



The propagation effect of commuting to work in the spatial transmission of COVID-19

Timo Mitze¹ · Reinhold Kosfeld²

Received: 23 December 2020 / Accepted: 31 March 2021 / Published online: 23 May 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

This work is concerned with the spatiotemporal dynamics of the coronavirus disease 2019 (COVID-19) in Germany. Our goal is twofold: first, we propose a novel spatial econometric model of the epidemic spread across NUTS-3 regions to identify the role played by commuting-to-work patterns for spatial disease transmission. Second, we explore if the imposed containment (lockdown) measures during the first pandemic wave in spring 2020 have affected the strength of this transmission channel. Our results from a spatial panel error correction model indicate that, without containment measures in place, commuting-to-work patterns were the first factor to significantly determine the spatial dynamics of daily COVID-19 cases in Germany. This indicates that job commuting, particularly during the initial phase of a pandemic wave, should be regarded and accordingly monitored as a relevant spatial transmission channel of COVID-19 in a system of economically interconnected regions. Our estimation results also provide evidence for the triggering role of local hot spots in disease transmission and point to the effectiveness of containment measures in mitigating the spread of the virus across German regions through reduced job commuting and other forms of mobility.

Keywords SARS-CoV-2 · COVID-19 · Mobility · Job commuting · Containment measures · Lockdown · Spatial econometrics

JEL Classification C23 · I18 · R23

✉ Timo Mitze
tmitze@sam.sdu.dk

¹ University of Southern Denmark, RWI and RCEA, Alsion 2, 6400 Sønderborg, Denmark

² University of Kassel, Nora-Platiel-Str. 4, 34109 Kassel, Germany

1 Introduction

Understanding the main transmission channels of the coronavirus disease 2019 (thereafter: COVID-19) pandemic is a key factor for effectively suppressing its spread until a pharmaceutical cure is available for large parts of the population. This work is concerned with an econometric analysis of the spatiotemporal dynamics of COVID-19 in Germany during the first pandemic wave in spring 2020. The goal is twofold: First, we identify the significance of commuting-to-work patterns between NUTS-3 regions (*Kreise und kreisfreie Städte*) as an explicit channel for spatial disease transmission before containment measures have been introduced in mid-March 2020. While it seems natural to assume that the spread of COVID-19 has a spatial dimension, estimating the magnitude of mobility effects related to professional and social interactions is challenging as these interactions are difficult to measure (e.g., social contacts with family and friends, shopping behavior, etc.) and disentangle—even with novel digital data sources such as mobile phone or social media data available (see Schlosser et al. 2020, for an overview).¹

The identification strategy proposed in this paper makes use of detailed information on interregional commuting flows across German NUTS-3 regions measured at the onset of the COVID-19 outbreak to proxy structural labor market interdependencies between regions. Although pre-COVID-19 data on interregional commuting cannot directly track behavioral changes in mobility patterns associated with the pandemic development (see, e.g., Lee et al. 2020; Pepe et al. 2020; Gao et al. 2020; Schlosser et al. 2020, for this type of analysis using mobility as key outcome), we argue that the main advantage of our identification strategy is precisely to exploit this exogeneity of using pre-pandemic mobility data to size the infection effects from mobility at the regional population level. To do so, we estimate a series of spatial panel error correction models (SP-ECMs), which allow us to combine elements from a theoretically grounded long-run specification of epidemic diffusion between regions together with a flexible short-run approach that accounts for the nonlinear dynamics of new COVID-19 cases.

Building on the SP-ECM specification, our second goal is to assess the effects of containment (lockdown) measures on the strength of interregional disease transmission through job commuting. In the case of Germany, several public health measures were targeted to reduce the mobility of the population in spring 2020 (e.g., Berlemann and Haustein 2020; Hartl and Weber 2020). While, for instance, almost all white-collar workers have been encouraged or ordered to work from home, education institutions, most retail stores and service activities (restaurants, bars, hairdressers, etc.) have been temporarily closed during the first lockdown between mid-March and mid-April 2020. We test for the effectiveness of such mobility constraints in slowing down the epidemic spread by measuring changes to the strength of spatial diffusion of COVID-19 across German regions with and without policy interventions in place. We argue that the analysis of the first pandemic wave and lockdown

¹ Tizzoni et al. (2020) also point to the potential problem of overestimation of mobility flows, such as commuting networks, when mobile phone rather than census data are used.

phase in Germany during spring 2020 is well suited to answer our research questions as both (a) the international arrival of COVID-19 and (b) the imposition of containment measures can be regarded as an exogenous shock to the local population so that early anticipation effects biasing our estimation results are likely to be small. This is less so the case for the second wave in autumn 2020 with lockdown measures being imposed in a much more staggered way accompanied by an inconsistent implementation of these rules (Warren et al. 2021).

Foreshadowing some key findings, consistent with prior results we find strong evidence for spatial dependence in regional COVID-19 data during the first pandemic wave. The main novel finding from our study is that this spatial correlation is significantly determined by job commuting—even when we control for other types of spatial dependence. Particularly, job commuting was found to be the first factor that significantly correlated with the spatial spread of COVID-19 after its international arrival in February 2020. With regard to the size of this propagation effect, we find that a doubling of new COVID-19 cases in neighboring regions interconnected through commuting flows is associated with an up to 20% higher number of new COVID-19 cases in a region. Importantly, we also find that the estimated spatial commuting channel (different from spatial dependence of general form) breaks down once mobility restrictions during the first lockdown are in place, which points to the effectiveness of such strict measures in mitigating spatial disease transmission. Our estimated effect size is comparable to recent other studies, such as for US cities reported in Glaeser et al. (2020).

From the perspective of public health policy, our results indicate that commuting linkages need to be considered and accordingly monitored as a relevant transmission channel when containment measures are planned. Though apparently effective, mobility constraints bear considerable socioeconomic costs, though, which are discussed in the end of this paper (Sect. 6). Before that, in the next section, we briefly summarize the related empirical literature on modeling the spatiotemporal dynamics of epidemic diseases, specifically COVID-19, that is used to guide our estimation approach. This is followed by a brief discussion of the empirical specification and estimator choice in Sect. 3. Section 4 presents the data and variables. In this section, we also refer to institutional details of containment measures in Germany and discuss the time span when new infections can be expected to be shown in the data. Empirical results are presented in Sect. 5. Section 6 concludes the paper. Additional topics are discussed in the online appendix.

2 Related literature

Our research approach draws on three strands of the empirical literature focusing on the spatial transmission channels of infectious diseases. The first strand deals with general empirical research on the role of human mobility (prior to COVID-19). The second strand relates to the nascent research field studying the link between mobility, mobility restrictions and disease spread during the first wave of the COVID-19 pandemic at an international scale. As third strand, we refer to recent contributions that model the spatiotemporal dynamics of COVID-19 and lockdown effectiveness

in Germany, particularly at the regional level. These strands of the literature are used as conceptual framework guiding our spatial econometric analysis.

2.1 Spatial transmission channels of infectious diseases

There is a broad consensus within the multidisciplinary field of epidemiological modeling that mobility and transport links, such as trade and commuting patterns, significantly affect the magnitude and speed of spatial disease transmission at national and global scales (see, for instance, Charaudeau et al. 2014; Cromley and McLafferty 2012; Wen et al. 2012). A particularly influential analysis tool for analyses in this context is the susceptible-infective-recovered (SIR) model (see Höhle 2016, for an overview).² SIR-type models can be spatialized by dividing the overall population into a number of sub-populations in different geographical areas and by allowing for various types of linkages between these sub-populations, for instance, proxied through the geographical distance between them. In the case of directly transmitted human diseases (such as COVID-19), a distinct focus is often placed on the temporary, recurrent movement of infectious individuals between sub-population, e.g., through job commuting and long-distance airplane travelling. A specific commuter-based SIR model is, for instance, presented in Tizzoni et al. (2014).

There are different ways to calibrate or estimate the key parameters of spatial SIR models.³ A particularly flexible empirical adaptation of the spatial SIR framework is the Held–Höhle–Hofmann (HHH) model (see Held et al. 2005, for a description and Lu and Meyer 2020, for a recent application), which decomposes the infection dynamics of regional sub-populations into an endemic and epidemic component. While the endemic component typically includes deterministic short- and long-run trends in new disease cases (e.g., seasonal variations), the epidemic dynamics is essentially estimated as an autoregressive specification of new disease cases (i.e., the reproduction of cases in the own region), regional population levels to proxy the stock-flow relationship between susceptibles, infectious and newly infected together with random regional effects and covariates capturing the infection dynamics in other, spatially interconnected regions. A standard way to incorporate spatial dependence in the HHH and related multivariate time-series SIR models is to rely on gravity-type mechanisms relating interregional disease transmission between sub-populations to the inverse geographical, social and/or economic distance between them (e.g., Xia et al. 2004).

2.2 Mobility and mobility restrictions during COVID-19

Several recent contributions have adopted SIR-type models to predict the development of COVID-19 cases at the country and regional levels (see, e.g., Atkeson 2020; Wang et al. 2020). In several of these models, mobility between sub-populations has been found as a significant factor predicting the spatiotemporal dynamics of

² Other types of region-scale models are, for instance, presented in Bertozzi et al. (2020).

³ We provide a summary of different methods in the online appendix.

COVID-19 (e.g., Chang et al. 2021; Liu et al. 2020; Wu et al. 2020, among others). Complementary approaches based on network analysis and spatial econometrics have confirmed the link between human mobility and the international transmission of COVID-19 for different connectivity measures such as flight connections and international trade relations (Krisztin et al. 2020).

