

Did Covid-19 Hit Harder in Peripheral Areas? The Case of Italian Municipalities

Francesco Armillei* Francesco Filippucci† Thomas Fletcher‡

March 2021

Abstract

The first wave of Covid-19 pandemic had a geographically heterogeneous impact even within the most severely hit regions. Exploiting a triple-differences methodology, we find that in Italy Covid-19 hit relatively harder in peripheral areas: the excess mortality in peripheral areas was almost double that of central ones in March 2020 (1.2 additional deaths every 1000 inhabitants). We leverage a rich dataset on Italian municipalities to explore mechanisms behind this gradient. We first show that socio-demographic and economic features at municipal level are highly collinear, making it hard to identify single-variable causal relationships. Using Principal Components Analysis we model excess mortality and show that areas with higher excess mortality have lower income, lower education, larger households, lower trade and higher industrial employments, and older population. Our findings highlight a strong centre-periphery gradient in the harshness of Covid-19, which we believe is highly relevant from a policy-making standpoint too.

Keywords

Covid-19 diffusion, health geography, geographical inequalities, periphery.

Highlights

- In Italy the first wave of Covid-19 hit harder in peripheral areas
- There is high correlation between socio-demographic and economic variables at municipal level, demanding great caution in the interpretation of correlations between Covid-19 and other local features
- Areas with higher excess mortality have lower income, lower education, lower trade, higher industrial employment, larger households and older population.

*Corresponding Author. Fondazione Ing. Rodolfo Debenedetti. Email: francesco.armillei@frdb.org

†Paris School of Economics. Email: francesco.filippucci@psemail.eu

‡University of Edinburgh. Email: T.Fletcher-6@sms.ed.ac.uk

We are indebted to the researchers of the localopportunitieslab.it project for data provision. We thank Caterina Alacevich, Valentin Kecht, Giulia Giupponi, Paolo Pinotti, Giacomo De Giorgi, Claudio Buongiorno Sottoriva, Paul Brandily-Snyers, Jacopo Bassetto, the Editor of *Economics&Human Biology* and two anonymous referees for their extremely helpful comments. All views expressed in this work are our own and do not represent the opinions of the entities with which we are affiliated.

1 Introduction

The impact of the Covid-19 virus was far from geographically homogeneous: not only were some countries hit harder than others, but even within those countries there were significant differences between various areas. When it comes to the diffusion of the virus, it seems that location matters. More specifically, the socio-economic features of different zones appear to have influenced the virus' diffusion even at a very granular scale, in a manner which is consistent with the great relevance of local characteristics in shaping socio-economic phenomena highlighted by recent economic literature (Chetty, Hendren and Katz, 2016).

In the aftermath of the first wave of pandemic, a burgeoning stream of literature studied what the relevant correlations are between Covid-19 spread (in most cases measured by confirmed Covid-19 cases and deaths) and the socio-demographic and economic features of different areas in hit countries; various covariates have been taken into consideration: income (Borjas, 2020; Brandily, Brébion, Briole and Khoury, 2020; Knittel and Ozaltun, 2020), population density (Hamidi, Ewing and Sabouri, 2020), social capital (Bartscher, Seitz, Siegloch, Slotwinski and Wehrhöfer, 2020; Francesca Borgonovi and Subramanian, 2020; Kuchler, Russel and Stroebel, 2020), health facilities (Alacevich, Cavalli, Giuntella, Lagravinese, Moscone and Nicodemo, 2020; Sussman, 2020), demographic profile (Borjas, 2020; Desmet and Wacziarg, 2020; Knittel and Ozaltun, 2020; Sa, 2020), intergenerational co-residence patterns (Fenoll and Grossbard, 2020), pollution (Coccia, 2020; Isphording and Pestel, 2020; Matthew A. Cole, 2020), ethnicity (Borjas, 2020; Hamman, 2021; Knittel and Ozaltun, 2020) and weather (Kapoor, Rho, Sangha, Sharma, Shenoy and Xu, 2020; Knittel and Ozaltun, 2020). This emerging stream of literature is closely connected to previous works studying the socio-demographic and economic features correlated at local level to the spread of other diseases (e.g. the seasonal influenza, as in Markowitz, Nesson and Robinson (2019)) or to previous pandemic episodes (e.g. the 1918 influenza, as in Clay, Lewis and Severnini (2019)). On the whole, however, it seems that the literature has still not reached a comprehensive understanding of the phenomenon.

Specific attention has been devoted to the geographical correlates of Covid-19 spread, to the urban structure and to the centre-periphery gradient. Indeed, understanding the spatial dynamics of Covid-19 is essential also for its mitigation, as it helps to clarify the extent and impact of the pandemic and it can aid decision making, planning and community action (Ivan Franch-Pardo and Billa, 2020). From a policy-making standpoint, many non-pharmaceutical interventions implemented by governments featured specific geographical boundaries, and they are part of the broader group of space-based policies which could help in facing the Covid-19 pandemic. Gerritse (2020), exploiting data on US counties, finds that population density is pos-

itively correlated to infection rates at the outbreak. Working on US counties as well, Knittel and Ozaltun (2020) find that higher amounts of commuting via public transportation, relative to working from home, is correlated with higher death rates; Desmet and Wacziarg (2020) find similar results. Sa (2020), leveraging data from England and Wales, finds that contagion is higher where more people make use of public transportation. Carozzi, Provenzano and Roth (2020) find that density has affected the timing of the outbreak in each county, with denser locations being more likely to have an early outbreak, but do not find any impact on Covid-19 cases and deaths. Hamidi et al. (2020) find that, after controlling for metropolitan size and other relevant features, county density leads to lower infection rates and lower death rates. Again, it seems that a comprehensive picture on the relationship between Covid-19 and the urban landscape still does not emerge.

In order to contribute to this literature, in the present work we leverage rich granular data on Italian municipalities. Italy constitutes an interesting case for exploring the relationship between Covid-19 spread and location, as it was the first western country to be severely hit by the pandemic, in a period of great unpreparedness. The first Covid-19 case in Italy was officially detected on the 30th of January 2020 in Rome. The first case of secondary transmission was detected on the 18th of February in Codogno, an intermediately peripheral municipality in the region of Lombardy. On the 8th of March the entire region of Lombardy was locked down and two weeks later, on the 22nd, the Italian government implemented the first national lockdown, prohibiting all individuals on Italian soil from travelling, except for working or health reasons; additionally, every non-necessary economic activity was shut down.

Our aim in this work is to enrich the current understanding of Covid-19 diffusion and to shed new light on the relationship between the first wave of the virus and the centre-periphery gradient. Firstly, we seek to understand whether the virus hit harder in peripheral areas and, exploiting a triple-differences methodology, we find that the excess mortality in peripheral areas was almost double that of central ones in March 2020 (1.2 additionally deaths every 1000 inhabitants). Then, we move to an analysis of the socio-demographic and economic features which characterize peripheral areas and correlate with Covid-19 diffusion. Using Principal Components Analysis we model excess mortality and show that areas with higher excess mortality have lower income, lower education, larger households, lower trade and higher industrial employments, and older population.

We believe that our work contributes to the existing literature in several ways. We highlight the existence of a strong centre-periphery gradient in Covid-19 harshness, a pattern that has not yet been mapped clearly. Besides, since our definition of periphery is multidimensional (and does not rely only on proxies as the

municipal number of inhabitants or population density as in most previous works), we believe that this pattern is highly relevant from both an academic and policy-making standpoint. As far as socio-demographic and economic features are concerned, our findings seem to corroborate those of existing studies, especially regarding the negative association between Covid-19 and income, education and household dimension. With respect to previous works our analysis also highlights the importance of the sectoral composition of the economy, even in the earliest phase of virus diffusion. We contribute to the existing literature from a methodological standpoint too. Firstly, we stress the importance of using as granular as possible data, since it allows for a better understanding of spatial dynamics in Covid-19 diffusion. In our case we make use of municipal level data, while many existing studies resort to aggregate national, regional or county level data. Secondly, we highlight the importance of using excess mortality as dependent variable, since the number of official Covid-19 cases and deaths might have been measured inaccurately, especially during the first wave of the virus. Thirdly, we highlight that there is high correlation between socio-demographic and economic variables at municipal level, demanding great caution in the interpretation of correlations between Covid-19 and other local features.

2 Data and measurement

For this study we rely on four main data sources, the first of which is the official dataset provided by the Italian National Statistical Institute (ISTAT) containing the daily death count in each Italian municipality. This dataset allows us to perform a very granular analysis of the period in which the Covid-19 crisis hit the country. For each municipality we observe the daily number of resident deaths between 2015 and 2020 by gender and age (without, however, knowing the cause of death). One should keep in mind that every death is recorded in the municipality of residence of the dead.

The second data source, also provided by ISTAT, is the dataset of travel-times (in minutes) between all municipalities, built from a commercial road system graph.

The third data source is key for our definition of the centre-periphery gradient. We decide to adopt a multidimensional definition of periphery and we do not rely only on proxies as the number of inhabitants of each municipality or the population density. For this purpose, we exploit the municipality classification developed by the Italian Agency for Territorial Cohesion, which divides all Italian municipalities into six classes (from *attractor* municipalities, which constitutes the “centre”, to ultra-peripheral ones). Consistently with this methodology, we classify a municipality as an *attractor* if it possesses the following three features: i)

educational facilities up to secondary education ii) a complete hospital (i.e. a hospital which guarantees the functions of First Aid, observation, short hospitalization and resuscitation, and which carries out diagnostic-therapeutic interventions in general medicine, general surgery, orthopedics and traumatology, and cardiology intensive care) iii) a medium-sized train station. The other municipalities are then classified based on their proximity to *attractors*¹. Table 1 presents a brief description of the classified municipalities, while Figure 1 maps their distribution.

Table 1: Definition of municipality classes

Class	Travel dist. from centres	N. of municipalities	%	Population	%
Attractor	0	219	2.7%	21223562	35.7%
Inter-Municipal Attractor	0	104	1.3%	2466455	4.1%
Belt	$t < 20'$	3508	43.4%	22203219	37.4%
Intermediate	$20' < t < 40'$	2377	29.4%	8952266	15.1%
Peripheral	$40' < t < 75'$	1526	18.9%	3671372	6.2%
Ultra-Peripheral	$t > 75'$	358	4.4%	916870	1.5%

Notes. The table reports the distribution of Italian municipalities along 6 classes defining a centre-periphery gradient. Attractor municipalities (and Inter-Municipal Attractors) are those which possess: i) educational facilities up to secondary education ii) a complete hospital iii) a medium-sized train station. The others are classified according to their travel distance from the closest Attractor municipality. The classification has been elaborated by the Italian National Agency for Territorial Cohesion.

