Article:
Rudkin, S., Webber, D. and Dlotko, P. (2026) The spatial spread of infection rates at the dawn of the pandemic: a socioeconomic approach. Papers in Regional Science. ISSN: 1056-8190

https://doi.org/10.1016/j.pirs.2026.100163

Journal Pre-proof

The spatial spread of infection rates at the dawn of a pandemic: a socioeconomic approach

Simon Rudkin, Don J Webber, Pawel Dłotko

PII: S1056-8190(26)00029-1

DOI: https://doi.org/10.1016/j.pirs.2026.100163

Reference: PIRS100163

To appear in: *Papers in Regional Science*

Received date: 9 September 2025
Revised date: 7 August 2026
Accepted date: 9 August 2026

Please cite this article as: Simon Rudkin, Don J Webber and Pawel Dłotko, The spatial spread of infection rates at the dawn of a pandemic: a socioeconomic approach, *Papers in Regional Science,* (2026) doi:https://doi.org/10.1016/j.pirs.2026.100163

The spatial spread of infection rates at the dawn of a pandemic: a socioeconomic approach

Simon Rudkin
University of Manchester, UK

Don J Webber
The University of Sheffield, UK
Email: d.j.webber@sheffield.ac.uk
Corresponding Author

Pawel Dłotko
Dioscuri Centre in Topological Data Analysis, Poland

Abstract

This paper presents an innovative examination of the spread of COVID-19 infection rates across England over a seven-week period commencing prior to the initial lockdown and continuing until the relaxation of some COVID-19 lockdown restrictions. We examine the evolution in infection rates across socioeconomic characteristics using topological data analysis to map the joint distribution of socioeconomic characteristics and identify where and when infection rates grew differentially. Although the over 65s and those with chronic health conditions were socioeconomic groups most at risk of experiencing morbidity and mortality from COVID-19, these socioeconomic correlates reflect consequences of the spread rather than the carriers of the virus and the reasons for the spread. Our empirical results present new findings that International Territorial Level 3 areas with higher proportions of younger inhabitants who worked longer hours for higher pay experienced earlier and faster rates of increase in infections, and that commuter towns and rural areas lagged their urban metropolises in terms of infection rates. These results underscore *who* was infected, *where* labour market characteristics associated with infection rates interact across space, and the need to improve understanding of *when* and *why* areas experienced higher infection rates. This is the first paper to find that the spread of the COVID-19 virus in the UK was greater where there was a relative abundance of skilled workers who have greater professional geographical mobility.

Keywords: COVID-19; Spatial contagion; Topological Data Analysis; Virus infections rates.

JELs: I18; C65; R58.

1. Introduction

The infrequency of pandemics means that we must learn as much as we can from each occurrence to enable us to guide policymaking and allocate resources efficiently and effectively to minimize resulting morbidity and mortality. Viruses impose enduring costs on society, such as greater use and depletion of health care resources, loss of stamina and agility due to morbidity, impaired economic growth due to time away from the purposeful accumulation of knowledge, shorter attention spans that reduce the efficiency of learning by doing, depressed social interactions during societal recovery with commensurate reductions in spending and consumption, and a permanent loss to potential output due to premature death. Although each pandemic can differ with respect to the rate of virus contagion and the probability of fatality, underpinning the likelihood of contagion, morbidity, and mortality are primarily socioeconomic characteristics and health status (Link and Phelan, 1995; Phelan et al., 2010). Guidance on how to reduce transmission and severity of impact of any virus is therefore of immense interest to regional development and health-focussed policymakers.

Socioeconomic status data are known to be useful when predicting the static distribution of COVID-19 infections and deaths (Consolazio et al., 2020; Hawkins, 2020; Andersen et al., 2021; Clouston et al., 2021; Ehlert, 2021; Olmo and Sanso-Navarro, 2021). Regions possess distinct combinations of industrial structures, transport links, workforces, and populations that are woven into the fabric of human geography (Iammarino and McCann, 2006), and it is these socioeconomic factors that underpin interregional mobility and human interaction (Yin et al., 2017) that spread viruses (Du and Holme, 2020). In the UK, the over 65s and those with chronic health conditions were the socioeconomic groups most at risk of experiencing morbidity and mortality due to COVID-19, but these socioeconomic correlates reflect consequences of the spread of the virus rather than the carriers of the virus; greater understanding is urgently needed on the socioeconomic and labour market characteristics associated with the evolutionary spread of the virus. In this paper we present an empirical assessment of the socioeconomic factors associated with the early temporal and spatial spread of COVID-19 across England's International Territorial Level 3 (ITL3)[1] areas, which could prove very useful in preventing the spread of viruses in the future. To achieve this research objective, we drew on data covering a seven-week period commencing prior to the initial lockdown and continuing until the UK Government's initial relaxation of some lockdown restrictions; crucially, data over this period also reflect the effectiveness of policies to halt the initial spread of the virus.[2]

Evolving contexts and differences between pandemics makes it difficult to deduce the correct functional form for any deterministic model used to predict the spatial spread of infection rates (Shi et al., 2020). Instead of guessing or using trial and error with in-sample data to identify the most suitable functional form for a model used to explain the temporal and spatial spread of a virus, we adopt two coalescing approaches to add rigor. First, we estimated descriptions of the spatial spread of infection rates using Global Moran's I, Geary's C, LISA, bivariate Moran, and bivariate LISA to identify spatial clustering, the co-movement between variables, and the presence of local hotspots. We then applied the novel Topological Data Analysis Ball-Mapper (TDA-BM) algorithm of Dlotko (2019), which again requires no assumptions relating to the functional form of the underlying deterministic model that would otherwise be needed in a regression approach. The TDA-BM technique is used to map the joint distribution of socioeconomic characteristics and then identify whether these were associated with the spatial asymmetries in the rate of growth of infection rates: that is, we use spatial descriptive statistics (LISA etc.) and newly developed statistical techniques (TDA-BM) to identify what the data actually shows without needing to know the true statistical form.[3] Critically, TDA-BM provides the ability to see the evolution of COVID-19 cases across the full joint distribution of variables, and this approach builds on the descriptive

1 International Territorial Levels (ITL) have replaced the Nomenclature d'Unités Territoraiales Statistiques (NUTS) levels that were used by the European Union to divide countries into spatial units.

2 In contrast, Olmo and Sanso-Navarro (2021) examined data on the second wave of the COVID-19 pandemic.

3 See Rudkin (2025) for an introductory guide to TDA-BM and its implementation in the *R* software.

spatial visualisations. Adoption of this combined approach enables us to make three novel contributions to the literature on the geographical spread of the COVID-19 disease.

First, we assert the importance of comprehending complex socioeconomic spatial heterogeneities to understand the spatial evolution of virus infection rates. Our analysis reveals nuance spatial patterns of socioeconomic characteristics associated with the speed and spatially evolving incidence of infections. Second, we find infection rates in commuter towns and rural areas lagged their metropolises, suggesting that spatial associations in the spread of the virus relating purely to transportation flows and physical geography were not of primary importance. Third, and crucially, we provide evidence suggesting that the spatial pattern of labour market characteristics correlates strongly with the spatial spread of infection rates. We contend that these results highlight *who* were infected and *where* they were affected, as labour market characteristics correlate with infection rates across space. Our results emphasise the urgent need to improve understanding of the mechanisms driving *when* and *why* areas experienced higher infection rates. Although the results highlight that high-wage, high-skilled workers were associated with the spread of the disease, what lies behind these socioeconomic correlates requires greater investigation, such as whether these correlations reflect greater connectivity and greater mobility of high-wage, high-skilled workers in areas that are more developed and more prosperous. Therefore, we contend, with hindsight, that policy attention to control the spread of COVID-19 should have been focussed on factors correlated with socioeconomic factors. These findings lead us to argue that designers of virus abatement policies need to be fully cognizant of socioeconomic factors, especially labour market characteristics, that vary across space and contribute to the interpersonal transfer of viruses.

The remainder of the paper is organised as follows. The next section summarises existing socioeconomically informed knowledge relating to COVID-19 infection rates. Sections 3 and 4 outline the data and methodology respectively. Section 5 presents results that illustrate how the power of the combined statistical approaches manifests within the study of spatial COVID-19 morbidity data. Section 6 discusses results and concludes.

2. Socioeconomics, mobility, and virus contagion

History teaches us that there are associations between socioeconomic characteristics and mortality during disease outbreaks, such as grave-digging practices during the plague in London (Howson, 1961; Harding, 1989; Mack, 1991), Black labour being employed intensively during pandemics in South Africa (Packard, 1989), and tribal traditions behind Ebola infections (Grépin et al., 2020). Deteriorations in health conditions of the poor are associated with greater exposures to foreign germs by the mobile rich (Kohn, 2007) and the international spread of disease is likely to be accelerated by interregional and international travel of the predominantly wealthy.

Fundamental cause theory (Link and Phelan, 1995; Phelan et al., 2010) suggests that socioeconomic characteristics affect how viruses spread, as those characteristics affect knowledge and understanding, money, power, and the beneficial social connections that positively affect health outcomes. For instance, Andersen et al. (2021) interrogated US county-level data to examine the contextual factors associated with higher rates of morbidity and mortality from COVID-19, and they identified clusters of infections and deaths that were driven by specific outbreaks in places like prisons and meat-packing plants, in areas where people speak English less well, and where there are higher rates of disabilities. Clouston et al. (2021) also examined US county-level data and found that counties with higher socioeconomic status populations experienced earlier infection cases, but once social distancing policies took place then the growth of infections was slower in higher socioeconomic status areas relative to counties with lower socioeconomic status. Hawkins (2020) and Andersen et al. (2021) show that people of colour were more likely to be employed in essential industries and in occupations with more exposure to other people in the UK, and they conclude that occupational segregation contributed to the differential risk of infection and death from COVID-19; however, they stopped short of affirming that the spatial and temporal spread of COVID-19 was due to issues related to occupational segregation.