COVID-19-related research is thereby particularly concerned with the role played by mobility restrictions in slowing down the disease spread during the first pandemic wave (e.g., Chinazzi et al. 2020; Gatto et al. 2020; McGrail et al. 2020; Klein et al. 2020a, b; Glaeser et al. 2020). While the focus in Chinazzi et al. (2020) is on the national and international COVID-19 transmission delay caused by the travel ban from Wuhan, China, in January 2020, Gatto et al. (2020) model the countrywide spread of the COVID-19 in Italy on the basis of provincial data and estimate the effects of restrictions posed to human mobility and human-to-human contacts. Their findings from a spatial SIR model indicate that lockdown measures associated with professional and social mobility constraints have significantly reduced COVID-19 disease transmission rates.

McGrail et al. (2020) investigate the effect of social distancing policies on community mobility. The authors find for US data from the Google mobility reports that average mobility rates in response to regional or national mitigation policies declined by 40–50%. Similarly, Klein et al. (2020a, b) find for interregional mobility data in the USA that COVID-related mobility constraints temporally coincide with a strong reduction in overall mobility patterns (commuting volumes between US census tracts and inter-city travel, i.e., visits of two US Census Bureau Combined Statistical Areas (CSAs) within 24 h) by up to 60%. Extending this descriptive evidence, Glaeser et al. (2020) study the causal impact of changes in mobility on the development of COVID-19 per capita for five US cities. The authors find that a decrease of roughly 20% in COVID-19 cases for every 10% points falls in mobility.

2.3 Predicting regional COVID-19 developments in Germany

Some first studies have also investigated the infection effects of COVID-19 in Germany and have assessed lockdown effectiveness. We focus on research with an explicit regional dimension here. Schlosser et al. (2020) use mobile phone data to investigate mobility patterns across German NUTS-3 regions (without distinguishing different types of mobility, though). Key findings from their network analysis of mobility changes are that (a) long-distance travel has been strongly reduced during the spring lockdown rendering mobility patterns more local and (b) the density of the German mobility network has been reduced. The authors use a spatial SIR model to assess the mediating effect of changes in a mobility network (although the model is not calibrated to COVID-19 parameters). They conclude that the reduced global mobility during a lockdown likely slows down the spatial disease spread.

Mixed evidence for lockdown effectiveness in Germany is reported in Berlemann and Haustein (2020). Their analysis is focused on predicting regional counts (NUTS-3 regions) of newly infected persons in the HHH-model framework and on

evaluating the effectiveness of three waves of containment measures during spring 2020. Estimation results indicate that particularly the first wave of containment measures (ban of mass events) contributed to flatten the curve of new infections. Similar to Berlemann and Haustein, results reported in Wieland (2020) provide mixed evidence for a regional trend change in infections around the timing of public interventions in spring 2020. As the author points out, in nearly two-thirds of German counties the flattening of the infection curve was found to occur before the full lockdown came into force on March 23 and one in eight counties even experienced a decline of infections even before the closures of schools, child day care facilities and retail facilities.

Kergassner et al. (2020) use a combined SIR-type and mobility network model to study the effect of infections stemming from selected “seeds” such as (a) Heinsberg in Germany during the Carnival season 2020 and (b) returning travelers from Ischgl in Austria. The authors conclude that refraining from traveling and large events are two key interventions that can effectively attenuate the spreading of COVID-19 at the regional level. Felbermayr et al. (2020) similarly report that the road distance to Ischgl is an important predictor for infection cases across German regions and that mobility restrictions have helped to suppress the further disease spread.

Taken together, while these earlier studies provide first important insights on the role of mobility and the effectiveness of containment measures in influencing the spread of COVID-19, no empirical investigation of the underlying transmission channels for these mobility effects has yet been undertaken. We will add to the scientific debate by providing an in-depth analysis of the role played by commuting-to-work patterns across NUTS-3 regions for the spatial spread of COVID-19.

3 Econometric specification

3.1 Spatial panel error correction model

The specification of our econometric model accounts for the key factors identified in the empirical literature of epidemic disease modeling: First, we follow the logic of HHH-type epidemiological models and place a space–time dynamic autoregressive mechanism at the heart of our short-run specification. Second, to cover disease transmission through stock–flow relationships (as stressed in SIR-type models), we propose an estimation approach that adapts the logic of space–time cointegration and error correction (e.g., Beenstock and Felsenstein 2010). Specifically, the latter framework allows us to add long-run information about the cumulative number of COVID-19 cases within and across regions in our short-run regression specification without the risk of running spurious, i.e., biased and regressions. The main driving force behind this model is the specification of a long-run co-movement (cointegration) path for the cumulative number of COVID-19 cases in interconnected regions,

whereby deviations from this long-run trend are expected to influence the regional short-run new case development.⁴

Formally, we estimate a spatial panel error correction model (SP-ECM) of the following form⁵

$$\begin{aligned} \Delta covid_{i,t} = & \sum_{l=1}^L \alpha_l \Delta covid_{i,t-l} + \sum_{m=1}^M \rho_m \left(\sum_{j=1}^N w_{ij} \Delta covid_{j,t-m} \right) \\ & + \sum_{m=1}^M \tau_m \left(\sum_{j=1}^N w_{ij}^{COM} \Delta covid_{j,t-m} \right) + \lambda u_{i,t-p} + D_{Weekday} + \mu_i + \Psi_{k(t)} + e_{i,t}, \end{aligned} \tag{1}$$

where $\Delta covid_{i,t}$ is the log-transformed number of new COVID-19 cases in region i at day t , which are regressed on $l=1, \dots, L$ time autoregressive and $m=1, \dots, M$ space-time lags of $\Delta covid_{i,t}$. Throughout this paper, log-transformed variables are written in lower-case letters, whereas the underlying original variables are written in upper-case letters, i.e., $\Delta covid_{i,t}$ is the log-transformation of the variable $\Delta COVID_{i,t}$ (details on variable definitions and the chosen log-transformation will be given below, particularly in Sect. 4).

We distinguish between two types of spatial transmission mechanisms in the short run: In first place, we are interested in identifying the role played by interregional disease transmission through interregional job commuting linkages. Accordingly, the included space-time lags $\left(\sum_{j=1}^N w_{ij}^{COM} \Delta covid_{j,t-m} \right)$ measure the (average) number of new COVID-19 cases in regions having commuting interdependencies with region i , where w_{ij}^{COM} is a measure for the strength of this interdependency used to calculate weighted averages for the $j=1, \dots, N$ regions (excluding region i). The magnitude of this channel of spatial disease transmission is captured by the regression coefficients $\sum_{m=1}^M \tau_m$. We thereby assume that spatial transmission does not take place instantaneously but with time lags. We apply different empirical concepts to operationalize w_{ij}^{COM} ; details are given in Sect. 4.

In order to avoid omitted variable biases stemming from other (latent) channels of spatial dependence, we also include a second spatial lag term $\left(\sum_{j=1}^N w_{ij} \Delta covid_{j,t-m} \right)$ in Eq. (1). Spatial weights w_{ij} thereby link region pairs (i, j) through the (inverse) geographical distance between them. The idea is that other types of social interactions, such as shopping behavior and friendship networks, can be proxied through

⁴ See the online appendix for a stylized simulation model that highlights the concept of (spatial) cointegration and error correction applied to the case of epidemic disease modeling at the regional level.

⁵ Different from the approach in Beenstock and Felsenstein (2010), we apply the SP-ECM framework to a univariate and not multivariate case. While the multivariate case allows the researcher to include two long-run residual adjustment terms in the SP-ECM equation, namely one for adjustment processes related to within-group cointegration across variables and a second “spatialized” residual term to account for between-group spatial adjustment processes, we only account for the latter through $u_{i,t-p}$. Moreover, while the inclusion of between-group adjustment processes often takes place on an ad hoc basis in spatial econometric applications, we provide theoretical priors with regard to the expected time path in the case of epidemic diffusion (see the online appendix).

these general space–time lags. We also control for region–fixed effects (μ_i), weekday effects (D_{Weekday}) and time–fixed effects for each $k=1, \dots, K$ calendar week in the sample period ($\Psi_{k(t)}$); finally, $e_{i,t}$ is the model’s *i.i.d.* error term. By including region–fixed effects, our empirical approach explicitly differs from related approaches such as the HHH model, which considers region–level effects to be at random (e.g., Berlemann and Haustein 2020). However, given distinct regional structures in the data such as a visible south–north divide in COVID–19 cases and urban–rural differences, we argue that such persistent structures are not a random representation and may lead to an estimation bias if the correlation of these structures with the included set of regressors (essentially historical information about COVID–19 cases) is nonzero.

The use of an (unrestricted) short–run dynamics measured through the space–time autoregressive (AR) components shown in Eq. (1) is a common feature of epidemiological models as outlined above. Maximum lag lengths are typically chosen by the researchers and should be ideally supported by a suitable test strategy.⁶ The underlying idea of this autoregressive short–run modeling approach is that historical information on new cases in region i and its spatial neighborhood captures important information on the local infection dynamics even if the maximum lag length chosen is shorter than theoretically expected incubation times at the individual level. In other words: The number of newly reported infections yesterday is assumed to be a suitable predictor for the number of newly reported infections today as they both—measured at the local population level—convey information about stocks of infectious persons back in time (together with a given transmission rate).