The fourth data source is the Local Opportunities Lab (LOL from now on) dataset². This newly available dataset gathers information from a number of public sources at a municipal granularity (mainly census, fiscal data and official statistics by the Italian National Institute of Statistics), with information that ranges from housing to education to income. Table 2 provides a summary of the covariates considered as a starting point for our analysis.

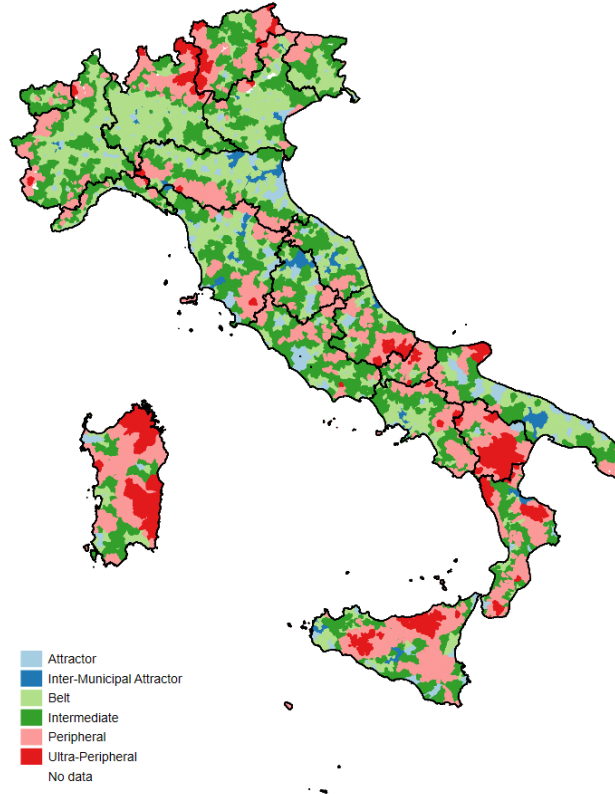
The datasets are merged by unique municipality ID. The most recent value of every covariate in the LOL dataset is then associated to each of the 7805 (out of 7907) municipalities observed in the ISTAT dataset, covering 99.96% of the Italian population.

Concerning the period of analysis, we focus on the month of March. In Italy, this corresponded to the peak of first wave of Covid-19. Moreover, deaths during this period are likely resulting from contagions having taken place before the national lockdown (beginning on March 22nd), reducing what could have otherwise acted as a policy confounding factor, undermining our analysis of the variables correlated with the spread of Covid-19. On the one hand, there is growing evidence that the timing and the stringency of lockdown policies were endogenous to the economic and political features of hit countries (Ferraresi, Kotsogiannis, Rizzo

¹For a more comprehensive explanation see the official document of the Italian National Agency for Territorial Cohesion: http://www.programmazioneeconomica.gov.it/wp-content/uploads/2017/02/Accordo-P-Strategia_nazionale_per_le_Aree_interne_definizione_obiettivi_strumenti_e_governance_2014.pdf

²For more information and to access the data, please visit <https://www.localopportunitieslab.it/>

Figure 1: Italian municipality distribution by class



and Secomandi, 2020). On the other, citizen behavioural response and the effects of lockdowns and other containment policies have been far from geographically homogeneous (di Porto, Naticchioni and Scrutinio, 2020) and correlated with several socio-economic features (Brodeur, Gray, Islam and Bhuiyan, forthcoming).

A key feature of our analysis is how we measure the harshness of Covid-19 diffusion in a municipality. The Italian government did not release municipal-level figures on Covid-19-related deaths; even more importantly, such a measure might be inaccurate and endogenous in any case, since the screening system covers a small and selected fraction of the population (as already highlighted in literature, see for example Borjas (2020)), particularly in the early phase of the pandemic, and such fraction is highly dependent on local health policy. We hence look at how the death rate (number of deaths divided by total population) varied in March 2020 with respect to the average of March 2017-19. From now on, we will refer to this rate difference as excess mortality (in deaths every 1000 inhabitants). Specifically, we define as our dependent variable the excess mortality rate at municipal level, computed as the difference between the mortality rate (deaths over total population) in March 2020 and the average in March 2017-2019 (multiplied by 1000). We compare March 2020 to the March 2017-2019 average (instead of just March 2019) to avoid excess volatility due to smaller

Table 2: Covariates considered in the analysis

Variable Category	Single variables included in the LOL dataset
<i>Income</i>	Income from buildings, labour, pensions, autonomous work, entrepreneurial profits, dividends, total income. Income frequencies in 7 brackets (<0, 0-10k, 11k-15k, 16k-26k, 26k-55k, 57k-75k, 76k-120k, >120k). Gini index.
<i>Population</i>	Number of residents, total, by 5y age brackets and by gender. Density, dependency rate, incidence of foreign-borns.
<i>Housing</i>	Square meters per inhabitant, crowding index, average house price.
<i>Household composition</i>	Average household dimension, incidence of families with no single reference residency, incidence of households with more than one reference residency, incidence of youths and elders living alone, incidence of young and elder single-parents, incidence of young and elder couples with and without children.
<i>Education</i>	Ratio of adults with bachelor degree or above to adults with middle school diploma.
<i>Labour Market</i>	Labour market participation by gender, young NEETs incidence, ratio of active/non active youths, unemployment and employment rates by gender, turnover index, youth employment, employment in agriculture, industrial, services trade and non-trade, incidence of high-specialization/low-specialization.
<i>Mobility</i>	Daily mobility for work or study by means of transportation.
<i>Urban structure</i>	Share of green spaces, presence of tertiary-education institutions, libraries, pharmacies every 10.000 inhabitants.
<i>Social capital</i>	Share of workers in APS and KIBS sectors, share of recycled waste. Employment in cultural sector and in associations (ateco 91 and 94). Incidence of volunteers.
<i>Hospital coverage</i>	Number of beds, public beds, and beds in ICU per inhabitant in the municipality and within a 30-minute range.

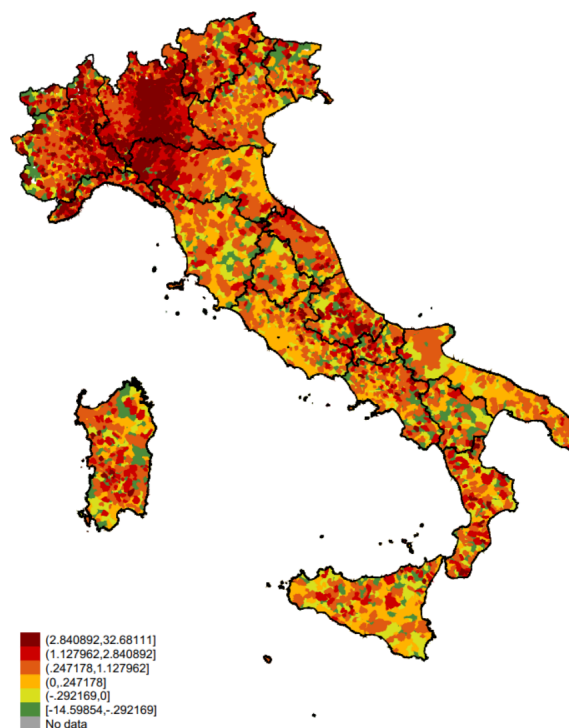
Notes. Data taken from the Local Opportunities Lab (LOL) dataset.

municipalities with zero deaths. At the same time, we refrain from using a longer average of lags to avoid capturing long-term trends of increase/decrease in mortality. In any case, our results are robust to defining the outcome variable of interest as the simple difference between mortality in 2020 and 2019, 2020 and 2018, 2020 and 2017.

It is important to note that the analysis of excess mortality captures at the same time heterogeneities in both the virus spread (i.e. more people getting infected in specific areas) and in its mortality (i.e. conditional on being infected, more people dying in specific areas). However, due to the mentioned data limitations, we believe it is not possible to disentangle these two channels at present, especially for the early-spread period we consider in this work.

As a validation of our measure, we compare our figures on excess mortality with those of the Italian govern-

Figure 2: Excess mortality (March on March) in Italy

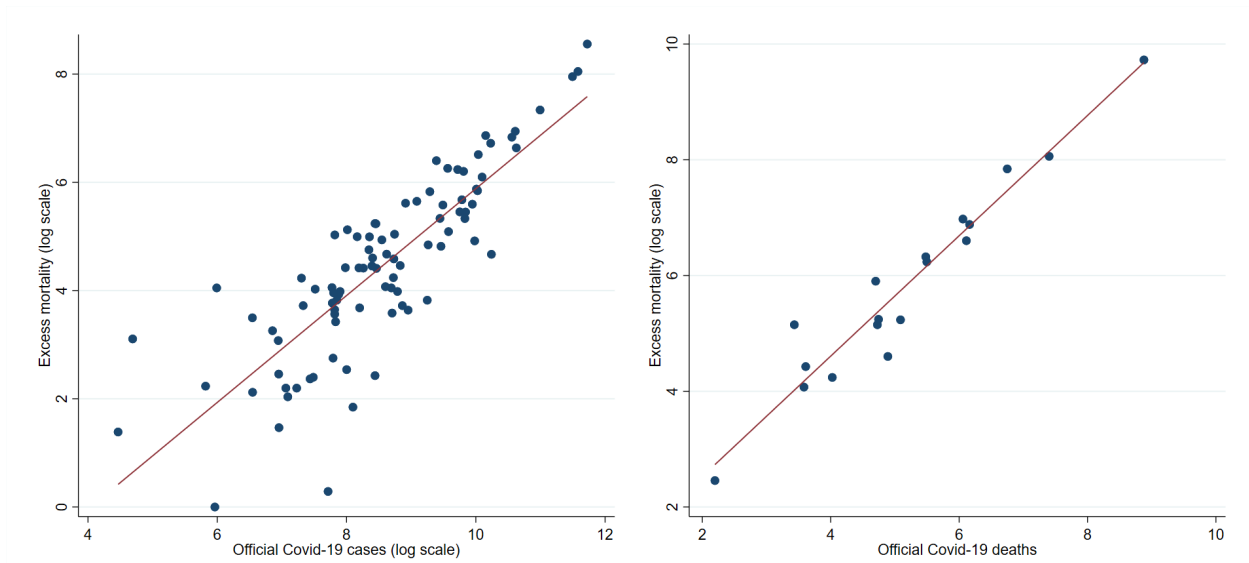


Notes. The figure reports March 2020 excess mortality in Italy, meaning the increase in deaths every 1000 residents between March 2020 and the average of March 2017, 2018 and 2019.

ment relative to the month of March, containing data on the number of Covid-19-related deaths aggregated at a regional level (NUTS-2) and data on the number of Covid-19 cases at provincial level (NUTS-3). These two levels of analysis would not be granular enough for the purposes of our work. Our measure lines up well with the data on regional deaths and province infection cases, as graphical evidence in Figure 3 shows. Interestingly, however, our excess mortality death toll counts around 26200 deaths in March 2020, while the official figures of the government only counts 12400 Covid-19 related deaths. We argue that this unaccounted excess-mortality stems from people dying from Covid-19 without a Covid-19 diagnosis, for example at home or in elderly-care centres, as found by Richterich (2020) and Sawano, Kotera, Ozaki, Murayama, Tanimoto, Sah and Wang (2020) among others. For the Italian case, Michelozzi, De'Donato, Scortichini, Pezzotti, Stafoggia, Sario, Costa, Noccioli, Riccardo, Bella and et al. (2020) find that official Covid-19 deaths accounted only for half of total excess mortality, with differences by age: among adults almost all excess deaths were reported as from Covid-19, while among the elderly only one third of the excess was reported as such.

