The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0144-3585.htm

Weather, mobility and the evolution of the Covid-19 pandemic

Corinna Ghirelli
Banco de España, Madrid, Spain
Andrea Gonzalez
Repsol SA, Madrid, Spain, and
Jose Luis Herrera and Samuel Hurtado
Banco de España, Madrid, Spain

Received 18 January 2022
Revised 7 March 2022
Accepted 10 April 2022

Abstract

Purpose – The authors investigate the effect of weather and mobility on the spread of the Covid-19 pandemic.
Design/methodology/approach – The authors first estimate the effective reproduction number (Rt) as a proxy of the spread of the Covid-19 pandemic and then study the relationship between the latter and weather and mobility in a panel data framework. The authors use US daily infections data between February and September of 2020 at the county level.
Findings – The authors find that lower temperatures are associated with a higher Rt, and this effect is greater at temperatures below 0°C. In addition, mobility reductions related to certain types of locations (retail and recreation, transit stations and workplaces) are effective at reducing Rt, but it is an increase in the time spent in parks that most helps reduce the spread of the pandemic.
Originality/value – The estimates imply that a 20°C fall in temperature from summer to winter would increase Rt by +0.35, which can be the difference between a well-controlled evolution and explosive behavior of the spread of the virus. Applying these coefficients estimated with US county data to aggregate series from other countries helps explain the resurgence of the pandemic in the Northern Hemisphere during the winter of 2020. The results show that mobility reduction and social distance are best policies to cope with the Covid-19 outbreak. This strong policy lesson will help facing similar outbreaks in the future.

Keywords Pandemic, Covid-19, Coronavirus, Weather, Mobility, Panel data

Paper type Research paper

1. Introduction

The Covid-19 pandemic is exacting a very heavy cost in terms of human lives, despite the extraordinary measures adopted to contain its spread. It has also caused far-reaching disruption to society and to the global economy throughout 2020. The economic literature has been studying the impact of this pandemic from different angles. Among others, a number of papers analyzed the policies adopted by governments to reduce the Covid-19 propagation (Baccini and Brodeur, 2021) or the effect of the pandemics on household consumption (Baker *et al.*, 2020b; Chetty *et al.*, 2020). Beyond economic effects, other researchers examined the effect of the pandemic in political elections (Baccini *et al.*, 2021; Fernandez-Navia *et al.*, 2021) or the impact of media on mobility during the pandemic (Ananyev *et al.*, 2021). With the pandemic still ongoing, assessing the determinants of its evolution and evaluating the effects of containment measures is of major importance.

To fight the spread of the virus, governments have put forward extremely costly policies, from both a human and economic perspective, such as reduced mobility and social distance.

JEL Classification — C23, I12, I18

The authors thank Angel Gavilán, José González, Enrique Moral, Matias Pacce, Alberto Urtasun and Ernesto Villanueva. The authors also thank all participants at the internal seminar of the Banco de España for their helpful comments. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Banco de España or the Eurosystem.

Journal of Economic Studies

0144-3585
DOI 10.1108/JES-01-2022-0032

The latter is particularly difficult to attain during bad weather (cold or rainy), when people tend to gather in closed spaces. However, despite the experts' recommendations, there is no political and social consensus yet on which have been the best policy practices to tackle the Covid-19 outbreak. Hence, this crisis provides us the opportunity to learn precious policy lessons to face similar outbreaks in the future.

A few economic papers address this research question (Wilson, 2020; Smith *et al.*, 2020) and find that reduced mobility impedes the spread of the virus. However, being this epidemic so recent, most of these studies use small datasets, which limit the ability to carry out complex analyses. We contribute to this literature by using a large database of daily data at the US county level, which allows us to tackle this research question with sound econometric analysis. In addition, we borrow from the epidemiological literature the appropriate measure for the spread of the virus, e.g. the effective reproduction number (Rt), and use it as a main variable of interest in our analysis. This enriches importantly the existing economic literature on Covid-19, which tends to rely on more easily available but imperfect proxies, such as the death toll or the number of contagions.

In this paper, we estimate Rt of the Covid-19 pandemic using US data on the daily number of infections between February and September of 2020 at the county level. This is then used to estimate the effect of weather and mobility variables on the spread of the virus. The panel estimation uses county dummies and dummies for state month interactions so that the estimated effects capture the differential evolution of, say, warmer or cooler counties within a state and in the same month [1].

Focusing on Rt provides two major advantages: first, it allows to study the dynamic effect of time-varying variables, such as temperature and mobility, without worrying about differences in the start of the pandemic across regions; second, it allows to get a more useful quantification of the effects of our covariates, since keeping $Rt < 1$ is the main target for a policymaker trying to keep the spread of the pandemic under control.

We find a strong and significant effect of the weather: lower temperatures are associated with a higher effective Rt. This effect is greater at temperatures below 0°C, and at low temperatures, precipitation is also associated with a higher Rt. These effects are sizable: according to our preferred linear specification, a summer to winter decrease of 20°C would imply an increase in Rt of approximately 0.35, which can be the difference between a well-controlled evolution and explosive behavior of the transmission of the virus.

We also find that mobility reductions related to certain types of locations (retail and recreation, transit stations and workplaces) are effective at reducing Rt. On the other hand, and in line with previous literature about the probability of infection in open and closed spaces, an increase in the time spent in parks also helps reduce the spread of the virus.

The estimates obtained using US data at the county level are also useful for interpreting the evolution observed in aggregate data for other countries. We show that the behavior observed in a wide range of countries is consistent with the estimated role of weather and mobility and can partially explain why it was so much difficult for many Northern Hemisphere countries to control their local epidemic over the winter of 2020, even before the emergence of the new strains. These figures suggest that many countries throughout 2020 converged towards the best practices (social distance and mobility reductions) in terms of control of the pandemic, set by some Asian countries, such as South Korea. All in all, our results provide strong policy lessons that will be useful to tackle similar outbreaks in the future.

The rest of the paper is organized as follows. Section 2 reviews the related literature. In Section 3, we present the data, while Section 4 describes the empirical framework. Results are reported in Section 5, and robustness exercises in Section 6. Finally, Section 7 presents counterfactual exercises for other countries using the estimated coefficients, and Section 8 concludes.