The spread of the COVID-19 virus in the UK has been examined in various papers. Building on the recognition that person-to-person interaction was the main mechanism for viral transmission and that airborne transmission is the main route to infection spread, Gibson-Miller et al. (2024) found social distancing behaviour to be key in restraining virus contagion in the UK.[4] Rahaman and Barik's (2023) assessment stressed the role of people movement in spreading the infection and infection susceptibility, while Hadjidemetriou et al. (2020) showed that reductions in human mobility reduced deaths, thereby supporting Government lockdown measures. Relatedly, Purves et al. (2023) argued that non-compliance

[4] Instigation of social-distancing guidelines is influenced by the expected probability that they will be followed by the public, with greater personal freedom identified to be positively associated with COVID-19 infection rates (Ang et al., 2023) and greater conservatism and lower education associated with lower vaccination rates (Bourdin et al., 2023).

with COVID-19 mitigation measures at the EURO 2020 UEFA soccer tournament (held in the UK in 2021) led to a significantly higher risk of virus transmission.

Elsewhere, Shi and Liu (2020) interrogated Chinese data[5] and contend that the mass migration in China due to their Spring Festival should not be held responsible for the spread of COVID-19. They illustrate empirically that local mobility between contiguous provincial capitals could not explain the long-distance spread of the virus, typically associated with transportation routes, and instead they conclude that the major factor driving the spread of COVID-19 was businesspeople who commute between regional capitals for work, and not migrant job seekers or Chinese Spring Festival attendees who migrate locally and less frequently.

COVID-19 migration and distancing measures were crucial to stem the transmission of the virus.[6] However, there have been strong criticisms in the UK of policies designed to reduce mortality and morbidity that stemmed from COVID-19. For instance, Rietveld et al. (2024) conclude that the UK was ill-prepared to respond to a pandemic and Paton (2020) emphasise that the UK Government did not follow the science and therefore made incorrect policy decisions. Aboukorin et al. (2023) go as far to say that implementing stricter measures earlier in the pandemic would have reduced the overall impact of COVID-19, as would also have been the case in comparable countries.

2.1. COVID-19 infections and socioeconomic correlates

COVID-19 affected the demand for labour differently in different regions (Amaros et al., 2025). When age-related issues concerning COVID-19 are discussed, emphasis rests on the higher probability of individuals over the age of 65 being severely affected by the disease, who often are outside of the labour market. Literature stresses that the severity of the effects

[5] Cai and Mason (2022) emphasise that the strong support for COVID-19 restrictions in China was due to a combination of self-interest, national pride, and a conscious indifference to transparency of government policy. Peer effects shape behaviours and are contagious (Argys and Rees, 2008), so we need to understand how to strengthen the positive contagion of socially beneficial behaviours and reduce the negative contagion of socially detrimental behaviours, such as decisions to obey / disobey social-distancing guidelines.

[6] Pandemic-related COVID-19 health policies were built on existing knowledge of the roles of socioeconomic forces that hinder pathogen transfer and augment productivity. For instance, in an effort to identify whether specific social policies restrict the effect of influenza, Adda (2016) shows that the closure of schools reduces the incidence of influenza by three to four weeks (with this benefit not restricted to children as the favourable effect is also sizeable for children's guardians and their wider social networks) and that temporary shutdowns of transportation networks reduce pathogen transmission rates. Although this is useful knowledge, it is unclear whether there are further useful lessons from the COVID-19 pandemic that will enable us to diminish the damaging effects of future pandemics.

of COVID-19 infections on the person is related to the presence of chronic health conditions, and particularly for those over the age of 65 with chronic health conditions. Bearth et al. (2023) show that the acceptance of and compliance with public health measures in response to COVID-19 were higher among older adults and those with pre-existing health conditions, perhaps because they saw themselves more at risk of morbidity or mortality due to COVID-19, and compliance with restrictions may explain why Ang et al.'s (2023) analysis identified that older populations were associated with lower COVID-19 infection rates. As acceptance of and compliance with public health measures to restrict mobility were lower for younger adults, we therefore suggest that areas with lower median age were more likely to experience earlier increases in COVID-19 infection rates.

Greater urbanisation and higher levels of income inequality have both been found to be positively associated with pathogen transfer (Ang et al., 2023). Although some assume that lower paid workers[7] in essential services (e.g., bus drivers and food delivery services) could be major spreaders of the virus, this line of thinking contrasts with Shi and Liu's (2020) analysis, highlighted above, of the spread of COVID-19 in China who emphasised the importance of higher wage, higher skilled, highly mobile workers, such as businesspeople, in spreading the virus.

Immediately prior to the lockdown, and concomitant with initial increases in infection rates in the UK, the highest paid (typically higher skilled) workers were the most mobile and tended to have more and longer distance business trips as well as longer commutes (Morris and Zhou, 2018). Although the impending lockdown restricted the human-to-human transfer of respiratory pathogens, some of the spread of COVID-19 may have already occurred through human-to-human contact immediately preceding the lockdown. As high wages and high commuting costs interact positively to affect workers' optimal commuting times (van Ommeren and Reitveld, 2005), a priori reasoning leads to the expectation that areas with greater proportions of highly skilled, higher wage, more mobile workers immediately prior to the lockdown would have stronger associations with the spread of this virus.

Population density affects commuting times and the speed of commuting though congestion effects that coalesce with the location choice of industry and workplace

[7] The UK Government's furlough and COVID-19 loan schemes reversed preceding austerity measure through Keynesian fiscal policies which sustained some income levels for employees unable to work due to lockdown restrictions. This policy enabled a level of consumer spending that bolstered the economy in the time of crisis. Yet, while the Government stance and furlough and loan schemes softened the effect of the disease on many businesses and supported the rapid switch to online working, which was more achievable in some industries than in others (Dingel and Neiman, 2020), there continued to be work-related mobility of essential workers during the lockdown (e.g., health workers and bus drivers).

employment levels. Commuting patterns enhance the probability of interpersonal contact over longer distances and are associated with the development and size of the local economy, with car ownership, carsharing, and public transport usage increasing with a location's wealth (Santos et al., 2013). So although population density is a natural key to respiratory pathogen infection rates and must be explicitly considered in any pandemic analysis (Jones et al., 2008), and indeed Ang et al. (2023) and González-Val and Sanz-Gracia (2022) found greater urbanisation to be positively associated with COVID-19 infection rates, population density is not the only spatially-relevant socioeconomic factor that influences human-to-human contact, mobility, and pathogen transfer but is one that we must include in any such analysis.[8] Thus, we suggest that areas with higher population density offer more opportunities for pathogen transfer and are likely to be associated with earlier increases in the spread of COVID-19.

The economic wealth of an area is also likely to be associated with higher infection rates. With more densely packed leisure activities present within wealthier areas, such hubs of activity may focus the transmission of the COVID-19 virus. Similarly, areas with higher average hours of work, which were primarily in-person prior to the shutdown, are also likely to have experienced higher rates of person-to-person pathogen transfer.

Although areas with higher proportions of workers on low average weekly wages may be more likely to take on extra shifts and be less likely to stay home if they become sick and therefore may have been thought to be associated with greater *local* pathogen transfer, Shi and Liu (2020) show that the *spread* of COVID-19 in China was more likely to be due to higher wage businesspeople who commute between regions for work and are therefore likely to be in contact with the pathogen earlier than others and also more likely to transfer the virus between areas.

Thus, we build on pre-COVID-19 pandemic knowledge provided by Böckerman et al. (2009) and Barnay (2016) – which suggest that working patterns and age profiles affect pathogen spread – to investigate the correlations of hours worked, income, and age with the temporal and spatial evolution in infection rates. We also investigate whether the spread of the COVID-19 disease during the early stages of the pandemic was associated with the dynamism of the local economy, in terms of GDP, and the concentration of the local population, in terms of population density. By focusing on these socioeconomic factors, which themselves are driven by an array of factors, we raise attention of the effects and

[8] For instance, the tolerance towards higher commuting times is affected by age, gender, education, household income, and the presence of children (He et al., 2016).

importance of the spatial social economy on the geographical spread of COVID-19 infection rates.

3. Data and descriptive statistics

Like other countries, policy response to COVID-19 was slow in the UK. Initial attention focussed on reducing the growth of infections by developing herd immunity while preserving the operation of the economy, even though this went against World Health Organisation (WHO) advice. When the modelling of COVID-19 processes started to indicate that the number of deaths could approach 250,000, the UK Government changed stance and proceeded to lockdown UK society.[9] Instruction for residents to stay at home and only leave their homes for essential groceries or exercise came on 23rd March 2020, which was almost two months after the WHO declared COVID-19 a global pandemic. The UK Government instigated lockdown started nine days after COVID-19 infection data began to be collated at the local authority level across England making it possible to explore the potential contemporaneous and subsequent effects of labour market characteristics on an area's infection rates during the early days of the pandemic.

The period of analysis is crucial in this study, as our focus is on the initial spread of infection rates which of course is the most important part of any pandemic. After identifying the virus in 2019, COVID-19 was first detected on the shores of mainland UK in January 2020. There were no official recordings of COVID-19 related deaths in the UK until March 2020, but it is thought that the virus entered the UK via international travellers bringing the virus to the UK.[10]

The nature of data and policy devolution means that the geography of our empirical analysis focuses on England between March 14th and May 2nd, 2020.[11] The first date precedes the introduction of the initial legally enforced lockdown on 23rd March, which banned all

9 When the UK went into lockdown on 23rd March 2020, the UK Government introduced the Coronavirus Act 2020. The British military was mobilised to assist where necessary and the National Health Service expanded to try to cope with the wave of admissions, albeit at the cost of the postponement of prior scheduled operations and treatments.

10 Efforts to employ contact tracing to map the path of the spread of the disease were overwhelmed and became ineffective to inform policy (Du Plessis et al., 2021). Limited testing and tracing meant that infections were underestimated perhaps on a non-random basis, and this may have resulted in underreporting of COVID-19 infections by local authority districts.

