Après-ski: The Spread of Coronavirus from Ischgl through Germany*

Gabriel Felbermayr† Julian Hinz‡ Sonali Chowdhry§

This version: March 2021

Abstract

The Austrian ski resort of Ischgl is commonly claimed to be ground zero for the diffusion of the SARS-CoV-2 virus in the first wave of infections experienced by Germany. Drawing on data for 401 German counties, we find that conditional on geographical latitude and testing behavior by health authorities, road distance to Ischgl is indeed an important predictor of infection cases, but — in line with expectations — not of fatality rates. Were all German counties located as far from Ischgl as the most distant county of Vorpommern-Rügen, Germany would have seen about 45% fewer COVID-19 cases. A simple diffusion model predicts that the absolute value of the distance-to-Ischgl elasticity should fall over time when inter- and intra-county mobility are unrestricted. We test this hypothesis and conclude that the German lockdown measures have halted the spread of the virus.

JEL Classification Codes: I18, R11

Keywords: Coronavirus, super-spreaders, Germany, spatial diffusion, negative binomial model, tourism

Acknowledgements: We thank Nils Rochowicz for helpful suggestions on our theoretical model. We are grateful to Jan Schymek, Oliver Falck and Wolfgang Dauth for providing us with German data on the Work-from-Home index and trade exposure to China, respectively. All remaining errors are our own.

†*Affiliation:* Kiel Institute for the World Economy & Kiel Centre for Globalization. *Address:* Kiellinie 66, 24105 Kiel, Germany

‡*Affiliation:* Universität Bielefeld, Kiel Institute for the World Economy & Kiel Centre for Globalization. *Address:* Kiellinie 66, 24105 Kiel, Germany

§*Affiliation:* Kiel Institute for the World Economy & EU Trade and Investment Policy ITN (EUTIP) project under the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement No 721916). *Address:* Kiellinie 66, 24105 Kiel, Germany

1 Introduction

By mid-May 2020, the highly contagious SARS-CoV-2 virus infected about 4.5 million people worldwide and led to almost 300,000 fatalities.[1] The outbreak prompted governments to impose lockdowns affecting nearly 3 billion people worldwide, in an unprecedented attempt to 'flatten the curve' of infections so that healthcare systems are not overwhelmed. In Germany, even though restrictions were phased in from March 9th to 23rd 2020, the number of confirmed cases increased to approximately 175,000 with almost 8,000 deaths by mid-May 2020.[2] However, the spread *within* Germany was far from homogeneous — the two southernmost states, Bayern and Baden-Württemberg, were amongst the most affected, and even within these states there was a lot of variation.

Figure 1 depicts the spatial distribution of confirmed COVID-19 cases per 100,000 inhabitants in each of the 401 'Kreise' (counties)[3] using data provided by the Robert-Koch Institute, the German federal government agency and research institute responsible for disease control and prevention.[4] The left-hand side map indicates that as early as March 13th 2020, 356 out of 401 counties already reported some confirmed cases. By May 9th, infections had increased across the country with counties in southern and eastern Germany experiencing significantly higher case burdens, as shown by the histogram on the right-hand side.

[Figure 1 about here.]

The county with the lowest case incidence rate (CIR) was Mansfeld-Südhartz in North-Eastern Saxony-Anhalt (0.03%). Tirschenreuth in Bavaria, the most affected county, had a CIR 52 times higher (1.53%).[5] Across counties, the standard deviation of the CIR was almost as large as its mean. A similar dispersion was observed for the case fatality rate (CFR), which was reported zero for 26 counties.[6]

Which factors explain this spatial distribution? In this paper we explore whether tourists visiting super-spreader locations, in particular the resort town of Ischgl in neighbouring Austria, brought home the virus from trips in February and March 2020, as hypothesized by German and international media outlets.[7] Another earlier hotspot in Germany, the county of Heinsberg,

[1] See e.g. https://ourworldindata.org/coronavirus-data.

[2] See https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4/page/page_1/.

[3] Strictly speaking, in Germany there are 294 so called 'Landkreise' (rural counties) and 107 'kreisfreie Städte' (cities not belonging to any 'Kreis').

[4] See https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Fallzahlen.html.

[5] Case incidence rate (CIR) is defined as the number of infected individuals divided by population size.

[6] Case fatality rate (CFR) is defined as the number of confirmed deaths divided by the number of confirmed cases.

[7] See e.g. "A Corona Hotspot in the Alps Spread Virus Across Europe", March 31st, 2020, Der Spiegel (https://www.spiegel.de/international/world/ischgl-austria-a-corona-hotspot-in-the-alps-spread-virus-across-europe-a-32b17b76-14df-4f37-bfcf-39d2ceee92ec).

located in the Carneval-celebrating Rhineland region, may have contributed to the diffusion of the virus. The transmission may have occurred from other neighbouring countries as well such as the highly affected town of Mulhouse in the French border region of 'Grand Est' and Bergamo, the northern Italian city.

[Figure 2 about here.]

We evaluate these claims by exploiting the exogenous variation in the road distances of German counties from these important clusters of infections — Ischgl, Heinsberg, Mulhouse and Bergamo. The road distance to Ischgl proxies well the probability of tourists visiting the Austrian ski and après-ski hotspot, as can be seen in Figure 2. The reported share of Ischgl tourist from origin regions in Germany in the previous years' skiing season (winter of 2018–2019), as well as the share of Google search queries in the same time period for the town of Ischgl from German states, are highly correlated with the road distance (-0.76 and -0.83, respectively). Whereas the distribution of Google searches for 'Ischgl' shows a gradual decline of search intensity with growing distance — underlining the town's Germany-wide tourist appeal, the same distributions for search queries for 'Mulhouse', 'Heinsberg' and 'Bergamo' are markedly different (see figure 7 in appendix A): Whereas for the former ones searches are highly concentrated in the near proximity of the respective towns, for the latter they are very evenly spread across all of Germany.

By estimating negative binomial regressions, we compute the elasticity of cases and mortality from COVID-19 with respect to distance from these initial European hotspots. The primary aim of our analysis is to explain the substantial spatial heterogeneity in the first wave of COVID-19 infections across German counties. By observing the spatial heterogeneity over the first wave, we indirectly evaluate the efficacy of early lockdown measures in halting the diffusion of the virus.

To guide our empirical analysis, we present a stylized two-period model where the mobility of persons drives infection transmissions. This simple model yields an insightful and testable proposition: the (absolute value of the) elasticity of COVID-19 cases with respect to distance from a super-spreader location is lower (higher) when individuals are more (less) mobile. We evaluate this proposition by examining the evolution of estimated distance elasticities over time. Finally, we demonstrate the significance of Ischgl as 'Ground Zero' for the outbreak in Germany during the first COVID-19 wave by performing a back-of-the-envelope counterfactual scenario with a hypothetical location for the town.

Crucially, all our regressions control for a host of possible confounding variables — including the relative latitude of a county. In robustness checks, we also examine how road distance to Ischgl affects case loads across counties that belong to the same latitude decile. Hence, our results do not simply capture general effects of a county's distance to the South, e.g. Lombardy in Northern Italy, the European region hit hardest and earliest by the pandemic. We also control for testing

by health authorities to account for the spatial pattern in the likelihood of detecting COVID-19 cases.

Our results paint a clear picture for the first wave of COVID-19: Cases increased strictly proportionally with population, but the share of the population infected was, amongst other factors, a function of the road distance to the major Austrian ski resort Ischgl. Were all German counties as far away from Ischgl as Vorpommern-Rügen, Germany would have 45% fewer COVID-19 cases. In contrast, distance to other hotspots was not important. Catholic culture appears to increase the number of cases — likely through Carnival celebrations in late February 2020.[8] We fail to find evidence for a host of socio-demographic determinants such as trade exposure to China, the share of foreigners, the age structure, or a work-from-home index. In line with expectations, fatality rates did not depend on distance to Ischgl. However, case fatality rates increased strongly with the share of population above 65 years and tended to fall in the number of available hospital beds. Finally, distance to Ischgl did not become irrelevant over time for observed cases, suggesting that early lockdown measures were effective in reducing mobility and avoiding further diffusion of the virus across German counties.

Studying the diffusion of the virus across space is of utmost importance to guide the pandemic response which has so far largely been framed and implemented at national levels. Yet, with substantial heterogeneities in the number of infections — both in absolute and per capita numbers — a more fine-grained approach may be required that can take into consideration the specificity of the diffusion. Our analysis also highlights that international tourism is a powerful channel for the spread of contagious diseases. Timely travel bans can therefore limit transmission paths and control the cross-border spillover of infections. Popular destinations such as Ischgl have a critical role to play in such containment strategies since they can rapidly turn into super-spreader locations.

Declared a global pandemic by the WHO on March 11th 2020, the SARS-CoV-2 virus and its associated disease COVID-19 present an enormous challenge to the world economy. Outside of China where the virus was first detected, several European countries such as Italy, Spain and the UK have been hit particularly hard by the outbreak. Within Europe, Germany was treated as an exception during the first wave due to its low case fatality rates (4.41%) in comparison to Italy (13.93%), Spain (11.84%) and the UK (14.91%).[9]

The absence of proven treatments and vaccines during the first wave necessitated quarantine measures which curtailed human mobility and halted economic activity such as industrial production, retail sales and tourism. Although there is a great degree of uncertainty, the economic costs are expected to be high. For 2020, the International Monetary Fund expects

[8] Carnival is a typical Catholic tradition. The German South and South-West are predominantly Catholic, the North and North-East predominantly Protestant, but there is substantial variation within those regions as well.

[9] Figures as of May 9th, 2020

global GDP to fall by 4.4%, more than in the world economic and financial crisis of 2009.[10] In the Eurozone, the Fund expects an output contraction of 8.3%. The World Trade Organization (WTO) projected the volume of international trade in goods to fall by by 9.2% in 2020, with substantial downside risks as the resurgence of the virus requires new lockdowns.[11]

Since the outbreak of disease, economists have worked on several strands of research. The literature is moving fast; here we present only a few characteristic papers. Macroeconomists have introduced optimizing behavior by economic agents into the basic epidemiological SIER (Susceptible-Infected-Exposed-Recovered) model to examine the economic consequences of pandemics under different policy choices (Eichenbaum, Rebelo, and Trabandt, 2020; Farboodi, Jarosch, and Shimer, 2020; Krueger, Uhlig, and Xie, 2020). Behavioral economists have started to examine the long-run effects of this crisis on preferences (Kozlowski, Veldkamp, and Venkateswaran, 2020). Trade economists are studying the diffusion of health-related shocks through trade networks (Sforza and Steininger, 2020). Economic historians are investigating past pandemics to search for patterns that may inform current policy making (Barro, Ursua, and Weng, 2020), whereas econometricians are working to fill data gaps in order to properly calibrate macroeconomic models (Stock, 2020).

