Spread of Covid-19 epidemic in Italy between March 2020 and February 2021: empirical evidence at provincial level

Fabrizio Antolini, Samuele Cesarini, Francesco Giovanni Truglia

1. Introduction

The literature on the determinants of Covid-19 contagion is evidently rather recent and does not always draw generally accepted conclusions in identifying the factors that may explain the differences between territorial areas in the severity of Covid-19 impact (Moosa and Khatatbeh, 2021). The rate of contagion is a phenomenon that depends on many and varied factors that are not easy to interpret and must be analysed considering their spatial component (Cutrini and Salvati, 2021).

To this end, convergence models were used, in which the initial level and growth of observed infections in a certain province were related to the level of infections and the relative growth rate of all other provinces. This model was implemented for all three waves that occurred in Italy from March 2020 to February 2021. The proposed convergence model was constructed by also including environmental (Azuma et al, 2020; Copat et al, 2020) and demographic (Goumenou et al, 2020) factors as controlling elements of a conditional β-convergence (Truglia, 2021).

In the literature, spatial regression models have been widely used in many epidemiological studies (Guo, G. et al., 2020; Liu, X. et al., 2020; Zhao, et al., 2020). To date, however, only a few studies are available that have investigated the close association between sociodemographic and environmental determinants and the spatial convergence of Covid-19 infection incidence. Therefore, this study aims to address the mentioned research gap.

This work further contributes to the study and understanding of the impact of demographic and environmental parameters on the spread of Covid-19 cases by adopting a spatial regression approach.

The work is divided into four sections. The first describes the construction of the panel of data used and their recoding into indicators and indices. The second part circumscribes the spatial approach in the implementation of the conditional β-convergence model to investigate any convergence processes observed in the transmission of contagion between the spatial areal units under study. The third part presents the results obtained. Finally, the fourth part proposes a discussion of the findings and introduces some final considerations and possible implications for future studies.

2. Data

In the following analysis, a balanced panel of data referring to the 107 Italian provinces was used. The data on contagion were retrieved and processed from the Civil Protection repository in the 'data-provinces' section. From these, for each of the 107 Italian provinces, the contagion rates for the three waves and their respective durations and distances (in days) were calculated. The spatial context data were collected from the ISTAT data warehouse and the ISPRA environmental data yearbook.

As for the infection rate, this was measured as the simple ratio of the total number of registered cases of Covid-19 infection at period t - where t represents the first (I), second (II) and third wave (III) respectively - to a standard reference population of 100,000 individuals.

The other indices relating to contagion (duration and distance), calculated for each province,

Fabrizio Antolini, University of Teramo, Italy, fantolini@unite.it, 0000-0002-3112-524X
Samuele Cesarini, University of Teramo, Italy, scesarini@unite.it, 0000-0001-7062-1580
Francesco Giovanni Truglia, ISTAT, Italian National Institute of Statistics, Italy, truglia@istat.it

do not require statistical formalisation, and represent, the first, the number of days elapsed between the beginning of one wave and its end, and the second, the number of days between the end of one wave and the beginning of the next. On the other hand, the indices that are assigned the role of explanatory variables and that will be the controlling factors for the convergence of infection are:

- The old-age index (old_index): defined as the simple ratio of the population aged 65 and over to the total population.
- The average temperature (temp_av): defined as the average annual temperature (expressed in C°).
- The density: defined as the simple ratio between the total population and the land area (expressed in number of inhabitants per square kilometre).
- Pollution (pollution_av): defined as the total daily average observed values of particulate matter pm10 and pm2.5 and nitrogen dioxide (NO2), expressed in µg/m3.

As these control variables have different units of measurement, they are standardised for use in the convergence model.