11 We focused on English areas because the other parts of the UK are devolved and applied their own policies during the pandemic.

non-essential travel and contact with other people even if those people were part of one's immediate family; as a result, businesses and schools closed with some businesses never reopening. Everyone with symptoms were instructed to self-isolate, and this applied to the whole household after 16th March throughout the rest of this study period even if only one person was symptomatic in the household. Three weeks after the lockdown started reports emerged that these restrictions had "flattened the curve" of the pandemic and that the peak of the pandemic had passed after 26,000 deaths.[12] By May 3rd, the UK's overall death toll and death rate surpassed Italy's, which made the UK the worst affected country in Europe and one of the worst affected countries on earth. By January 2025, there were over 25 million confirmed infections and over 230,000 deaths from COVID-19 in the UK at a death rate of 409 per 100,000 inhabitants (Mathieu et al., 2026).

Data recording the number of COVID-19 infections across local authority districts were reported by the UK Government and collated for researchers on the Github website of Tom White (https://github.com/tomwhite/COVID-19-uk-data)[13]. The number of people locally carrying the virus is likely to be an underestimate of the actual number of local people infected with the virus, not least because some people did not have symptoms and therefore were unlikely to test whether they had COVID-19.[14] We accessed and then merged this COVID-19 data with local authority district level data on workplace-based earnings and hours worked sourced from the Annual Survey of Hours and Earnings (ASHE). Further data on population density, median age, and GDP data at ITL3 were obtained from Eurostat and then merged using lookup queries from the Office for National Statistics, supplemented by some manual work on cases where no match was found.[15]

[12] Due to the need to inform families of a person's death, it was possible that the official statistics were slightly inaccurate with delays of a few days possible.

[13] Data were obtained from the UK Government Gateway through scrapers developed for repository. The UK Government Gateway now only includes ITL3 level data for 1st April onwards, but the old data is available via the GitHub site used in this paper.

[14] Asymptomatic and pre-symptomatic patients are known transmitters of COVID-19 (Wilder-Smith et al., 2020). Previously infected individuals were less compliant with protective health behaviours (e.g., mask wearing and social distancing) rendering this population group a potential reservoir of viral spread (Kaim et al., 2021), and previously infected individuals who successfully recovered may have had a diminished perceived health risk to the virus (Kaim and Saban, 2022).

[15] Look up queries for Local Authority Districts to ITL3 level data are available from the Open Geography Portal of the Office for National Statistics at https://geoportal.statistics.gov.uk/datasets/0de287b886954f54b4c2fffcfd514079_0/explore. The data covers the Local Authority Districts in 2018 and applies to ITL3 level data for 2018. Where there were name changes between 2018 and 2020, manual alignment was performed using former boundaries. In the event of small boundary changes, a lack of microscale data on the areas that change Local Authority District means that merging applied to the old boundaries. The impact of these manual adjustments is expected to be small.

We chose to use the raw number of infections rather than the local infection rate because the infection rate at the dawn of the pandemic will show very small percentages of an area's inhabitants (particularly in higher populated areas), and yet someone only needed to come into close contact with one person to become infected, and each infected person could infect multiple people. In our study it is the presence of the small and increasing number of infected actual people, rather than the population-weighted average, that is of key interest when attempting to understand the pattern of virus spread, as has also been analysed by Hadjidemetriou et al. (2020), González-Val and Sanz-Gracia (2022), Olmo and Sanso-Navarro (2021), and others. Following these authors, although mortality prevention is of prime importance, it is COVID-19 related morbidity that is more strongly related to the spread of the virus and should be the focus of prevention policy, while it is also the case that the chance of mortality is affected by age and underlying health status.

Although it is contact with only one person emitting an airborne COVID-19 pathogen that could lead to its spread, González-Val and Sanz-Gracia (2022) argue that the density of population will likely increase the chance of that spread. Moreover, although it was the over 65s who were most affected by COVID-19, even though they were the least mobile with low incidence of commuting, it would be particularly interesting to identify whether the median age of an area affected the number of infections, because transmission and mobility are likely to be greater when an area's median age is lower.

Table 1 presents summary statistics of the variables. Average hours worked do not vary greatly but we do see a large regional income disparity, as has been documented in the literature (see McCann (2016) for an illustration of these and other economic imbalances across the UK). Raw values of GDP were used as our aim is to distinguish tangibly between the economic sizes of regions (use of a log-scaling would create greater spread lower down the distribution). Higher GDP and wage values closer to London and other regional economic centres reflect the geography of their national and regional spread. The rise in the number of COVID-19 recorded infections is clear, while the standard deviation reminds us of large inter-regional disparities. The trend in the number of infections continues upwards over the period of analysis and emphasizes the need to understand more about the spatial and socioeconomic factors underlying the temporal pattern of infection rates.

{Table 1}

Chen et al. (2021) provides a geo-visualisation of the regional disparities in the severity of the COVID-19 outbreak across England and shows that the epicentres shifted from Greater London to Leicester and finally to the North of England during pre-lockdown, post-lockdown, easing lockdown, and second national lockdown phases. This is useful evidence of a spatial evolution of a disease, and it shows COVID-19 infection rates evolving at varying rates across heterogeneous populations.

Understanding the factors that affected the rate of contagion enables more effective policy formation in future pandemics, but as the speed of each pandemic depends on the numbers who are susceptible to the virus and the degree of human connectivity, simply imposing an a priori physical or human geography model on the data may generate an incorrect spatial understanding of the virus, though of course this is something that others can test. For instance, Figure 1 presents maps of England that show the number of infections on the 25th of March and the 5th of May, and the initial pattern reveals that most areas had very few infections outside London, with an indication of some geographical spread out of London and into Berkshire to the west and into Hertfordshire to the north, plus an anomaly of Sheffield in South Yorkshire. By the 5th of May, infections had spread along the geographical spine of England up through rural Oxfordshire, into densely populated Birmingham, through rural Staffordshire, across the wider Manchester urban metropolis, and up to the city of Newcastle. Many other cities including Bristol, Cambridge, Nottingham, Leicester, Coventry, and Southampton appear to be relatively unscathed. Commuter zones west of London towards Reading were highly affected and yet commuter zones in the east of London remained largely unscathed at this time. These heterogenous spatial patterns of infection have hitherto not been explained with any rigour and are the focus of our attention.

{Figure 1}

4. Methodology

The effectiveness of policy interventions hinges on making the most informed decisions possible on where to allocate those resources. If policymakers are to achieve this maximal impact, then they must bring together insights from data at appropriate and practical spatial scales and take due care not to discard information that may be critical for obtaining optimal outcomes. In practice, this places considerable pressure on statistical techniques and modelling approaches: under-fit the data and the inference will not inform any policy effectively but over-fit the data and there is a risk of missing critical cases out of the sample which would outweigh the benefits achieved from a more accurate in-sample picture. Often what is required is a truer picture of what is in the data rather than a claim to extrapolate therefrom.

Similarly, imposition of an a priori theoretical model for the interpretation of data corresponding to the infection growth path, as is required when using a regression approach, could generate an incorrect overly generalized understanding of the spatial variation in the evolution of infection rates. Put simply, we do not know the correct functional form and model specification of the determinants of the spread of COVID-19 infection rates. For this reason, in this paper we draw together two statistical approaches to reveal statistical patterns in the data.

The first set of statistical approaches applies Moran's I, Geary's C, Local Indicators of Spatial Association (LISA), and bivariate Moran and bivariate LISA to reveal spatial clustering, co-movement between variables, and local hotspots with statistical significance testing. These are now standard techniques and remain simple to estimate and provide clear interpretation of spatial associations.

Our second approach complements and builds on these established spatial statistics by drawing on data science and adopting a statistical method that provides an understanding of the typology of the data: Dłotko (2019a)'s Topological Data Analysis – Ball Mapper (TDA-BM). Although this is a relatively simple method and generates figures that are can be easy to interpret, this is a relatively new method and so we provide preliminary details to enable the interpretation of the results. Examples of academic papers using this analytical method include Qiu et al. (2020), Rudkin and Webber (2023), Dłotko et al. (2021), Rudkin et al. (2024) and Tubadji and Rudkin (2025).

TDA-BM is an iterative process that begins with the selection of one datapoint from the full set of points in a data cloud. Once the point cloud is assembled, a ball of fixed radius is drawn around a randomly selected datapoint with all other datapoints within that ball being considered as covered. Covered datapoints are added to the covered set. Random selection of new datapoints from the uncovered set continues until there are no longer any points in the data cloud that do not belong to at least one ball. Each ball covers one or more datapoints, and the full set of balls provides a cover for the entire dataset.[16]

Representation of the balls in two dimensions for ease of visualisation requires the construction of an abstract graph where the balls are presented as discs. For consistency with their true status, we refer only to balls when talking about the TDA-BM graph. The size of

[16] Although construction of the cover vector requires consideration of the choice of variables on the axes, TDA-BM is not affected by correlations between axes and hence usual concerns about having too many axes relative to the number of observations becomes superfluous, and in this respect TDA-BM is far superior to regression analysis. Moreover, outliers tend to be considered inconvenient in regression, whereas the TDA-BM approach highlights these outliers so that we can learn more from their existence.

the balls refers to the number of areas possessing the same socioeconomic characteristics. Secondly, the colour of the balls provides information on the value of a variable of interest relating to the ball members. We colour the balls according to the average value across ball members of a variable of interest. Connections between balls – shown as edges – are added when a datapoint appears in more than one neighbouring ball. Edge length and thickness has no interpretation in this paper. The TDA-BM graph thus conveys information about the density of the point cloud through the size, the values of a variable of interest through colouration, and the relationships between balls through the edges. Numbering the balls allows cross-reference between the ball, its members, and inspection of underlying data.