As is well known, the increase in mortality was concentrated in northern provinces; however, it is very

Figure 3: Excess mortality comparison with official Covid-19 cases and deaths



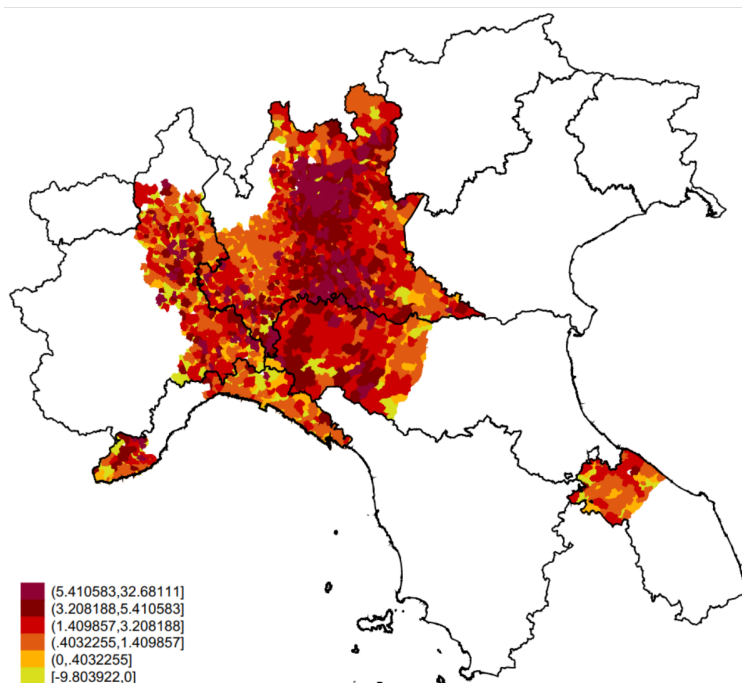
Notes. The graph on the left compares excess mortality (the increase in deaths every 1000 residents between March 2020 and the average of March 2017-2019) with official Covid-19 cases at provincial level in the month of March (correlation: 0.96). The graph on the right compares excess mortality with official Covid-19 related deaths at regional level in the month of March (correlation: 0.99). Official data on cases and deaths released by the Italian government available here: <https://github.com/pcm-dpc/Covid-19>

difficult, if not hardly possible, to causally explain such occurrence at this stage. The regions in which the first outbreaks occurred (Lombardy, Piedmont and Emilia-Romagna) are different from the rest of the country in terms of several characteristics, including income, weather and demographic profile, but also in terms of international trade relationships and internal mobility, therefore it may be the case that clusters of contagion developed there early and quickly because of it; still, we cannot exclude that these regions were the first hit just because of chance, and since early outbreaks were concentrated in just a few regions, we lack a sufficient number of observations to study how the evolution of outbreaks differed in different regions. Later outbreaks are not comparable and not a good measure of how the virus spreads independently from policy, since policy and population behavior changed dramatically to contain the spread after the first few outbreaks. Further research will be needed to explore these aspects of the pandemic. In light of this reasoning, it is key to understand that we do not focus on heterogeneities between Italian regions, but instead on heterogeneities within regions: we strongly believe that this granular level of analysis is the most suited to study the local determinants of the spread on Covid-19 and to assess whether it hit peripheral areas harder.

In order to account for the uneven inter-regional distribution of the initial Covid-19 exogenous shock, we distinguish between high and low infection areas. We classify a municipality as high (low) infection if it

belongs to a province with excess mortality above (below) the 75th percentile. We assume that the Covid-19 exogenous shock was homogeneous in high infection provinces; in other words, we believe that high infection provinces were equally exposed to the initial arrival of the virus. Consequently, the heterogeneity in the outcomes (i.e. excess mortality) at municipal level has to stem from the heterogeneous centre-periphery gradient of municipalities and from heterogeneous socio-demographic and economic features. Excess mortality in high-infection municipalities is visualised in Figure 4, showing that the Covid-19 death toll varied significantly even within the most affected provinces, to the point of including municipalities with negative excess mortality, i.e. with decreased death rate compared to March 2017-2019.

Figure 4: Excess mortality (March on March) in most affected Italian provinces



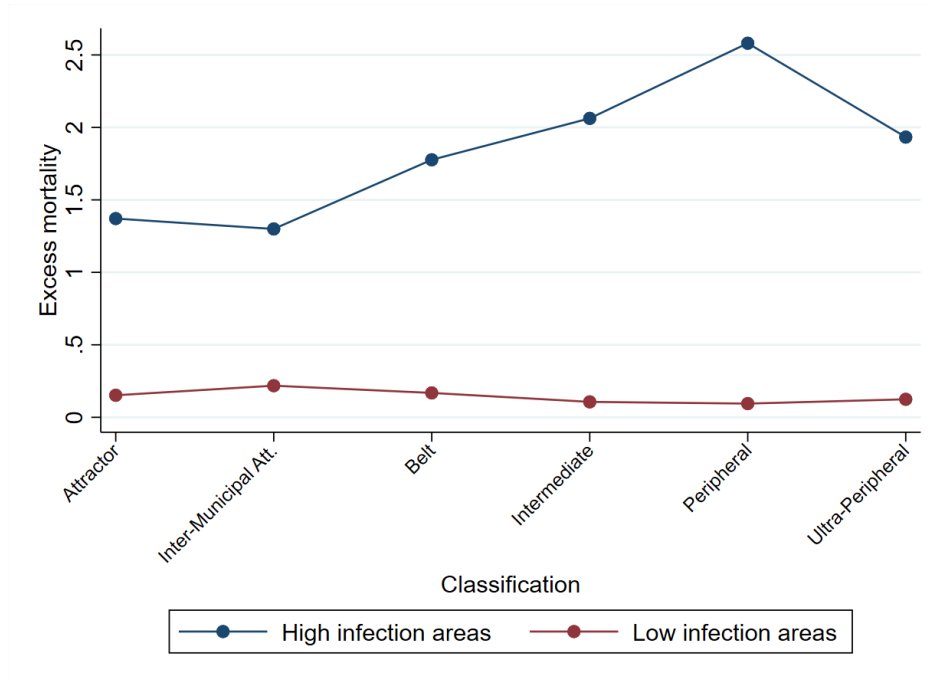
Notes. The map reports the March 2020 excess mortality, meaning the increase in deaths every 1000 residents between March 2020 and the average of March 2017-2019. Only high infection provinces are shown, meaning those with an excess mortality above the 75th percentile.

3 A triple-differences approach

3.1 Method

Figure 5 reports, from a merely descriptive point of view, the excess mortality of the six municipality classes for both high and low infection areas. A very clear gradient is present between central and peripheral areas in high-infection provinces, while no differences appear between areas within low-infection provinces.

Figure 5: Extra-mortality by municipality class



Notes. The graph shows the difference in death-rate (i.e. the extra-mortality) between March 2020 and the average of March 2017-2019. Figures are reported for high and low infections areas and for the six municipality classes.

Starting from this first graphical intuition, we rely on a triple-differences approach to get a precise estimate of how the Covid-19 virus impacted different municipalities within high infection provinces, in the spirit of Brandily et al. (2020). The intuition is as follows: the response variation in low-infection areas is used to isolate the Covid-19 effect in the high-infection ones. In the early phase of the first wave, Covid-19 cases were almost absent in low-infection areas, while country-wide confoundings affected both regions roughly uniformly. Hence, net of the variation in low infection areas, the increase in mortality in high infection areas isolates the effect of Covid-19. We compare how this effect varies for central and peripheral sub-populations. In practice, we regress the mortality in each municipality on a set of dummy variables accounting for: i) time (2020 vs 2017-2019 average) ii) infection intensity of Covid-19 (high vs low infection) iii) degree of centre-periphery (according to the aforementioned six classes). We hence estimate through OLS the following

equation:

$$\begin{aligned}
Y_{i,t} = & \beta_1 \mathbb{1}(t = 2020) + \sum_d \beta_2^d \mathbb{1}(D_i = d) + \sum_i R_i \\
& + \beta_3 [\mathbb{1}(t = 2020) \cdot \mathbb{1}(G_i = high)] + \sum_d \beta_4^d [\mathbb{1}(t = 2020) \cdot \mathbb{1}(D_i = d)] + \sum_d \beta_5^d [\mathbb{1}(D_i = d) \cdot \mathbb{1}(G_i = high)] + \\
& + \sum_d \beta_6^d [\mathbb{1}(t = 2020) \cdot \mathbb{1}(D_i = d) \cdot \mathbb{1}(G_i = high)] + \\
& + \delta_1 S_i + \delta_2 [\mathbb{1}(t = 2020) \cdot S_i] + \delta_3 [S_i \cdot \mathbb{1}(G_i = high)] + \delta_4 [S_i \cdot \mathbb{1}(t = 2020) \cdot \mathbb{1}(G_i = high)] + \epsilon_{i,t} \quad (1)
\end{aligned}$$

in which t is time (note that we consider only two periods: 2020 and the 2017-2019 average), $G_i \in \{high, low\}$ is the infection group, $D_i \in \{Attractor, \dots, Ultra-Peripheral\}$ is the categorical variable for the six centre-periphery class; the model, consistently with the triple differences approach, also includes all the possible interactions between these variables. We also add R_i , unit-level fixed effects (at local labour market or municipal level depending on the model³), and S_i , a control for the share of population above the age of 80 (together with its interaction with the time and infection intensity dummies). It is worth noting that in this section we do not control for more covariates (such as hospital presence, education or transports) since these public services constitute part of the variation of interest; their role (and that of all other mentioned ones) is instead explored later in Section 4. In the regression we weight each municipality observation by its population. Our coefficient of interest are β_6^d s, which capture differences in excess mortality between municipalities with different centre-periphery class in high infection provinces. Under the assumption that in the absence of Covid-19 the average difference in the evolution of mortality in March (2020 vs 2017-2019 average) between central and peripheral areas would have been the same in high and low infection provinces, this model identifies the reduced-form relationship between centre-periphery gradient and Covid-19 excess mortality.

