., Scott, K. (2013a). *Measuring financial capability: a new instrument and results from low-and middle-income countries*, World Bank, Washington, DC, (USA).

Kempson, E., Perotti, V., Scott, K. (2013b). *Measuring Financial Capability: Questionnaires and Implementation Guidance for Low-, and Middle-Income Countries*, World Bank, Washington, DC, (USA).

Kempson, E., Finney, A., & Poppe, C. (2017). Financial well-being, a conceptual model and preliminary analysis, *SIFO Working Paper* No. **3**-2017, Oslo and Akershus University College of Applied Sciences, Oslo, (NO)

Health and well-being

Determinants of social startups in Italy

Lucio Palazzo, Pietro Sabatino, Riccardo Ievoli

1. Introduction

The so called Startup Act (Decree Law 179/2012, converted into Law 221/2012), has introduced in Italy the notion of innovative companies with a high technological value, i.e., the innovative start-ups. Among them, the Italian government includes the category of social startups, i.e., "*startup innovative a vocazione sociale*" (hereafter SIAVS), representing a relatively new field of interest in both scientific and normative perspective.

SIAVS must satisfy the same requirements of other innovative startups, but operate in sectors such as social assistance, education, health, social tourism and culture, enjoying also some tax benefits. Furthermore, they have a possible direct (social) impact on the collective well-being, measured through a self-evaluation document named: "Documento di Descrizione dell'Impatto Sociale" published yearly by each SIAVS (Vesperi, Lenzo). Today, social startups are more than doubled with respect to five years ago[1].

Within Italian academic debate concerning startups and innovative economic enterprises, SIAVS have been considered for their hybrid nature, balancing between profit and non-profit model of business, and for their role of producing value for local communities (Vesperi et al, 2015). Although there are some recent empirical studies on social entrepreneurship intentions (Bacq et al., 2016), little is known about territorial pattern of SIAVS, even if a certain similarity has been observed, at regional scale, with the territorial distribution of overall startups (Maglio, 2019). Italian non-profit organizations present different characteristics compared to innovative companies, notably on gender balance in workforce and territorial diffusion (Istat, 2019; Forum del Terzo Settore, 2017).

The aim of this paper is to investigate the relevant factors influencing the presence of social startups in Italy at the provincial level. The outcome variable is the number of active social startups in Italian provinces while the set of explanatory variables is composed by economic and demographic indicators at the provincial level.

Regarding the explanatory variables, unemployment rate and number of incubators have been used as predictors of the number of startups at regional level in Colombelli (Quartaro), while Hoogerndoorn (2016) considers the GDP per capita. Information regarding registered firms at the provincial level can be found also in the work of Colombelli et al. (2019) to predict the number of new firms at the provincial level (NUTS 3 regions). Furthermore, the effectiveness of incubators for Italian startups is still under debate (Deidda Gagliardo et al., 2017), while Sansone et al. (2020) have introduced a new taxonomy, distinguishing between *business*, *mixed* and *social* incubators. We also consider other variables as *broadband*, which can be viewed as a proxy of the technological level of a province, and the percentage of *NEET* (neither in employment or in education or training between 15 and 29 years) which is a measure of non-attractiveness of a territory for the young people.

Generalized linear models (GLM) for discrete outcomes are applied and compared, even taking into account the zero-inflated issue arising due to the distribution of these particular data.

---

[1] "Relazione Annuale al Parlamento sullo stato di attuazione e l'impatto delle policy a sostegno di startup e PMI innovative" `www.mise.gov.it`

Lucio Palazzo, University of Naples Federico II, Italy, lucio.palazzo@docenti.unisob.na.it, 0000-0001-7529-4689
Pietro Sabatino, University of Naples Federico II, Italy, piesabatino@gmail.com, 0000-0002-3538-6201
Riccardo Ievoli, University of Ferrara, Italy, riccardo.ievoli@unife.it, 0000-0001-9489-3564

2. Materials and Methods

Data

Information regarding startups and certified incubators are retrieved from the Italian Chambers of Commerce[2], updated to the third quarter 2020. Other additional variables, at the provincial (NUTS 3) and regional (NUTS 2) level, and the spatial coordinates of these provinces, are obtained through the Italian National Institute of Statistics[3] (ISTAT) and European Statistical Office [4] (EUROSTAT).

A possible drawback is that some variables suffer from timeliness issue. Moreover, for the purpose of this explorative study, this issue seems less severe considering the reasonably not too high variations occurring in the short term period at provincial level. Thus, we retrieved the latest update (i.e., the value for the last available year) for all considered covariates. In some cases we consider the geometric mean to avoid problems related to possible temporal variations.

Measurement Variables

The dependent variable is the count of SIAVS in Italian provinces. Therefore, the sample size is equal to $n = 105$, composed by all Italian provinces except for "Sud Sardegna" and "Andria-Trani-Barletta", which do not include any kind of startup in their territory.

As mentioned, we identified the following candidates as possible determinants for the presence of SIAVS (the latest update is in brackets):

Population density, number of inhabitants divided by the area of province (reference year: 2017). A logarithm transformation is applied;

GDP per capita, in thousand of euros (reference year: 2017);

Incubators, count of *certified* incubators in each province. These particular companies (Decree Law 179/2012) are registered in the Italian Chambers of Commerce and offer services for the developing of startups (reference year: 2020);

Unemployment rate, at the provincial level (reference year: 2019);

Registration rate, companies registered in a year divided by the total number of companies registered in the previous year in the Italian "registro imprese" at the provincial level (geometric mean between 2015 and 2018);

Broadband, number of ultra-broadband subscriptions as a percentage of inhabitants in each provinces (geometric mean between 2015 and 2017);

Social employees, rate of workers in social cooperatives (reference year: 2019)

NEETs, percentage of neither in employment or in education or training pepole between 15 and 29 years (reference year: 2019).

Statistical Models

The number of SIAVS in Italian provinces can be modelled applying GLM family (see e.g. Nelder, Wedderburn; McCullagh, Nelder, among others). The general formulation of GLM (Agresti, 2003) is carried out through a link function $g(\cdot)$, which transforms the expectation of the response variable, i.e. $\mu_i = E(Y_i)$, to the linear predictor:

$$g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}, \qquad i = 1, \ldots, n. \tag{1}$$

where $p = 8$ is the number of variables previously discussed.

---

[2]https://www.registroimprese.it/
[3]https://www.istat.it/
[4]https://ec.europa.eu/eurostat

In this context, two main competing models can be considered: Poisson (POI) and Negative Binomial (NB) regression. In the former case, $Y_i \sim Poi(\lambda_i)$ and the corresponding log-link function is $g(\mu_i) = \log \lambda_i$, while in the latter case $Y_i \sim NegBin(\mu_i, \omega)$. In the POI model, the observed counts are equidispersed, i.e. $E(Y_i) = Var(Y_i) = \mu_i$. Moreover, the scale parameter ω in NB model takes into account for the presence of overdispersion i.e. $Var(Y_i) = \mu_i + \mu_i^2/\omega$.

A possible issue related to the count of SIAVS (and startups in general) is the possible presence of excess of zeros in the data, i.e. provinces without any registered SIAVS. Thus, previously introduced models may be modified to take into account the zero inflation. The zero inflated Poisson (ZIP) model is derived as a mixture of a binary logistic and POI (Lambert, 1992). The responses Y_i are independent and $Y_i \sim 0$ with probability π_i and $Y_i \sim Poi(\lambda_i)$ with probability $1 - \pi_i$. The resulting link function can be written as follows:

$$g(\mu_i) = \begin{cases} \log \frac{\pi_i}{1-\pi_i}, & \text{if } y_i = 0 \\ \log \lambda_i, & \text{if } y_i > 0 \end{cases} \tag{2}$$

The zero inflated NB (ZINB) model, introduced in Greene (1994), is derived by substituting the POI link function with the NB (when responses are not equal to zero). We remark that ZIP and ZINB assume that the zero inflation effect is generated by a separate process apart from the count values.

3. Results and Discussion

The number of SIAVS is equal to 240 and the 87.5% of them are classified in the service sector. The remaining 12.5% is divided in industry and/or craft sector (7.9%) and sectors such as agriculture, tourism and commerce (4.6%). Registered SIAVS present almost 40 activity codes. The main activities of SIAVS can be divided in: *a)* software production and IT consultancy (17.5%), *b)* scientific research and development (12.9%), *c)* education (10.4%), *d)* information and other services (9.6%), *e)* non-residential social assistance (8.8%), *f)* activities related to libraries, archives and museums (3.8%) *g)* art and entertainment (2.9%). The remaining 34.2% of SIAVS are classified in 33 different activity codes.

Almost a quarter of SIAVS (24.2%) is located in the province of Milan (58), while provinces of Rome and Turin include respectively 27 (11.2%) and 13 (5.4%) SIAVS. In general, 65 provinces (62%) contain almost a SIAVS but only 20 (19%) of them registered more than 2 social startups. SIAVS also present a higher frequency of female prevalence (measured in terms of at least 50% of women in the company) compared to other startups, exceeding them by the 10%. Moreover, differences can not be found in practice (with respect to other startups) regarding the proportion of young people (under 35) and foreigners.

In Figure 1 we can observe the distribution of SIAVS (left panel), the distribution of startups (center panel) at the provincial level and the distribution of non-profit institutions at the regional level (right panel). Main differences between startups and SIAVS can be viewed in the provinces of Centre Italy. Nonetheless, startups and SIAVS are concentrated in the metropolitan areas (especially in the provinces of Milan and Rome) and also the non-profit subjects can be found especially in the North-East (Lombardia Region). In addition, the provinces of Sardinia present the lower counts of startups and SIAVS, even if the number of non-profit institutions appears comparable with respect to the other regions.

Table 1 summarizes the main results of statistical models discussed in Section 2. First of all, we check the usefulness of the whole set of regressors in all models, by observing the decreasing of the Bayesian Information Criterion (BIC) between the null models (BIC_0), including only the intercept, and the models with all considered covariates. For each model, the BIC is function

of a different likelihood, and the decreasing is more (numerically) evident in the POI and ZIP models than in the NB and ZINB. Another similar check can be also carried out (only for the first two models) through the McFadden's *Pseudo* R^2. We also report, for each model specification, the likelihood ratio test statistic to formally test for the departure from the "null" model (which only includes the intercept) and its associated p-value. This check also confirms the usefulness of proposed regressors. We have to remark that it is not possible to make a proper comparison between the four models in terms of likelihood-based statistics. Therefore, we use a leave-one-out cross-validation (CV) approach to compare the prediction of four models, estimating $R = 105$ times the model and then computing $MSE(CV) = n^{-1} \sum_r (\hat{y}_r - y_r)^2$. Regarding this performance indicator, the conventional POI exhibits the lower MSE(CV), followed by the ZIP and ZINB. Finally, the (here not reported) results of two Vuong tests (Vuong, 1989) suggest the rejection of null hypothesis of POI and NB in favour of ZIP and ZINB.
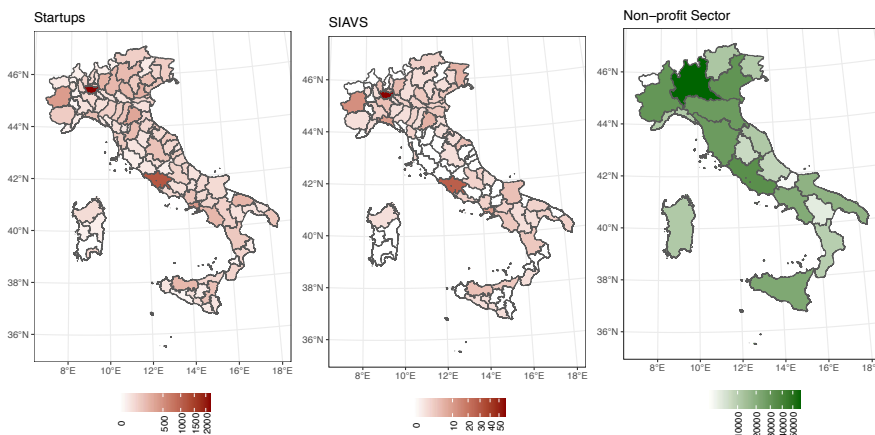
Figure 1: Geographical distribution of number of startups, number of SIAVS (provincial level) and non-profit sector (regional level).

Conventional GLM models help to identify log population density, (certified) incubators and broadband as positive determinants of the counts of SIAVS at the provincial level considering a nominal error rate of the 1%. Conversely, in more robust zero-inflated regressions, the coefficient of population density is no longer statistically significant. Therefore, in ZIP and ZINB, unemployment rate is identified as a possible positive driver for the arise of SIAVS, while the percentage of young people neither in employment or in education or training can be considered as a negative indicator for the arise of SIAVS. Surprisingly, GDP per capita and social employees are not statistically significant in any considered model.

Certified incubators appears fundamental for the presence of SIAVS. At a descriptive level, 64% of SIAVS (153) is located in provinces including almost a certified incubator. This percentage is slightly lower considering all innovative startups (56%).

To conclude, SIAVS arise in provinces with higher technological levels, including ecosystem to develop and assist startups. Basing on our results, also population density and unemployment may have an influence on the presence of SIAVS, but further investigation will be

Table 1: Input variables setting scheme used in each model.

	POI	NB	ZIP	ZINB
Intercept	0.7569	-0.3718	0.7480	0.7441
log(Population density)	0.3491 ***	0.3949 ***	0.1636	0.1637
Incubators	0.2373 **	0.2326 *	0.2269 **	0.2269 **
Registration rate	-0.2995 .	-0.2228	-0.1976	-0.1972
Unemployment rate	0.0468	0.0409	0.1512 ***	0.1512 ***
Broadband	0.1969 ***	0.1810 ***	0.1987 ***	0.1987 ***
GDP per capita	-0.0304	-0.0184	-0.0146	-0.0146
Social employees	-0.0746	-0.0591	-0.0500	-0.0499
NEETs	-0.0312	-0.0247	-0.0764 **	-0.0764 **
BIC_0	803.8809	415.8049	714.4403	420.4589
BIC	359.6001	362.5449	357.1889	361.8049
McFadden's R^2	0.6025	0.2226	-	-
LR Test	481.5100 ***	90.4920 ***	431.7100 ***	133.1200 ***
MSE(CV)	6.1835	25.5918	10.7048	10.7284

Significance codes: $0 \leq `***` < 0.001 \leq `**` < 0.01 \leq `*` < 0.05 \leq `.` < 0.1 \leq ` ` < 1$

conducted at the territorial level.

Future interesting analysis will concern the trend of new SIAVS in time (using quarterly data), even considering autoregressive models for integer data (see e.g Palazzo, 2019).

References

Agresti, A. (2003). Categorical data analysis. Second Edition (Vol. 482), John Wiley & Sons, New York (US).

Bacq, S., Hartog, C., Hoogendoorn, B. (2016). Beyond the moral portrayal of social entrepreneurs: An empirical approach to who they are and what drives them. *Journal of Business Ethics*, **133**(4), pp. 703–718.

Colombelli, A., Quatraro, F. (2019). Green start-ups and local knowledge spillovers from clean and dirty technologies. *Small Business Economics*, **52**(4), pp. 773–792.

Colombelli, A., Orsatti, G., Quatraro, F. (2019). Local knowledge composition and the emergence of entrepreneurial activities across industries: evidence from Italian NUTS-3 regions. *Small Business Economics*, **56**, pp. 613–635.

Deidda Gagliardo, E., Gobbo., G., Papi. L., Bigoni, M. (2017). L'efficacia dei programmi di incubazione nel successo e nel fallimento delle startup. *RIREA*. (5-8), pp. 225–239.

Forum del Terzo Settore (2017). Le reti del Terzo Settore. Terzo Rapporto. Eds. Peruzzi, G. Forum Nazionale del Terzo Settore, Rome (IT), pp. 1-72.

Greene, W.H. (1994). Accounting for excess zeros and sample selection in Poisson and negative binomial regression models. Working paper No. 94-10 Stern School of Business, New York University, New York (US), pp.1–36.

Hoogendoorn, B. (2016). The prevalence and determinants of social entrepreneurship at the macro level. *Journal of Small Business Management*, **54** (sup1), pp. 278–296.

ISTAT (2019). Struttura e profili del settore non profit.
Report censimento permanente su istituzioni non profit anno 2017. ISTAT, Rome (IT), pp. 1–15.

Maglio, M. (2019). Il framework territoriale nel processo di diffusione del fenomeno delle startup a vocazione sociale. *XIII Colloquio Scientifico sull'impresa sociale*, 24-25 maggio 2019, Università degli Studi di Roma Tor Vergata, Rome (IT), pp. 1–14.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**(1), pp. 1–14.

McCullagh, P., Nelder, J. A. (1989). Generalized linear models. Second Edition, *Chapman and Hall*, New York (US).

Nelder, J. A., Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, **135**(3), pp. 370–384.

Palazzo, L. (2019) Integer Autoregressive modeling: a new gretl routine, in *Proceedings of the International Conference on the GNU Regression, Econometrics and Time-series Library*, eds. F. Di Iorio, R. Lucchetti, Università degli studi di Napoli Federico II, Naples (IT), pp. 95–103.

Ventura, M., Vesperi, W., Reina, R. (2018). Innovazione sociale nelle startup innovative: un approccio narrativo. *Iris Network*, pp. 1–18.

Sansone, G., Andreotti, P., Colombelli, A., Landoni, P. (2020). Are social incubators different from other incubators? Evidence from Italy. *Technological Forecasting and Social Change*, **158**, pp. 120–132.

Vesperi, W., Lenzo, P. (2016). La start-up innovativa a vocazione sociale e il documento di descrizione dell'impatto sociale: una comparazione di casi. *Colloquio scientifico sull'impresa sociale, 10-11 giugno 2016*, Università degli studi di Napoli Federico II, Naples (IT), pp. 1–18.

Vesperi, W., D'Aleo, V., Morabito, G., Lo Bue, S. (2015). Imprenditorialità vs. responsabilità sociale: La startup innovativa a vocazione sociale. *Colloquio scientifico sull'impresa sociale, 22-23 maggio 2015*, Università degli studi Mediterranea di Reggio Calabria, Reggio Calabria (IT), pp. 1–14.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, **57**(2) pp. 307–333.

Multipoint vs slider: a protocol for experiments

Venera Tomaselli, Giulio Giacomo Cantone

1. Introduction

Since the 1990s, in all fields involving survey tools aimed at collecting data from a sample of a target population, computer-assisted technologies of data recording replaced the old *paper-&-pen*. The speed of technological shift was not paired by methodological innovations.

Multipoint scales, indeed, are still among the most employed numerical (or semantic) supports for many variables in psychological, health, socio-economic research, and even in engineering (e.g., user experience design). With the spread of 'Big Data', an old issue in statistical measurement gained a new relevance. It can be shortly summarized: tons of Big Data from self-reports of taste and perception are recorded every day. While these data are reported through multipoint scales, almost all the relevant inferences are made through families of methods with parametric assumption, for example, one of the most notorious methodology to infer human preferences through analysis of similarity, *collaborative filtering* (Kluver, Ekstrand, and Konstan 2018).

The debate about the plausibility of an estimation of central value in ordinal variables (which is the core of the debate about parametric methods for analysis of 'ratings') is well summarised by Velleman and Wilkinson (1993). Kampen and Swyngedouw (2000) expanded the issue relating it the consequential debate about derivative measures of association and correlation among variables (also, see, Agresti 2010). Tomaselli and Cantone (2020) highlighted a more recent issue in data analysis: when the number of items compared (e. g, a ranking) exceeds too much the categories of the supporting ordinal scale, the comparison is made impossible by the high amount of tie cases. Therefore, statistics constrained in the support scale (i.e., the median) are unfeasible to index distributions from very large samples, or populations. This problem of ranking statistics could be interpreted as an extreme case of 'ceiling effect' (Austin and Brunner 2003).

Slider scales, which are technological advancements not previously available on paper-&-pen survey but now enhanced by surveying with web tools, can overcome the issues of ordinal scales. A slider scale ('slider') is a bar representing a visually continuous segment of numerical points through 1 to m (sometimes through 0 to m, or to -m to m). While the number of points is finite, for any analytical purpose this measurement is considered continuous and not ordinal, therefore m should not be a small number. A very common case is $m = 100$.

The respondent moves an indicator ('it slides') among the values in the bar. If the bar is drawn on a paper, as in the case for Visual Analogue scales (VAS), the respondent can only appoint a mark on the bar. The estimate of VAS may be considered continuous, and more accurate than multipoint scales (Voutilainen *et al.* 2016), but the value would be technically harder to record. For years the absence of proper computing, visualizing, and recording technologies impacted the developments of statistical science. Could multipoint and Likert scales be reputed obsolete because they were designed for *paper-&-pen* data collection? Results from Fryer and Nakao (2020) validate this thesis, while a web experiment by Funke (2015) criticizes sliders. Other results (see, Roster, Lucianetti, and Albaum 2015; Bosch *et al.* 2018) bring further arguments on the evaluation of sliders, in particular reporting a longer time of completion of tasks. A comprehensive review of the debate is provided by Chyung *et al.* (2018).

Matejka *et al.* (2016) performed an experiment testing the accuracy of sliders compared to a Likert scale and on the impact of marks with percentages ('ticks') on the bar of sliders. Participants

(n = 2000) were recruited through *Amazon's service Mechanical Turk*. Participants were asked to estimate the blackness of a shade of grey through sliders or Likerts. Results show that sliders without ticks have better performances in both accuracy of the judgements and bias reduction. Even if authors do not mention it directly bias observed in their results is coherent with the psychological phenomenon of *heaping*, a connection rarely mentioned (an exception: Couper *et al.* 2006).

To monitor heaping effects is important because, while in scales with ticks heaping is due to psychological attachment, there is evidence that heaping is also related to fabricated data in data collections (Finn and Ranchos 2015).

2. Experimental protocol

The sample of respondents is recruited through a web open procedure, like the aforementioned Mechanical Turk. The survey tool is therefore a website. The data collection process is segmented in 3 phases. After completion of 1st phase, a new record is added to a connected database while 2nd and 3rd phases add more data to the record.

In the 1st phase participants are randomly assigned to two random treatment groups. Both the groups are assigned to a task or 'trial': they have to estimate the colour of a square. This trial is repeated for 10 times. The treatment difference among the two groups is that the control group has to estimate the colour through a 0-10 multipoint scale, while the experimental group has to estimate it through a 0-100 slider bar.

As showed in Matejka *et al.* (2016), estimation of shades of colours through a sequence of trials is among the best for objective evaluation of measurement tools (i.e., scales). Instead of presenting to respondents 50 fixed shades of grey squares, we propose a random generator of a shades of Red and Blue. A square of Yellow is superimposed with an opacity randomly distributed between 0% and 10%. Therefore, any randomly coloured square is a realization of the combination of: (i) a randomly generated parameter ξ of shade, uniformly distributed between 0% (full Red) and 100% (full Blue) and (ii) a randomly generated parameter ζ of noise, uniformly distributed between 0% and 10%.

In the 1st phase participants are requested to estimate only shade, with opacity being a possible factor of controlled noise. In the original experiment of Matejka *et al.* there was no mechanism to control noise in the estimation process, even if authors accounted that differences in participants' devices should have been factors of noise out of experimental control. Another difference from Matejka *et al.* is that participants should be free to refuse to complete any trial. The default option in a Likert scale, signalled through a button under (not adjacent) the multipoint scale, is 'no answer'. The best equivalent to let "no answer" in a slider would be setting invisible the indicator on the bar before interaction to it, providing a button 'no answer' to remove it again. This does not push a heaping bias inflation towards initial positions of the indicator (Liu and Conrad 2018). In this case, if the respondent avoids interacting with the slider, a 'no answer' is recorded.

The software must record not only the final choice of the participants but also every single interaction with the tool, tracing their decisional process. Continuous sliders are very well suited for this tracing because there is a large support of values to pick on.

When a participant completes 1st phase, data recorded is: (i) random generated shade parameter ξ for the 10 trials; (ii) random generated opacity parameter ζ for the 10 trials; (iii) participant's estimations x for the 10 shades; (iv) time of completion t_x for each of the 10 trials; (v) number of clicks k_x for each of the 10 trials.

In the 2nd phase participants are asked to report their *taste*-response of 10 well-known leisure products through the scale (to rate) of their treatment groups in 1st phase. When the participant completes the 2nd phase, further information can be added to the record: (vi) participant's rating r for each of the 10 products; (vii) time of completion t_r for each of the 10 ratings; (viii) number of clicks k_r for each of the 10 trials. If the rating process is interrupted, no data is added to the record.

In the 3rd phase standard demographic variables are collected from participants, whereas they provide consent.

3. Methods of data analysis

Heaping is a relevant bias in applied statistical studies on scales of measurement. Even if they do not mention it directly, the statistic adopted in Matejka *et al*. (2016) to measure heaping is a normalised score of the mean deviation from the expected difference of observed frequency among adjacent values:

$$\sqrt{\frac{\Sigma\left((n_x - n_{x-1}) - \frac{\Sigma\, n_x - n_{x-1}}{|M|}\right)^2}{|M|}} \qquad (1)$$

where $|M|$ is the cardinality of the support, x is the observed value from the M scale and n is the absolute frequency associated to x[1]. Matejka *et al*. reported a score of heaping ~ 2 (± 0.1 at CI 95%) for sliders, while the introduction of 'ticks' that imitate multipoint scales in the slider significantly increases the heaping bias (Fig 1, see "no ticks"). The relation is not linear to the number of ticks.
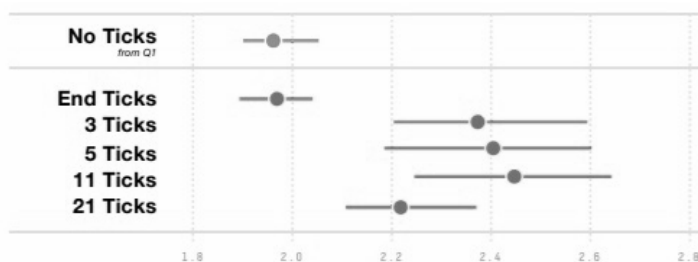
Figure 1. Mean heaping scores for varying number of tick marks.
Error bars show 95% CIs. (Matejka *et al*., p. 5)

We make the hypothesis that control group (multipoint) induces more heaping than experimental group (sliders).

Since values (x for estimates of shades, r for ratings on products) from sliders and multipoint scales are constrained in a finite support, they can be normalised into a [0,1] interval. The distribution of errors $\xi - x$ is the main statistic and is assumed to be normally distributed. A Shapiro-Wilk test is performed on the sample of $\xi - x$ values of all the trials *per* group to confirm this assumption. Since noise factors ζ are all sampled from the same population, we expect no significant difference in the distribution of values. This assumption is tested through a Kolmogorov-Smirnov test. If violated, $\xi - x$ values will be controlled *per* ζ. Times of completion t_x are assumed to be normally distributed. This assumption is tested through a Shapiro-Wilk test.

Null hypotheses on the *objective task* of shade estimation with random noise are:
  i. sliders induce a distribution of mean absolute errors (MAE) from randomised parameters over the 10 trials which is not superior to multipoint scales' MAE.

     Absolute errors $\mid \xi - x \mid$ are never assumed to be distributed normally: if $\xi - x$ values were

---

[1] Is there an implicit consensus of statistical science on this measure? Roberts and Brewer (2001) provide 2 different approaches to measure heaping: (i) H_1 is technically only a minor improvement over (1) while (ii) C_2 is based on the probability to observe local modes. The (ii) approach raises issues on the confidence threshold to assert that an observed local mode is *likely a true* local mode and not a local noise. For a modern approach to heaping models, see Zinn and Würbach (2015)

normally distributed, then their absolute values would be distributed as half-normal distribution (Folded Normal). Given the structure of the hypothesis, a non-parametric 1-tailed test (i.e., Mann-Whitney test) on the samples of participants' MAE in the two groups (a MAE *per* participant) seems suited to check the hypothesis.

ii. sliders induce less variance and not superior skewness than multipoint scales. If $\xi - x$ values of all the trials *per* group are normally distributed, the exact 1-tailed Fisher's test of variance (*F*-test) is suited to check the hypothesis on variance. If the errors are not normally distributed, the non-parametric alternative will be 1-tailed Levene's test. The simpler test to check if treatment variable induces a systemic error in objective estimation is the test of signs of $\xi - x$. A significant difference from null hypothesis of sum of signs equal to 0 for both groups will need to be commented.

iii. sliders induce a not superior t_x than multipoint scales. If time values of all the trials *per* group are normally distributed, a 1-tailed *z*-test of means will check the hypothesis. If time values are not normally distributed the non-parametric alternative is 1-tailed Mann-Whitney test.