In this paper, colouration of the balls is a three-step process. First, we calculate the average number of accumulated infections for the local authorities who are included within a ball on a given day and then divide this by the sum of accumulated infections for all balls in that day, and this gives a proportion of the accumulated cases within the considered ball relative to all accumulated cases on the given day. These numbers changed daily so we stress that all calculations are made from the same day's data, as represented in that day's TDA-BM graph. Alternative approaches, such as when high infection areas report low cases on a particular day, are more volatile. Using our adjustment allows the continued charting of dynamics and the resulting variable is subsequently plotted in the TDA-BM graph. Altering colouration in this way reminds that the total number of accumulated cases increased every day and ensures the balls only show their relative positions.

Application of the TDA-BM graph enables the visualization of important pieces of information.[17] First, a ball represents a subset of the entire dataset that have similar values of datapoints from a higher dimensional dataset. Second, the size of a ball reflects the number of local authorities that possess similar socioeconomic characteristics. Third, the colouration of the balls reflects the average value of the infection rate for all the local authorities included in that ball. Fourth, connections between balls[18] illustrate that at least one local authority is a member of both balls, and this is made possible when local authorities have socioeconomic characteristics that are common to both groups (balls) of local authorities. Observation of

[17] The choice of ball radius is important. Too low a radius and balls will be small, may not overlap, and will be affected greatly by random variations in the data. Choosing a larger radius controls for these factors but reduces the detail in the overall picture. Balancing informativeness with maintaining detail from the underlying data is a challenge and to date no optimisation algorithm has been generated to guide this decision. Therefore, application of the TDA-BM method requires a degree of care and researchers are advised to review a series of values to ensure the result is stable.

[18] The distance between the balls is meaningless and is the result of the abstract choice presentation.

pairs of connected balls with very different colouration attracts attention and emphasizes similarity in socioeconomic characteristics but dissimilarity in the level of infection rates; similarly, if the pattern of colour evolves steadily across the TDA-BM graph then this emphasises that socioeconomic characteristics are associated with the spatial evolution of COVID-19 infection rates.[19] Where a data cloud contains high and increasing infection levels so attention must focus on the combination of each socioeconomic characteristic that is associated therewith. By knowing combinations of socioeconomic characteristics in such detail, it is possible to design policy that fits the best those most affected areas given that they face the greatest challenge. Further details of the method are available in Rudkin (2025).

TDA-BM is thus creating a representation of a multidimensional dataset that is topologically faithful.[20] All data points are defined by their fixed values on each axis. Consequently, although the cover may vary slightly based upon the selection of the landmarks, the existence of LADs with high infection rates close to LADs with low infection rates in the data space will always manifest in the TDA-BM graph.[21]

Our application of the TDA-BM approach to map the variables' joint distribution creates an abstract summary of the importance of the interaction of socioeconomic characteristics, and we allow for the variation of the variables over time by looking at the maps coloured according to each day's number of cases.[22] As the aggregate socioeconomic characteristics do not change substantially during this short time period, the TDA-BM plot

---

[19] There are similarities between TDA-BM plots and clustering algorithms, as the aim in both approaches is to classify points according to their coordinates. TDA-BM places points into one or more balls, whereas clustering methodologies might only place each point into one set. Fuzzy and standard clustering techniques attempt to minimise the distance between points and the centroid of the cluster by segmenting denser areas into many blocks and then pulling in outliers to the main block. BM is better suited to covering data sets where there are meaningful groups that are not part of the main data cloud precisely because it does not pull such observations into bigger groups.

[20] As with a map of physical geography, the scale of the map is determined by the objective of the plot. A small radius provides a detailed, zoomed in, view of the data. A larger radius zooms out and shows the global structure. We sought to balance two issues: maintaining the structure and obtaining enough balls to show the variation in cases across the characteristic space. The balance was obtained through a radius of 0.9. Recalling that the data is standardised onto [0,1] for each variable, a radius of 0.9 means that a ball cannot contain the full range of values for more than two of the six variables. A circle of radius 0.9 will cover a unit square, but a ball of radius 0.9 can only cover all the values of two variables if the other four variables are equal.

[21] Consider a set of LADs arranged along a line: the LADs at the left end of the line have low cases, those at the right have high, and the midpoint of the line sees a step change from low to high. If a point near the middle is selected then the ball would contain some high, and some low infection LADs. However, the balls needed to cover the right and left ends of this spectrum would be high and low respectively showing the pattern. If a point near one end is chosen, then it is possible to obtain a cover with just two balls that has one high and one low ball. Again, the notion of a difference is preserved.

[22] Where there are underlying coefficients linking those demographic characteristics to the observed outcomes, we do not seek to estimate for precisely the reason that there are too many parameters for the number of observations.

also does not change. The changing colours over time provide a neat picture of the evolution of cases across the socioeconomic space.

5. Results

5.1 Spatial statistics

Our first step examines how spatial dependence evolves over time through application of Moran's I, Geary's C, Local Indicators of Spatial Association (LISA), bivariate Moran, and bivariate LISA. Table 2 reports selected-day values of Global Moran's I, Geary's C, and bivariate Moran statistics for the total number of infection cases. These results indicate clear positive spatial clustering in the early phase of the epidemic. Moran's I is 0.245 on 15 March and 0.247 on 25 March, with both values highly statistically significant; Geary's C yields the same conclusion. Figure 2 provides a time series plot of the bivariate Moran's I and reveals that early case spatial clustering is positively associated with the neighbouring distribution of hourly pay and population density, with the strongest effects observed in mid-March and late March. However, by 15 April these associations are substantially weaker, and by 15 May they are no longer statistically significant.

{Figure 2 and Table 2}

Consideration of per capita cases reflect the assertion that counts of raw cases may mechanically favour larger regions. Nevertheless, the qualitative patterns remain present in all spatial statistics. Application of bivariate LISA statistics reveal a declining number of regions in the High-High cluster for the earnings and hours worked variables. Temporal patterns are consistent with the contention that pay, and hence high-skilled workers, are linked to the spread of cases in the early days of the COVID-19 pandemic.

5.2 TDA-BM

The evolution of COVID-19 infection rates for each area can be understood by examining the socioeconomic characteristics at the dawn of the pandemic, a time series of observed infections, a time series of confounding factors, and combinations thereof. Interactions between local socioeconomic factors and local infection trajectories should be jointly understood to guide policymaking, and thus our work centralises the notion of a concert – or joint distribution – of regional characteristics. We approach this issue using an abductive quantitative methodological approach where the data observations tell us what to look for, rather than entering the analysis with a fixed vision of what we wish to see.

Figure 2 presents seven plots that chart the evolving picture of COVID-19 infections. These abstract BM graphs are drawn with six axes and demonstrate a broad connectivity across areas. Nevertheless, there are a series of balls that do not connect at this radius value, and these balls capture areas with particularly low population density (coloured red as they have a low number of infections) and areas with particularly high earnings (such as ball 19, which includes the areas of Hammersmith, Fulham, Kensington, and Chelsea). This finding is consistent with the proposition that the international spread of COVID-19 was associated with (perhaps facilitated or accelerated by) international travel of the predominantly wealthy community. Note that although larger balls represent a greater number of local authorities – i.e., areas with similar socioeconomic characteristics – it does not follow that they also have a greater proportion of infections. The highest proportions of infections are concentrated in balls that are smaller and thus infection rates are highly spatially concentrated, illustrating that the infection rate varies enormously across English local authorities .

{Figure 3}

All the panels in Figure 3 show significant colour variation across balls, suggesting that socioeconomic characteristics are associated with the spatial evolution of infections. When reporting deaths from the virus, it is common to talk of regions that are much larger than ITL3 regions, such as the ITL1 region of England's Northwest. However, the BM plot reveals that there is no restriction that places ITL3 regions near to each other based on their ITL1 hierarchical affiliation. Hence, we can quickly conclude that reporting infection rates at the larger ITL1 geographical area would be an inappropriate generalisation of the patterns of infections.

By early May, the highest concentration of infections appears in ball 26, which contains only Berkshire. This ITL3 region has a high population density like many London boroughs where infection cases are also high, but Berkshire is distinct from the London boroughs because it has higher average pay levels and a higher number of average hours worked along with a high proportion of commuters. The only other blue ball contains Hertfordshire, another similarly affluent commuter area for London. Green balls in panel (a) are predominantly the boroughs of London, incorporating the Outer London areas of Lewisham, Brent, Ealing, and the Inner London areas like Hackney, Lambeth, and Islington. The former set are revealed to be outliers for their comparatively low population densities and high average hours worked, and it is these balls that had previously housed most cases and appear dark blue in many of the earlier plots. Outer London boroughs are seen in balls 22, 24 and 25. Ball 31 has the lowest proportion of infections and represents five areas of the Southwest (Cornwall, Devon, Dorset, Somerset, and Torbay), which has traditionally been a region of England where people relocate for their retirement, with fewer high-paid jobs, lower population density, and it is too far for a regular commute to London.

Panels (b) to (d) reveal a consistency in the relative proportions of infections in each ball through the early weeks of the pandemic. Infections started to rise and spread, as shown by greater blue and orange colourations across the plot in panel (b), while panel (d) shows an early sign of a reduction in the concentration of infections across space. While these new hotspots have similar population densities, median ages, GDP, and hours worked, the notable differential is in the average rates of pay. It is this large differential in pay rates that leaves ball 19 disconnected from the main space. From a socioeconomic standpoint, this temporal transition reflects an important evolution of hotspots from high to low wage areas.

Observation of the evolution in colour from panels (b) to (d) reveals that many balls are catching up with ball 19 and this is particularly true of the balls to the right of the connected shape. Ball 1 is unusual as it remains red while its neighbour, ball 2, becomes progressively bluer. Ball 1 contains commuter towns that serve London (Swindon, Milton Keynes, Peterborough, and the Thames Gateway) that are all at least a half an hour fast train journey from London, and therefore require more expensive tickets than the outer London boroughs that have similar commute times. The results do not corroborate expectations of a rapid transmission of infections from London to these expensive commuter areas. By the end of March, the outbreak appears largely contained within England's cities and Berkshire.

