Impact of socioeconomic determinants on the speed of epidemic diseases. a comparative analysis

Gilles Dufrénot✉[1,2], Ewen Gallic[1], Pierre Michel[1], Norgile Midopkè Bonou[1], Ségui Gnaba[1], and Iness Slaoui[3]

[1]Aix Marseille Univ, CNRS, AMSE, Marseille, France
[2]CEPII and Institut Louis Bachelier
[3]Paris School of Economics

Oxford Economic Papers, 2024, 1–19,
https://doi.org/10.1093/oep/gpae003

Abstract

We study the impact of socioeconomic factors on two key parameters of epidemic dynamics. Specifically, we investigate a parameter capturing the rate of deceleration at the very start of an epidemic, and a parameter that reflects the pre-peak and post-peak dynamics at the turning point of an epidemic like COVID-19. We find two important results. The policies to fight COVID-19 (such as social distancing and containment) have been effective in reducing the overall number of new infections, because they influence not only the epidemic peaks, but also the speed of spread of the disease in its early stages. The second important result of our research concerns the role of healthcare infrastructure. They are just as effective as anti-COVID policies, not only in preventing an epidemic from spreading too quickly at the outset, but also in creating the desired dynamic around peaks: slow spreading, then rapid disappearance.

JEL Classifications: C23, I12, I18

The project leading to this publication has received funding from the French government under the 'France 2030' investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University—A*MIDEX.

✉Corresponding author: gilles.dufrenot@univ-amu.fr. Aix-Marseille School of Economics, 5 Bd Maurice Bourdet CS 50498 – 13205 Marseille Cedex 1 – France.

1 Introduction

The beginning of 2020 witnessed the emergence of COVID-19 worldwide. For almost a year, while awaiting the discovery of an effective vaccine (the vaccination campaign began globally in 2021), various governments adopted restrictive measures of varying severity to try to contain the epidemic. The availability of open data on the number of confirmed cases and deaths has made it possible to study the evolution of this pandemic from the outset. After a year and a half, the COVID-19 pandemic was responsible for over 180 million confirmed cases and more than 4 million deaths according to the WHO. During this period of epidemic growth, the observed dynamics can be highly heterogeneous between different countries around the world. This source of heterogeneity is multidimensional and linked on the one hand to the response of governments in crisis situations through the implementation of restrictive policies. However, other factors also have a more intrinsic influence, such as different healthcare systems, the health status of populations, different natural environments, and socioeconomic conditions.

The aim of this article is to understand how some factors related to the socioeconomic environment of individuals are likely to accentuate or mitigate the different epidemic phases of viral diseases. We are interested in the SARS-COV-2 (COVID-19) pandemic of 2020/2021. International observations suggest geographic heterogeneity in the impact of social distancing and mitigation measures. It is conceivable that beyond the microbiological and physical characteristics of the viruses, contagion depends on other factors related to the living conditions of the populations, *e.g.*, densities in urban areas, comorbidities related to metabolic diseases, preventive measures related to hygiene, etc. These factors can reduce (or amplify) the effectiveness of barrier measures imposed by health authorities.

Our study lies at the intersection of two literatures. The first is epidemiological and is related to the modelling of viral disease epidemic waves. The second is related to health economics and human behaviour variables in the epidemic contagion phenomena. We study the impact of socioeconomic variables on the estimated parameters of epidemic curves that provide information on the dynamics of epidemics: the growth in the number of cases during the beginning and restart phases of epidemics, and the speed of deceleration in transmission rate near the peak or after the peak. We consider the statistical uncertainty in the estimates by computing the confidence intervals.

We conduct a comparative study, distinguishing between countries according to their level of development (industrialised countries versus emerging and developing countries) and accounting for the heterogeneity of socioeconomic behaviours from one geographical area to another: Asia, Latin and Central America, sub-Saharan Africa, and MENA.

To measure the heterogeneity between countries in the spread of the epidemic and the influence of socioeconomic factors, we wish to introduce differentiated

effects that take into account the different phases of the epidemic. This requires first obtaining measures of epidemic dynamics that can then be introduced into a parametric regression model, making it possible to explain the evolution of these epidemic measures as a function of different socioeconomic factors. Our methodology is therefore based on a two-stage approach: 1) constructing indicators of epidemic dynamics during epidemic growth, and 2) assessing the impact of socioeconomic factors on epidemic dynamics.