An exclusive focus on the short–run times–series dynamics may, however, neglect some important structural features of the data generation process, which is contained in variations in the number (stock) of infectious persons in a region and its spatial neighborhood (here: stocks denote the cumulative number of COVID–19 cases). Hence, Eq. (1) also considers that transmission is triggered by deviations from a long–run co–movement of regional stocks of infected individuals, i.e., cumulative COVID–19 cases. Specifically, in our SP–ECM framework, deviations from a long–run cointegration path in the infection curve of neighboring regions are captured through the term $u_{i,t-p}$ (with $p > l, m$), which is calculated as the residuals from a long–run regression equation of the following form

$$u_{i,t} = \left[\text{covid}_{i,t} - \hat{\phi} \sum_{j=1}^N w_{ij} \text{covid}_{j,t} - \Psi_{k(t)} - \mu_i \right]. \quad (2)$$

In Eq. (2), $\text{covid}_{i,t}$ is the (log–transformed) cumulative number of COVID–19 cases in region i at time t . The coefficient $\hat{\phi}$ expresses the estimated long–run spatial association between the stock of COVID–19 infections in region i and its spatial

⁶ For instance, Berlemann and Haustein (2020) use two lags ($L=2, M=2$) in the epidemic component of their HHH model for Germany arguing that this lag is sufficient to capture the time–series properties of new infections. However, the authors do not explicitly test their choice against some information criteria for evaluating the model fit.

neighborhood $\sum_{j=1}^N w_{ij} covid_{j,t}$. As an extension, we also estimate Eq. (2) including $\sum_{j=1}^N w_{ij}^{COM} covid_{j,t}$ to explicitly account for commuting interdependences in the long-run co-movement of regions along their infection curves.⁷ With regard to the expected transmission lag running from the stock of infectious to new cases through $u_{i,t-p}$, we set lag length for p as $p = \{8, \dots, 11\}$ days to capture the median incubation time for the case of COVID-19 of 5.2 days (Linton et al. 2020; Lauer et al. 2020) and a similarly sized median reporting lag for the German RKI data (Mitze et al. 2020). See also Sect. 4 for further details.

The long-run residual $u_{i,t}$ takes positive values if the cumulative number of COVID-19 cases in region i exceeds the average value in the spatial neighborhood. In this case, region i can be characterized as a relative hot spot of COVID-19 cases (vis-à-vis its spatial neighborhood) and we expect that this excess infection level marks another channel that helps predicting the number of new cases emerging in region i as outcome variable of our SP-ECM. As shown in the online appendix in greater detail, we can formulate ex ante expectations about the strength of spatial (dis-)equilibrating forces induced by this adjustment mechanism which depend on the state of regions along an S-shaped infection curve during an epidemic wave. That is, if regions move along their S-shaped infection curves below its turning point, we expect that the coefficient λ for $u_{i,t-p}$ is positive and exerts a disequilibrating force, i.e., it increases the differences in new cases between region i and its geographical neighborhood in the short run. Beyond this turning point, however, we expect that an equilibrating mechanism comes into play. Accordingly, tracking the evolution of λ over time allows us to predict the peak of an epidemic wave.

In terms of estimator choice, we estimate the short- and long-run specification shown in Eqs. (1) and (2) on the basis of linear panel data models for log-transformed data. Specifically, the short-run SP-ECM in Eq. (1) is estimated straightforwardly as dynamic fixed effects model (DFE) by ordinary least squares (OLS) given that it only includes predetermined space-time lags of $\Delta covid_{i,t}$ as right-hand-side regressors, i.e., we do not need to account for the endogeneity of the included spatial lag terms. We refrain from applying bias-correction or generalized method of moments (GMM) estimation methods for the short-run equation since earlier research has shown that the estimation bias in the AR coefficients becomes negligible for sufficient time periods (Judson and Owen 1999). Similarly, Korniotis (2010) has shown through Monte Carlo simulations that with increasing number of time observations the reduction in the percentage mean bias in $\sum_{m=1}^M \rho_m$ is larger for the DFE estimator compared to a bias corrected estimator or the Andersen-Hsiao-type IV estimator. Finally, GMM methods are typically found to suffer from instability with regard to the choice of overidentifying restrictions used for estimation (e.g., Ziliak 1997; Windmeijer 2005).

The long-run equation as shown in Eq. (2) is estimated by (quasi-)maximum likelihood including region- and time-fixed effects (Elhorst 2014). The choice of maximum likelihood estimation is mainly motivated by the contemporaneous inclusion

⁷ See the online appendix for conceptual details.

of the spatial lag of $covid_{i,t}$ as endogenous right-hand-side regressor. For the case that the long-run specification in Eq. (2) contains more than one spatial weighting matrix, we impose exogeneity assumptions on the spatial lag terms included to facilitate estimation. Finally, as stated above, for both the short- and long-run specification, we argue that the inclusion of two-way (region- and time-) fixed effects is crucial factor for ensuring estimator consistency in the presence of unobserved regional heterogeneity. We thereby essentially deviate from the estimation practice chosen in related studies on the HHH model such as Berlemann and Hausteiner (2020), which assume that time-constant differences in regional infection cases are randomly distributed across regions. The inclusion of region-fixed effects together with space–time dynamics motivates our choice for using a linear panel data model instead of count data regression approaches.⁸

While it seems natural on first sight to apply count data estimators for our analysis of new COVID-19 infections at the regional level, we follow the methodological literature arguing that standard count data estimators, such as the Poisson conditional maximum likelihood estimator with individual specific constants, are inconsistent in settings that involve (two-way) fixed effects together with a dynamic panel data structure (Blundell et al. 2002) and may thus result in severe estimation biases.⁹ As Manning and Mullahy (2001) report, the log-transformed OLS estimator deals particularly better with heavy-tailed data than many generalized linear model (GLM) alternatives, such as Poisson-like and gamma models. While the log-transformation of data with zero entries is not without limitations either (Bellégo and Pape 2019), we carefully account for the pitfalls of this linearized dynamic panel data approach and test for biases associated with the transformation and back-transformation of the data for model predictions.

3.2 Analysis of the effectiveness of containment measures

To study the effectiveness of containment measures on spatial disease transmission, we test if the spatiotemporal dynamics in newly registered COVID-19 cases changes over time with and without containment measures in place. A similar identification approach for the estimation of time-dependent parameters in SIR models has, for instance, been adapted in Hong and Li (2020) studying the COVID-19 development in Hong Kong. We make use of an interaction term approach which multiplies the binary dummy variables for individual weeks ($\Psi_{k(t)}$) with the set of regressors shown in Eq. (1). While the additional number of estimated coefficients makes the model cumbersome for forecasting or simulation purposes, from an (ex-post) policy evaluation perspective it allows us to identify trends in the space–time COVID-19 spread within and across regions and thus to test if the spatial transmission dynamics

⁸ We, though, apply count data models as a robustness check. Details are given in the online appendix.

⁹ As discussed in Blundell et al. (2002), a solution to this inconsistency problem would be to apply a quasi-differences GMM estimator, which may, however, suffer from a weak instrument bias if time series are highly persistent.

Table 1 Summary statistics for newly reported and cumulative COVID-19 cases by regions

Variable	Mean	SD	Min	Max	Stationary time series	Spatial dependence
$\Delta\text{COVID}_{i,t}$	4.13	10.66	0	310	Yes	Yes
$\text{COVID}_{i,t}$	120.86	289.07	0	5795	No	Yes

$\Delta\text{COVID}_{i,t}$ =newly registered COVID-19 cases for region i at day t ; $\text{COVID}_{i,t}$ =cumulative number of registered COVID-19 cases. The dataset relates to 401 NUTS-3 regions and 95 days between January 28 and May 1, 2020. Data from RKI (2020). Details on panel unit root and Moran’s I tests for spatial dependence are given in the online appendix. These tests have been applied to log-transformed variables. Summary statistics for log-transformed variables are presented in the replication files available for this study.

has slowed down after the introduction of mobility constraints. Formally, our time-dependent parameter specification has the following form

$$\begin{aligned}
 \Delta\text{covid}_{i,t} = & \underbrace{\sum_{k=1}^K \sum_{l=1}^L \alpha_{l,k} (\Delta\text{covid}_{i,t-l} \Psi_{k(t)})}_{\text{Component 1}} \\
 & + \underbrace{\sum_{k=1}^K \sum_{m=1}^M \tau_{m,k} \left(\left(\sum_{j=1}^N w_{ij}^{\text{COM}} \Delta\text{covid}_{j,t-m} \right) \Psi_{k(t)} \right)}_{\text{Component 2}} \\
 & + \underbrace{\sum_{k=1}^K \sum_{m=1}^M \rho_{m,k} \left(\left(\sum_{j=1}^N w_{ij} \Delta\text{covid}_{j,t-m} \right) \Psi_{k(t)} \right)}_{\text{Component 3}} \\
 & + \underbrace{\sum_{k=1}^K \varphi_k (u_{i,t-p} \Psi_{k(t)})}_{\text{Component 4}} + \underbrace{\Psi_{k(t)}}_{\text{Component 5}} + D_{\text{Weekday}} + \mu_i + e_{i,t}.
 \end{aligned} \tag{3}$$

Components 1–3 allow us to gain insights on the temporal heterogeneity in the short-run dynamics of the COVID-19 spread in Germany; *Component 4* tests for time-specific effects in the deviation/adjustment toward a long-run cointegration path between regions, which can be used to detect local hot spots and identify the peak of the pandemic wave; and finally, *Component 5* tracks the overall development of new cases over time, i.e., it is a general measure of the overall dynamic development of COVID-19 in Germany (given that we account for differences across regions through region-fixed effects). This latter component can be seen as the unrestricted version of a deterministic time trend plus seasonal components typically included in the HHH model.