Our paper is most closely linked to the emerging literature on the geographical dispersion of the SARS-CoV-2 virus. Harris (2020) shows how the subway system was critical for the propagation of infections in New York City and identifies several distinct hotspot zip codes from where the virus subsequently spread. Jia et al. (2020) also examine the geographical distribution of COVID-19 cases by using detailed mobile phone geo-location data to compute population outflows from Wuhan to other prefectures in China. Pluemper and Neumayer (2020), in a related research endeavor also using German county-level data, find a positive association between the wealth of a district and a negative association with social deprivation in the initial phase of the pandemic up until mid-April 2020, which turned positive for the former and disappeared for the latter afterwards. Also related is Cuñat and Zymek (2020), who combine the SIR model with a structural gravity framework to simulate the spread of contagion in the UK. Our work contributes to the literature by (i) using exogenous variation in the distance to a super-spreader location to identify the role of tourism in the spatial diffusion of COVID-19 in the first wave and; (ii) providing a very simple test for the effectiveness of lockdown measures.

The remainder of this paper is structured as follows. Section 2 provides the relevant context to this analysis by describing the circumstances of the outbreak in Ischgl, Heinsberg, the French region of Grand Est and Bergamo. Section 3 outlines a simple theoretical model which underpins our empirical analysis. In Section 4, we describe our empirical strategy, the datasets used and the construction of key variables. Section 5 presents the main regression results followed by a counterfactual analysis in Section 6. Finally, Section 7 concludes.

[10] World Economic Outlook, October 2020.

[11] See WTO press release dated October 6th, 2020 accessible via https://bit.ly/3i1yXlr.

2 Context

By mid-May 2020, there were around 16,000 confirmed cases of COVID-19 in Austria. The largest cluster of infections, comprising more than 20% of total cases, was located in the alpine province of Tyrol that is home to approximately 8% of Austria's population. The province's capital city, Innsbruck, was the first to report COVID-19 infections in the country, on February 25th, 2020. In Tyrol, the ski resort town of Ischgl is considered to be one of the epicentres, where the virus spread within après-ski bars, restaurants and shared accommodation.

A highly popular destination for international tourists, Ischgl was first flagged as a risk zone by Iceland on March 5th, 2020 after infection tracing revealed it as an important origin for COVID-19 cases. By March 8th, Norway's testing results also revealed that 491 of its 1198 cases had acquired the infection in Tyrol.[12] Despite these early warnings, skiing in Ischgl continued for nine more days. It was only on March 13th, 2020 that the town was placed under quarantine measures. On the same day, Germany's leading centre for epidemiological research, Robert Koch Institute (RKI), also designated Ischgl as a high risk area — alongside Italy, Iran, Hubei Province in China, North Gyeongsang Province in South Korea, and the Grand Est region in France.

As the caseload of infections increased, Austrian authorities finally announced a lockdown in Tyrol on March 19th, 2020. This substantial delay in response is likely to have exacerbated the spread of the pandemic in Austria and other European countries, given the timing of the ski season and the location of the province which is bordered by Italy, Germany and Switzerland. As of March 20th, one-third of all cases in Denmark and one-sixth of those in Sweden were traced to Ischgl.[13] With data on mobile phone usage, a software company was also able to trace the movement of tourists after their visit to Ischgl. Returning skiers went to several destinations such as Munich, Frankfurt, Berlin, Cologne and Hamburg.[14]

In Germany, the states of Bavaria, Baden-Württemberg and North Rhine-Westphalia (NRW) reported the highest number of confirmed cases of the disease in the first wave. Together, they accounted for about two thirds of Germany's total 175,000 COVID-19 cases as of mid- May 2020. Besides Ischgl, the district of Heinsberg in NRW has emerged as another important cluster that may have intensified the outbreak in Germany. The virus was reported to have spread there through Carnival celebrations, with an attendant testing positive on February 25th, 2020.

The northeastern French region of Grand Est was also heavily affected by the pandemic. Close to

[12] See "How an Austrian ski paradise became a COVID-19 hotspot", March 20th, 2020, Euractiv,(https://www.euractiv.com/section/coronavirus/news/ischgl-oesterreichisches-skiparadies-als-corona-hotspot/)

[13] See "Austrian Ski Region Global Hotspot for Epidemic", March 19th, 2020, Financial Times, (https://www.ft.com/content/e5130f06-6910-11ea-800d-da70cff6e4d3)

[14] See "So bahnte sich das Virus seinen Weg von Ischgl nach Deutschland", March 29th, 2020, Die Welt, (https://www.welt.de/wirtschaft/article206879663/Corona-Pandemie-So-hat-Ischgl-das-Virus-in-die-Welt-getragen.html)

France's border with Germany, the spread of infections in the area were largely traced to a mass church gathering in Mulhouse.[15] Given the region's proximity to the hard-hit German state of Baden-Württemberg and the regular cross-border movement of German and French workers, we incorporate the town of Mulhouse in the Grand Est region into our analysis. We also include the distance to Bergamo in our regressions. The city is located in the densely inhabited region of Lombardy, which became the epicenter of the COVID-19 outbreak in Italy. Therefore, these four locations — Ischgl, Heinsberg, Mulhouse and Bergamo — constitute interesting candidates as 'super-spreader locations' for studying the transmission of infections within Germany.

3 Theoretical Model

In this section, we sketch a stylized two-period model where the virus is transmitted through the mobility of population. This simplified set-up is only used to guide our econometric strategy and does not form the basis of any structural estimations.

Let there be two rounds of infections. In the first round, people can be infected by visiting a super-spreader location such as Ischgl. Let P_i be the (time-invariant) population of county i and I_i^0 the number of infected individuals at the end of period 0. Let $f\left(D_i^0\right)$ denote the likelihood that an individual from county i has visited the super-spreader location in period 0 and has become infected, with f being a function of county i's distance to the super-spreader location. Let $f : [1, \infty) \to [0, 1]$ be a continuous and twice differentiable function with $f' < 0$ and $f'' < 0$.[16]

Hence,

$$I_i^0 = P_i f\left(D_i^0\right)$$

$$\Leftrightarrow \quad \iota_i^0 = I_i^0 / P_i = f\left(D_i^0\right),$$

with $\iota_i^0 \in [0, 1]$ being the initial infection rate in county i.

In the second round, individuals randomly meet within Germany. If an infected person comes into contact with a susceptible person, the latter is also infected. Thus, in the absence of outside mobility between counties, new infections in period 1 would be given by

$$\iota_i^1 - \iota_i^0 = \gamma \iota_i^0 (1 - \iota_i^0)$$

$$\Leftrightarrow \quad I_i^1 = I_i^0 + \gamma P_i \iota_i^0 (1 - \iota_i^0),$$

[15] See e.g. "Special Report: Five days of worship that set a virus time bomb in France", March 30th, 2020, Reuters, (https://www.reuters.com/article/us-health-coronavirus-france-church-spec/special-report-five-days-of-worship-that-set-a-virus-time-bomb-in-france-idUSKBN21H0Q2)

[16] The underlying intuition being that a person is less likely to visit Ischgl when the road distance is greater.

where $\gamma \in [0, 1]$ is the probability that an infection occurs when a susceptible individual meets an infected one.

However, individuals tend to move — within and across counties.[17] Let M_{ij} denote those individuals from county i that meet other individuals from county j, with $\sum_j M_{ij} = P_i$.[18] Assuming symmetry in mobility between counties, i.e. $M_{ij} = M_{ji}$, we have

$$I_i^1 = I_i^0 + \gamma M_{ii} \left(1 - \iota_i^0\right) \iota_i^0 + 2\gamma \sum_{j \neq i} M_{ji} \left(1 - \iota_i^0\right) \iota_j^0. \tag{1}$$

The elasticity of the infection rate with respect to the distance to the super-spreader location is given by $\delta^t \equiv \frac{\partial I_i^t}{\partial D_i^0} \frac{D_i^0}{I_i^t}$. Assuming $\frac{\partial M_{ii}}{\partial D_i^0} = 0$, it can be shown that

Proposition. *If any $M_{ij} > 0 \; \forall \; i, j$, then $\delta^0 < \delta^1$.*

Proof. See appendix B.

As both δ^0 and δ^1 are negative, we expect the elasticity of infections with respect to distance from the super-spreader location to be greater (i.e. closer to zero) with mobility than without mobility. When there is no inter-county geographical mobility after period 0, then $M_{ij} = 0$ for all $j \neq i$, the elasticity is larger in absolute terms than when mobility is allowed; when even intra-county mobility is not permitted, then the elasticity is time invariant: $\delta^1 = \delta^0$ as $I_i^1 = I_i^0$.

The intuition for this result is simple: as mobility between and within counties spreads the virus further over time, the role of distance to Ischgl in explaining the spatial variation of infections decreasesgoes down. We assume mobility between counties i and j, M_{ij}, to be exogenous to i's and j's distance to Ischgl, believing this to be a rather innocuous assumption.[19]

4 Empirical Model and Data

4.1 Model and Hypotheses

As reflected in our stylized model, we are interested in understanding the number of COVID-19 patients (I_i^0 and I_i^1) and fatalities registered in a county. For this reason, the appropriate econometric strategy is to estimate a count data model, such as a Poisson or negative binomial model. In this context, we expect the variation of our dependent variable to exceed that of

[17] For simplicity we assume there is no mobility outside of Germany.

[18] Note that one could assume gravity-type micro-foundations with frictions to interactions between i and j, e.g. à la Anderson (2011).