In a first step, we estimate the parameters of a generalised phenomenological Richards model using the nonlinear least squares (NLS) method. In our estimations, the two key parameters are the rate of expansion of the infectious disease during the ascending phases of the pandemic and the rate of deceleration or acceleration with respect to a sigmoid function (which gives an idea of the severity of rebounds, for example, during episodes of epidemic recovery, or on the speed of disappearance of a disease). In a second step, we explain heterogeneous epidemic parameters based on socioeconomic variables. These variables are summarised by factors (through principal component analysis) capturing several dimensions of social and economic environments: health (*e.g.*, prevalence of comorbidities, access to sanitation, and nutrition), poverty and precariousness (*e.g.*, multidimensional poverty indicators, informal employment, and variables of inequality) demographic characteristics (*e.g.*, urban population density, median age, and death indicators), mitigating policies (the Oxford COVID Government Response Tracker), natural environment (weekly average precipitation and temperature, air pollution, and air transport) healthcare infrastructure, economic performance (*e.g.*, GDP per-capita, unemployment rate, human capital, and education), social connectedness (tourism, digitalisation), and governance variables. To study their impact, we adopt a framework of ordered multinomial models.

The second step is motivated by several arguments. First, it allows us to understand the role of societal structural factors in the different forms of epidemic waves across different geographical regions. Second, the significant impact of socioeconomic variables illustrates that the spread of epidemics is the result of interactions between several ecosystems, with those related to human behaviour playing a significant role.

The contribution of this article to the existing literature on the relationship between socioeconomic variables and epidemics is the following. Rather than studying the impact on the number of cumulative cases, we focus our interest on the dynamics of pandemics, *i.e.*, on the speed of spread and deceleration of a disease at different stages of a pandemic. The importance of this approach is to capture the variables that influence the transition states of epidemic curves. This leads us to propose a methodological innovation by estimating generalised Richards curves with time-varying coefficients, which are usually assumed to be constant. This is interesting to assess how, in the future, we could curb the curve of new cases as the pandemic progresses by acting on socioeconomic causes.

The key findings of the paper are the following.

First, the implementation of policies to combat COVID-19 (*e.g.*, social distancing) only slows down an epidemic if it spreads very quickly at the start of a new epidemic wave.

Second, richer countries experience faster epidemic outbreaks than poorer ones, despite the fact that the former have better living conditions and more developed health infrastructure. When sub-Saharan Africa is excluded from the sample of countries, epidemic factors do little to explain the inertia or speed at which an epidemic starts. This is probably due to predominant biological and epidemiological factors. Finally, when sub-Saharan Africa is excluded from the sample, we also find that the turning point of epidemics depends on the countries' vulnerability to environmental conditions, as well as the socioeconomic characteristics of the population.

The remainder of the article is organised as follows. Section 2 presents our contributions to the literature. Section 3 presents our methodological framework *i.e.*, the phenomenological epidemic model and the estimation methods. Section 4 presents the results of our investigation of the relationship between the epidemiological parameters and socioeconomic variables. Section 5 concludes.

2 Contribution to the literature

Our paper lies at the intersection of two literatures, one relating to the mathematical modelling of the expansion of a viral disease epidemic, and the other related to social epidemiology. We contribute to both strands of the literature by understanding how some socioeconomic determinants have a predictive effect on the probability that a population evolves in a regime of high or low epidemic diffusion at the beginning of a contagious viral disease or around an epidemic peak. Regarding the modelling of COVID-19, two classes of models are used in the epidemiological literature to investigate the spread of epidemics, such as COVID-19.

First, researchers have proposed scenarios and simulation models based on the interactions between people whose degree of contamination by the disease varies. Such models are theory-based. These include SIR and SEIR models. The population is divided into several exclusive categories (susceptible, infected, exposed, and recovered) and the disease spreads according to how the groups interact with each other. These models are referred to as compartmental models. Their aim is to measure the reproduction rate R_0, *i.e.*, the risk of a new pathogen causing an epidemic, the final size of the epidemic, and the effect of public health intervention, such as the impact of vaccination on herd immunity (for illustrations and extension of compartmental models, see Capasso, 2008, Porwal et al., 2015, Rodrigues, 2016, Atkeson, 2020, Bernadi and Manuchehr, 2021, and Section A of the special issue of the Journal of Mathematical Economics edited by Boucekkine et al., 2021).