Panels (e) to (g) and (a) portray growing infection rates through April and an overall brightening into orange. The highest number of cases transfers from ball 19 to ball 22 and then to ball 24 and finally to ball 26. The stark contrast between balls 1 and 22 remains despite both experiencing rising infection rates. The growth in infections far from London appears most noticeably through the darker colourations first in ball 9 and then in balls 7 and 8. Ball 9 contains areas that serve the main conurbations of Manchester and Leeds, which are far from the London boroughs in the plot, so we must consider what prompts the BM algorithm to create this separation.

To understand the plots further we colour the balls according to the values of each axis instead of the number of COVID-19 infections, as shown in Figure 4. Panel (a) shows that it is the balls to the upper left of the main shape that have the highest population densities, but the corresponding colouration for the infection rates in Figure 3 is mixed with the lower-left set experiencing faster growth in infections. In contrast to the findings of González-Val and Sanz-Gracia (2022), this result suggests that population density is not necessarily the main factor underlying the contagion of infections and that further confounding factors are at play. Panels (c), (d) and (e) in Figure 4 offer a nuanced explanation of our findings as areas with the lower median age, greater hours of work, and higher pay correlate neatly with the areas that have the highest reported number of COVID-19 infections.

{Figure 4}

There is further nuance at the other end of the connected shape where ball 9 becomes greener in Figure 3 which highlights non-uniform changes in other parts of the shape. Ball 9 contains the industrial north, including the Northern cities of Blackburn, Bradford, and Sheffield as well as outer suburbs of Manchester. The cities of Liverpool, Nottingham, Birmingham, and Bristol sit in ball 7, which lies between the major growth balls and ball 9; these cities serve as a conduit to their satellites. Our plots help greatly to reveal these dynamic characteristics.

The overarching message from the BM graphs is that there were major spatial disparities in the evolution of COVID-19 infection rates across England, and that much of the evolving differentiation was associated with socioeconomic characteristics. Given this observable heterogeneity, policies should recognise the importance of the interactions of socioeconomic factors when determining how to allocate scarce resources in the fight against pandemics.

Both the spatial statistics and the TDA-BM analysis speak to the importance of highly skilled workers at the start of the pandemic. To evaluate the importance of labour market variables in comparison to population density and GDP we use ANOVA analysis and explicitly consider the incremental R^2 from including labour variables. The incremental effect on the R^2 from including labour market variables is shown in Figure 5, and these results illustrate the qualitative importance of labour market variables for improving the explanatory power of the ANOVA model for total cases and cases per capita respectively.[23] Figure 5

[23] Although we follow Hadjidemetriou et al. (2020), González-Val and Sanz-Gracia (2022), Olmo and Sanso-Navarro (2021) and others by examining raw counts of infections rather than population-normalised rates, it is the case that population areas with larger population size and density have experience faster growth. Using absolute numbers mechanically inflates the apparent severity in larger regions even when the per capita risk is similar or lower. We chose to follow those authors and use the raw number of infections, rather than the local infection rate, because the infection rate at the dawn of the pandemic shows very small percentages of an area's inhabitants (particularly in higher populated areas), and yet someone only needed to come into close contact with one person to become infected, and each infected person could infect multiple people. Future research could explore these associations in more detail.

shows that adding labour market variables to a model containing population density and GDP to the 25th March 2020 model yields a larger improvement in fit than when adding population density and GDP to a labour model (ANOVA F=9.31 versus F=3.69). Across the estimations from 15th March 2020 to 15th May 2020, labour market variables provide the larger incremental contribution to total cases model on 51 of the 62 days, and to the per capita cases model on 43 of the 62 days. Thus Figure 5 provides evidence to support the claim that labour market variables had strong explanatory power at the start of the pandemic.[24] These results collectively point to the primacy association with the evolution in infection rates of labour market characteristics over population density and GDP, and they illustrate that the baseline results demonstrated in the TDA-BM analysis are neither a population size artefact nor the result of economic size.[25]

{Figure 5}

6. Discussion and conclusion

This paper sought to improve understanding of the importance of socioeconomic characteristics that underpin the spatially heterogeneous evolution of COVID-19 infections across England's ITL3 regions at the start of the pandemic. The small spatial scale for analysis underscores the importance of not making sweeping conclusions about infection cases across large swathes of geography, such as nationwide or ITL1 spatial definitions. Although each pandemic may differ with respect to the ease and rate of contagion and incidence of mortality, socioeconomic interactions associate with pathogen transfer.

Zhou et al. (2021) observed a relatively even distribution of cases in the early days of the pandemic in Hong Kong, and the reason for this may be that Hong Kong is dominated by one city. In contrast, our findings show that the early days of the COVID-19 pandemic in the UK illustrated strong concentration in infections with most cases occurring within London,

[24] The average incremental R^2 from labour variables compared to the incremental R^2 contribution of population density and GDP is 0.102 for total cases and 0.111 for per-capita cases.

[25] A criticism of ANOVA models used to compute the incremental R^2s is that they fail to account for the spatial correlation potentially present in the error terms. It is readily verifiable that the adoption of a spatial modelling approach does little to change the qualitative conclusions. Table A in the appendix reports the coefficients on hourly gross pay using OLS, spatial lag models (SAR), and spatial error models (SEM). Spatial dependence parameters are reported for SAR and SEM. The coefficients on hourly gross pay do not change greatly across specifications. The spatial parameter is significant in the SAR and only marginally insignificant for total cases in the SEM. However, the inference of key relevance in this paper is not that spatial effects are absent, which could of course be the focus of future research, rather that the inclusion of labour market variables is important when understanding the spread of COVID-19 infections at the start of the pandemic.

and particularly in the relatively affluent area of Hammersmith, Fulham, Kensington, and Chelsea. That area stands apart from the main set of ITL as it is an area with high GDP, high wages, and high numbers of hours worked. At this point, there was much to link working practices and population density to the outbreak. The most economically similar regions were also in London but were not sufficiently associated for the TDA-BM to connect them based on COVID-19 infections. The first two weeks of the pandemic in the UK were characterised by increasing infection cases in the London boroughs, which is entirely consistent with movements of key workers and support staff for the economic activity within London. As the UK approached lockdown, there were growing infection cases in other major English cities, although London boroughs still led the way. Hence, the data strongly support the UK Government's policy of restricting movement, though there are arguments to suggest this decision should have been made earlier.

Although the early stages of the pandemic showed isolated infections in pockets of high population densities, primarily in London, subsequent data reveal increasing dispersion based on socioeconomic similarity rather than geographic proximity. Very quickly new curves of total infections began to turn upwards in Manchester, Birmingham, and some other larger English cities, but as the spread of the pandemic unfolded so infections dispersed more intensively across London boroughs rather than to commuter towns or other larger cities. As the first month of data ended, data revealed that infections then dispersed more evenly across the economic characteristics space. The rationale for targeting resources had moved progressively away from containing infection cases within a limited number of regions to slowing the growth in infections in areas where hospital provision was far lower.

Even regions that are very similar to London boroughs in terms of their socioeconomic characteristics were not exhibiting many cases relative to the maximum. Our analysis shows how the curve of total infections in commuter towns that serve London did not start trending upwards until after the commencement of the lockdown, and the total infections in other London commuter towns were following significantly lower upward trends (such as in Essex). Restrictions on travel had come too late to prevent the explosion in infection rates, but equally there is evidence to suggest that the natural flows of people had not been transporting the virus. Without the data on individual movements such inference cannot be confirmed, but the observable contrasts in the TDA-BM graph are unambiguous. These results add to and nuance existing research on the socioeconomic s of COVID-19, such as the findings of Ascani et al. (2021), who found that geographically concentrated economic activities act as a vehicle for virus transmission, suggesting a core-periphery pattern of COVID-19 following the local economic landscape.

As the outbreak progressed, our analysis points to a second group of regions developing in the lower pay end of the plot where smaller towns with lower median ages, lower population density and which deliver lower GVA began experiencing increasing infection rates. They are areas where the hours worked is similar to the London boroughs and so we posit a role for workplaces in this pandemic, but equally we see that no single variable is informing us where the next cluster will form in the characteristics space. Although the early policy focus was on enhancing capacity to nurse infected people in London, now there was an urgent need for more regionalised provision. In more remote areas, effective accessibility to hospital services were vitally important, especially in areas with harder-hitting austerity measures and poorly maintained public transport links to hospitals.

Understanding of the evolution of COVID-19 infection rates presented in this paper was derived from the power of data visualisation methodologies associated with the TDA-BM algorithm (Dłotko, 2019a). Premising the need to gain understanding from raw data rather than the testing of hypotheses, our results point to different actions suiting different locations at different stages of the pandemic.

Our evidence is an original and critical call for heterogeneous actions in future pandemics: first, as COVID-19 pathogen transfer is rapid, we need a rapid early warning system to fight against pandemic outbreaks, and this needs to be emphasised clearly to all interconnected groups within society. Second, we need an effective and transparent comprehension of the complex localised but spatially interlocking socioeconomic patterns of human interaction that correlate with pathogen transfer; this is necessary if we are going to formulate and enable policies to abate the spread and consequences of future pandemics. Our detailed and rigorous contribution directs attention beyond trends and averages, both in the fight against the COVID-19 pandemic and to learn nuanced patterns that may help form policies to fight future pandemics.

Our focus on locality engenders a more efficient spatial resource allocation, which is pivotal at times of competing needs. Yet, at small spatial scales, there may be other factors relevant to the spread of pathogen transfer that we have not included in our analysis, such as local trust, social capital or civic engagement, that could be included in future analyses. As our results correspond to the early stages of the 2020 pandemic in the UK when the binding affinity of COVID-19 (wild type) to human cells was lower than under the Omicron strain

(Nguyen et al., 2022), our identified spatial patterns of pathogen transfer may be stronger or weaker during the next pandemic.

References:

Aboukorin, S.A.A., Han, H.Y., and Mahran, M.G.N. (2023) Pandemic resilience planning: NPI measures and COVID-19 impacts in UK, Germany and Italy, *Cities*, 143.