2. Previous studies

A few papers have already studied the effect of weather and mobility on the spread of the Covid-19 pandemic. In the areas where they intersect, our results are in line with what these previous articles have found, but in contrast, we put together in one estimation the following important factors: (1) Rt as the variable of interest, instead of the cumulative number of cases or deaths, (2) data for an area and time period with high heterogeneity in all observed dimensions, (3) a large data panel with very granular temporal and spatial information and (4) both weather and mobility information.

Three of these factors can be observed in Smith *et al.* (2020). We differ from this work in the granularity of the data: they run regressions with data for US states before and after lockdown, while we use daily information at the county level. In addition, they focus on the basic reproduction number (R0) and the initial wave of the pandemic, whereas we use a longer time period extending until the end of August. Apart from the methodological differences, in terms of the estimation results, both studies are in line: mobility is a major factor in determining the speed of the spread of the virus, but temperature also matters, and further reductions in mobility may be needed to slow contagions in the winter.

A second paper that is very relevant to ours is Wilson, which uses basically the same data as we do (Rt, weather and mobility in the USA at the county level) and finds that holding mobility fixed, temperature reduces Covid-19 infections and holding weather fixed, mobility increases infections. The main difference is the variable of interest: Wilson (2020) studies the average of the growth rate of cases over long periods, while we skew the methodology in the direction of what is more common in epidemiology, and focus on Rt. This is in line with the request of Smith *et al.* (2020), who conclude that "the role of environment in transmission has become controversial, in part because of the application of models to case prevalence, rather than fundamental epidemiological parameters such as *R*." One downside of our approach is that it leaves more space for statistical problems in the estimation, which is the reason Wilson (2020) skews his methodology towards the best practices in economics, producing sound estimates that should be more robust to possible issues of identification or endogeneity, at the cost of deviating slightly from the use of "fundamental epidemiological parameters." Nevertheless, the fact that both papers reach similar conclusions is reassuring.

In addition, Kapoor *et al.* (2022) look at the effect of social distancing in the initial spread of the pandemic, with differences in rainfall with respect to the local average as their instrument for social distancing. In terms of the previous discussion on best practices in epidemiology and economics, this paper falls very close to the latter, using instrumental variables that ensure exogeneity of the regressors but also using cumulative numbers of cases and deaths as their variables of interest.

The work of Desmet and Wacziarg (2022) is also closely related to our research. They analyze the correlates of Covid-19 cases and deaths across US counties and find that variables like population density, public transportation, age structure, nursing home residents and connectedness to source countries are important predictors of variation in disease severity.

Other papers study the relationship between mobility and the spread of the virus based on small samples of US data. For instance, Badr *et al.* (2020) use data from 25 US counties that had large Covid-19 outbreaks and find a strong relationship between mobility and the rate of increase in the number of cases. In addition, Chang *et al.* (2021) relies on location data from mobile phones in ten US cities and show that a small minority of "superspreader" points of interest account for a large majority of infections and that restricting maximum occupancy at each point of interest is more effective than uniformly reducing mobility. Similarly, recent papers rely on mobile phone data to quantify mobility and social contact, and hence assess the evolution of the pandemic, showing that out-of-home activity is strongly associated with Covid-19 infections (e.g. Almagro *et al.*, 2020; Chevalier *et al.*, 2022).

Another strand of literature focused on both mobility and environmental factors in relation with the spread of the virus. For instance, Poirier *et al.* (2020) examine the spatial variability of (a proxy to) the basic reproductive numbers of Covid-19 across provinces and cities in China and assess the effect of mobility and of temperature and humidity. Compared to them, we use much more granular data which allow providing a very robust assessment of the effects of our covariates. Similarly, Rosario *et al.* (2020) evaluate the relationship between weather and Covid-19 infection, using data from six cities in the state of Rio de Janeiro, Brazil. Their database only spans 47 days and does not offer much geographical heterogeneity, but they still find significant effects from several weather factors.

3. Data

We construct a large database of US data at the county level from February to September of 2020, combining information from several publicly available sources:

(1) Epidemiological data: We use the daily number of new confirmed Covid-19 cases in the USA at the county level. These data can be accessed through the American Center for Disease Control and Prevention website but is actually elaborated by the USA Facts initiative, which compiles information published by different official local sources.

(2) Effective reproduction number: We estimate Rt following the approach of Cori *et al.* (2013), as implemented in the EpiEstim software package, and Tindale *et al.* (2020) [2]. Following Soucy *et al.* (2020), we use rolling windows of 14 days of the number of confirmed Covid-19 cases (longer than the common window of seven days, in order to smooth out noise from the county data).

(3) Weather data: We focus on the average temperature (degrees Celsius) and precipitation (mm). We use daily observations by weather station, which can be accessed through the National Oceanic and Atmospheric Administration (NOAA)'s National Centers for Environmental Information and the National Weather Service website. The data at the weather station level are assigned to every county by combining the measurements from the ten stations that are closest to the coordinates of the center of the county, with a weighted average where the weight is the inverse of that distance. This ensures that counties where there is no weather station get a measurement based on several nearby stations and that stations close to the center of the county receive a disproportionately high weight in the average.

(4) Mobility data: We collect these data from Google Global Mobility Reports, which is a database that describes how people move in their everyday life during the pandemic. Google elaborates these data using information from mobile phones collected through their applications, such as Google Maps, which allows the company to record changes in the locations of phones anonymously. It is defined as the change with respect to baseline days before the outbreak of the pandemic (the median between January 3rd and February 6th). The information is sorted into different location categories: grocery stores and pharmacy, retail and recreation, transit stations, parks, workspaces and homes; in our estimations, we used the indicator for parks and, an average of the indicators for retail and recreation, transit stations and workplaces, as proxy of general mobility.

(5) Literacy and numeracy indicators: they are provided by Program for the International Assessment of Adult Competencies (PIAAC), which can be accessed through the National Center for Education Statistics. The PIAAC is a survey that collects data about adults' performance in literacy, numeracy and problem-solving

and allows sorting the population into three main categories: people with a low (P1), intermediate (P2) or high (P3) information-processing level. We use the share of the population that only attains low scores (P1), and we use the average of the shares for literacy and numeracy.