3.2 Results: Covid-19 Hit Harder in Peripheral Areas

The results of our regression are reported in Table 3, showing only the breakdown of the set of coefficients of interest, $\{\beta_6^d\}$, in which the attractor class works as baseline. A clear centre-periphery pattern emerges, with peripheral areas (and, to a lesser extent, intermediate ones) being hit harder by Covid-19: after the Covid-19 shock and within high-infection provinces, the death rate of these areas increased more than in attractor municipalities. This increase is quite sizable, as the baseline average excess mortality (attractor municipalities in high-infection provinces) was of 1.37 deaths per 1000 inhabitants during the month of March 2020,

³In the specification with municipal level fixed-effects the second term on the right-hand side of the Equation, $\sum_d \beta_2^d \mathbb{1}(D_i = d)$, clearly is not present because of perfect collinearity.

making the increase in death-rate in peripheral areas almost double that of central ones. Figure 6 visualises the results of column 3 in Table 3.

The coefficients are reassuringly stable when the unit-level fixed effects are at either local labour market (LLM) or municipal level. As far as the statistical significance of our estimates is concerned, we report standard errors obtained with both clustering levels (LLM and municipal). Which level of clustering is preferable is not an easy call, as it should account for serial correlation in the errors at a unit level and for unobserved components in outcomes for units within clusters (Abadie, Athey, Imbens and Wooldridge, 2017; Bertrand, Duflo and Mullainathan, 2004). Comfortingly, our coefficients of interest remain statistically significant across the different specifications and clustering choices (growing more conservative towards right in Table 3), even if at different conventional levels.

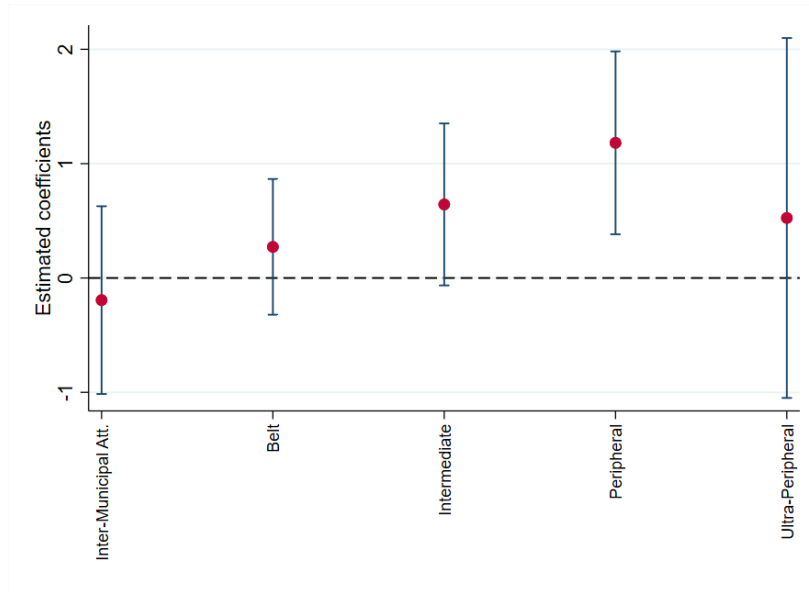
Table 3: Main results (triple-diff)

Triple-Interaction Value	(1) Death rate	(2) Death rate	(3) Death rate	(4) Death rate
Inter-Municipal Attractor # post # high infection	-0.168 (0.319)	-0.168 (0.239)	-0.194 (0.419)	-0.194 (0.347)
Belt # post # high infection	0.395* (0.238)	0.395** (0.154)	0.272 (0.303)	0.272 (0.259)
Intermediate # post # high infection	0.792*** (0.268)	0.792** (0.365)	0.643* (0.361)	0.643 (0.451)
Peripheral # post # high infection	1.338*** (0.304)	1.338*** (0.517)	1.182*** (0.408)	1.182* (0.683)
Ultra-Peripheral # post # high infection	0.686 (0.579)	0.686 (0.786)	0.525 (0.803)	0.525 (1.058)
Observations	15,602	15,602	15,602	15,602
R-squared	0.726	0.726	0.814	0.814
Unit FEs	LLM	LLM	Municipal	Municipal
Cluster level	Municipal	LLM	Municipal	LLM
Controls	Over 80	Over 80	Over 80	Over 80

Notes. The table presents the results of the triple-differences analysis. We regress mortality in each municipality on a set of dummy variables accounting for: i) time (2020 vs 2017-2019 average) ii) infection intensity of Covid-19 (high vs low infection) iii) degree of centre-periphery (six classes). All the interactions among these dummies are also included. The table reports only the coefficients of the triple-interaction term, where the attractor class works as baseline. Coefficients are estimated with OLS. Observations are weighted by the municipal population. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Our key identifying assumption is that, in the absence of the Covid-19 shock, the difference in evolution of mortality in March (2020 vs 2017-2019 average) between central and peripheral areas would have been the same in provinces affected by Covid-19 in March 2020. An implication of this assumption is that before the Covid-19 shock we should observe a parallel trend in March mortality rates for different classes. Figure 7

Figure 6: Estimated coefficients of Covid-19 impact

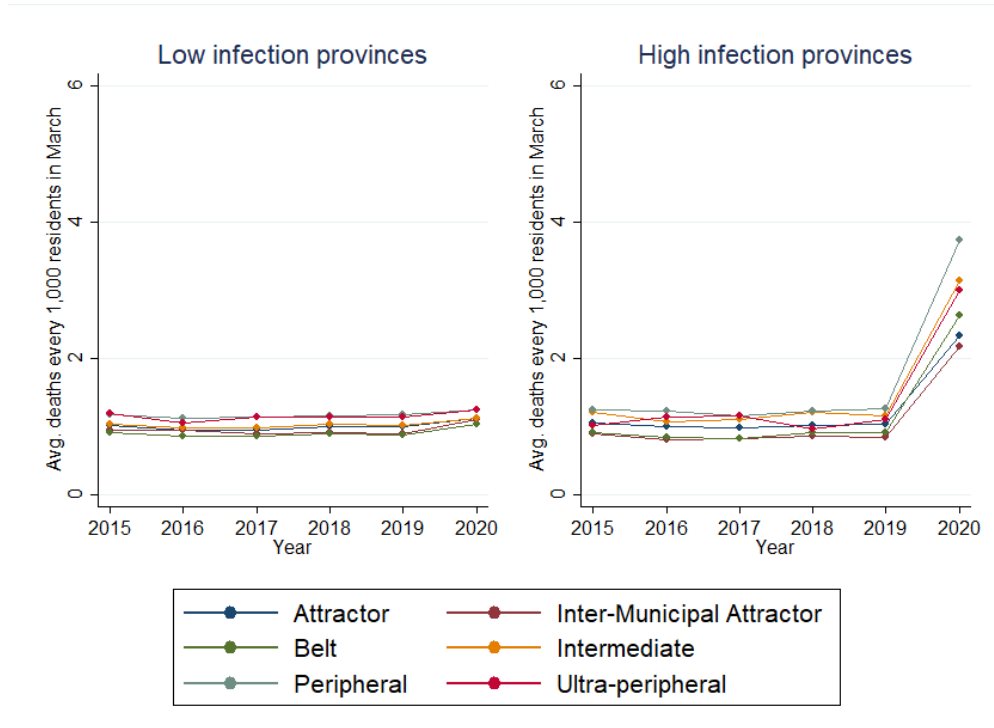


Notes. The graph shows the municipality class coefficients estimated through the triple-differences approach with local labour market fixed effects, where the *attractor* class works as baseline. The reported standard errors are clustered at local labour market level.

provides graphical evidence of this: in both low and high infection provinces the death rates of municipalities belonging to different centre-periphery classes were clearly on parallel trends before 2020. Also, the levels of pre-Covid-19 death rates are similar in high and low infection provinces.

To formally test the parallel trends we conduct a placebo analysis: Table A.1 replicates column 1 of Table 3 (our least conservative specification), revealing no significant effect in pre-Covid-19 years, reassuring us on the validity of the parallel trend assumption. Note also that in Figure 7 low infection provinces do not show any significant jump in mortality in March 2020, as they were not reached by the Covid-19 shock: this further justifies our choice to focus on high infection provinces (defined as those above the 75th percentile of extra-mortality) as our treatment group.

Figure 7: Death-rate time trend



Notes. The graph shows the time trend of the death-rate in the month of March for high and low infections areas and for two broad municipality classes.

We provide some further checks to our analysis. Firstly, we address spatial correlation by estimating a Leroux Conditional AutoRegressive (CAR-Leroux) model (Lee, 2013; Leroux, Lei and Breslow, 2000): a type of generalised linear mixed model with directly incorporated spatial dependence handling and with no prior assumption on its strength, thus addressing our needs (note also that its conception context is similar to our own). The metric we use for this spatial modelling is not as-the-crow-flies distance but rather effective road-travel-time, which is much more relevant to this analysis. The estimation shows that the spatial correlation is predictably very high ($\rho = 0.945$), but the ultimate results (shown in Table A.2) are consistent with those obtained with the previous model estimated with OLS, both in terms of point estimates and in terms of inference.

Then, in order to provide a robustness check with respect to our dependent variable choice, we estimate our model using the differences between mortality in March 2020 and in March 2017, 2018 and 2019 separately. Without loss of generality, in this case and in the following ones we use the specification with municipal level fixed effects and we cluster standard errors at municipal level as well. The results of this exercise, shown in columns 1-3 of Table A.3, are consistent with those in Table 3, both in terms of point estimates and of inference.

As a further robustness check with respect to our high/low infection provinces choice, we estimate the same specification using different high/low infection threshold. Specifically, in columns 4-6 of Table A.3 we firstly show that the relationship we find is robust if we define as high infection those provinces above the 70th and the 80th percentile. Then we define as “high infection” those provinces for which the level of mortality in March 2020 is an outlier compared to the distribution of death rates (we use the standard outlier definition of exceeding the 3rd quartile by more than 1.5 times the inter-quartile range). Again, results are in line with those obtained with the preferred definition of the 75th percentile shown in Table 3. Lastly, we check our method of selecting provinces hit by the Covid-19 shock in March 2020 by splitting them in high and low infection based on February provincial level contagions. The results, shown in column 7 of Table A.3, are once again reassuring. Although there is an attenuation of the coefficient for Intermediate municipalities, the coefficient for Peripheral areas remains almost unchanged despite the fact that contagions in February are a very noisy estimate of infections due to the testing system being unprepared and concentrated in specific areas.