[19] Essentially, the spatial distribution of counties, mobility costs between them, and their population sizes are assumed to be independent of the counties' distance to Ischgl.

a true Poisson since (i) counts will not be independent in a pandemic; and; (ii) there may be unobserved heterogeneity. Therefore, we employ a negative binomial model in which the variance is assumed to be a function of the mean (NB-2 model; see Cameron and Trivedi, 2013). Since the NB-2 model nests the simple Poisson model, one can test for over-dispersion.[20] One handy feature of the negative binomial model is that its coefficients can be interpreted exactly as in a linear model in which the dependent variable is logarithmic.

We exploit the variation in cases and deaths across the numerous counties as of May 9th, 2020, and estimate elasticities with respect to road distances from Ischgl, Heinsberg, Mulhouse and Bergamo. We run cross-sectional regressions which are specified as follows:

$$\text{cases}_i = exp(\alpha + \sum_{k \in \{0,1,2,3\}} \delta_k \log(D_i^k) + \boldsymbol{\gamma}\mathbf{Z}_i + \varepsilon_i) \tag{2}$$

$$\text{deaths}_i = exp(\alpha + \rho \log(\text{lagged cases}_i) + \sum_{k \in \{0,1,2,3\}} \delta_k \log(D_i^k) + \boldsymbol{\gamma}\mathbf{Z}_i + \varepsilon_i) \tag{3}$$

The main coefficients of interest in the above regressions are the set of δ coefficients which capture the elasticity of COVID-19 cases or deaths with respect to the road distance of any given county i from Ischgl, Heinsberg, Mulhouse or Bergamo. In equation (2), these distance elasticities enable us to test our first hypothesis — namely, that COVID-19 cases decay as distance from an infection cluster increases. Therefore, as per our theoretical model, we expect the δ coefficients to be negative.

Equation (3) takes the number of COVID-19 deaths as the dependent variable and introduces log cases lagged by 18 days as an additional explanatory variable. We control for cases with a lag since mean time between onset of symptoms and death is estimated at 17.8 days (Verity et al., 2020). This allows us to test our second hypothesis — that distance to super-spreader locations should not matter for the number of deaths in a county, controlling for the number of infections in a county. Proximity to any of the hotspots may have affected the incidence rate, but should not determine the medical severity of cases and therefore the fatalities.

Our third and final hypothesis is that the distance of a county from these towns is more crucial for spreading infections in the initial phase of the epidemic — in the absence of restrictions on the movement of people. With time, COVID-19 expands its reach to more locations and the role of these initial clusters may become less relevant. A test of this hypothesis can be conducted by introducing time variation in the number of cases and deaths at the county-level. By repeatedly estimating equation (2) for each day within this period, we obtain a time series of coefficients for the distance variables. These time series can then be examined graphically in order to determine

[20] As frequently observed with negative binomial models, as in our exercise, it does not matter substantially whether dispersion is assumed constant across observations (NB-1 model) or is a function of the expected mean (NB-2 model).

when and for how long distance to initial infection clusters mattered in the propagation of COVID-19.

Clearly, distance to Ischgl correlates with other potential determinants of infections. Hence, while we trivially have no issues with reverse causality, our exercise is potentially subject to substantial omitted variable bias. In our exercise, we have no other way to deal with this problem than to load the vector $\mathbf{Z_i}$ with a rich and well-design array of control variables. The most important is geographical latitude, relative to the southernmost point of Germany. This rules out that the coefficient δ_0 simply captures proximity to Italy. The control also captures climatic variation, as well as other factors, e.g. cultural practices, that tend to have a north-south gradient and may influence infection rates. Moreover, we add further county-specific characteristics such as population and population density, GDP per capita, share of population that is older than 65 years, shares of Protestants and Catholics, share of foreigners, a work-from-home index that captures the prevalence of home office work, exposure to trade with China and the number of hospital beds in a county. All these controls may exhibit non-zero correlation with distance to Ischgl. For example, the share of Catholics is much higher in the South than in the North and Catholic festivities, e.g. Carnival, may propagate infections.

The $cases_i$ variable in equations (2) and (3) refer to diagnosed cases rather than to a full count of the infected population, or a random draw. There could be many more undetected cases in the German population than diagnosed ones. For instance, Li et al. (2020) find that in early stages of pandemics, the number of infected people was six higher than official statistics revealed. To deal with this issue, we control for the number of tests per county. Interestingly, there is substantial variation across counties in the share of population tested.

Despite these efforts to contain omitted variable bias, we adopt a cautious reading of our results and refrain from interpreting them as causal. Nonetheless, our evidence on the spatial determinants of the COVID-19 spread in Germany reveals interesting correlations and strong indications of a link between the COVID-19 burden of a county and its distance from a super-spreader location.

Finally, in Appendix C we analyse the sensitivity of our results to the choice of distance measures by switching from road distance to travel time and great circle distances. We also introduce fixed effects for deciles of counties' latitude and great circle distance to Ischgl to study the East-West spread of the virus, as well as ensuring the analysis does not simply pick up North-South variations. Note that by subtracting the log of population from both sides of equation (2) and the log of lagged number of cases from both sides of equation (3), one can interpret the estimated coefficients as elasticities (or semi-elasticities) of case incidence rates (CIRs) or of case fatality rates (CFRs), respectively. Therefore, as an additional robustness check, we estimate models with CIR and CFR as dependent variables using OLS.

4.2 Data

We use publicly available data on COVID-19 cases in Germany provided by the Robert Koch Institute (RKI). The RKI database reports confirmed cases as well as fatalities from COVID-19, although it should be noted that these numbers may under-represent the actual spread of infections due to limitations in testing. A valuable feature of the RKI dataset for our purposes is its level of geographic disaggregation. Information is available not just at the country-wide or Bundesländer (state) level, but at the county-level in Germany. The data spans from March 10th to May 9th, 2020 and relates to the first wave of COVID-19 infections experienced by Germany.[21] In this paper, we work with cumulative confirmed cases and COVID-19 related deaths as of May 9th, 2020.

[Table 1 about here.]

We merge this database with information on the county-level from the Regionaldatenbank Deutschland. We include data on the local population,[22] which allows us to control for the demographic structure of each county, given the higher risk of hospitalisation and fatalities from COVID-19 amongst older populations. We also control for another population characteristic, namely religious affiliation, that may indicate whether Carnival gatherings — largely a Catholic festival — may have contributed to the spread.

To control for the levels of economic activity, we utilize GDP per capita at the county-level for the latest available year, 2018. We further include a variable that describes the regional intensity of jobs that can be performed from home, the 'work-from-home' index at the county-level computed by Alipour, Falck, and Schüller (2020). Ability to work from home, and thus avoid public spaces and offices, may have played an important role in determining the local spread of the virus (Fadinger and Schymik, 2020). As another possible channel for the transmission of the virus within Germany, we incorporate the exposure of counties to international trade with China, where the outbreak was first reported. The trade (export and import) exposure measures are taken from Dauth et al. (2017).[23]

The number of confirmed cases may also be dependent on the testing capacity and healthcare infrastructure of the county. However, there is no reliable data available as of the time of writing on the number of tests conducted daily in each county. Given this limitation, we use the number of tests performed in each of the 16 German Bundesländer. This information is provided by the

[21] The RKI data have been criticized for inaccurate timing of reported cases. For example, there are differences between weekends and weekdays. Since we do not exploit daily variation in cases, our estimations should be largely free from this problem. Besides these issues, German administrative data are generally perceived as being of high quality.

[22] As of December 31st, 2017.

[23] The measure is constructed from national sector-level import and export data and regional sector-level employment shares.

RKI. For healthcare capacity, which may impact the prevalence of testing and the possibility of adequate treatment, we use the number of hospital beds in each county as an indicator. This is again drawn from Regionaldatenbank Deutschland database for the year 2018.

In order to examine the impact the three hotspots had on the spread of the virus, we exploit each of the county administrative centers' distance to the towns of Ischgl, Heinsberg and Mulhouse. We compute road distance and travel times based on the the shortest path in road networks with data from the OpenStreetMap project. In a robustness exercise we additionally use the great circle distance between the respective locations; see Table 4 in Appendix C.

5 Regression Results

In this section, we analyse regression results based on specifications described in equations (2) and (3) and assess the evolution of estimated coefficients such as distance elasticities over time.

5.1 COVID-19 Cases

Table 2 reports results for confirmed COVID-19 cases, where we introduce a richer set of controls with each successive regression. Starting with Column (1), we find that the coefficient on population is statistically identical to 1, implying that cases rose proportionately with population size.[24] Counties with bigger populations did not have higher case rates (infections per number of inhabitants). This finding is robust across all our specifications.

[Table 2 about here.]

The coefficient for the number of tests is positive and statistically significant — i.e. counties located in states that conducted more tests reported more confirmed cases. The estimated coefficient is large; it suggests that an increase in the number of tests by 1% correlates with an increase in the number of cases by 0.441%. This implies that increasing the number of tests by 10% reveals about 12 more cases of infected persons in the median county.[25] This emphasises the vital importance of testing in understanding the spread of infections and its role in the policy response. In all columns of Table 1, we also report the θ parameter which indicates the extent of over-dispersion in the data. If the θ parameter were to approach infinity, the negative binomial distribution would approach a Poisson distribution. However, the parameter is seen to be finite

[24] For this reason, we can interpret the coefficients in our regressions as also measuring the effect on the log CIR of counties.

[25] $(1.1^{0.441} - 1) \times 279 = 11.98$; adding further covariates reduces the importance of tests by about half.

across specifications. Hence, our choice of negative binomial regressions over Poisson estimation is indeed valid. Not surprisingly, infection data exhibits over-dispersion.

In column (2), we introduce road distance to Ischgl as an additional explanatory variable. In doing so, we find that the pseudo-R^2 increases by 6 percentage points or 9%, indicating the relevance of this variable for the overall fit of the model. The resulting coefficient implies that a county whose road distance to Ischgl is by 1% lower than that of another county has a count of infections that is higher by 0.68%. However, Ischgl may not be the only cluster from where the virus may have spread through Germany. To examine this possibility, column (3) introduces the road distances to other clusters — Heinsberg, Mulhouse and Bergamo — as controls. By additionally controlling for the latitude of each county, we exploit precisely the variation in road distance and not the geographical location of a county on the North-South axis. As such, latitude has no measurable effect on the case load.[26] The coefficient on the distance to Ischgl remains statistically significant. Proximity to Ischgl also appears to be more important than proximity to the other hotspots. For the purpose of illustration, compare the city of Munich, that is about 190 km away from Ischgl, to Hamburg, 935 km away. Everything else equal, Hamburg should have 73.7% fewer COVID-19 cases than Munich.[27] The high elasticity implies a fast decay of infections as one moves away from Ischgl.

In column (4), we control for a wide range of county-level variables that could also predict infections. Notably, the distance elasticity for Heinsberg is no longer statistically significant whereas the distance elasticity to Ischgl is stable. Examining the demographic characteristics, factors such as population density, share of the elderly (65 years and older) and foreign residents in total population are not significant determinants of the spread. In contrast, a 1% point increase in the share of Catholics is associated with a 0.712% increase in cases — probably attesting to the role of carnival celebrations in February 2020, which are typical for Catholic regions but not for Protestant ones, in propagating the virus. To illustrate the importance of this correlation: increasing the share of Catholics in the county with the smallest share (0.02, county of Weimar in Thuringia) to the share observed in the most Catholic county (share of 0.88, county of Freyung-Grafenau in Bavaria) almost doubles the case count.[28]

Our baseline specification additionally controls for economic factors and is reported in column (5). In comparison to the minimalist specification reported in column (2), controlling for demographic and economic factors increases rather than decreases the distance elasticity to Ischgl; adding additional socio-economic controls keeps it approximately constant. A 1% reduction in road

[26] If latitude is included in specification (1), its coefficient is observed to be negative (coefficient of -0.45) and highly statistically significant; if distance to Ischgl is added (without the distances to other super-spreader locations, the coefficient on latitude remains negative but turns statistically insignificant while distance to Ischgl appears significant (with a coefficient of -0.40).

[27] $100\% \times [(935/190)^{-0.84} - 1] = -73.7\%$. In fact, Hamburg had about 20% fewer infections in May 2020, but 24% more population in comparison to Munich.

[28] $[\exp(0.712 \times 0.86) - 1] \times 100\% = 84.4\%$. The case incidence rate is 0.10% in Weimar and 0.24% in Freyung-Grafenau.

distance to Ischgl corresponds to a 0.8% increase in the number of confirmed cases.[29]

Looking at the coefficient on a county's trade exposure to China, where the virus first appeared, we observe that the transmission of virus in Germany was not driven by the strength of economic ties to China. Our results therefore undermine possible claims that the participation of local firms in global production chains involving China may have led to the import of the virus and therefore propagated contagion.[30] We also find that the 'Work-from-Home' (WFH) Index is not a significant factor in the diffusion process. This runs counter to the results reported by Fadinger and Schymik (2020) – possibly because we control for WFH at the more disaggregated county (NUTS-3) level as opposed to the NUTS-2 level. Rather, infections are seen to be dependent on population size and the proximity to local hotspots. All together, the models have relatively high values for pseudo-R^2, which offers a rough measure of the variation in infection rates that our models are able to explain.

For the sake of checking robustness, Table 5 in the Appendix reports regressions analogous to those in Table 2, but with the dependent variable being the case incidence rate and the estimation method being OLS. This regression design is more restrictive than our preferred one, but we generally find that our findings are confirmed. In our most comprehensive regression, Hamburg is predicted to have a CIR that is 0.23 percentage points lower than Munich's (in the data, Hamburg's CIR is 0.27% and Munich's 0.46%).

5.2 COVID-19 Fatalities

Having examined confirmed infection cases, in Table 3 we address the observed spatial heterogeneity in COVID-19 deaths across counties. All regressions contain the log of confirmed infections 18 days prior as a major predictor of the death count.[31] As in Table 2, regressions also include the log of the number of tests conducted in a county and the log of population.

In all specifications, the coefficient on log lagged cases is observed to be statistically significant and greater than 1, implying that deaths are increasing more than proportionately to the number of reported cases in a county. An underlying issue of congestion in healthcare facilities may

[29] Table 4 column (2) in the Appendix uses travel time instead of road distance as a measure of distance; the pseudo-R^2 goes down slightly, but our main results remain intact. Importantly, when distance variables are constructed using great circle distances, distance to Ischgl is no longer statistically significant, but log latitude changes sign and becomes large in absolute value. This is not surprising, as latitude almost perfectly predicts geodesic distance to Ischgl (coefficient of correlation $\rho = 0.989$); latitude highly correlates to travel time, too, but the ρ is somewhat lower at 0.972. We view this as supportive of our identification strategy which relies on road distance conditional on latitude.

[30] While the first patient in Germany contracted COVID-19 from a visiting colleague from China in January 2020, this case was well identified. The systematic spread of the virus however, began in Germany much later from March 2020 onwards and is not observed to be linked to exposure to China. See https://www.spiegel.de/wissenschaft/medizin/corona-virus-erster-fall-in-deutschland-bestaetigt-a-19843b8d-8694-451f-baf7-0189d3356f99.

[31] As noted above, Verity et al., 2020 find that the average time between a confirmed infection and a death is approximately 18 days.

explain this relationship. Importantly this relation is not driven by population: across all specifications, we find that more populous counties tend to have lower number of fatalities, holding the case load constant. But note that the two variables are strongly correlated, as the previous section has shown. The number of tests has no measurable influence on death counts.

[Table 3 about here.]

Without adding the controls introduced in column (1), distance to Ischgl has a large, negative effect on the dependent variable; however, this would be a meaningless result as it only reflects the geography of case counts. Once we control for confirmed cases, distances to the super-spreader locations cease to have a negative effect; if at all, there is a positive effect which is, however, only marginally statistically significant. This is reasonable since health outcomes are likely to depend more on the individual case or county's demographic and economic characteristics than on the distance to a ski resort.[32] However, mortality rises sharply with the share of the elderly in county populations (see columns (4) and (5)), conforming with medical findings that case fatality ratios are higher for older age groups (Verity et al. (2020)). For the purpose of illustration, comparing the county with the smallest share of elderly (0.15, county of Vechta, Lower Saxony) to the county with the greatest (0.29, county of Dessau-Roßlau in Saxony-Anhalt), model (5) predicts more than a doubling of the death count.[33]

Variables such as the share of Catholics that had an important effect in Table 2, are no longer significant. This indicates that the capacity of the health system does not depend on a county's predominant religious group. Also, the share of foreigners is not significant. In contrast, healthcare infrastructure, as proxied by the number of hospital beds, turns out as a statistically significant predictor of COVID-19 morbidity. A 10% increase in number of beds in a county lowers deaths by approximately 1.36%.[34] Thus, access to quality medical care is imperative for minimising the loss of human life due to the pandemic. While this finding warrants further investigation, we would like to stress that the number of beds is predetermined in our specification, so we do not face the issue of reverse causality. Moreover, the effect is estimated conditional on a number of variables that explain both fatalities and the number of hospital beds, such as density (population per area, capturing the urban/rural divide) or GDP per capita. Also note that mobility from counties with few beds to others with more beds would attenuate the effect; hence, we are likely to identify a lower bound of the true effect.

For robustness, Table 6 in the Appendix reports regressions analogous to those in Table 3, but with CFR as the dependent variable and OLS as the estimation method. Results are broadly

[32] Note that, in the robustness checks (CFR model) presented in Table 6 distance to Ischgl is never statistically significant.

[33] $\exp[6.705 \times (0.29 - 0.15)] - 1$.

[34] $1.1^{-0.144} - 1$.

robust. For example, increasing the number of beds by 10% in a county, lowers the case fatality rate by 0.092% points;[35] the median CFR being 4.18%.

5.3 Super-spreader Effects Over Time

So far, we have focused on examining a cross-section of the RKI database by running regressions on a snapshot of COVID-19's impact across counties as of May 9th, 2020. Now, we move towards analysing the time dimension as well. Our question here is: Did the role of super-spreader locations like Ischgl diminish over time during the first wave of infections? This is addressed clearly by Figure 3. It depicts the evolution of the 'daily distance elasticities' that are computed by repeatedly estimating our baseline specification for confirmed cases. To the extent that tourists returning from Ischgl explain an initial distribution of infections but subsequent mobility spreads the virus further, one would expect the measured elasticity to decline in absolute value. This corresponds to the proposition of our theoretical model. If the lockdown (phased in from March 9th to March 23rd 2020) has been effective in restricting mobility, our model predicts that distance elasticity will remain highly negative as initial exposure continues to be important.

Strikingly, we observe distinct phases in the behaviour of the Ischgl elasticity that broadly corresponds with the timeline of Germany's lockdown. Over the initial period, this elasticity reduces in absolute value as individuals continue to be mobile. Once mobility is severely restricted with the imposition of a lockdown, it remains stable, significantly different from zero and strongly negative. Thus, distance from Ischgl is a relevant predictor of cases not just over varying specifications as pointed out in Table 2, but also over time.[36]

The same exercise is carried out for other control variables that were observed to be significant in Table 2 to construct Figure 4. It shows that the positive relationship between cases and testing capacity is consistent and statistically significant over time. In the case of population size, results are in alignment with the cross-sectional regression as elasticities remain close to 1. How well does the baseline model explain the variation in cases across counties? As shown in Figure 5, the pseudo-R^2 is high and improves substantially with time up to 0.80 on March 25th 2020, and has only slightly fallen from there. Again, if the infection would have spread geographically after the containment measures, we would expect a sizeable decline in R^2 from our model; however, we do not observe this pattern.

We conclude that restrictions in mobility after March 23rd 2020 helped contain the virus imported from Ischgl in those counties where it first arrived.

[35] $-0.963 \times \ln(1.1)$.

[36] The Mulhouse elasticity is never statistically significant whereas the Heinsberg elasticity becomes statistically insignificant by March 28th 2020. The elasticity of the distance to Bergamo is initially positive but not statistically significant, with large standard errors.

[Figure 3 about here.]

6 Counterfactual Scenario

The previous section described how the prevalence of infections in Germany during the first COVID-19 wave was related to counties' geographic proximity to a super-spreader location such as Ischgl. To further gauge the impact of proximity to Ischgl on case counts, we now perform a simple back-of-the-envelope counterfactual exercise. We predict the number of confirmed cases were Ischgl located 1,134 km away from all counties, the distance at which the Kreis Vorpommern-Rügen, the northeastern-most county, is actually located from Ischgl. This assumes a situation in which no county is located close to the resort town, and hence simulates a situation in which fewer German tourists may have returned from their ski trip infected with the virus.

[Figure 4 about here.]

Using the baseline negative binomial regression, we compare the predicted number of confirmed cases against the number of cases with the new, hypothetical location of Ischgl. The experiment leads to the total number of cases in Germany (as of May 9th 2020) dropping from the predicted level of 172,275 to the counterfactual level of 94,304 i.e. a 45 % reduction. This back-of-the-envelope calculation validates our prior findings and offers a compelling demonstration of the spatial aspects of virus transmission. Figure 6 below presents maps for the predicted and counterfactual scenarios, with a histogram that captures the differences in number of cases by latitude. The south, in reality located relatively closely to Ischgl, would have seen far fewer cases.

7 Conclusion

This paper studies the geographical distribution of COVID-19 cases and fatalities across the 401 German counties. It tests the hypothesis that returning visitors from super-spreader locations like Ischgl, a popular ski resort in Tyrol, Austria, have played a major role in spreading the disease. Indeed, distance to Ischgl turns out to be an important predictor for case incidence rates, but not for case fatality rates. Were all German counties situated as far from Ischgl as the most distant county of Vorpommern-Rügen, Germany would have 45% fewer COVID-19 cases. Distance to Ischgl does not become irrelevant over time, suggesting that lockdown measures have avoided further diffusion of the virus across German counties. In contrast, distances to other hotspots are unimportant.

Catholic culture, likely capturing local Carnival festivities in late February 2020, appears to increase the number of cases while other socio-demographic determinants such as trade exposure to China, the share of foreigners, the age structure, GDP per capita, or a work-from-home index do not add any explanatory power. Case fatality rates increase strongly in the share of population above 65 years and fall in the number of available hospital beds.

We view our results as evidence towards confirming the role of super-spreader locations for the diffusion of a pandemic. Additionally, we find evidence for the efficacy of the lockdown measures put in place in reducing the spread of the virus. Further improvements of the analysis will be possible as more data become available, for example on testing strategies at the county-level.

A Further figures

[Figure 5 about here.]

B Proof of proposition

Consider

$$\delta^0 < \delta 1$$

$$\Leftrightarrow \quad \frac{\partial I_i^0}{\partial D_i}\frac{D_i}{I_i^0} < \frac{\partial I_i^1}{\partial D_i}\frac{D_i}{I_i^1}.$$

Cancelling D_i, and applying the chain rule yields

$$\frac{\partial I_i^0}{\partial D_i}\frac{1}{I_i^0} < \frac{\partial I_i^1}{\partial I_i^0}\frac{\partial I_i^0}{\partial D_i}\frac{1}{I_i^1}.$$

Dividing by $\frac{\partial I_i^0}{\partial D_i}$ flips the sign because it is negative, yielding

$$\frac{I_i^1}{I_i^0} > \frac{\partial I_i^1}{\partial I_i^0}. \tag{4}$$

We now turn to equation (1) that describes the number of new infections in county i in period 1:

$$I_i^1 = I_i^0 + (1-\iota_i^0)M_{ii}\iota_i^0 + 2\left(1-\iota_i^0\right)\sum_{j\neq i} M_{ji}\iota_j^0 \tag{5}$$

Dividing by I_i^0 gives us the left-hand side of condition (4)

$$\frac{I_i^1}{I_i^0} = 1 + (1-\iota_i^0)\frac{M_{ii}}{P_i} + 2\frac{1-\iota_i^0}{P_i}\sum_{j\neq i}\frac{M_{ij}}{P_i}\iota_j^0 \tag{6}$$

Taking the derivative of (5) with respect to I_i^0 yields

$$\frac{\partial I_i^1}{\partial I_i^0} = 1 + (1-\iota_i^0)\frac{M_{ii}}{P_i} - \frac{M_{ii}}{P_i}\iota_i^0 - 2\sum_{j\neq i}\frac{M_{ij}}{P_i}\iota_j^0 \tag{7}$$

Inserting equations (7) and (6) into condition (4) and rearranging yields

$$1 + (1 - \iota_i^0)\frac{M_{ii}}{P_i} + 2\frac{1 - \iota_i^0}{\iota_i^0}\sum_{j \neq i}\frac{M_{ij}}{P_i}\iota_j^0 > 1 + (1 - \iota_i^0)\frac{M_{ii}}{P_i} - \frac{M_{ii}}{P_i}\iota_i^0 - 2\sum_{j \neq i}\frac{M_{ij}}{P_i}\iota_j^0$$

$$2\frac{1 - \iota_i^0}{\iota_i^0}\sum_{j \neq i} M_{ij}\iota_j^0 > -M_{ii}\iota_i^0 - 2\sum_{j \neq i} M_{ij}\iota_j^0$$

$$2\frac{1}{I_i^0}\sum_{j \neq i} M_{ij}\iota_j^0 > -\frac{M_{ii}}{P_i}\iota_i^0$$

Since the left side is always positive, and the right side is always negative, this proves that $\delta^0 < \delta^1$.

■

C Robustness checks

In this appendix, we perform a number of robustness checks to determine whether distance elasticities are sensitive to variable definitions or model choice. In all prior regressions, we used continuous measures of distance. However, we can divide the measure into bins in order to test whether the relationship between case counts and distance from Ischgl is non-linear. We therefore alter our baseline specification by introducing a series of dummies for the various deciles of road distance to Ischgl. The estimated coefficients then capture cases relative to the first decile i.e. relative to counties that are nearest to Ischgl. Figure 8 below plots this sequence of coefficients and reveals a close to linear relationship. To explain with an example, counties belonging to the 10th decile that are farthest away from Ischgl have approximately 0.5% fewer cases in comparison to the reference group of counties closest to Ischgl.

[Figure 6 about here.]

Table 4 below compares our baseline negative binomial specification for confirmed cases in column (1) with regressions that employ alternative measures of distance. We find that Ischgl dominates over Heinsberg and Mulhouse as a super-spreader location even when switching from road distance to travel time. The results for other controls closely follow the pattern observed in Table 2. While the elasticities on population size, testing and share of Catholics are highly significant and comparable across specifications, the coefficients on other demographic and economic factors remain largely insignificant. There in no marked improvement in the model's

Pseudo R^2 either when estimating with alternative definitions of distance. Switching to a great circle distance, which should not matter for the spread of the disease, yields a much smaller and statistically insignificant elasticity. Note that now relative latitude to Ischgl has a negative, albeit statistically insignificant coefficient, as it is highly collinear to the great circle distance. All other coefficients remain largely unchanged.

In column (4) we include fixed effects for great circle distance decile. We therefore compare counties with very similar great circle distance to Ischgl, yet varying road distances. The results are still very similar to those in column (1). The same holds true for column (5). Here we include latitude decile fixed effects, hence comparing counties at similar latitudes. While the distance to Ischgl elasticity is somewhat smaller, the effect remains highly significant.

[Table 4 about here.]

The following robustness checks relate to the choice of the dependent variable and the estimation strategy. In Table 5, we move towards analysing CIR as opposed to the number of cases. With CIR as our outcome variable, we are now no longer in a count-model and can estimate regressions with simple OLS. Consistent with prior findings, we observe that distance to Ischgl is a significant predictor for infections. In a similar vein, we move from count models for fatalities in Table 3 to estimating OLS regressions for CFR in Table 6. This change does not undermine our main results. While testing capacity and share of the elderly influence CFR, distances of counties from super-spreader locales do not.

[Table 5 about here.]

[Table 6 about here.]

References

Alipour, Jean-Victor, Oliver Falck, and Simone Schüller (2020). "Germany's Capacities to Work from Home". In: *CESifo Working Paper* 8227.

Anderson, James E (2011). "The Gravity Model". In: *Annual Review of Economics* 3.1, pp. 133–160.

Barro, Robert, Jose Ursua, and Joanna Weng (2020). "The Coronavirus and the Great Influenza Pandemic. Lessons from the "Spanish Flu" for the Coronavirus's Potential Effects on Mortality and Economic Activity". In: *NBER Working Paper 26866*.

Cameron, Colin and Pravin Trivedi (2013). *Regression Analysis of Count Data*. Cambridge, UK: Cambridge University Press.

Cuñat, Alejandro and Robert Zymek (2020). “The (Structural) Gravity of Epidemics”. In:

Dauth, Wolfgang, Sebastien Findeisen, Jens Südekum, and Nicole Wössner (2017). “The Adjustment of Labor Markets to Robots”. In: *CEPR Discussion Paper* 12306.

Eichenbaum, Martin, Sergio Rebelo, and Mathias Trabandt (2020). “The Macroeconomics of Epidemics”. In: *NBER Working Paper 26882*.

Fadinger, Harald and Jan Schymik (2020). “The Effects of Working from Home on Covid-19 Infections and Production - A Macroeconomic Analysis for Germany”. In: *Covid Economics* 9, pp. 107–134.

Farboodi, Maryam, Gregor Jarosch, and Robert Shimer (2020). “Internal and External Effects of Social Distancing in a Pandemic”. In: *Covid Economics* 9, pp. 22–58.

Harris, Jeffrey E (2020). “The Subways Seeded the Massive Coronavirus Epidemic in New York City”. In: *NBER Working Paper* 27021.

Jia, Jayson S, Xin Lu, Yun Yuan, Ge Xu, Jianmin Jia, and Nicholas A Christakis (2020). “Population Flow Drives Spatio-Temporal Distribution of COVID-19 in China”. In: *Nature*, pp. 1–11.

Kozlowski, Julian, Laura Veldkamp, and Venky Venkateswaran (2020). “Scarring Body and Mind: The Long-Term Belief-Scarring Effects of Covid-19”. In: *Covid Economics* 8, pp. 1–26.

Krueger, Dirk, Harald Uhlig, and Taojun Xie (2020). “Macroeconomic Dynamics and Reallocation in an Epidemic”. In: *Covid Economics* 5, pp. 21–55.

Li, Ruiyun, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman (2020). “Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2)”. In: *Science* 368.6490, pp. 489–493.

Pluemper, Thomas and Eric Neumayer (2020). “The COVID-19 Pandemic Predominantly Hits Poor Neighborhoods, or does it? Evidence from Germany”. In: *medRxiv*. DOI: 10.1101/2020.05.18.20105395. eprint: https://www.medrxiv.org/content/early/2020/05/22/2020.05.18.20105395.full.pdf. URL: https://www.medrxiv.org/content/early/2020/05/22/2020.05.18.20105395.

Sforza, Alessandro and Marina Steininger (2020). “Globalization in the Time of COVID-19”. In: *CESifo Working Paper* 8184.

Stock, James (2020). “Data Gaps and the Policy Response to the Novel Coronavirus”. In: *Covid Economics* 3, pp. 1–11.

Verity, Robert, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick GT Walker, Han Fu, et al. (2020). “Estimates of the Severity of Coronavirus Disease 2019: A Model-based Analysis”. In: *The Lancet Infectious Diseases*.

List of Figures

1 Confirmed cases in Germany on March 13th and May 9th, 2020 23

2 Road distances, tourist origins and Google searches 24

3 Distance coefficents 25

4 Control variable coefficients 25

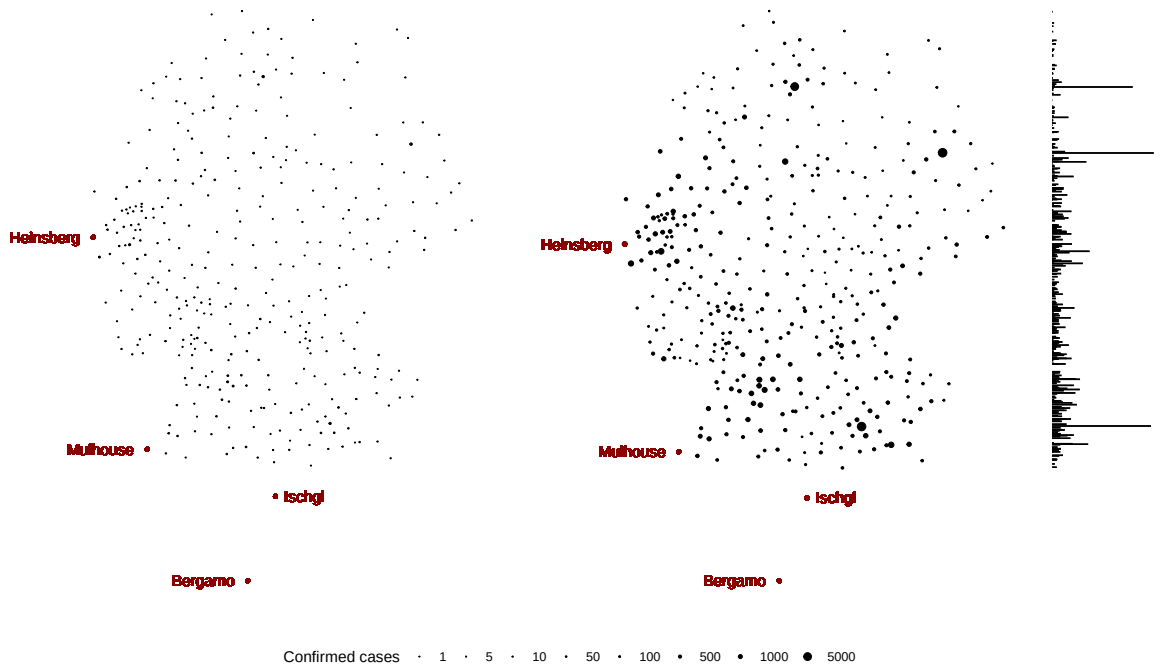
5 Pseudo-R^2 by date of data 25

6 Predicted confirmed cases on May 9th vs. back-of-the-envelope counterfactual . . 26

7 Google searches for Mulhouse, Heinsberg and Bergamo 27

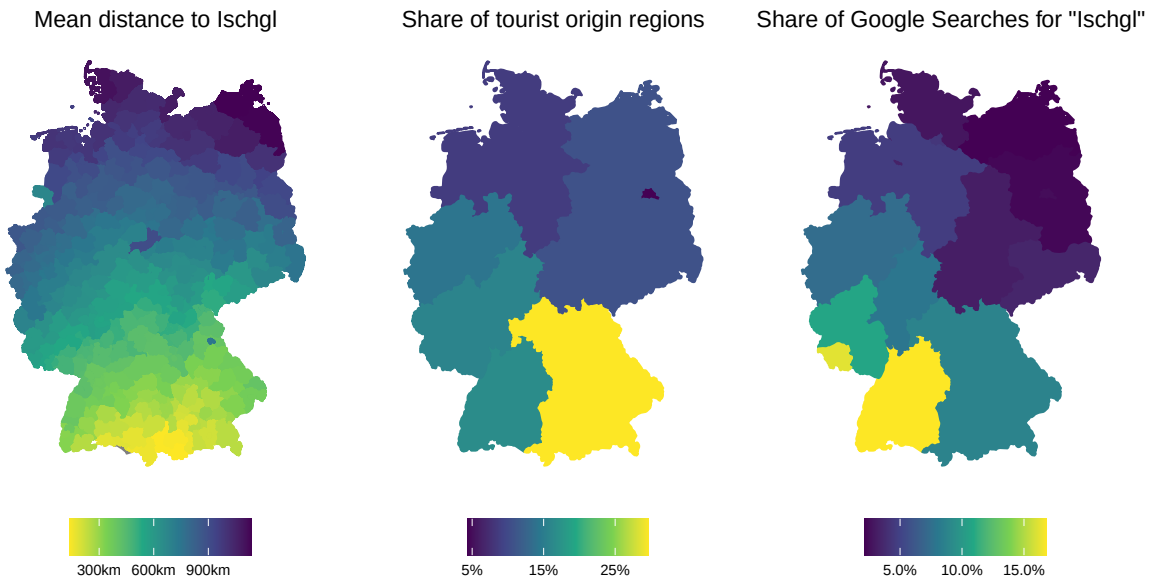
8 Robustness check: Distance coefficient by decile 28

Figure 1: Confirmed cases in Germany on March 13th and May 9th, 2020



Note: Map on the left-hand side shows confirmed cases on March 13th 2020, map on the right-hand side shows those on May 9th 2020. The histogram to the very right shows the change by latitude, binned by county.

Figure 2: Road distances, tourist origins and Google searches



Note: The map on the left-hand side shows each counties' mean distance to Ischgl (own computation based on OpenStreetMap data). The map in the center depicts the share of each region in the total number of German tourists who visited Ischgl over the winter season of 2018-2019 (own computation based on data shared by Statistik Austria). Note that there are only seven regions provided in the raw data: Bavaria, Baden-Württemberg, Nordrhein-Westphalia, Berlin, Middle Germany (Saarland, Rhineland-Palatinate, Hesse), Eastern Germany (Brandenburg, Saxony-Anhalt, Thuringia, Saxony, Mecklenburg-Vorpommern) and Northern Germany (Lower Saxony, Bremen, Hamburg, Schleswig-Holstein). The map on the right-hand side shows the share of Google searches for the town of Ischgl from all 16 German states (own computation based on data from Google Trends).

Figure 3: Distance coefficents

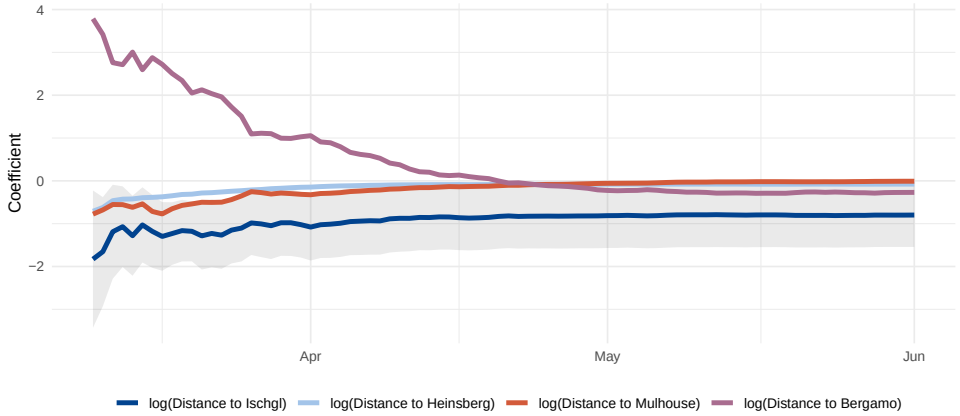


Figure 4: Control variable coefficients

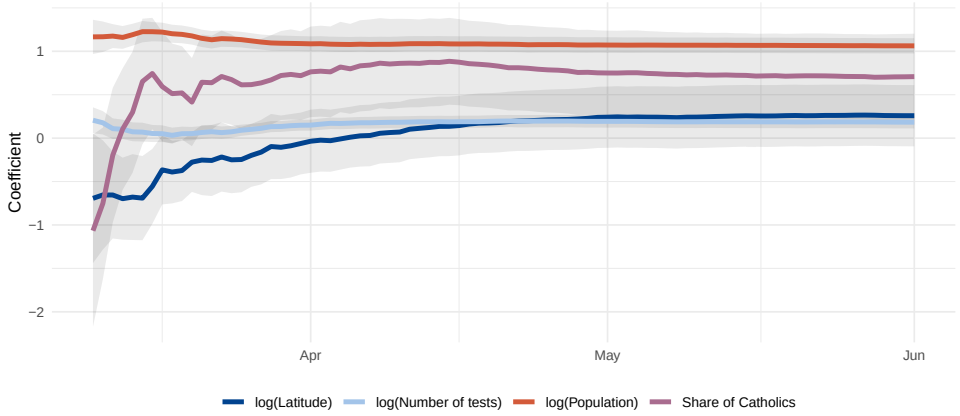


Figure 5: Pseudo-R^2 by date of data

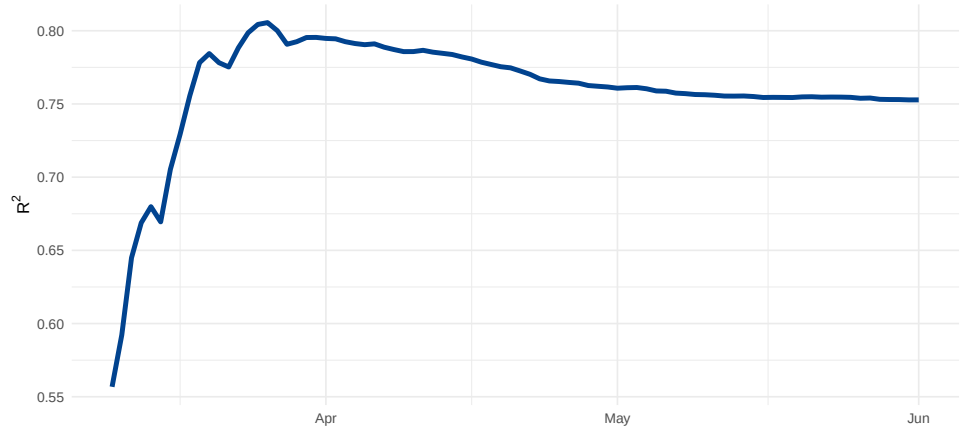
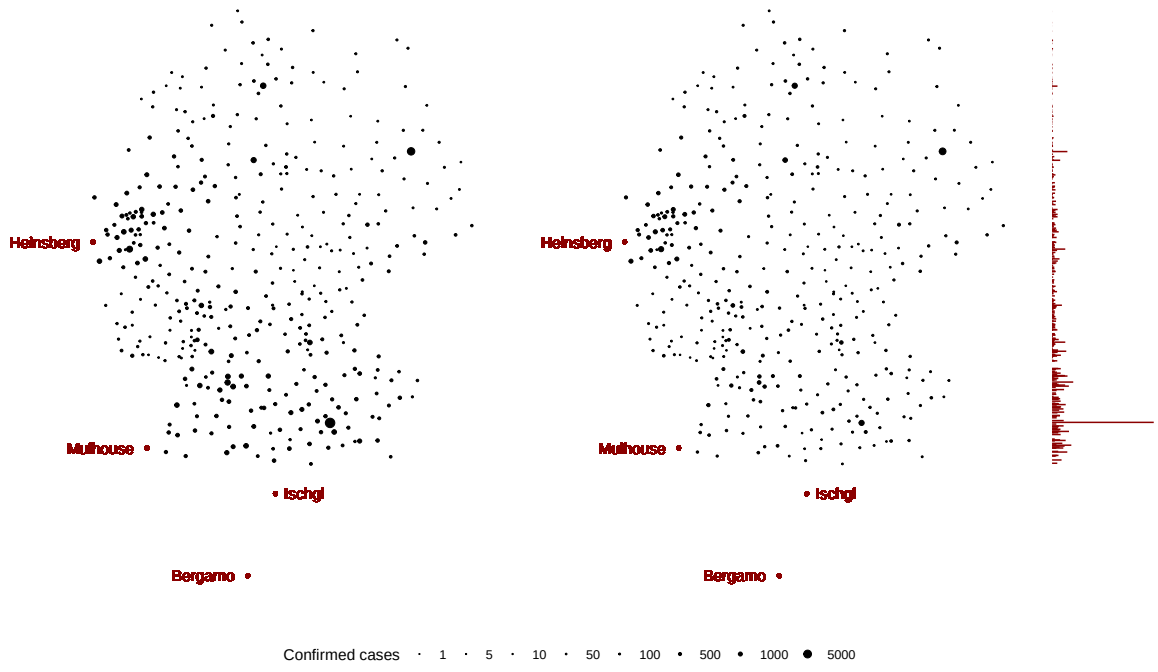
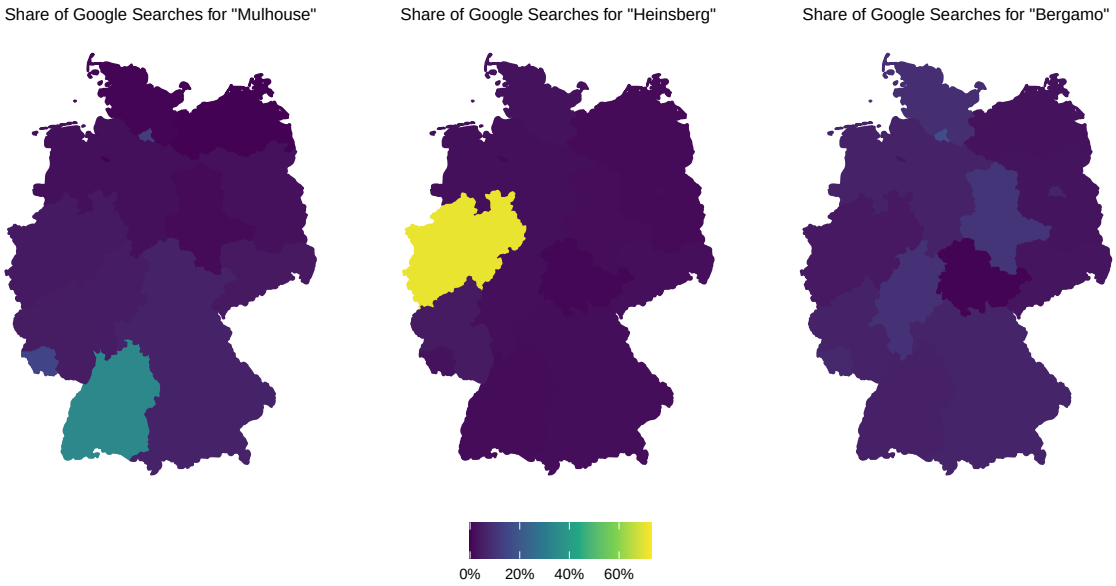


Figure 6: Predicted confirmed cases on May 9th vs. back-of-the-envelope counterfactual



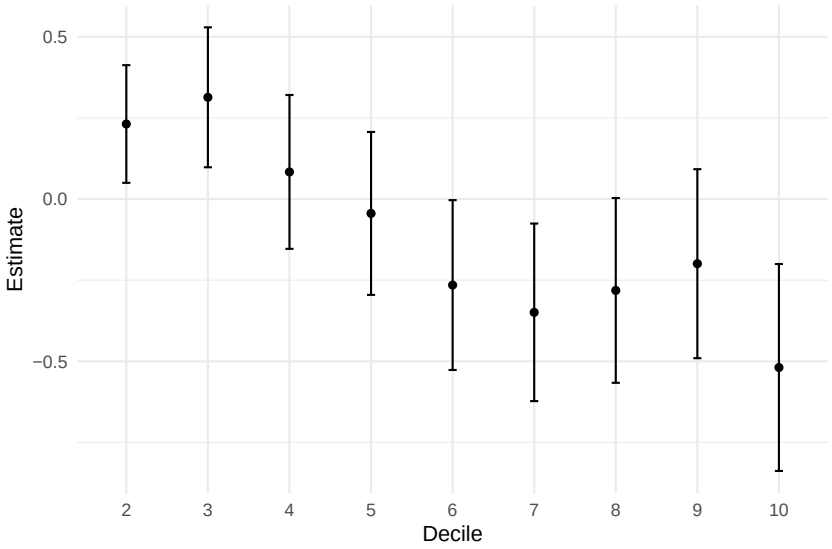
Note: Map on the left-hand side shows predicted confirmed cases on May 9th, map on the right-hand side shows those predicted in counterfactual scenario for May 9th. The histogram to the very right shows the difference by latitude, binned by county.

Figure 7: Google searches for Mulhouse, Heinsberg and Bergamo



Note: The maps show the share of Google searches about the towns of Mulhouse, Heinsberg and Bergamo from all 16 German states (own computation based on data from Google Trends).

Figure 8: Robustness check: Distance coefficient by decile



List of Tables

1 Summary statistics 30

2 Count of Confirmed Cases: Negative Binomial Regressions 31

3 Count of Deaths: Negative Binomial Regressions 32

4 Robustness check: Distance measure 33

5 Case Incidence Rate, OLS Regressions 34

6 Case Fatality Rate, OLS Regressions 35

Table 1: Summary statistics

Variable	Mean	Std. dev.	Median	Max	Min
Number of confirmed cases, current	423.03	591.92	278.00	6261.00	13.00
Number of confirmed cases, 18 day lag	364.91	515.48	240.00	5295.00	13.00
Case incidence rate (CIR), in %	0.21	0.16	0.17	1.52	0.03
Deaths	21.66	30.10	12.00	261.00	0.00
Case fatality rate (CFR), in % &	5.63	3.92	5.06	24.05	0.00
Population (in thousands)	201.23	231.06	149.07	3421.83	34.08
Number of tests (in thousands)	257.90	176.51	208.50	500.48	24.86
Road distance to Ischgl (in km)	609.75	237.28	610.95	1134.26	138.69
Road distance to Heinsberg (in km)	428.40	184.31	433.05	805.37	0.27
Road distance to Mulhouse (in km)	521.01	211.44	507.22	1069.63	56.81
Road distance to Bergamo (in km)	782.55	228.83	768.56	1319.22	326.32
Population / Area	517.74	676.47	195.78	4531.17	36.47
Log of relative latitude	3.35	1.75	3.32	7.51	0.22
GDP per capita (in thousand Euros)	37.16	16.14	33.11	172.44	16.40
Share of foreigners	0.07	0.05	0.06	0.31	0.01
Share of 65+	0.21	0.02	0.21	0.29	0.15
Share of Catholics	0.32	0.24	0.29	0.88	0.02
Share of Protestants	0.30	0.17	0.26	0.72	0.04
Work-from-Home Index	0.53	0.04	0.52	0.67	0.46
Trade with China measure	6338.35	4079.31	5321.70	30228.97	470.53
Number of hospital beds	1255.41	1598.54	851.50	20390.00	42.00

Note: Epidemological data refer to May 9, 2020; other data to year of 2019 or latest available year. Case fatality rate calculated on the basis of reported cases 18 days earlier.

Table 2: Count of Confirmed Cases: Negative Binomial Regressions

	Dependent variable:				
	Number of confirmed cases				
	(1)	(2)	(3)	(4)	(5)
log(Population)	0.991***	1.073***	1.053***	1.070***	1.070***
	(0.048)	(0.053)	(0.055)	(0.053)	(0.054)
log(Number of tests)	0.441***	0.286***	0.258***	0.190***	0.188***
	(0.039)	(0.044)	(0.040)	(0.043)	(0.043)
log(Distance to Ischgl)		−0.679***	−0.840*	−0.787*	−0.795*
		(0.075)	(0.460)	(0.459)	(0.458)
log(Distance to Heinsberg)			−0.138***	−0.064	−0.077
			(0.050)	(0.098)	(0.096)
log(Distance to Mulhouse)			−0.053	−0.019	−0.032
			(0.201)	(0.212)	(0.211)
log(Distance to Bergamo)			−0.256	−0.294	−0.250
			(0.757)	(0.713)	(0.722)
log(Latitude)			0.174	0.243	0.236
			(0.169)	(0.204)	(0.216)
log(Population / Area)				0.038	0.001
				(0.045)	(0.047)
Share of Catholics				0.712**	0.734**
				(0.297)	(0.295)
Share of Protestants				0.166	0.187
				(0.259)	(0.252)
Share of 65+				−1.183	−0.739
				(2.343)	(2.222)
Share of Foreigners				−0.547	−0.810
				(1.116)	(1.142)
log(GDP p.c.)					0.069
					(0.122)
Work-from-Home Index					1.078
					(1.216)
log(China Trade)					−0.006
					(0.070)
Pseudo R2	0.66	0.73	0.74	0.76	0.76
Observations	401	401	401	401	401
θ	3.202***	3.925***	(0.287)	(0.306)	(0.308)
	(0.219)	(0.273)	(0.287)	(0.306)	(0.308)

Note: Constant not reported. Robust standard errors: *p<0.1; **p<0.05; ***p<0.01.

Table 3: Count of Deaths: Negative Binomial Regressions

	Dependent variable:				
	Number of deaths				
	(1)	(2)	(3)	(4)	(5)
log(Lagged number of confirmed cases)	1.301***	1.394***	1.395***	1.443***	1.441***
	(0.062)	(0.064)	(0.064)	(0.065)	(0.065)
log(Population)	−0.372***	−0.487***	−0.487***	−0.490***	−0.387***
	(0.074)	(0.078)	(0.079)	(0.088)	(0.097)
log(Number of tests)	−0.052	−0.003	−0.003	0.038	0.016
	(0.043)	(0.047)	(0.047)	(0.050)	(0.050)
log(Distance to Ischgl)		0.549	0.548	0.448	0.523
		(0.350)	(0.353)	(0.351)	(0.354)
log(Distance to Heinsberg)		0.091**	0.091**	0.067	0.064
		(0.037)	(0.037)	(0.042)	(0.043)
log(Distance to Mulhouse)		0.005	0.004	−0.043	−0.049
		(0.175)	(0.176)	(0.173)	(0.176)
log(Distance to Bergamo)		−0.149	−0.151	−0.187	−0.018
		(0.660)	(0.661)	(0.682)	(0.683)
log(Latitude)		−0.086	−0.085	0.040	−0.073
		(0.187)	(0.190)	(0.194)	(0.195)
log(GDP p.c.)			−0.008	0.071	0.142
			(0.117)	(0.142)	(0.139)
log(Population / Area)				−0.118**	−0.076
				(0.050)	(0.051)
Share catholics				−0.098	−0.068
				(0.273)	(0.273)
Share protestants				0.009	0.021
				(0.315)	(0.316)
Share population 65+				6.705***	7.248***
				(1.637)	(1.642)
Share foreigners				2.711	2.560
				(1.666)	(1.695)
log(Number of hospital beds)					−0.144***
					(0.049)
Pseudo R2	0.76	0.77	0.77	0.78	0.78
Observations	401	401	401	401	396
θ	4.098***	4.341***	4.341***	4.717***	4.829***
	(0.419)	(0.454)	(0.454)	(0.503)	(0.520)

Note: Constant not reported. Robust standard errors: *p<0.1; **p<0.05; ***p<0.01.

Table 4: Robustness check: Distance measure

	Dependent variable:				
	Number of confirmed cases				
	(1)	(2)	(3)	(4)	(5)
log(Population)	1.068***	1.070***	1.044***	1.090***	1.120***
	(0.053)	(0.055)	(0.055)	(0.060)	(0.055)
log(Number of tests)	0.186***	0.192***	0.207***	0.206***	0.228***
	(0.044)	(0.045)	(0.044)	(0.052)	(0.050)
log(Distance to Ischgl)	−0.879***	−0.817***	−0.153	−0.800**	−0.586***
	(0.293)	(0.249)	(0.204)	(0.402)	(0.189)
log(Distance to Heinsberg)	−0.080	−0.041	−0.036	−0.124	−0.123
	(0.094)	(0.157)	(0.110)	(0.090)	(0.093)
log(Distance to Mulhouse)	−0.086	−0.091	0.003	0.071	0.017
	(0.111)	(0.127)	(0.108)	(0.124)	(0.093)
log(Latitude)	0.209	0.082	−0.237	0.073	
	(0.232)	(0.179)	(0.192)	(0.218)	
log(Population / Area)	0.001	0.006	0.003	0.028	0.022
	(0.047)	(0.049)	(0.049)	(0.050)	(0.048)
Share of Catholics	0.733**	0.812**	0.754**	0.643**	0.515*
	(0.295)	(0.336)	(0.324)	(0.278)	(0.294)
Share of Protestants	0.187	0.168	0.212	0.164	0.064
	(0.252)	(0.262)	(0.262)	(0.284)	(0.276)
Share of 65+	−0.769	−0.577	−1.172	−0.188	0.007
	(2.220)	(2.247)	(2.308)	(1.979)	(1.986)
Share of Foreigners	−0.845	−0.714	−0.842	−1.031	−1.190
	(1.152)	(1.193)	(1.196)	(1.141)	(1.224)
log(GDP p.c.)	0.066	0.047	0.064	−0.011	0.100
	(0.123)	(0.122)	(0.121)	(0.127)	(0.121)
Work-from-Home Index	1.137	0.987	1.326	1.148	0.199
	(1.204)	(1.204)	(1.245)	(1.197)	(1.263)
log(China Trade)	−0.002	0.006	0.041	0.010	−0.014
	(0.069)	(0.068)	(0.068)	(0.069)	(0.069)
Distance measure	Road	Travel time	Great circle	Road	Road
Fixed effects	-	-	-	Great circle decile	Latitude decile
Pseudo R2	0.76	0.75	0.75	0.77	0.77
Observations	401	401	401	401	401
θ	4.404***	4.347***	4.261***	4.701***	4.556***
	(0.308)	(0.303)	(0.297)	(0.330)	(0.319)

Note: Constant and great circle decile fixed effects in column (4) not reported. Robust standard errors: *p<0.1; **p<0.05; ***p<0.01.

Table 5: Case Incidence Rate, OLS Regressions

	Dependent variable:				
	Number of confirmed cases / Population x 100.000				
	(1)	(2)	(3)	(4)	(5)
Number of tests	0.090***	0.057***	0.050***	0.043***	0.043***
	(0.008)	(0.008)	(0.009)	(0.011)	(0.011)
log(Distance to Ischgl)		−0.134***	−0.224***	−0.213**	−0.213**
		(0.016)	(0.084)	(0.084)	(0.084)
log(Distance to Heinsberg)			−0.026**	−0.014	−0.018
			(0.012)	(0.014)	(0.014)
log(Distance to Mulhouse)			−0.025	−0.029	−0.032
			(0.042)	(0.045)	(0.045)
log(Distance to Bergamo)			0.167	0.166	0.182
			(0.158)	(0.157)	(0.160)
log(Latitude)			−0.013	0.010	0.006
			(0.043)	(0.045)	(0.046)
log(Population / Area)				−0.010	−0.018
				(0.010)	(0.012)
Share of Catholics				0.102*	0.110**
				(0.052)	(0.054)
Share of Protestants				−0.024	−0.019
				(0.057)	(0.058)
Share of 65+				−0.157	−0.060
				(0.369)	(0.382)
Share of Foreigners				0.142	0.073
				(0.278)	(0.288)
log(GDP p.c.)					0.011
					(0.031)
Work-from-Home Index					0.300
					(0.296)
log(China Trade)					0.001
					(0.014)
Observations	401	401	401	401	401
R^2	0.252	0.359	0.367	0.386	0.389

Note: Constant not reported. Robust standard errors: *p<0.1; **p<0.05; ***p<0.01.

Table 6: Case Fatality Rate, OLS Regressions

	Dependent variable:				
	Number of deaths / Confirmed cases 18 days ago				
	(1)	(2)	(3)	(4)	(5)
Log(Lagged Number of confirmed cases)	1.335***	1.864***	1.861***	2.239***	2.227***
	(0.411)	(0.463)	(0.464)	(0.477)	(0.475)
log(Population)	−1.902***	−2.539***	−2.540***	−2.689***	−2.007***
	(0.489)	(0.537)	(0.538)	(0.609)	(0.655)
log(Number of tests)	−0.135	0.060	0.054	0.305	0.161
	(0.280)	(0.305)	(0.308)	(0.356)	(0.358)
log(Distance to Ischgl)		2.107	2.131	1.504	1.970
		(2.274)	(2.290)	(2.263)	(2.273)
log(Distance to Heinsberg)		0.492*	0.491*	0.352	0.281
		(0.252)	(0.253)	(0.298)	(0.306)
log(Distance to Mulhouse)		0.172	0.181	−0.226	−0.251
		(1.180)	(1.188)	(1.186)	(1.193)
log(Distance to Bergamo)		−1.214	−1.208	−1.074	−0.235
		(4.318)	(4.325)	(4.202)	(4.222)
log(Latitude)		0.155	0.137	0.827	0.162
		(1.209)	(1.226)	(1.272)	(1.300)
log(GDP p.c.)			0.077	0.518	1.114
			(0.768)	(0.971)	(1.039)
log(Population / Area)				−0.589*	−0.344
				(0.332)	(0.329)
Share catholics				−0.947	−1.038
				(1.801)	(1.828)
Share protestants				−0.321	−0.424
				(1.991)	(2.020)
Share population 65+				43.298***	46.561***
				(10.891)	(11.013)
Share foreigners				16.140	14.921
				(11.118)	(11.256)
log(Number of hospital beds)					−0.963***
					(0.343)
Observations	401	401	401	401	396
R^2	0.060	0.077	0.077	0.127	0.136

Note: Constant not reported. Robust standard errors: *p<0.1; **p<0.05; ***p<0.01.