4 Data, descriptive analysis and institutional details

4.1 Data, variables and spatial weighting schemes

We use daily statistics on newly reported COVID-19 cases at the level of German NUTS-3 regions obtained from the Robert Koch Institute (RKI 2020) which is in charge of disease control in Germany. The RKI database is fed from information of local health authorities and only covers cases confirmed by a testing laboratory. We use the RKI data as main input to build a balanced panel dataset for 401 NUTS-3 regions and 95 days between January 28 and May 1, 2020 (38,095 region-time observations). Our outcome variable of interest are newly reported COVID-19 cases ($\Delta\text{COVID}_{i,t}$) and their cumulative stocks, where the latter are constructed as

$$\text{COVID}_{i,t} = \text{COVID}_{i,t-1} + \Delta\text{COVID}_{i,t} \quad (\text{for } t > 1).$$

Note that log-transformed variables introduced in Sect. 3 are written in lower-case letters, i.e., $\text{covid}_{i,t}$ and $\Delta\text{covid}_{i,t}$ are logarithmic transformations of $\text{COVID}_{i,t}$ and $\Delta\text{COVID}_{i,t}$, respectively. Summary statistics for both variables together with an indication of the time-series properties and the significance of spatial dependence in the data are given in Table 1 (for details on the applied panel unit root tests and Moran's I tests see the online appendix). The pre-estimation tests indicate that we can clearly reject stationarity of cumulative stocks of COVID-19 cases, which supports our chosen SP-ECM approach. We also find persistent evidence for spatial dependence in the data.

As discussed in Sect. 3, the reader has to note that we use log-transformations of both outcome variables for estimation, where the log-transformation for any variable X is computed as $x = \log(X + 1)$. This type of log-transformation considers that newly reported COVID-19 cases may be zero at the regional level (McDonald 2014) while assuming that the variable follows a log-normal distribution (e.g., Olsson and Zhang 2020, for the case of COVID-19 time series). We also account for the problem of back-transforming the data in order to use our SP-ECM specification for predictions of new and cumulative COVID-19 cases at the regional and national level. Specifically, we make sure that level predictions from log-transformed data do not suffer from a re-transformation bias that arises when predictions of the log-dependent variable are exponentiated. We apply the re-transformation approach suggested by Duan (1983) for potentially non-normally distributed residuals ε from a log-transformed linear regression equation as $y = \mathbf{x}'\Psi + \varepsilon$. In this case, the adjusted re-transformation is given by $Y = \exp(\mathbf{x}'\Psi)E[\exp(\varepsilon)]$, with $E[\exp(\varepsilon)] = \frac{1}{NT} \sum \exp(\varepsilon_{i,t})$. We finally make sure that predicted values are corrected for the add factor (+1) introduced during the logarithmic transformation.

When it comes to the measurement of interregional commuting linkages, we construct w_{ij}^{COM} as elements of an $N \times N$ spatial weighting matrix \mathbf{W}^{COM} in different ways. As default specification, we use binary dummies indicating if region i and j belong to a joint commuting zone ($w_{ij}^{\text{COM}} = 1$) or not. We employ two concepts defining joined commuting zones in Germany. These relate to (a) the German Federal Institute for Research on Building, Urban Affairs and Spatial Development

(Bundesinstitut für Bau-, Stadt- und Raumforschung, BBSR), and (b) Kropp and Schwengler (2011). The BBSR concept builds on a narrow definition of commuting zones and segments the 401 German NUTS-3 regions into 258 local labor markets with strong commuting flows. In all cases, regions belonging to a joint commuting zone also share a common territorial border, i.e., are immediate spatial neighbors.¹⁰ Kropp and Schwengler (2011) adopt a wider definition and make use of graph theory to arrive at 50 larger commuting zones.¹¹ Different from the BBSR concept, Kropp and Schwengler also group regions further away from each other into joint commuting zones if they have strong commuting interdependencies. In both cases, we row-standardize elements of \mathbf{W}^{COM} . This allows us to interpret the spatial lag terms generated as the average number of newly registered COVID-19 cases in the other regions belonging to the same commuting zone as region i .

As an alternative to these binary indicators, we also measure the intensity of labor market interaction directly through gross number of commuter flows between region i and j (summed over in- and out-commuting for each pair of regions). Data on gross commuter flows have been obtained from German labor market statistics (provided by the Institut für Arbeitsmarkt- und Berufsforschung, IAB) and are based on annual data for 2019.¹² As before, we row standardize the obtained matrix \mathbf{W}^{COM} . Importantly, the use of data on gross commuting flows at the onset of the COVID-19 outbreak ensures the exogeneity of space–time lags calculated on the basis of \mathbf{W}^{COM} . Summary information for the three different commuting-based weight matrices to track the spatial association of newly registered COVID-19 cases across regions is reported in the online appendix.

With regard to distance-based spatial dependence (w_{ij}), as default, we construct a binary dummy which takes a value of one if region i and j share a common territorial border (i.e., queen-type first-order contiguity). We employ a measure of inverse distances between region centroids as an alternative; weights are row standardized.¹³

Importantly, the inclusion of two different types of spatial transmission channels allows us to test for additionalities in the spatial dynamics of the epidemic diffusion in Germany. When we interpret the resulting spatial effects, we carefully account for the relationship between the two weighting schemes w_{ij} and w_{ij}^{COM} . For instance, when we include space–time lags constructed on the basis of the narrow BBSR definition for joint commuting zones together with spatial lags for regions sharing a common territorial border, we can exploit the fact that the former weighting scheme is perfectly nested within the latter. Accordingly, the coefficient for the spatial lag term for job commuting then measures the additional disease transmission effect for regions being in a joint commuting zone on top of a general spatial association in

¹⁰ See <https://www.bbsr.bund.de/BBSR/DE/forschung/raumbeobachtung/Raumabgrenzungen/deutschland/regionen/AMR/Arbeitsmarktregionen.html> for further details.

¹¹ Updated data for the most recent definition of commuting zones in Germany in 2016 are obtained from the supplementary online materials available at: <https://www.iab.de/389/section.aspx/Publikation/k110222301>.

¹² For further information on the IAB gross commuting database, see also <http://www.iab.de/infoplattform/pendler>.

¹³ Summary information on spatial weight matrices is given in the online appendix.

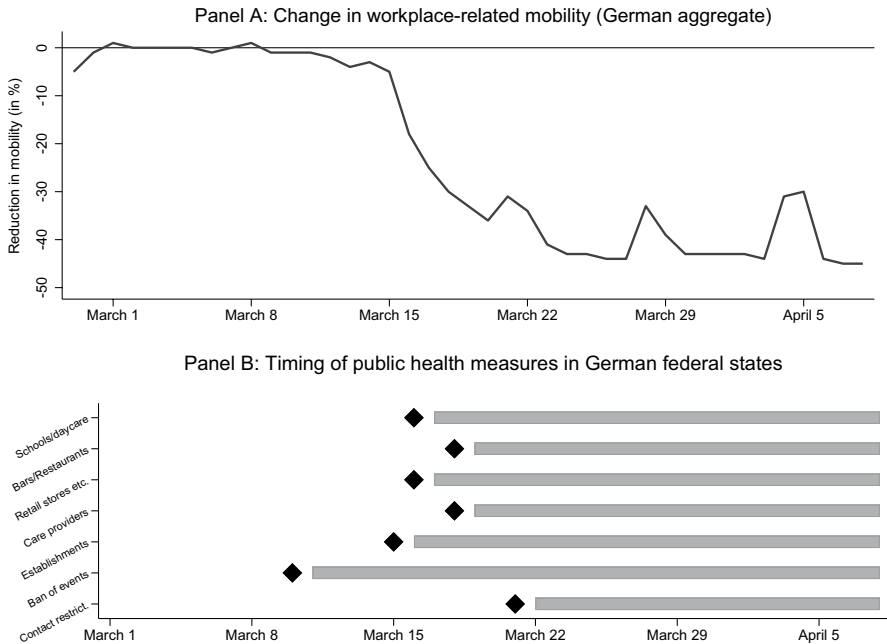


Fig. 1 Changes in workplace-related mobility and timing of containment measures. *Notes:* Data for Panel A are obtained from Google mobility reports (Google 2020). The timing of the start of public health measures (black diamonds) indicates the point in time when the first federal state imposed a particular containment measure

COVID-19 cases. While other combinations for w_{ij} and w_{ij}^{COM} are only partly overlapping, e.g., spatial lag variables defined on inverse distances and gross commuter flows, the online appendix shows in greater detail, that for no pair of spatial lag variables we observe an almost perfect correlation. This is important to rule out the problem of multicollinearity.

4.2 Timing of containment measures and visibility of potential effects in the data

An assessment of the effectiveness of containment measures during the first pandemic wave chiefly depends on (a) exact information about the timing of restrictions and (b) precise expectations about their likely visibility in the data. Elaborating on the first aspect, we have systematically reviewed public health measures implemented in German federal states as responsible decision-making body for public health regulations. This allows us to separate our sample period into two sub-periods with and without containment measures in place. While no (systematic) measures were undertaken until March 9, federal states imposed various restrictions on economic activities and social contacts afterward. As shown in Panel B of Fig. 1, this imposition has been done in a staggered manner, i.e., diamonds in Panel B of Fig. 1 indicate the point in time when the measure was implemented by the first German federal state. Most measures have

been implemented during the period March 15 to 22 and include the closing of restaurants, shopping centers, retail stores and further private and public establishments (including museums, theaters, parks, etc.). This was completed by a ban of events and even personal contacts constraints up to strict stay-at-home orders in some German federal states.