Correlations between degrees of controlled noise ζ, errors $\xi - x$, times of completion t_x, and clicks k_x are graphically represented through scatterplots and visualised through a generalised model if the fit is sufficiently good. The effect of noise on $\xi - x$ is supposed to be non-linear and possibly not even symmetrical around the value of $\xi - x = 0$, although it can be symmetrical around a different value. Noise can similarly affect t_x and k_x, too.

Does the same structure of hypotheses A, B, and C hold for measures collected in 2nd phase? Since the 10 leisure products have to be chosen among well-known, a prior value ρ of expected taste can be elicited through an expected value computed from rating statistics of online rating platforms. Although arguably biased for both small and large samples (Askalidis, Kim, and Malthouse 2017), these priors are likely the most reliable predictors of expected *taste* at least from a population of subjects very interested in the product category[2].

Even accounting for aforementioned biases, the statistic $r - \rho$ can be interpreted as a *deviation* of biased raters *vs.* randomised raters. Even if $|r - \rho|$ and $|\xi - x|$ are technically the same operation of *distance*, their arguments are conceptually distinct, as reflected through the order of minuends and in the semantic difference between an *error* (there is always a true parameter ξ) and a *deviation* (two procedures to evaluate the same *evaluando*). As a consequence, the hypotheses on $r - \rho$ cannot be 1-tailed. However, although *tastes* are not *objective*, hypotheses on the differences in values, variances, and skewness among groups can still be asserted.

Moreover, means of $r - \rho$ values can be both correlated and compared to paired (intra-participant) means of $\xi - x$ values (controlled on ζ). Correlating and comparing times of completions (t_x with t_r) and clicks (k_x with k_r) is even less ambiguous since they measure both the same physical quantities. Differences and ratios between the two phases can be compared *per* group, too.

Finally, whereas the sample sizes on demographics collected in 3rd phase support it, associations between demographic variables to aforementioned statistics can be asserted as a control procedure but no causal explanation emerges from literature about trials on the colour perception.

4. Conclusions

While this protocol partly replicates the experiment of Matejka *et al*. (2016), we propose some relevant improvements to define a general experimental protocol for data collection and analysis on web-tool of human perception and tastes:

---

[2] For example, the rating platform *Letterboxd* reports that the movie *The Godfather* (directed by Francis F. Coppola, released in 1972) received more than 300,000 ratings from all over-the-world raters. According to Lorenz (2006) even in presence of local peaks, the best models to represent movies have only one location parameter, which the author interprets as an "evidence of universality in processes of continuous opinion dynamics about taste" (p. 251).

- we generalise the structure of hypotheses that tests the statistical efficiency of the measurement tool through web trials. While hypotheses A (location) and C (duration) were already well-covered in literature, hypothesis B (variance) is often neglected. The definition of statistical assumptions makes explicit some elements of potential fragility of previous literature on the topic of evaluation of measurement tools for social sciences, i.e., to our knowledge no research on sliders mentioned the potential need of non-parametric tests for variance of errors or deviations.
- the previous issue is likely the consequence of a general under-recognition of research of heaping bias. Matejka *et al.* (2016) did not acknowledge literature on heaping. We connected their empirical work to the at-state-of-art mathematical alternatives for measurement of heaping bias. We also re-wrote (1) in a less ambiguous and friendlier formalism for statisticians and psychometricians.
- we see improvements in the experimental procedures, since we introduced a noise parameter ζ that affects the coloured square inducing visual opacity. This inclusion reproduces better extra-experimental situations of perception.
- the inclusion of a data collection on tastes in the 2nd phase provides not only a better assessment on scales' performance but it could also highlight insights on the relationships between *perception* and *taste*. Of course, we assumed that an experiment focuses on a particular taste for *something* (e.g., movies are convenient) but further experiments could pair perceptions and ratings on different objects (arts, languages, etc.).

The major rationale to adopt sliders has sprouted from the theoretical debates mentioned in Section 1, so far. For applied research, even in absence of evidence of remarkable improvements (see, hypotheses A, B, and C in Section 3) in the reduction of coarseness in data, inaccuracies of self-report, and biases through adoption of sliders, the evidence that sliders reduce scale-induced heaping (Figure 1) is extremely insightful. Better measurement scales can minimize the confounding effect in those research programmes aimed to investigate data fabrication (i.e., fraud reports) through tests on heaping.

References

Agresti A. (2010). *Analysis of Ordinal Categorical Data*, Wiley, Hoboken, (NJ).

Askalidis, G., Kim, S.J., Malthouse, E.C. (2017). Understanding and overcoming biases in online review systems. *Decision Support Systems*, **97**, pp. 23-30.

Austin, P.C., Brunner, L.J. (2003). Type I error inflation in the presence of a ceiling effect. *The American Statistician*, **57**(2), pp. 97-104.

Bosch, O.J., Revilla, M., DeCastellarnau, A., Weber, W. (2018). Measurement reliability, validity, and quality of slider versus radio button scales in an online probability-based panel in Norway. *Social Science Computer Review*, doi:10.1177/0894439317750089.

Chyung, S.Y.Y., Swanson, I., Roberts, K., Hankinson A. (2018). Evidence-based survey design: The use of continuous rating scales in surveys, *Performance Improvement*, **57**(5), 38-48.

Couper, M.P., Tourangeau, R., Conrad, F.G., Singer, E. (2006). Evaluating the effectiveness of visual analog scales. *Social Science Computer Review*, **24**(2), pp. 227-245.

Finn A., Ranchhod V. (2015). Genuine fakes: The prevalence and implications of data fabrication in a large South African survey. *The World Bank Economic Review*. doi:10.1093/wber/lhv054.

Fryer, L.K., Nakao, K. (2020). The future of survey self-report: An experiment contrasting Likert, VAS, slide, and swipe touch interfaces. *Frontline Learning Research*, **8**(3), pp. 10-25.

Funke, F. (2015) A web experiment showing negative effects of slider scales compared to visual analogue scales and radio button scales, *Social Science Computer Review,* **34**(2), pp. 244-254.

Kampen, J., Swyngedouw, M. (2000). The ordinal controversy revisited. *Quality & Quantity,* **34,** pp. 87-102.

Kluver, D., Ekstrand, M. D., Konstan, J. A. (2018). Rating-based collaborative filtering: algorithms and evaluation. In *Social Information Access,* eds. P. Brusilovsky and D. He, Springer, Charm, (SW), pp. 344-390.

Liu M., Conrad, F.G. (2018). Where should I start? On default values for slider questions in web surveys, *Social Science Computer Review*. doi:10.1177/0894439318755336.

Lorenz, J. (2006). Universality in movie rating distributions. *The European Physical Journal B.* **71**, pp. 251-258.

Matejka, J., Glueck, M., Grossman, T., Fitzmaurice, G. (2016). The effect of visual appearance on the performance of continuous sliders and visual analogue scales. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. doi: 10.1145/2858036.2858063.

Roberts, J.M., Brewer, D.D. (2001). Measures and tests of heaping in discrete quantitative distributions. *Journal of Applied Statistics*, **28**(7), pp. 887-896.

Roster, C.A., Lucianetti L., Albaum, G. (2015). Exploring slider vs. categorical response formats in web-based surveys, *Journal of Research Practice*, **11**(1), Article D1. Retrieved from http://jrp.icaap.org/index.php/jrp/article/view/509/413.

Tomaselli, V., Cantone, G.G. (2020). Evaluating Rank-Coherence of Crowd Rating in Customer Satisfaction. *Social Indicators Research*. doi: 10.1007/s11205-020-02581-8.

Velleman, P.F., Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician*, **47**(1), pp. 65-72.

Voutilainen, A., Pitkäaho, T., Kvist, T., Vehviläinen-Julkunen, K. (2016). How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *Journal of Advanced Nursing*, **72**(4), pp. 946-957.

Zinn, S., Würbach, A. (2015). A statistical approach to address the problem of heaping in self-reported income data. *Journal of Applied Statistics*, **43**(4), pp. 682-703.

Life satisfaction of refugees living in Germany

Daria Mendola, Anna Maria Parroco

1. Lives of refugees in high-income countries

Since 2015, Germany has become home to significant numbers of refugees and asylum seekers which has led to it topping the ranking of European destination countries. German is now the fifth in the world for the number of accommodated refugees (UNHCR, 2021). The large influx of refugees during these last few years put great strain on German receiving system that struggled with offering full services to newly arrived refugees and asylum seekers (Hinger, 2016).

Despite the fact that the quality of life of refugees is expected to have been improving in the aftermath of their arrival to Germany, refugees and asylum seekers must still face several problems of integration and economic deprivation as well as concerns and worries for their lives (e.g., about 90% are unemployed and nearly 54% are worried that they will be unable to stay in Germany- own elaborations on data from the 2016 IAB-BAMF-SOEP[1] Survey of Refugees in Germany).

Whereas academic research is traditionally devoted to examining the objective pillars of the integration of immigrants and refugees (their educational accomplishments, language skills, or labour market positioning), immigrants' subjective evaluation of their life situation -and subjective well-being more in general- has only started to draw scholarly attention in recent years (Colic-Peisker, 2009; Kogan *et al*., 2018; Schiele, 2020). Nowadays, life satisfaction (LS) of refugees is still an under-explored theme. Amongst the main predictors of refugees' well-being, we find mental and general health, family ties, and housing conditions, all widely reported in the literature (Phillips, 2006; Belau, 2019; Gambaro *et al*., 2018; Walther *et al*., 2020).

Issues of mental health (such as depression, anxiety, or post-traumatic distress) are reported in a recent and increasing strand of literature for refugees hosted also in highly developed countries (see, e.g., the Leiler *et al*. (2019) study on Sweden; Walther *et al*. (2020) and Georgiadou *et al*. (2018) in Germany).

Family ties strongly influence subjective well-being, especially in the case of refugees, whose family members often remain in their homeland or have died due to conflicts or during the migration (Gambaro *et al*., 2018; Busetta and Mendola, 2018).

Despite the clearly improved objective living conditions of the migrants, whether migration has significant and long-lasting effects on life satisfaction of those who have moved, is still debated in the scientific literature. Indeed, Hendriks (2015) underlined contradicting results in his review of cross-sectional studies on immigrants, that compared subjective well-being of "movers" to that of "stayers".

The aim of this paper is to contribute to the ongoing literature on the quality of life of refugees in host nations. Using the first wave of the IAB-BAMF-SOEP survey of refugees (carried out in 2016), we estimate ordinal regression models for LS levels and offer some preliminary statistical investigations into life satisfaction and its components in the context of refugees who arrived in Germany between 2013 and 2016.

---

[1] The Institute for Employment Research (IAB), the Socio-Economic Panel (SOEP) at the German Institute for Economic Research (DIW Berlin), and the Research Centre on Migration, Integration, and Asylum of the Federal Office of Migration and Refugees (BAMF-FZ). See Brücker *et al*. (2016).

The following Section 2 presents in details data and methods; Section 3 proposes our statistical analyses, and Section 4 concludes discussing the main results from this study.

2. Data and methods

Data are from the IAB-BAMF-SOEP Survey of Refugees in Germany that is a survey of people who entered Germany between 2013 and 2016 and applied for asylum, whatever the result of the application. It includes information on individual socio-demographic characteristics and household level information. The survey is longitudinal and provides yearly interviews of household members aged 18 and over. In this study, we rely on the first wave of the survey (2016).

Using a sample of 3,408 individuals, we present some preliminary analyses on the life satisfaction of these vulnerable individuals. Life satisfaction is understood to be a subjective aspect of the quality of life (see Cummins, 2000); the main variable consists of people's self-assessment of their overall life satisfaction ("How satisfied are you currently with your life in general?" arranged on an 11-point scale). LS answers show the usual negatively skewed distribution with a generally high mean ($Q_1 = 6$, $Q_2 = 8$, $Q_3 = 9$, mean = 7.28, standard deviation = 2.31, skewness = -0.88). Given this, we arranged LS levels by quartile (slightly rounding the cut-points in order to guarantee about 25% of observations for each interval) and an ordinal regression model was estimated to focus on the association among levels of LS and main individual and household level characteristics.

Analyses include *sociodemographic control variables*: such as sex of the respondents, their education level (arranged in three ordinal levels, according to ISCED standards), nationality proxied by the country of origin (Syria, Afghanistan, Iraq, former USSR; Africa; Balkan region, other countries), and age (including a quadratic effect). Then a set of *post-migration personal factors* are considered: time in Germany (as the number of years passed between arrival in Germany and the time of the interview); legal residence permit (dummy variable in which we combined refugees, entitled to asylum and holder of subsidiary/humanitarian and other forms of international protection into one category, and placing into the other one those awaiting the response to asylum application ad those whose application was dismissed), concerns about their own economic situation (a lot, somewhat, not at all). *Post-migration family related factors* include family arrangements (household size, presence of a partner/spouse possibly cohabitant); kind of accommodation (shared with others or private).

In the end, we also considered a selected set of life domains for which satisfaction evaluations were available: satisfaction with current living arrangements, with the quality of the food, with the privacy that they have, with the safety of their neighbourhood and with their own current health. These were assumed to be post-migration subjective well-being factors.

3. Results

Descriptives

Our sample is made up of 3,408 adults, with a prevalence of men (62%), a mean age of 33.5 years, with four nationalities (Afghan, Eritrean, Iraqi, and Syrian) accounting for about 83% of the sample. Among them, 85.67% do not have any form of international protection, in part because their application was dismissed and partly because they still have a pending request; the others being granted by some form of international protection like refugee status (73.66%), international protection or status of tolerance.

Satisfaction with life was generally rated lower by men (average score is 7.15; IC$_{95\%}$: 7.05-7.25) than by women (7.50; IC$_{95\%}$: 7.38-7.61), by people without or with a pending legal status (7.01; IC$_{95\%}$: 6.88-7.15) than by refugees and holders of international protection (7.45; IC$_{95\%}$: 7.36-7.55).

Figure 1 shows a comparison among nationalities on the average LS score, along with 95% confidence intervals. Former USSR countries have the highest LS mean (7.95); neatly higher than the average LS of Africans, Iraqis, and Syrians.

Figure 1: *Mean life satisfaction score (and 95% confidence intervals) by main nationalities*
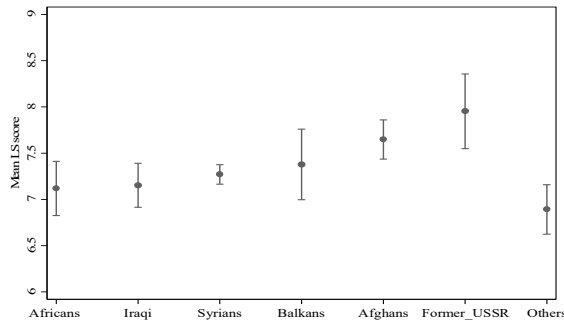
Multivariate analysis

Ordinal regression model was estimated in order to provide possible explanations of different levels of LS through the sets of covariates presented above.[2] Table 1 displays the ordinal regression model estimates for LS scores arranged in quartiles.

Socio-demographic factors: while there is not any difference between men and women on LS, age has an effect which is slightly non-linear: youngsters show higher values of LS than elderly. Education shapes life satisfaction too: highly educated respondents are less satisfied than those with low levels of education, other things being equal; instead, respondents with low and medium level of education experience the same LS.

The country of origin is significantly associated with life satisfaction: when compared to Syrians, Afghans people as well as Balkans, Iraqis, those from former USSR, or from "other countries" have a higher level of life satisfaction. No statistically significant differences emerge between Syrians and people coming from African countries.

Post-migration personal factors: As expected, even controlling for main socio-demographic characteristics, respondents' LS is higher among those who obtained any kind of legal protection than among those who had not (yet) received their residence permit.

LS is negatively associated with the extent of financial concerns. Particularly, people partially concerned or not concerned at all with financial issues show higher level of LS.

Post-migration family related factors: the two covariates accounting for family arrangements are associated significantly with LS. Indeed, according to international studies (see, e.g., Busetta and Mendola, 2019), higher household size and having a cohabiting partner/spouse -which are both proxies of social support and, more in general, of social capital- increase refugees' LS. Particularly, not having a partner or living separated from him/her (that is not in the same house nor in the same

---

[2] An ordinal logit model with parallel lines assumption was first estimated. The violation of this assumption, that was assessed via a Brant test (Brant 88.31, df=48, p=0.000), is due to the coefficients for higher education (odds ratios: 0.858; 0.727***; 0.578***), being with a legal residence permit (1.458***; 1.255**; 1.010), and not concerned at all about ones' own economic situation (1.922***; 1.350***; 1.105). Since the other estimates were almost identical, we decided to present in table the PL model.

city) lowers the life satisfaction, even controlling for other personal and family characteristics. Unexpectedly, respondents who live in private houses have a lower level of satisfaction than those who live in shared ones, other things being equal. These last results could be related to the feeling of loneliness even if this hypothesis would need a further in-depth analysis.

Post-migration subjective well-being: As accounted for in many studies, also perceived well-being measures related to specific life domains are highly significantly associated with overall life satisfaction. Increasing levels of satisfaction with health, living arrangements, feeling safe with neighbourhood, and privacy in the current living arrangements, positively affect LS.

Table 1: *Ordinal regression model for Life Satisfaction quartiles (odds ratio estimates)*

Covariates	Odds ratio	Sign.	Covariates	Odds ratio	Sign.
SOCIO-DEMOGRAPHIC FACTORS			**POST-MIGRATION FAMILY FACTORS**		
Gender (ref. male)			**Family arrangements**		
Female	1.097		Household size	1.076	***
Age	0.953	**	Partner or Spouse (ref. none)		
Age squared	1.001	**	cohab. partner/spouse	1.407	***
Nation group (ref. Syrians)			not cohab. partner/spouse	0.969	
Afghans	1.752	***	**Accommodation** (ref. shared acc.)		
Africans	1.237		private apartment	0.757	***
Balkans	1.603	***			
Former USSR	1.617	**	**POST-MIGRATION SUBJECTIVE WELL-BEING**		
Iraqis	1.211	*	Sat_Health	1.173	***
Other nationalities	1.236	*	Sat_Living arrangements	1.251	***
Education (ref. low)			Sat_Safety neighbourhood	1.063	***
middle school	0.900		Sat_Quality of Food	1.081	***
high school or more	0.694	***	Sat_Privacy	1.055	***
POST-MIGRATION PERSONAL FACTORS					
With a legal residence permit	1.251	***	**Cut-points**	16.18	***
Years in Germany	1.080			70.81	***
Economic concerns (ref. a lot)				210.80	***
somewhat concerned	1.428	***			
not concerned at all	2.384	***			
Aic	8229.2		N = 3,408		
Bic	8394.8		*** p<0.001; ** p<0.05; * p<0.10		

4. Discussion and conclusions

When they arrived in highly developed host nations, refugees face new challenges for their integration and successful settlement, and often experience material deprivation, isolation, uncertainty, and bad quality of life. However, life satisfaction of refugees in the post-migration phase, in high-developed hosting countries, is an under-investigated theme.

Using the results from the first wave of the German survey of refugees, we provide preliminary analyses of the determinant of their life satisfaction.

Our estimates pointed out how lower life satisfaction levels are associated with the condition of being older, Syrian, alone or with few family members, highly educated, without a partner or a spouse or without a cohabiting ones, and without a legal permit to stay in Germany.

Furthermore, our analyses highlight the fact that those factors addressing a greater stability in people lives (e.g., the status of refugee or the international protection, as well as living as a couple and without financial concerns) appear to be correlated with greater life satisfaction (Nesterko *et*

al., 2012; Colic-Peisker, 2009). Hence, to foster social integration and increase LS of refugees and asylum seekers, it stands out as crucial to shorten the process for the issue of the status of refugees or of the international and humanitarian permits (which are also related to the possibility of family reunification) and foster opportunities for economic independence (pre-requisite for the formation of new family unions).

As expected, LS is positively associated with satisfaction with some specific life domains, which hence play an important role in shaping the overall life satisfaction (Amint, 2010). Not trivially, being satisfied with these specific life domains (such as safety, privacy, food), related to the new life conditions in Germany, tell us about the process of acculturation (Berry, 2017) which involves changes in social structures and institutions and in people's behaviours, towards an integration pathway that accounts for cultural traits of both the origin and host country. It is indisputable that satisfied immigrants have a much better integration in society and can give a greater contribution to its development. Thus, understanding and fostering life satisfaction is widely seen as a central goal.

Among the limitations of this contribution, we acknowledge the lack of a deeper analysis of the migratory history. Indeed, since immigrants, and refugees in particular, are a heterogeneous group with a great variety of immigration-related experiences, their past experiences can affect current evaluation of life satisfaction both in terms of *inertia* of negative feelings accumulated during the travel phase of their migration, and in terms of resilience.

Moreover, the cultural dimensions of the acculturation process, mentioned above and herein accounted for by means of subjective well-being proxies, could be better argued including some other or additional descriptors of the quality of life in Germany.

From a methodological point of view, given the practice to assimilate 11-point scale variables to numerical ones, models for skewed variables (e.g., using a Gamma link) could be tested for a better prediction of LS scores, allowing for parsimony in the number of estimated coefficients.

References

Amint, K., (2010). Determinants of Life Satisfaction Among Immigrants from Western Countries and from the FSU in Israel, *Social Indicators Research,* **96**, pp. 515–534.

Belau, M. (2019). The impact of the housing situation on the health-related quality of life of refugees located in North Rhine-Westfalia, Germany. *European Journal of Public Health*, **29**(Supplement_4), ckz185-175.

Berry, J.W. (2017). Theories and Models of Acculturation, The Oxford Handbook of Acculturation and Health, Edited by Seth J. Schwartz and Jennifer Unger. DOI:10.1093/oxfordhb/9780190215217.013.2

Brücker, H., Rother, N., Schupp, J. (2016). *IAB-BAMF-SOEP-Befragung von Geflüchteten: Überblick und erste Ergebnisse* (Vol. 29, p. 77). DEU.

Busetta, A., Mendola, D. (2018). The effect of family networks on refugees health conditions. In: *Giornate di Studio sulla Popolazione 2019,* January 24-26, 2019 – Milan, Italy.

Colic-Peisker, V. (2009). Visibility, settlement success and life satisfaction in three refugee communities in Australia. *Ethnicities*, **9**(2), pp. 175-199.

Cummins, R.A. (2000). Objective and Subjective Quality of Life: An Interactive Model. *Social Indicators Research,* **52**, pp. 55–72.

Gambaro, L., Kreyenfeld, M., Schacht, D., Spieß, C. K. (2018). Refugees in Germany with children still living abroad have lowest life satisfaction. *DIW Weekly Report*, **8**(42), pp. 415-425.

Georgiadou, E., Zbidat, A., Schmitt, G. M., Erim, Y. (2018). Prevalence of mental distress among Syrian refugees with residence permission in Germany: a registry-based study. *Frontiers in psychiatry*, **9**, 393, pp. 1-12.

Hendriks, M. (2015). The happiness of international migrants: A review of research findings. *Migration Studies*, **3**(3), pp. 343–369.

Hinger, S. (2016). Asylum in Germany: The making of the 'crisis' and the role of civil society. *Human Geography*, **9**(2), pp. 78-88.

Kogan, I., Shen, J., Siegert, M. (2018). What makes a satisfied immigrant? Host-country characteristics and immigrants' life satisfaction in eighteen European countries. *Journal of Happiness Studies*, **19**(6), pp. 1783-1809.

Leiler, A., Bjärtå, A., Ekdahl, J., Wasteson, E. (2019). Mental health and quality of life among asylum seekers and refugees living in refugee housing facilities in Sweden. *Social psychiatry and psychiatric epidemiology*, **54**(5), pp. 543-551.

Nesterko, Y., Braehler, E., Grande, G., Glaesmer, H. (2013). Life satisfaction and health-related quality of life in immigrants and native-born Germans: the role of immigration-related factors. *Quality of Life Research*, **22**(5), pp. 1005-1013.

Phillips, D. (2006). Moving towards integration: the housing of asylum seekers and refugees in Britain. *Housing Studies*, **21**(4), pp. 539-553.

Walther, L., Fuchs, L. M., Schupp, J., von Scheve, C. (2020). Living Conditions and the Mental Health and Well-being of Refugees: Evidence from a Large-Scale German Survey. *Journal of Immigrant and Minority Health*, pp. 1-11.

Schiele, M. (2020). Life satisfaction and return migration: analysing the role of life satisfaction for migrant return intentions in Germany. *Journal of Ethnic and Migration Studies*, pp. 1-20.

UNHCR (2021). Refugee data finder. Available at: https://www.unhcr.org/refugee-statistics (Accessed 5th January 2021).

A quantitative study to measure the family impact of e-learning

Cristina Davino, Marco Gherghi, Domenico Vistocco

1. Introduction

The Covid-19 emergency has forced universities around the world to transfer teaching activities online. Even if online teaching allowed to carry out the planned teaching activities, it is necessary, in retrospect, to evaluate the impact of this teaching method on the different types of students, in terms of preparation, characteristics and social background. The switch from offline to online learning caused by Covid-19 is expected to exacerbate existing educational inequalities penalising more vulnerable students. The social and economic conditions of families have a major influence on the e-learning experience because less advantaged students are less likely to have access to relevant learning digital resources (e.g. laptop/computer, broadband internet connection) and less likely to have a suitable home learning environment (e.g. a quiet place to study or their own desk) (Di Pietro et al., 2020). Furthermore, according to the 2020 European Commission's annual report on the levels of digitalisation achieved by the various member states[1], Italy ranks 25th among the 28 EU Member States[2].

The aim of this paper is to analyse whether and how the distance learning activities impacted on the students' families both in terms of the organisation of spaces and daily rhythms and from an economic point of view, having required additional expenses. The study is based on the analysis of data collected at the University of Naples Federico II in June 2020. More than 19,000 students took part in a survey, carried out to monitor distance learning activities and perceptions. The paper is organised into two sections. In the first, a factorial method is exploited to obtain a composite indicator measuring the family impact of distance learning. Then, we try to explain if the family impact takes different forms and intensity depending on the students' characteristics, the availability of computer equipment and the type of teaching used. Finally, quantile regression allow to differentiate the study of effects for different levels of family impact. Some considerations on the distance learning experience in terms of family impact and the evaluation on the preferred teaching method for the future are also enclosed.

2. Measuring family impact of E-learning

The measurement of family impact is carried out following the classic steps used in research methodology for the measurement of a multidimensional and abstract concept (Freudenberg, 2003), i.e. a latent variable not directly observable and expressed as a combination of several components. The construction of such a latent variable, often referred to as Composite Indicators (CI), is done through the use of an aggregation method appropriate to the nature of the observed variables (Lebart et al., 2000).

The study proposed in this paper is based the survey conducted by the University of Naples considering only students who attended at least one distance learning course in the 2019/2020

---

[1] $https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi$
[2] The data in the report refer to 2019 and therefore do not take into account all the initiatives taken by governments to counter pandemic.

Cristina Davino, University of Naples Federico II, Italy, cristina.davino@unina.it, 0000-0003-1154-4209
Marco Gherghi, University of Naples Federico II, Italy, marco.gherghi@unina.it, 0000-0003-0363-145X
Domenico Vistocco, University of Naples Federico II, Italy, domenico.vistocco@unina.it, 0000-0002-8541-6755

academic year. The sample of responses received reflects the distribution of the student population by degree course. A special section of the questionnaire was dedicated to the detection of the family impact of the e-learning experience. The following is a list of questions relating to this section with an indication of the percentage of answers for each category (labels that will be used in the tables and graphs in italics, percentages in parentheses):

- *Place* - When attending a distance learning class, you were mainly: *anywhere* (12.9), *alone* (78.6), *with_relatives*(8.3);
- *Expenses* - Did you and/or your family incur any expenses in order to follow the distance learning lessons?: *equipment* (17.1), *equipment&network* (3.2), *network* (11.1), *no* (67.0), *other* (1.6);
- *use_equipment* - To follow lessons at a distance, the device you mainly used was: *exclusive* (67.5), *shared_other* (10.5), *shared_teaching_working* (22);
- *family_habits* - Distance learning courses affected your normal habits and those of the rest of the family.: *strongly_agree* (19.3), *agree* (26.3), *neither_agree_nor_disagree* (29), *disagree* (14.7), *strongly_disagree* (10.7).