(6) Density: inhabitants per square kilometer, as found in the US census.

(7) Services: share of employment in the services sector in 2014–2018, as found in the US census.

To model infections, we make the following timing assumptions: (1) temperature and mobility today affect contagions today (they affect how many contacts there are and how many of them end up infected), and (2) contagions at t become identified cases at $t + 7$ (the lag would probably be slightly shorter for cases found through contact tracing and slightly longer for the rest). Since Rt is estimated with a 14-day window of case data, we apply the analogous transformation to temperature and mobility data, calculating a 14-day moving average that then enters the estimations with a lag of one week. Therefore, average temperatures and mobility between t and $t + 13$ affect the Rt estimated with cases data from $t + 7$ to $t + 20$.

Since data are only available at the county level, most of the observations refer to relatively large population groups (Los Angeles county is home to almost ten million people), whereas many counties with a small population have to be left out of the estimation because Google has not published mobility data due to privacy concerns or because they have too few cases and Rt cannot be estimated; this is very common in counties with small populations (almost no counties with fewer than 32,000 people have all the necessary data). Figure 1 represents, first, the histogram of counties by population for all counties and for those that could be used in the estimation because of data availability, and second, the evolution of the number of counties used in the estimation across time (some counties disappear from the sample, typically because Google did not publish mobility data for them during the summer because of increased privacy concerns or because the number of cases became too small to provide an estimate of Rt). The estimation sample ends on the 31st of August of 2020 because Google drastically reduced the coverage of the mobility indicators for the second half of August, providing data only for a much smaller number of counties. Since this indicator enters the estimation as a two-week average lagged by one week, the estimation could still be extended to the end of the month.

Figure 2 represents the counties that appear at least once in the estimation. They are mainly populous areas that have had significant Covid-19 outbreaks. They constitute approximately 18% of counties but represent 49% of the population of the US.

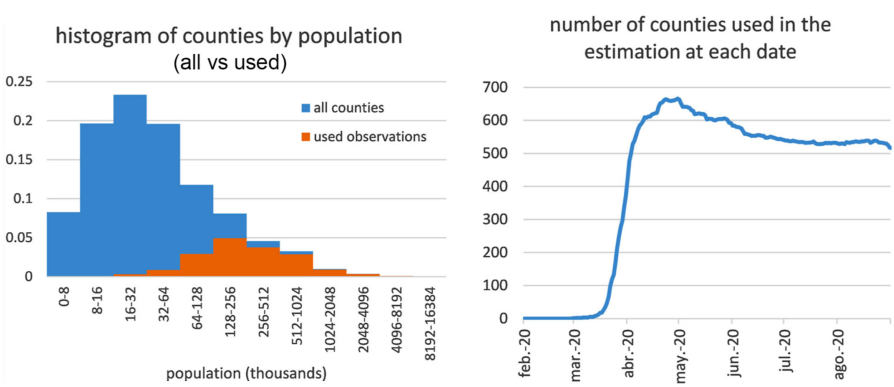

Figure 1.
Counties used in the estimation

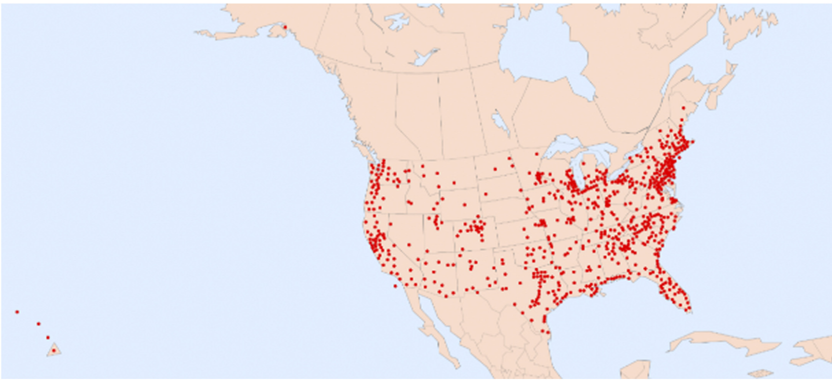

Figure 2.
Counties used in the estimation

As implied by Figure 2, our analysis exploits high variability in the weather data. This is also the case for the epidemiological data, as shown in Figure 3, which presents all the values of Rt used in the analysis, compared with the national average. This diversity of observations helps identify the effect of weather on the evolution of the pandemic without having to rely on broad North vs South or spring vs summer comparisons, where the comparability of the situations would be more questionable.

4. Econometric framework

We rely on a typical fixed effects model, specified as follows:

$$R_t^c = a^c + b_1 TEM_{t-7}^c + b_2 PRE_{t-7}^c + b_3 MOB_{t-7}^c + b_4 PAR_{t-7}^c + \sum_i m_i + u_t^c \qquad (1)$$

where

(1) Superscript "c" refers to counties, and subscript "t" refers to days so that the units of observation vary at the county and day level.

(2) The dependent variable is Rt.

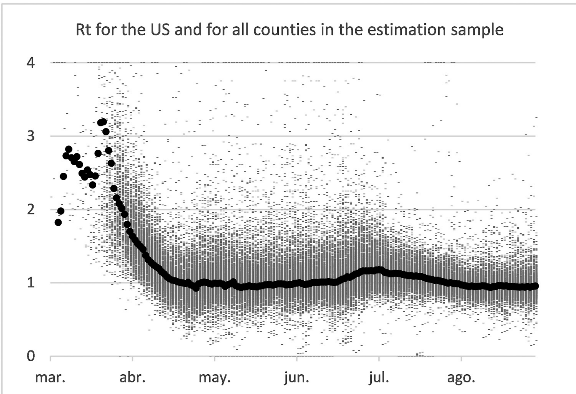

Figure 3.
Rt for all the US counties used in the estimation

(3) The main regressors are temperature (TEM), precipitation (PRE), average mobility towards locations in retail and recreation, transit stations and workplaces (MOB) and mobility in parks (PAR). These are all time-varying. All regressors are lagged by one week to reduce endogeneity and correctly incorporate the lag according to which contacts become infections and then cases.

(4) m_i are monthly fixed effects that we add in the equation to control for all unobservable common actors affecting all counties over time.