3.3 Method Discussion

Before moving on, it is worth discussing the definition of G_i used in the triple-difference approach and its robustness with respect to possible endogeneity risks. Our goal is to investigate the potential centre-periphery gradient of a shock, Covid-19. To do this, we could have run a simple double difference, comparing the evolution over time of death rates in March in all Italian municipalities:

$$Y_{i,t} = \beta_1 \mathbb{1}(t = 2020) + \sum_d \beta_2^d (\mathbb{1}(t = 2020) \cdot \mathbb{1}(D_i = d)) + \sum_i R_i + \delta_1 S_i + \delta_2 (\mathbb{1}(t = 2020) \cdot S_i) + \epsilon_{i,t} \quad (2)$$

This reduced-form model compares mortality in central and peripheral municipalities in the whole country, under the assumption that mortality is not on a different trend in centres vs peripheries. However, this pooled strategy lacks power in estimating the differing effects of Covid-19 in centres and peripheries, as attested by the non-significant results in the first column of Table 4. This is not surprising, as almost 3/4 of the sample reports no major change in mortality in March 2020 compared to previous years since that month Covid-19 infections were concentrated in a specific area of the country and were practically absent elsewhere. One solution is to run a separate diff-in-diff for only the macro-areas which were actually exposed to Covid-19 in March 2020. As mentioned above, we define these areas by provinces (i.e. administratively defined groups of municipalities) and we compare different municipalities within them. This can be seen as

an heterogeneity analysis of Equation 2. The results of this restricted diff-in-diff are reported in column 2 of Table 4, and they clearly exhibit the centre-periphery gradient identified in the previous part of this section. Conversely, regressions on the remaining low-infection provinces display no significant gradient.

Table 4: Difference in differences results

Interaction Value	(1) Death rate	(2) Death rate	(3) Death rate
Inter-Municipal Attractor # post	-0.0798 (0.153)	-0.149 (0.433)	-0.0186 (0.0949)
Belt # post	0.0986 (0.129)	0.299 (0.317)	-0.0843 (0.0688)
Intermediate # post	-0.0260 (0.131)	0.810** (0.373)	-0.00783 (0.0712)
Peripheral # post	0.0740 (0.136)	1.403*** (0.419)	0.128* (0.0763)
Ultra-Peripheral # post	-0.0144 (0.187)	0.678 (0.849)	0.126 (0.137)
Observations	15,602	4,014	11,588
R-squared	0.722	0.761	0.847
Sample	All provinces	High infection provinces	Low infection provinces
Cluster level	Municipality	Municipality	Municipality
Unit FEs	Municipality	Municipality	Municipality
Controls	Over 80	Over 80	Over 80

Notes. The table reports the results of a difference in difference approach as in Equation 2. Column (1) reports results for the whole sample, column (2) restricts to high infection provinces and column (3) only to low infection provinces. The table reports only the coefficients of the interaction term, where the attractor class works as baseline. Coefficients are estimated with OLS. Observations are weighted by the municipal population. Clustered standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Our preferred specification is instead a triple difference on the results obtained for high and low infection provinces, i.e. Equation 1 and the results presented in columns 2 and 3 of Table 4. This approach relies on more general assumptions and is more conservative, as it removes potential confounders occurring at the same time as Covid-19 (such as the economic crisis or the first effects of the national lockdown).

While our goal is to estimate effects conditional on being a high-infection area ($G_i = 1$), endogeneity issues should be investigated in both the restricted diff-in-diff and in the triple-diff case since we could only define G_i according to the evolution in mean ex-post outcomes in provinces (which, we stress again, are pre-determined administrative agglomerations of municipalities). This is why it would be naïve to interpret the interaction of high infection and post variables, β_3 . We are, however, allowed to interpret the triple interaction coefficient between high infection provinces and peripheral status (i.e. β_6 in Equation 1) under the identification assumption that Covid-19 presence in the province (captured by province-level death rate increase) is exogenous to the *within province* death rate increase gradient. The relevant assumptions are of

parallel trend between mortality in centre-periphery for the restricted diff-in-diff, and of equality of relative trends for the triple difference⁴.

The first three columns of Table A.3 suggest that the equal relative trends assumption is credible, as placebo tests are all non significant. Moreover, Figure 7 suggests that also parallel trends between centre and periphery is a credible assumption, and that high and low infection provinces share not only a clear parallel trend in mortality in the years preceding Covid-19, but also very similar levels in mortality rates. Finally, reassuring evidence of low risk of bias arising from the definition of G_i comes from the robustness checks we run in Table A.3, where the results hold with different definition of high-infection areas.

4 Mechanisms of the Centre-Periphery Gradient: a Dimensionality Reduction Approach

4.1 The caveat of highly correlated regressors

It is important to underline that the previous section shows a reduced-form relationship, a visible heterogeneity stemming from a hidden causal relationship. In other words, the relationship between peripheral areas and excess mortality from Covid-19 is not causal per sé: the causal origin of the higher spread of the virus has to be looked for in the social, demographic and economic features which characterize central and peripheral areas. Indeed, this is what many works have tried to do so far; two broad approaches are clearly possible in this case: univariate or multivariate analysis.

As an example of the former, Armillei and Filippucci (2020) find that when regressing Covid-19 excess mortality on *single* socio-demographic and economic regressors, average income, education, use of public transports and employment in the service sector (all with negative sign) are the most significant and robust

⁴Formally, define potential outcomes $Y_{i,t}(1)$ if there is Covid-19 in the province i at time t and $Y_{i,t}(0)$ if not. Also define the evolution in mean outcomes for $D_i = d, G_i = g$ as

$$\begin{aligned} & \mathbb{E}(Y_{i,t}(0)|D_i = d, G_i = g, T_i = 2020) \\ & - \mathbb{E}(Y_{i,t}(0)|D_i = d, G_i = g, 2017 \leq T_i \leq 2019) = \Delta_d^g \quad \forall D_i, G_i = high, low \end{aligned}$$

Identification in the diff-in-diff case requires the evolution in potential outcomes to be the same in centre and periphery, namely the parallel trend assumption:

$$\begin{aligned} \Delta_d^{high} &= \Delta^{high} \quad \forall D_i \\ \Delta_d^{low} &= \Delta^{low} \quad \forall D_i \end{aligned} \tag{3}$$

Adapting Olden and Møen (2020) to our setting, identification of our triple difference requires instead only equality between relative trends:

$$\Delta_d^{high} - \Delta_{Attractor}^{high} = \Delta_d^{low} - \Delta_{Attractor}^{low} \quad \forall D_i \tag{4}$$

Note that (3) is a sufficient (and yet not necessary) condition for (4). Note also that G_i does not need to be independent from $Y_{i,t}$, but only mean independent from the relative difference in the evolution of *potential* outcomes $Y_{i,t}(0)$.

correlates. Brandily et al. (2020) propose instead a horse-race approach, where the coefficient of interest (poverty in their case) is regressed each time including a different control while checking how much the magnitude of the coefficient of interest decreases. Yet, taking control covariates singularly delivers neat but not particularly informative correlations, due to an evident problem of omitted variable bias.

On the other hand, any multivariate analysis method can be quite problematic as well: socio-demographic and economic variables are highly multicollinear, especially at local level, and this is likely to both jeopardise inference and to make estimates interpretation difficult. By virtue of the richness of the LOL dataset we can provide evidence of the magnitude of this issue; Figure 1 in Appendix A.1 shows just how collinear our full dataset is by plotting all covariates between which strong correlations exist (i.e. of absolute value over 0.7). Given the high number of available covariates, correlations of lower magnitude are not plotted here because including even just moderate ones (i.e. lowering the inclusion threshold to 0.5) would drastically reduce readability. Indeed, 48 out of our 104 variables exhibit at least one correlation stronger than 0.7, while 79 have at least one correlation stronger than 0.5. Hence, we argue that when using granular data and when lacking a proper instrument for the variables of interest, the exercise of tracking causality down to single socio-demographics and economic factors is flawed, and that many papers in the literature underestimate the problem of such highly correlated regressors.

4.2 Principal Component Analysis

For the aforementioned reasons it is difficult to identify single economic and socio-demographic characteristics of peripheral areas which cause higher mortality; we nonetheless try to narrow down the field to few dimensions that explain most of the variance in excess mortality, and then analyze how they relate to the centre-periphery gradient. First, we perform Principal Component Analysis (PCA) on the dataset (the Singular Value Decomposition variety) and use its result to model excess mortality with the goal of interpreting the highest weights of original covariates in the orthogonal components which explain most of its variance. We then fit a proportional-odds model for the municipality class using the components which are found to be significant in modelling excess mortality in order to identify which ones are also capable of discriminating between classes.

PCA iteratively selects covariate linear combinations with maximal variance with the constraint of being orthogonal to the previous ones. These linear combinations (rather, their coefficient vectors) are called Principal Components (PCs). This method's sensitivity to covariate scale differences is remedied by stan-

standardising (centring and scaling) the dataset. PCA requires all observations to be complete, therefore, in order not to drop observations which have very few missing values, full imputation was performed by Data INterpolating Empirical Orthogonal Functions (DINEOF) Beckers and Rixen (2003) given the method's direct use of empirical orthogonal functions (like PCs) and its spatial statistics original context. The output of a PCA is as many PCs as there are original covariates, but the usefulness lies in being able to select a smaller subset without losing significant information. The most common criterion with which to perform this selection is by proportion of retained variance, i.e. selecting the first n PCs such that together they account for at least $X\%$ of the original dataset's variance; the threshold of 90% was chosen in our case.

The result of applying this procedure to the dataset of high-infection provinces was a dataset of 44 realised PC scores in place of the original 104 covariates, representing their linear combinations of greatest variance. Although analysing the top PCs is a worthy pursuit on its own (i.e. understanding what combinations of covariates constitute the most variation in the dataset), what is relevant to this work is analysing only the ones which are also significant in modelling excess mortality. For this purpose, linear models were iteratively fit to the new dataset by Maximum Likelihood Estimate in backwards steps of model selection: starting from a model with all 44 PCs, a model with one PC fewer is fit and compared with the previous one in order to determine whether the step is statistically justified (ordered candidate models are generated from the current one in a nested iteration by removing from it a single PC, trying them in order of decreasing within-model t-test p-value, but the main iteration moves forward only when an F-test against the current model succeeds).