To further narrow down the time window when policy measures can be expected to impact on the spatial disease transmission, we have additionally screened the COVID-19 community mobility reports published by Google (2020). These reports map movement trends over time across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces and residential for Germany (aggregate). We are particularly interested in screening workplace-related mobility trends. Unfortunately, data are only available at higher aggregation levels and not for NUTS-3 regions. As Panel A of Fig. 1 shows for the German aggregate, we can see a distinct drop in workplace-based mobility from March 16 onwards (calendar week 12). At around March 23, i.e., calendar week 13, most mobility adjustments had taken place resulting in a relative reduction of mobility of 40% compared to a five weeks reference period between January 3 and February 6, 2020. This drop on overall mobility provides an important prior for our analysis of time heterogeneities in the estimated SP-ECM as shown in Eq. (2).

Finally, to be able to formally test for the effect of policy interventions on the evolution of new COVID-19 cases, we need to consider that reported cases only appear in the data with a certain time lag. Two aspects determine this time lag: First, we need to consider an incubation time. Prior research has documented a median incubation time of 5.2 days for COVID-19 infections (Linton et al. 2020 and Lauer et al. 2020). Secondly, we need to account for a reporting lag, i.e., the time period between the outbreak of the disease (first symptoms) and the recording of a new COVID-19 case in the data. The median reporting lag for German data is found to be of almost similar size as the incubation time, which implies a median overall transmission lag of approximately 10.5 days (see Mitze et al. 2020).

5 Empirical results

Table 2 reports the estimated short- and long-run coefficients for alternative SP-ECM specifications. Optimal lag selection has been guided by statistical information criteria (see Table S6 in the online appendix for details).¹⁴ The table reports coefficient sums for the AR component and space-time lags together with associated standard errors calculated on the basis of the Delta method. For all specifications, we find that newly reported COVID-19 cases are autoregressively distributed over time; in most cases, the estimated coefficient sum for $L=4$ is about 0.5, i.e., a doubling of new cases in the last four days leads to an increase in new cases by roughly

¹⁴ The model's explanatory power increases with the inclusion of additional lags starting from a simple AR(1) specification but becomes smaller for a lag size larger than $L=4$ and $M=4$. We thus argue that the latter lag structure is sufficient to capture both the short-run temporal and spatial dynamics of new infections.

Table 2 Spatiotemporal estimation of SP-ECM for newly reported COVID-19 cases in German NUTS-3 regions

Dep. Var.: $\Delta\text{covid}_{i,t}$	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Weighting matrix 1 (w_{ij})	First-order contiguity	First-order contiguity	First-order contiguity	First-order contiguity	Inverse distances	Inverse distances	Inverse distances	Inverse distances
Weighting matrix 2 (w_{ij}^{COM})	None	Commuting zones (BBSR)	Commuting zones (K&S)	Commuting flows	None	Commuting zones (BBSR)	Commuting zones (K&S)	Commuting flows
SP-ECM								
Time autoregressive lag	0.541*** (0.0172)	0.536*** (0.0175)	0.538*** (0.0177)	0.538*** (0.0174)	0.579*** (0.0171)	0.564*** (0.0175)	0.563*** (0.0178)	0.562*** (0.0171)
$\sum_{j=1}^L \Delta\text{covid}_{j,t-l}$								
General space-time lags	0.192*** (0.0158)	0.169*** (0.0170)	0.171*** (0.0170)	0.152*** (0.0206)	0.137*** (0.0376)	0.101*** (0.0387)	-0.027 (0.0423)	-0.079* (0.0424)
$\sum_{m=1}^M \left(\sum_{j=1}^N w_{ij} \Delta\text{covid}_{j,t-m} \right)$								
Commuting space-time lags		0.033*** (0.0095)	0.037** (0.0173)	0.075*** (0.0234)		0.060*** (0.0089)	0.155*** (0.0181)	0.240*** (0.0210)
$\sum_{m=1}^M \left(\sum_{j=1}^N w_{ij} \Delta\text{covid}_{j,t-m} \right)$								
L.R. adjustment parameter $u_{i,t-p}$	0.148*** (0.0135)	0.150*** (0.0134)	0.149*** (0.0136)	0.146*** (0.0133)	0.158*** (0.0127)	0.158*** (0.0126)	0.158*** (0.0131)	0.151*** (0.0130)
Long-run equation								
General spatial lag $\sum_{j=1}^N w_{ij} \text{covid}_{j,t}$	0.606*** (0.0047)	0.489*** (0.0053)	0.425*** (0.0061)	0.361*** (0.0065)	0.456*** (0.0156)	0.428*** (0.0158)	0.393*** (0.0164)	0.330*** (0.0161)
Commuting-based spatial lag		0.038*** (0.0019)	0.234*** (0.0059)	0.552*** (0.0086)		0.068*** (0.0022)	0.351*** (0.0064)	0.625*** (0.0087)
$\sum_{j=1}^N w_{ij}^{\text{COM}} \Delta\text{covid}_{j,t}$								
Observations ($N=401, T=85$)	34,085	34,085	34,085	34,085	34,085	34,085	34,085	34,085
Region-fixed effects	YES	YES	YES	YES	YES	YES	YES	YES
Time-fixed effects (weekly)	YES	YES	YES	YES	YES	YES	YES	YES
Dummies for weekdays	YES	YES	YES	YES	YES	YES	YES	YES
Panel unit test for $u_{i,t-p}$ (p val.)	0.01	0.02	0.00	0.00	0.99	0.93	0.85	0.10
Breusch-Godfrey (p val.)	0.00	0.00	0.00	0.97	0.06	0.23	0.35	0.56
Relative RMSE	0.83	0.83	0.83	0.82	0.82	0.82	0.82	0.82

BBSR, Bundesinstitut für Bau, Stadt- und Raumforschung; K&S, Kropp and Schwengler (2011). The CIPS test is applied to check for the stationarity of the residuals ($u_{i,t-p}$) from the long-run regression equation; the Breusch-Godfrey LM test checks for serial correlation in the residuals of the AR(1) and SP-ECM specifications; the relative root mean square error (RMSE) as a measure for the in-sample predictive performance is calculated as the ratio of RMSE [model] / RMSE [AR(1)].

***, **, * Statistical significance at the 1%, 5% and 10% level. Standard errors (S.E.) in brackets are clustered by regions

50% on day t . We also find a statistically significant degree of spatial dependence in interregional disease transmission once we include general space–time lags $\left(\sum_{j=1}^N w_{ij} \Delta \text{covid}_{j,t-m}\right)$ in the SP-ECM. In our baseline specification in column (I), the estimated overall effect of 0.192 indicates that a doubling in the average number of newly reported COVID-19 cases in the spatial neighborhood of region i (in the last four days) is associated with a roughly 20% higher number of new cases in region i on day t . This significant degree of spatial dependence for new COVID-19 cases also holds for alternative spatial weighting schemes applied, although effect size differs moderately (compare, for instance, column (I) and column (V)).

Having established the fact that spatial dependence strongly matters in the data generating process of new COVID-19 cases, our focus is on identifying the role played by interregional commuting linkages (see columns (II) to (IV) as well as columns (VI) to (VIII) for alternative weighting schemes and combinations with general spatial lag terms). In all cases, the included space–time lags measuring interregional commuting linkages turn out to be statistically significant and positive – even if we control for spatial dependence of general form and thus other types of mobility effects.¹⁵ With regard to the interpretation of effect size, the following can be stated: If we take the example of column (III), which measures commuting linkages through a binary dummy for regions belonging to a common local labor market, the estimated coefficient sum in Table 2 for $\left(\sum_{j=1}^N w_{ij}^{\text{COM}} \Delta \text{covid}_{j,t-m}\right)$ measures the additional degree of spatial association in newly reported COVID-19 cases for regions in a common commuting zone on top of the spatial association observed for spatial neighboring regions (measured through the general spatial lag terms). Thus, the total strength of the commuting channel can be calculated as the sum to individual effects from the two types of space–time lags as $0.169 + 0.033 = 0.20$. This indicates that a doubling in the average number of newly reported COVID-19 cases in other regions of the same commuting zone are associated with an 20% higher number of newly reported COVID-19 cases in region i .

Using the broader definition of commuting zones suggested by Kropp and Schwengler (2011), as shown in column IV, delivers almost identical estimates. When we look at the estimation results that combine spatial lags calculated on the basis of metric information (gross commuter flows and inverse distances), the results in Table 2 further point to (a) larger spatial transmission effects when gross commuter flows are used instead of a binary dummy for a common commuting zone (given that higher order neighbors are considered as well) and (b) that commuting-based interregional dependence dominates the overall effect size stemming from spatial dependence in the data generating process. This can, for instance, be seen in columns (VII) and (VIII) showing that space–time lags calculated on the basis of inverse distances turn nonsignificant or negative once we include space–time lags calculated on the basis of gross commuter flows. This is a clear indication that

¹⁵ Regression results only including commuting-based space–time lags are reported in the online appendix.

geographical interactions through job commuting are a significant determinant of the observed degree of spatial disease transmission in the data.

Taken together, the reported short-run parameters for the included space–time lags point to a significant propagation effect of commuting to work on the development of new COVID-19 cases. The overall effect size is found to be quite stable across the different specifications shown in Table 2 (once we account for the degree of nestedness of spatial lag terms) and ranges between 0.15 and 0.23. This indicates that a doubling of the (average) number of new cases in the last four days in regions have spatial commuting interdependences with region i leads to an increase in new cases in region i of approximately 15 to 23% on day t .¹⁶ Together with the within-region autoregressive component of 0.5, the SP-ECM parameter estimates thus underline that the short-run spatiotemporal variation of COVID-19 is a significant channel of disease transmission at the regional level. When interpreting these effects, one should keep in mind, though, that all regression parameters in Table 2 measure average effects over the entire sample period, i.e., without and with containment measures in place.