(5) a_c are county fixed effects that account for all county-specific characteristics that are constant.

(6) u_t^c is the error term.

Standard errors are clustered at the county level to account for the fact that the observations are correlated over time. Notably, the presence of serial correlation is due to our construction procedure, since all values of variables are 14-day moving averages (because by definition Rt can be estimated considering a lengthy period of time). Clustering at the county level allows to tackle this problem.

Including county fixed effects allows us to control for all county-specific characteristics that are constant over time. To deal with this, we implement the fixed effect estimator (within estimator).

In addition, in our preferred specification we add to Equation (1) interactions between time (m) and state (s) fixed effects ($\sum_i \sum_j m_i s_j$), which take into account all unobservable factors that change over time at the state level, such as the regional economic situation and state-level policies implemented to cope with the spread of the virus. The inclusion of these fixed effects ensures that the estimated coefficients reflect changes that are not correlated with major changes in circumstances or behavior. For example, the estimated effect of temperature is not the result of comparing the summer with the spring (when schools were open and people did not have a lot of information about the pandemic) or comparing Arizona with Alaska; instead, it captures the much subtler fact that counties with higher temperatures than other counties in the same state and in the same month observed a more negative deviation of their Rt with respect to the usual level for that county and with respect to the evolution that was common to the whole state in that month.

The issue of possible reverse causality or variable endogeneity also needs to be discussed: if mobility responds to Rt, our estimates could be biased. It is certain that when the situation worsens and the news is full of reports about the carnage of the pandemic, people reduce their visits to restaurants and shops and work from home if possible. However, we expect that mobility today will respond to cases and deaths today, and this does not depend on Rt today, but it does depend on long-past values of Rt. In terms of our estimation, there would be an endogeneity problem if mobility today responded to the Rt calculated with the subsequent weeks' data on identified Covid-19 cases. This is not completely implausible because Rt can have strong autocorrelation.

To formally test this, we run a panel version of the Granger causality test, using the daily data about Rt and mobility at the county level, allowing up to 28 lags of both variables. The test is run for the unbalanced panel of observations used and includes county fixed effects as in Holtz-Eakin *et al.* (1988):

$$MOB_{t-7}^{c} = a^{c} + \sum_{k=1}^{28} b_1^k MOB_{t-7}^{c} + \sum_{k=1}^{28} b_2^k R_{t-k}^{c} + u_t^c \qquad (2)$$

We find that no b_2^k is statistically different from zero for any k between 1 and 28, meaning that Rt does not Granger-cause mobility in our county-level database. Thus, we can dismiss reverse causality or covariate endogeneity issues in our estimation.

5. Estimation results

5.1 Linear specifications

Table 1 reports the estimation results of our baseline model. We move from a less demanding specification (Column 1), in which we use state fixed effects instead of county fixed effects, to the more demanding specification (Column 4), in which we control for county fixed effects (our units of observations) and for interactions between time and state fixed effects.

Overall, the estimated coefficients show the expected sign: temperature and mobility in parks are negatively associated with Rt, while higher precipitation and higher mobility towards recreational locations, transit stations and work-related locations increase the spread of the virus. Results are robust across specifications, although there are some interesting differences. The effect of temperature is larger whenever the estimation controls for state-level policies against the spread of Covid-19 (Column 2 and 4), which suggests that part of the effect of temperature would be biased by the omission of state policy responses to Covid-19 in Columns 1 and 3, if the latter were implemented in periods of high temperature. The opposite holds for rainfall: the effect of precipitation is reduced when the state-by-month interacted fixed effects are introduced. The effects of mobility variables are stable across all specifications.

In the rest of this subsection, we discuss the estimated coefficients of the specification reported in Column 4, the most demanding one, which is our preferred linear specification. The estimated effect for temperature Rt is sizeable. Since our estimation assumes linearity, a summer to winter decrease of 20 °C (the national average difference between July and January) would imply an increase in Rt of approximately 0.35, which can be the difference between well-controlled evolution and explosive behavior of the spread of the virus. In contrast, the estimate for precipitation shows the expected sign, but the coefficient is small and not significant once we control for interactions between state and month fixed effects.

The first mobility variable measures the number of visits to three specific categories of locations (retail and recreation, transit stations and workplaces), and the estimate for its coefficient is positive. To contextualize this figure, for one week in April, mobility decreased

Variables	(1) Rt	(2) Rt	(3) Rt	(4) Rt
Temperature	−0.0047*** (0.0016)	−0.0154*** (0.0016)	−0.0052*** (0.0017)	−0.0173*** (0.0017)
Precipitation	0.0004*** (0.0001)	0.0001 (0.0001)	0.0004*** (0.0001)	0.0001 (0.0001)
Mobility	0.0145*** (0.0006)	0.0138*** (0.0006)	0.0167*** (0.0007)	0.0169*** (0.0008)
Mobility parks	−0.0012*** (0.0002)	−0.0009*** (0.0001)	−0.0014*** (0.0002)	−0.0012*** (0.0002)
Fixed effects	State dummies and month dummies	State dummies interacted with month dummies	County dummies and month dummies	County dummies interacted with month dummies
Observations	89,669	89,669	89,669	89,669
R-squared	0.302	0.365	0.3028	0.3664
Cluster	County	County	County	County

Note(s): Standard errors are cluster-robust in parentheses and are clustered at country level. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table 1.
Main specifications

to around 49% in the USA as a whole, which is the lowest point in the estimation sample. Our estimated coefficient implies that this led to a decrease in Rt of 0.83 units.

On the other hand, mobility in parks is negatively associated with the rate of contagion. The magnitude of the effect is much smaller than for the measure of general mobility, but the effect is still highly statistically significant and quantitatively sizable: a reduction of the mobility indicator regarding parks from its level over the summer (around +60%) back to zero would be associated with an increase in Rt of approximately 0.07.

Overall, these results are in line with what previous studies have found about the virus and its transmission. For instance, it has been shown that the SARS-CoV-2 virus that causes Covid-19 survives much longer at low temperatures (e.g. Chin and Poon, 2020). On the other hand, many studies, and most notably Shen *et al.* (2020), found that the transmission of the virus is much more likely indoors than outdoors. Our results show that these results from laboratory experiments and contact-tracing studies can also translate to the evolution of the epidemic at the macro-level.