Relevant metrics with which to compare models are the following: the proportion of original dataset intrinsic variance captured by the used PCs (indicating how much/little information is leveraged), the Bayesian Information Criterion (BIC) score (a relative model selection score balancing closeness-of-fit with complexity, where lower values indicate better models), and the prediction Mean Absolute Error (MAE) (though prediction is not the purpose of this analysis, it is relevant for empirical closeness of fit). The starting model contained 44 PCs (90.2% orig.var.) and had 9839 BIC and 1.87 MAE, while the refined model contains 12 PCs (27.7% orig.var.) and has 9647 BIC and 1.89 MAE, thus providing essentially the same quality of fit by using less than a third of the available dataset variance.

Reporting detailed interpretations of these components is impractical (they are 12 linear combinations of 104 covariates after all), therefore only the most significant 4 are analysed here; they capture 16.2% of the original dataset variance, and if put in a model on their own they achieve a BIC of 9694 and a MAE of 1.95. The components and their model coefficients are reported below in order of captured variance proportion (i.e. their numbering) for general simplicity and because their significances are roughly the same; Figure 8

shows the highest original-covariate weights (“loadings”) within each of them. Remembering that since the original covariates are standardised (centred and scaled) their lower/higher values truly are negative/positive (and not 0), looking at said weights it can be seen that:

Figure 8: Highest covariate loadings of the most significant PCs in predicting excess mortality



Notes. Blue loadings are positive and red loadings are negative.

C2 (-0.125 , 10.8% orig. var.) places high and positive weight on income-based covariates (red_* are also income-related) and education/development ones; given its negative sign, lower incomes and education correspond to higher excess mortality.

C7 (-0.276 , 2.75% orig. var.) places most weight on family/household-related covariates, along with a contrast between trade and industrial employment; given its negative sign, larger households and higher industrial employment correspond to higher excess mortality

C10 (0.242 , 1.97% orig. var.) places most weight on youth and industrial employment along with a contrast of some age-based demographics; given its positive sign, higher youth activity rate, lower turnover index⁵, higher incidence of industrial activity and lower 45-54-range vs 65-69-range population correspond to higher excess mortality

⁵This is defined by as the ratio of workers over 45 to workers aged 15-29

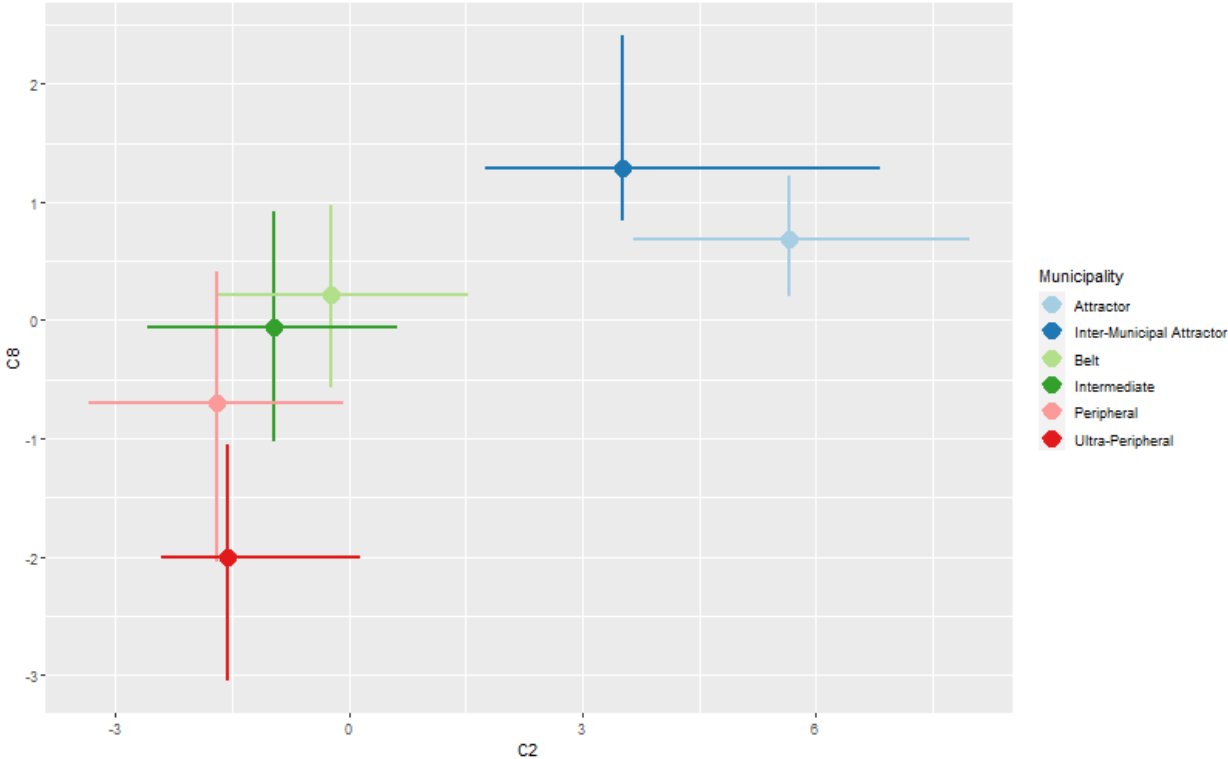
C35 (-0.557 , 0.616% orig. var.) places most weight on some age-range contrasts and the presence of small young families, however looking at the pattern of lower-loading covariates (not depicted) it becomes clear that this component is in fact representing population extremes: young families and everyone over 80 years old; the peculiar contrasts of age-ranges between these two extremes are most likely an artifact of the limited demographic combinations the used covariates capture; given its negative sign, more young families and very old population correspond to higher excess mortality

Summarising, it appears that the most severely hit municipalities within high infection provinces are those with lower income & education (C2), larger family sizes & prevailing industrial employment (C7), younger population & higher youth employment (C10) and more young families & very old population (C35). These findings are consistent with part of the aforementioned literature, especially with respect to the role of the population demographic profile and income levels. We highlight that the different sectoral composition of the economy, so far studied only in relation to the heterogeneous effect of the lockdown in Italy, is also significant in the very initial spread of the virus. More generally, even if it is not possible to claim any causal link or deterministic mechanism, the correlations emerging from our analysis of significant PCs are coherent and hint at a story consistent with the idea of Covid-19 hitting less developed and peripheral areas harder, as discussed in previous sections.

In order to see how our previous classification and findings on the centre-periphery gradient relate with the 12 PCs which are significant in predicting excess mortality, we use the latter to fit a proportional-odds model of the 6 municipality classes, with the aim of characterising central and peripheral areas with the variables considered in the second part of the previous analysis. The final proportional-odds model (refined as in the previous analysis) shows that out of the 12 PCs which are significant in predicting excess mortality only 7 are also significant in discriminating between classes, and the previously-described top-4 (C2 - income/education, C7 - family size & employment sector, C10 - youth presence/employment, and C35 - youth/elderly presence) are among them, though not in the same order. The centre-periphery gradient defined in Section 2 appears to indeed relate well with PCs which hint at less developed areas. Figure 9 provides some graphical intuition of the above by showing medians and inter-quartile ranges for each municipality class plotted over the space of the two PCs which are most significant in discriminating between them: C2 and C8 (both contributing negatively, i.e. towards *attractors*, and accounting for 13.3% of the original variance). C2 was described above, so here lower incomes and education correspond to higher periphery, while C8 places most weight on 50-65-range population and business income/tax (all negative), meaning that more older-working-age population and lower business presence correspond to more peripheral areas.

The separation between classes is evident. It is also worth noting that the two components better distinguish between classes at opposite ends of the scale (i.e. C2 separates attractors better than peripherals and vice-versa for C8), thus also highlighting which characteristics matter most along the spectrum. The directions in which these two PCs contribute with respect to excess mortality and periphery degree are consistent, showing how correlated they are to similar economic and socio-demographic features, and also acting as yet another verification that excess mortality is higher in more-peripheral municipalities (the bottom-left region of the plot)⁶.

Figure 9: Municipality classes over significant PCs



Notes. The plot reports municipality class medians and inter-quartile ranges over the two PCs which best model class among those which are significant in predicting excess mortality.

5 Conclusion

In this work we explore the existence and the magnitude of a geographic centre-periphery gradient in the harshness of the first wave of Covid-19 in Italy. Firstly, we find that peripheral municipalities were hit harder than central ones in March 2020: this result enriches the current literature understanding of the relationship

⁶As a final corroboration we run a sensitivity analysis to the addition of each single covariate in Appendix Table A.4.

between Covid-19 and location.

Secondly, thanks to the rich dataset we leverage, we can show that there is severe multicollinearity between economic and socio-demographic variables at municipal level, suggesting caution on previous literature which considers such factors without a solid strategy for causality.

Thirdly, we deploy a dimensionality reduction approach to explore the mechanisms behind the centre-periphery gradient, and we show that peripheral areas being highly affected mainly correlates with lower income, lower education, larger households, lower trade and higher industrial employments, and older population. We believe that for the time being it is not possible to disentangle the effects of these features on the spread of the virus (i.e. on contagions) from those on its mortality (i.e. on deaths). This is due first and foremost to the unreliability and lack of contagion data, especially during the first wave of Covid-19. Secondly, using aggregate data to this end might suffer the limitations of ecological inference. Hence, in order to better understand the mechanism behind the aggregate-level relationships we highlight, individual level data is of paramount importance.

In our view, further research of chief interest would be that exploring the existence of the centre-periphery gradient in other countries and that analysing heterogeneities during the subsequent waves of Covid-19 to see whether the same results hold. This is crucial in determining whether such a gradient is somehow structural or was instead determined by the unpreparedness of many countries during the first wave. Such work might also leverage more reliable swab testing data, and hence reasonably distinguish between a centre-periphery gradient in contagions (measuring the spread of the virus) and one in terms of deaths (measuring the mortality of the virus). The latter aspect will be essential in future research, as it will also help in understanding the role played by the economic and socio-demographic factors associated with peripheral areas and in designing better suited policy responses. Separately, future research is needed to explore the role in the spread of Covid-19 of those local features when taken singularly, exploiting *ad hoc* and solid identification strategies.