An alternative objective to building scenarios from theory-based models is to fit the curves suggested by these models to clinical and epidemiological data and use the estimates to obtain data-based predictions. For instance, this involves inferring from past clinical observations what proportion of patients are likely to be positive to detection tests. The models used for this purpose belong to the category of linear and nonlinear time series models, neural networks, random forests, support vector machines, logistic regressions, etc. (see, for example, Maleki et al., 2020, Andrade et al., 2021, Doornik et al., 2021, Chyon et al., 2022). Among the empirical models, one of the most widely used is the Richards model (known as the theta-logistic model), which allows the prediction of the cumulative number of new infected people. In this model, the cumulative number of new cases is described by a bell-shaped function with a specific diffusion mechanism. When an epidemic breaks out, it spreads at an exponential rate. However, the rate of reproduction is slowed by negative feedback mechanisms. A recent literature has motivated the use of this model because the Richards model has a one-to-one correspondence with SIR models in the sense that the latter give an explicit formula for the parameter that appears in the Richards equation (see the proof by Wang et al., 2012).

However, such a model is based on assumptions that contradict reality and whose limitations we take into account. First, the number of people finally infected is equal to the size of the population, which is interpreted as reflecting a situation where the disease stops spreading when there is herd immunity in the population as a whole (*i.e.*, when someone meets an infected person, he or she necessarily becomes ill). There are two possibilities: either ill people die or survive having acquired immunity). However, in reality, some people never become ill. This means that we need to take a closer look at the feedback mechanisms that slow down the reproduction rate of the disease. This may depend on factors other than immunological or biological causes. In this paper, we show that, taking a general version of Richard's model (a beta logistic function), not only is the number of contaminated people far below the population size of the countries examined, but the parameter describing the negative feedback mechanism is also linked to factors that have the same effect as immunity mechanisms, but are of a different nature. In particular, we show that socioeconomic factors can explain the speed at which epidemics expand or slow down.

Moreover, a restrictive assumption of the standard Richards model is that the logistic shape is symmetrical around the peak of the epidemic. This assumes that, on either side of the peak, the rates of rise and fall are identical. Our model shows that, in reality, there are different forms of endemic behaviours to the left and right of an epidemic peak, but also different start-up speeds when the disease begins. The disease spread dynamics before and after the peak are asymmetric. Our work falls within a second field of literature, that of social epidemiology, which has seen renewed interest since the emergence of Covid-19. Surprisingly, the economic literature on the links between social factors and Covid-19 has fo-

cused mainly on reverse causality, *i.e.*, on the socioeconomic implications of this viral disease. This has been motivated by the high long-term costs of COVID-19. In contrast to this literature, we also wish to contribute to the literature on social medicine and social epidemiology by showing that the environment in which people live can explain the dynamics of the spread of COVID-19 beyond biological factors.

Our aim is to determine whether the factors usually identified as socioeconomic determinants of health in the literature (see Braveman and Gottlieb, 2014) have been conducive to the spread of COVID-19. Research to date has focused on the role of the following factors: a) demographic structure to explain morbidity and mortality rates (see Ascani et al., 2020, Sannigrahi et al., 2020, Amdaoud et al., 2021a, Bourdin et al., 2021 for studies of European regions; Borjas, 2020, Schmitt-Grohé et al., 2020 for studies of US data), b) wealth and income inequalities insofar as they reflect inequalities in quality of life, medical care (see Amdaoud et al., 2021a), c) social capital as a factor in the trust people place in social distancing policies (see Van Bavel et al., 2020, Amdaoud et al., 2021b, Barrios et al., 2021, Bartscher et al., 2021, Brodeur et al., 2021), d) anti-COVID policies (see Chen et al., 2021, Gallic et al., 2021), or e) the structure of labour markets (see Mongey et al., 2020). Instead of taking only a few variables, we consider a large number (41) that concern six essential dimensions of socioeconomic health factors taken from several fields of health sciences (medicine, health economics, medical health, and international organisations). The areas covered are as follows.