5.2 Nonlinear specifications

Adding some interactions between variables to the specification in Column 4 of Table 1 allows us to assess some possible nonlinear effects. Table 2 summarizes these results. All specifications retain the general characteristics of the specification in Column 4 of Table 1: county fixed effects, state month interacted fixed effects and clustering by county.

Column 1 of Table 2 adds an interaction between temperature and precipitation, which allows us to decompose the nonstatistically significant coefficient of precipitation in Column 4 of Table 1 into two statistically significant ones: the effect of precipitation on Rt is approximately null at high temperatures (when the average is around 25 °C) but is negative at lower temperatures. Additionally, a nonlinearity for the effect of temperature is allowed at freezing point, and this shows that the spread of the virus becomes particularly fast when average temperatures fall below 0 °C.

Column 2 of Table 2 introduces an interaction between temperature and mobility and reveals that the effect of mobility is reduced in warm weather, both for parks and for general mobility. The implications of these nonlinearities are complex: in winter, a given reduction in mobility generates a greater fall in Rt, but a relaxation of mobility restrictions also has larger negative effects. If we move from a summer situation with an average temperature of 25 °C, general mobility of 25% and park mobility of 60% to a winter situation with an average temperature of 5 °C and the same 25% and +60% mobility, this estimation indicates an increase in Rt of +0.33, which is very similar to the +0.35 that the linear estimation would imply. With this change in temperature, the effective coefficient for general mobility goes from 0.013 to 0.023. This makes it easier to compensate, via additional reductions in mobility, for the effect of colder weather, but also increases the repercussions of any relaxation in mobility during winter.

Column 3 of Table 2 adds interactions with characteristics of the counties and finds that the effect of mobility is enhanced in counties with high population density, where a larger share of the population attained the lowest scores (P1) in the adult numeracy and literacy PIAAC tests and where the service sector represents a lower percentage of employment (possibly because firms in the industry sector have more workers in larger enclosed spaces).

All of the coefficients presented in Column 1–3 of Table 2 are statistically significant, including the ones that identify nonlinear effects.

Finally, Column 4 of Table 2 puts all of these elements together and checks that the signs of the estimated coefficients are robust to the simultaneous inclusion of all covariates; the quantification does change in some cases, however, and a few become statistically non-significant.

Table 2. Nonlinear specifications

Variables	(1) Rt	(2) Rt	(3) Rt	(4) Rt
Temperature	−0.0158*** (0.0018)	−0.0327*** (0.0023)	−0.0167*** (0.0017)	−0.0336*** (0.0024)
Temp. negat. (< 0C)	−0.0210** (0.0103)			−0.0127 (0.0100)
Precipitation	0.0010*** (0.0004)	0.0007* (0.0004)	0.0010*** (0.0004)	0.0008** (0.0004)
Temp × Precipitation	−0.00004** (0.0000)	−0.00003* (0.0000)	−0.00004*** (0.0000)	−0.00004** (0.0000)
Mobility	0.0168*** (0.0008)	0.0251*** (0.0009)	0.0210*** (0.0017)	0.0260*** (0.0018)
Mobility × Temp		−0.0005*** (0.0000)		−0.0005*** (0.0000)
Mobility parks	−0.0011*** (0.0002)	−0.0028*** (0.0004)	−0.0010*** (0.0001)	−0.0028*** (0.0004)
Mob. parks × Temp		0.0001*** (0.0000)		0.0001*** (0.0000)
Mobility × Density			0.1820* (0.0000)	0.0619 (0.0000)
Mobility × PIAAC			0.0153*** (0.0052)	0.0249*** (0.0048)
Mobility × Services			−0.0004*** (0.0001)	−0.0003*** (0.0001)
Fixed effects	County dummies and state dummies interacted with month dummies	County dummies and state dummies interacted with month dummies	County dummies and state dummies interacted with month dummies	County dummies and state dummies interacted with month dummies
Observations	89,669	89,669	89,669	89,669
R-squared	0.3668	0.3729	0.3681	0.3750
Cluster	County	County	County	County

Note(s): Standard errors are cluster-robust in parentheses and are clustered at country level. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

6. Robustness

First, we reestimate our preferred specification (Column 4 in Table 1, which is also reported in Column 1 of Table 3 to ease comparisons) by replacing monthly dummies with dummies of two-week periods. This allows us to control for state-level policy changes in a more fine-grained way, accounting for policies that may have been implemented within the month (see Column 2 of Table 3). Overall, results are consistent with our main specification, although the size of the effect of temperature and mobility in parks is reduced. However, this is expected since we are now estimating twice as many time fixed effects.

Second, we check that our results are not affected by the serial correlation induced by the specific construction of our variables. This is because Rt is estimated by construction using a window and refers to a lengthy period of time, and for consistency, we apply an analogous procedure to all time-varying regressors on the right-hand side of the equation, calculating 14-day moving averages [3]. This use of moving averages generates serial correlation in our sample since the values of two consecutive days are based on common information. To control for this, all estimations in Table 1 allow for serial correlation of the residuals by means of clustering at the county level. Still, one may worry that our results are affected by our construction procedure. Therefore, we estimate a version of our preferred specification that now uses only one set of observations every two weeks so that each observation corresponds to a two-week average. Results are reported in Column 3 and 4 of Table 3, where time fixed effects are defined each month and each two-week period, respectively. Overall, results are robust, although the coefficients are slightly smaller in magnitude and less significant. This is reasonable since the number of observations drops considerably and the specification is very demanding.

Finally, we check that extending the sample period beyond the period of low data availability in September does not alter the results. To do that, we show that the estimates of our preferred specification are qualitatively similar when extending the sample until the end of October [4].

7. Decomposition and counterfactual scenarios

In order to illustrate the quantitative effect associated with each explanatory variable, Figure 4 presents a decomposition of the nationwide evolution of Rt. For most of the variables, the aggregate national data are used, but temperature and rainfall are population-weighted averages of the county data. The size of each contribution is calculated using the coefficients estimated in Column 4 of Table 1.