Since the indicators are all qualitative and/or ordinal, a multiple correspondence analysis (MCA) was used to provide a CI measuring the family impact of e-learning. MCA can be considered one of the best known and most effective tools for the simultaneous analysis of questionnaire data. Proposed in the late 1970s by J.P. Benzècri for the case of two qualitative variables (Binary Correspondence Analysis), it has been extended to the case of many qualitative variables. MCA is a Factor Analysis that allows to identify a reduced number of variables (also called factors or latent variables or CIs) as a linear combination of the original variables. Each CI is able to explain a part of the variability of the phenomenon.

The first factor obtained from the MCA accounts for 91.34% of the total variability. It can therefore be considered an adequate measure of the Family Impact of the experience of E-Learning (from now on FIEL). The distribution of FIEL (Figure 1, left-hand side) shows a phenomenon almost equally distributed around the average value (represented by the value 35.58[3]) even if with different characteristics in the two parts of the distribution: students with a low family impact are more concentrated, while the right tail of the distribution is more dispersed. This is a signal of greater heterogeneity among those who have a family impact above the average.

The interpretation of the FIEL indicator can be deepened by considering also the contributions of the categories on the first factor, not focusing only the coordinate represented by the indicator itself. The contribution of a category to the explanation of a factor is provided by the product of the weight of the category, represented by its frequency, and the square of the coordinate of the category on the factor. Indeed, in MCA categories with vey low or high coordinates do not necessarily contribute to the explanation of the factor itself, because if they had a very low frequency, they would have a very low contribution. Similarly, categories that are more "central", but with a very high frequency, may have an important contribution to the explanation of the factor. The joint visualisation of coordinates and contributes (Figure 1, right-hand side) highlights that students who predominantly experienced a quiet e-learning experience without changing family habits (they already had all the equipment available for their exclusive use) are separate from students who were forced to share both the workstation and the device with family members engaged in smart working or other learning activities. This second group of students is forced to study in makeshift places, sometimes with other family members and distance learning has also affected their families financially.

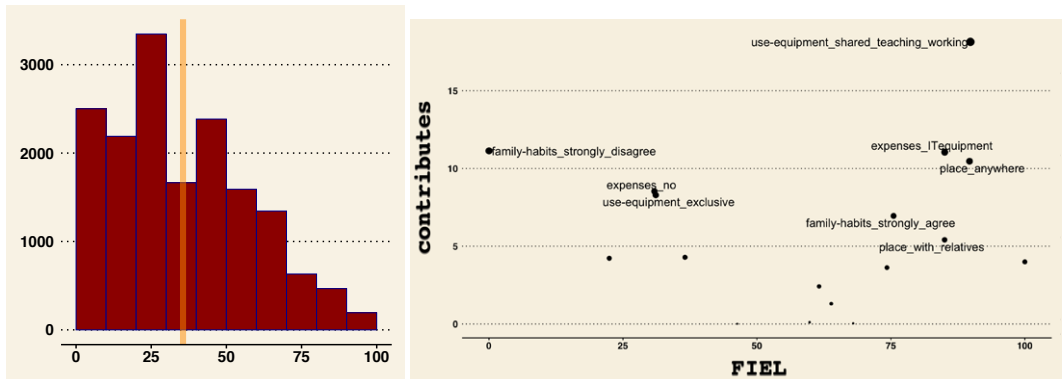[3]The indicator has been rescaled in the range 0-100.

Figure 1: Distribution of the family impact of the experience of e-learning (left-hand side) and scatter plot of the categories measuring the FIEL according to the MCA coordinates and contributes (right-hand side)

3. Explaining family impact of E-learning

The interpretation of the FIEL indicator can be deepened by considering additional variables that did not contribute to its determination and that concern both personal characteristics of students, issues related more specifically to the availability of computer and network equipment (IT equipment) and also to the modality of distance learning. The former features are represented in the upper panel of Figure 2 while the latter in the bottom panel. Each point is located at the average values of FIEL in the correspondent category, the size being proportional to the frequency[4]. The vertical line represents the FIEL average. The family impact seems stronger (higher than the general average) for female students in the first years of the university experience. As might be expected, a wi–fi connection and a mobile study station (linked to the use of smartphones and tablets) can explain more complicated family situations.
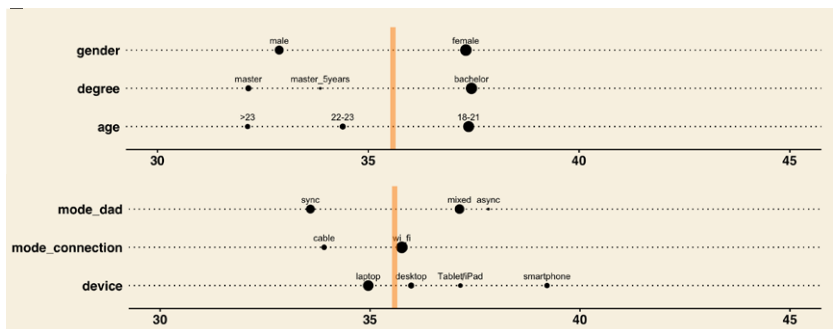
Figure 2: FIEL averages according to socio-demographic features and type of IT equipment

A comparison among the average values of FIEL does not allow to capture possible differences in the impact of the considered variables for different levels of family difficulties. Quantile regression (Koenker and Basset, 1978) allows us to complement the results of a classical OLS regression by exploring the effects of the regressors on the entire distribution of FIEL. In fact, although the number of quantiles that can be explored is theoretically infinite, it is shown that a sufficiently dense grid can be enough to reconstruct the entire dependent variable (Davino et

---

[4]For each variable considered, the averages are significantly different.

al., 2013). Nevertheless, in many cases we explore a small number of quantiles that represent parts of the distribution important for the particular analysis. In Figure 3, QR coefficients equal or greater than the conditional median are graphically represented for the different considered regressors. The horizontal axis displays the different quantiles, while the effect of each feature holding the others constant is represented on the vertical axis. The horizontal solid lines show the OLS results while the piecewise lines refer to the coefficients at different quantiles. The aim is to graphically catch the coefficient trends moving from lower to upper quantiles.
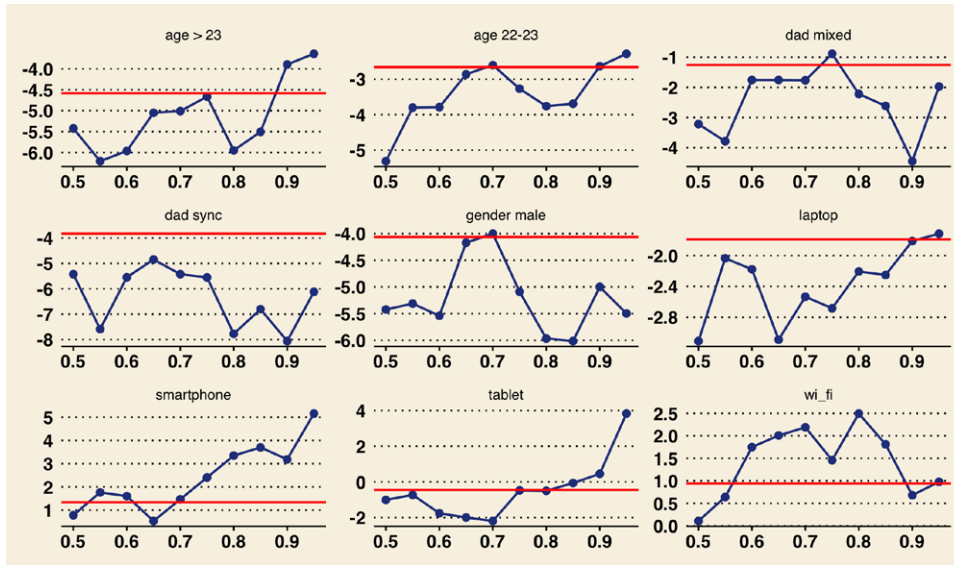

Figure 3: OLS (horizontal solid lines) and QR (piecewise lines) coefficients

Coefficients have been estimated for a sequence of quantiles from 0.5 to 0.9 with a step of 0.5. It was decided to explore the results of only the top 50% of the distribution as the aim is to investigate the situations of discomfort in order to understand what levers can be used to intervene. By the way, in the remaining part of the distribution (students with FIEL below the median) the effects of the regressors considered are practically null. A positive trend of the quantile curves emerges from the plot. This correspond to an increased variability of the FIEL variable (i.e. an increased difference between for instance the 25% and 75% conditional quantiles) with increasing values of the regressor or when the category changes. The interpretation of the results must take into account, apart from possible fluctuations in the values of the coefficients at the different quantiles, the sign of these coefficients and the possible presence of patterns from the lowest to the highest quantiles. For example, the negative effect of age on FIEL is less amplified in cases where the family impact is very high. The increasing trend suggests that this effect is gradually disappearing. More interesting is the interpretation of the results concerning the device used for distance learning (the reference category for the regression is desktop). In particular, the use of a smartphone compared to a fixed location has a consistently positive and increasingly strong effect moving towards the top of the distribution. As regards the use of Tablet/Ipad, the sign is even reversed starting from quantile 0.85. In addition to the above information, it should be noted that all the coefficients are always significant, with the exception of tablets and mixes, which are never significant, and wi–fi and smartphones, which contribute significant coefficients at the top of the distribution, at quantile 0.65 and 1uantile 0.85 respectively.

The results shown in this paper, although in many cases expected, allow to quantify and

visualise relationships among different elements that can contribute to highlight heterogeneity in the conditions and characteristics of students, an element that, in non-emergency conditions, is ignored when the same teaching strategies are adopted for all the students. Moreover, a complete understanding of a phenomenon cannot be achieved without measuring it. In this sense the results here illustrated can provide a quantitative measure of a multidimensional and abstract concept, the family impact of e-learning. The use of quantile regression allows to explore if student characteristics or IT equipment have different effect among those who have suffered a stronger family impact.

Looking to the future, students' preference for the different teaching modes changes according to the family impact of the experience. The boxplots in Figure 4 show the distribution of FIEL in the group of those who would prefer lessons exclusively at a distance (online), who believe that they can still benefit from an appropriate combination of the two modes (mixed) or who would prefer a total return to normality (onsite). There is an increase in the FIEL quartiles from the online category to the mixed and then onsite category, a sign that lived experience influences, hopefully only in part, the vision of the future.
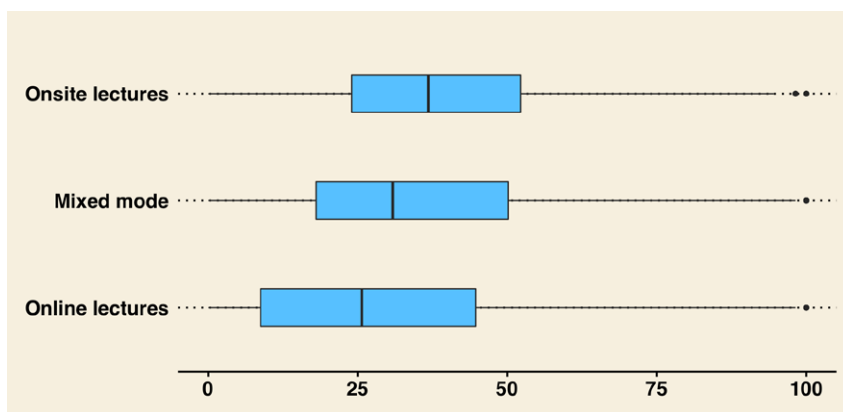
Figure 4: FIEL distribution according to the future vision of the students

References

Davino, C., Furno, M., Vistocco, D. (2013). *Quantile Regression: Theory and Applications.* Wiley, Chichester, UK.

Di Pietro, G., Biagi, F., Costa, P., Karpiski Z., Mazza, J. (2020). *The likely impact of COVID-19 on education: Reflections based on the existing literature and international datasets.* EUR 30275 EN, Publications Office of the European Union, Luxembourg.

Freudenberg, M., (2003). *Composite Indicators of Country Performance: A Critical Assessment.* OECD Science, Technology and Industry Working Papers 2003/16, OECD Publishing.

Koenker R., Basset G.W. (1978). Regression Quantiles. *Econometrica*, **46**, pp. 33–50.

Lebart, L., Morineau, A., Piron, M. (2000). *Statistique Exploratoire Multidimensionnelle.* Dunod, Malakoff Cedex, FR, pp. 181–184.

Thematic atlas of Italian oncological research: the analysis of public IRCCS

Corrado Cuccurullo, Luca D'Aniello, Maria Spano

1. Introduction

This paper has been developed in the frame of the research project "V:ALERE 2019" focused on Italian public-owned Academic Medical Centers (AMCs - that is 16 public AMCs as "Aziende Ospedaliere Universitarie", 9 public AMCs as "Ex Policlinici Universitari a gestione diretta", 21 public-owned "Istituti di Ricovero e Cura a Carattere Scientifico" (IRCCS) (Ministry of Health - http://www.salute.gov.it/, 2018)). These institutions have a triple mission: research, teaching, and care, having an enormous impact on society and the nation's health.

The main aim of the project is to provide new evidences and proposals to support and advise Italian public AMCs in their quest to address their challenges.

In recent years, there is increasing recognition of the potential value of research evidence as one of the many factors considered by policymakers and practitioners. Even more, in the case of medical science, the analysis of research and its impact is indispensable, in light of its implications for public health.

The starting point for mapping a research area is to review the related scientific literature by synthesizing past research findings and then, effectively use the existing knowledge base and advanced lines of future researches. In this sense, bibliometrics becomes useful, by introducing a systematic, transparent, and reproducible review process based on the statistical measurement of science, scientists, or scientific activity (Cuccurullo *et al.*, 2016). Many research areas use bibliometric methods to explore the impact of their field, the impact of a set of researchers, the impact of a particular paper, journals taken as a reference by researchers, the input knowledge, research gaps, trends, and future opportunities (Zaho, 2010). Performance analysis and science mapping (Noyons *et al.*, 1999) are the two main bibliometric approaches for investigating a research area.

In this work, we focus on science mapping as it allows identifying and displaying themes and trends with a synchronic (Callon *et al*., 1983) or a diachronic perspective (Cobo *et al*., 2011). By means of science mapping techniques, namely the term co-occurrence networks, and strategic/thematic maps, we aim at providing a data visualization of strategic positioning of the different Italian public AMCs in terms of their research positioning.

In particular, we identify the research-front of different AMCs and then, we visualize them in a joint representation, useful for comparing their main research themes and at the same time their different specializations, by considering also their evolution during the years.

Mapping the dynamic positioning of Italian medical research at various levels (i.e. national, regional, AMCs type, AMC) will provide a conceptual framework for policymakers and managers to understand and manage the problems of the AMCs (e.g. appropriate funding mechanisms for financing the triple-mission). Moreover, this tool could be useful for the institutions themselves to direct their research efforts towards increasingly innovative fronts taking into account the general landscape and at the same time exploiting this information to establish collaborations with other AMCs dealing with the same research topics.

Here, the effectiveness of our strategy is showed by considering the scientific production of the last 20 years of IRCCSs specialized in the oncology research.

Corrado Cuccurullo, University of Campania Luigi Vanvitelli, Italy, corrado.cuccurullo@unicampania.it, 0000-0002-7401-8575
Luca D'Aniello, University of Campania Luigi Vanvitelli, Italy, lucadaniello94@gmail.com, 0000-0003-1019-9212
Maria Spano, University of Naples Federico II, Italy, maria.spano@unina.it, 0000-0002-3103-2342

2. Data and methodology

IRCCSs are Italian healthcare organizations of relevant national interest that drive clinical assistance in strong relation to research activities. Their mission is the continuous upgrade of healthcare. The IRCCS title is granted by the Italian Ministry of Health to a very limited number of institutes throughout the nation, and their activities are federally regulated by Legislative Decree 288/2003. They are committed to being a benchmark for the whole public health system for both the quality of patient care and the innovation skills in the field of the organization. The activity of IRCCSs relates to well-defined research areas whether they received recognition for a single subject (monothematic IRCCS) or for multiple integrated biomedical areas (polythematic IRCCS).

Among the 21 public IRCCSs in Italy, we considered the nine institutions specialized in the oncology research area (6 monothematic and 3 polythematic IRCCSs).

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) was used for the selection process of the publications (Liberati *et al.*, 2009). We retrieved on Web of Science (WoS) indexing database – launched by the Institute for Scientific Information (ISI) and now maintained by Clarivate Analytics – all the publications from January 2000 to December 2019. To identify the publications related to each IRCCS, we searched by full name, part of the organization name's or by its commonly known abbreviation from the Organizations – Enhanced List available on WoS (e.g. "IRCCS FND MILANO" for the Fondazione IRCCS Istituto Nazionale Tumori Milano). We limit our search by document type and selected only Articles, Proceedings Papers, Review Articles, and Book Chapters in the English language. The records were exported into PlainText format.

Starting from our final collection, we loaded the data and converted it into R data frame using bibliometrix, an open-source tool for quantitative research in scientometrics and bibliometrics that includes all the main methods for performance analysis and science mapping (Aria and Cuccurullo, 2017).

In this preprocessing phase, for the polythematic IRCCSs (*Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Istituto Nazionale Tumori Regina Elena (IRE), IRCCS Ospedale Policlinico San Martino*) we considered only the publications dealing with oncological topics, by filtering the records with respect to the metadata "Research Areas" (SC) included in WoS.

In order to consider the publications that have a major impact in the field of oncological research, we calculated the normalized citation score (NCS), one of the most frequently used field-normalized indicators (Bornmann and Haunschild, 2016). It has been calculated by dividing the citation count of a focal paper by the average citation count of the papers published in the same field (and publication year). The normalization procedure is based on all articles published within one year (and must be repeated for publications from other years).

The citation count of the article is divided by the average number of citations in the field of the article, yielding the normalized citation score for the paper. The overall normalized citation impact of each IRCCS can be analyzed on the basis of the mean value over the publication set. This results in the mean NCS (MNCS) for the paper set. In the end, following the percentile approach, we performed our analysis only on the publications with an MNCS greater than 75% (the top 25% publications).

To map the conceptual structure of each IRCCS we conducted two related analyses: a term co-occurrence network analysis and a strategic or thematic map. The combined use of these techniques allows us to illustrate: how terms relate to each other, the main research themes within each institution, and how they develop.

The basic idea behind the term co-occurrence network analysis (Wang *et al*., 2018) is that each research field or topic can be represented as a set of terms (e.g. keywords, terms extracted from titles, or abstracts). Network representation is used to understand the themes

covered by a research field, to define which are the most important and the most recent ones; i.e., the research front. Following the network approach, we built a term co-occurrence matrix, in which each cell outside the principal diagonal contains the number of times two terms appear together in the articles (co-occur). Then, the co-occurrences among terms were normalized by the association index as proposed by Van Eck and Waltman (2009). This measure assumes values in the interval [0,1] and reflects the strength of the association among terms. Co-occurrence matrices can be seen as undirected weighted graphs; therefore, we can build a network in which each term is a node and the association between linked terms is expressed as an edge, visualizing both single terms and subsets of terms frequently co-occurring together. To detect subgroups of strongly linked terms, where each subgroup corresponds to a center of interest or to a theme of the analyzed collection, we refer to community detection algorithms (Fortunato, 2010). Here, to this end, we carried out a community detection procedure by using Louvain algorithm (Blondel *et al*., 2008).

Strategic or Thematic map (Cobo *et al*., 2011) allows plotting the themes, identified through community detection, in a bi-dimensional matrix where axes are functions of the Callon centrality and density, respectively (Callon *et al.*, 1983). Centrality can be read as the importance of the theme in the research field; while density can be read as a measure of the theme's development.

In this way, we identified the conceptual structure of each IRCCS in the three different considered time slices. Then, we standardized centrality and density values, in order to make a comparison among the research fronts of the different institutions by plotting themes in a joint map. As in classical analysis, the obtained strategic map allows defining four typologies of themes (Cahlik, 2000) according to the quadrant in which they are placed. Themes in the upper-right quadrant are known as the motor themes. They are characterized by both high centrality and density. This means that they are both developed and important for the research field. Themes in the upper-left quadrant are known as isolated themes or niche themes. They have well developed internal links (high density) but unimportant external links and so are of only limited importance for the field (low centrality). Themes in the lower-left quadrant are known as emerging or declining themes. They have both low centrality and density meaning that are weakly developed or marginal. Themes in the lower-right quadrant are known as basic and transversal themes. They are characterized by high centrality and low density. These themes are important for a research field and concern general topics transversal to the different research areas of the field. In each temporal interval, we considered the KeyWords Plus (ID) used in the different documents. The ID are words or phrases that frequently appear in the titles of an article's references but do not appear in the title of the publication itself.

Their generation is based upon a special algorithm (Garfield, 1990) that is unique to *Clarivate Analytics* databases.

3. Main results

To highlight the main research themes of oncological IRCCSs and evaluating their evolution over time, we decided to divide our timespan (2000–2019) into three-time slices.

In Table 1 the distribution of the selected publications per IRCCS in the three different periods is reported. The scientific production of institutions has increased over time. The production is constant in the three-time slices for two IRCCSs (i.e. *IRCCS Ospedale San Martino* and *Istituto Nazionale Tumori Regina Elena (IRE) IRCCS*). However, some IRCCSs produced a great number of publications in the third period with respect to the previous ones (e.g. *Istituto Tumori Bari "Giovanni Paolo II" IRCCS* and *IRCCS Centro di Riferimento Oncologico della Basilicata (CROB))*.
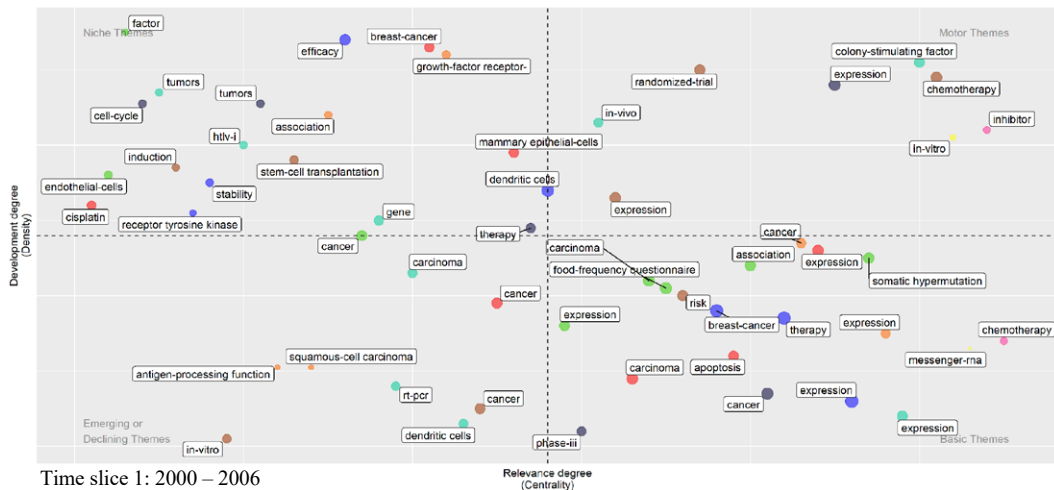
Table 1 Publications distribution per IRCCS in the three different time slices

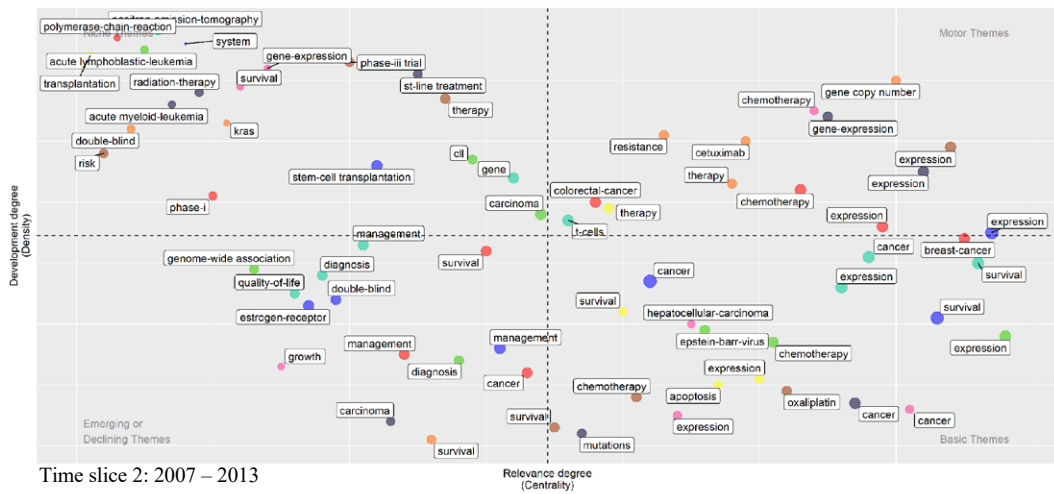| Period	2000 – 2006		2007 – 2013		2014 – 2019	
Organizations	*No. Of Doc*	*% of Doc*	*No. Of Doc*	*% of Doc*	*No. Of Doc*	*% of Doc*
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico	48	18.60	73	28.29	137	53.10
Centro di Riferimento Oncologico (CRO AVIANO)	175	28.83	186	30.64	246	40.53
Fondazione IRCCS Istituto Nazionale dei Tumori	466	22.40	753	36.20	861	41.39
IRCCS Ospedale Policlinico San Martino	147	35.25	135	32.37	135	32.37
Istituto Oncologico Veneto (IOV) IRCCS	97	13.04	338	45.43	309	41.53
Istituto Tumori Bari "Giovanni Paolo II" IRCCS	16	6.53	59	24.08	170	69.39
Istituto Nazionale Tumori IRCCS – Fondazione Pascale	147	18.85	265	33.97	368	47.18
Istituto Nazionale Tumori Regina Elena (IRE) IRCCS	121	31.35	140	36.27	125	32.38
IRCSS Centro di Riferimento Oncologico della Basilicata (CROB)	11	6.51	65	38.46	93	55.03

In Figure 1 the thematic Atlas of IRCCSs' oncological research is shown. It is worth noting that each theme, identified with the community detection, is labelled with the corresponding most frequent ID.

In the three-time slices, the production of IRCCSs is rich but they have three main themes in common: *expression*, *survival,* and *chemiotherapy*. In the first time slice (2000 – 2006) *expression* was a basic theme for many IRCCSs and only for *IRE RO* was a motor theme. The position of this theme changes over the years. In the second time slice (2007 – 2013) *expression* becomes a motor theme - high density and high centrality – for many IRCCSs and starting to shift from the upper-right quadrant to the lower-right quadrant in the third slice (2014 – 2019), consolidating its role as traditional theme - low density and high centrality. Since 2007 studies focus on s*urvival* that appeared as an emerging theme on the lower-left quadrant - low density and low centrality. In the third period, *survival* becomes a traditional theme, indicating great interest in the health care of patients by many IRCCSs.

Chemiotherapy is also a theme treated by many IRCCSs over time, always positioned to the right of the map - high centrality - in the three-time slice. From the second to the third period the *chemiotherapy* theme shift from the upper-right quadrant to the lower-right quadrant, becoming a basic theme. On the upper-left quadrant, we have observed that niche themes - low centrality and high density - have increased over time. This means that the oncological research of IRCCSs is oriented towards studies more and more specialized from 2000 to 2019.
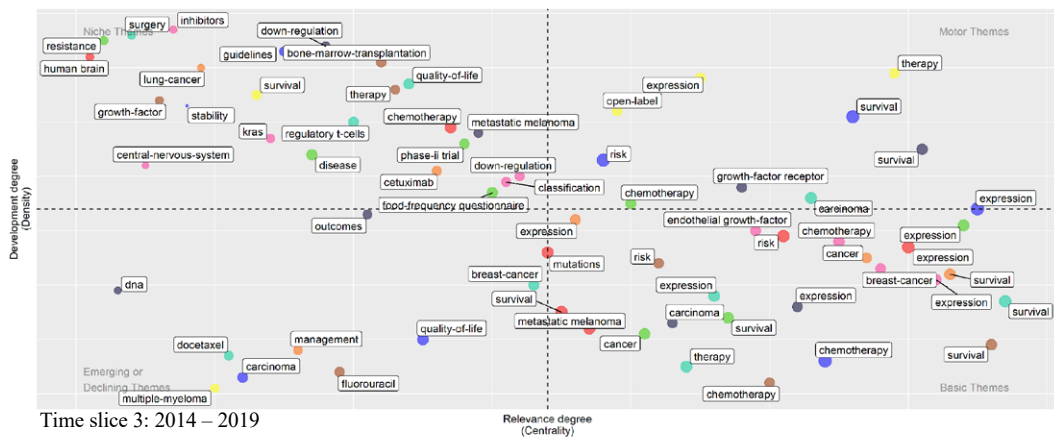
Fig. 1. Thematic Atlas of IRCCSs' oncological research

4. Conclusion and future developments

In this paper, we propose to jointly represent the dynamic research positioning of the different Italian public IRCCSs specialized in Oncology. These graphical representations summarize many aspects of the cancer research landscape in Italy. Obviously, the presented results are only a small part of what could be observed starting from the thematic maps.