Adda, J. (2016). Economic activity and the spread of viral diseases: Evidence from high frequency data. *Quarterly Journal of Economics*, 131(2), 891-941.

Amaros, F.G., Ascani, A., Faggian, A., and Vermeulen, W.N. (2025) Labour demand in the wake of a shock: a dose-response approach, *Papers in Regional Science*, 104(2), 100083.

Andersen, L.M., Harden, S.R., Sugg, M.M., Runkle, J.D., and Lundquist, T.E. (2021) Analysing the spatial determinants of Covid-19 transmission in the United States, *Science and the Total Environment*, 754, 142396, 1-10.

Ang, J.P., Gao, G., and Sparks, A. (2023). A dichotomy between democracy and personal freedom on the spread of COVID-19. *International Journal of Social Economics*, 50(3), 321-334.

Argys, L.M. and Rees, D.I. (2008). Searching for peer group effects: A test of the contagion hypothesis. *Review of Economics and Statistics*, 90(3), 442-458.

Ascani, A., Faggian, A., and Montresor, S. (2021). The geography of COVID-19 and the structure of local economies: the case of Italy. *Journal of Regional Science*, 61, 407-441.

Barnay, T. (2016). Health, work and working conditions: a review of the European economic literature. *European Journal of Health Economics*, 17(6), 693-709.

Bearth, A., Luchsinger, L., and Siegrist., M. (2021). Reactions of older Swiss adults to the COVID-19 pandemic: a longitudinal survey on the acceptance of and adherence to public health measures. *Social Science and Medicine*, 280, 114039.

Böckerman, P., Johanssen, E., Helakorpi, S. and Uutela, A. (2009) Economic inequality and population health: looking beyond aggregate indicators, *Sociology of Health and Illness*, 31(3), 422-440.

Bourdin, S., Tuzcu, S.E., and Satici, E. (2023). Explaining COVID-19 vaccine uptake: a spatial sociodemographic study in Turkey, *Papers in Regional Science*, 102(2), 307-330.

Cai, Y.T. and Mason, K.A. (2022). Why they willingly complied: ordinary people, the big environment, and the control of COVID-19 in China. *Social Science and Medicine*, 309, 115239.

Chen, Y.W., Ni, L., Xu, D.L. and Yang, J.B. (2021) Visualising regional disparities in the risk of COVID-19 at different phases of lockdown in England. *Environment and Planning A: Economy and Space*, 53(5), 883-886.

Clouston, S.A.P., Natale, G., and Link, B.G. (2021) Socioeconomic inequalities in the spread of coronavirus-19 in the United States: A examination of the emergence of social inequalities, *Social Science and Medicine*, 268, 113554, 1-6.

Consolazio, D., Murtas, R., Tunesi, S., Gervasi, F., Bennasi, D., and Rosso, A.G. (2021) Assessing the impact of individual characteristics and neighbourhood socioeconomic status during the COVID-19 pandemic in the provinces of Milan and Lodi, *International Journal of Health Services*, 51(3), 311-324.

Dingel, J.I. and Neiman, B. (2020) How many jobs can be done at home? *Journal of Public Economics*, 189, 104235.

Dłotko, P. (2019a). Ball mapper: a shape summary for topological data analysis. arXiv preprint arXiv:1901.07410.

Dłotko, P. (2019b). *BallMapper: Create a Ball Mapper graph of the input data*. R package version 0.1.0.

Dłotko, P., Qiu, W. and Rudkin, S.T. (2021). Financial ratios and stock returns reappraised through a topological data analysis lens. *The European Journal of Finance*, 1-25.

Du, Z. and Holme, P. (2020). Coupling the circadian rhythms of population movement and the immune system in infectious disease modelling. *PLoS ONE*, 15(6), e0234619

du Plessis, L., McCrone, J.T., Zarebski, A.E., Hill, V., Ruis, C., Gutierrez, B., Raghwani, J., Ashworth, J., Colquhoun, R., Connor, T.R. and Faria, N.R. (2021). "Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK". *Science*, 371 (6530), 708–712.

Ehlert, A. (2021) The socio-economic determinants of COVID-19: A spatial analysis of German county level data, *Social-Economic Planning Sciences*, 78, 101083, 1-9.

Gibson-Miller, J., Zavlis, O., Hartman, T.K., Bennett, K.M., Butter, S. Levita, L., Martinez, A.P., Mason, L., McBride, O., and McKay, R. (2024) A network approach to understanding social distancing behaviour during the first UK lockdown of the COVID-19 pandemic, *Psychology and Health*, 39(1), 109-127.

González-Val, R. and Sanz-Gracia, F. (2022) Urbanization and COVID-19 incidence: A cross-country investigation, Papers in Regional Science, 101(2), 399-416.

Grépin, K.A., Poirier, M.J., and Fox, A.M. (2020). The socio-economic distribution of exposure to Ebola: Survey evidence from Liberia and Sierra Leone. *SSM Population Health*, 10:100472.

Hadjidemetriou, G.M., Sasidharan, M., Kouyialis, G., Parlikad, A.K. (2020) The impact of government measures and human mobility trend on COVID-19 related deaths in the UK, *Transportation Research Interdisciplinary Perspectives*, 6, 100167, 1-6.

Harding, V. (1989). 'and one more may be laid there': the location of burials in early modern London. *The London Journal*, 14(2), 112-129.

Hawkins, D. (2020) Differential occupational risk for COVID-19 and other infection exposure according to race and ethnicity, *American Journal of Industrial Medicine*, 63, 817-820.

He, M.W., Zhao, S.C., and He, M (2016). Tolerance threshold to commuting time: evidence from Kunming, China. *Journal of Transport Geography*, 57, 1-7.

Howson, W. (1961). Plague, poverty, and population in parts of north-west England, 1580-1720. *Transactions of the Historic Society of Lancashire and Cheshire*, 112, 29-55.

Iammarino, S. and McCann, P. (2006) The structure and evolution of industrial clusters: transactions, technology, and knowledge spillovers. *Research Policy*, 35(7), 1018-1036

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L. and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990-993.

Kaim, A., and Saban, M. (2022). Are we suffering from the Peltzman effect? Risk perception among recovered and vaccinated people during the COVID-19 pandemic in Israel. *Public Health*, 209, 19-22.

Kaim, A., Gering, T., Moshaiov, A., and Adini, B. (2021). Deciphering the COVID-19 health economic dilemma: a scoping review. *International Journal of Environmental Research and Public Health*, 18(18), 9555.

Kohn, G.C. (2007). *Encyclopedia of plague and pestilence: From ancient times to the present*. Infobase

Link, B.G., and Phelan, J. (1995) Social conditions as fundamental causes of disease, *Journal of Health and Social Behaviour*, 80-94.

Mack, A. (1991). In time of plague: the history and social consequences of lethal epidemic disease. NYU Press.

Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., and Roser, M. (2020-2026f). Coronavirus pandemic (COVID-19). *Our World in Data*, https://ourworldindata.org/coronavirus.

McCann, P. (2016). *The UK regional-national economic problem: Geography, globalisation, and governance*. Routledge.

Morris, E.A., and Zhou, Y. (2018) 'Are long commutes short on benefits? Commute duration and various manifestations of well-being,' *Travel Behaviour and Society*, 11, 101-110.

Nguyen, H.L., Nguyen, Q.T., Nguyen, P.H., and Li, M.S. (2022) SARS-CoV-2 Omicron variant binds to human cells more strongly than the Wild Type: evidence from molecular dynamics simulation. *Journal of Physical Chemistry B*, 126(25), 4669-4678

Olmo, J. and Sanso-Navarro, M. (2021) Modeling the spread of COVID-19 in New York City, *Papers in Regional Science*, 100(5), 1209-1230.

Packard, R.M. (1989). White plague, black labor: Tuberculosis and the political economy of health and disease in South Africa, vol. 23. University of California Press.

Paton, C. (2020) Insouciance and inexperience: a deadly combination when dealing with COVID-19, *International Journal of Health Planning and Management*, 35(5), 983-987.

Phelan, J.C., Link, B.G., Tehranifar, P. (2010) Social conditions as fundamental causes of health inequalities: theory, evidence and policy implications, *Journal of Health and Social Behaviour*, 51(1), S28-S40.

Purves, R.I., Maclean, J., Rocha, C., Philpott, M., Fitzgerald, N., Piggin, J., and Hunt, K. (2023) Attending sporting mega events during COVID-19: mitigation and messaging at UK EURO 2020 matches, *Health Promotion International*, 38(1), 1-12.

Qiu, W., Rudkin, S. and Dłotko, P. (2020). Refining understanding of corporate failure through a topological data analysis mapping of Altman's Z-score model. *Expert Systems with Applications*, 156, 113475.

Rahaman, H., and Barik, D. (2023) Investigation of airborne spread of COVID-19 using a hybrid agent-based model: a case study of the UK. *Royal Society Open Science*, 10(7), 230377.

Rietveld, J., Hobson, T., Mani, L., Avin, S., and Sundaram, L. (2024) The UK's pandemic preparedness and early response to the COVID-19 pandemic, *Global Public Health*, 19(1), 2415499.

Rudkin, S. (2025). An Introduction to Topological Data Analysis Ball Mapper in R. *arXiv preprint arXiv:2504.14081*.

Rudkin, S. and Webber, D.J. (2023). Regional growth paths and regional resilience. *Available at SSRN 4333276*.

Rudkin, S., Barros, L., Dlotko, P. and Qiu, W. (2024). An economic topology of Brexit. *Regional Studies*, 53(3), 601-618.

Santos, G., Maoh, H., Potoglou, D. and von Brunn, T. (2013) Factors influencing modal split in commuting journeys in medium-sized European cities. *Journal of Transport Geography*, 30, 127-137.

Shi, Q.J. and Liu, T. (2020) Should internal migrants be held accountable for spreading COVID-19? *Environment and Planning A*, 52(4), 695-697.