3. Method

There are various procedures for analysing territorial convergence. In the present study, the most well-known convergence concepts were used to which reference is made in the bibliography (Barro e Sala-I-Martin, 1992; Mankiw, 1992; Arbia, 2005), including β-convergence. In short, in the literature, this approach originates directly from the neoclassical theory of economic growth theorised by Solow-Swan (Solow et Swan, 1956). This type of convergence describes an economic environment in which a poorer country develops faster than a richer country, in terms of per capita income level. Unlike formal models that require a measure of physical and/or human capital, greater freedom is granted by informal models that are not required to be traceable to the variables brought into play by growth accounting (Alexiadis, 2010). The conditional β-convergence model can therefore be rewritten as follows (equation 1):

$$ln(Y_{i,t}/Y_{i,0}) = \beta_o + \beta_1 ln(Y_{i,0}) + Y Z_i + \varepsilon_i \qquad (1)$$

Where,
i, and t denote respectively, the spatial unit and the time reference in which the
observation Y is measured
β_0 is the intercept
Z is the matrix of the *n* control variables that are assumed to influence the growth rate
ε_i is the error term at zero mean and variance σ^2
ln($Y_{i,t}/Y_{i,0}$) is the natural logarithm of the growth rate
ln($Y_{i,0}$) is the natural logarithm of the initial level

The β_1 coefficient, if statistically significant and of negative sign, indicates the existence of the β-convergence hypothesis.

The β-convergence model thus captures whether territorial gaps, in relation to a specific aspect, increase or decrease over a certain time span (in our study the beginning and end of the three successive waves). This research adopts a method that differs from the conventional convergence strategy by instead focusing on the spatial convergence aspect. In fact, an interesting issue to consider in the territorial convergence analysis is the recognised need to introduce elements that consider functional relationships between provinces. For these reasons, it is therefore appropriate to make use of specific procedures capable of considering the structure of

connections between the units of analysis (Guliyev, 2020). Translated into other terms, the β-convergence model can be transformed in such a way that it considers the spatial proximity of the N observations by means of a proximity matrix W consisting of elements w_{ij} that take on value 1 or 0, respectively in the case that units i and j are contiguous or non-contiguous.

The spatial methods that can be constructed from this common basis are many and varied depending on the spatial effects to be investigated. Below we propose the conditional β-convergence model (in matrix form) in the case of spatial autoregressive lag of the dependent variable (SAR) (equation 2).

$$y = \rho Wy + \beta X + YZ + \varepsilon \tag{2}$$

Where,

y is the matrix containing the natural logarithm of the growth rate at time *t* and province *i*

X is the matrix containing the natural logarithm of the initial level

Z is the matrix of the *n* control variables that are assumed to influence the growth rate

ρ (Rho) denotes the spatial autoregressive coefficient

W represents the contiguity matrix of the provinces

β and Y are the coefficients to be estimated

ε is the error term with zero mean and variance σ^2.

It was decided to use a W contiguity matrix of the queen contiguity type. In this typology, provinces that share at least one side or vertex are considered contiguous (LeSage, 1998).

4. Results

Table 1 show the results obtained through the estimation of the spatial autoregressive SAR model implemented for the conditional β-convergence model.

Table 1. Results conditional β-convergence (SAR): (a) first wave; (b) second wave; (c) third wave

		Estimate	Std. Error	z value	Pr (>|z|)
(a)	B_0	-3.747	0.414	-9.048	< 2.2e-16 ***
	β_1	-0.489	0.042	-11.429	< 2.2e-16 ***
	old_index	-0.013	0.047	-0.282	0.777
	temp_av	0.027	0.047	0.585	0.558
	density	0.126	0.049	2.550	0.010 *
	pollution_av	0.212	0.057	3.692	2.22e-04 ***
	gg_I	0.011	0.004	2.347	0.018 *
(b)	B_0	-0.327	0.062	-5.217	1.811e-07***
	β_1	-0.109	0.010	-10.545	< 2.2e-16 ***
	old_index	-0.021	0.010	-2.093	0.036 *
	temp_av	0.002	0.010	0.252	0.800
	density	0.041	0.011	3.713	2e-04 ***
	pollution_av	0.021	0.013	1.649	0.099 .
	gg_II	0.001	0.000	4.253	2.1e-05 ***
(c)	B_0	-0.581	0.092	-6.307	2.841e-10 ***
	β_1	-0.093	0.022	-4.077	4.562e-05 ***
	old_index	-0.016	0.016	-0.998	0.318
	temp_av	0.029	0.017	1.745	0.080 .
	density	0.002	0.019	0.150	0.880
	pollution_av	0.043	0.020	2.158	0.030 *
	gg_III	0.003	0.000	4.288	1.8e-05 ***

Signif. codes: 0 <= '***' < 0.001 < '**' < 0.01 < '*' < 0.05 < '.' < 0.1 < '' < 1

Source: author's elaboration of collected data

The regression results show that the coefficient of the initial level of the infection rate β_1 is less than 0 and significant for all three waves analysed in this study. This implies the existence of the convergence hypothesis (Baumol, 1986).