In addition to these spatiotemporal short-run dynamics, in all model specifications shown in Table 2, we find a significant long-run correlation of cases between regions associated through spatial distance and commuting flows¹⁷ and further that the coefficient for the adjustment parameter ($u_{i,t-p}$) is statistically significant and positive in all cases. This points to the fact that the COVID-19 dynamics in Germany is significantly determined by local hot spots that antecede the infection development in their broader geographical neighborhood. When we accordingly apply the SP-ECM approach for hot spot detection, Fig. 2 plots the resulting spatial distribution of the long-run adjustment coefficient $u_{i,t}$ for selected points in time.

Accordingly, if we look at Panel A plotting the spatial distribution of long-run residuals for March 7, the three NUTS-3 regions with the largest deviations of cumulative COVID-19 cases from their geographical neighborhood are Heinsberg, the Main-Tauber-Kreis and Freising. All three regions have been epidemic epicenters during the first COVID-19 wave in Germany, for which the local disease dynamics can be traced back to individual superspreading events, for instance, related to

¹⁶ The online appendix additionally assesses the robustness of the estimated effects for alternatively specified outcome variables in the linear dynamic panel data model and alternative estimators including a negative binomial (NegBin) panel data model with random and fixed effects. However, since model properties of the NegBin specification are unclear in dynamic panel settings (e.g., related to the initial condition problem in nonlinear models), the obtained (unadjusted) estimates should only be seen as a rough proxy for comparison purposes. All in all, these specifications provide similar (or even higher) estimates for the included space–time lag coefficients for the short-run dynamics and thus underline parameter robustness in the baseline SP-ECM rendering the reported DFE coefficients conservative estimates of the propagation effect of commuting to work.

¹⁷ In the estimation of the long-run specification as shown in Eq. (2), we avoid estimating specifications with higher order spatial weights by assuming strict exogeneity of one of the two included spatial lag terms. The results show that the estimated coefficients remain very stable regardless which spatial lag term is treated as exogenous right-hand side regressor. For the results shown in Table 2, we have maximized the concentrated log likelihood for the general spatial weighting matrix \mathbf{W} while included the commuting-based spatial lag term as an exogenous regressor.



Fig. 2 Local hot spot identification from residuals of long-run spatial regression equation. *Notes:* In each Panel, names indicate the top-3 regions with largest residual values for the respective dates. Residual values are calculated for the model specification with first-order contiguity spatial weighting matrix and commuting links measured through gross commuter flows (see the long-run equation in Column (IV) of Table 2)

vacationers returning from ski holidays (in the Main-Tauber-Kreis, the first 21 out of 23 infections were detected among travelers from a tour group returning from

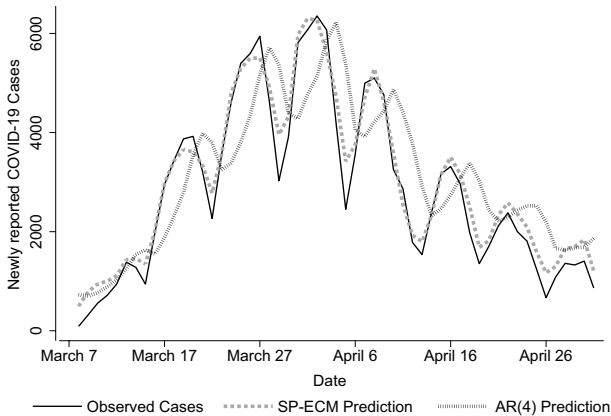


Fig. 3 In-sample predictions for newly reported COVID-19 cases (March–May 2020). *Notes:* Reported level predictions from log-transformed model specifications are based on the Duan (1983) re-transformation method to reduce the re-transformation bias that arises when predictions of the log-transformed dependent variable are exponentiated; predicted values have further been corrected for the add factor $(x + 1)$ in the logarithmic transformation. Model predictions for SP-ECM based on first-order contiguity spatial weighting matrix and commuting linkages measured on the basis of gross commuter flows; AR(4)=time-series autoregressive specification with $L=4$; daily count along x-axis

South-Tyrol); in the case of Heinsberg, the COVID-19 spread was triggered through social events in the carnival season (see Kuebart and Stabler 2020).

Over time, Fig. 2 indicates that larger urban areas, such as Hamburg and Hannover, exhibit a more dynamic COVID-19 development compared to their respective hinterlands. Finally, the residual plots also highlight why certain policy choices were taken. For instance, the local administration in Jena was Germany's first city to mandate the mandatory use of face masks in public transport and sales shops in early April (Mitze et al. 2020). As Panel B in Fig. 2 shows, this policy intervention coincides with a relatively high number of infections in Jena relative to its spatial neighborhood in late March. The visualization of long-run residuals in Fig. 2 thus indicates that our proposed SP-ECM model can be used as a tool for the space–time surveillance of hot spots to identify locations for targeted policy interventions (Desjardins et al. 2020). In addition, the inclusion of $u_{i,t-p}$ as additional source of newly reported COVID-19 cases adds a further predictive element to our SP-ECM approach, which complements the short-run parameter estimates.

Taken together, we find that the in-sample predictive performance of the spatially augmented SP-ECM increases by up to one fifth compared to a purely autoregressive specification (evaluated on the basis of the relative root mean square error [RMSE] as reported in Table 2).¹⁸ Moreover, for most SP-ECM specifications we do not detect residual autocorrelation indicating a good fit (see Breusch–Godfrey LM test in Table 2). And lastly, when we evaluate the model's in-sample predictive

¹⁸ The online appendix additionally shows the contribution of the individual model components for the in-sample fit on the basis of proportional reduction in error (PRE) tests.

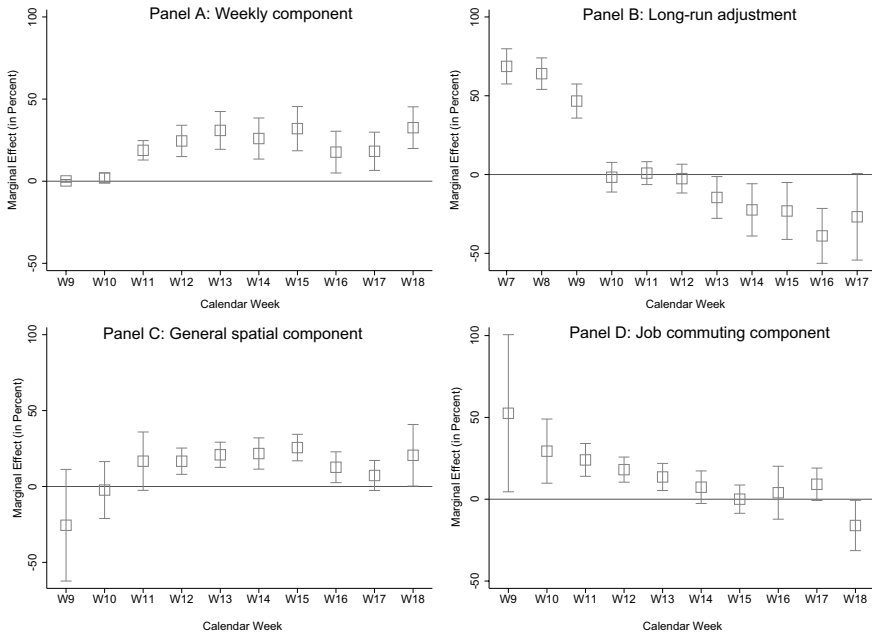


Fig. 4 Temporal heterogeneity of estimated SP-ECM components by calendar weeks. *Notes:* Results are shown for the SP-ECM model specification with general spatial dependence based on first-order contiguity weighting matrix and commuting linkages measured on the basis of gross commuter flows; coefficient plots show point estimates together with 99% confidence intervals. Calendar week 9 (W9) starts on February 24, 2020, W10 on March 2, W11 on March 9, W12 on March 16, W13 on March 23, W14 on March 30, W15 on April 6, W16 on April 13, W17 on April 20, W18 on April 27

performance, Fig. 3 shows that the SP-ECM properly tracks the dynamic pattern of new COVID-19 cases for the aggregated German development. Importantly, the figure also shows that the chosen log-transformation and the subsequent back-transformation do not bias the predictions new cases in levels. The SP-ECM clearly outperforms the predictive performance of an AR(4) benchmark specification.

By accounting for time heterogeneity in the parameters of the SP-ECM, we finally use the model to assess the effects of containment measures on the spatial transmission of COVID-19. Figure 4 therefore plots the estimated coefficients (rescaled to percentage effects) of the different model components by calendar weeks. Panel A and B plot the coefficients for the level (trend) component and the adjustment to the long-run spatial association between regions, respectively. While the trend component reflects that the number of new COVID-19 cases has increased over time between calendar weeks 10 and 13, we observe a side-ward trend afterward and a consecutive reduction in effect size between calendar weeks 15 and 17. Accounting for incubation times and a reporting lag, this period coincides with the timing of imposed containment measures. The evolution of the estimated trend component in Panel A thus indicates that these measures were successful in suppressing the overall spread of the virus, although

effects stay significant over time (evaluated against the baseline period before calendar week 9).