We believe that our work is also highly relevant from a policy-making standpoint. In particular, our analysis reinforces the idea that location-based policies represent valid tools to tackle Covid-19 and its consequences. As already shown in literature, local and spatially targeted lockdowns might prove to be more effective and less costly (Fajgelbaum, Khandelwal, Kim, Mantovani and Schaal, 2020; Karatayev, Anand and Bauch, 2020). During the second wave of Covid-19 in autumn 2020, Italy implemented local lockdowns region by region; this level still seems too aggregated, since, as shown in this work, a lot of variability exists even within single regions. Implementing sheltering orders at the level of local labour market might be a wiser approach (Tortuga, 2020), even if fine-tuning of these policies is still required (Fontán-Vela, Gullón

and Padilla-Bernáldez, 2021). The same line of reasoning applies to vaccination campaigns, which should primarily target those who are exposed to a higher risk: in light of our results, governments should prioritize peripheral municipalities, which appear to be the most fragile and least equipped to cope with the spread of the virus.

The place-based policy approach is also relevant from a post-pandemic recovery point of view: while governments implement unprecedented fiscal stimuli, they might want to address peripheral areas first and invest in spatially targeted policies aimed at mitigating the centre-periphery gradient we highlighted. Most importantly, interventions might regard health-care infrastructures, which in Italy can be reinforced following the *hub-and-spoke* approach, which is particularly suited for addressing the needs of peripheral areas (Elrod and Fortenberry, 2017) and consistent with the results of our analysis. A *hub-and-spoke* organization involves the establishment of a main campus (or hub), which supplies the most intensive medical services, complemented by satellite campuses (or spokes), which offer more limited service arrays at sites distributed across the served market. Basic healthcare needs are addressed locally through the network’s satellite facilities, but in cases where more intensive medical interventions are required, patients are routed to the main campus or hub for treatment. This approach has been advocated in the Italian debate (CERGAS, 2020) and is indeed part of the Italian Government strategy for the post-pandemic recovery. As the post-pandemic world creates new scenarios with respect to the geographical distribution of work (affected by the rapid spread of remote-working), adequate policy responses might exploit such a trend to foster the development of peripheral areas. This would help to mitigate and counterbalance the relatively higher losses that peripheral areas suffered during the pandemic, as highlighted in our work, and it would be an approach consistent with the ongoing efforts to revitalize rural and mountainous areas in Italy. Investments might prioritize better education and broadband infrastructure, with the aim of attracting highly-skilled workers in service-oriented occupations. As we have shown, more severely hit areas feature higher share of manufacture-oriented activities, which is consistent with the fact that these occupation are more exposed to contagion risk (Basso, Boeri, Caiumi and Paccagnella, 2020). On the whole, our work highlights the importance of the centre-periphery gradient, which seems to encompass other gradients proposed in the literature and offers a straightforward reading key to study the impact of Covid-19 and its consequences, from both a positive and normative standpoint.

References

- Abadie, A., Athey, S., Imbens, G. W. and Wooldridge, J. (2017), ‘When should you adjust standard errors for clustering’, *National Bureau of Economic Research* .
- Alacevich, C., Cavalli, N., Giuntella, O., Lagravinese, R., Moscone, F. and Nicodemo, C. (2020), ‘Exploring the relationship between care homes and excess deaths in the covid-19 pandemic: Evidence from italy’, *IZA discussion paper* (13492).
- Armillei, F. and Filippucci, F. (2020), ‘The heterogenous impact of covid-19: Evidence from italian municipalities’, *Working Paper - Local Opportunities Lab* .
- Bartscher, A. K., Seitz, S., Siegloch, S., Slotwinski, M. and Wehrhöfer, N. (2020), ‘Social capital and the spread of covid-19: insights from european countries’, *Covid economics* (26).
- Basso, G., Boeri, T., Caiumi, A. and Paccagnella, M. (2020), ‘The new hazardous jobs and worker reallocation’, *OECD Social, Employment and Migration Working Papers* .
- Beckers, J.-M. and Rixen, M. (2003), ‘Eof calculations and data filling from incomplete oceanographic datasets’, *Journal of Atmospheric and Oceanic Technology* **20**(12), 1839–1856.
- Bertrand, M., Duflo, E. and Mullainathan, S. (2004), ‘How much should we trust differences-in-differences estimates?’, *The Quarterly journal of economics* **119**(1), 249–275.
- Borjas, G. J. (2020), ‘Demographic determinants of testing incidence and covid-19 infections in new york city neighborhoods’, *National Bureau of Economic Research* .
- Brandily, P., Brébion, C., Briole, S. and Khoury, L. (2020), ‘A poorly understood disease? the unequal distribution of excess mortality due to covid-19 across french municipalities’, *medRxiv* .
- Brodeur, A., Gray, D., Islam, A. and Bhuiyan, S. J. (forthcoming), ‘A literature review of the economics of covid-19’, *Journal of Economic Surveys* .
- Carozzi, F., Provenzano, S. and Roth, S. (2020), ‘Urban density and covid-19’, *IZA Discussion Papers* .
- CERGAS (2020), ‘Oasi report 2020’.
URL: <https://www.cergas.unibocconi.eu/observatories/oasi/oasi-report-2020>
- Chetty, R., Hendren, N. and Katz, L. F. (2016), ‘The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment’, *The American Economic Review* **106**(4), 855–902.
- Clay, K., Lewis, J. and Severnini, E. (2019), ‘What explains cross-city variation in mortality during the 1918 influenza pandemic? evidence from 438 u.s. cities’, *Economics Human Biology* **35**, 42–50.
- Coccia, M. (2020), ‘Factors determining the diffusion of covid-19 and suggested strategy to prevent future accelerated viral infectivity similar to covid’, *Science of The Total Environment* **729**.

- Desmet, K. and Wacziarg, R. (2020), ‘Understanding spatial variation in covid-19 across the united states’, *National Bureau of Economic Research* .
- di Porto, E., Naticchioni, P. and Scrutinio, V. (2020), ‘Partial lockdown and the spread of covid-19: Lessons from the italian case’, *IZA discussion paper* (13375).
- Elrod, J. K. and Fortenberry, J. L. (2017), ‘The hub-and-spoke organization design revisited: a lifeline for rural hospitals’, *BMC Health Services Research* **17**(S4).
- Fajgelbaum, P., Khandelwal, A., Kim, W., Mantovani, C. and Schaal, E. (2020), ‘Optimal lockdown in a commuting network’, *Nber Working Paper Series* .
- Fenoll, A. A. and Grossbard, S. (2020), ‘Intergenerational residence patterns and covid-19 fatalities in the eu and the us’, *Economics Human Biology* **39**, 100934.
- Ferraresi, M., Kotsogiannis, C., Rizzo, L. and Secomandi, R. (2020), ‘The ‘great lockdown’ and its determinants’, *Economics Letters* **197**, 109628.
URL: <https://www.sciencedirect.com/science/article/pii/S0165176520303888>
- Fontán-Vela, M., Gullón, P. and Padilla-Bernáldez, J. (2021), ‘Selective perimeter lockdowns in Madrid: a way to bend the COVID-19 curve?’, *European Journal of Public Health* . ckab061.
URL: <https://doi.org/10.1093/eurpub/ckab061>
- Francesca Borgonovi, E. A. and Subramanian, S. V. (2020), ‘Community-level social capital and covid-19 infections and fatality in the united states’, *Covid Economics* (32).
- Gerritse, M. (2020), ‘Cities and covid-19 infections: Population density, transmission speeds and sheltering responses’, *Covid economics* (37).
- Hamidi, S., Ewing, R. and Sabouri, S. (2020), ‘Longitudinal analyses of the relationship between development density and the covid-19 morbidity and mortality rates: Early evidence from 1,165 metropolitan counties in the united states.’, *Health Place* **64**, 102378.
- Hamman, M. K. (2021), ‘Disparities in covid-19 mortality by county racial composition and the role of spring social distancing measures’, *Economics Human Biology* **41**, 100953.
- Isphording, I. E. and Pestel, N. (2020), ‘Pandemic meets pollution: Poor air quality increases deaths by covid-19’, *IZA Discussion Papers* .
- Ivan Franch-Pardo, Brian M. Napoletano, F. R. and Billa, L. (2020), ‘Spatial analysis and gis in the study of covid-19. a review’, *Science of The Total Environment* **739**.
- Kapoor, R., Rho, H., Sangha, K., Sharma, B., Shenoy, A. and Xu, G. (2020), ‘God is in the rain: The impact of rainfall-induced early social distancing on covid-19 outbreaks’, *Covid economics* (24).
- Karatayev, V. A., Anand, M. and Bauch, C. T. (2020), ‘Local lockdowns outperform global lockdown on the far side of the covid-19 epidemic curve’, *Proceedings of the National Academy of Sciences*

117(39), 24575–24580.

- Knittel, C. R. and Ozaltun, B. (2020), ‘What does and does not correlate with covid-19 death rates’, *National Bureau of Economic Research* .
- Kuchler, T., Russel, D. and Stroebel, J. (2020), ‘The geographic spread of covid-19 correlates with structure of social networks as measured by facebook’, *National Bureau of Economic Research* .
- Lee, D. (2013), ‘Carbayes: An r package for bayesian spatial modeling with conditional autoregressive priors’, *Journal of Statistical Software* **55**(1), 1–24.
- Leroux, B. G., Lei, X. and Breslow, N. (2000), ‘Estimation of disease rates in small areas: A new mixed model for spatial dependence’, *Statistical Models in Epidemiology, the Environment, and Clinical Trials The IMA Volumes in Mathematics and its Applications* p. 179–191.
- Markowitz, S., Nesson, E. and Robinson, J. J. (2019), ‘The effects of employment on influenza rates’, *Economics Human Biology* **34**, 286–295.
- Matthew A. Cole, Ceren Ozgen, E. S. (2020), ‘Air pollution exposure and covid19’, *IZA discussion paper* (13367).
- Michelozzi, P., De’Donato, F., Scortichini, M., Pezzotti, P., Stafoggia, M., Sario, M. D., Costa, G., Noccioli, F., Riccardo, F., Bella, A. and et al. (2020), ‘Temporal dynamics in total excess mortality and covid-19 deaths in italian cities’, *BMC Public Health* **20**(1).
- Olden, A. and Møen, J. (2020), ‘The triple difference estimator’, *NHH Dept. of Business and Management Science Discussion Paper* (2020/1).
- Richterich, P. (2020), ‘Severe underestimation of covid-19 case numbers: effect of epidemic growth rate and test restrictions’, *medRxiv* .
- Sa, F. (2020), ‘Socioeconomic determinants of covid-19 infections and mortality: Evidence from england and wales’, *IZA policy papers* .
- Sawano, T., Kotera, Y., Ozaki, A., Murayama, A., Tanimoto, T., Sah, R. and Wang, J. (2020), ‘Underestimation of covid-19 cases in japan: an analysis of rt-pcr testing for covid-19 among 47 prefectures in japan’, *QJM: An International Journal of Medicine* **113**(8), 551–555.
- Sussman, N. (2020), ‘Time for bed(s): Hospital capacity and mortality from covid-19’, *Covid economics* (20).
- Tortuga (2020), Fase 2: Sistemi locali del lavoro. https://www.tortuga-econ.it/wp-content/uploads/2020/05/SLL-REPORT-WORD-aggiornato-10-aprile_v2.pdf.