Since the spatial regression parameters, unlike with the OLS method, were estimated using the maximum likelihood (ML) method, this does not allow the R^2 index to be used to assess the goodness of fit of the model. In this case, therefore, the goodness of fit of the model is assessed by comparing the AIC statistics (Akaike, 1974) calculated for the OLS and SAR models (Table 2).

Table 2. Goodness of fit conditional β-convergence (SAR): (a) first wave; (b) second wave; (c) third wave

		Estimate	**p-value**
(a)	*Rho (ρ)*	0.412	-
	LR test value	27.52	(1,555e-07)
	Wald statistic	35.288	2.843e-09
	AIC	149.06 (OLS: 174.58)	-
	LM test for residual autocorrelation	0.019	0.888
(b)	*Rho (ρ)*	0.291	-
	LR test value	12.436	0.0004
	Wald statistic	13.497	0.0002
	AIC	-180.52 (OLS: -170.09)	-
	LM test for residual autocorrelation	1.101	0.294
(c)	*Rho (ρ)*	0.263	-
	LR test value	5.254	0.021
	Wald statistic	6.042	0.014
	AIC	-72.22 (OLS: -68.973)	-
	LM test for residual autocorrelation	0.272	0.601

Source: author's elaboration of collected data

The AIC calculated for SAR is always lower than the same measured for OLS. The Rho (ρ) is statistically significant as is its relationship to the dependent variable (Wald test). Therefore, the spatial model best fits the data and most accurately interprets the observed convergence process.

5. Discussion

The results obtained are robust and consistent with the established body of literature in previous medical studies suggesting that poor air quality creates chronic exposure to respiratory disease. On the other hand, population density, the old-age index and average temperature were not always found to be conditional elements of the observed convergence processes, varying in significance depending on the wave taken as the period of observation, and thus partly confirming what emerged from the reference literature. As far as the spatial delays are concerned, the spillover effects recorded by the parameter ρ (Rho) for all three waves are significant and are respectively equal to 0.41 for the first wave, 0.29 for the second, and 0.26 for the third. According to these results, therefore, it is possible to state that increases and decreases in the average growth rate in the *i-th* province can also be attributed to changes in growth levels in its neighbouring provinces. According to the estimated SAR model, *spillover effects* calculated for population density (0.12) and pollution (0.21) for the first wave are also significant. It would thus appear that provinces with a high population density over the available surface area and above-average presence of substantial air pollutants are directly responsible for the growth of contagions in neighbouring areas. Density retains its spatial influence even during the second wave by

significantly reducing its magnitude (0.04). Pollution (0.02) becomes slightly significant (p-value just under 10%) and decreases its influence in exerting an effect on the growth of contagions in neighbouring provinces. During the second wave there emerges a restraining effect due to the old index (old_index = -0.02) according to which in provinces in which there is a high presence of individuals aged 65 years or over, relative to the resident population, there is a negative relationship with the growth rate of contagions in the contiguous provinces. Finally, as regards the third wave, a weak (p-value of just under 10%) positive spatial relationship emerges between the observed temperature (0.02) and the level of contagions in the neighbouring areas. Confirmed, on the other hand, is the significance of pollution (0.04) in producing an increase in contagions in provinces sharing a border with a province characterised by high levels of this variable. Finally, all three waves share the significance of the observed durations, respectively 0.01 the first, 0.001 the second and 0.003 the third wave, showing, however, a weak spatial influence on the average rate of contagion growth.