According to these estimates, the increase of Rt above one in the second wave of infections, at the end of June, was mainly caused by an increase in mobility, and the higher temperatures of the summer helped to contain it. The third wave that started in October, on the other hand, happened without an increase in general mobility and was instead associated with the fall in temperatures (and also to some extent to a reduction in visits to parks). The containment of this third wave was based on both lower mobility and the continuous reduction of the grey contribution labeled “others”, which captures all improvements in the fight against the pandemic since its onset (larger testing capacity, better contact tracing, more widespread use of masks, stronger emphasis on ventilation and any other factors not related to either weather or mobility). This residual contribution has continued its reduction in 2021, thanks to the effect of the vaccines, which, together with the moderate temperatures of spring, has made $Rt < 1$ possible even with much smaller reductions in mobility, even after the arrival of more transmissible strains that would otherwise have pushed this residual upwards.

Figure 5 presents a counterfactual scenario where we assess, according to the estimates from Column 4 of Table 1, the evolution of Rt that would have been observed if temperatures had remained constant at the levels of January. The difference implied by this exercise is very

Table 3.
Robustness: frequency and time fixed effects

Variables	(1) Sample frequency: daily (each obs. is a 14-day moving average) Rt	(2) Rt	(3) Sample frequency: two weeks (simple average) Rt	(4) Rt
Temperature	−0.0173*** (0.0017)	−0.00528*** (0.0019)	−0.00644** (0.00295)	−0.00662* (0.00383)
Precipitation	0.0001 (0.0001)	0.00005 (0.00009)	−0.00012 (0.00014)	0.00010 (0.00018)
Mobility	0.0169*** (0.0008)	0.01216*** (0.00076)	0.00649*** (0.00138)	0.00335*** (0.00119)
Mobility parks	−0.0012*** (0.0002)	−0.00078*** (0.00015)	−0.00050*** (0.00017)	−0.00033* (0.00017)
Fixed effects	County dummies and state dummies interacted with month dummies	County dummies and state dummies interacted with month dummies	County dummies and state dummies interacted with month dummies	County dummies and state dummies interacted with month dummies
Observations	89,669	89,669	6,612	6,612
R-squared	0.36644	0.43024	0.43232	0.51951
Cluster	County	County	County	County

Note(s): Standard errors are cluster-robust in parentheses and are clustered at country level. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

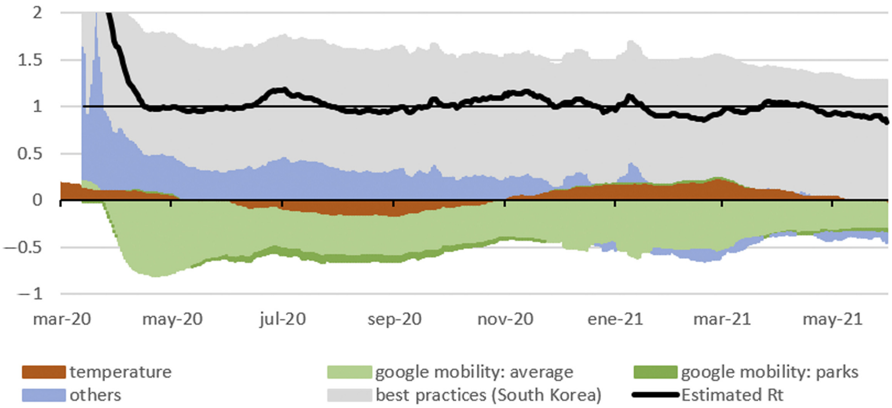

Figure 4.
Contributions to the evolution of Rt in the USA

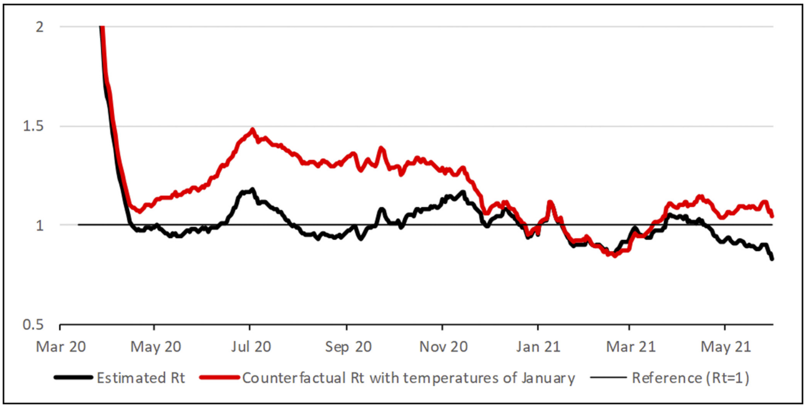

Figure 5.
Counterfactual scenario for Rt in the USA with constant January temperature

large: with an average of 1.43 in July instead of the observed 1.09, the second wave of Covid-19 in the USA would have implied a doubling of the number of deaths approximately every 9 days instead of every 37 days. The exercise also shows that, in 2021, vaccination alone would not have been enough to allow the fast normalization of mobility that has been observed: with the colder temperatures of winter and the still incomplete vaccination process, more substantial reductions in mobility would still have been needed in order to keep the virus contained.

The coefficients estimated with county data from the USA may not accurately capture the characteristics of the spread of the virus in other countries with different institutions and characteristics, but, nonetheless, a decomposition exercise using these coefficients provides an interesting narrative of the evolution of the local epidemic in different countries.

Figure 6 presents this exercise, with three differences with respect to Figure 4: (1) we display the results at monthly frequency to abstract from short-term events; (2) the residual is divided in two parts: one that is equal to the average residual observed in South Korea between April 2020 and April 2021 and which is designated as a reference of "best practices" in the fight against the spread of the virus, and the rest which identifies the distance from what is observed in each country and that South Korean reference residual; (3) for the first months of 2021, the effect of vaccination is subtracted from the residuals [5].