A Appendix

A.1 Additional Tables and Figures

Table A.1: Placebo Results

Triple-Interaction Value	(1) Death_rate	(2) Death_rate	(3) Death_rate
Inter-Municipal Attractor # post # high infection	0.00469 (0.0586)	0.0226 (0.0505)	0.0199 (0.0578)
Belt # post # high infection	-0.0469 (0.0325)	0.0345 (0.0295)	0.0239 (0.0306)
Intermediate # post # high infection	-0.0235 (0.0508)	0.0624 (0.0457)	0.0129 (0.0465)
Peripheral # post # high infection	-0.0575 (0.0804)	0.0634 (0.0703)	0.0512 (0.0948)
Ultra-Peripheral # post # high infection	0.0757 (0.210)	-0.0956 (0.199)	0.0125 (0.158)
Observations	15,602	15,602	15,602
R-squared	0.219	0.234	0.234
Unit FEs	LLM	LLM	LLM
Cluster level	Municipal	Municipal	Municipal
Controls	Over 80	Over 80	Over 80
Placebo year	2017	2018	2019

Notes. The table presents the results of placebo analysis. The table reports only the coefficients of the triple-interaction term, where the attractor class works as baseline. Clustered standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: Spatial analysis results

Triple-Interaction Value	(1) Death_rate
Inter-Municipal Attractor # post # high infection	-0.270
Belt # post # high infection	0.064
Intermediate # post # high infection	0.558**
Peripheral # post # high infection	1.384***
Ultra-Peripheral # post # high infection	0.308
Observations	7801
Controls	Over 80

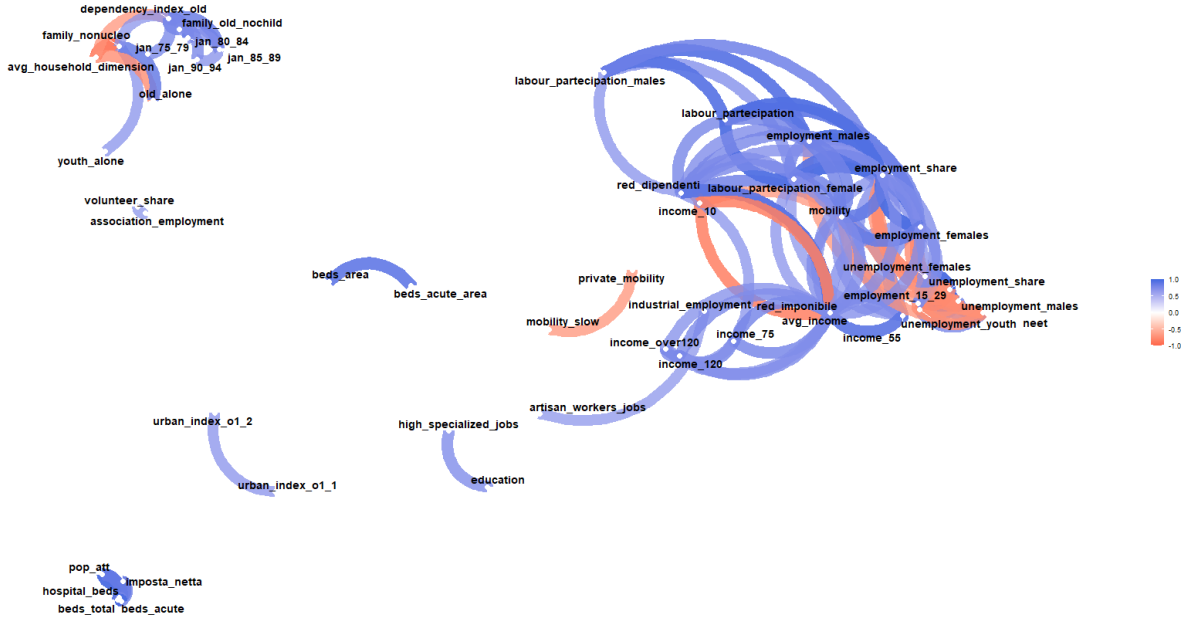
Notes. The table presents the result for the spatial analysis using a CAR Leroux model. The model is estimated on the delta in mortality between March 2020 and the average of March 2017-2019. The model is estimated by MCMC. In this Bayesian setting we use the graphical symbol of the star with the following meaning: *** when 0 is not contained in the 99% credible interval; ** when 0 is not contained in the 95% credible interval; * when 0 is not contained in the 90% credible interval.

Table A.3: Robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Triple-Interaction Value	Death_rate	Death_rate	Death_rate	Death_rate	Death_rate	Death_rate	Death_rate
Inter-Municipal Attr. # post # high inf.	-0.112 (0.406)	-0.131 (0.415)	-0.124 (0.419)	-0.176 (0.407)	-0.384 (0.675)	-0.116 (0.428)	-0.343 (0.338)
Belt # post # high infection	0.330 (0.287)	0.282 (0.286)	0.249 (0.289)	0.287 (0.308)	0.462 (0.568)	0.373 (0.303)	0.0995 (0.295)
Intermediate # post # high infection	0.645* (0.349)	0.594* (0.354)	0.598* (0.352)	0.688* (0.357)	0.943 (0.613)	0.727** (0.353)	0.276 (0.351)
Peripheral # post # high infection	1.242*** (0.410)	1.166*** (0.410)	1.132*** (0.404)	0.998** (0.398)	1.278** (0.640)	1.247*** (0.404)	1.119** (0.480)
Ultra-Periph. # post # high infection	0.486 (0.617)	0.632 (0.757)	0.483 (0.717)	0.554 (0.855)	0.296 (0.986)	0.566 (0.852)	0.884 (0.832)
Observations	15,602	15,602	15,602	15,602	15,602	15,602	15,602
R-squared	0.696	0.690	0.694	0.802	0.813	0.805	0.779
R-squared	0.462	0.462	0.462	0.802	0.813	0.805	0.779
Unit FEs	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality
Cluster level	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality	Municipality
Controls	Over 80	Over 80	Over 80	Over 80	Over 80	Over 80	Over 80
Base year	2017	2018	2019				
High inf. threshold				> 70 th percentile	> 80 th percentile	Outlier rule	
High inf. definition							February contagion

Notes. The table presents the results of the triple-difference analysis of different specifications. We regress mortality in each municipality on a set of dummy variables accounting for: i) time (2020 vs base year specified in the table) ii) infection intensity of Covid-19 (high vs low infection, defining the groups according to the threshold specified) iii) degree of centre-periphery (six classes). All the interactions among these dummies are also included. We also include interactions between the specified controls and the infection and time dummies. The table reports only the coefficients of the triple-interaction term, where the attractor class works as baseline. Coefficients are estimated with OLS. Observations are weighted by municipal population. Clustered Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure 1: Variable correlations of absolute value over 0.7



Notes. This figures visualises the correlations among the variables in our dataset. Every point corresponds to a variable and every link shows the presence of a correlation of absolute value over 0.7. Blue links imply a positive correlation, while red links a negative one. The darker the link, the stronger the correlation.

A.2 Sensitivity Analysis Adding Single Control Variables

A final step in order to highlight potential mechanisms behind the centre-periphery gradient we describe is to run a sensitivity analysis. This means checking by how much the effect of the Periphery dummy on excess mortality (in Table 3) changes when adding single covariates. More precisely, we add covariates to the triple-difference model in Equation 1 in the same way as we added controls for share of over 80 S_i : as simple values, as interaction with the high-infection dummy, as interaction with the post dummy, and as triple interaction between the control variable, the high infection dummy and the post dummy.

Table A.4 reports (in columns 1 and 2) the coefficients of the triple interaction of the control variable and of the periphery dummy when both are included in the triple difference model (with municipality FE). Column 3 instead simply reports the baseline results (as in Table 3, columns 3-4). We run this exercises for all variables summarized by Table 2, but report only the ten variables for which the coefficient in col-

umn 2 falls the most. These variables are to be considered as the ones which capture most of the variation described in our centre-periphery gradient, hence indicating which could be the most important mechanisms.

Although this approach is simpler, the interpretation of covariate coefficients may suffer from omitted variable bias; this is why we prefer to use the formal dimensionality reduction approach reported in Section 4. In any case, the results of the two methods are consistent: using the sensitivity analysis, we find that youth employment, employment in industry and artisan jobs, and low education correlate significantly to higher excess-mortality and entail the largest drops in the coefficient of the Periphery dummy.

Table A.4: Horse-race between mechanism variables

Added covariate	(1) Covariate	(2) Periphery dummy, baseline	(3) Periph. dummy, controlled spec.
active_inactive_ratio_youth	0.031 [0.825]	0.222 [0.826]	1.182 [0.814]
artisan_workers_jobs	0.065 [0.825]	0.018 [0.826]	1.182 [0.814]
education	-0.006 [0.822]	0.498 [0.824]	1.182 [0.814]
employment_15_29	0.110 [0.825]	0.006 [0.825]	1.182 [0.814]
family_old_nochild	-0.264 [0.824]	0.515 [0.826]	1.182 [0.814]
industrial_employment	0.059 [0.829]	0.030 [0.830]	1.182 [0.814]
labour_partecipation_female	-0.186 [0.829]	0.357 [0.830]	1.182 [0.814]
service_employment	-0.047 [0.825]	0.047 [0.826]	1.182 [0.814]
turnover_index	-0.010 [0.829]	0.090 [0.829]	1.182 [0.814]
unemployment_youth	-0.094 [0.825]	0.260 [0.825]	1.182 [0.814]

Notes. The table presents the results of the horse-race between potential mechanisms, selecting the ten covariates reporting the largest impact on baseline coefficients of the Periphery dummy (i.e. those for which the estimate in column 2 is smaller). Column 1 reports the coefficient of a regression of death rate on the interaction between the covariate, high-infection dummy and 2020 dummy, plus municipality and time FEs. Column 2 reports the coefficient of the interaction between the dummy for Peripheral municipality, high-infection dummy and 2020 dummy in a regression which also includes the interaction between the covariate, high-infection dummy and 2020 dummy, plus municipality and time FEs. Column 3 reports baseline results for the dummy for Peripheral municipality as in Table 3. R squares are reported in square brackets.