Although consistent with the initially hypothesised framework, however, the results obtained have several limitations and implications for future research. Firstly, some critical elements should be noted in the nature of the dependent variable used. These reflections arise from the fact that it is not possible to know the true population that has been exposed to the virus. A further investigation could examine the actual number of people tested. These data are currently not available at the provincial level, and those at the regional level suffer from multiple counting due to repeated testing of positive cases. Secondly, there are some provinces that have reallocated some positive cases to other provinces due to health facility capacity or registration errors. To address these concerns, the paper proposes an analysis on aggregated wave-level data, but possible biases may still exist. Future studies could implement estimation control procedures, potentially including some dummy variables and retesting the model. A further possible source of bias may be introduced by potential outliers. Results could potentially be driven by a few provinces showing several new cases that are exceptionally far from the average. In addition to all this, it must be remembered that the Covid-19 testing policy in Italy, especially at the beginning of the pandemic, was different over time and in the various provinces. Initially, the tests were performed on suspected patients who presented themselves in hospital and/or on persons who had been in contact with positive cases, later only patients with severe symptoms were tested, and finally the tests were also performed on suspects without severe symptoms. Finally, it should be added that the statistical significance of conditional factors does not necessarily imply causality in the recorded convergence process and based on the characteristics of the data, there is no possibility of testing causality by means of a suitable counterfactual trend (in fact, it is impossible to construct a suitably randomised control group for a phenomenon that is already occurring at the time of the evaluation).

References

Alexiadis, S. (2010) Regional Convergence-clubs and Dynamic Externalities. *Italian Journal of regional Science*, **9**(1).

Akaike, Hirotugu (1974). A new look at the statistical model identification. In *IEEE Transactions on Automatic Control* **19**(6), pp. 716–723.

Arbia, G. (2005). Convergence in Per-capita GDP across European Regions using Panel Data Models Extended to Spatial Autocorrelation Effects. *ISAE Working Paper,* No.51. *SSRN: http://ssrn.com/abstract=936327*.

Azuma, K., Yanagi, U., Kagi, N., Kim, H., Masayuki, M., & Hayashi, M. (2020). Environmental factors involved in SARS-CoV-2 transmission: effect and role of indoor environmental quality in

the strategy for COVID-19 infection control. *Environmental Health and Preventive Medicine volume*, **25**(1).

Barro, R. J., Sala-I-Martin, X. (1992). Convergence. *Journal of Political Economy*, **100**, pp. 223-251.

Baumol, W. J. (1986). Productivity Growth, Convergence and Welfare: What the Long-Run Data Show. *American Economic Review*, **76**, pp. 1072-1085.

Copat, C., Cristaldi, A., Fiore, M., Grasso, A., Zuccarello, P., Signorelli, S. S., Conti, G., & Ferrante, M. (2020). The role of air pollution (PM and NO2) in COVID-19 spread and lethality: A systematic review. *Environmental Research*, **191.**

Cutrini E., Salvati L. (2021). Unraveling spatial patterns of COVID-19 in Italy: Global forces and local economic drivers. *Regional Science Policy and Practice*, **13**(S1), pp. 73-108.

Goumenou, M., Sarigiannis, D., Tsatsakis, A., Anesti, O., Docea, A. O., Petrakis, D., and Calina, D. (2020). COVID 19 in Northern Italy: An integrative overview of factors possibly influencing the sharp increase of the outbreak (Review). *Molecular Medicine Reports*, **22**, pp. 20-32.

Guliyev, H. (2020). Determining the Spatial Effects of COVID-19 Using the Spatial Panel Data Model. *Spatial Statistics*, **38**.

Guo, G., Wu, Z., and Chen, Y. (2020). Evaluation of spatially heterogeneous driving forces of the urban heat environment based on a regression tree model. *Sustainable Cities and Society*, **54**.

LeSage, J.P. (1998). *Spatial Econometrics*. Department of Economics. University of Toledo.

Liu, X., Sun, T., and Feng, Q. (2020). Dynamic spatial spillover effect of urbanization on environmental pollution in China considering the inertia characteristics of environmental pollution. *Sustainable Cities and Society*, **53**.

Mankiw, N.G. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, **107**(2), pp. 407–437.

Moosa, I. A., Khatatbeh, I. N. (2021). Robust and fragile determinants of the infection and case fatality rates of Covid-19: international cross-sectional evidence. *Applied Economics*, **53**(11), pp. 1225-1234.

Solow, R. M., T. W. Swan (1956) Economic Growth and Capital Accumulation. *Economic Record*, **32**, pp. 334-361.

Truglia, F.G. (2021). La nuda città. Spillover territoriali dei contagi da Covid-19 a Roma nella "seconda ondata", in *La metropoli continua Storia e vita sociale del quadrante Sud di Roma*, Materiali e Documenti n.78 ed., La Sapienza University Press, pp. 403-437.

Zhao, R., Zhan, L., Yao, M., and Yang, L. (2020). A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2. 5. *Sustainable Cities and Society*, **56**.