The temporal distribution of the spatial adjustment component in Panel B of Fig. 4 supports our theoretical considerations that a spatial disequilibrating force (driven by emerging local hot spots) is observed if regions move along an exponential growth path of infections. Our estimation results indicate that this was particularly relevant in the initial phase of the first pandemic wave in Germany during calendar weeks 7 to 9. Hereafter, the effects turn nonsignificant pointing to a flattening of the German infection curve. With the imposition of mobility constraints, we further get evidence for an emerging equilibrating process as indicated by negative coefficients for $u_{i,t-p}$. Together with the general trend results in Panel A, this indicates that the first wave of new COVID-19 cases peaked around calendar week 13 (starting on March 23, 2020).

Finally, Panels C and D allow us to make statements about the degree of spatial association between regions during this first wave of COVID-19 cases in Germany. With regard to general spatial effects, Panel C indicates that spatial correlation (first-order contiguity) gradually builds up and turns out to be positive and statistically significant from calendar week 12 onward. In the following weeks, the COVID-19 infection dynamics in a region's immediate geographical neighborhood constitutes a relevant transmission channel until week 17 (with effect size gradually declining after week 14).

In comparison, for spatial dependence determined by commuter flows, Panel D points to an earlier, though, temporally limited time window of significant transmission effects. Here, we find a significant spatial association already from calendar week 9 onwards, which phases out after calendar week 14 (starting on March 30, 2020). Taking a median incubation time and a reporting lag of 10.5 days into account, this phasing out of spatial association through commuting flows clearly coincides with the imposition of mobility constraints in mid-March as shown in Fig. 1. This result also holds if we disaggregate newly reported COVID-19 cases by age groups (see the online appendix for more details). The latter disaggregation shows that COVID-19 transmission follows a similar spatial pattern across age groups for aggregate regional data. Taken together, our SP-ECM approach provides important empirical evidence concerning the effectiveness of lockdown measures in suppressing the spatial spread of COVID-19 in Germany. The commuting-based channel of disease transmission quickly breaks down after their imposition.

6 Closing remarks

In this paper, commuting-to-work patterns have been identified as the first significant spatial transmission channel of COVID-19 across German regions after its international arrival in February 2020. We have estimated a spatial panel error correction model (SP-ECM) that allows us to track the short-run dynamics of the spatiotemporal evolution of new COVID-19 cases in Germany together with a theoretically grounded adjustment processes accompanying a long-run

co-movement of cumulative cases between spatially associated regions. We find that spatial dependence generally matters for the regional evolution of newly reported cases in German NUTS-3 regions and that commuter flows between regions mark a relevant channel of epidemic diffusion.

With regard to the overall effect size of spatial transmission, we find that a doubling of newly reported COVID-19 in regions interconnected with region i through job commuting (and other types of interregional mobility) is associated with an approximately 20% higher number of newly reported COVID-19 cases in region i . In addition to the spatiotemporal short-run dynamics, we also find that disease trajectories for German NUTS-3 regions show a distinct spatial component in their long-run development. Importantly, deviations from this spatial long-run relationship can be used to identify and surveil emerging local hot spots constituting a significant effect channel triggering the short-run dynamics of COVID-19.

Another key finding from our empirical analysis is that spatial transmission channels vary over time. While spatial disease transmission across NUTS-3 regions strongly matters during the initial phase of the first pandemic wave of COVID-19 in Germany, particularly correlating with commuting-to-job patterns early on, the timing containment measures during the German lockdown in mid-March 2020 coincides with a significant decline in the degree of spatial dependence across regions. This is most strongly shown for the commuting-based transmission channel, which quickly breaks down after lockdown measures have been initiated. Our findings of a decline in short-distance job commuting thus complement related studies pointing to a drop in long-distance mobility during the German lockdown and a reduction in the German mobility network (Schlosser et al. 2020).

Importantly, however, we would like to stress that our study does not come without limitations and therefore should be interpreted with some caution. As we have pointed out in the main text, for instance, our choice of applying linear dynamic panel data models for log-transformed data relies on several assumptions about the underlying data generating process. Moreover, our analysis should not be seen as a strict causal investigation of policy effectiveness as our primary focus is on revealing space–time correlations in COVID-19 data; however, the significant temporal heterogeneity in estimated parameters strongly points to the fact that mobility constraints have been effective in slowing down the spread of COVID-19 across regions.

Finally, seen from a socioeconomic perspective, restricting mobility and spatial economic interaction may be costly compared to other measures of social distancing such as the wearing of face masks (see, e.g., Mitze et al. 2020).¹⁹ The costs of constraints to workplace-related mobility depend, among other aspects, on the ability of firms to run businesses remotely and also the relative productivity

¹⁹ In addition to imposing higher economic costs, Espinoza et al. (2020) argue on the basis of a historical analysis of earlier epidemics that more extreme mobility restrictions may also bear undesirable epidemiological consequences.

performance of home office compared to on-site work. With regard to the former, first survey results for Germany reported in Hofmann et al. (2020) indicate that the majority of firms in Germany was able to rapidly adapt to digitalized business processes. About 70% of surveyed firms reported that employees were able switch to major or full use of home office work during the lockdown period. Prior to the lockdown, only 15% of these firms allowed employees to use of home office work (to a large or full extent). With regard to the relative productivity performance, the available empirical results are mixed. Though some studies point at low or moderate productivity losses of home office work (Hofmann et al. 2020; Valoir 2020), effects may considerably vary by education levels (Bartik et al. 2020). Given limits to workplace digitalization, Morikawa (2020) predicts that a productivity gap between home office and regular on-site work is likely to remain also in the mid-run. Further, risks for firm innovativeness and worker satisfaction need to be considered (OECD 2020).

The importance of gaining continuous insights into the effectiveness and efficiency of public interventions surely calls for more research on this matter in order to fully assess the right dose of public actions to fight the spread of COVID-19 through non-pharmaceutical interventions. For policy makers, our findings stress that commuting-to-work linkages should closely monitored and considered in policy strategies that target to suppress the spread of COVID-19. Promising avenues for future research on the spatiotemporal dynamics of COVID-19 should consider the use of multiple data sources (including real- or near-time data) on interregional mobility to better surveil the spatiotemporal dynamics of local hot spots and foci of infection (Peixoto et al. 2020) and analyze policy effectiveness of public health measures (Vinceti et al. 2020).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10109-021-00349-3>.

Acknowledgements The authors thank Annkatrin Niebuhr from the Institut für Arbeitsmarkt- und Berufsforschung (IAB) for the kind provision of data on interregional commuting flows. We are also grateful to Klaus Wälde and Johannes Rode for discussing earlier versions of this manuscript and to three anonymous referees for giving us important comments and advice on how to improve this analysis. Finally, we would like to thank the editor-in-chief, Manfred M. Fischer, for his constant support during the review process. The usual disclaimer applies.

Funding No funding to report.

Availability of data and material Data and supplementary research materials (online appendix) are available under the <https://doi.org/10.6084/m9.figshare.12789941>.

Code availability STATA codes and documentations for replication purposes are available under the <https://doi.org/10.6084/m9.figshare.12789941>.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

- Atkeson A (2020) On using SIR models to model disease scenarios for COVID-19. *Quart Rev Fed Reserve Bank Minneapolis* 41(1):1–35
- Bartik A, Cullen Z, Glaeser EL, Luca M, Stanton C (2020) The impact of COVID-19 on small business outcomes and expectations. *Proc Natl Acad Sci* 177(30):17656–17666
- Beenstock M, Felsenstein D (2010) Spatial error correction and cointegration in nonstationary panel data: regional house prices in Israel. *J Geogr Syst* 12(2):189–206
- Bellégo C, Pape LD (2019) Dealing with the log of zero in regression models. Working Papers 2019–13. Center for Research in Economics and Statistics (CREST), for download at: <http://crest.science/RePEc/wpstorage/2019-13.pdf>. Accessed 02 Aug 2020
- Berlemann M, Hausteine E (2020) Right and yet wrong: a spatio-temporal evaluation of Germany's COVID-19 containment policy. CESifo Working Paper Series 8446, CESifo
- Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D (2020) The challenges of modeling and forecasting the spread of COVID-19. *Proc Natl Acad Sci* 117(29):16732–16738
- Blundell R, Griffith R, Windmeijer F (2002) Individual effects and dynamics in count data models. *J Econ* 108(1):113–131
- Chang S, Pierson E, Koh PW, Gerardin J, Redbird B, Grusky D, Leskovec J (2021) Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* 589:82–87
- Charaudeau S, Pakdaman K, Boëlle PY (2014) Commuter mobility and the spread of infectious diseases: application to influenza in France. *PLoS ONE* 9(1):e83002
- Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, Pastore y Piontti A, Mu K, Rossi L, Sun K, Viboud C, Xiong X, Yu H, Halloran ME, Longini IM, Vespignani A (2020) The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* 368(6489):395–400
- Cromley EK, McLafferty SL (2012) GIS and public health, 2nd edn. Guilford Press, New York
- Desjardins MR, Hohl A, Delmelle EM (2020) Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: detecting and evaluating emerging clusters. *Appl Geogr* 118(2020):102202
- Duan N (1983) Smearing estimate: a nonparametric retransformation method. *J Am Stat Assoc* 78(383):605–610
- Elhorst JP (2014) Spatial econometrics. From cross-sectional data to spatial panels. SpringerBriefs in Regional Science. Springer, Berlin, Heidelberg
- Espinoza B, Castillo-Chavez C, Perrings C (2020) Mobility restrictions for the control of epidemics: When do they work? *PLoS ONE* 15(7):e0235731
- Felbermayr G, Hinz J, Chowdhry S (2020) Après-ski: the Spread of Coronavirus from Ischgl through Germany. *Covid Econ* 22:177–204
- Gao S, Rao J, Kang Y, Liang Y, Kruse J (2020) Mapping county-level mobility pattern changes in the United States in response to COVID-19. *SIGSPATIAL Spec* 12(1):16–26
- Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A (2020) Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc Natl Acad Sci* 117(19):10484–10491
- Glaeser E, Gorbach C, Redding SJ (2020) How much does COVID-19 increase with mobility? Evidence from New York and four other U.S. cities. *J Urban Econ* JUE Insights. <https://doi.org/10.1016/j.jue.2020.103292>
- Google (2020). Google COVID-19 community mobility reports. <https://www.google.com/covid19/mobility>. Accessed 07 July 2020
- Hartl T, Weber E (2020) Welche Maßnahmen brachten Corona unter Kontrolle? Available at: <https://www.oekonomenstimme.org/artikel/2020/05/welche-massnahmen-brachten-corona-unter-kontrolle/>
- Held L, Höhle M, Hofmann M (2005) A statistical framework for the analysis of multivariate infectious disease surveillance data. *Stat Model* 5(3):187–199
- Hofmann J, Piele A, Piele C (2020) Arbeiten in der Corona-Pandemie– Auf dem Weg zum New Normal. Studie des Fraunhofer IAO in Kooperation mit der Deutschen Gesellschaft für Personalführung DGFP e.V. <http://publica.fraunhofer.de/dokumente/N-593445.html>. Accessed 05 Nov 2020
- Höhle M (2016) Infectious disease modelling. In: Lawson AB, Banerjee S, Haining RP, Ugarte MD (eds) *Handbook of spatial epidemiology*. Chapman and Hall/CRC Press, New York, pp 478–500