Therefore, they are powerful decision support tools for the different agents involved in the health system. However, it is important to highlight that this approach could be used for different purposes in a more general bibliometric framework (e.g. comparison of topics covered by different sources, by different countries, or as in this case by different institutions).

On the one hand, future developments will be devoted to extending our analysis to the other Italian AMCs in order to completely mapping their research positioning; on the other hand working on the graphical representations to improve the readability of the results.

References

Aria, M., Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, **11**(4), pp. 959-975.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008. URL: http://stacks.iop.org/1742-5468/2008/i=10/a=P10008.

Bornmann, L., Haunschild, R. (2016). Citation score normalized by cited references (CSNCR): The introduction of a new citation impact indicator. *Journal of Informetrics*, **10**(3), pp. 875-887.

Cahlik, T. (2000). Comparison of the maps of science. *Scientometrics*, **49**(3), pp. 373-387.

Callon, M., Courtial, J. P., Turner, W. A., Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, **22** (2), pp. 191-235.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, **5**(1), pp. 146–166.

Cuccurullo, C., Aria, M., Sarto, F. (2016). Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics*, **108**(2), pp. 595-611.

Eck, N. J. V., Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American society for information science and technology*, **60**(8), pp. 1635-1651.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, **486**(3), pp. 75-174.

Garfield, E. (1990). Keywords Plus®: ISI's breakthrough retrieval method. Part 1. Expanding your searching power on Current Contents on Diskette, *Current Contents*, **32**, pp. 5-9.

Liberati A., Altman D. G., Tetzlaff J., Mulrow C., Gøtzsche P. C., Ioannidis J. P. A., Clarke M., Devereaux P. J., Kleijnen J., Moher D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Journal of clinical epidemiology*, **62**(10), pp. e1-e34.

Noyons, E., Moed, H., van Raan, A. (1999). Integrating research performance analysis and science mapping. *Scientometrics*, **46**(3), pp. 591-604.

Wang, H., Zhao, Y., Dang, B., Han, P., Shi, X. (2019). "Network centrality and innovation performance: the role of formal and informal institutions in emerging economies", *Journal of Business & Industrial Marketing*, **34**(6), pp. 1388-1400.

Zhao, D. (2010). Characteristics and impact of grant-funded research: a case study of the library and information science field. *Scientometrics*, **84**(2), pp. 293-306. DOI= 10.1007/s11192-010-0191-y.

Frameworks and inequalities in healthcare: some applications

Pietro Renzi, Alberto Franci

1. Introduction

There is increasing recognition of the importance of social determinants of health (SDOH), which encompass social, behavioural, and environmental influences on one's health. Indeed, SDOH have taken centre stage in many recent health policy discussions; particularly those relating to the Covid-19 pandemic, accountable care organizations, and other initiatives focusing on improving population health (Townsend et al.,1982). Furthermore, existing literature (Vian, 1982) and current research (Marmot et al., 2020) clearly suggest that a focus on SDOH can enable improvements in the health of populations. Therefore, giving greater attention to SDOH may help both improve Italians' health and reduce health care costs.

This paper:

1.   Identifies and investigates the principal conceptual frameworks for action relating to SDOH;

2.   Analyses possible relationships between SDOH and health outcomes (life expectancy, mortality rates, morbidity rates etc.) using the Quadrant Analysis technique; and

3.   Contributes to the ongoing debate about practicable measures which could be used to alert regions to inequalities in health and healthcare.

2. Methodology, data, interpretation and use

Quadrant charts were used to plot SDOH against other indicators of interest on health outcomes (life expectancy, mortality rates, morbidity rates, quality of care, access and physical resources, etc.). These showed percentage differences from the Italian averages for each indicator; with the intersection of the axes representing the Italy average for both indicators. Therefore, deviations from the midpoint readily highlight which regions perform above or below the Italy average for both indicators. A simple correlation line was included.

There are many methods to measure health inequalities. Those chosen to quantify the degree of inequality in a specific health variable in this research were the **slope index of inequality** (SII) and **the concentration index** (CI), which the authors consider to be the most relevant and important. According to O'Donnell et al. (2008), the CI is able to measure the association between socio-economic and health inequalities; and it should be noted that the CI directly relates to Concentration Curves (Kakwani et al., 1997). Given there are various methods proposed to calculate the CI, the authors applied that deemed most relevant to their research, i.e., that for grouped data proposed by Brown (Fuller and Lury, 1977):

$$CI = (p_1 L_2 - p_2 L_1) + (p_2 L_3 - p_3 L_2) + \cdots + (p_{T-1} L_T - p_T L_{T-1})$$

Italian data refers to the year 2016 and was sourced from:

•   Health for All and I.Stat from Istituto Nazionale di Statistica (ISTAT);

•   Osservasalute from Osservatorio Nazionale sulla Salute nelle Regioni Italiane dell'Istituto di Sanità Pubblica-Sezione di Igiene dell'Università Cattolica del Sacro Cuore;

Pietro Renzi, University of San Marino, San Marino, pietro@unirsm.sm, 0000-0001-6200-7265
Alberto Franci, University of Urbino Carlo Bo, Italy, alberto.franci@uniurb.it, 0000-0002-8157-9792

- Rapporto sulla situazione sociale del Paese from Centro Studi Investimenti Sociali (CENSIS)
- Passi d'Argento from Istituto Superiore di Sanità.

Each region was colour-coded based on a simple (unweighted) risk factors index, which averaged smoking, alcohol and overweight variables. 'Blue' indicates that a region's performance is close to the Italian average; 'Green' indicates that it is significantly better (with 'low' risk factors); and 'Red' indicates that it is considerably worse (with 'high' risk factors).

3. Results

The first investigation examined the underlying conditions and root causes contributing to health inequities, and the interdependent nature of the factors that create them. After a holistic analysis of the various frameworks available in the literature (Canadian council on social determinants of health, 2015), it was considered that the conceptual model for community-based solutions to promote health equity (fig. 1) was the most appropriate and informative. Unlike a logic model, which is linear and progresses neatly from inputs to outputs and outcomes, the model in figure 1 is circular, thereby reflecting the topic's complexity. Inputs are shown in the outer circle and background, depicting the context of structural inequities, socio-economic and political drivers, and the determinants of health, in which health inequities and community-driven solutions exist.
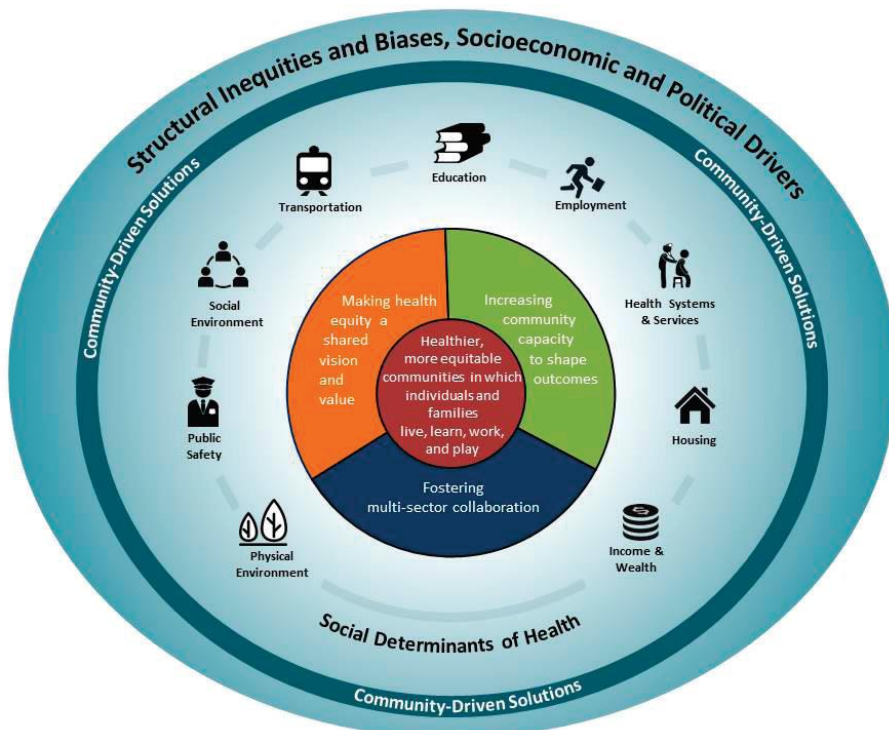
Figure 1 - A conceptual model for community-based solutions to promote health equity.
SOURCES: National Academies of Sciences, Engineering, and Medicine (2017), Communities in Action: Pathways to Health Equity

The quadrant charts were used to measure the extent by which SDOH influence better access, quality and health outcomes (OECD, 2019). These analyses illustrated the relationship between a variable linked to the health and social care system and another variable of interest; the latter included health risks factors, income (or other economic variables) and environmental quality.

The main results are presented in the figure 2 and 3.
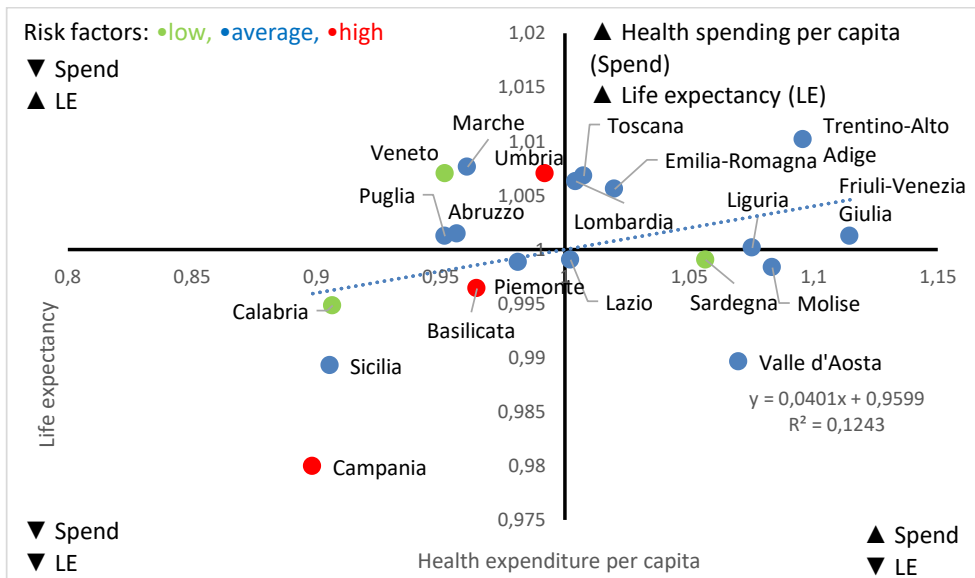
Figure 2 - Life expectancy and health expenditure

Figure 2 illustrates the extent to which regions that spend more on health have better health outcomes (noting such associations do not guarantee a causal relationship). There is a clear positive association between health spending per capita and life expectancy. Amongst the twenty regions, six spend more and also have higher life expectancy than the Italy average (top right quadrant). A further five regions spend less and have lower life expectancy at birth (bottom left quadrant). Of particular interest are regions that deviate from this basic relationship. Five regions spend less than average but achieve higher life expectancy overall (top left quadrant); these are Marche, Umbria, Veneto, Puglia, and Abruzzo. The four regions in the bottom right quadrant present higher spending, but lower life expectancy than the Italy average; these are: Lazio, Sardegna, Molise, and Valle d'Aosta.

It is noticed that two of the three regions with high overall risk factors (red dots) have lower life expectancy than the Italy average; and are also typically below the trend line, which shows the average spending to life expectancy ratio across Italy regions. Further interesting results were obtained using the same quadrant analysis technique applied to different SDOH and health outcomes. The great strength of this diagram is that it enables the simultaneous consideration of the three variables being studied, viz. life expectancy, per capita health expenditure and risk factors. Added value is created by deliberately using colour to reflect whether a region's figures were above or below the range of $M \pm \sigma$. Furthermore, the diagram serves to highlight that outcomes (such as life expectancy, mortality, morbidity, infant mortality etc.) can be influenced by variables that are outside the National Health System, i.e., it is not certain that increasing health expenditure per capita will necessarily enable improvements in health outcomes. This conclusion is consistent not only with the oldest literature (Bruno et al., 1978; Vian, 1982; Fabbris, 1990; Biggeri and Grisotto, 2005) but also with the most recent literature (Marmot et al., 2020). The results presented in this quadrant analysis invite the reader and the regional health authorities to broaden their approach to improving health and look beyond the National Health System itself by investing in the social determinants of health. Particular attention should be given to the key risk factors, viz. smoking, alcohol consumption and overweight. The need for a wider approach is reinforced by the statement that non-medical factors play a substantially larger role than do medical factors in the maintenance of health, with medical factors only weighted from 10%-20% (Remington et al., 2015; National

Academies of Sciences, Engineering, and Medicine, 2017).

The SII and CI methods were applied to an outpatient department in the Marches region. The aim was to analyse inequalities among women, classified according to their level of education, with regards to their degree of access to qualified gynaecological staff. The results are presented in Tables 1 & 2 and Figures 3 & 4 below:

Level of education	f women	f_r women	Cum. f_r	f of visits	f_r
Primary school	594	0,12	0,12	253	12,75
Secondary school	861	0,17	0,29	271	13,65
High school	1704	0,34	0,63	526	26,50
Bachelor's degree	1212	0,24	0,87	459	23,12
Master's degree	646	0,13	1,00	476	23,98
Total	**5017**	**1**		**1985**	**100**

Table1–SII. Classification of the women population by level of education and by number of obstetric visits received by a gynaecologist in an outpatient department in the Marches region
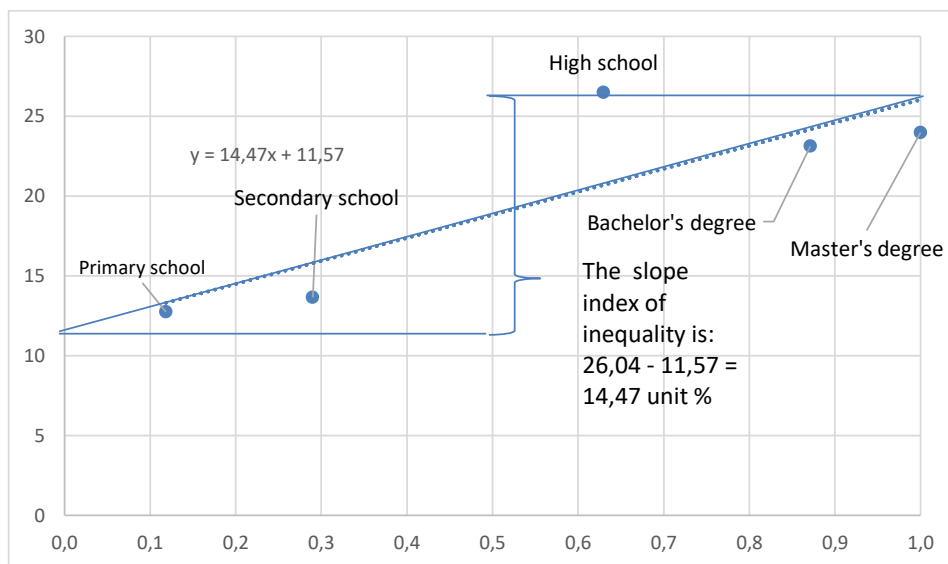
Figure 3–SII representing inequalities in obstetric needs compared to the level of education

These results show that women who have a higher level of education have a 14.47% better chance of receiving obstetric care from qualified personnel than those who have a lower level of education.

Level of education	f women	f_r women	Cum. f_r	f of visits	f_r of visits	Cum. f_r	CI
Primary school	594	0,12	0,12	253	0,13	0,13	-0,006
Secondary school	861	0,17	0,29	271	0,14	0,27	-0,013
High school	1704	0,34	0,63	526	0,26	0,53	0,018
Bachelor's degree	1212	0,24	0,87	459	0,23	0,76	0,111
Master's degree	646	0,13	1,00	476	0,24	1,00	0,000
Total	**5017**	**1**		**1985**	**1**		**0,110**

Table2 – CI using Brown formula. Classification of the women population by level of education and by number of obstetric visits received by a gynaecologist in an outpatient department in the Marches region
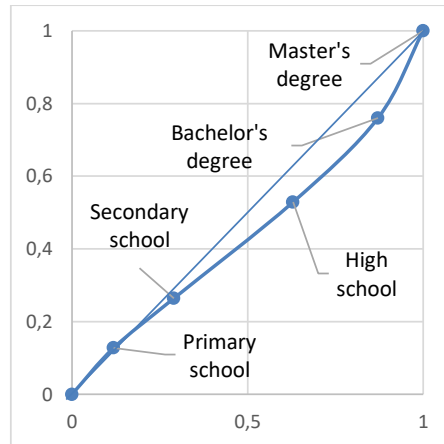
Figure 4 - Concentration curve representing inequalities in obstetric needs compared to the level of education

The obtained value of the CI is 0.11. Its positive value indicates the existence of a weak inequality that is favourable to the more educated female population. This is inferred from the fact that the 63% of the female population with lower levels of education (i.e., up to high school) accounted for only 53% of the obstetric visits. In summary, females with a higher level of education have a greater chance of obtaining obstetric visits from a qualified gynaecological staff.

This conclusion is consistent with the idea that higher levels of scholarship enable people to better understand health literacy, and innovations in medical and food hygiene fields. Also, more educated people are arguably more able to deal with disadvantageous situations. In synthesis, better education can facilitate better health (Feinstein et al., 2006; Zajacova et al., 2018).

4. Conclusion

Pandemics are arguably more of a social problem than a healthcare problem. A population that lives in poverty and in neighbourhoods that are overcrowded, with poor maintenance and sanitation, is being disproportionately affected by COVID-19. This serves to highlight the importance and weight assumed by SDOH in the health of populations.

To this end, following a brief review of the available main frameworks, our research identified three types of variables that are identified in any health system, namely: the final variables (outcomes), the instrumental variables (linked to the healthcare sector) and the current variables (linked to the characteristics of socio-economic systems).

The Quadrant Analysis showed some relationships between a final variable (life expectancy) and an instrumental variable (per capita health expenditure). The existing low correlation between these two variables was already known in the literature, but the simultaneous visualization of risk factors in the quadrants suggests a need to look more widely than the health system alone and develop/invest in socio-health policies that address the SDOH.

The work identified some important measures of inequality in healthcare through the use of SII and CI (the last calculated according to Brown's formula). The applications involved social and health facilities in an *area vasta* of the Marche Region, and highlighted how the use of an obstetric outpatient department by the female population varied according to women's level of education (which is a key SDOH).

A further application of the CI (using Erreygers formula) examined the evolution of inequalities in the Marche region, and suggested their weakening over the years investigated. The application of the methodology on the *area vasta* of the Marche region has limits linked to the specificity of the demographic and social context; yet its transferability to other contexts is straightforward. Therefore, it is considered that this research has made a methodological contribution to the visualization of SDOH and to the measurement of inequalities.

References

Bambra, C., Riordan, R., Ford, J., Matthews, F., (2020), The COVID-19 pandemic and health inequalities, *J Epidemiol Community Health,***74**, pp. 964-968

Biggeri, A., Grisotto, L., (2009), Fonti di distorsione nelle misure delle disuguaglianze di salute, in *Disuguaglianze sociali di salute. Problemi di definizione e di misura.*, eds. Costa, G., Cislaghi, C., Caranci, N., Salute e Società, Franco Angeli Editore, Milano, (I), pp. 78-89

Bruno, S., Sgritta, G. B., Trivellato, U., Zuliani, A., (1978), Stratificazione economico-sociale e distorsioni distributive nei servizi pubblici, in *Atti della XXIX° riunione scientifica della Società Italiana di Statistica, Bologna 20-22/03/1978,* CLEUP, Padova, (I), **I/1** pp. 251-295

Canadian council on social determinants of health (2015), *A Review of Frameworks on the Determinants of Health*, http://ccsdh.ca/images/uploads/Frameworks_Report_English.pdf

Fabbris, L., (1990), *Analisi esplorativa dei dati multidimensionali*, CLEUP, Padova, (I)

Feinstein, L., Sabates, R., Anderson, T. M., Sorhaindo, A., Hammond, C., (2006), What are the effects of education on health? in *Proceedings of the Copenhagen Symposium Measuring the Effects of Education on Health and Civic Engagement,* eds. Desjardins, R., Schuller T., OECD/CERI, Copenhagen (DK), pp. 171-354

Fuller, M. F., Lury, D. A., (1977), *Statistics Workbook for Social Science Students*, Philip Allan Publishers Ltd, London, (UK)

Kakwani, N., Wagstaff, A., van Doorslaer, E., (1997), Socioeconomic inequalities in health: measurement, computation, and statistical inference. *J. Econom.,* **77**(1), pp. 87-103

Marmot, M., Allen, J., Boyce, T., Goldblatt, G., Morrison, J., (2020), *Health equity in England: The Marmot review 10 years on*, The Health Foundation, London, (UK)

National Academies of Sciences, Engineering, and Medicine, (2017), *Communities in Action: Pathways to Health Equity*. The National Academies Press, Washington DC (USA)

O'Donnell, O., van Doorslaer, E., Wagstaff, A., Lindelow, M., (2008), *Analyzing Health Equity Using Household Survey Data: A Guide to Techniques and Their Implementation*, World Bank, Washington, DC (USA)

OECD, (2019), *Health at a Glance 2019*: OECD Indicators, OECD Publishing, Paris, (F)

Remington, P. L., Catlin, B. B., Gennuso, K. P., (2015), The County Health Rankings: rationale and methods, *Population Health Metrics,* **13**

Townsend, P., Davidson, N., (1982), *Inequalities in Health: The Black Report*, Penguin, London (UK)

Vian, F., (1982), Analisi statistica degli squilibri territoriali in *Le statistiche dello sviluppo. Atti del Convegno (Salerno, 15-16-17 ottobre 1980)*, Edizioni Scientifiche Italiane, Napoli, (I), pp. 463-477

Zajacova, A., Lawrence, E. M., (2018), The Relationship Between Education and Health: Reducing Disparities Through a Contextual Approach, *Annual Review of Public Health,* **39**(1), pp. 273-289

An analysis of the transaction towards sustainable food consumption practises during the Italian lockdown for SARS-CoV-2: the experience of the Lombardy region

Marco D'Addario, Massimo Labra, Silvia Mari, Raffaele Matacena, Mariangela Zenga

1. Introduction

In the pandemia for SARS-CoV-2, food occupies a central position (Galimberti et al., 2020). During the lockdown period, the attention has been devoted to the activities and behaviors related to nutrition, considering also the acts of purchasing, cooking and consuming food. In a context in which working-day out-of-home and school meals were no longer available, people forcely prepared and consumed their meals at home.

This work is part of an ongoing research that analyzes the effects of the pandemic on the healthiness and sustainability of food-related behaviors. It does so by means of an empirical investigation carried out in Lombardy region, the region severely hit by the coronavirus pandemic in Italy. Within this frame, the specific objective of this work is to assess whether behavioral and attitudinal patterns related to consuming food have changed with respect to the established habits of 'ordinary' periods, and how these transformations are linked to socio-demograhic information of respondents.

2. The survey and the sample

An online survey was administered in May-June 2020, employing the Computer Assisted Web Interview (CAWI) methodology. The survey was designed to link data about socio-demographics and living conditions, with self-reported changes in practices related to food consumption, cooking and food shopping. Moreover, data about the psychological condition during the lockdown, weight management, physical activity and health status, and food- and sustainability-related opinions, attitudes and future intentions were recorded.

Of the 2288 complete responses recorded, we consider only $n = 1540$ respondents living in Lombardy that was the region most affected by the SARS-CoV-2 during the period February-May 2020. As shown in Table 1, 51.6% of respndents were provided by participants who identify themselves as females. The average age was 48.79 years (sd=17.43). The level of education of the sample is imbalanced towards higher educational attainments: 63.8% of respondents hold a graduate or a post-graduate degree. The sample was characterized by higher-than-average levels of socio-economic well-being (measured by MacArthur Scale of Subjective Social Status (Adler et al. (2000)) with a mean value of 6.24 (sd=1.33). Most respondents (51.7%) had a normal weight, while the 45.1% was overweighted or obese. Moreover, 80.1% of respondents declared to follow an omnivore diet.

A large proportion of respondents (34.4%) declared a worsening effect of the SARS-CoV-2 emergency on their economic conditions. From the point of view of work, 53.8% of the sample reported having worked from home, while 9.4% declared not having worked at all in the period

and 7.6% of the sample were employed in essential sectors. Most respondents (39.2%) lived as a couple, 18.9% of cases consisted of three people, and 24.5% of cases comprised four or more people. The individuals who lived the period alone were 15.5%.

	n	%		n	%
Gender			*Liv. cond. in 1.st lockdown*		
Male	746	48.4	Single	238	15.5
Female	794	51.6	Couple	604	39.2
			3 persons	291	18.9
Level of education			4 or more	377	24.5
Up to secondary school	85	5.5			
High school	473	30.7	*Work cond. in 1.st lockdown*		
Graduation or more	983	63.8	Working at home	828	53.8
			Essential sector	117	7.6
BMI			Not working	145	9.4
Underweight	49	3.2	Other	450	29.2
Normal weight	796	51.7			
Overweight	420	27.2	*Ec. cond. after 1.st lockdown*		
Obese	275	17.9	Much worse	118	7.7
			A little worse	412	26.7
Usual Dietary Regime			No influence	904	58.7
Omnivore	1234	80.1	A little better	104	6.7
No red meat	181	11.7	Much better	3	0.2
Pescatarian, vegetarian, vegan	125	8.1			

Table 1: The sample (n=1540)

3. The transition towards sustainable foods consumption practises

With the aim to analyze the transition towards sustainable foods consumption practises, we considered the multiple choice (single answer) item: "In comparison to your 'ordinary' life habits, how often have you consumed the following dishes and foods during the lockdown? Answer: Never as before, less frequently, as usual, more frequently". For the purpose of this paper, the four categories on food consumption were collapsed in 3 categories as follow: Less than before; Never or equal than before; More than before. Now, the second category is thought to underline behaviors that have not changed since before the lockdown. Table 2 reports the consumption's food habits during the first Italian lockdown. A closer look at the results reveals how certain food groups have been favored over others in the timeframe investigated. Among these, sweets and desserts, vegetables, carb dishes and fresh fruit recorded the highest percentages of consumption increase, since they were eaten more frequently than usual by, respectively, 43.3%, 35.8%, 27.5% and 26.5% of the sample. Other foods that were privileged by lockdown eaters belong to the categories of legumes (21.1%) and dairy (20.8%). Interestingly, meat does not seem to have played a leading role within lockdown diets. Despite the overarching tendency pointing towards increased variety and quantity of food consumption, the proportion of respondents who in fact reduced the consumption frequency of meat-based dishes (15.9%) is slightly higher than that of those who consumed meat more frequently (12.3%). A trend of reduction is also highlighted in the cases of sugary beverages (sodas and juices) and alcoholic drinks, most likely linked to the supervened impossibility to experience social gatherings and/or celebration moments. Nevertheless, it is important to notice that 19.1% of the sample - a significant propor-

tion - increased their alcohol consumption while under lockdown. In sum, the lockdown seems to have had a double effect on diets: on the one hand, it spurred the consumption of ingredients that are typical of the Mediterranean diet (vegetables, legumes, fruit) and also deeply associated with traditional patterns of cooking and eating in Italy; on the other, it underscored the 'comforting' effect of certain foods, which brought many people to indulge, in our case, on pasta, sweets and dairy, perhaps as an attempt to cope with boredom and/or other negative subjective consequences of social confinement.

	Less than before		Equal than before		More than before	
Food	**n**	**%**	**n**	**%**	**n**	**%**
Carb-based dishes	141	9.1	975	63.3	424	27.5
Meat-based dishes	245	15.9	1106	71.8	189	12.3
Dairy products	151	9.8	1069	69.4	321	20.8
Sweets and desserts	220	14.3	654	38.2	666	43.3
Alcholic beverage	285	18.5	961	62.4	294	19.1
Sugary beverage	175	11.3	1251	81.2	115	7.5
Vegetables-based dishes	120	7.8	868	56.4	552	35.8
Legumes	162	10.5	1054	68.4	325	21.1
Whole-grain cereals	142	9.2	1220	79.2	179	11.6
Nuts and oily seeds	188	12.2	1165	75.7	187	12.2
Fresh fruit	122	7.9	1010	65.6	408	26.5

Table 2: The consumer's food habits during the first Italian lockdown.