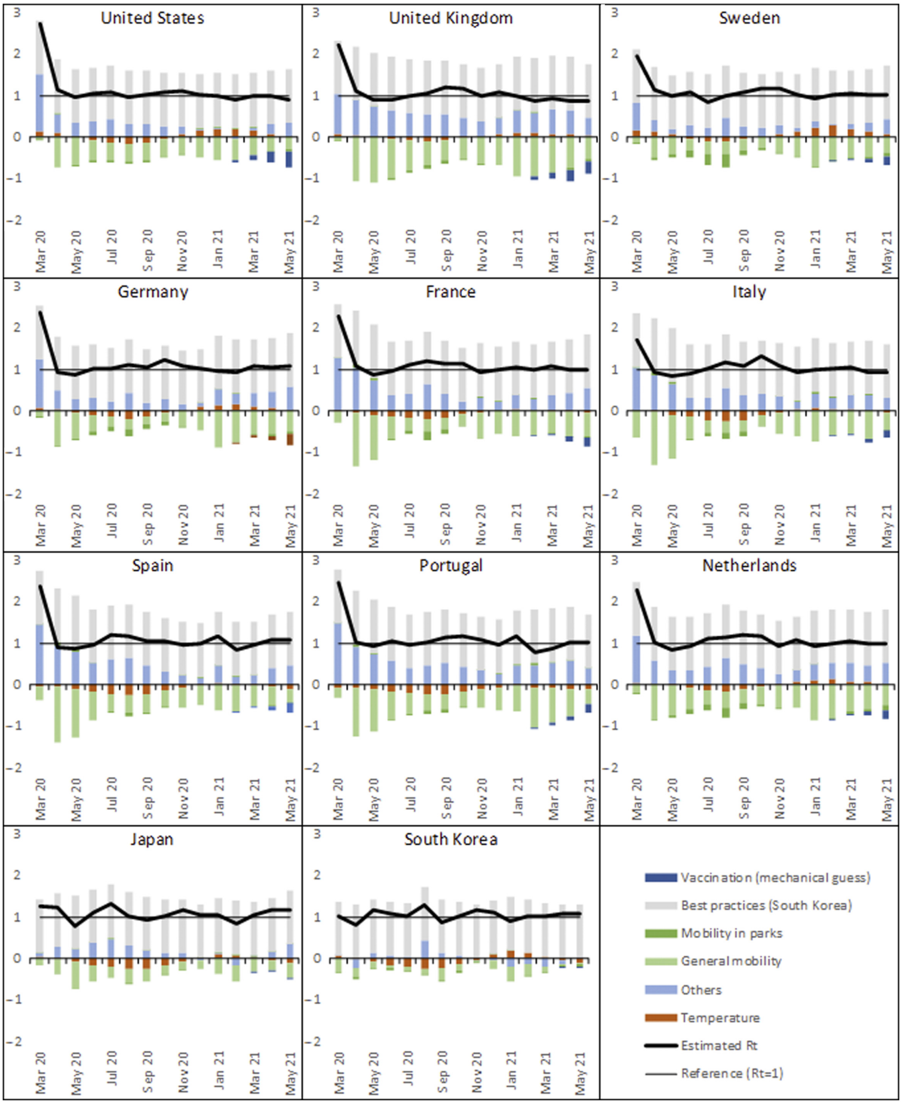

Figure 6.
Contributions to the evolution of Rt in other countries

The decomposition shows three main facts for all countries: (1) residuals were larger than those of the South Korean "best practices" reference level until the last months of 2020, which suggests that general efforts to stop the spread of the virus (other than through reductions in mobility) were not as effective, but steadily improved throughout 2020; (2) the larger residuals in the most recent months could be explained by the emergence of new strains; (3) vaccination has had a very significant role in the reduction of Rt in the spring of 2021. The model can also explain the difficulties observed in Japan and South Korea since November 2020, as bigger reductions in mobility were needed to counteract the effect of the lower temperatures of winter.

8. Conclusions

We estimate the effective reproduction number (Rt) of the Covid-19 pandemic using US daily infections data between February and September of 2020 at the county level. This is then used to estimate the effect of weather and mobility on the spread of the pandemic. We find a strong and significant effect of the weather: lower temperatures are associated with a higher Rt, and this effect is greater at temperatures below 0 °C. At low temperatures, precipitation is also associated with a higher Rt. We also find that mobility reductions related to certain types of locations (retail and recreation, transit stations and workplaces) are effective at reducing Rt, but it is an increase in the time spent in parks that most helps reduce the spread of the pandemic. Quantitatively, our estimates imply that a 20 °C fall in temperature from summer to winter would increase Rt by +0.35, which can be the difference between a well-controlled evolution and explosive behavior. Applying these coefficients estimated with US county data to aggregate series from other countries helps explain the resurgence of the pandemic in the Northern Hemisphere during the winter, in terms of the contributions from weather and mobility.

Moreover, the estimates obtained using US data at the county level are useful to analyze the evolution of the pandemic observed in aggregate data for other countries. The model is able to explain the tenser situation experienced by Japan and South Korea since November 2020, as bigger reductions in mobility were needed to counteract the effect of the lower temperatures of winter. Moreover, it shows the improvement in the fight against contagion in many countries as they converged towards the international best practices. Additionally, the emergence of more-transmissible strains would be captured by the increased residuals that are visible in 2021.

All in all, our results show that many countries throughout 2020 converged towards the best practices (social distance and mobility reductions) in terms of control of the pandemic set by some Asian countries, such as South Korea. This provides strong policy lessons that will be useful to tackle similar outbreaks in the future.

Notes

1. Similarly, Goswami *et al.* (2020) uses panel data methods to examine which factors affect the transmission of Covid-19 in a panel of several countries.
2. When an outbreak is discovered, estimates for Rt *c* are capped at 4, which is a conservative estimate of the basic reproduction number (R0).
3. Instead of the seven-day window that is common in the literature, we extend it to 14 in order to avoid noise when using county data about Covid-19 cases.
4. Results are not shown for a matter of space and are available upon request.
5. We assume that the vaccine efficacy is 70% (except for the United Kingdom [UK] where it is set at 50% due to large share of Astrazeneca vaccines and B.1.1.7 strain). This is in line with what found in the literature, e.g. Richterman *et al.* (2021).

References

Almagro, M., Coven, J., Gupta, A. and Orane-Hutchinson, A. (2020), "Racial disparities in frontline workers and housing crowding during COVID-19: evidence from geolocation data", Opportunity and Inclusive Growth Institute Working Papers, Federal Reserve Bank of Minneapolis, Vol. 37.