- Hong HG, Li Y (2020) Estimation of time-varying reproduction numbers underlying epidemiological processes: a new statistical tool for the COVID-19 pandemic. *PLoS ONE* 15(7):e0236464
- Judson RA, Owen A (1999) Estimating dynamic panel data models: a guide for macroeconomists. *Econ Lett* 65(1):9–15
- Kergassner A, Burkhardt C, Lippold D, Nistler S, Kergafner M, Steinmann P, Budday D, Budday S (2020) Meso-scale modeling of COVID-19 spatio-temporal outbreak dynamics in Germany. medRxiv. <https://doi.org/10.1101/2020.06.10.20126771>
- Klein B, LaRock R, McCabe S, Torres L, Privitera F, Lake B, Kraemer M, Brownstein J, Lazer D, Eliassi-Rad T, Scarpino S, Chinazzi M, Vespignani A (2020a), Assessing changes in commuting and individual mobility in major metropolitan areas in the United States during the COVID-19 outbreak. Preprint available at last accessed 15 Mar 2021. <https://www.networkscienceinstitute.org/publications/assessing-changes-in-commuting-and-individual-mobility-in-major-metropolitan-areas-in-the-united-states-during-the-covid-19-outbreak>
- Klein B, LaRock T, McCabe S, Torres L, Privitera F, Lake B, Kraemer M, Brownstein J, Lazer D, Eliassi-Rad T, Scarpino S, Chinazzi M, Vespignani A (2020b) Reshaping a nation: mobility, commuting, and contact patterns during the COVID-19 outbreak. Preprint available at last Accessed 15 Mar 2021. https://www.mobs-lab.org/uploads/6/7/8/7/6787877/covid19mobility_report2.pdf
- Korniotis GM (2010) Estimating panel models with internal and external habit formation. *J Bus Econ Stat* 28(1):145–158
- Krisztin T, Piribauer P, Wögerer M (2020) The spatial econometrics of the coronavirus pandemic. *Lett Spat Resour Sci* 13(3):209–218
- Kropp P, Schwengler B (2011) Abgrenzung von Arbeitsmarktregionen – ein Methodenvorschlag. *Raumforsch Raumordn* 69:45–62
- Kuebart A, Stabler M (2020) Infectious diseases as socio-spatial processes: the COVID-19 outbreak in Germany. *Tijdschr Econ Soc Geogr* 111(3):482–496
- Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J (2020) The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 172(9):577–582
- Lee M, Zhao J, Sun Q, Pan Y, Zhou W et al (2020) Human mobility trends during the early stage of the COVID-19 pandemic in the United States. *PLoS ONE* 15(11):e0241468. <https://doi.org/10.1371/journal.pone.0241468>
- Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung SM, Yuan B, Kinoshita R, Nishiura H (2020) Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J Clin Med* 9(2):538
- Liu M, Thomadsen R, Yao S (2020) Forecasting the spread of COVID-19 under different reopening strategies. *Sci Rep*. <https://doi.org/10.1038/s41598-020-77292-8>
- Lu J, Meyer S (2020) Forecasting flu activity in the United States: benchmarking an endemic-epidemic beta model. *Int J Environ Res Public Health* 17(4):1381
- Manning WG, Mullahy J (2001) Estimating log models: to transform or not to transform? *J Health Econ* 20(4):461–494
- McDonald JH (2014) Handbook of biological statistics, 3rd edn., for download at: <http://www.biostathandbook.com/HandbookBioStatThird.pdf>. Accessed 02 Aug 2020
- McGrail DJ, Dai J, McAndrews KM, Kalluri R (2020) Enacting national social distancing policies corresponds with dramatic reduction in COVID19 infection rates. *PLoS ONE* 15(7):e0236619
- Mitze T, Kosfeld R, Rode J, Wälde K (2020) Face masks considerably reduce COVID-19 cases in Germany: a synthetic control method approach. *Proc Natl Acad Sci* 117(51):32293–32301
- Morikawa M (2020) COVID-19, teleworking, and productivity. *VoxEU*, April 14, 2020, for download at: <https://voxeu.org/article/covid-19-teleworking-and-productivity>
- OECD (2020) Productivity gains from teleworking in the post COVID-19 era: How can public policies make it happen? OECD Policy Responses to Coronavirus (COVID-19), for download at: <http://www.oecd.org/coronavirus/policy-responses/productivity-gains-from-teleworking-in-the-post-covid-19-era-a5d52e99/>. Accessed 03 Aug 2020
- Olsson S, Zhang J (2020) The ongoing COVID-19 epidemic curves indicate initial point spread in china with log-normal distribution of new cases per day with a predictable last date of the outbreak. Preprints, 2020: 2020030077, for download at: <https://www.preprints.org/manuscript/202003.0077/v3>. Accessed 02 Aug 2020

- Peixoto PS, Marcondes D, Peixoto C, Oliva SM (2020) Modeling future spread of infections via mobile geolocation data and population dynamics. An application to COVID-19 in Brazil. *PLoS ONE* 15(7):e0235732
- Pepe E, Bajardi P, Gauvin L, Privitera F, Lake B, Cattuto C, Tizzoni M (2020) COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Sci Data* 7:230. <https://doi.org/10.1038/s41597-020-00575-2>
- Robert Koch Institute (2020) Covid-19 Infektionen, General Website (NPGeo Corona Hub): <https://npgeo-corona-npgeo-de.hub.arcgis.com/>
- Schlosser F, Maier BF, Jack O, Hinrichs D, Zacharias A (2020) COVID-19 lockdown induces disease-mitigating structural changes in mobility networks. *Proc Natl Acad Sci* 117(52):32883–32890
- Tizzoni M, Bajardi P, Decuyper A, Kon Kam King G, Schneider CM, Blondel V, Smoreda Z, González MC, Colizza V (2014) On the use of human mobility proxies for modeling epidemics. *PLoS Comput Biol* 10(7):e1003716. <https://doi.org/10.1371/journal.pcbi.1003716>
- Valoir (2020) The real productivity impact of remote work. Valoir Report May 2020, available for download at: www.valoir.com. Accessed 03 Aug 2020
- Vinceti M, Tommaso F, Rothman KJ, Ferrari F, Goffi A, Maffei G, Orsini N (2020) Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking. *E Clin Med* 25:100457. <https://doi.org/10.1016/j.eclinm.2020.100457>
- Wang N, Fu Yu, Zhang H, Shi H (2020) An evaluation of mathematical models for the outbreak of COVID-19. *Precis Clin Med* 3(2):85–93
- Warren GW, Lofstedt R, Wardman JK (2021) COVID-19: the winter lockdown strategy in five European nations. *J Risk Res.* <https://doi.org/10.1080/13669877.2021.1891802>
- Wen TH, Lin MH, Fang CT (2012) Population movement and vector-borne disease transmission: differentiating spatial–temporal diffusion patterns of commuting and non-commuting dengue cases. *Ann Assoc Am Geogr* 102(5):1026–1037
- Wieland T (2020) Flatten the curve! Modeling SARS-CoV-2/COVID-19 growth in Germany on the county level. *REGION* 7(2):43–83
- Windmeijer F (2005) A finite sample correction for the variance of linear efficient two-step GMM estimators. *J Econ* 126(1):25–51
- Wu N, Ben X, Green B, Rough K, Venkatramanan S, Marathe M, Eastham P, Sadilek A, O'Banion S (2020) Predicting onset of COVID-19 with mobility-augmented SEIR model. medRxiv. <https://doi.org/10.1101/2020.07.27.20159996>
- Xia Y, Bjørnstad ON, Grenfell BT (2004) Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am Nat* 164(2):267–281
- Ziliak JP (1997) Efficient estimation with panel data when instruments are predetermined: an empirical comparison of moment-condition estimators. *J Bus Econ Stats* 15(4):419–431

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.