Shi, Q.J., Dorling, D., Cao, G.A., Liu, T. (2020) Changes in population movement make COVID-19 spread differently from SARS, *Social Science and Medicine*, 255, 113036, 1-3.

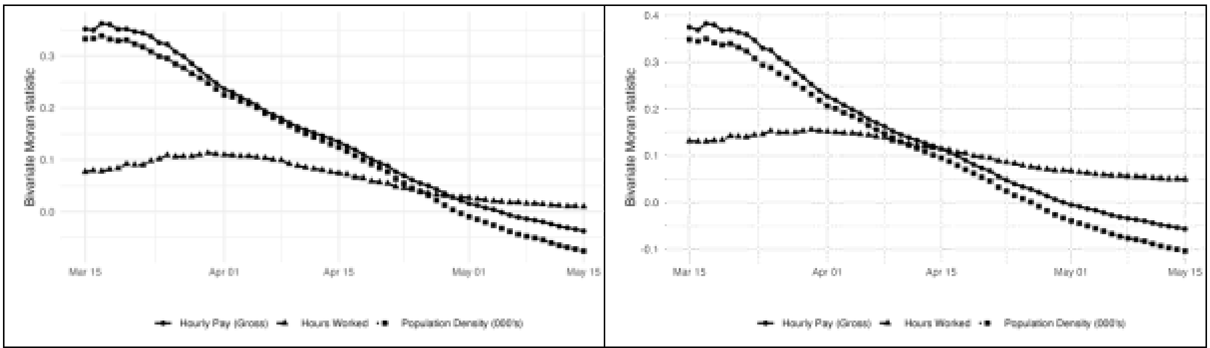
Tubadji, A., & Rudkin, S. (2025). Cultural gravity and redistribution of growth through migration: Cohesion lessons from spatial econometrics and topological data analysis. *Papers in Regional Science*, 104(1), 100064.

van Ommeren, J. and Reitveld, P. (2005). The commuting time paradox. *Journal of Urban Economics*, 58(3), 437-454

Wilder-Smith, A., Chiew, C.J., and Lee, V.J. (2020). Can we contain the COVID-19 outbreak with the same measures as for SARS? *The Lancet Infectious Diseases*, 20, e102-107.

Yin, J.J., Soliman, A., Yin, D.D., and Wang, S.W. (2017). Depicting urban boundaries from a mobility network of spatial interactions: a case study of Great Britain with geo-located Twitter data. *International Journal of Geographical Information Science*, 31(7), 1293-1313.

Zhou, J.P, Ho, S.K.S., Lei, S.Y. and Pang, V.C.K. (2021). Population density, activity centres, and pandemic: visualizing clusters of COVID-19 cases in Hong Kong. *Environment and Planning A: Economy and Space*, 53(7), 1594-97.



(a) Total Cases	(b) Cases per 100,000 people

Figure 1: Daily Bivariate Moran statistics

Notes: The figure plots the daily bivariate Moran statistic between total case levels, panel (a), and population normalised case levels, panel (b), and the neighbouring distribution of hourly pay (gross), hours worked, and population density from 15 March to 15 May 2020. The strongest positive spatial association is observed in the early phase of the epidemic, particularly for hourly pay and population density, and weakens as diffusion becomes more geographically widespread.

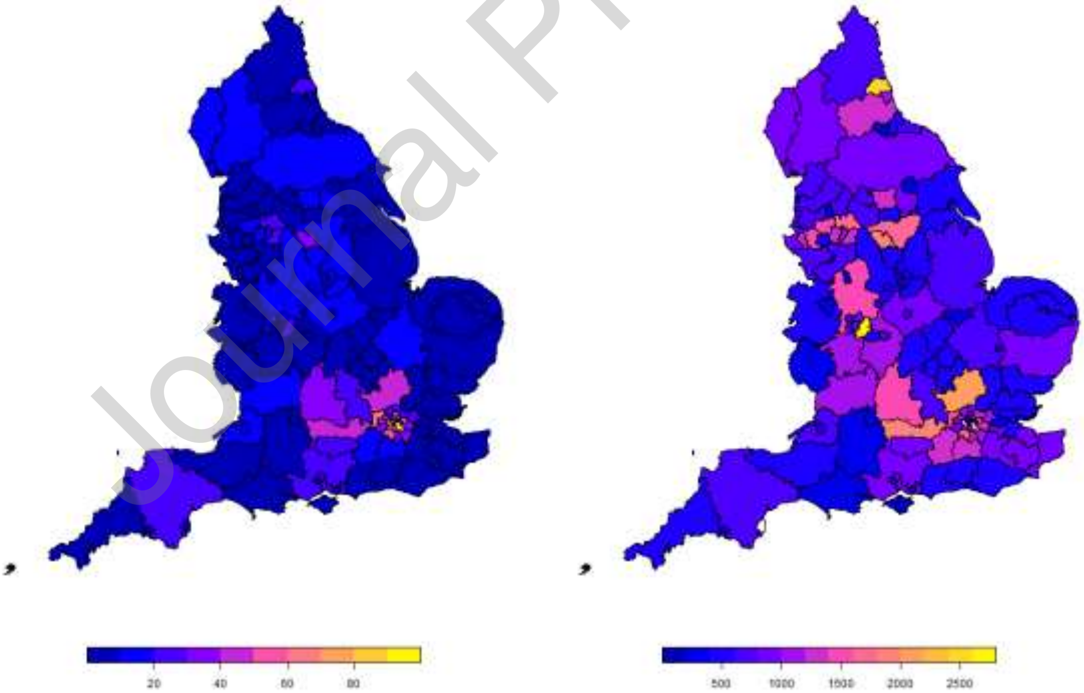


Figure 2: Cumulative case numbers by ITL3 regions on 25 March (left) and 5 May

Notes: Maps generated using shapefiles of ITL3 regions downloaded from the Office for National Statistics. Colouration is as indicated in the bar below the plot and captures the total number of cases observed within the regions in the period up to the last date stated.

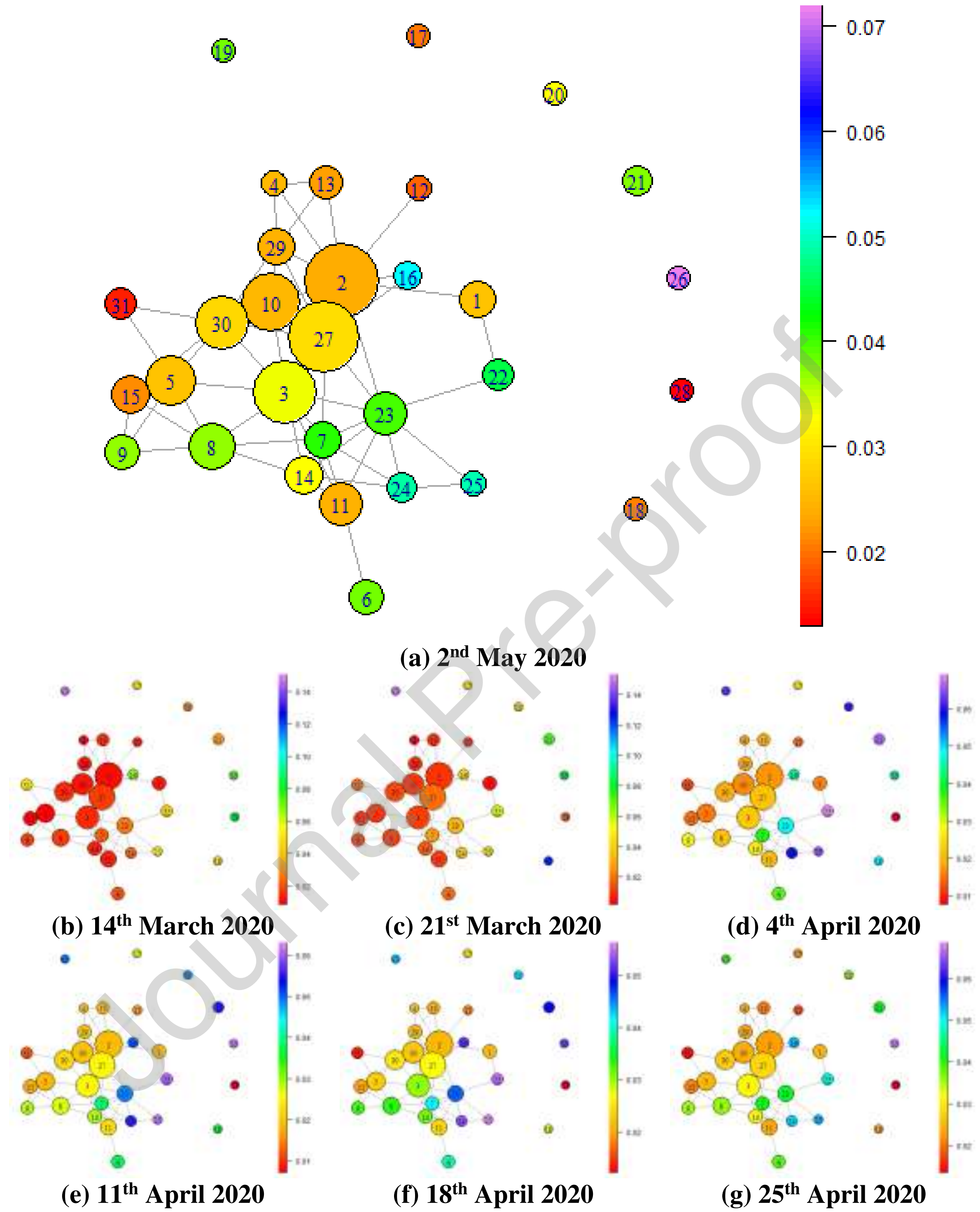
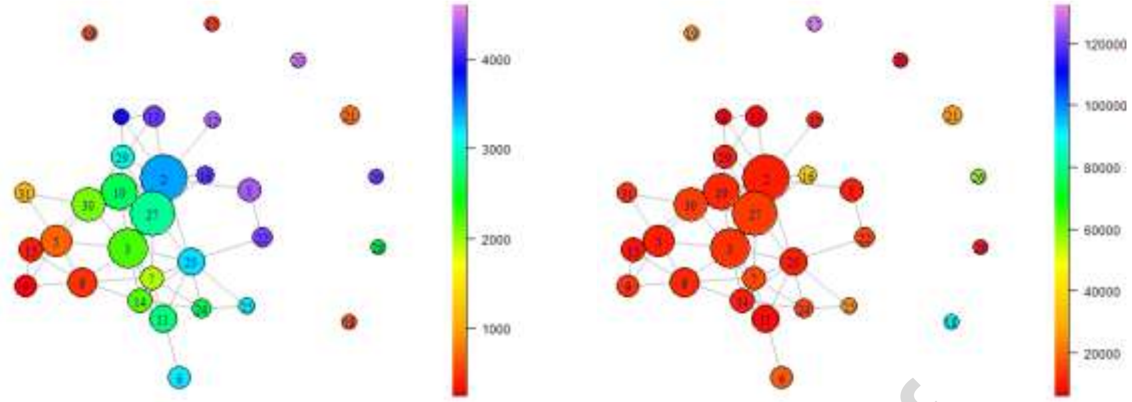
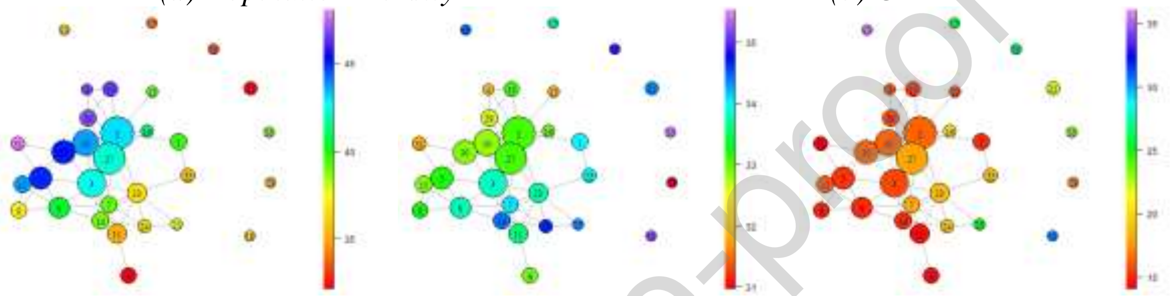