Since the nature of the scale of the previous items, we applied the categorical principal component analysis (CatPCA, Linting & van der Kooij (2012)) to examine the component structure of the latent construct. Following the EATLancet Commission's dietary recommendations (Willet et al. (2019)), we considered two groups of foods: the sustainable and healthy foods (vegetables-based dishes, legumes, whole-grain cereals, nuts and oily seed and fresh fruit) and, on the contrary, unsustainable and unhealty foods (carb-based dishes, meat-based dishes, dairy products, sweet and desserts, alchoolic beverage, sugary beverage). This choice was confirmed by the application of a preliminary CatPCA on the eleven items. The application of the CatPCA separately on the two groups allowed us to obtain an index of transition for the sustainable foods' consumption (TSF) and an index of transition for the unsustainable foods' consumption (TUF).

4. The transition for the sustainable foods' consumption

We performed a CatPCA with the five items of the sustainable foods' consumption. According to the "eigenvalue greater than one" criterion only the first component was retained (first eigenvalue equal to 1.895).

Food	Comp. loading
Legumes	0.716
Whole-grain cereals	0.635
Vegetables-based dishes	0.624
Nuts and oily seeds	0.518
Fresh fruit	0.475

Table 3: Component loadings for the CatPCA of transition for sustainable consumer's foods.

The related Cronbach's alpha was 0.555. Table 3 reports the factor loadings of the five foods: it is clear that the first component is highly influenced by the increase in the consumption of legumes, whole-grain cereals and vegetables-based dishes. This new latent construct is interpretable as the transition towards sustainable food consumption practises (TSF): the more the value is positive the more a person realized a transition to sustainable foods' consumption. In analysing the transition towards sustainable foods' consumption practises, a linear regression model was fitted. We obtained the model reported in Table 4, where R^2 equals 8.3% (adjusted $R^2 = 7.4\%$).

Parameter	Estimate	SE	p-value
Intercept	-0.272	0.273	0.320
Male (vs Female)	-0.078	0.061	0.199
Working condition (reference: Other)			
Working at home	0.092	0.077	0.228
Not working	0.247	0.110	0.024
Essential sector	0.168	0.112	0.134
Educational level (reference: University)			
Up to secondary school	-0.170	0.122	0.164
High school	-0.090	0.063	0.155
Living condition (refence: 4 persons or more)			
Single	-0.125	0.095	0.188
Couple	0.123	0.073	0.094
3 persons	-0.034	0.093	0.714
Pescatarian/veg (vs no pescatarian/veg)	0.148	0.061	0.016
BMI	-0.014	0.007	0.040
Economic Well-being	0.019	0.021	0.366
Age	0.009	0.002	<0.0001

Table 4: Parameters estimates, standard errors (se) and p-values of the predictors for the linear regression model with the dependent variable being the TSF.

The TSF index resulted to be affected (statistically significant at 90%) by: age, BMI and food diet. In particular:

- the older people resulted to have realized a greater transition to a sustainable foods' consumption;
- people with a lower BMI realized a greater transition to sustainable foods' consumption;
- respect to omnivorous people, pescatarian/vegerarian/vegan respondents had a greater transition to sustainable foods' consumption.

5. The transition for the unsustainable foods' consumption

The results of the CatPCA on the six items of the unsustainable foods' consumption showed that only the first component had a eigenvalue greater than one (eigenvalue equal to 1.888). The related Cronbach's alpha was 0.564. Table 5 reports the factor loadings of the six foods: the first component is highly influenced by the increase in the consumption of carb-based dishes, sweets and desserts. This new latent construct is interpretable as the transition towards unsustainable foods' consumption practises (TUF): the more the value is positive the more a person realized a transition to unsustainable foods' consumption. We fitted a linear regression model on TUF and we obtained the model reported in Table 6, where R^2 equals 13.5% (adjusted $R^2 = 12.5\%$).

Food	Comp. loading
Carb-based dishes	0.693
Sweets and desserts	0.675
Dairy products	0.578
Alcoholic beverage	0.567
Sugary beverage	0.416
Meat-based dishes	0.350

Table 5: Component loadings for the CatPCA of transition for unsustainable consumer's foods.

Parameter	Estimate	SE	p-value
Intercept	0.917	0.246	<0.0001
Male (vs Female)	-0.387	0.062	<0.0001
Working condition (reference: Other)			
Working at home	0.099	0.072	0.169
Not working	0.337	0.110	0.002
Essential sector	0.066	0.115	0.569
Educational level (reference: University)			
Up to secondary school	0.171	0.122	0.161
High school	-0.061	0.062	0.323
Living condition (refence: 4 persons or more)			
Single	0.128	0.093	0.168
Couple	0.131	0.079	0.097
3 persons	0.139	0.087	0.111
Pescatarian/veg (vs no pescatarian/veg)	-0.185	0.071	0.009
BMI	0.003	0.007	0.634
Economic Well-being	-0.064	0.021	0.003
Age	-0.010	0.002	<0.0001

Table 6: Parameters estimates, standard errors (se) and p-values of the predictors for the linear regression model with the dependent variable being the TUF index.

The TUF index resulted to be affected (statistically significant at 90%) by: gender, age, food diet and economic well-being. In particular:

- respect to females, males showed a higher transition to unsustainable foods' consumption;
- the transition to unsustainable foods' consumption decreased with increasing age;
- people with a higher level of economic well-being realized a lower transition to unsustainable foods' consumption;
- respect to omnivorous people, pescatarian/vegerarian/vegan respondents had a lower transition to unsustainable foods' consumption.

6. Conclusion

The outbreak of the SARS-CoV-2 pandemic caused major perturbations to the food environment in many localities of the world, further exacerbated by the introduction of social isolation and business shutdown measures intended to slow down the transmission of the virus.

This research investigated the profiles of sustainability of the transformations that occurred in the daily nutritional choices and behaviors of Italian households during the March-May 2020 general lockdown. Home confinement affected the food behaviors of our respondents and the health crisis seemed to be an occasion for a large section of interviewees to rethink food and nutrition.

During lockdown weeks, food was appreciated in its raw, fresh, seasonal, local-bound and unprocessed form, (re-)gaining relevance not only as a pleasurable hobby (cooking as a leisure activity) but also as a cornerstone of pro-health behaviors and shared social practices. This led to an improvement of the healthiness and sustainability of diets which we measured and compared through the elaboration of the transition for the sustainable foods' consumption index and the transition for the unsustainable foods' consumption index.

The evidence gathered by this research suggests that the trajectories towards such a transition are already plotted, but it will take an adequate support from cultural, political and economic institutions to create the conditions for sustainable food production and consumption to take hold as the 'new' normal in the post-pandemic era.

References

Adler, N. E., Epel, E. S., Castellazzo, G., Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, White women. *Health Psychology*, **19**(6), pp. 586-592, doi:10.1037/0278-6133.19.6.586.

Galimberti, A., Cena, H., Campone, L., Ferri, E., Dell'Agli, M., Sangiovanni, E., et al. (2020). Rethinking Urban and Food Policies to Improve Citizens Safety After COVID-19 Pandemic. *[Perspective].Frontiers in Nutrition*, **71**(181), doi:10.3389/fnut.2020.569542.

Linting, M., & van der Kooij, A. (2012). Nonlinear Principal Components Analysis With CATPCA: A Tutorial. *Journal of Personality Assessment*, **94**(1), pp. 12-25, doi:10.1080/00223891.2011.627965.

Willett, W., RockstrÃ¶m, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., et al. (2019). Food in the Anthropocene: the EAT-Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, **393**(10170), pp. 447-492, doi:10.1016/S0140-6736(18)31788-4.

Tourism and gastronomy

Wine preferences based on intrinsic attributes:
A tasting experiment in Alto Adige/Südtirol province

Luigi Fabbris, Alfonso Piscitelli

1. Introduction

Consumers choose a wine according to the information they possess regarding its intrinsic and extrinsic attributes (Charters and Pettigrew, 2003). Price, brand, region of origin, type of grapes, and awards achieved are the basic key extrinsic attributes used by different consumer groups when choosing wine (Combris et al., 1997; Batt and Dean, 2000; Lockshin et al., 2006; Martínez et al., 2006; Chrea et al., 2011; Brentari et al., 2011; D'Alessandro and Pecotich, 2013). Physical characteristics of the wine, such as taste, color, and flavor, are intrinsic attributes that play an important role in consumers' wine quality perception (Dodd et al., 2005; Carbonell et al., 2008; Rahman et al., 2014). Research evidence suggests that consumers tend to use both intrinsic and extrinsic attributes concurrently when choosing wine (Jover et al., 2004; Charters and Pettigrew, 2007; Veale and Quester, 2009; Mueller et al., 2010; Brentari and Zuccolotto, 2011). Different consumption situations may amplify or change the perception of wine characteristics (Hall and Lockshin, 2000); consumer drinking frequency also significantly and positively influences the perceptive ability of wine consumers (Rahman and Reynolds, 2015).

The classification of wine attributes into extrinsic and intrinsic refers to the hierarchical and multi-dimensional models, which in turn refer to a higher-level Total Food Quality model for product choice (Grunert, 1997). A model is multi-dimensional if the consumers' final evaluation includes more than one quality dimension and is hierarchical if each dimension of quality includes at least one product characteristic (Olson and Jacoby, 1972).

Most wine purchases do not provide the opportunity to taste them before purchasing. Nevertheless, consumers place the most emphasis on taste when it comes to wine evaluation, preference, and purchase because the intrinsic characteristics of previously experienced wines play a major role in repurchasing. Moreover, scholars (Oomen, 2015; Mueller et al., 2010) suggest that wine tasting may have such an important role in the purchase process that could ultimately lead to more sampling in wine shops. The tasting and repurchase decision process may be considered a first step towards predicting the market uptake of new wines (Mueller et al., 2001).

The goal of this study, given that the taste of wine plays an important role in people's choices, is to determine which intrinsic attributes influence wine preferences. For this, we have held an experiment on how a sample of wine consumers evaluate a set of intrinsic attributes in case they can taste the available wines. Also, we measured the impact the attributes have on consumers' preferences.

In September 2016, a sensory evaluation experiment was conducted on twelve white wines originating from six different grape varieties (Chardonnay, Müller-Thurgau, White Pinot, Sauvignon, Gewürztraminer, Riesling) of the Alto Adige/Südtirol province in Italy.

The pool of tasters included 33 individuals who typically consumed mild amounts of wine. They were selected on the basis of their interest in and availability for the experiment, as well as of their experience in wine consumption. Moreover, they were not connected to the "wine and spirits" business sector, nor were they wine makers. Neither the tasters nor the person pouring the wines knew the grape variety or cellar of any wine; hence, the tasting procedure was double-blind so as not to introduce bias or otherwise skew the results (Rivers and Webber,

1907). Just the researcher (not involved in wine preparation) knew the symbols of the experimental design. This procedure was aimed to eliminate any emotional conditioning and address the assessors' attention directly and exclusively towards the technical aspects of wines.

The wine characteristics considered in this sensory evaluation experiment were collected through an anonymous paper questionnaire. This questionnaire asked for participants to make judgments on 11 intrinsic attributes of appearance, nose, and palate. After that, they were also asked to give an overall judgment for each wine. In addition, data on background characteristics of tasters, his or her drinking habits, and the relevance of wine in his or her diet and social life were also collected.

The experiment compared wines of the same terroir and of the same vintage and then belongs to the class of the so-called horizontal tasting. This way, it is possible to obtain comparative judgements between the selected wines.

The remainder of this paper is organized as follows: Section II introduces the sensory experimental procedure, Section III introduces the statistical approach for data analysis, and Section IV reviews the main results obtained in the study. Section V concludes the presentation of this research.

2. Fractional experiment

Each taster was administered four randomly selected wines from different grapes, in accordance with a fractional factorial experiment. A fractional, or partial-profile, design is an experimental design consisting of a carefully chosen fraction of the experimental runs of a full factorial design (Box et al., 2005). In our wine-tasting experiment, the sampling was carried out at the grape-variety level, administering just four of the six possible varieties to any taster, and selecting one of the two possible cellars. This is a case in which possible choices rather than choosers are sampled (Manski and Lerman, 1977).

The sampling design followed a systematic pattern. For each grape variety, $15 = [(6 \cdot 5)/2]$, different random sets were created so that each grape variety appeared 10 times in 15 trials. This way, each wine variety had 20 repetitions after 30 tasters performed their task, though the number of repetitions of each variety by cellar is 10. With the number of tasters being 33, the number of repetitions of each variety by cellar is slightly above 10.

Wines were randomly divided into two groups of grape varieties and placed in brown bags. The first group was identified by A1 through F1, while the second group was identified by A2 through F2. The tasters were rapidly trained to familiarize them with the terms of the experiment and with the scales used.

Each taster had five glasses, one for water and the remaining for wines. The four wines were poured in a flight, and then the tasting began. In the tasting session, the judges were given 6 centilitres of each of the four randomly selected wine varieties which were served at the same cold temperature. The protocol was open, meaning that tasters could taste and re-taste before assigning preferential judgments; for each tasted wine, they evaluated also the intrinsic attributes of each.

3. Estimation method

A conditional logit regression was performed on the judgment data of intrinsic attributes of the wine in order to model the participants' choices (McFadden, 1974; 1980; Soofi, 1992). This model is consistent with economic theory and allows the relation of choices to the characteristics of the possible alternatives. According to random utility theory, individuals who choose an alternative or a profile tend to maximize their own utility. Wine utility refers both to nutritional and emotional aspects. Utility is considered a function of observed characteristics (attribute levels) and unobserved characteristics of the alternative.

The utility function is specified by the attribute levels of the alternative and by a random error term:

$$U_i = V(\beta, X_i) + \epsilon_i,$$

where V is a function linking the attribute levels of the alternative i to the utility of the alternative, and ϵ_i is a random term following an i.i.d. type-1 extreme-value distribution (McFadden, 1974). The probability of choosing the alternative i is:

$$P(choice = i) = \frac{e^{V(\beta, x_i)}}{\sum_{i=1}^{I} e^{V(\beta, x_i)}},$$

where $V(\beta, x_i)$ is the utility function, also called *part-worth utility*, for alternative i, with $i = 1, \ldots, I$. In other words, the probability of choosing an alternative i depends on both attribute levels of the profile i and attribute levels of all other profiles.

The vector of unknown utility parameters β is estimated through maximum likelihood of regularized weights. The solution is typically found using some non-linear, iterative maximization algorithm. The attribute levels are constrained, imposing that their sum equals zero. The resulting set of estimated parameters is unique, and the model is robust to violation of the assumption (Louviere et al., 2000).

The goodness-of-fit conditional logit model is evaluated through both the log likelihood ratio test and McFadden's *pseudo R^2*. The log likelihood ratio chi-square test determines whether including attribute-level variables significantly improves the model fit compared with a trivial model with no attribute. This highlights whether one or more preference weights are expected to be different from 0.

Test statistic D, log likelihood ratio, is calculated as:

$$D = 2 \log\left(\frac{L(M_{fit})}{L(M_0)}\right) = -2(LL(M_0) - LL(M_{fit}))$$

where $L(M_0)$, $L(M_{fit})$, $LL(M_0)$ and $LL(M_{fit})$ are the likelihood and the log likelihood values of the trivial and the fitted models, respectively. The log likelihood ratio follows a chi-square distribution with degrees of freedom equal to the number of parameters to be estimated. McFadden's *pseudo R^2* is calculated as:

$$Pseudo\ R^2 = 1 - \frac{LL(M_{fit})}{LL(M_0)},$$

Pseudo R^2 varies between 0 and 1. A value of *pseudo R^2* from 0.2 on can be considered a good model fit, while a value of 0.4 indicates an extremely good fit (McFadden, 1978).

The relative importance of an attribute (RIA) can be calculated as the percentage of estimated utility parameters of the levels of an attribute (the difference between parameters of the most preferred level of an attribute and the least preferred level of the same attribute):

$$RIA_j = 100 \frac{\{\max(\beta_j) - \min(\beta_j)\}}{\sum_{j=1}^{J}\{\max(\beta_j) - \min(\beta_j)\}},$$

where j indicates an attribute and J the total number of attributes used in the profile definition. RIA measures may be influenced by number of levels composing an attribute (Orme, 2010). A RIA measure varies between 0 and 100.

4. Results

Our model has been fitted using the "clogit" function from the "survival" package in R (Therneau, 2015). Table 1 shows the utility parameter estimations of conditional logit models for the intrinsic attributes of the wines. Positive significant parameter estimation means a positive effect of the attribute (level) on the choice. On the contrary, a negative significant value implies an adverse effect of that attribute (level) on the choice. Attribute levels without significant estimates do not play any role into the choice process. In addition, the RIA estimates of the 11 attributes are also shown in Table 1.

First, the pseudo R^2 equals 32.3%, which shows that the intrinsic attributes successfully explain the preferences of the involved consumers for the 6 wine grapes. Moreover, the coefficient estimates highlight that wine choices were driven chiefly by the following:

- a clear perception of wine flavors in the mouths of tasters (flavor intensity);
- a good balance of flavors, that is alcohol, acidity, tannins, sweetness and the possible fruits are in *harmony*;
- a complex bouquet (bouquet complexity) and the perception of an aroma.

The relevance of the three variables is confirmed by the estimates of the attribute importance, since the intensity of flavor, out of a hundred importance points, received 33.6, while overall harmony received 19.2, and the complexity of the bouquet received 14.3. Another slightly relevant attribute is evolutionary state, e.g. the classification of wines according to their aging potential, in fact RIA is 12.2%, but the coefficient is not statistically significant.

Unexpectedly, also appearance did not influence wines rankings: neither differences in color nor in clarity had a role in determining the final rankings. Another unexpected result is that aroma – an aspect that characterizes Gewürztraminer that the large majority of tasters judged as the most preferable among the administered wines (Table 2) – was not significant at all. This may mean that tasters evaluated the assessed wines giving much more importance to their palatal sensations than to the olfactory and visual ones. Indeed, palatal sensations refer particularly to the pleasure of eating and health implications related to wine consumption, while the others are merely aesthetic and transitory.

Table 1. Estimates of model coefficients applying conditional logit regression on wine choices and the relative (percent) importance (RIA) of attributes (n=33)

	$\hat{\beta}$	$se(\hat{\beta})$	Z	$Pr(>|z|)$	Signific.	RIA
V Clarity	0.183	0.480	0.382	0.498		3.2
V Color	-0.2292	0.474	-0.484	0.321		4.0
Visual attributes						*7.2*
O Intensity	0.490	0.470	1.042	0.810		8.5
O Complexity	0.822	0.470	1.750	0.030	*	14.3
O Aroma description	0.157	0.488	0.322	0.750		2.7
Olfactory attributes						*25.5*
G Body	-0.128	0.487	-0.264	0.950		2.2
G Balance	-0.006	0.540	-0.012	0.944		0.1
G Intensity	1.936	0.631	3.065	0.003	**	33.6
G Persistence	0.006	0.561	0.011	0.866		0.1
G Evolutionary state	0.702	0.525	1.337	0.348		12.2
Taste-Olfactory attributes						*48.2*
Harmony	1.109	0.517	2.144	0.026	*	19.2

** 0.001 < α^{oss} < 0.01; * 0.01 < α^{oss} < 0.05; R^2= 0.323 (max possible=0.694); Likelihood ratio test= 111.6; Wald test = 55.47 on 11 df; p=6.366e-08; Score (logrank) test = 91.9 on 11 df, p=7.105e-15.

Table 2. Evaluation of the tested wines, by assessors' characteristics (n=33)

	Chardonnay	Gewürztraminer	Müller-Thurgau	Pinot Bianco	Riesling	Sauvignon
Overall	4	1	3	6	2	5
Women	6	1	5	3	4	2
Men	4	1	3	6	2	5
Younger	3	1	4	6	2	5
Older	5	1	3,5	6	2	3,5

5. Final remarks

The tasting experiment described in this paper draws the conclusion that mild wine-consumers chose wines according to all sensorial dimensions, and in particular flavor and odor, but the perception of harmony among wine attributes is relevant as well. We suggest that wine was evaluated according to easy-to-perceive (that is, non-technical), general-type attributes. In fact, the attributes highlighted by respondents, on top of the overall harmony, are the intensity of flavor and the complexity of the wine's bouquet. In contrast, the more an intrinsic attribute

is peculiar of a grape – for instance, the aromas that could identify it, its color, the balance between opposing components, and the persistence of flavor – the less it factors into people's choices.

This outcome is consistent with the results that Rahman et al. (2014) obtained using a convenience sample (i.e., students, faculty and staff). Their research has highlighted that individuals place the most emphasis on taste when it comes to wine evaluation, preference, and purchase. Though, the easier aspects of wine likely dominate consumers' judgement. The authors state that, in fact, when a person is trying a wine for the first time, appearance might influence the perception of aroma and taste, and aroma might also influence the perception of taste. While our results may not be encouraging for wineries, it should be kept in mind by people who "construct" and sell wines because this knowledge is vital for increasing the success of their wine.

Going forward, we are prepared to repeat this experiment with other grapes and other participants to determine if this outcome is replicated when the study is conducted with different factors and a larger subject pool. Moreover, in order to improve the concentration of assessors on the tasting experiment, it may help if tasting is targeted to the intention to buy and also to the economic value of the tasted wines. Finally, since assessors tended to agree just on first position of the wine ranking, there is room for future analyses of the reasons why people showed such large variability on preferences.

References

Batt, P.J., Dean, A. (2000). Factors influencing the consumer's decision. *Australia and New Zealand Wine Industry Journal Marketing Supplement*, **15**(4), pp. 34–41.

Box, G.E., Hunter, J.S., Hunter, W.G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd Edition. Wiley Interscience, New York, NY.

Brentari, E., Levaggi, R., Zuccolotto P. (2011). Pricing strategies for Italian red wine. *Food Quality and Preference*, **22**(8), pp. 725–732.

Brentari, E., Zuccolotto P. (2011). The impact of chemical and sensorial characteristics on the market price of Italian red wines. *Electronic Journal of Applied Statistical Analysis*, **4**(2), pp. 265–276.

Carbonell, L., Izquierdo, L., Carbonell, I., Costell, E. (2008). Segmentation of food consumers according to their correlations with sensory attributes projected on preference spaces. *Food Quality and Preference*, **19**(1), pp. 71–78.

Charters, S., Pettigrew, S. (2003). The intrinsic dimensions of wine quality: An exploratory investigation. In L., Lockshin, C., Rungie (eds.) *Proceedings of the International Colloquium in Wine Marketing* (pp. 1-21). University of South Australia, Adelaide, AUS.

Charters, S., Pettigrew, S. (2007). The dimensions of wine quality. *Food Quality and Preference*, 18(7), pp. 997–1007.

Chrea C., Melo, L., Evans, G., Forde, C., Delahunty, C., Cox, D.N. (2011). An investigation using three approaches to understand the influence of extrinsic product cues on consumer behavior: An example of Australian wines. *Journal of Sensory Studies*, **26**(1), pp. 13–24.

Combris, P., Lecocq, S., Visser, M. (1997). Estimation of a hedonic price equation for Bordeaux wine: Does quality matter? *Economic Journal*, **107**(441), pp. 390–402.

D'Alessandro, S., Pecotich, A. (2013). Evaluation of wine by expert and novice consumers in the presence of variations in quality, brand and country of origin cues. *Food Quality and Preference*, **28**(1), pp. 287–303.

Dodd, T.H., Laverie, D.A., Wilcox, J.F., Duhan, D.F. (2005). Differential effects of experience, subjective knowledge, and objective knowledge on sources of information used in consumer wine purchasing. *Journal of Hospitality & Tourism Research*, **29**(1), pp. 3–19.

Grunert, K.G. (1997). What's in a steak. A cross-cultural study on the quality perception of beef. *Food Quality and Preference*, **8**(3), pp. 157–174.

Hall, J., Lockshin, L. (2000). Using means-end chains for analysing occasions-not buyers. *Australasian Marketing Journal*, **8**(1), pp. 45–54.

Jover, A.J.V., Montes, F.J.L., Fuentes, M.D.M.F. (2004). Measuring perceptions of quality in food products: The case of red wine. *Food Quality and Preference*, **15**(5), pp. 453–469.

Lockshin, L., Jarvis, W., d'Hauteville, F., Perrouty, J.P. (2006). Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice. *Food Quality and Preference*, **17**(3), pp. 166–178.

Louviere, J.J., Hensher, D.A., Swait, J.D. (2000). *Stated Choice Methods: Analysis and Applications*, Cambridge University Press, Cambridge, UK.

Manski, C.F., Lerman S.R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica*, **45**(8), pp. 1977–1988.

Martínez, L.M.C., Mollá-Bauzá, M.B., Gomis, F.J.D.C., Poveda, Á.M. (2006). Influence of purchase place and consumption frequency over quality wine preferences. *Food Quality and Preference*, **17**(5), pp. 315–327.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior, in P. Zarembka (ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press, New York, NY.

McFadden, D. (1978). Quantitative methods for analyzing travel behaviour of individuals: Some recent developments, in D., Hensher, P., Stopher (eds.), *Behavioural Travel Modelling* (pp. 279–318). Croom Helm London, London, UK.

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, **53**(3), pp. S13–S29.

Mueller, S., Osidacz, P., Francis, I.L., Lockshin, L. (2001). The relative importance of extrinsic and intrinsic wine attributes: Combining discrete choice and informed sensory consumer testing. *5th International Conference of the Academy of Wine Business Research, February 2001, Auckland, NZ.*

Mueller, S., Osidacz, P., Francis, I.L., Lockshin, L. (2010). Combining discrete choice and informed sensory testing in a two-stage process: Can it predict wine market share?. *Food Quality and Preference*, **21**(7), pp. 741–754.

Olson, J.C., Jacoby, J. (1972). Cue utilization in the quality perception process. In M., Venkatesan (ed.), *Proceedings of the Third Annual Conference of the Association for Consumer Research* (pp. 167–179). Association for Consumer Research, Chicago, IL.

Oomen, R. (2015). The role of tasting in the purchasing process. *BIO Web of Conference* (DOI: 10.1051/bioconf/20150503010).

Orme, B. (2010). Interpreting the results of conjoint analysis. In: *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research,* Second edition (pp. 77–88). Research Publishers LLC, Manhattan Beach, CA.

Rahman, I., Reynolds, D. (2015). Wine: Intrinsic attributes and consumers' drinking frequency, experience, and involvement. *International Journal of Hospitality Management*, **44**, pp. 1–11.

Rahman, I., Stumpf, T., Reynolds, D. (2014). A comparison of the influence of purchaser attitudes and product attributes on organic wine preferences. *Cornell Hospitality Quarterly*, **55**(1), pp. 127–134.

Rivers, W.H.R., Webber, I.N. (1907). The action of caffeine on the capacity for muscular work. *The Journal of Physiology*, **36**(1), pp. 33–47.

Soofi, E.S. (1992). A generalizable formulation of conditional logit with diagnostics. *Journal of the American Statistical Association*, **87**(419), pp. 812–816.

Therneau, T. (2015). *A Package for Survival Analysis in S. version 2.38*. available at: https://CRAN.R-project.org/package=survival.

Veale, R., Quester, P. (2009). Do consumer expectations match experience? Predicting the influence of price and country of origin on perceptions of product quality. *International Business Review*, **18**(2), pp. 134–144.

Profiling visitors of a national park in Italy through unsupervised classification of mixed data

Giulia Caruso, Adelia Evangelista, Stefano Antonio Gattone

1. Introduction

The success of a tourism destination, among other things, relies on the implementation of a strategic marketing plan. Since the identification and understanding of customers features and needs are essential for a correct market segmentation, the use of inappropriate techniques could result in missing strategic marketing opportunities (Bloom, 2004, Thompson & Schofield, 2009). Furthermore, any subsequent marketing activity would incur the risk to disappoint customers' expectations, producing their dissatisfaction. Moreover, the segmentation of markets based on visitor features and their motivations enables the identification of strengths and opportunities of a market (Lee & Lee, 2001).

The main benefit of market segmentation lies in knowledge acquisition. Profiling visitor allows to identify current consumers travel behaviour and to forecast future ones (Suleiman & Mohamed, 2011), enabling to acquire a competitive advantage (Hsu & Kang, 2003; Bui & Le, 2016, Koshy et al, 2019).

The aim of our study is to determine visitors characteristics and their satisfaction toward facilities of the National Park of Majella, in Italy. The outcome of our analysis is expected to serve as a guide for tourism operators, in order to facilitate plans toward formulating robust marketing strategies aimed to enhance visitors satisfaction. Our data have been collected on-site, from a sample of park visitors, and include both continuous and categorical features. In order to cluster such kind of data, we used an unsupervised classification method, specific for mixed data.

The paper is articulated as follows: in Section 2 we explain our data and consider the main clustering approaches for mixed variables, whereas in Section 3 we show the results obtained by the application of these methods to our dataset, providing an evaluation of the clustering results, by means of internal and external validity indexes. Finally, in Section 4, we draw some conclusions and discuss some suggestions for future research.

2. Data and method

Our dataset results from a questionnaire which has been collected on-site, from a sample of visitors of the Park, during the period from July 16 until October 27, 2020. A total of 523 tourists has been interviewed.

The Majella National Park is in Abruzzo, central Italy, and incorporates the provinces of Chieti, L'Aquila and Pescara, including 39 municipalities, characterized by a high spatial heterogeneity. This natural area is crucial for the protection of the natural ecosystem and for the socio-economic development of the area.

These data allow to perform a qualitative analysis on visitors of the Majella National Park, and consequently to assess their satisfaction level on the Park services.

The variables analysed are 16 (9 numerical - 7 categorical) and the entries are 523. The numerical variables concern the visitors perceived quality (measured in a 5 point Likert scale) on the following aspects: the web site, the naturalistic heritage conservation, the adequate presence of signage, of public transport, of children amenities, of footpaths maintenance, of accommodation facilities, of restaurant services and of food and wine products. The qualitative variables, instead, involve the following variables:

Giulia Caruso, Gabriele d'Annunzio University, Italy, giulia.caruso@unich.it, 0000-0003-0236-6201
Adelia Evangelista, Gabriele d'Annunzio University, Italy, adelia.evangelista@unich.it
Stefano Antonio Gattone, Gabriele d'Annunzio University, Italy, antonio.gattone@unich.it, 0000-0002-6143-9012

customers' expectations, the aim of their trip, the chosen location and how they came to its knowledge, the number of overnight stays, the type of chosen accommodation and, finally, the daily average expenditure per person.

In literature, most clustering approaches are limited to numerical or categorical data only. The traditional approach, instead, when dealing with both quantitative and qualitative variables, is to convert the latter values into numerical ones, and then apply the quantitative value based clustering methods (Foss et al, 2016; Ichino et al, 1994, Caruso et al, 2018). However, this approach would ignore the similarity information enclosed in the qualitative attributes, producing a loss of knowledge (Ahmad, A. & Dey, L. 2007). Finding a unified similarity metric for both kind of data, instead, would allow to remove the metric gap between them. Therefore, in order to detect different clusters, we compared two of the most used mixed data clustering methods, namely, the methods of Huang (Huang, Z., 1997) and Cheung & Jia (Cheung, Y. & Jia, H., 2013).

For sake of brevity, we will not describe in detail the methods we adopted to analyse the variables; the reader may consult our previous works for details (Caruso et al 2018-2019).

3. Results

We implemented a cluster analysis with a number of clusters equal to 3. Table 1 displays, for each cluster, the mean value of the 9 quantitative attributes analyzed and shows that the patterns produced by the two performed methods, specific for mixed data, are quite similar among them. The Huang one, in particular, highlights a slightly stronger clustering structure, meaning that the dissimilarity between clusters is higher.

Method	Cluster	Size	Signage	Footpaths	Restaurant services	Food & wine products	Public transports	Accommodation facilities	Web site	Children amenities	Natural heritage conservation
Huang	1	246	3.43	3.46	4.16	4.25	2.71	4.11	3.74	3.52	4.06
	2	171	4.44	4.53	4.75	4.82	3.84	4.71	4.57	4.39	4.68
	3	106	2.53	2.44	3.17	3.38	1.96	3.02	2.95	2.52	2.87
Cheung	1	201	2.96	2.91	3.69	3.85	2.31	3.54	3.36	2.95	3.44
	2	161	3.49	3.54	4.22	4.28	2.77	4.18	3.80	3.71	4.11
	3	161	4.45	4.53	4.65	4.76	3.84	4.67	4.52	4.31	4.66

Table 1. Cluster mean values for quantitative variables.

Figures 1 and 2 show the boxplot of the variables "Signage" and "Footpaths" in each cluster. The visual analysis highlights different median values in each group. Similar behaviours have been observed for the remaining quantitative variables.

Table 2 reports the results for the variable "**overnight stays**". The mode of the marginal distribution is represented by the value "1-3 nights stays" (42%). The clusters identified by the Cheung method are characterized with three different modes "1-3 nights stays" (Cluster 2), "4-7 nights stays" (Cluster 3) and "more than 7 nights stays" (Cluster 1).

The Huang method produced a slightly different result with two clusters out of three having mode "1-3 nights stays".
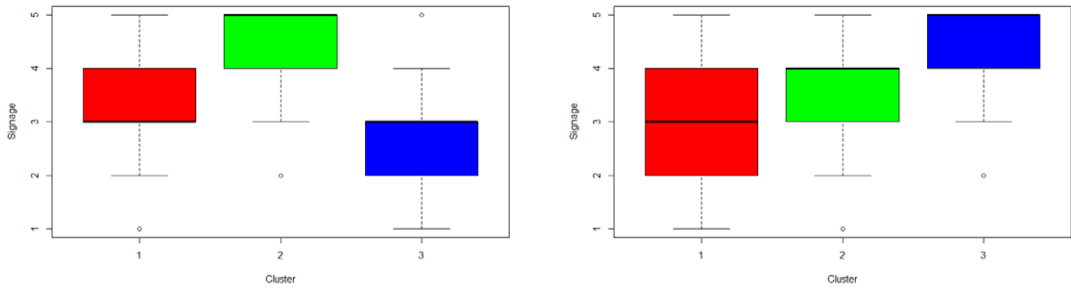
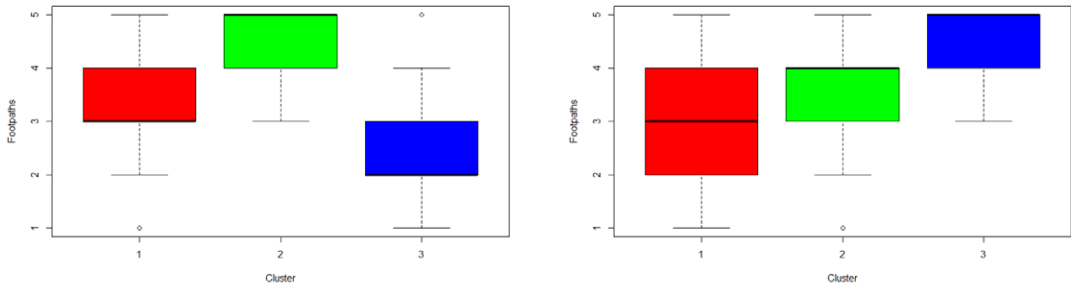**Figure 1.** Quantitative variable Signage: boxplots for Huang (left panel) and Cheung (right panel) method.

Figure 2. Quantitative variable Footpaths: boxplots for Huang (left panel) and Cheung (right panel) method.

OVERNIGHT STAYS	Huang			Cheung			
Cluster	**1**	**2**	**3**	**1**	**2**	**3**	**Marginal**
1-3 nights stays	**0.54**	**0.35**	0.24	0.23	**0.71**	0.35	**0.42**
4-7 nights stays	0.26	0.33	0.15	0.17	0.23	**0.40**	0.26
More than 7 nights stays	0.20	0.32	**0.61**	**0.60**	0.06	0.25	0.32

Table 2. Categorical variable Overnight stays: marginal and conditional distribution for each cluster.

A similar pattern can be observed with regards to the variable "Accommodation" (Table 3). The clusters identified by the Cheung method have different modes, i.e. "Other" (Cluster 3), "Second house" (Cluster 1) and "Hotel" (Cluster 2) while Clusters 2 and 3 of Huang have the same mode "Second house".

ACCOMMODATION	Huang			Cheung			
Cluster	1	2	3	1	2	3	Marginal
Other	0.22	0.22	0.17	0.17	0.17	**0.29**	0.21
Rented apartment	0.07	0.11	0.14	0.13	0.02	0.12	0.09
B&B/ rented rooms	0.20	0.23	0.16	0.13	0.25	0.24	0.20
Second house	0.22	**0.30**	**0.43**	**0.53**	0.06	0.21	**0.29**
Hotel	**0.31**	0.13	0.09	0.03	**0.50**	0.14	0.21

Table 3. Categorical variable Accommodation: marginal and conditional distribution for each cluster.

With regards to the variable "Expenditure" (Table 4), the mode of the marginal distribution is represented by "10-30 Euros" (36%). The same result is observed in two out of three clusters for both methods.

EXPENDITURE	Huang			Cheung			
Cluster	1	2	3	1	2	3	Marginal
10-30 €	0.26	**0.46**	**0.42**	**0.42**	0.14	**0.50**	**0.36**
30-50 €	0.33	0.29	0.35	0.41	0.23	0.30	0.32
Less than 10€	0.06	0.08	0.05	0.06	0.04	0.08	0.06
More than 50€	**0.35**	0.17	0.18	0.10	**0.60**	0.11	0.26

Table 4. Categorical variable Expenditure: marginal and conditional distribution for each cluster.

With regards to the variable "Expectation" (Table 5), most tourists visited the park in order to take "guided tours for environmental education" (45%). This result is in line with all clusters produced by the Huang method and by two clusters obtained by the Cheung method.

EXPECTATION	Huang			Cheung			
Cluster	1	2	3	1	2	3	Marginal
Other	0.31	0.20	0.27	0.23	**0.43**	0.15	0.27
Flora observation	0.29	0.30	0.22	0.22	0.38	0.25	0.28
Guided tour for environmental education	**0.40**	**0.50**	**0.51**	**0.54**	0.19	**0.60**	**0.45**

Table 5. Categorical variable Expectation: marginal and conditional distribution for each cluster.

INTERNAL INDEXES	Huang	Cheung
CH	**189.4317**	106.4502
SHI	**0.1853808**	0.07521456
H	1.062639	**1.00835**

Table 6: Internal indexes for each method.

Synthetizing, by using the **Huang** method, cluster 1 differs from the others because it is characterized by tourists which stay in hotel, from 1 up to 3 nights, with an average daily expenditure of Euro 50,00. Cluster 2, instead, includes visitors which choice falls on B&B or rented rooms, for a period from 1 to 3 nights and which the average daily expenditure ranges from Euros 10 to 30. Visitors belonging to cluster 3, instead, choose their second house and they stay for more of 7 nights and with an average daily expenditure which ranges from 10 to 30 Euros.

When using the **Cheung** method, cluster 1 includes tourists which stay in their second houses, for more than 7 nights, and which daily expenditure ranges from 10 and 30 Euros. The aim of their visit is to take guided tours for the environmental education and their final goal is relaxation. Tourists inside cluster 2, instead, choose to stay in hotel, from 1 up to 3 nights, and they spend more than 50 Euros per day. Both in case of expectation and motivation they selected the option "other". The tourists of cluster 3 choose an alternative kind of accommodation and they stays from 4 to 7 nights. Their daily expenditure goes from 10 to 30 Euros. Their expectation is to take guided visits for the environmental education and their aim is to relax.

Internal validity Indexes were computed in order to evaluate the quality of the cluster solutions. Results are shown in Table 6.

For numerical variables, the Calinski-Harabasz and the Silhouette Indexes are reported. Higher values correspond to better results; thus, the method of Huang is the one performing better when it comes to quantitative variables. With regards to the Internal Index for categorical variables, we used the Entropy Index. In this case a lower value of H corresponds to the best clustering result. The best (lowest) result for Entropy is obtained by using the Cheung method.

4. Conclusions

In order to detect clusters in a more efficient way, it is very useful to dispose also of qualitative variables. Our main aim was to observe the results of each method and to detect which one performs better. From our analysis it appears clearly that it corresponds to the Huang one as for the numerical variables, whereas the method of Cheung allows to obtain better results when it comes to qualitative ones.

Our objective for the future research is to develop new clustering analysis techniques for mixed data, which will consider an interesting insight provided by the work of Diday & Govaert, proposing an adaptive dynamic clustering procedure useful to calibrate the weights between qualitative and quantitative variables.

References

Basak, J., De, R., Pal, S. (1998). Unsupervised feature selection using a neuro-fuzzy approach, *Pattern Recognition Letters*, **19** (11), pp. 997-1006.

Bloom, J. Z. (2004). Tourist market segmentation with linear and non-linear techniques. *Tourism Management*, **25**, pp. 723–733.

Bui, H. T., Le, T. A. (2016). Tourist satisfaction and destination image of Vietnam's Ha Long Bay. *Asia Pacific Journal of Tourism Research*, **21**(7), pp. 795-810.

Caruso, G., Gattone, S.A., Balzanella, A., Di Battista, T. (2019). Cluster analysis: an application to a real mixed-type data set, in *Models and Theories in Social Systems. Studies in Systems, Decision and Control*, eds. C. Flaut, S. Hoskova-Mayerov´a and D. Flaut, vol. **179**, Springer, Heidelberg, pp. 525–533.

Caruso, G., Gattone, S.A., Fortuna, F., Di Battista, T. (2018). Cluster analysis as a decision-making tool: a methodological review, in *Decision Economics: In the Tradition of Herbert A. Simon's Heritage. Advances in Intelligent Systems and Computing*, eds. E. Bucciarelli, S. Chen, J.M. Corchado, vol. **618**, Springer, Heidelberg, pp. 48–55.

Cheung, Y., Jia, H. (2013). Categorical and numerical attribute data clustering based on a unified similarity metric without knowing cluster number, *Pattern Recognition*, **46** (8), pp. 2228-2238.

Diday, E., Govaert, G. (1997). Classification automatique avec distances adaptatives. R.A.I.R.O. *Informatique Computer Science,* **11** (4), pp. 329–349.

Foss A., Markatou, M., Ray, B., Heching, A. (2016). A semiparametric method for clustering mixed data, *Machine Learning*, **105**, pp. 419-458.

Hsu, C. H., Kang, S. K. (2003). Profiling asian and western family independent travelers (FITS): An exploratory study. *Asia Pacific Journal of Tourism Research*, **8** (1), pp. 58-71.

Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values, in: *Proceedings in the First Pacific-Asia Conference on Knowledge Discovery and Data Mining,* Singapore: World Scientific, pp. 21–34.

Ichino, M., Yaguchi, H. (1994). Generalized minkowski metrics for mixed feature type data analysis, *IEEE Transactions on Systems, Man and Cybernetics,* **305** (24), pp. 698-708.

Lee, C. K., Lee, T. H. (2001). World culture EXPO segment characteristics. *Annals of Tourism Research*, **28** (3), pp. 812–816.

Nitanan Koshy, M. et al. (2019). Profiling the segments of visitors in adventure tourism: comparison between visitors by recreational sites. *International Journal of Business and Society*, **20** (3), pp. 1076-1095.

Suleiman, J. S., & Mohamed, B. (2011). Profiling visitors to Palestine: the case of Bethlehem city. *The Journal of Tourism and Peace Research,* **1** (2), pp. 41-52.

Thompson, K., Schofield, P. (2009). Segmenting and profiling visitors to the Ulaanbaatar Naadam festival by motivation. *Event Management*, **13**, pp. 1-15.

Yeung, D., Wang, X. (2002). Improving performance of similarity-based clustering by feature weight learning, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24** (4), pp. 556-561.

Using eye-tracking to evaluate the viewing behavior on tourist landscapes

Gianpaolo Zammarchi, Giulia Contu, Luca Frigau

1. Introduction

According to World Travel & Tourism Council (WTTC), tourism's direct and indirect impact accounted for 10.3% of global GDP, and one over ten jobs around the World are tourism-related (WTTC, 2020). In the last years, a sheer number of people started to use Internet as a primary source to search for travel information and choose their travel destination (Garín-Muñoz et al., 2011). In this sense, digital media now exert a relevant influence on tourism management. Several hotels, travel agencies, or other entities (e.g., municipalities, cultural sites, or leisure destinations) use websites, social media accounts, or pages on travel fare aggregators/search engines to attract clients. All these resources make use of a high number of images to transmit the attractiveness of their destinations (Ruhanen et al., 2013). The image can influence travel choice and behavioral intention (Wang & Sparks, 2016). The effectiveness of these tools might be enhanced by exploiting information on user viewing behavior, which can be provided by eye-tracking technology (Scott et al., 2019). Eye-tracking allows measuring the exact position of the eyes during the visualization of images, texts, or other visual stimuli. Consequently, eye-tracking data can be used to compute quantitative measures of viewing behavior that can provide information useful for many applications, such as improving the effectiveness of a website or consumer segmentation.

The first aim of this study is to analyze viewing behavior on images depicting natural and city landscapes. The visual processing of tourism image is investigated in order to evaluate the tourists' perceived destination image and the capacity to impact on the tourist decision making process (Li et al., 2016). The second goal is to compare performances of different widely used supervised and unsupervised models in the classification of these two classes of images.

2. Materials

The dataset used in this study comprises 1003 images (779 in landscape mode and 228 in portrait mode), mostly depicting natural indoor or outdoor scenes, obtained from the MIT saliency benchmark repository (freely available online) (Judd, 2009). Data were collected from a group of 15 participants (ages: 18-35). Each participant looked at each image for 3 seconds in free viewing (no specific instruction given to the subjects prior to the experiment) with a 1 second pause (gray screen) between images. Viewers were seated in a dark room two feet apart from the screen (19" and 1280x1024 resolution), and a chin rest was used to stabilize the head (to limit the range of motion). The eye-tracker used for the study was an ETL 400 ISCAN 240Hz model. Data do not contain the first fixation (point observed) of each participant on each image to correct for the central fixation bias (Busswell, 1935; Mannan et al., 1996; Parkhurst & Niebur, 2003; Itti, 2004). The images were collected from two online repositories: Flickr and LabelMe are very different in nature (e.g., people, animals, objects, buildings, mountains, and so on). In this study, we assigned each image to one of three possible classes: (i) natural landscapes, (ii) city landscapes, (iii) other. To assign each image to one of these three classes, we have taken into account the main element of the image. Since

Gianpaolo Zammarchi, University of Cagliari, Italy, gp.zammarchi@unica.it, 0000-0002-9733-380X
Giulia Contu, University of Cagliari, Italy, giulia.contu@unica.it, 0000-0001-9750-1896
Luca Frigau, University of Cagliari, Italy, frigau@unica.it, 0000-0002-6316-4040

our focus was the behavior of people looking at natural or city landscapes, we selected only images where the main element depicted on the scene was a natural landscape or a city landscape. For example, if the image depicts a valley or a desert, it would be classified as "natural landscape". Conversely, if the whole image was focused on a single flower, even if flowers are typical elements of natural environments, that image would be classified as "other". At the end of the manual labelling, we removed every image classified as "other" (591 images), and the remaining 412 images (187 classified as "city landscape" and 225 classified as "natural landscape") were used for subsequent analyses. Figure 1 represents an example of each of the two classes: (a) city landscapes and (b) natural landscapes.

Figure 1. Examples of (a) city landscapes and (b) natural landscapes
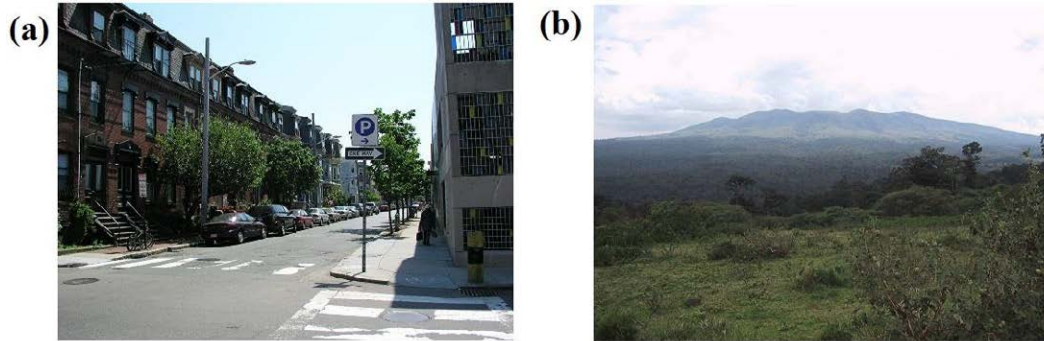
The landscape is considered as a "factor of attraction and development for tourism" (Jiménez-García et al., 2020). Our hypothesis was that an average user (e.g., a visitor of a touristic website) tends to look at a city landscape shifting from one object to another (e.g., from a car to a building to a road sign), while a natural environment might represent a more homogenous picture with fewer different stimuli to focus on. In accordance, if we measure the path followed by the observer's eye on a picture, we should expect a longer path in city landscapes than in natural environment pictures.

For each image, we calculated two metrics reflecting the viewing behavior of participants: number of fixations and path length covered by the eye gaze of each participant during observation of each image (computed for each image, using X and Y coordinates of each fixation, as the sum of the Euclidean distances between fixations). The normality of distribution for both variables was assessed using Shapiro-Wilk test. Homogeneity of variance was assessed using Levene's test. Based on results from these tests, Mann Whitney's U test and Welch's t-test were used to compare the number of fixations and the path length between the two classes of images, respectively.

Next, we used a classification approach using the path length and the number of fixations as predictors and the image class as the outcome. We applied supervised and unsupervised methods and compared the results for logistic regression (LR) with a decision rule, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and K-nearest neighbours (KNN). The four models are trained using 80% (n = 330) of the images and tested over the remaining 20% (n = 82) using k-fold cross-validation (k = 5). We also compared the hard clustering performed using K-Means Clustering algorithm (K-means) with the soft clustering performed using Gaussian Mixture Model clustering method (GMM) to show which one provides better visualization. K-means and GMM are both popular clustering methods which work following an iterative procedure, but the former is non-probabilistic and performs hard assignments, that is, each point can only belong to one class while the latter is a probabilistic algorithm based on multivariate Gaussian distributions as in eq. (1)

$$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\} \tag{1}$$

so that, when the EM (expectation-maximization) algorithm converges, each point is assigned to a class with a certain probability. GMM is more flexible than K-means because it allows decision boundaries to assume an elliptical shape while K-means only a circular shape. All analyses were carried out with R (v. 3.6.3, R Core Team, 2020) using the packages mclust (Scrucca et al., 2016), MASS (Venables & Ripley, 2002), class (Venables & Ripley, 2002), factoextra, and ggplot2 (Wickham, 2009).

3. Results

We observed a significant difference in both path length and number of fixations between natural and city images. Namely, we observed shorter path length ($p < 0.001$) and number of fixations ($p < 0.001$) in natural compared to city landscapes (Table 1).

Table 1. Summary statistics for path length and number of fixations

	Path length (pixel)		Number of fixations	
	Natural (n=187)	City (n=225)	Natural (n=187)	City (n=225)
Min	4668	8011	70	79
Q1	14267	18522	103	116
Median	17504	21317	112	123
Mean (± SD)	17766 (± 4942)	21431 (± 4322)	111.2 (± 12.84)	123 (± 12.67)
Q3	21287	24298	120	131
Max	31938	32020	148	160

Next, we applied several widely used classification methods to assess if path length and number of fixations could be used to automatically separate pictures of natural and city landscapes. The results of LR, LDA, QDA, and KNN are showed in Table 2.

Table 2. Performance of four models (LR, LDA, QDA, and KNN) in the classification of landscapes

	LR	LDA	QDA	KNN
Sensitivity	**0.743**	0.724	0.719	0.662
Specificity	0.608	0.616	0.642	**0.662**
Accuracy	**0.680**	0.672	0.680	0.621
F1-score	**0.714**	0.704	0.707	0.653

Best performances are reported in bold.

As shown in Table 2, the four classification methods showed very similar results. In particular, sensitivity ranged from slightly above 66% to 74%, and specificity had the lowest values (with best performance achieved by KNN with 66%). This means that most misclassification errors are made when we try to predict the "city landscapes" class. The accuracy ranged from 62% to 68% and that means that, overall, we make many errors when we try to assign images to one of the classes. The results show that the highest accuracy was obtained by logistic regression, which also reached the highest sensitivity and F1-score, so overall can be considered as the best classification method for this task. Finally, we compared the results of two unsupervised classification methods. Since we have two classes of images, we set the number of clusters equal to two. This number was confirmed to be the optimal number of clusters by the plot shown in Figure 2, obtained using the silhouette method.

Figure 2. Optimum number of clusters based on the silhouette method
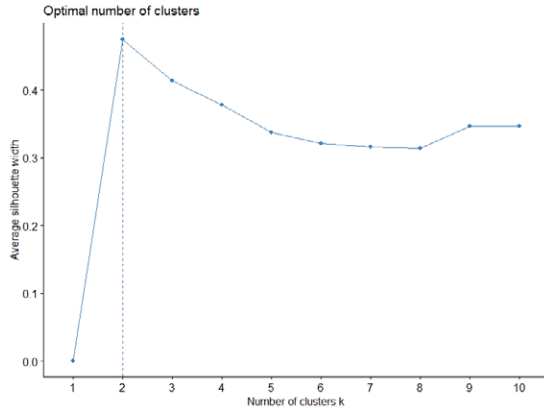
K-means and GMM provided very similar results, as we can see from Figure 3. Both in K-means clustering and GMM plots, the "city landscapes" class is colored in blue and the "natural landscapes" class in red. We used different symbols for correctly classified points (an empty circle for city and an empty square for nature) and misclassified points (a filled circle for city and a filled square for nature). If we compare the two plots from panel (a) and panel (b) we can see that the two methods produce very similar results as regards to misclassification errors.

Figure 3. Comparison of clustering using (a) K-means and (b) GMM

(a)   (b)
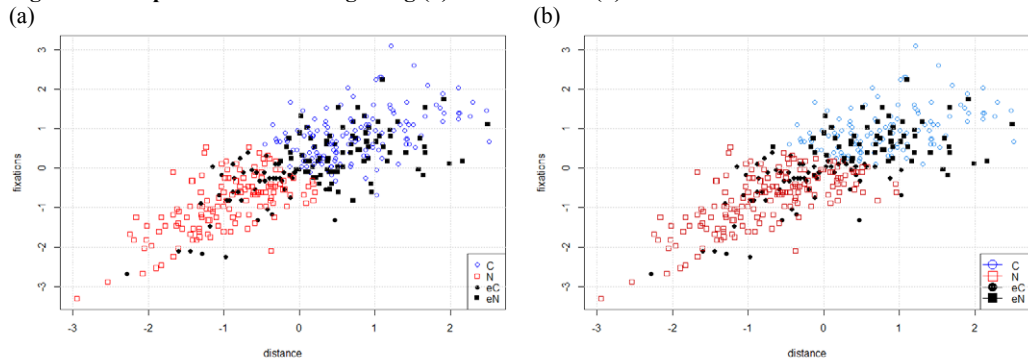
Legend: C: city landscapes, N: natural landscapes, eC: fixations erroneously classified as city landscapes, eN: fixations erroneously classified as natural landscapes

4. Discussion

In our study we showed that, given a set of images depicting a city or natural environment, it is possible to perform an automatic classification in the two classes using only path distance and number of fixations. To do this we used a subset (412 images) of the MIT dataset (1003 images depicting a large variety of subjects) available online on a public repository, selecting only those images manually labelled as "natural landscapes" or "city landscapes". We used the path length and the number of fixations in our preliminary statistical analysis showing that both metrics were significantly lower in natural compared to city landscapes. This result is in accordance with our hypothesis that natural landscapes are easier to visually explore, possibly due to a generally lower number of objects of interest and a more homogeneous background compared to city images. This result is in line with Wang & Sparks (2016), who have underlined how nature images are easier to comprehend, and with Dupont

et al (2013) who have discovered that a panoramic photograph may be easier to recognize and memorize.

We also compared four widely used classification methods (LR, LDA, QDA and KNN) in the classification of images in natural and city landscapes. Performances were very similar, but logistic regression proved to be the best method based on the highest sensitivity, accuracy and F1-score and a slightly lower specificity compared to KNN. Our results can be useful for example, for stakeholders involved in tourism management who have to decide whether to insert images depicting "city landscapes" or "natural landscapes" in their web portals. The choice could fall on images of "natural landscapes" as these can be observed with a lower number of fixations (therefore leaving more time for the user to explore a higher number of pictures or other parts of the website), or on images of the city with a reduced number of elements, in order to simplify their perception. In general, the results suggest the necessity to simplify the communication through images which should be clear, simple and with few elements that can attract the viewers' attention.

5. Conclusions

In the last two decades, tourism promotion is deeply changed and the use of images through websites and travel aggregators for the travel and tourism industry has become crucial to promote travel destinations. Particular attention has been posed on the literature to identify the best images to insert in websites. In this paper, we have investigated the different viewing behavior on images depicting natural and city landscapes. The aim was to evaluate how different classes of images are observed and which images can be easily processed by our brain, thus being potentially more effective in the engagement of viewers. In order to reach this aim, we analyzed eye-tracking data focusing on two metrics: number of fixations and path length. The results showed significant differences in viewing behavior between images picturing natural and city landscapes. The natural images were perceived as easier to visually explore. Moreover, the results have highlighted a relevant utility of the analysis of eye-tracking data to gain insights into the use of images in tourism promotion. The comparison of the performances of different supervised models showed similar performances in the classification of the two classes of images with logistic regression achieving slightly better results. Finally, two commonly used unsupervised methods produced very similar results as regards to misclassification errors when dividing the observations in two clusters. The main limitations of our study include the small number of participants for which viewing behavior data were available as well as the limited number of metrics that we were able to analyze. For instance, as time of observation was fixed to 3 seconds for each image, it was not possible to use this variable as a predictor. Additionally, removal of images not depicting city or natural landscapes resulted in a relatively small dataset (especially when we divided it into training and test set). However, this limitation was partially addressed using a k-fold cross-validation approach, that allows to exploit the entire dataset. Nonetheless, our results should be confirmed in larger and independent datasets. Future developments of this study will involve the analysis of images from different datasets to assess whether other variables (e.g., time of observation) might be helpful to reduce the misclassification errors.

References

Busswell, G.T. (1935). *How people look at pictures: A study of the psychology of perception in art*. University of Chicago Press, Chicago, (US).

Dupont, L., Antrop, M., Van Eetvelde, V. (2013). Eye-tracking analysis in landscape perception research: influence of photograph properties and landscape characteristics. *Landscape Research*, **39**(4), pp. 1-18.

Garín-Muñoz, T., Amaral, T. (2011). Internet Usage for Travel and Tourism. The Case of Spain. *Tourism Economics,* **17**, pp. 1071-1085.

Itti, L. (2004). Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Transactions on Image Processing*, **13**, pp. 1304–1318.

Jiménez-García, M., Ruiz-Chico, J., Peña-Sánchez, A.R. (2020). Landscape and Tourism: Evolution of Research Topics. *Land*, **9**(12), pp. 1-17.

Judd, T., Ehinger, K., Durand, F., Torralba, A. (2009). Learning to Predict where Humans Look. ICCV 2009.

Li, Q., Huang, Z., Christianson, K. (2016). Visual attention toward tourism photographs with text: An eyetracking Study. *Tourism Management*, **54**, pp. 243-258.

Mannan, S.K., Ruddock, K.H., Wooding, D.S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, **10**, pp. 165–188.

Parkhurst, D.J., Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, **16**, pp. 125–154.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, (AT). URL https://www.R-project.org/.

Ruhanen, L.M., McLennan, C.-L. J., Moyle, B.D. (2013). Strategic issues in the Australian tourism industry: A 10-year analysis of national strategies and plans. *Asia Pacific Journal of Tourism Research*, **18**(3), pp. 220–240.

Scott, N., Zhang, R., Le, D., Moyle, B. (2019) A review of eye-tracking research in tourism. *Current Issues in Tourism*, **22**(10), pp. 1244-1261.

Scrucca, L., Fop, M., Murphy, T.B., Raftery, A.E. (2016). mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, **8**(1), pp. 289–317.

Venables, W.N., Ripley, B.D. (2002). Modern Applied Statistics with S, Fourth edition. Springer, New York, (US). ISBN 0-387-95457-0.

Wang, Y., Sparks, B.A. (2016). An Eye-Tracking Study of Tourism Photo Stimuli: Image Characteristics and Ethnicity. *Journal of Travel Research*, **55**(5), pp. 588-602.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, New York, (US).

WTTC, GLOBAL ECONOMIC IMPACT & TRENDS 2020. https://wttc.org/Research/Economic-Impact/moduleId/1445/itemId/91/controller/DownloadRequest/action/QuickDownload. Last access: 8 December 2020.

Decomposing tourists' sentiment from raw NL text to assess customer satisfaction

Maurizio Romano, Francesco Mola, Claudio Conversano

1. Introduction

Starting from Natural Language text corpora, considering data that is related to the same context, we define a process to extract the sentiment component with a numeric transformation. Considering that the Naïve Bayes model, despite is simplicity, is particularly useful in related tasks such as spam/ham identification, we have created an improved version of Naïve Bayes for a NLP task: Threshold-based Naïve Bayes Classifier (Romano et al. (2018) and Conversano et al. (2019)).

The new version of the Naïve Bayes classifier has proven to be superior to the standard version and the other most common classifiers. In the original Naïve Bayes classifier, we face two main problems:

- A response variable is needed: we need to know a priori the "Positive" ("Negative") label of a consistent amount of comments in the data;
- There is some hand-work to be done: consistently reducing the dimensionality of the problem, is a keystone for a sentiment classification task. That means to "merge words by their meanings", and usually it is done by hand. This leads to major problems in terms of subjectivity while those words are merged, moreover it prevents to consistently run an automatic program.

2. The data

For this study, we have collected two separated – but related – datasets obtained from: Booking.com and TripAdvisor.com. More in detail, with an ad hoc web scraping Python program, we have obtained from Booking.com data about:

- 619 hotels located in Sardinia
- 66,237 reviews, divided in 106,800 comments (in Italian or English): 44,509 negative + 62,291 positive
- Period: Jan 3, 2015 – May 27, 2018

Furthermore, for a comparison purpose, we have downloaded additional data from TripAdvisor.com:

- 1,450 hotels located in Sardinia
- 39,780 reviews (in Italian or English): 879 rated 1/5 stars; 1,205 rated 2/5 stars; 2,987 rated 3/5 stars; 10,169 rated 4/5 and 24,540 rated 5/5 stars
- Period: Feb 10, 2006 – May 7, 2020

3. The framework

Considering that the downloaded raw data is certainly not immediately usable for the analysis, we start with a data cleaning process. We start with some basic filtration of the words to

remove the meaningless ones (i.e. stopwords). Next, we convert emoticons and emoji and we reduce words to their root or base form (i.e., "fishing," "fished," "fisher" are all reduced to the stem "fish").

We use Word Embeddings to reduce the dimensionality of text data.

We recall few fundamentals concepts and terminologies, mostly related to the lexical database WordNet (Miller (1995)), to better understand the next steps:

- Words Embeddings: a vectorial way for representing words. "Each word is associated with a vector in $\mathbb{R}^d$, where the "meaning" of the word with respect to some task is captured in the different dimensions of the vector, as well as in the dimensions of other words." Goldberg (2017)
- Synsets: a collection of words that have a similar meaning. These inbuilt vectors of words are used to find out to which synset belongs a certain word.
- Hypernyms: These are more abstract terms concerning the name of particular synsets. While organizing synsets in a tree-like structure based on their similarity to each other, the hypernyms allow to categorize and group words. In fact, such a structure can be traced all the way up to a root hypernym.
- Lemmas: A lemma is a WordNet's version of an entry in a dictionary: A word in canonical form, with a single meaning. E.g., if someone wanted to look up "mouses" in the dictionary, the canonical form would be "mouse", and there would be separate lemmas for the nouns meaning "animal" and "pc component", etc.
- Words merging by their meaning: we iterate through every word of the received text and, for each word, we fetch the synset which it belongs to. Using the synset name, we fetch the hypernym related to that word. Finally, the hypernym name is used to find the most similar word, replacing the actual word in the text.

Moreover, while using the hypernyms proprieties, we adopt a newspaper pre-trained Words Embeddings produced by Google with Word2Vec SkipGram (Mikolov et al. (2013)) for obtaining the vectorial representation of all the words in the dataset (after the data cleaning process). Finally, to finalize the "merging words by their meaning" step, we use K-Means clustering.

As a result, a λ number of clusters in produced, and the centroid-word is chosen as the word that replaces all the other words present in a cluster. In this way the model is trained using, in place of a general Bag-of-Words, a Bag-of-Centroids (of the clusters produced over the Word Embeddings representation of the dataset).

The value of λ is estimated by cross validation, considering the best accuracy (or others performance metrics) within a labelled dataset (E.g. Booking.com or TripAdvisor data).

Once the data is correctly cleaned and all the words with the same meaning are merged in a single one, it is finally possible to compute the overall sentiment score for each observation.

For this purpose, the Lexical Database SentiWordNet (Esuli and Sebastiani (2006)) allows us to obtain the positive as well as the negative score of a particular word. The sentiment score ($neg\_score - pos\_score$) allows us to determine the polarity of each word. So, the overall score of a specific text (i.e. a comment, a review, a tweet) is defined as the average of all the scores of all the words included in the parsed text.

In that way, with this framework (Fig. 1) we create a temporary sentiment label while using a simple threshold over the so produced overall score. Such a temporary label is the useful base for training the Threshold-based Naïve Bayes Classifier.
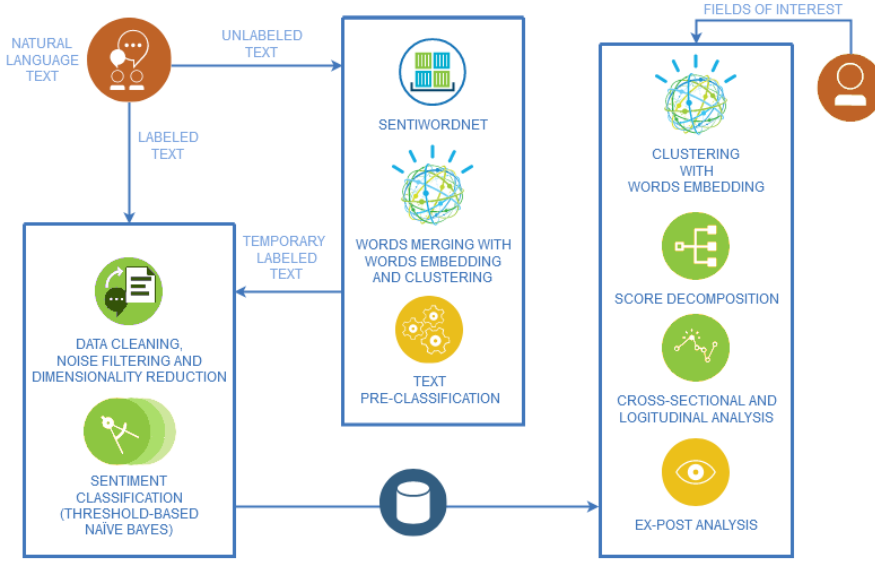
Figure 1: General Sentiment Decomposition framework

4. Threshold-based Naïve Bayes Classifier

Considering a Natural Language text corpora as a set of reviews **r** s.t.:

$$r_i = comment_{pos_i} \cup comment_{neg_i}$$

where $comment_{pos}$ ($comment_{neg}$) are set of words (a.k.a. comments) composed by only positive (negative) sentences, and one of them can be equal to $\emptyset$, the basic features of Threshold-based Naïve Bayes classifier applied to reviews' content are as follows. For a specific review *r* and for each word *w* ($w \in$ *Bag-of-Words*), we consider the log-odds ratio of *w*,

$$
\begin{aligned}
LOR(w) &= \log\left[\frac{P(c_{neg}|w)}{P(c_{pos}|w)}\right] \approx \\
&\approx \log\left[\frac{P(w|c_{neg})}{P(\bar{w}|c_{neg})} \cdot \frac{P(w|c_{pos})}{P(\bar{w}|c_{pos})} \cdot \frac{P(c_{neg})}{P(c_{pos})}\right] = \ldots = \\
&\approx pres_w + abs_w
\end{aligned}
$$

where $c_{pos}(c_{neg})$ are the proportions of observed positive (negative) comments whilst $pres_w$ and abs_w are the log-likelihood ratios of the events ($w \in r$) and ($w \notin r$), respectively.

While calculating those values for all the *w* ($w \in$ *Bag-of-Words*) words, it is possible to obtain an output such that reported in Table 1, where we have c_{pos}, c_{neg}, $pres_w$ and abs_w for each words in the considered *Bag-of-Words*.

	w_1	w_2	w_3	w_4	w_5	...
$P(w_i|c_{neg})$	0.011	0.026	0.002	0.003	0.003	...
$P(w_i|c_{pos})$	0.007	0.075	0.005	0.012	0.001	...
$pres_{w_i}$	0.411	-1.077	-1.006	-1.272	1.423	...
abs_{w_i}	-0.004	0.052	0.003	0.008	-0.002	...

Table 1: Threshold-based Naïve Bayes output

We have then used cross-validation to estimate a parameter τ such that: c is classified as "negative" if $LOR(c) > \tau$ or as "positive" if $LOR(c) \leq \tau$.

While comparing the performances on Table 2 and Table 3, we can then ensure that using the Threshold-based Naïve Bayes Classifier in this framework can definitely lead to more precise predictions.

ME	ACC	TPR	TNR	F1	MCC	BM	MK
0.092	0.908	0.936	0.398	0.951	0.268	0.334	0.215

Table 2: Performance metrics obtained using the temporary sentiment label to predict the "real" label. Notice that to estimate the temporary sentiment label only text data is used, and the "real" label it is not provided in the training phase.

ME	ACC	TPR	TNR	F1	MCC	BM	MK
0.055	**0.945**	**0.973**	**0.503**	**0.973**	**0.475**	**0.476**	**0.474**

Table 3: Performance metrics obtained with Threshold-based Naïve Bayes and 10-fold CV while predicting the real label – trained with the temporary sentiment label

5. Conclusions

Compared to other kinds of approaches, the log-odds values obtained from the Threshold-based Naïve Bayes estimates are able to effectively classify new instances. Those values have also a "versatile nature", in fact they allows to produce plots like in Fig. 2a and Fig. 2b, where customer satisfaction about different dimensions of the hotel service is observed in time.
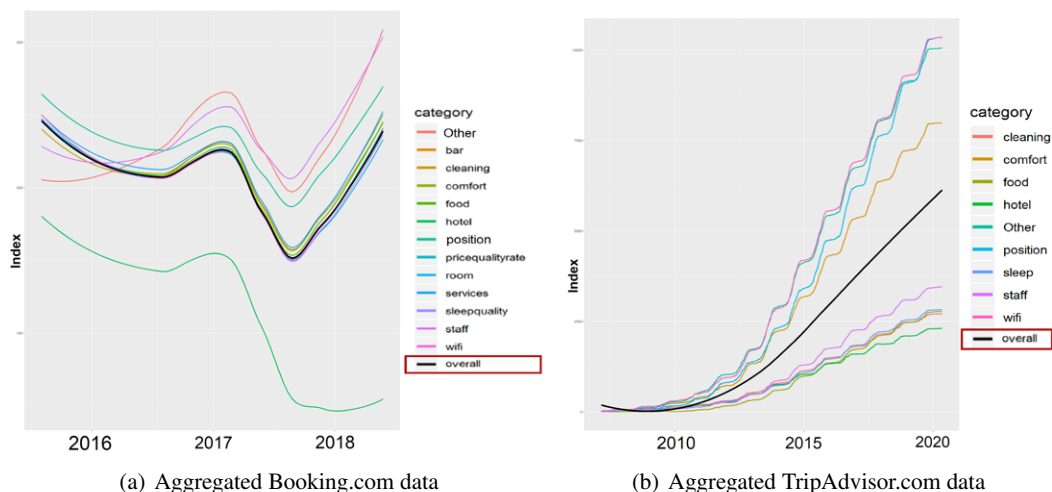

(a) Aggregated Booking.com data   (b) Aggregated TripAdvisor.com data

Figure 2: Category scores observed in time (overall sentiment in black).

References

Conversano, C., Romano, M., Mola, F. (2019). Hotel search engine architecture based on online reviews' content, in *Smart Statistics for Smart Applications. Book of Short Papers SIS2019*, eds. G. Arbia, S. Peluso, A. Pini, and G. Rivellini, Pearson, Milan, (IT), pp. 213–218.

Esuli, A., Sebastiani, F. (2006). SENTIWORDNET: A publicly available lexical resource for opinion mining, in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, eds. N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijkn, and D. Tapias, European Language Resources Association (ELRA), Genoa, (IT), pp. 417–422.

Goldberg, Y. (2017). Neural Network Methods in Natural Language Processing. *Synthesis Lectures on Human Language Technologies*, **10**(1), pp. 1–309.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality, in *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, eds. C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, Curran Associates Inc., Lake Tahoe, Nevada, (USA), pp. 3111–3119.

Miller, G.A. (1995). Wordnet: A lexical database for English. *Communications of the ACM*, **38**(11), pp. 39–41.

Romano, M., Frigau, L., Contu, G., Mola, F., Conversano, C. (2018). Customer Satisfaction from Booking, in *Selected papers Conferenza GARR_18 Data (R)evolution*, eds. M. Mieli, and C. Volpe, Associazione Consortium GARR, Cagliari, (IT), pp. 111–118.

Exploring the intention to walk: a study on undergraduate students using item response theory and theory of planned behaviour

Carla Galluccio, Rosa Fabbricatore, Daniela Caso

1. Introduction

Physical activity is one of the most basic human functions, and it is an important foundation of health throughout life. Physical activity apports benefit on both physical and mental health, reducing the risk of several diseases and lowering stress reactions, anxiety and depression (Penedo, Dahn, 2005). More specifically, physical activity is defined as "any bodily movement produced by skeletal muscles that require energy expenditure" (WHO, 2018), including in this definition several activities. Among them, walking has been shown to improve physical and mental well-being in every age group.

In this regard, the World Health Organization has suggested taking a goal of about $10,000$ steps per day. However, achieving this goal may be difficult for many. For this reason, Tudor-Locke and Bassett (2004) proposed to lower the threshold at least $7,000$ steps a day. Despite that, insufficient walking among university students has been increasingly reported (Sun et al., 2015), requiring walking promotion intervention (e.g. Caso et al., 2020). In order to do this, dividing students based on their intention to walk might be useful, since intention is considered the best predictor of behaviour. In this regard, the main theoretical framework used to explain physical activity is the Theory of Planned Behaviour (TPB; Ajzen, 1991).

In this theory, behavioural intention is determined by three factors. The first predictor of intention is the attitude toward behaviour (both affective and instrumental; see Lowe, Eves, Carroll, 2002 for details), that is the evaluation of the behaviour as favourable or unfavourable. The second factor are subjective norms, which refer to individual's beliefs about whether an important person or group of people approved or not the behaviour. Finally, the third antecedent of intention is the perceived behavioural control (PBC), which can be defined as the individual's perception of the easiness or difficulty of performing the behaviour (Ajzen, 1991).

Herein, we decided to extend the traditional TPB model adding two additional variables as walking intention's predictors, namely self-identity and risk perception. The former is defined as salient and prominent aspects of one's self-perception, whereas the latter refers to the subjective judgement about the severity of a risk. In this regard, some studies have shown that self-identity emerged as a significant predictor of intention to walk in different population (e.g. Ries et al., 2012). Besides, past research (e.g. Stephan et al., 2011) has also shown that risk perception could affect physical activity motivation and behaviour. For these reasons, may be reasonable to suppose that these predictors could be significant also for university students.

In this work, we investigated the university students' intention to walk by exploiting Item Response Theory (IRT) models (Bartolucci, Bacci, Gnaldi, 2015). In particular, we inspected the predictors of intention by means of Rating Scale Graded Response Model (RS-GRM; Muraki, 1990). Afterwards, we used the Latent Class RS-GRM (Bacci, Bartolucci, Gnaldi, 2014) to divide students according to their intention to walk, including predictors' scores as covariates.

Carla Galluccio, University of Florence, Italy, carla.galluccio@unifi.it, 0000-0003-1154-3601
Rosa Fabbricatore, University of Naples Federico II, Italy, rosa.fabbricatore@unina.it, 0000-0002-4056-4375
Daniela Caso, University of Naples Federico II, Italy, daniela.caso@unina.it, 0000-0002-6579-963X

2. Participants and procedure

Data was collected administrating an online self-report questionnaire to undergraduate students enrolled in the Psychology course at Federico II University of Naples. The final sample included $N = 146$ students.

Regarding the questionnaire, for the traditional TPB variables we adapted the scale proposed by Ajzen (2002): *intention* was assessed by 3 items (e.g. "I intend to walk $7,000$ steps a day"); *subjective norms* were assessed by 5 items (e.g. "Most people who are important to me think that I should do $7,000$ steps a day"); *PBC* was assessed by 4 items (e.g. "Doing $7,000$ steps a day is under my control"). For these variables we used a 7-point Likert response scale (1 = strongly disagree to 7 = strongly agree). About *attitude*, it was assessed by 8 items on a semantic differential scale, with 4 items for both instrumental and affective attitude (e.g. "disadvantageous-advantageous" and "unpleasant-pleasant", respectively). On the other hand, we assessed *self-identity* using 4 items (1 = strongly disagree to 7 = strongly agree response scale), e.g. "I think of myself as a physically active subject" (Fishbein, Ajzen, 2010). Finally, *risk perception* was assessed by 6 items (1 = not at all to 7 = very much response scale), e.g. "I think I am personally exposed to the risk of heart disease" (Petrillo, Caso, Donizzetti, 2004).

3. Statistical analysis

IRT model for ordinal polytomous items was carried out for measuring all the TPB variables. In particular, the analysis made up of two steps. Firstly, we estimated the predictors of intention exploiting the RS-GRM as the best model selected among several others according to the BIC index (Schwarz, 1978). For the attitude variable we carried out a bi-dimensional RS-GRM since attitude consists of both instrumental and affective dimensions, whereas for the other variables we used a uni-dimensional RS-GRM. In the second step of our analysis we divided students according to their intention to walk by using a Latent Class RS-GRM. We considered the TPB predictors of intention to walk as individual covariates by using the scores obtained in the first step of the analysis. The analyses were computed using R statistical software.

Let Y_{ij} the response of individual i (with latent trait θ_i) to a polythomous item j with l_j response categories indexed from 0 to $l_j - 1$, the formulation of the GRM (Samejima, 2016) can be expressed as:

$$\mathrm{g}_x[P(Y_{ij} = x | \theta_i)] = \log \frac{P(Y_{ij} \geq x | \theta_i)}{P(Y_{ij} < x | \theta_i)} = \gamma_j(\theta_i - \beta_{jx}), \quad j = 1, \ldots, r, \ x = 1, \ldots, l_j - 1, \ (1)$$

where $\mathrm{g}_x(\cdot)$ is the global logit link function. The item parameters γ_j and β_{jx} represent the *discrimination* and the item-step *difficulty* parameter, respectively. It is worth noting that in this context a useful tool to evaluate the goodness of an item or a test as a whole is the Fisher information (Bartolucci, Bacci, Gnaldi, 2015).

A multidimensional extension of IRT models has been proposed to taking into account the correlation between multiple latent traits (Reckase, 2009). Therefore, each subject i is described by a vector of latent variables $\theta_i = (\theta_{i1}, \ldots, \theta_{iD})$, where D indicates the number of dimensions in the model. According to the between-item multidimensional approach, each item measures only one latent trait. In particular, for the GRM we have:

$$\log \frac{P(Y_{ij} \geq x | \theta_i)}{P(Y_{ij} < x | \theta_i)} = \gamma_j(\sum_{d=1}^{D} \delta_{jd}\theta_{id} - \beta_{jx}) \tag{2}$$

where δ_{jd} is a dummy variable indicating if the item j measures the latent trait d ($\delta_{jd} = 1$) or not ($\delta_{jd} = 0$), with $d = 1, \ldots, D$.

In this vein, the RS-GRM, adopted in the first step of our analysis, represents a constrained version of the GRM in which β_{jx} is expressed in an additive way, namely $\beta_{jx} = \beta_j + \tau_x$. According to this formulation, items may have different general difficulty level (β_j), but equal response category difficulty level (τ_x).

In the second step of the analysis we exploited a Latent Class IRT model, a semi-parametric extension of the IRT model allows to detecting sub-populations of individuals that are homogeneous with respect to the latent trait. The latter is represented through a discrete distribution with $\xi_1, \ldots, \xi_k$ support points defining k latent classes with weights $\pi_1, \ldots, \pi_k$. It is worth noting that $\pi_c = P(\Theta = \xi_c)$ represents the prior probability of belonging to the latent class c ($c = 1, \ldots, k$) with $\sum_{c=1}^{k} \pi_c = 1$ and $\pi_c \geq 0$. The discreteness of the latent trait leads to express the manifest distribution of the response vector $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ir})'$ as:

$$P(\mathbf{Y}_i) = \sum_{c=1}^{k} P(\mathbf{Y}_i | \xi_c) \pi_c \tag{3}$$

where $P(\mathbf{Y}_i | \xi_c) = \prod_{j=1}^{r} P(Y_{ij} = x | \xi_c)$ due to the *local independence* assumption.

In particular, in this work we refers to the RS-GRM parameterisation, selected again as the best model by the BIC, so that:

$$g_x[P(Y_{ij} = x | \xi_c)] = \log \frac{P(Y_{ij} \geq x | \xi_c)}{P(Y_{ij} < x | \xi_c)} = \gamma_j [\theta_i - (\beta_j + \tau_x)]. \tag{4}$$

When a vector of individual covariates $\mathbf{Z}_i$ is considered, as in our analysis, the weight π_c is replaced with the individual weight $\pi_{ci} = P(\Theta = \xi_c | \mathbf{Z}_i = z_i)$. About that, according to the global logit formulation, possible only when latent classes are ordered with respect to the latent trait, we have:

$$\log \frac{\pi_{ci} + \pi_{(c+1)i} + \ldots + \pi_{ki}}{\pi_{1i} + \pi_{2i} + \ldots + \pi_{(c-1)i}} = \beta_{0c} + \mathbf{z}_i' \beta_1, \tag{5}$$

where β_{0c} is the class-specific constant term and β_1 is the vector of regression coefficients describing the effect of individual covariates (Dayton, Macready, 1988).

The estimation of the model parameters is obtained using the *Maximum Marginal Likelihood* (MML) approach (see Bartolucci, Bacci, Gnaldi, 2015 for details). The number of latent classes k was chosen by comparing the fit of models using different values of k.

4. Results

The latent trait analysis in the first step pointed to a good test Fisher information for all the predictors of the intention to walk we considered (see Figures 1 and 2). In particular, items measuring PBC, self-identity and attitude are maximally informative for students with low levels of the latent trait; whereas the test information curve for the risk perception is shifted on the right (greater information for high levels of the latent trait).

Regarding the Latent Class IRT model, the BIC indicated the RS-GRM with $k = 4$ number of classes as the best model. The standardised support points and the average of the individual weights π_{ci} are reported in Table 1. Looking at support points, we notice that latent classes are increasing ordered according to the levels of intention to walk $7,000$ steps a day. On the other hand, the average weights indicated that Class 3 is the largest one, followed by Class 2. Thus, the majority of the students reported a medium level of intention to walk $7,000$ steps a day.

Besides, in Table 2 we reported the TPB predictors that significantly affect the class weights. To estimate this effect, we adopted the global logit specification (see Equation 5) since the
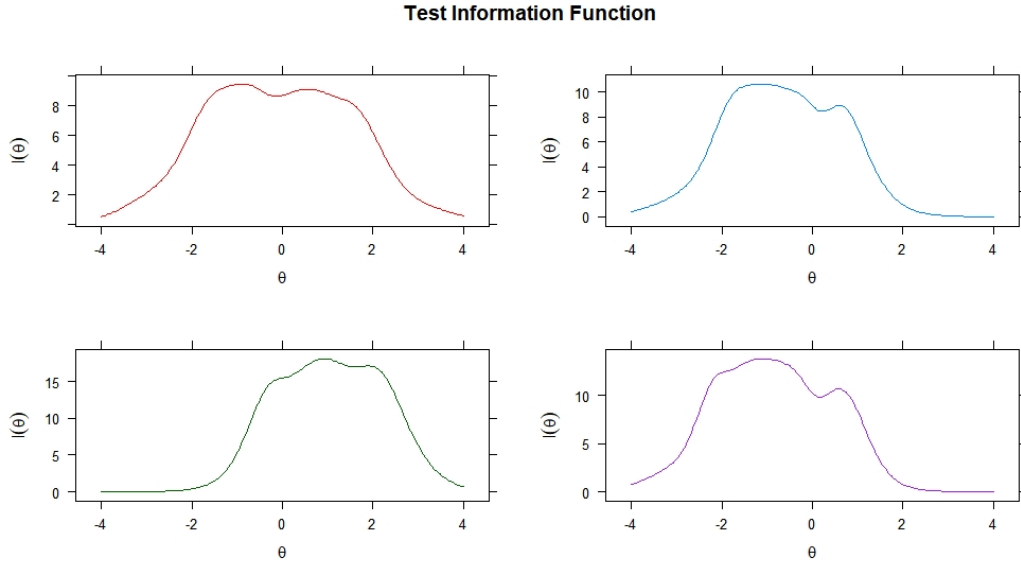
155

Test Information Function

Figure 1: Test Fisher information curve for the predictors of the intention to walk: subjective norms (red line), PBC (blue line), risk perception (green line), and self-identity (purple line).

Table 1: Standardised support points and average class weights for the Latent Class RS-GRM with covariates.

	Latent class			
	1	**2**	**3**	**4**
Intention to walk 7, 000 steps a day	-1.79	-0.62	0.36	1.65
Average weights	*0.13*	*0.26*	*0.46*	*0.15*

Table 2: Regression coefficient ($\hat{\beta}_1$), standard error ($\hat{se}$), t-value, and p-value for the individual covariates.

Covariate	$\hat{\beta}_1$	$\hat{se}$	t-value	p-value
Affective Attitude	0.97	0.21	4.62	0.00
Subjective Norms	0.38	0.19	2.04	0.04
Self-identity	0.42	0.18	2.31	0.02

support points were increasingly ordered. We removed from the final model all the variables resulted not significant for $\alpha = 0.10$, namely instrumental attitude, PBC, and risk perception.

We can conclude that the most significant covariate affecting positively the student's intention to walk 7, 000 steps a day is affective attitude ($\hat{\beta}_1 = 0.97$, p-value < 0.01), followed by self-identity ($\hat{\beta}_1 = 0.42$, p-value < 0.05) and subjective norms ($\hat{\beta}_1 = 0.38$, p-value < 0.05).

5. Discussion and conclusion

The present study aimed to detect homogeneous groups of university students according to their intention to walk exploiting IRT models. We found that students could be divided into four ordered classes: Class 1 is made up of students with the lowest intention to walk, whereas Class 4 includes students with the highest intention to walk 7, 000 steps a day. Besides,
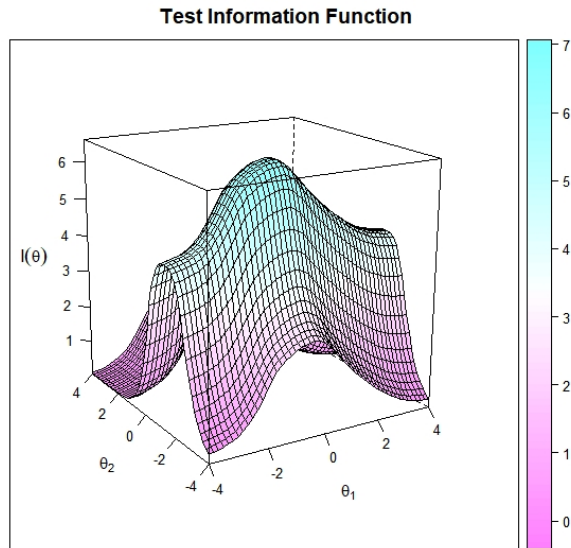
Figure 2: Test Fisher information curve for the attitude variable: θ_1 refers to the affective dimension, whereas θ_2 to the instrumental one.

results showed that the best predictors of intention to walk were affective attitude, subjective norms and self-identity. In contrast, instrumental attitude, risk perception and PBC were not significant. Regarding affective and instrumental attitudes, several studies on health behaviours have shown that affective attitude was a strong predictor of intention, often at the expense of instrumental attitude (e.g. Lowe, Eves, Carroll, 2002). Usually, health promotion programmes emphasised the instrumental benefits of physical activity, such as improved health, which are not immediately apparent to the individual due to the delay between doing physical activity and its results. Conversely, affective components of physical activity, such as its pleasant nature, are immediate consequences of involvement. Concerning subjective norms, results showed a moderate and positive influence on students' intention to walk. This finding is consistent with those in literature (Wing Kwan, Bray, Martin Ginis, 2009), where it is supposed that social influences on physical activity intention were stronger among younger populations. Besides, as we expected, self-identity resulted as a significant predictor of intention to walk in university students. In fact, according to the literature (e.g. Ries et al., 2012), our results are consistent with an interpretation that who identify themselves as physically active persons are more likely to practise regular physical activity. Finally, regarding risk perception and PBC, we found that in our model these variables are not significant. About risk perception, it is reasonable to suppose that the perception of the riskiness correlated with the physical inactivity, such as physical and mental diseases, is more likely in older than younger populations. On the contrary, the finding that PBC is not a significant predictor of intention was quite surprising. We speculate that university commitment leads students to not fully perceived the extent of their control on other activities, such as physical activity, especially during the first year.

In conclusion, we believe that the Latent Class IRT models represent a useful statistical tool for dividing students according to their intention to walk in order to define a more tailored walking promotion programmes. Indeed, we could support students in Class 4 in maintaining their intention to walk, whereas a different walking promotion intervention could be implemented for students in Class 1, focusing on the TPB variables that resulted as significant predictors.

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, **50**(2), pp. 179–211.

Ajzen, I. (2002). Constructing a TPB questionnaire: Conceptual and methodological considerations.

Bacci, S., Bartolucci, F., Gnaldi, M. (2014). A class of Multidimensional Latent Class IRT models for ordinal polytomous item responses. *Communications in Statistics-Theory and Methods*, **43**(4), pp. 787–800.

Bartolucci, F., Bacci, S., Gnaldi, M. (2015). *Statistical analysis of questionnaires: A unified approach based on R and Stata*. CRC Press, Boca Raton, (FL).

Caso, D., Carfora, V., Capasso, M., Oliano, D., Conner, M. (2020). Using Messages Targeting Psychological versus Physical Health Benefits to Promote Walking Behaviour: A Randomised Controlled Trial. *Applied Psychology: Health and Well-Being*.

Dayton, C.M., Macready, G.B. (1988). Concomitant-variable latent-class models. *Journal of the american statistical association*, **83**(401), pp. 173–178.

Fishbein, M., Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. Taylor & Francis, New York, (NY).

Lowe, R., Eves, F., Carroll, D. (2002). The influence of affective and instrumental beliefs on exercise intentions and behavior: A longitudinal analysis. *Journal of Applied Social Psychology*, **32**(6), pp. 1241–1252.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, **14**(1), pp. 59–71.

Penedo, F.J., Dahn, J.R. (2005). Exercise and well-being: a review of mental and physical health benefits associated with physical activity. *Current opinion in psychiatry*, **18**(2), pp. 189–193.

Petrillo, G., Caso, D., Donizzetti, A.R. (2004). Questionario "Salute e benessere nel mondo giovanile", in *Promuovere la salute nei contesti educativi: comportamenti salutari e benessere tra gli adolescenti*, eds. G. Petrillo, D. Caso, Franco Angeli, Milano, (IT), pp. 401–407.

Reckase, M. (2009). *Multidimensional Item Response Theory*. Springer, New York, (NY).

Ries, F., Hein, V., Pihu, M., Armenta, J.M.S. (2012). Self-identity as a component of the theory of planned behaviour in predicting physical activity. *European Physical Education Review*, **18**(3), pp. 322–334.

Samejima, F. (2016). Graded response model, in *Handbook of Item Response Theory*, ed. W.J. van der Linden, Springer, New York, (NY), pp. 95–108.

Stephan, Y., Boiche, J., Trouilloud, D., Deroche, T., Sarrazin, P. (2011). The relation between risk perceptions and physical activity among older adults: a prospective study. *Psychology & health*, **26**(7), pp. 887–897.

Sun, G., Acheampong, R.A., Lin, H., Pun, V.C. (2015). Understanding walking behavior among university students using theory of planned behavior. *International journal of environmental research and public health*, **12**(11), pp. 13794–13806.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**(2), pp. 461–464.

Tudor-Locke, C., Bassett, D.R. (2004). How many steps/day are enough?. *Sports medicine*, **34**(1), pp. 1–8.

World Health Organization. (2018). Physical activity. `https://www.who.int/news-room/fact-sheets/detail/physical-activity`.

Wing Kwan, M.Y., Bray, S.R., Martin Ginis, K.A. (2009). Predicting physical activity of first-year university students: An application of the theory of planned behavior. *Journal of American College Health*, **58**(1), pp. 45–55.

Determinants of spatial intensity of stop locations on cruise passengers tracking data

Nicoletta D'Angelo, Mauro Ferrante, Antonino Abbruzzo, Giada Adelfio

1. Introduction

A tourism destination can be seen as a mix of tourist attractions and of tourist supporting elements, such as accommodation, transport and tourist-related services, which make it attractive and accessible and, in turn, determine its value. Various authors have highlighted the importance of managing key locations and of understanding tourist spatial behaviour and its main determinants (Cooper, 1981; Liu et al., 2017; Russo, 2002). Tourist services characteristics and the spatial distributions of attractions represent supply-side factors which have an influence on tourists' spatial behaviour (Zheng et al., 2017). It is acknowledged that the spatial movements of tourists in a destination are also influenced by demand-side factors, such as time budget, motivations, and destination knowledge, to mention but a few (Lew and McKercher, 2006). Moreover, human interactions may have a role in tourists' spatial behaviour, these may include tourists-residents as well as tourist-tourist interactions.

Despite the importance of understanding tourist movements within a destination, collecting data on tourist mobility is not an easy task (Stopher, 2012). Traditional methods are generally based on post-visit questionnaire or trip diaries, which rely on the accurate recall of the places visited and activities made. Moreover, they may introduce a bias on participant's behaviour, who knows which is being observed (East et al., 2017). Nowadays, GPS technology allows to collect information on human mobility at a very high temporal and spatial detail, with no effort required from the participant in recalling the places visited. Since the influential book of Shoval and Isaacson (2009) many studies in tourism field have been conducted by using GPS technology [see Shoval and Ahas (2016) for a review of the first decade].

This paper expands the knowledge of tourists' spatial behaviour within a destination – considering cruise tourists as a case study – by analyzing their stop location pattern in order to highlight the main determinants of spatial intensity of stops at their destination. To this end, a stochastic point process modelling approach on a linear network is proposed. We refer to Baddeley et al. (2020) for a review of spatial point processes on networks.

In this paper, we fit a Gibbs point process model adapted on the network, that takes into account individual-related variables, contextual-level information, and spatial interaction among stop points. From an applied perspective, this allows to determine the attractiveness of various places in the destinations, as well as the influence of destination-related characteristics and of individual-level variables on stop location pattern. Moreover, the use of Gibbs point process approach allows for the analysis of interactions among points, in order to check whether attraction or repulsive relationships exist among tourists' stop location choice. From a more methodological perspective, while most of the recent literature on this topic is concerned with non-parametric intensity estimation, both in space (Moradi et al., 2019) and space-time (Moradi and Mateu, 2020; Mateu et al., 2019), our approach contributes to the framework of point processes on networks by proposing a parametric model.

Nicoletta D'Angelo, University of Palermo, Italy, nicoletta.dangelo@unipa.it, 0000-0002-8878-5986
Mauro Ferrante, University of Palermo, Italy, mauro.ferrante@unipa.it, 0000-0003-1287-5851
Antonino Abbruzzo, University of Palermo, Italy, antonino.abbruzzo@unipa.it, 0000-0003-2196-3570
Giada Adelfio University of Palermo Italy, giada.adelfio@unipa.it 0000-0002-3194-4296

2. Data

The cruise tourism segment was selected for the analysis in consideration of the single exit/entry point and the relatively brief visiting time, which characterize cruise passengers' experience at their destination. These features make the use of GPS technology particularly suitable for the analysis of such a relevant phenomenon (Shoval, 2008). Data have been collected in Spring 2014 in the city of Palermo through an integration of questionnaire-based survey and GPS technology [see Ferrante et al. (2018) for details on data collection procedures]. For the purposes of the present study, due to computational reasons, only two days of survey have been considered, referred to cruise passengers visiting the City after disembarking from the cruise ship. After pre-processing of GPS tracking data, stop locations were derived through the implementation of the *dbscan* algorithm on individual trajectories, according to the procedure described in Abbruzzo et al. (2020).

The final spatial point pattern considered consists of 429 stops made by 58 visitors, stopping 7 times on average during their visit in the downtown of Palermo city on the 27th and 28th April 2014. In order to properly account for the constrained structure of the space support, the road network of selected area was considered, providing a linear network L with 4473 vertices and 5399 lines. Other information have been derived both from destination-related characteristics, questionnaire-based survey, whereas synthetic information on cruise passengers' spatial mobility at the destination have been derived from individual trajectories. As for destination-related characteristics, beyond the geographical configuration of the destination, determined by the road network, also the shortest-path distance of each stop location from the nearest tourist attraction was computed. In Figure 1, the locations of stop locations are displayed in red, along with the main attractions considered, displayed in green. Among socio-demographic characteristics, according to the literature on tourist mobility, age, education level, and income are supposed to be the main potential determinants of the spatial studied phenomenon. In addition, synthetic information derived from individual trajectories includes: total length of tour, total duration of the visit, maximum distance from the port location, and average speed.

Figure 1: In red: the spatial point pattern. In green: the location of the touristic attractions.

3. Model proposal

We here introduce a novel modelling approach for describing the spatial behaviour of the visitors. In detail, we fit a parametric model to the visitors' stops accounting for both the underlying network and the individual tourists' choices by introducing a random subject-specific effect. At this aim we refer to the Gibbs point process models with mixed effects (Illian and Hendrichsen, 2010), conforming the procedure to the linear networks context. Let M be the number of visitors on a linear network L, each generating the point patterns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M$ that can be thought as the individual pattern of stops. This flexible procedure allows to account for the individual information both by suitable random and fixed factors, and by external covariates. We therefore assume, for each $\boldsymbol{x}_m$ with $m = 1, \ldots, M$, a pairwise interaction process (Van Lieshout, 2000) with conditional intensity (Kallenberg, 1984) given by:

$$\lambda_{\theta, \phi_m}(u; \boldsymbol{x}_m) = b_{\theta, \phi_m}(u) \prod_{i=1, x_{mi} \neq u}^{n(\boldsymbol{x}_m)} h_{\theta, \phi_m}(u, x_{mi})$$

where $n(\boldsymbol{x}_m)$ is the number of points in $\boldsymbol{x}_m$, that is, the number of stops per visitor, $b_{\theta, \phi_m}(u)$ and $h_{\theta, \phi_m}(u, v)$ are two functions that model the intensity and the interaction, respectively. For estimation purposes, the Berman-Turner device for maximum pseudolikelihood is considered. The final quadrature scheme used for model fitting consists of the analysed 429 data points, representing the visitors' stops, and of 10798 dummy points, obtained generating the quadrature scheme on the analysed network. This leads to a dataset of 651166 quadrature points, that is equal to the number of data points plus the number of dummy points, all replicated for the number of marks M. In this paper, we fit the proposed model to these new quadrature points, in order to enable the inclusion of random effects and subject-specific covariates. We denote by u_{im} the location of the new set of points.

As for the intensity function $b_{\theta, \phi_m}(u)$, we set $B_1(u_{im}) = \mathbf{1}$, with $\mathbf{1}$ the identity function and $B_3(u_{im})$ is the *distance from the nearest attraction* (see Figure 1). In addition, $B_2(u_{im})$ denotes the `ID` of the tourist, included as a random effect. $B_4(u_{im})$ is a non-parametric function for $u_{im} \in L$, estimated through thin plate regression splines with a chosen number of 29 knots for our analysis. Therefore, for the intensity function we have:

$$b_{\theta, \phi_m}(u_{im}) = \exp(\theta_1 + \phi_{1m} B_2(u_{im}) + \theta_3 B_3(u_{im}) + B_4(u_{im})).$$

To describe the interaction function $h_{\theta, \phi_m}(u, v)$, we propose a smooth interaction function $H(\cdot, \cdot)$ which is assumed dependent only from the shortest-path distance between any pairs of points, i.e. the length of the shortest path between the location of the two points on the network. For two points occurring on the network, with location u and v, we define:

$$H_k(u, v) = \begin{cases} \left(1 - \left(\frac{d(u,v)}{R}\right)^2\right)^2 & \text{if} \quad 0 < d(u, v) \leq R \\ 0 & \text{else} \end{cases} \tag{1}$$

where $d(u, v)$ is computed as the shortest-path distance, and $R \geq 0$ defines the radius of interaction. Therefore, for the interaction function we have:

$$h_{\theta, \phi_m}(u_{im}, v_{im}) = \exp(\theta_2 H(u_{im}, v_{im}) + \phi_{2m} H(u_{im}, v_{im})).$$

In this application, the interaction radius is set to $R = 100$ meters, as a reasonable threshold of distance up to which we assume that there may be interaction among visitors' stop location choice.

In order to explain the spatial inhomogeneity and to consider the characteristics of the visit, socio-economic characteristics and synthetic information on the itinerary undertaken are included as covariates. These are:

- income: *yearly income*, dichotomized in <40000 and $\geq~40000$ euro;
- education: *education level*, dichotomized in low (High school diploma or Bachelor degree) and high (Master or Ph.d.);
- visit: *independent visit*, indicating whether the visitor is travelling independently (yes) or by an organized visit (no);
- dist: *maximum distance from the port*, dichotomized in >3.5 and ≤ 3.5 km.

Thus, we propose to model the spatial intensity as:

$$\log \hat{\lambda}_{\theta,\phi_m}(u_{im}) = \hat{\theta}_1 + \hat{\phi}_{1m}B_2(u_{im}) + \hat{\theta}_2 v_{im} + \hat{\theta}_3 Z(u_{im}) + B_4(u_{im}) + \hat{\phi}_{2m}v_{im}$$
$$+ \hat{\theta}_4 \text{income} + \hat{\theta}_5 \text{education} + \hat{\theta}_6 \text{visit} + \hat{\theta}_7 \text{dist} \tag{2}$$

where: $v_{im} = \sum_{j=1}^{n(\boldsymbol{x}_m)} H(u_{im}, x_{jm})$; θ_2 is the fixed effect of the smooth function in (1); θ_3 is the fixed effect of the *distance from the nearest attraction*; ϕ_{1m} is the random effect of the ID; and ϕ_{2m} represents the random effects for the interaction smooth function.

4. Results

In Table 1 the estimates of the fixed effects and the summary of the random effects of the final selected model are reported.

Table 1: Model coefficients and approximate significance of smooth terms of the Gibbs model

	Estimate	Std. Error	z value	Pr($>$ $\|z\|$)	
$\hat{\theta}_1$ Intercept	-10.808	0.519	-20.834	0.000	***
$\hat{\theta}_2$ Interaction	0.152	0.010	15.305	0.000	***
$\hat{\theta}_3$ Distance from the nearest attraction	-0.005	0.001	-3.369	0.001	***
$\hat{\theta}_4$ Income (<40000)	0.412	0.148	2.775	0.006	**
$\hat{\theta}_5$ Education (low)	0.296	0.144	2.057	0.040	*
$\hat{\theta}_6$ Independent visit (yes)	0.796	0.266	2.993	0.002	**
$\hat{\theta}_7$ Max distance from the port (> 3.5 km)	0.627	0.190	3.304	0.001	***
	edf	Ref.df	Chi.sq	p-value	
s(lat,long)	21.036	24.91	119.22	0.000	***
$\hat{\phi}_{1m}$ s(id)	0.000	44.00	0.00	0.000	***
$\hat{\phi}_{2m}$ s(v,id)	19.269	45.00	40.16	0.000	***

When $\exp(\hat{\theta}_1)$ is multiplied by the length of the network, the estimated stops for each individual are 2.4, lower than the original average stops. This is likely due to the sparsity of the original points in certain regions of the network. Regarding the fixed part of the model, among socio-demographic characteristics, cruise passengers with higher level of education and higher income tend to stop more. This is in line with expectations, by considering both a more detailed enjoyment of cultural attractions for people with a higher education level, and a potential association of stops with spending activities, such as purchasing of food and beverage, visit to museums, etc. Also being and independent cruise passengers increases the stop intensity, compared to organized cruise passengers. This is likely due to the fixed scheduling of activities of the organized tour. Still, maximum distance from the port has been considered as a *proxy* of the degree of exploration of the destination (Jaakson, 2004), and it resulted positively associated also with stop intensity. The positive interaction parameter $\exp(\hat{\theta}_2) = 1.164$ indicates that

overall the visitors' stops attract each other. Therefore, visitors tend to stop in the same spots. Furthermore, $\exp(\hat{\theta}_3) = 0.995$ indicates that moving away from any tourist attraction slightly decreases the probability of visitor stopping. From the significant random effects, we notice that not only the intensity varies among visitors ($\hat{\phi}_{1m}$), but also the interaction ($\hat{\phi}_{2m}$). This opens new research perspectives on the modeling of human behaviour, and on the application of ecological theories (Meekan et al., 2017). Finally, the inclusion of the smooth term $B_4(u_{im})$ accounting for the spatial coordinates improves significantly the fitting of the model.

In order to make the estimator unbiased, that is, given the expected number of points $\mathbb{E}[\int_L \lambda(u)d(u)] = n$, the intensity obtained by (2) has been normalized $\hat{\lambda}(u) = \frac{n\hat{\lambda}(u)}{\int_L \hat{\lambda}(u)d(u)}$. Therefore, in Figure 2 the estimated intensity is shown, displaying the expected number of stops for each location. We report only those estimated intensities higher than the 99th percentile, to facilitate reading and to highlight the regions where visitors are most likely to stop.
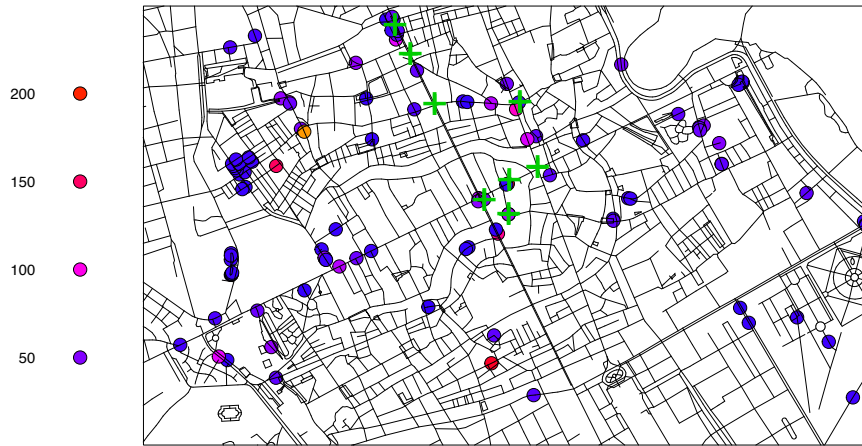

Figure 2: Estimated pointwise intensities above the 99th percentile: the lighter the colour the higher the intensity. The intensity has been normalized in order to obtain the expected number of stops for each location. Location of the tourist attractions are displayed in green.

5. Conclusion

In this paper, we have proposed a novel model to analyze the main determinants of spatial intensity of cruise passengers' stop locations during their visit. The proposed model allows taking into account the linear network determined by the street configuration of the destination under analysis. The results show an influence of both socio-demographic and trip-related characteristics on the stop location patterns, as well as the relevance of distance from the main attractions, and potential interactions among cruise passengers in stop configuration. The proposed approach represents an improvement both from the methodological perspective, related to the modelling of spatial point process on a linear network, and from the applied perspective, given that a better knowledge of the determinants of spatial intensity of visitors' stop locations in urban contexts may orient destination management policy. A limit of the present study is not accounting for the temporal component. Also, the analysis is here focused in a restricted area of the destination. Considering a wider study area would allow to better account for covariates related to the individuals trajectories. Indeed, the total length of the tour, as well as the duration of the visit, represent useful information that could influence visitor's stop location choice.

References

Abbruzzo, A., Ferrante, M., & De Cantis, S. (2020). A pre-processing and network analysis of GPS tracking data. *Spatial Economic Analysis*, pp. 1–24, doi: 10.1080/17421772.2020.1769170.

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., & Davies, T.M. (2020). Analysing point patterns on networks – a review. *Spatial Statistics*, 100435.

Cooper, C. P. (1981). Spatial and temporal patterns of tourist behaviour. *Regional Studies*, **15**(5), pp. 359–371.

East, D., Osborne, P., Kemp, S., & Woodfine, T. (2017). Combining GPS & survey data improves understanding of visitor behaviour. *Tourism Management*, **61**, pp. 307–320.

Ferrante, M., De Cantis, S., & Shoval, N. (2018). A general framework for collecting and analysing the tracking data of cruise passengers at the destination. *Current Issues in Tourism*, **21**(12), pp. 1426–1451.

Illian, J.B., & Hendrichsen, D. K. (2010). Gibbs point process models with mixed effects. *Environmetrics*, **21**, pp.341–353.

Jaakson, R. (2004). Beyond the tourist bubble?: cruiseship passengers in port. *Annals of tourism research*, **31**(1), pp. 44–60.

Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes. *International Statistical Review/Revue Internationale de Statistique*, pp. 151–164.

Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of tourism research*, **33**(2), pp. 403–423.

Liu, B., Huang, S. S., & Fu, H. (2017). An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tourism Management*, **58**, pp. 132–141.

Mateu, J., Moradi, M., & Cronie, O. (2019). Spatio-temporal point patterns on linear networks: Pseudo-separable intensity estimation. *Spatial Statistics*, 100400

Meekan, M. G., Duarte, C. M., Fernández-Gracia, J., Thums, M., Sequeira, A. M., Harcourt, R., & Eguíluz, V. M. (2017). The ecology of human mobility. *Trends in Ecology Evolution*, **32**(3), pp. 198–210.

Moradi, M.M., Cronie, O., Rubak, E., Lachieze-Rey, R., Mateu, J., & Baddeley, A. (2019). Resample-smoothing of voronoi intensity estimators. *Statistics and computing*, **29**, pp. 995–1010.

Moradi, M.M., & Mateu, J. (2020). First- and second-order characteristics of spatio-temporal point processes on linear networks. *Journal of Computational and Graphical Statistics*, **29**(3), pp. 432–443.

Russo, A. P. (2002). The 'vicious circle' of tourism development in heritage cities. *Annals of tourism research*, **29**(1), pp. 165–182.

Shoval, N. (2008). Tracking technologies and urban analysis. *Cities*, **25**(1), pp. 21–28.

Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: the first decade. *Tourism Geographies*, **18**(5), pp. 587–606.

Shoval, N., & Isaacson, M. (2009). *Tourist mobility and advanced tracking technologies*. Routledge, London.

Stopher, P. (2012). *Collecting, managing, and assessing data using sample surveys*. Cambridge University Press, Cambridge.

Van Lieshout, M. N. M. (2000). *Markov point processes and their applications*. World Scientific

Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place?. *Tourism Management*, **59**, pp. 267–280.

This book includes 25 peer-reviewed short papers submitted to the Scientific Opening Conference titled "Statistics and Information Systems for Policy Evaluation", aimed at promoting new statistical methods and applications for the evaluation of policies and organized by the Association for Applied Statistics (ASA) and the Department of Statistics, Computer Science, Applications DiSIA "G. Parenti" of the University of Florence, jointly with the partners AICQ (Italian Association for Quality Culture), AICQ-CN (Italian Association for Quality Culture North and Centre of Italy), AISS (Italian Academy for Six Sigma), ASSIRM (Italian Association for Marketing, Social and Opinion Research), Comune di Firenze, the SIS – Italian Statistical Society, Regione Toscana and Valmon – Evaluation & Monitoring.

Bruno Bertaccini is Associate Professor of Statistics at the "G. Parenti" Department of Statistics, Computer Science and Applications of the University of Florence. After earning his PhD in Applied Statistics, he became an expert in the effectiveness assessment of public policies, with particular attention to the evaluation of the quality of higher education and academic policies.

Luigi Fabbris is a freelance researcher in the field of statistics and social science; President of the ASA, (Italian) Association for Applied Statistics; formerly Professor of Social Statistics at the University of Padua; he has authored or co-authored more than 400 papers or books.

Alessandra Petrucci is Professor of Social Statistics and she holds a PhD in Applied Statistics. Her research interests include survey sampling methods, spatial statistics, multivariate statistical analysis applied to social and environmental issues. She is the author and co-author of numerous scientific papers published in national and international journals, and she has been a member of research projects and scientific committees for several national and international conferences.