Ananyev, M., Poyker, M. and Tian, Y. (2021), "The safest time to fly: pandemic response in the era of fox news", *Journal of Population Economics*, Vol. 34 No. 3, pp. 775-802.

Baccini, L. and Brodeur, A. (2021), "Explaining governors response to the COVID-19 pandemic in the United States", *American Politics Research*, Vol. 49 No. 2, pp. 215-220, doi: 10.1177/1532673X20973453, available at: https://journals.sagepub.com/description/apr.

Baccini, L., Brodeur, A. and Weymouth, S. (2021), "The COVID-19 pandemic and the 2020 US presidential election", *Journal of Population Economics*, Vol. 34 No. 2, pp. 739-767.

Badr, H.S., Du, H., Marshall, M., Dong, E., Squire, M.M. and Gardner, L.M. (2020), "Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study", *The Lancet Infectious Diseases*, Vol. 20 No. 11, pp. 1247-1254.

Baker, S.R., Farrokhnia, R.A., Meyer, S., Pagel, M. and Yannelis, C. (2020), "How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic", *The Review of Asset Pricing Studies*, Vol. 10 No. 4, pp. 834-862.

Chang, S., Pierson, E., Koh, P.W., Gerardin, J., Redbird, B., Grusky, D. and Leskovec, J. (2021), "Mobility network models of COVID-19 explain inequities and inform reopening", *Nature*, Vol. 589 No. 7840, pp. 82-87.

Chetty, R., Friedman, J., Hendren, N., Stepner, M. and the Opportunity Insight Team (2020), "How did covid-19 and stabilization policies affect spending and employment? A new real-time economic tracker based on private sector data", NBER Working Paper No. 27431.

Chevalier, J.A., Schwartz, J.L., Su, Y. and Williams, K.R. (2022), "JUE insight: distributional impacts of retail vaccine availability", *Journal of Urban Economics*, Vol. 127, 103382, doi: 10.1016/j.jue.2021.103382.

Chin, A.W. and Poon, L.L. (2020), "Stability of SARS-CoV-2 in different environmental conditions–authors' reply", *The Lancet Microbe*, Vol. 1 No. 4, p. e146.

Cori, A., Ferguson, N.M., Fraser, C. and Cauchemez, S. (2013), "A new framework and software to estimate time-varying reproduction numbers during epidemics", *American Journal of Epidemiology*, Vol. 178 No. 9, pp. 1505-1512.

Desmet, K. and Wacziarg, R. (2022), "JUE insight: understanding spatial variation in COVID-19 across the United States", *Journal of Urban Economics*, Vol. 127, 103332.

Fernandez-Navia, T., Polo-Muro, E. and Tercero-Lucas, D. (2021), "Too afraid to vote? The effects of COVID-19 on voting behaviour", *European Journal of Political Economy*, Vol. 69, 102012.

Goswami, G.G., Ali, A.M. and Islam, S. (2020), "A panel path analysis approach to the determinants of COVID-19 transmission: does testing matter for confirmed cases?", *Journal of Economic Studies*, Vol. 48 No. 8, pp. 1624-1648.

Holtz-Eakin, D., Newey, W. and Rosen, H.S. (1988), "Estimating vector auto regressions with panel data", *Econometrica: Journal of the Econometric Society*, Vol. 56 No. 6, pp. 1371-1395.

Kapoor, R., Rho, H., Sangha, K., Sharma, B., Shenoy, A. and Xu, G. (2022), "God is in the rain: the impact of rainfall-induced early social distancing on covid-19 outbreaks", *Journal of Health Economics*, Vol. 81, 102575, ISSN 0167-6296, doi: 10.1016/j.jhealeco.2021.102575.

Poirier, C., Luo, W., Majumder, M.S., Liu, D., Mandl, K.D., Mooring, T.A. and Santillana, M. (2020), "The role of environmental factors on transmission rates of the COVID-19 outbreak: an initial assessment in two spatial scales", *Scientific Reports*, Vol. 10 No. 1, pp. 1-11.

Richterman, A., Meyerowitz, E.A. and Cevik, M. (2021), "Indirect protection by reducing transmission: ending the pandemic with SARS-CoV-2 vaccination", in *Open Forum Infectious Diseases*.

Rosario, D.K., Mutz, Y.S., Bernardes, P.C. and Conte-Junior, C.A. (2020), "Relationship between COVID-19 and weather: case study in a tropical country", *International Journal of Hygiene and Environmental Health*, Vol. 229, 113587.

Shen, Y., Li, C., Dong, H., Wang, Z., Martinez, L., Sun, Z., Handel, A., Chen, Z., Chen, E., Ebell, M.H. and Wang, F. (2020), "Community outbreak investigation of SARS-CoV-2 transmission among bus riders in eastern China", *JAMA Internal Medicine*, Vol. 180 No. 12, pp. 1665-1671.

Smith, T.P., Flaxman, S., Gallinat, A.S., Kinosian, S.P., Stemkovski, M., Unwin, H.J.T., Watson, O.J., Whittaker, C., Cattarino, L., Dorigatti, I. and Tristem, M. (2020), "Environment influences SARS-CoV-2 transmission in the absence of non-pharmaceutical interventions", medRxiv 2020.09.12.20193250, doi: 10.1101/2020.09.12.20193250 2020.09.12.20193250.

Soucy, J.-P.R., Sturrock, S.L., Berry, I., Daneman, N., MacFadden, D.R. and Brown, K.A. (2020), "Estimating the effect of physical distancing on the COVID-19 pandemic using an Urban Mobility Index", medRxiv 2020.04.05.20054288, doi: 10.1101/2020.04.05.20054288.

Tindale, L.C., Stockdale, J.E., Coombe, M., Garlock, E.S., Lau, W.Y.V., Saraswat, M., Zhang, L., Chen, D., Wallinga, J. and Colijn, C. (2020), "Evidence for transmission of COVID-19 prior to symptom onset", *Elife*, Vol. 9, e57149.

Wilson, D.J. (2020), "Weather, social distancing, and the spread of COVID-19", medRxiv 2020.07.23.20160911, doi: 10.1101/2020.07.23.20160911.

Corresponding author
Corinna Ghirelli can be contacted at: corinna.ghirelli@bde.es