Figure 3: Evolution of cumulative case numbers

Notes: Axes used to construct the TDA-BM graph are population density, median age, gross domestic product, hours worked, and gross pay. Data on average hours worked and average annual gross income is taken from the ASHE table 7 and is collated at the local authority level. Covid-19 case data is collected by the UK Government and is downloaded from the Covid-19 dashboard on www.gov.uk. Merging between local authorities and ITL3 regions is performed using the UK Government lookup query. Where matches are not found manual search is performed. Plots are generated using the R package Ballmapper (Dłotko, 2019b). Colouration is the proportion of the cumulative total of English COVID-19 cases which are within each ball. Ball radius 0.9.



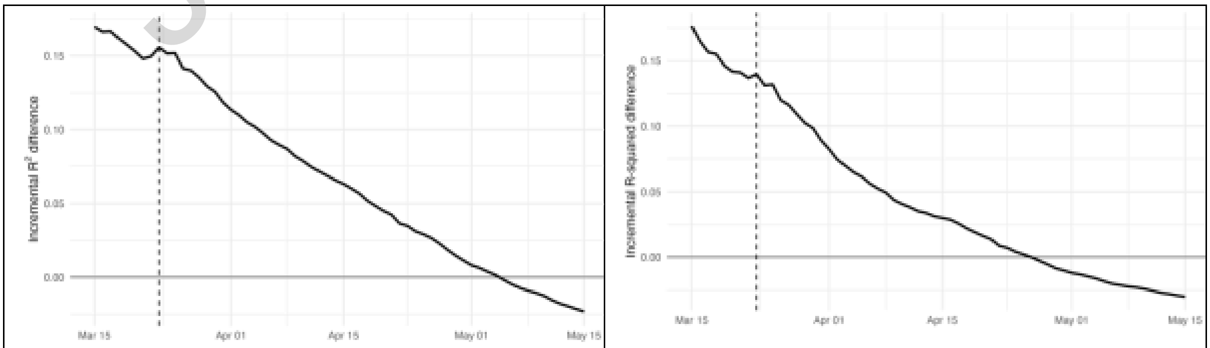
(d) *Population Density* (b) *GDP*



(c) *Median Age* (d) *Hours Worked* (e) *Gross Pay*

Figure 4: Axis variables

Notes: Population density, median age, and Gross Domestic Product (GDP) are sourced at the ITL3 level from Eurostat. Data on average hours worked and average annual gross income is taken from the ASHE table 7 and is collated at the local authority level. Covid-19 case data is collected by the UK Government and is downloaded from the Covid-19 dashboard on www.gov.uk. Merging between local authorities and ITL3 regions is performed using the UK Government lookup query. Where matches are not found manual search is performed. Plots are generated using the R package Ballmapper (Dłotko, 2019b). Ball radius 0.9. (Obs=134).



(a) Total Cases (b) **Cases per 100,000 people**

Figure 5: Axis variables

Table 1: Summary statistics

	Mean	S.D.	Min	Max
Panel (a): Axis variables				
Population density (Population per Km^2)	2454	1337	15	4374
Median age (in years)	40.52	4.742	29.90	50.40
Gross Domestic Product (in £000s)	15719	15067	2517	131476
Average hours worked per week	33.37	0.854	31.00	35.50
Average annual gross income (in £000s)	17.03	3.501	12.25	35.92
Panel (b): Average number of infection cases by ITL3 region:				
14th March	2.341	2.828	0	16
18th March	5.788	6.709	0	37
21st March	10.77	12.90	0	76
25th March	30.27	33.76	0	194
28th March	50.05	50.12	1	278
1st April	105.0	92.52	5	507
4th April	153.0	127.0	11	690
8th April	253.0	194.0	25	1138
11th April	334.5	240.7	36	1372
15th April	454.6	299.4	45	1712
18th April	532.3	336.1	53	1884
22nd April	704.7	424.3	72	2390
25th April	762.3	450.0	79	2559
29th April	836.3	481.3	103	2758
2nd May	878.8	499.3	116	2836

Notes: Data on population density, median age, and regional Gross Domestic Product (GDP) are collected at the ITL3 level from Eurostat. Data on average hours worked and average annual gross income are taken from the ASHE table 7 and are collated at the local authority level. Covid-19 case data are collected by the UK Government and are downloaded from the COVID-19 dashboard on www.gov.uk. Merging between local authorities and ITL3 regions are performed using the UK Government lookup query. Where matches are not found manual search were performed. A pre-processing step is used to transform each variable in our dataset onto [0,1] so that all variables are on an equivalent scale. (Obs=134).

Table 2: Spatial Diagnosis Statistics

	Date	Moran's I	Geary's C	Bivariate Moran's I		
				Hourly Pay	Hours Worked	Population Density
Total Cases	15/03/2020	0.2450***	0.7361***	0.3523	0.0763	0.3331
	25/03/2020	0.2468***	0.7203***	0.3226	0.1086	0.2952
	15/04/2020	0.0747	0.9125	0.1344	0.0736	0.01231
	15/05/2020	0.0357	0.9950	-0.0374	0.0100	-0.0762
Cases per 100,000 people	15/03/2020	0.3704***	0.6352***	0.3750	0.1310	0.3483
	25/03/2020	0.3625***	0.6662***	0.3253	0.1527	0.2886
	15/04/2020	0.2086***	0.08863	0.1145	0.1137	0.0948
	15/05/2020	0.1700**	0.9361	-0.0569	0.0495	0.1042

Appendix

Modern spatial econometrics considers that there may be influence on outcomes from proximate regions. To evaluate the importance of spatial effects on the inferences drawn in this paper, we estimate spatial lag models (SAR) and spatial error models (SEM) and then compare these with coefficients generated using a linear Ordinary Least Squares (OLS) model. We estimate the models for both the count of cases and the number of cases per 100,000 residents for data from March 25th 2020. Table A provides the estimates on the average Hourly Pay - Gross (HPG) within ITL3 regions though for brevity we do not report the coefficients on the other variables.

Table A reveals that HPG is significant under all specifications. Applications of spatial econometrics do not adjust the conclusion that labour market variables matter. For the two spatial models, we report the spatial parameter estimate and the associate *p*-value. Only for the total number of cases in the SEM model is the spatial parameter not significant at the 5% level (*p*=0.069). Spatial models do help if the goal is to predict case numbers, but they do not change the primary inference of our paper. Results for other days within the sample reveal a qualitatively similar picture. The importance of labour market variables decline as the pandemic goes on and infections spread further. However, the significance of labour market variables in the early days of the pandemic continues to be substantively important and this conclusion is not altered when applying spatial econometrics.

Table A: Spatial regression coefficients

Outcome Variable	Model Coefficients OLS	SAR	SEM	Spatial Parameter SAR	SEM	AIC OLS	SAR	SEM
Log cases	0.177*** (0.000)	0.144*** (0.000)	0.167*** (0.000)	0.258** (0.008)	0.215 (0.065)	353.4	347.9	351.7
Log cases per 100k	0.163*** (0.000)	0.120*** (0.000)	0.141*** (0.000)	0.343*** (0.000)	0.371*** (0.000)	324.0	312.8	314.6

Notes: Hourly pay coefficients are reported for spatial models of COVID-19 infections on 25th March 2020 at the ITL3 level. Models included are Ordinary Least Squares (OLS), a spatial lag model (SAR), and a Spatial Error Model (SEM). The table reports the model coefficient and, for the SAR and SEM models, the estimated spatial dependence parameter; AIC values for the three models are also reported. Figures in parentheses are the associated *p*-values. Significance is denoted by * (5%), ** (1%), and *** (0.1%). Obs=134.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: