I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14669

COVID-19 Spread in Germany from a Regional Perspective

Olaf Hübler

AUGUST 2021

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14669

COVID-19 Spread in Germany from a Regional Perspective

Olaf Hübler
Leibniz University Hannover and IZA

AUGUST 2021

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

COVID-19 Spread in Germany from a Regional Perspective

This paper investigates the regional differences in the spread of COVID-19 infections in Germany. A machine learning selection procedure is used to reduce variables from a pool of potential influencing variables. The empirical analysis shows that both regional structural variables and regionally aggregated personality traits are significant for the different corona spread. The latter characteristics express differences in mentality between the federal states. The north-east of the country shows a lower degree of affectedness. Regions with a high proportion of migrants show a higher incidence than others. If personality traits are neglected, the migrants' influence is overestimated. With school education and the risk of poverty, two further important regional characteristics are identified. Federal states that have a disproportionately high share of the population with low school education tend to have fewer COVID-19 cases. With regard to poverty, no clear statement can be made. The more the population tends to be responsible towards fellow human beings, the higher is the risk of a more pronounced spread. Where there is a tendency towards altruism, which consists of helping other people, a higher level of COVID-19 infections is revealed. A significant positive correlation between infections and testing is shown by the estimates. The link between vaccinations and the number of infections is less clear. Across the three corona waves,significant changes emerge. This relates in particular to the proportion of migrants and the proportion of families at risk of poverty. The effects decrease over the course of the pandemic.

JEL Classification: C21, C23, I12, R12

Keywords: COVID-19, states, regional characteristics, personality traits, vaccinations, testing, machine learning, cluster-robust estimation, unobserved characteristics, heterogeneity, corona waves, structural break

Corresponding author:
Olaf Hübler
Institut für Empirische Wirtschaftsforschung
Leibniz Universität Hannover
Königsworther Platz 1
30167 Hannover
Germany

E-mail: huebler@ewifo.uni-hannover.de

1 Introduction

It is obvious that the distribution of Covid-19 varies strongly regionally and internationally. Within Germany, Schleswig-Holstein and Mecklenburg-Vorpommern in particular show a favorable development. Only recently, northern regions show an increasing number of infections. The relative affectedness of other federal states is characterised by fluctuating patterns in the course of the pandemic. Various reasons have been and continue to be blamed in the public and in the media for the varying spread of COVID-19. During the first wave, for example, an imported virus was often cited. Cross-border traffic and holiday trips abroad are still cited as explanations. In the east and south of Germany, particularly severely affected areas have appeared at the borders. Hotspots within Germany triggered in individual companies or in individual municipalities have received much attention in the press.

When the partial lockdown in November 2020 did not lead to the desired success, individual attitudes and behaviors were increasingly used as arguments. Empirically, hardly any clear patterns have been identified so far. This could partly be due to the fact that parts of the population are in principle in favor of measures and also stricter requirements to contain the pandemic, but are themselves less likely to adhere to commandments and prohibitions. In the course of the pandemic the individual willingness to forego civil liberties in the short term decreases, in order to contain health damage to society as a whole. With these personal explanations, it is difficult to explain regional incidence differences. However, it can be argued that politically and historically shaped attitudes, mentality-related characteristics, which are unequally distributed regionally, influence individual compliance with the requirements and thus the spread of infection. It remains unclear whether the regional distribution in Germany is subject to constant change or whether the hypotheses, which are usually only put forward one-dimensionally, are unsatisfactory because they do not fully grasp the complex web of effects. In this context, the Robert Koch Institute (RKI) has spoken of a diffuse spread of the corona virus. Clusters may be involved, but we do not (yet) know what they are.

There is a lack of empirical analyses that systematically show why regional disparities occur, which forces are decisive for this, and to what extent the causes are subject to change since the beginning of the pandemic. The aim of this contribution is to find out empirically, taking into account as broad a field as possible of potential direct and indirect influencing variables, which determinants make a significant explanatory contribution statistically and which are negligible, why the pandemic is unevenly distributed in space, whether or not the sign of the determined effect is specification- and method-dependent, whether different influences can be identified in the course of the pandemic. Due to the insufficient data available, the primary aim cannot be to uncover causal relationships, although this would be desirable. Rather, only regional structural patterns can be identified that are closely statistically related to the number of infections. The descriptive character of the contribution should thus be emphasized.

In contrast to previous empirical analyses on the regionally different spread of the corona virus, the following problems are also investigated here. First, cluster robust instead of classical standard errors are used in the estimates to address the Moulton problem (Moulton 1990) that arises when using repeated uniform or constant regional data. Second, the importance of tests and vacchine is investigated. Thirdly, not all available information is used, as there are strong statistical relationships between them that cannot be isolated. Rather, information is condensed and thus only a few observable variables are selected. Fourthly, it is examined how the remaining determinants relate to each other and what significance in the condensed form unobserved compared to observed influences have. Fifthly, the question is examined whether a temporally largely constant pattern of regional infection distribution can be depicted or whether systematic wave movements and structural breaks can be identified.

2 Reasons and hypotheses for regional differences in the spread of the corona virus

When looking for causes of regional differences, simple explanations are not enough. Various factors always play a role, which must be considered depending on the region. Population density is an important reason, but not sufficient for the spatial differences. In various international studies, theoretical considerations and empirical results can be found on possible determinants for the regional unequal distribution of infections, without a clear theory having emerged. The contributions of Akbarpour et al. 2020, Benitez et al (2020), Brown/Ravallion (2020), Chang et al. (2021), Desmet/Wacziarg (2020), Galasso et al. (2020), Goldstein/Lee (2020), Knittel/Ozaltun (2020), Krekel et al. (2020), McLaren (2020), Papageorge et al. (2021), Qui et al. (2020) and Sa (2020) should be mentioned here. In each case, only the influence of a few determinants is examined or vague assumptions are made. In empirical studies, the limitations are often predetermined depending on the available data and depend on the level of observation (individuals, households, companies, regions, national economies).

In the studies mentioned, the following characteristics are used: individual mobility, risk aversion, prosocial motives, pre-existing conditions, age, gender, population density, health status, migrant status, poverty risk, inequality. Brown and Ravallion (2020) highlight that income poverty and income inequality increase infection rates. They find that strong effects come from origin country of immigration, that poor people are less able to protect themselves against infection, that the elderly population and young families have a greater social distance to other people. Furthermore they emphasize that as population density increases, the risk of infection increases significantly. Toya/Skidmore (2021) conclude that countries with higher income, lower population density, older population, fewer hospital beds, more freedoms and more PCR testing have higher corona infection rates than others. Transport restrictions are thus associated with a weaker spread of the pandemic. No clear evidence is found on the effects of a lockdown. Desmet/Wacziarg (2020) also highlight the importance of population density, show that infections increase with family size, that age structure is significant, that minorities are more affected, that people in care facilities have higher infection rates than others. Cultural differences may also be important for unenqual regional dispersion of COVID-19. Deopa/Fortunato (2021) find that German-speaking cantons in Switzerland decreased their mobility since the early stages of the pandemic significantly less than French-speaking cantons. From this, in the latter a lower risk of corona infection could be expected. According to Galasso et al. (2020), women are more likely to see COVID-19 as a serious health problem than men, advocate for more restrictive measures to combat the pandemic. They also protect themselves more against infection.

The importance of weather for the spread of corona infections is examined by Puhani (2020) and Qui et al. (2020). In this context, Burdett et al. (2021) emphasize that temperature, sun and rain had an impact on outdoor activities during the first lockdown in the UK. Therefore, more corona infections may follow. Knittel/Ozaltun (2020) include climate and environmental variables as well as health indicators in their analysis of death risks from corona infection. They find that people who do not work and therefore do not commute are more affected by serious COVID-19 illness and death. Papageorge et al. (2020) elaborate on regional differences. People in regions with high average incomes, with high summer and low winter temperatures have a higher COVID-19 death risk. The authors cannot find any correlation with obesity, the number of ICU beds or poverty rates. Regionally, however, quite different death rates are reported.

Chang et al. (2021) conclude that disadvantaged groups are unable to greatly reduce their mobility. As in the times before the corona pandemic, they stay in crowded places and are therefore exposed to a high risk of infection. McLaren (2020) also finds that minorities are significantly more affected by COVID-19. However, for Hispanics and people of Asian descent, the associations otherwise found for the US are fragile and disappear when educational attainment, occupation and commuting behavior are included as control variables. For citizens of the US, regardless of origin, the correlation does not seem to be tangential when income, poverty rate, schooling and occupation are included in the empirical investigation. Access to health insurance is also irrelevant in relation to the risk of infection. What is significant, on the other hand, is whether public transport is used.

It remains unclear whether the above-mentioned influences on the spread of COVID-19 are significant not only at the individual or macroeconomic level or in an international comparison, but also at the regional level. The aim of this paper is to investigate this. In order to be able to define the analytical framework more clearly, it seems sensible to first separate between different areas, between structures and indicators from a substantive point of view in order to summarize features that have been discussed in the literature so far. Subsequently, these are considered together. The direction of impact is by no means always clear a priori. The following blocks offer starting points:

(1) Geographical structures, measured by population density, by the distinction between area and city states, by regions separated by cardinal points, by commuters from abroad and by contiguity with other countries;
(2) demographic structures, measured by the proportion of women, the proportion of younger and older people, by the proportion of migrants and by household size;
(3) educational-economic indicators, measured by the share of the population without a school-leaving certificate, the share of the population with a university degree, the childcare rate of young children;
(4) climatic influences, measured by precipitation, hours of sunshine and air temperature;
(5) economic indicators, measured by income, unemployment rate and proportion of households at risk of poverty;
(6) voter behavior, measured by the share of voters for individual parties;
(7) policy decisions, measured by the number of asylum seekers, by the number of deportations, by the clearance rate of crimes, by the number of prisoners and the size of the national debt;
(8) health economic influences, measured by average life expectancy, probability of death and health status;
(9) personality-related behavior, measured by the BIG5 variables, by level of satisfaction, by risk-taking, self-confidence, tendency towards optimism, impulsivity, patience and attachment to a region.

Personality-related behavior was given little consideration in the context of corona. Assignment to the various categories is not always clear-cut. For example, the size of the national debt can be assigned to both the economic indicators block and the policy decisions block. There are overlaps between the variables mentioned. Interactions in terms of effects are to be expected. The first step is to record a larger number of possible influences in order to then use statistical-econometric methods, such as machine learning approaches, to separate important influence variables from less important ones, observable from unobserved influence variables. Based on this, it can then be analyzed whether the specification selected in this way reacts sensitively to changes, whether there are structural breaks, whether and, if so, which differences can be identified between the first, second and third wave. In addition to the effects of specific characteristics, it should be increasingly investigated what influence measures serving to control COVID-19 have on the regional spread of corona infections. Lockdowns, travel restrictions, compulsory masks and distance regulations should be mentioned here, i.e. measures that are primarily valid for the entire national economy. However, in a federal state like Germany, the individual federal states have scope for decision-making. Regionally differentiated information is only available to a limited extent, e.g. the number of testing carried out and the number and extent of vaccinations carried out.

3 Data basis

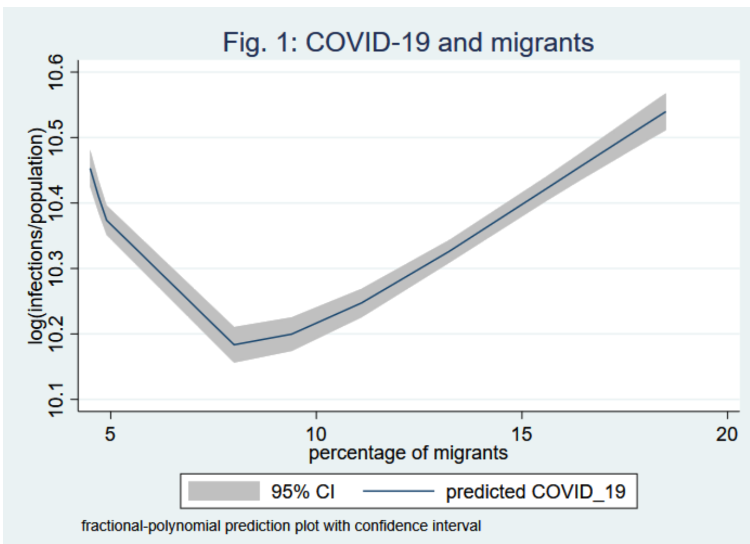
The following empirical analysis primarily draws on official surveys conducted by the Federal Statistical Office. In order to exclude reverse causality as far as possible, the regional data is based on data collected before the outbreak of the Corona virus. The analysis level is the federal states. In states such as the Federal Republic of Germany, the federal states are a suitable level of analysis. A characteristic of the German federal system is the close cooperation between the Federal Government and the states. The former determines the framework conditions in many areas, and in individual fields the latter have autonomy. During the pandemic phase, the frequent meetings between Chancellor Merkel and the Prime Ministers of the states on the subject of COVID-19 showed that uniform agreements were sought, but that the federal states used leeway in implementation.

An empirical analysis at the district level is less suitable not only because not all the necessary information is available but also because the scope for action at this level is smaller and because the socio-economic dividing lines between districts are less clear. However, one problem at the states level also remains. Namely, the heterogeneity within the federal states cannot be depicted with it. Different patterns of infection in sub-regions that deviate from those in the region as a whole remain undetected.

For the following empirical study, the daily recorded COVID-19 infection cases, the number of vaccinations and the weekly reported number of testing published by the Robert Koch Institute are used in particular. The observation period is March 2020 to June 2021. Information on testing is included in the analysis from 9 May 2020. For vaccinations, the recording begins on 1 January. The federal state ranking number for the initial vaccination rate is used here. Regional structural data from the Federal Statistical Office are only recorded once a year. They form the essential framework for the determinants of corona infections. If individual data from the Socio-Economic Panel (SOEP) are used, aggregation takes place at the federal state level. Acronyms, the definition, measurement and descriptive statistics of the variables used can be found in Tables 1 and 2. As an example, Table 3 shows the average values for some characteristics, split by federal states, which could be of importance for further analysis and provide a first insight into regional variability.

4 COVID-19 and regional characteristics: simple regression models

Simple correlation coefficients between the total number of infections and the variables mentioned for blocks (1)-(9) serve as the first level of analysis. Almost all correlations are significant - cf. Table 2, column R. This applies equally to the absolute number of infections (INF), to the variable related to the population size (INF/POP) and to the logarithmic measurement (lnINF, lnINF/POP). For individual variables, however, different signs result for the correlation coefficients depending on whether INF/POP or lnINF/POP is used. In the following, we work with the logarithmic measurement, related to the population (COVID-19). Likelihood ratio tests, however, show neither a clear superiority of the logarithmic approach nor of the linear approach for Box-Cox models. Fractional-polynomial prediction plots suggest a non-linear approach for some individual variables, e.g. for migrants, as Fig.1 shows.



In federal states with a migrant share of up to 8 percent, infections decrease the higher the migrant share in this group. Thereafter, in the majority of the federal states, the infections, in relation to the population, increase as the proportion of migrants increases. Decreasing integration, combined with insufficient knowledge of the German language and consequently insufficient adherence to protective rules against infections, can be an explanatory reason.

In Table 4, estimates are shown separately for the blocks listed in Section 2, with t-values based on classical standard errors on the one hand and on cluster-robust standard errors on the other. It is recommended to work with the latter standard errors, since the underlying regressors are not available in the desired periodicity, but are essentially time-invariant for the period under consideration. Only annual values at the federal state level are available. In this case, the variance is underreported. This can be countered with cluster-robust estimates (Hübler 2014). The comparison of the two t-values should make it clear that the two approaches produce considerable differences. In the further estimates, only the cluster-robust standard errors are shown. It should not go unmentioned that cluster-robust estimates are not always preferable (MacKinnon 2019).

From a geographical perspective (block (1)), it appears that the population in the northern and eastern federal states is less vulnerable than the rest. The west is more affected than the east. Federal states that have relatively many commuters from abroad (COMMUTER) have significantly higher infection rates. The length of a federal state's foreign borders (BORDER) is another indicator pointing in the same direction.

In the case of demographic influences (block (2)), a statistically confirmed correlation between COVID-19 infections and the proportion of migrants ($MIG-RANTS$) can be seen. In other words, migrants are at greater risk of infection than the native population. If the nonlinear term of migrants ($MIGRANTS^2$) is added following Fig.1 we find that the linear and the squared term of migrants are significant ($t_{MIGRANTS}$=5.23; $t_{MIGRANTS^2}$=-4.66).

Younger persons (at least 15 years old, but younger than 25) are less affected by COVID-19 than older persons (65 years or older) in the period under consideration. Only recently, this pattern is changing. With increasing household size, statistically significantly more COVID-19 cases are observed. The higher the proportion of women in a province, the greater the number of infections per inhabitant.

From the point of view of education economics (block (3)), it can be said that if the childcare rate for under-threes is high, the risk of infection is comparatively low, whereas in states with a high proportion of persons with a higher education entrance qualification, COVID-19 is registered more frequently in relation to their age group. Nothing can be said about persons without a school-leaving qualification in this respect, if cluster-robust standard errors are used as a basis.

Climatic factors (block (4)) are certainly significant for the risk of infection at the federal state level. Heavy precipitation is associated with a high risk of infection. No reliable statement can be made for the annual hours of sunshine and the average annual temperature per province. High temperatures and a lot of sunshine mean a lot of outdoor contact, and this could result in an increasing risk of infection. However, it should be noted that there are fewer contacts in a more confined space. Which effect results on balance requires more detailed investigation.

The economic indicators (block (5)) are the gross domestic product (GDP), the unemployment rate (UR), the youth unemployment rate (Y_UR) and the rate of households at risk of poverty (POVERTY). A priori, on the one hand, it would be expected that with higher (average) income there would be better opportunities for protection against corona infections. On the other hand, increasing prosperity results in increasing mobility, both occupationally and privately, from which an increasing risk of infection follows. Conversely, the link between unemployment rates and COVID-19 infections should look the other way round. Those who are unemployed have less money than those who are employed to protect themselves against infection, but they will also be less mobile because there is little reason to do so occupationally and because mobility is associated with costs that people try to avoid, except when looking for new employment. It is possible that young people behave differently from family men. The effect of being at risk of poverty could be similar to that of unemployment. However, Table 4 provides only insignificant effects for this. A weakly significant correlation follows only between COVID-19 and GDP.

In order to uncover a possible link between voting behavior (block (6)) and COVID-19 infections, the voter share of CDU, SPD, GREENS, LEFT, FDP and AfD in the last state election before 2020 is used. According to this, no clear direction of effect can be identified. A traditional party classification according to a right-left pattern leads just as little further as a division into parties in the centre and extremes. It is conceivable that the population tends towards the governing parties in elections because they expect them to make more efforts to combat the pandemic or because they reward their efforts. However, if they are dissatisfied with this, there could conversely be an influx towards the opposition parties. An interdependence of voter share and corona infections would be the result. In order to exclude this to a large extent, the empirical analysis maps voter behavior before the emergence of corona.

Policy decisions (block (7)), which at least directly have nothing to do with corona, are nevertheless remarkably related to the current pandemic on a descriptive level. Estimates with cluster-robust standard deviations show significant associations only for PRISON. The more people there are in prisons in a federal state, the more strongly the region is affected by the pandemic. A direct explanation for this cannot be provided. If not COVID-19, but only the absolute number of infections were used as a basis, the statistical link would be clear. In a federal state with a high population, both the number of infections and the number of prisoners should be greater than in other federal states.

Variables that explicitly or implicitly express health status (block (8)) are linked to the number of corona infections, relative to the population size, as follows: The higher the average life expectancy (LIFE) and the more deaths per year (DEATHS) occur in a federal state, the more infections are observed. In federal states where the average health status of the population is rather poor (SICK), a higher incidence of corona is observed. This is quite consistent with expectations. People who are less well or even poorly off are at a higher risk of infection than others. Somewhat unexpected, on the other hand, is the result that federal states in which many people describe their state of health as good to very good have a high level of contagion. One explanation could be that people in this group are less cautious, less wary. They believe they are robust to infection and even if they do get infected, they expect a mild course. It should be noted that when cluster-robust standard deviations are used, a significant correlation only follows in relation to deaths (DEATHS).

In the preceding partial explanations, possible differences in behavior have occasionally been used as reasons for regional variations in the incidence of corona. The last explanatory block only deals with attitudes and behaviors (block (9)), since such reasons are more frequently put forward in the second and third corona wave than at the beginning of the pandemic, without, however, this phenomenon having been systematically investigated so far. The focus is on personality traits that have been studied extensively in psychology and are summarized under the abbreviation Big FIVE: open to experience (open, full of ideas, curious -OPEN), conscientiousness (thorough, efficient, goal-oriented, organized, responsibly -CONSC), sociability (extroverted, communicative-EXTRA), agreeableness (willingness to cooperate, considerate, compatible-AGREE) and neuroticism (emotional lability, tense, easily nervous, vulnerable-NEURO).

A priori, it can be assumed that openness is associated with multiple contacts with other people and that this results in a high risk of contagion. Emotionally unstable persons are easily upset, often worry, are anxious and tense. They will be overcautious and follow commandments closely to avoid contagion. Empiricism at the aggregate level, however, comes to a different conclusion. According to this, a tendency towards neuroticism is associated with a higher prevalence of the corona virus. If neuroticism is more widespread in one federal state than in others, then politicians will increasingly rely on appeasement, downplaying the dangers of corona in order to allay fears. And this may result in a more carefree approach to the pandemic among the population.

Sociable people live from contact with other people and are therefore more exposed to the risk of infection than others. In the case of emotionally unstable people on the one hand and conscientiousness, level-headed people guided by a striving for achievement and a sense of responsibility on the other, it is less clear a priori whether these personality traits provide an indication of a corona risk. The former often complain of physical pain and could therefore also be susceptible to contracting the corona virus. However, their tendency towards increased irritability and sadness could also lead them to isolate themselves from the outside world and thus provide less of a target for infection. Prudent people, guided by self-discipline, will avoid risks of infection because of this basic attitude. However, because of their competence, and their sense of responsibility, they will be encouraged by others to have outside contacts and seek them out themselves. They might, for example, feel responsible for fellow human beings living in collective accommodation. Empirically, none of the BIG FIVE characteristics are shown to be significant for COVID-19 spread at this level of observation. The cluster-robust t-values are all insignificant. Interactions beween the Big Five variables might be responsible. Further investigations using condensed information are necessary to explore whether single characteristics are mainly relevant.

Beyond the BIG FIVE, other personality traits are also included in the study: satisfaction, optimism, patience, risk-taking, impulsivity and attachment to home. If the population of a federal state tends towards patience (PATIENT) or impulsive behavior (IMPULSE), then a disproportionate distribution of corona is found there. This also seems to be more the case with low life satisfaction (SATIS).

If the population in a state is largely patient, possesses long-suffering and perseverance, one would think that they would adhere to measures taken to combat the corona pandemic over a longer period of time. Or, to put it another way, patience strengthens resilience to corona. Impatient people, on the other hand, do not want to give up leisure activities even in corona times. Although they generally advocate strict measures to combat the pandemic, they hardly adhere to them themselves. Here, too, the empirical evidence points in the opposite direction. If the population is patient, a greater spread of COVID-19 is observed. If policymakers assess their population as predominantly patient, they will not immediately enact very strict measures, combined with sanctions for non-compliance, because they assume that even minor requirements will be largely complied with. Now, a population never consists entirely of patient people, and the impatient will not comply with even weak conditions.

If we not only look separately for the nine blocks with different contents to see which of the selected variables are statistically significant for COVID-19 infections, but also start from more than one block, some deviating patterns emerge. We suppose that there are links between the variables of the different blocks. However, it makes little sense to consider all variables from Table 4 together. A high degree of multicollinearity with insignificant influences and/or less plausible regression coefficients would be the result.

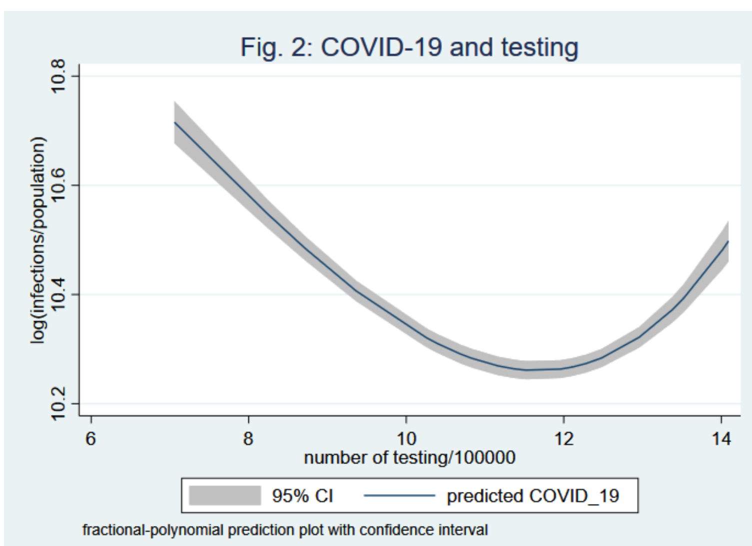
In summary, it can be said up to this point that there are statistical correlations between the regressors within the explanatory blocks, that the signs are specification-dependent and that unobserved influences can distort the actual effect relationship. More complex relationships are presumed. These problems must be investigated in order not to run the risk of misinterpreting results. It should also be emphasized that the results are not causal statements, but merely descriptive results that refer to a given time period. A whole bundle of causes can be responsible for this. Differences in mentality such as love of life, satisfaction, attitudes towards work and family, but also historically developed differences due to religious affiliation and industrial development can be significant.

5 Testing and vaccination

In addition to the characteristics mentioned above, the empirical analysis should also examine the effects of measures that serve to combat COVID-19. On the one hand, the number of testing (TEST) or the normalized number of tests and, on the other hand, the federal state ranking for the first vaccination rate (VACCINE) are available. Tests are used to identify infections. The resulting quarantine is intended to prevent the further spread of COVID-19.

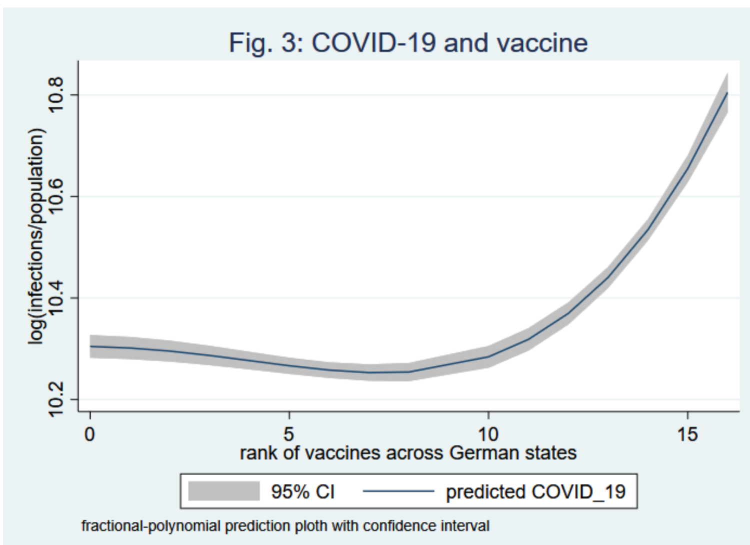
Table 5 shows the regression estimates for COVID-19 with the two measures as regressors. Note that data on vaccination are available from calendar week 1 Janurary, 2021. For reasons of comparison, data on testing are also limited to this period, namely the number of tests in one week for Germany as a whole. Note that COVID-19 and TEST show a common trend. Therefore, a linear trend variable (TREND) is included as a supplement to avoid attributing effects to both measures on the still increasing total number of COVID-19.

Column (1) shows that the more testing is done, the higher the number of infections detected relative to the population. Testing is what detects otherwise undetected corona infections in the first place. Acemoglu et al. (2020) model the effect of testing on the spread of corona. This results in a non-monotonic relationship. Fig.2 makes it clear that the connection between these two variables does not increase linearly throughout.



If a lot of testing is done, the correlation reverses. The trend is not taken into account. But column (2) of Table 5 also shows a non-linear course. One reason could be that test persons may be more careful with multiple tests in order not to have to go into quarantine in case of a positive test result. The probability of a positive test result increases with the number of tests.

There is also a positive relationship between the cumulative COVID-19 infections and the regional rank of the first vaccination rates by federal state at a point in time (VACCINE=1, if highest vaccination rate, ... , =16, if lowest vaccination rate). The higher the rank of a federal state, the lower the relative vaccination rate, the higher the cumulative relative number of infections follows. This is an hint that vaccinations are successful. The same result could probably be demonstrated even more convincingly if a delayed vaccination rate would be included as a regressor in the estimation. However, so far there are no reliable studies on how long it takes for the protection of the vaccinations to become fully effective. In contrast to the tests, no curved course can be determined for the vaccinations (column (4)), although Fig.3 provides an indication for this.



In column (5) of Table 5, both measures enter the estimates simultaneously. Siginificance and signs do not change compared to columns (2) and (3). Which of the two measures is more successful in detecting infections or in containing them cannot be said on the basis of the estimated coefficients, as they are scale-dependent. However, a clue is provided by the $BETA$ coefficients that result when the variables are standardized. The estimates are: $BETA_{TEST}$=0.0166 and $BETA_{VACCINE}$=0.0245. Vaccination is therefore more promising, especially since not much is ultimately gained by detecting infections, unless the affected persons are quarantined and further spread is thus curbed. The further step is to analyze whether and how these relationships change when the influencing variables examined separately in Sections 4 and 5 are considered together.

It should be noted in the empirical analysis of the relationship between COVID-19, tests and vaccinations that mutual dependencies may exist. For example, in the first vaccination deliveries, the distribution was not generally per capita, but partly according to incidence. For example, Saarland and especially Tirschenreuth initially received more vaccine doses than other federal states because of the high incidences and the proximity to Lorraine, a virus variant region. This means that the number of infections can also determine the number doses and in consequence the number of vaccines.

To address this methodologically, IV estimates are conducted as robustness checks. The instrument used for COVID-19 is the cumulative number of corona deaths before the start of the vaccination period. Results for this are provided in column (1), Table 5a. Although the test for exogeneity rejects the null hypothesis, the coefficients of the IV estimate are similar to those of the cluster-robust estimate in column (5), Table 5, except the vaccacine coefficient.

It cannot be ruled out that tests and infections also show reverse causality. The assumption is that if more people become ill, then more people will be tested. As a robustness check, IV estimates are performed for TEST and VACCHINE. In addition to the number of pre-corona deaths, the doubling time (DOUBLING) before the introduction of the vaccination period is used as an instrument. The results in column (2), Table 5a indicate endogeneity and the IV estimation results are in this case close to those in column (1), Table 5a.

6 Variable selection by machine learning

Overall, the block by block modelling in Section 4 remains unsatisfactory. Connections between the blocks are obvious. There are limits to a simple expansion to include all available regressors. A condensation of the information should provide a remedy. Various statistical-econometric methods are available, such as multivariate statistics or the machine learning methods that have recently gained in importance. Here, a LASSO approach is followed. It is used to select relevant regressors from a given set of determinants. All variables that have already been used in the separate block approach are used for preselection. Alternatives would be selection using LARS (Efron et al. 2004) and principal component analysis (PCA) as a statistical technique of reducing variables to describe uncorrelated linear combinations from all available variables. However, preliminary studies have shown that they are less suitable than the LASSO method (Hübler 2021). The latter does not lead to substantively unambiguous interpretations of the extracted components and the former reacts very sensitive to alternative preselections.

The Least Absolute Shrinkage and Selection Operator (LASSO) was developed by Tibshirani (1996). The estimation is based on

$$\hat{\beta} = argmin \sum_{i=N}^{N} (y_i - \sum_{j=1}^{p} x_{ij})^2$$

subject to

$$\sum_{j=1}^{p} |\beta_j| <= t,$$

where $t >= 0$ is a tuning parameter. We follow Belloni et al. (2012). This robust LASSO approach allows an estimation under heteroscedastic non-Gaussian and clustered disturbances (RLASSO)

$$\hat{\beta} = argmin \sum_{i=1}^{N} (y_i - \sum_{j=1}^{p} x_{ij} b_j)^2 + \lambda \sum_{j=1}^{p} |b_j| \gamma_j,$$

where $\lambda > 0$ is the "penalty level" and γ_j are the "penalty loadings".

This procedure selects fewer significant variables than LARS. A problem with RLASSO remains also the pre-selection of variables. It has an influence on the result, which variables are finally selected. This is shown in Table 6, where a preselection is made according to three different criteria:

(1) all variables from Table 4;

(2) only those variables that are significant at the classical standard error in Table 4;

(3) all variables selected according to LARS.

The specification in column (1) is supplemented with

- the interaction between MIGRANTS and SCHOOL, each formed as dummy variables in column (4),

- TEST, VACCINE and trend variables from Table 5 in column (5) as additional determinants.

The central determinants in columns (1)-(3) of Table 6 turn out to be the dummy variable northern and eastern states versus the rest, the migrant share, schooling, the poverty variable, hours of sunshine per year and population density. Plenty of sunshine contributes to more close contacts between people. This increases the risk of infections. Among personality traits, two of the Big Five characteristics are significant, namely CONSC and AGREE according to the

RLASSO analysis. Both, regional variables derived from official statistics and personality traits aggregated at the federal state level, are found to be relevant for the unequal spread of COVID-19.

We find that RLASSO selects $MIGRANTS$ as an important determinant of COVID-19. In Table 6 the link between these two variables is positively significant. If RLASSO is supplemented by the squared term of migrants as Section 4 suggests $MIGRANTS^2$ is, however, not selected and the cluster robust estimation of the extended version of the RLASSO specification shows: $t_{MIGRANT}$=-0.40 and $t_{MIGRANTS^2}$=0.90. Therefore, $MIGRANTS^2$ is neglected in the following.

Why federal states with a comparatively higher proportion of migrants are more affected by the pandemic than others can be explained by various arguments. Especially in the first phase of the pandemic, not all fellow citizens with a migrant background have fully understood the restrictions imposed due to language difficulties. Therefore, they did not comply with them. After initial difficulties, the decrees on the containment of corona infections were also announced by the government in various languages. As effect it would have been expected a downward trend in the disproportionately high number of migrants affected. This cannot be clearly demonstrated. However, the results in Section 7 provide indications for this. Even during the third wave, language difficulties were still blamed in the media for the fact that migrants are affected by corona to a greater extent than citizens of German origin.

Poorer economic conditions, combined with less favorable housing conditions, may be decisive for the fact that distance rules are less easily observed among migrants. It is also possible that the behavior of migrants from different countries differs significantly from that of the native population. More intensive contacts with other people, especially with people from the same countries of origin, from the same community, may be given a higher priority. If the proportion of migrants of the same original nationality is particularly high in one federal state, then the tendency to integrate, to give up habits from the country of origin, is lower than elsewhere.

If the above arguments are viable in the context of migrants, and migrants tend to have lower schooling and be poorer, then it would be expected that low schooling and poverty would tend to be associated with a higher risk of contagion. However, the results in columns (1) and (2) of Table 6 do not initially point in this direction, but provide strong evidence to the contrary. One explanation for this could be that people with a better school education are more mobile than others, both professionally and privately, have more contacts than those without a school-leaving qualification, and are thus also exposed to a higher risk of infection.

In addition, older people are more likely than younger people not to have a school-leaving qualification and, regardless of this, they are limited in their mobility. It should be noted, however, that the schooling effect on COVID-19 is reversed for the migrant group. In column (4) of Table 6, this becomes clear when the interaction effect of MIGRANTS and SCHOOL is taken into account

$$COVID - 19 = x'\beta + (MIGRANTS \times SCHOOL)\gamma + u$$

The extension of the model in column (4), Table 6 including the TREND, TEST and VACCINE variables from column (5), Table 5 shows, that although the signs of the effects are preserved, they are insignificant for migrants (MIGRANTS, MIGRANTS_D*SCHOOL_D) and the coefficients are lower. Tests and vaccinations that reduce the risk of infection are less common among migrants. If they are not taken into account as regressors, a partial effect is attributed to migrants. Or, in other words, the migrants' risk of infection is overestimated.

A priori, the relationship between poverty and COVID-19 is by no means clear. Wright et al. (2020) derive in a theoretical model that in poorer regions the distance rules are less observed during the pandemic and therefore lead to more infections. A similar argument can be made for poorer people as for school education. Poorer families are less mobile due to scarce resources, more location-bound and thus less exposed to the risk of infection. They are less able to afford, for example, trips abroad where they encounter corona hotspots, so that infection risks are reduced. This is less true for migrants, even if they are in a poor economic position. The attachment to their home countries, which usually remains, prompts them to travel to their home countries and to the relatives and acquaintances, despite having few resources.

Of the personality traits assessed, the RLASSO selection procedure identifies conscientiousness (CONSC) as an important trait. If a person's conscientiousness trait is weak, reflecting carelessness and negligence, a high risk of infection should be associated because less attention is paid to compliance with infection prevention measures. Empirical evidence suggests the opposite. This result can be reconciled with the view held by personality researchers that conscientiousness is simultaneously the product of a motivating and a disciplining psychological force. On the one hand this is motivating, because such people can concentrate entirely on a job. And on the other this is disciplining because they can ascetically block tempting distractions (Bonelli 2014, pp. 62). The dangers of corona infections are simply ignored by such people.

A second selected personality trait by RLASSO is agreeableness. People with such a characteristic (AGREE) are attracted to others, have many contacts professionally and privately and are therefore more exposed to corona risks.

RLASSO-based specifications and the resulting estimates should examine the extent to which behavioral characteristics influence non-behavioral influences. The overall statistical influence can be quickly tested with a simple F-test, i.e.

the residual sum of squares of the restricted model without behavioral variables ($\tilde{u}'\tilde{u}$) is compared with that of the extended model ($\hat{u}'\hat{u}$), weighted with the degrees of freedom.

$$\frac{\tilde{u}'\tilde{u} - \hat{u}'\hat{u}}{\hat{u}'\hat{u}} \times \frac{N-k}{l} = T.$$

The residuum of the restricted model (r model) is $\tilde{u}$, that of the unrestricted model (u-model) ist $\hat{u}$. N is the number of observations, k is the number of regressors of the unrestriced model and l is the difference of regressors of the u and r model. If the hypothesis T=0 is rejected, the behavior variables correlate with the regressors of the r model. It follows that F=8.58 > F(3; 4809). This means the effect cannot be neglected. Following the partitioned model

$$y = X_1'\beta_1 + X_2'\beta_2 + u,$$

where X_1 are the regressors of the r model and X_2 are the personality traits, it can be shown (Hübler 1989, S.108), how strongly the OLS estimates of β_1 are biased, if this vector is only determined for the r model

$$y = X_1'\beta + u_1.$$

Via the auxiliary model $X_2 = X_1F + V$ it follows

$$\tilde{\beta}_1 = (X_1'X_1)^{-1}X_1'y = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 =: \beta_1 + \hat{F}\beta_2.$$

Nevertheless, it is still unclear, which contribution have to be allocated the single regressors of the r model. This can be done with the Gelbach decomposition (Gelbach 2016). For this purpose we need the decomposition of the OLS estimates of the auxiliary model

$$\hat{F}_k = (X_1'X_1)^{-1}X_1'X_{2k}\hat{\beta}_{2k},$$

where $k = 1, \ldots, l$. From the OLS estimates of the u model we can determine $\hat{\beta}_{2k}$. If the classical conditions of the error term are not fulfilled and generalized LS estimates are applied, we have to note, that in contrast to the classical model the EGLS estimation of the u model usually does not lead to the same result as the OLS estimation of

$$(I - (X_1'X_1)^{-1}X_1')y \neq (I - (X_1'X_1)^{-1}X_1')X_2\beta_2 + (I - (X_1'X_1)^{-1}X_1')u$$

(Fiebig et al 1996).

When the significance of the behavioral variables CONSC and AGREE in Table 6, column (1) and column (3) is investigated, Table 7a shows that the influence of MIGRANTS on COVID-19 is overestimated if CONSC and AGREE are disregarded. The latter exert a joint significant influence on the regressor

MIGRANTS. The coefficient for MIGRANTS decreases from 0.0604 to 0.0475 in the transition from the restricted to the unrestricted model. The absolute change effect is significant and amounts to 0.0129 - see Table 7b. Significant changes are also caused with respect to NE-SW, SUN and POVERTY. This is not the case for SCHOOL.

The next step is to examine whether the regional differences are primarily due to observed determinants or primarily to unobserved influences. This can be tested with the help of a decomposition known from discrimination theory (Blinder 1973). According to Table 8, with the selected specification from Table 6, column (1) without the variable NE-SW, 71.5% of the COVID-19 differences between northern/eastern states on the one hand (NE-SW=1) and southern/western states on the other (NE-SW=0) are explained by the model. The rest (28.5%) is due to unexplained variation.

A specification error in the information aggregation by RLASSO may result from the fact that unobserved regional heterogeneity is not taken into account. Whether this results in less reliable differences is to be tested by panel data analyses. Fixed effects estimates are not appropriate here, however, because the data are based on time-invariant regressors, especially for personality traits, which are not taken into account as regressors and cannot be separated from the individual effects. Random effects estimates hardly differ from the pooled estimates due to the measurement of the regressors, as the comparison of column (1) and (4) from Table 6 with column (1) and (2) in Table 9 shows. However, regional effects (federal state effects) can also occur, which can be captured by regional dummy variables and are quite predominantly significant. Estimates for this can be found in Table 9, columns (3) and (4). Compared to the estimates without federal state effects, fewer significant influences can be identified. This suggests that the significance of MIGRANTS, MIGRANTS*SCHOOL and SUN for COVID-19 is driven more by structural differences between the federal states than directly by these characteristics.

7 Changes over the course of the pandemic

The next step is to examine whether the results reported so far are time-invariant or whether there are structural breaks. For this purpose, the RLASSO specifications are first determined, separated by month (March 2020 to June 2021) for a five-day interval in the middle of the month (Table 10). There are changes at this level, but no clear structural breaks. It should be mentioned that at the beginning of the pandemic (March 2020), the lower affectedness of the northern and eastern federal states compared to the rest does not yet show through. Otherwise, the correlation is statistically secure and shows increasing significance over time. The estimated coefficients are larger in the last months of 2020 and then until the end of the period under consideration, i.e. until 06/2021, than before.

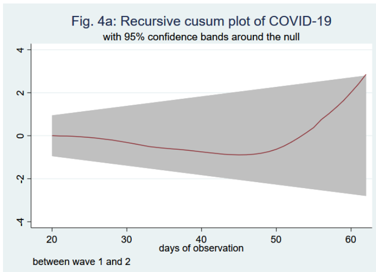
The higher is the proportion of migrants in a federal state, the higher the number of infections. The significance of this factor initially remains largely unchanged in the course of the pandemic, but then declines and no significant correlations are shown from 02/2021 onwards. For all months, the coefficients determined for the SCHOOL variable are negative and significant until the end of 2020. Thus, in federal states with a relatively high proportion of persons without a school-leaving certificate, fewer Corona infections were observed. While families at risk of poverty were exposed to a lower risk of infection until autumn 2020, a change of sign becomes apparent for the later period. A statistically reliable increase in the risk of infection of the poorer population can only be seen for the last two months. In the later course of the pandemic from 2021 onwards, the explanatory pattern of COVID-19 by the RLASSO model appears to be less stable. This is manifested by an increase in insignificant regressors.

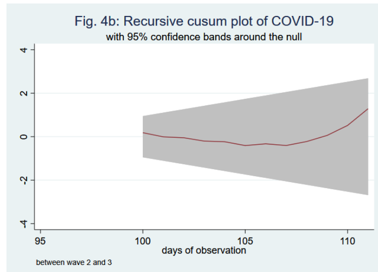
It should be noted that the month-by-month separation of the estimates is not without arbitrariness. Against the background of the course of the pandemic a separation between a first, second and third wave (W1, W2 and W3) is appropriate, as the RKI graph suggests (https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Corona−virus/Situationsberichte/Jul_2021/2021−07−15−en.pdf ?_blob = publicationFile). However, a clear demarcation of the phases is not possible. In some cases, virologists have already recognized tendencies of a new wave in the course of the pandemic, when this has not yet been expressed in the indicators. After looking at the RKI graph, the following provisional classification can be made

W1: until the end of May 2020
W2: from the beginning of June 2020 to the beginning of March 2021
W3: from mid-March 2021.

Due to the fuzziness of the demarcation between the different phases, it is appropriate to test for structural breaks. The tests are only based on data for Germany as a whole on a day-by-day basis using the Johns Hopkins University international data. The break between wave 1 and wave 2 is determined within the period 27 April.2020 to 28 September 2020. Analogously for wave 2 and wave 3, the test is conducted for data between 1 February 2021 and 18 March 2021.

Table 11a shows that a significant structural break is detected between W1 and W2 based on a Wald test after 48 observation points, i.e. on 18 August 2020. This is not already detected at the low point of the 7-day incidence, but only later. Then the increase can no longer be interpreted as random. Between the second and the third wave, a break is indicated after 108 observation points on 8 March 2021 (Table 11b). Again, the break is not indicated at the low point of the 7-day incidence, but with a certain lag. Alternatively, the breaks are revealed by recursive cusum plots in Fig. 4a and 4b





An open question remains. Have the determinants of the regional infection gradient remained unchanged for the entire period or are there wave-specific structural patterns? Results on this can be found in Table 12. From there, the following should be noted:

(1) The differences between the north-eastern federal states and the south-western ones have increased from wave to wave.
(2) Federal states with a high proportion of migrants were clearly more affected by the pandemic than others during the first and second waves.
(3) In the federal states with a comparatively high proportion of residents without a school-leaving certificate, the incidence of infection was lower than in other federal states across all three waves. However, the coefficients were declining from wave-to-wave.
(4) During the first wave, federal states with a high proportion of families at risk of poverty were at lower risk of infection than others. This relationship reversed during the pandemic.
(5) No clear pattern across the three waves can be identified in terms of hours of sunshine per year.
(6) In states with a comparatively high proportion of conscientious COVID-19 cases are different than elsewhere. The coefficients of this factor have decreased from wave to wave.

8 Conclusions

The article has chosen regions as the level of observation. This is not synonymous with analyses at the individual or company level, nor at the macroeconomic or international level. International comparisons are difficult, as e.g. Alain Berset, member of the Swiss Federal Council since 2012, explains in an interview (https://www.welt.de/politik/ausland/plus231733177/Erfolgreich-in-der-Pandemie-Warum-hat-die-Schweiz-es-ganz-anders-gemacht-als-Deutschland .html) and the statistician, Walter Krämer, emphasizes ($2021-02-25\_Unsta-tistics\_February.pdf$). In the case of federal states, the focus is on the meso view, which also includes elements from the micro and macro levels. The advantage of this approach in the context of COVID-19 for a federal state like Germany is above all that the situation of a social experiment can be depicted. The federal government and the prime ministers of the federal states (MPK) create the basic regulations and the states implement them, taking regional conditions into account.

This study has worked out which possible influences are rather negligible and which are not. The latter include above all the agglomeration of regions, migration background, school education, poverty and aggregated personality characteristics. There are correlations between these. A clean delimitation is difficult. Contrary to widespread opinion, poverty risk and school education without a diploma are not the decisive infection drivers. The greater is the willingness to cooperate, the higher is the risk of infection. The empirical investigation provides evidence that unobserved heterogeneity remains a major determinant. A significant but non-linear correlation between infections and tests is depicted by the estimates. The more tests are performed in a province, the more infections are detected. A positive linear link is shown between vaccinations and the number of infections. The more vaccinations are carried out, the fewer new infections are observed. IV estimates are similar to classical estimates with cluster robust standard errors.

The most important result of the course analysis is that the significance of the migrant share for the spread of new infections has decreased in the course of the pandemic. The effect of the poverty factor is also subject to change. While families at risk of poverty were exposed to a comparatively low risk of infection in the first phase of the pandemic, this result is reversed later on. The lower exposure of the northern and eastern federal states compared to the rest became more evident in the second and third waves.

Overall, the empirical analyses show a clear specification and pandemic duration dependency. A very careful analysis is needed to arrive at reliable results. Unobserved influences that change over time give the impression of a diffuse appearance and make it difficult to uncover clear relationships. Whether and to what extent the occurrence of corona mutants plays a significant role in this must remain open at the current state of knowledge.

Further investigations should also examine to what extent the results react sensitively to a change in the measurement of the variables used. It needs to be clarified whether the correlations found are robust at the district level, with an extension of the observation period, with alternative data sets and the use of further variables. The warnings of the health authorities in the U.S., the health politician Lauterbach, the virologist Streeck and in the media of a fourth wave in autumn 2021 should be taken seriously. In the U.S. and Israel we observe currently a clear tendency to a fourth wave. At the regional level in Germany, the question must be pursued as to where and from when clear indications can be recognized in order to be able to take countermeasures at an early stage. Not every temporary improvement in the incidence of infection, regardless of which indicator is used, should be used immediately to ease the situation. The course of the previous three waves has shown that the sooner political measures are taken, the sooner a deterioration of the situation can be stopped.

References

Acemoglu D, Makhdoumi A, Malekian A, Ozdaglar A (2020), Testing, Voluntary Social Distancing and the Spread of an Infection, NBER Working Paper 27483

Akbarpour M, Cook C, Marzuoli A et al. (2020), Socioeconomic Network Heterogeneity and Pandemic Policy Response, Becker Friedman Institute for Economics, Working Paper Series, Vol. 11 No. 17

Belloni A, Chen D, Chernozhukov V, Hansen C (2012), Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain, Econometrica 80:6, 2369-2429

Benitez J, Courtemanche C, Yelowitz A (2020) , Racial and Ethnic Disparities in COVID-19: Evidence from Six Large Cities, IZA Discussion Paper No. 13521

Bonelli R M (2014), Perfektionismus: Wenn das Soll zum Muss wird. Pattloch-Verlag, München

Blinder A (1973), Wage Discrimination: Reduced Form and Structural Estimates. Journal of Human Resources 18, 436-455

Brown C S, Ravallion M (2020), Inequality and the Coronavirus: Socioeconomic Covariates of Behavioral Responses and Viral Outcomes across US Counties, NBER Working Paper 27549.

Burdett A, Davillas A and Etheridge B (2021), Weather, Psychological Wellbeing and Mobility during the first wave of the COVID-19 Pandemic, IZA DP No. 14119

Chang S, Pierson E, Koh P, Gerardin J, Redbird B, Grusky D (2021), Mobility network models of COVID-19 explain inequities and inform reopening, Nature 589

Deopa N and Fortunato P (2021), Coronagraben in Switzerland: culture and social distancing in times of COVID-19, Journal of Population Economics 34, 1355-1383

Desmet K, Wacziarg, R (2020), Understanding Spatial Variation in COVID-19 across the United States, NBER Working Paper 27329

Efron B, Hastie T, Johnstone I, Tibshirani R (2004), Least Angle Regression, Annals of Statistics 32:2, 407-451

Fiebig D G, Bartels R, Krämer W (1996), The Frisch-Waugh theorem and generalized least squares.Econometric Reviews, 15(4), 1996, 431 443

Galasso V, Pons V, Profeta P, Becher M, Brouard S, Foucault M (2020), Gender Differences In COVID-19 Related Attitudes And Behavior: Evidence From A Panel Survey In Eight OECD Countries, NBER Working Paper 27359

Gelbach J B (2016), When Do Covariates Matter? And Which Ones, and How Much? Journal of Labor Economics 34:2, 509-543

Goldstein J R, Lee R D (2020), Demographic Perspectives on Mortality of COVID-19 and Other Epidemics, NBER Working Paper 27043

Hübler O (1989), Ökonometrie, UTB Gustav Fischer, Stuttgart

Hübler O (2014), Estimation of standard errors and treatment effects in empirical economicsmethods and applications, J Labour Market Res. 47, 4362

Hübler O (2021), Ungleich verteilte Corona-Infektionen zwischen den Bundesländern, Hannover Economic Papers 687.

Knittel C R, Ozaltun B (2020), What Does and Does Not Correlate With COVID-19 Death Rates, NBER Working Paper 27391

Krekel C, Swanke S, De Neve J-E, Fancourt D (2020), Are Happier People More Compliant? Global Evidence from Three Large-Scale Surveys during Covid-19 Lockdowns, IZA Discussion Paper 13690

MacKinnon J G (2019), How cluster-robust inference is changing applied econometrics, Canadian Journal of Economics 52:3, 851- 881

McLaren J (2020), Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census Data, NBER Working Paper 27407

Moulton B R (1990), An illustration of a pitfall in estimating the effects of aggregate variables on micro units. Rev. Econ. Stat. 72, 334338

Papageorge N W, Zahn M V, Belot M, van den Broek-Altenburg E (2021), Socio-Demographic Factors Associated with Self-Protecting Behavior during the COVID-19 Pandemic, J Popul Econ 34(2):691738

Puhani P A (2020), France and Germany Exceed Italy, South Korea and Japan in Temperature-Adjusted Corona Proliferation A Quick and Dirty Sunday Morning Analysis, Hannover Economic Papers 668

Qui Y, Chen X and Shi W. (2020), Impacts of Social and Economic Factors on the Transmission of Coronavirus Disease 2019 (COVID-19) in China, J Popul Econ 33:11271172

Sa F (2020), Socioeconomic Determinants of COVID-19 Infections and Mortality: Evidence from England and Wales, IZA Policy Paper No.159

Tibshirani R (1996), Regression Shrinkage and Selection via the Lasso', Journal of the Royal Statistical Society, B 58(1), 267-288

Toya H and Skidmore M (2021), A Cross-Country Analysis of the Determinants of Covid-19 Fatalities, CESifo Working Papers 9028

Wright A L, Sonin K, Driscoll J, Wilson J (2020), Poverty and Economic Dislocation Reduce Compliance with COVID-19 Shelter-in-Place Protocols, Becker Friedman Institute, Working Paper No. 2020-40

Table 1: Definition, measurement of variables and source

Acronym	Definition	Measurement	Source
COVID-19	infections/population	in logarithm	Robert-Koch-Institute
UR	unemployment rate	in percent; 2019	Federal Statistical Office
GDP	gross domestic product	in mill. euros p.c.; 2019	Federal Statistical Office
WOMEN	proportion of women	in percent; 2016	Federal Statistical Office
UNI	higher education	in percent; 2020	Federal Statistical Office
SCHOOL	no degree in schooling	in percent; 2020	Federal Statistical Office
SATIS	life satisfaction	0=no,. . .,10=yes; 2019	Federal Statistical Office
DEATHS	number of deaths	absolute number; 2019	Federal Statistical Office
Y_UR	youth unemployment	in percent; 2020	Federal Statistical Office
LIFE	life expectancy	in years; 2019	Federal Statistical Office
CRIMES	crimes in one year	absolute number; 2020	Federal Statistical Office
DEPT	total dept	in euros; 2020	Federal Statistical Office
CARE	child care rate	in percent; 2020	Federal Statistical Office
POVERTY	risk of poverty	in percent; 2019	Federal Statistical Office
ASYLUM	asylum application	absolute number; 2020	Federal Statistical Office
DEPORT	deportations Germany	absolute number; 2019	Federal Statistical Office
CDU	Christian Democrats	voter share in percent	STATISTA until 2020
SPD	Social Democrats	voter share in percent	STATISTA until 2020
GREEN	The Greens	voter share in percent	STATISTA until 2020
FDP	Free Democrats	voter share in percent	STATISTA until 2020
LEFT	The Left	voter share in percent	STATISTA until 2020
AfD	Alternative for Germany	voter share in percent	STATISTA until 2020
CLARIFIC	clarification of crimes	in percent	Federal Statistical Office
PRISON	number of prisoners	in percent	Federal Statistical Office
TEMP	average temperature	in degrees Celsius	Federal Statistical Office
SUNSHINE	sunshine	p..a. in hours	Federal Statistical Office
RAIN	rainfall	litres/square metre	Federal Statistical Office
BORDER	foreign border	in kilometres	Federal Statistical Office
YOUNG	people 15-25 years old	in percent	Census 2011
OLD	people >= 65 years	in percent	Census 2011
SIZE	average(av.) household size	absolute size	Microcensus 2018
MIGRANTS	migrants share	in percent	Genesis 2016
POPULAR	popularity index	neg./pos.number of votes	YouGov Institute 2016
HEALTHY	health status is good	in percent	GSOEP 2018, wave 35
SICK	health status is bad	in percent	GSOEP 2018, wave 35
EXTRA	av. extraversion	1=no,. . . , 7=yes; 3 items	GSOEP 2013, wave 30
OPEN	av. openness	1=no,. . . , 7=yes; 3 items	GSOEP 2013, wave 30
AGREE	av. agreeableness	1=no,. . . , 7=yes; 3 items	GSOEP 2013, wave 30
NEURO	av. neuroticism	1=no,. . . , 7=yes; 3 items	GSOEP 2013, wave 30
CONSC	av. conscientiousness	1=no,. . . , 7=yes; 3 items	GSOEP 2013, wave30
SELF	av. self-confident	1=no,. . . , 3=yes	GSOEP 2018, wave 35
OPTIM	av. optimistic	1=yes,ldots, 4=no	GSOEP 2018, wave 35
PATIENT	av. patient	0=no,. . . , 10=yes	GSOEP 2018, wave 35
RISK	av. risk-taking	0=no,. . . , 10=yes	GSOEP 2018, wave 35
IMPULSE	av. impulsive	0=no,. . . , 10=yes	GSOEP 2018, wave 35

Table 2: Descriptive statistics of regional variables

Acronym	Obs	R	Mean	Std. Dev.	Min	Max
DENSITY	7,722	0.0771*	689.4375	1077.032	69	4090
BORDER	7,722	0.0832*	3.635756	3.064795	0	8.8917
WEST-EAST	7,722	0.1019*	.6875	.4635424	0	1
NO-SW	7,722	-0.0938*	.5	.5000324	0	1
AREA-CITY	7,722	-0.0647*	.8125	.3903376	0	1
UR	7,722	-0.0416*	5.94375	1.747242	3.2	10.2
GDP	7,722	0.1233*	39193.69	9842.448	27905	65603
YOUNG	7,722	0.0905*	10.6875	1.004353	9	12
OLD	7,722	-0.1013*	21.16875	1.691159	19	24.8
LIFE	7,722	0.1062*	78.075	.8750557	76.3	79.7
UNI	7,722	0.0303*	40.8375	5.847266	32.1	54.8
SCHOOL	7,722	-0.0231	7.4625	4.288907	4.7	18.5
MIGRANTS	7,722	0.1317*	10.925	4.995383	4.5	18.5
SIZE	7,722	0.0171	1.951875	.0832384	1.79	2.09
HEALTHY	7,722	0.0253	45.48563	5.115019	37	59.41
SICK	7,374	0.0509*	20.01196	2.548764	14	24.27
WOMEN	7,722	0.0011	50.70625	.1819076	50.4	51
POPULAR	7,722	0.0874*	21.66875	18.93673	-13	59.1
DEATHS	7,722	0.0734*	58720	51808.78	7704	206479
Y_UR	7,722	-0.0809*	5.80625	2.029746	2.5	9.3
DEPT	7,722	0.0518*	37650	39597.78	1388	177670
CARE	7,722	-0-0949*	40.225	11.73343	28.2	58
POVERTY	7,722	-0.0828*	16.55625	2.776186	11.7	22.7
ASYLUM	7,722	0.0761*	2955.625	2781.009	449	11578
DEPORT	7,722	0.0845*	1347.438	1578.32	93	6359
CDU	7,722	0.0272	27.25	7.954759	11.2	40.7
SPD	7,722	-0.0343	23.2875	10.38882	7.7	39.2
GREEN	7,722	0.0958*	12.26875	7.592981	64	30.3
FDP	7,722	0.0186	66.31875	2.608861	3	12.6
LEFT	7,722	-0.0326*	9.94375	7.032966	2.8	31
AfD	7,722	-0.0609*	13.85625	7.493062	5.3	27.5
CRIMES	7,722	0.0848*	339900.1	287115.4	74719	1227929
CLARIFIC	7,722	-0.0383*	57.175	6.588246	44.7	67
PRISONS	7,722	0.0938*	3723.563	3736.475	697	14490
TEMP	7,722	-0.0171	10.525	.5190471	9.5	11.7
SUN	7,722	0.0315*	1807.188	105.4518	1645	1970
RAIN	7,722	0.0843*	707.5	143.4818	475	980

Note: R - correlation coefficient COVID-19 and a regional variable; * p≤.05.

Source: Robert-Koch-Institute, Federal Statistical Office, Statista, Microcensus, Genesis

Table 2a: Regional averages of personality traits - descriptive statistics

Acronym	Obs	R	Mean	Std. Dev.	Min	Max
SATIS	7,722	0.0541*	7.11	.171366	6.76	7.44
EXTRA	7,722	0.0405*	15.03515	.2137706	14.7091	15.5341
OPEN	7,722	0.0315*	13.63865	.7605371	12.4307	15.5563
AGREE	7,722	0.0797*	14.41021	.5829772	13.3142	15.6557
NEURO	7,722	0.0449*	12.53352	.4466929	11.787	13.414
CONSC	7,722	0.0784*	14.20893	.4002733	13.6367	15.2323
OPTIM	7,722	-0.0362*	2.276363	.0938235	2.1194	2.4374
PATIENT	7,722	0.0334*	6.061288	.1549098	5.862	6.5528
RISK	7,722	-0.0002	4.229663	.1377881	3.9711	4.4671
IMPULSE	7,722	0.0938*	4.834106	.1585452	4.5505	5.118
SELF	7,722	0.0839*	2.327325	.0645483	2.2113	2.415
POPULAR	7,722	0.0874*	21.66875	18.93673	-13	59.1

Note: R - correlation coefficient COVID-19 and a regional variable; * $p \leq .05$.

Source: GSOEP, wave 30-35, aggregated on German states level and YouGov Institute 2016

Table 3: Averages of regional variables, split by German states

GERMAN STATE	SH	MV	HH	HB
COVID_19	7586.8	8114.0	8114.0	14352.4
DENSITY	183.0	69.0	69.0	1629.0
UR	5.4	7.6	7.6	10.2
YOUNG	11.0	9.7	9.7	11.8
OLD	21.7	22.1	22.1	21.2
SCHOOL	7.0	13.7	13.7	4.9
POVERTY	15.3	20.9	20.9	22.7
MIGRANT	8.0	4.5	4.5	18.1
GERMAN STATE	NI	NW	RP	SL
COVID_19	11241.8	16322.5	13236.4	15207.0
DENSITY	167.0	526.0	206.0	385.0
UR	5.0	6.7	4.7	6.5
YOUNG	11.4	11.5	11.6	11.0
OLD	20.8	20.3	20.3	22.0
SCHOOL	6.0	5.5	5.8	5.8
POVERTY	15.9	18.1	15.4	16.0
MIGRANT	9.4	13.3	11.1	11.1
GERMAN STATE	BW	BY	BE	BB
COVID_19	16217.7	18492.7	14626.7	18788.0
DENSITY	310.0	185.0	85.0	4090.0
UR	3.5	3.2	5.9	8.0
YOUNG	12.0	11.6	9.0	10.6
OLD	19.4	19.5	22.7	19.2
SCHOOL	5.1	5.5	4.7	18.5
POVERTY	11.9	11.7	15.2	18.2
MIGRANT	15.6	13.2	4.7	18.5
GERMAN STATE	TH	SN	ST	HE
COVID_19	18829.4	23239.5	13681.5	16016.1
DENSITY	132.0	221.0	108.0	297.0
UR	4.4	5.6	7.4	4.6
YOUNG	9.4	9.1	9.3	11.1
OLD	21.7	24.8	24.3	19.7
SCHOOL	4.9	4.9	4.9	16.2
POVERTY	16.4	16.6	19.5	15.8
MIGRANT	4.9	4.9	4.9	16.2

Note: SH-Schleswig-Holstein; MV-Mecklenburg-Western Pomerania; HH-Hamburg; HB-Bremen; NI-Lower Saxony; NW-North Rhine-Westphalia; RP-Rhineland-Palatinate; SL-Saarland; BW-Baden-Wuerttemberg; BY-Bavaria; BE-Berlin; BB-Brandenburg; TH-Thuringia; SN-Saxony; ST-Saxony-Anhalt; HE-Hesse

Source: Robert-Koch-Institute and Federal Statistical Office

Table 4: COVID-19 estimates, separated for substantial blocks

	coef.	t_c	t_{cr}	Variable	coef.	t_c	t_{cr}
DENSITY/1000	.1415	4.74	2.89	MIGRANTS	.0925	8.60	2.28
POPULATION	-.0096	-1.37	-1.76	YOUNG	-.2062	-4.20	-1.49
WEST-EAST	.1386	1.77	0.97	OLD	.0406	1.74	0.59
NE-SW	-.4402	-5.71	-2.95	WOMEN	.0324	1.16	0.10
NORTH-SOUTH	.2310	2.15	1.25	SIZE	1.9131	4.48	1.46
BORDER	.0405	3.38	2.54				
COMMUTER/1000	.0178	6.03	4.97				
constant	8.1643	79.22	42.26		3.5597	0.52	0.19
N	7,722			7,722			
R^2	0.03			0.02			
TEMP	.2187	4.35	1.22	GDP/1000	.01730	5.99	1.79
SUN	.0012	5.90	1.73	UR	.0862	2.25	0.74
RAIN	.0018	9.60	2.68	Y_UR	-.0094	-0.30	-0.09
				POVERTY	-.0845	-3.91	-1.44
constant	2.7618	3.51	0.94		8.8679	36.73	13.56
N	7,722				7,722		
R^2	0.01				0.02		
ASYLUM/1000	-.0861	-1.12	-0.30	LIFE	.1264	4.33	1.25
DEPORT/1000	-.2549	-2.75	-1.22	HEALTHY	.0175	4.09	1.15
CRIMES/10000	-.0019	-0.33	-0.11	SICK	.0488	5.51	1.43
CLARIFIC	-.0271	-4.74	-1.13	DEATHS/1000	.0018	3.87	1.81
PRISON/1000	.25581	5.44	2.34				
DEPT/1000	-.0021	-1.29	-0.52				
constant	9.9441	9.9441	7.46		-3.1787	-1.39	-0.40
N	7,722				7,722		
R^2	0.02				0.01		
UNI	.0178	4.74	1.67	CDU	.0120	1.49	0.63
SCHOOL	.0138	-2.74	-0.76	SPD	-.0075	-1.07	-0.53
CARE	-.0162	-8.95	-2.82	GREEN	.0254	3.01	1.25
				FDP	-.0173	-1.68	-0.43
				LEFT	.0094	1.24	0.63
				AfD	-.0176	-2.02	-0.96
constant	8.6300	55.65	17.13		8.3959	12.31	5.95
N	7,722				7,722		
R^2	0.01				0.02		

Table 4: Continuation - COVID-19 estimates under regional averages of personality traits only as determinants

	coef.	t_c	t_{cr}
SATIS	-.2479	-2.96	-7.28
EXTRA	-.7377	-3.47	-1.28
OPEN	.0522	1.04	0.33
AGREE	.0249	0.40	0.14
NEURO	.1429	1.24	0.49
CONSC	.2642	2.93	1.09
OPTIM	1.8900	3.51	0.95
PATIENT	.8442	3.50	1.57
RISK	-.7934	-2.17	-0.92
IMPULSE	1.5225	7.40	2.29
POPULAR	.0155	6.85	1.45
cons	.8890	0.26	0.10
N	7,722		
R^2	0.02		

Note: t_c - t-statistics based on classical standard errors; t_{cr} - t-statistics based on cluster-robust standard errors

Source: GSOEP, wave 30-35, aggregated on German states level and YouGov Institute 2016

Table 5: COVID-19 regressions with testing and vaccine

	(1) coef.	(2) coef.	(3) coef.	(4) coef.	(5) coef.
TREND	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
TESTING	0.017***	0.161***			0.161***
	(0.00)	(0.01)			(0.01)
TESTING^2		-0.007***			-0.007***
		(0.00)			(0.00)
VACCINE			0.024*	-0.031	0.024*
			(0.01)	(0.04)	(0.01)
VACCINE^2				0.003	
				(0.00)	
constant	9.750***	8.958***	9.765***	9.900***	8.763***
	(0.10)	(0.14)	(0.12)	(0.15)	(0.17)
N	2848	2848	2848	2848	2848
R^2	0.396	0.400	0.497	0.535	0.507

Note: Cluster-robust standard errors in parentheses;
* $p \leq .10$, ** $p \leq .05$, *** $p \leq .01$

Table 5a: COVID-19 IV regressions with trend, testing and vaccine

IV	(1) DEATHS coef	(2) DEATHS DOUBLING . coef.
TREND	0.005***	0.006***
	(0.00)	(0.00)
TESTING	0.154***	0.360***
	(0.01)	(0.12)
TESTING^2	-0.006***	-0.016***
	(0.00)	(0.01)
VACCINE	0.083***	0.082***
	(0.03)	(0.03)
constant	8.343***	7.203***
	(0.32)	(0.70)
N	2800	2800
EXO-TEST	11.92***	13.93***

Note: Cluster-robust standard errors in parentheses; * p≤.10, ** p≤.05, *** p≤.01

Table 6: COVID-19 regressions, specifies by RLASSO

	(1) coef.	(2) coef.	(3) coef.	(4) coef.	(5) coef.
NE-SW	-0.212***	-0.303***	-0.403***	-0.196***	-0.339***
	(0.07)	(0.06)	(0.09)	(0.06)	(0.06)
MIGRANTS	0.052***	0.036***	0.036***	0.044***	0.013
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SCHOOL	-0.035***	-0.034***		-0.064***	-0.047***
	(0.01)	(0.01)		(0.01)	(0.01)
POVERTY	-0.009	-0.023**		-0.006	0.049***
	(0.01)	(0.01)		(0.01)	(0.01)
SUN	0.001**	0.001**		0.001**	0.001**
	(0.00)	(0.00)		(0.00)	(0.00)
CONSC	0.247***			0.268***	0.260***
	(0.07)			(0.06)	(0.07)
DENSITY		0.159***			
		(0.00)			
AGREE			0.245***		
			(0.07)		
MIGRANTS*SCHOOL				0.454***	0.245
				(0.15)	(0.18)
TESTING					0.161***
					(0.01)
TESTING^2					-0.007***
					(0.00)
VACCINE					0.030***
					(0.01)
TREND					0.005***
					(0.00)
constant	3.516***	7.700***	4.883***	3.781***	2.994**
	(1.16)	(0.57)	(0.93)	(1.09)	(1.254)
N	7,722	7,722	7,722	7,722	2,848
R^2	0.033	0.034	0.029	0.034	0.901

Note: Cluster-robust standard errors in parentheses;
* p≤.10, ** p≤.05, *** p≤.01

Table 7a: Restricted and unrestriced model, specified by RLASSO, without and with personality traits, measured by CONSC and AGREE

COVID-19	restricted coef.	model std. err.	t	unrestricted coef.	model std. err.	t
NE-SW	-.1600673	.048854	-3.28	-.2586916	.0536829	-4.82
MIGRANTS	.0604417	.0047593	12.70	.047537	.0055255	8.60
SCHOOL	-.0282565	.0054865	-5.15	-.02819	.0063105	-4.47
POVERTY	-.0158326	.0087528	-1.81	-.0107017	.0089535	-1.20
SUN	.00106	.0002209	4.80	.0007581	.0002303	3.29
CONSC				.1654412	.072474	2.28
AGREE				.1243467	.0470686	2.64
constant	6.580057	.4589055	14.34	3.088023	.945747	3.27
N	7,722			7,722		
R^2	0.0303			0.0336		

Note: Cluster-robust standard errors

Table 7b: GELBACH DECOMPOSITION - effects of average personality traits, measured by CONSC and AGREE on regional variables

COVID-19	coef.	std. err.	z
NE-SW	.0986243	.022531	4.38
MIGRANTS	.0129047	. 0028283	4.56
SCHOOL	-.0000665	. 0031428	-0.02
POVERTY	-.005131	.0019886	-2.58
SUN	.0003019	. 0000671	4.50
constant	3.492034	.8276144	4.22

Table 8: BLINDER DECOMPOSITION -
Endowment and unobserved differences between
north-east and south-west German states

Amount attributable:	126.5
- due to endowments (E):	-24.8
- due to coefficients (C):	151.3
Shift coefficient (U):	-161.2
Raw differential (R) E+C+U:	-34.7
Adjusted differential (D) C+U:	-9.9
Endowments effects as percent total (E/R):	71.5
Unobserved effects as percent total (D/R):	28.5

U = unexplained portion of differential
(difference between NORTH-EAST and SOUTH-WEST model constants)
D = portion due to unobserved variables (C+U)

positive number indicates advantage to NORTH-EAST German states
negative number indicates advantage to SOUTH-WEST German states

Table 9: COVID-19 RE panel estimates, specifies by RLASSO

	(1) coef.	(2) coef.	(3) coef.	(4) coef.
NO-SW	-0.212***	-0.196***		
	(0.07)	(0.06)		
MIGRANTS	0.052***	0.044***	0.070***	0.005
	(0.01)	(0.01)	(0.01)	(0.05)
SCHOOL	-0.035***	-0.064***	-0.029***	-0.087**
	(0.01)	(0.01)	(0.00)	(0.04)
POVERTY	-0.009	-0.006	0.008	0.032
	(0.01)	(0.01)	(0.01)	(0.02)
SUN	0.837***	0.636**	0.138	-0.020
	(0.30)	(0.27)	(0.37)	(0.84)
CONSC	0.247***	0.267***	0.461***	0.519***
	(0.07)	(0.06)	(0.03)	(0.12)
MIGRANTS*SCHOOL		0.453***		0.931
		(0.15)		(0.60)
constant	3.519***	3.785***	0.826	0.752
	(1.16)	(1.09)	(0.66)	(2.79)
STATE DUMMIES	no	no	yes	yes
N	7,722	7,722	7,722	7,722

Note: RE - random effects estimates; cluster-robust standard errors in parentheses; * p≤.10, ** p≤.05, *** p≤.01

Table10: Monthly COVID-19 regressions of RLASSO specification

	(1)	(2)	(3)	(4)
	coef.	coef.	coef.	coef.
MONTH	03/2020	04/2020	05/2020	06/2020
NE-SW	-0.103	-0.231**	-0.185*	-0.170*
	(0.09)	(0.08)	(0.10)	(0.10)
MIGRANTS	0.081***	0.057***	0.070***	0.081***
	(0.01)	(0.00)	(0.01)	(0.01)
SCHOOL	-0.038**	-0.048***	-0.053***	-0.055***
	(0.01)	(0.01)	(0.01)	(0.01)
POVERTY	-0.013	-0.079***	-0.064***	-0.045***
	(0.02)	(0.01)	(0.01)	(0.01)
SUN	-0.000	0.000	0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)
CONSC	0.388***	0.474***	0.287**	0.192
	(0.12)	(0.10)	(0.11)	(0.12)
constant	-1.029	1.315	3.520**	4.484**
	(1.43)	(1.31)	(1.57)	(1.79)
N	80	80	80	80
R^2	0.616	0.958	0.936	0.917

Note: Cluster-robust standard errors in parentheses;
* p≤.10, ** p≤.05, *** p≤.01

Table 10: Continuation - monthly COVID-19 regressions of RLASSO specification

	(1)	(2)	(3)	(4)
	coef.	coef.	coef.	coef.
MONTH	07/2020	08/2020	09/2020	10/2020
NE-SW	-0.180*	-0.185*	-0.193**	-0.189**
	(0.10)	(0.10)	(0.09)	(0.09)
MIGRANTS	0.082***	0.081***	0.086***	0.091***
	(0.01)	(0.01)	(0.01)	(0.01)
SCHOOL	-0.053***	-0.046***	-0.041***	-0.033***
	(0.01)	(0.01)	(0.01)	(0.01)
POVERTY	-0.040**	-0.038**	-0.042***	-0.019
	(0.01)	(0.01)	(0.01)	(0.01)
SUN	0.000	0.000	0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)
CONSC	0.222*	0.230*	0.220**	0.237**
	(0.12)	(0.11)	(0.10)	(0.11)
constant	3.907**	4.193**	4.563***	3.870**
	(1.77)	(1.64)	(1.45)	(1.55)
N	80	80	80	80
R^2	0.916	0.923	0.941	0.932

Note: Cluster-robust standard errors in parentheses;
* p≤.10, ** p≤.05, *** p≤.01

Table 10: Continuation - monthly COVID-19 regressions of RLASSO specification

	(1)	(2)	(3)	(4)
	coef.	coef.	coef.	coef.
MONTH	11/2020	12/2020	01/2021	02/2021
NE-SW	-0.276***	-0.353***	-0.282**	-0.249*
	(0.09)	(0.11)	(0.13)	(0.12)
MIGRANTS	0.083***	0.055***	0.026*	0.019
	(0.01)	(0.01)	(0.01)	(0.01)
SCHOOL	-0.030***	-0.028**	-0.023	-0.021
	(0.01)	(0.01)	(0.01)	(0.01)
POVERTY	0.007	0.020	0.020	0.020
	(0.02)	(0.02)	(0.02)	(0.02)
SUN	0.001	0.001**	0.002***	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)
CONSC	0.253**	0.243	0.197	0.180
	(0.12)	(0.14)	(0.15)	(0.14)
constant	3.474*	3.051	3.836	4.369*
	(1.67)	(2.16)	(2.26)	(2.08)
N	80	80	80	80
r2	0.908	0.793	0.649	0.626

Note: Cluster-robust standard errors in parentheses;
* p≤.10, ** p≤.05, *** p≤.01

Table 10: Continuation - Monthly COVID-19 regressions of RLASSO specification

MONTH	(1) coef. 03/2021	(2) coef. 04/ 2021	(3) coef. 05/2021	(4) coef. 06/2021
NE-SW	0.233*	-0.227*	-0.255**	-0.263**
	(0.12)	(0.12)	(0.12)	(0.12)
MIGRANTS	0.013	0.010	0.009	0.009
	(0.01)	(0.01)	(0.01)	(0.01)
SCHOOL	-0.019	-0.017	-0.016	-0.016
	(0.01)	(0.01)	(0.01)	(0.01)
POVERTY	0.021	0.023	0.026*	0.026*
	(0.02)	(0.02)	(0.01)	(0.01)
SUN	0.001***	0.001**	0.001**	0.001**
	(0.00)	(0.00)	(0.00)	(0.00)
CONSC	0.170	0.153	0.142	0.143
	(0.14)	(0.14)	(0.13)	(0.13)
constant	5.010**	5.578**	5.907**	5.956**
	(2.07)	(2.17)	(2.11)	(2.07)
N	80	80	80	80
R^2	0.573	0.528	0.562	0.574

Note: Cluster-robust standard errors in parentheses;
* p≤.10, ** p≤.05, *** p≤.01

Table 11a: Test for a structural break between waves - unknown break date

(a) between first and second wave
Estimated break date: 48
Ho: No structural break

Test	Statistic	p-value
swald	822.8963	0.0000

Table 11b: Test for a structural break - unknown break date

(b) between second and third wave
Estimated break date: 108
Ho: No structural break

Test	Statistic	p-value
swald	52.1065	0.0000

Table 12: Wave by wave COVID-19 estimates, specifies by RLASSO

WAVE	(1) coef.	(2) coef.	(3) coef.
NE-SW	-0.155**	-0.243***	-0.244***
	(0.07)	(0.04)	(0.01)
MIGRANTS	0.073***	0.061***	0.010***
	(0.01)	(0.00)	(0.00)
SCHOOL	-0.052***	-0.030***	-0.017***
	(0.01)	(0.00)	(0.00)
POVERTY	-0.042***	-0.004	0.024***
	(0.01)	(0.01)	(0.00)
SUN	0.000	0.001***	0.001***
	(0.00)	(0.00)	(0.00)
CONSC	0.348***	0.214***	0.150***
	(0.10)	(0.06)	(0.02)
constant	1.571	4.036***	5.658***
	(1.36)	(0.80)	(0.24)
N	2538	3344	1840
R^2	0.087	0.112	0.447

Note: cluster-robust standard errors in parentheses; * p≤.10, ** p≤.05, *** p≤.01