The first is the quality of health infrastructure. For example, during the acute phases of COVID-19, we observed differences in the extension of countries whose hospitals were provided with oxygen, hospital beds, and anti-retroviral drugs. The second and third are sensitivity to comorbidities and vulnerability to the environment, respectively. In the medical and epidemiological literature, these are called stress mediators, *i.e.*, factors that can accelerate physiological changes if someone is exposed to COVID-19. For example, they worsen inflammatory responses to the virus or dysregulation of the metabolic system. The fourth and fifth are the living conditions and societal characteristics of the populations, respectively. These are usually cited as biomarkers of health risks. For example, living in poverty reduces immune resistance and causes more psychosocial stress, which accentuates cellular ageing and increases the likelihood of exposure to chronic disease. Similarly, the higher a population's level of education, the more important healthy behaviours are. Finally, we consider policy variables to combat COVID-19.

3 Estimation of a generalised Richards model of epidemic curve

Our objective is to assess the impact of socioeconomic determinants on the spread of the COVID-19 epidemic. These factors can influence the growth dynamics of

an epidemic, and their effects may vary depending on whether we are examining the early stages of the epidemic or its peak. To effectively quantify these effects, it is essential to initially model the epidemic, which we undertake in this section using a phenomenological model. Once this model is estimated, we can analyse the parameters that characterise the propagation dynamics of the epidemic.

3.1 The estimated model

As explained in Section 2, there are two approaches for modelling the transmission of an epidemic. The first approach is to test specific mechanisms based on theoretical behavioural models, such as the SIR and SEIR family of models. The second approach, adopted here, focuses on key parameters that are responsible for the spread of the disease,regardless of the cause. The models used in the second case are referred to as phenomenological. They are used to investigate the growth dynamics of an epidemic on based on a small number of parameters, whereas theoretical models involve a large number of parameters and require the researcher to simulate trajectories by making numerous assumptions.

Here, we use a growth model known as the generalised Richards model, which extends the standard Richards model based on logistic growth to more realistic cases of potential epidemic regimes. This is the case for so-called re-emerging viral infectious diseases. Indeed, for the same disease, the characteristic profiles of epidemic curves can vary. An epidemic can occur where the number of ill people rises rapidly, reaches a peak, and then falls back just as quickly. We can also observe a situation in which a lightning ascension phase is followed by a plateau dynamic around the peak. The propagation of the disease can also be characterised by a very slow decline after a very gradual rise, which is generally the case with dissemination epidemics. To account for this variety, we need a flexible form of the growth function. The one considered here is as follows:

$$\Delta C_t = C_{t+1} - C_t = rC_t^P \left[1 - \left(\frac{C_t}{K}\right)^{\alpha}\right] + \varepsilon_t, \quad P \in [0, 1]. \tag{1}$$

where ε_t is the model error and is distributed as a Gaussian noise process with a zero mean and variance $\sigma^2_{\varepsilon_t}$.

A continuous time equivalent of Equation 1 has been used in the literature to simulate different past viral pandemics (see Wang et al., 2012, Chowell et al., 2016, Viboud et al., 2016, Pell et al., 2018).

ΔC_t is the number of new infected, C_t is the cumulative number of infected. P, the deceleration rate parameter, captures the different growth dynamics in the early stages of an epidemic episode. During the early ascending phases of an infectious disease, the incidence curve dynamics are 'dominated' by rC_t^P. They can be characterised by different types of growth: linear growth (or constant incidence, when $P = 0$), exponential (or Malthusian growth, when $P = 1$), and

sub-exponential growth when $0 < p < 1$ (polynomial growth). r is the intrinsic growth rate, *i.e*, the growth rate of the disease that would prevail if there were no limitations to the disease. Parameter α measures the extent of deviation from a sigmoid function. K is the final epidemic size (final outbreak size).

Equation 1 encompasses several submodels. When $\alpha = 1$, the generalised Richard's model reduces to the standard Richard's model. When $\alpha = P = 1$, it reduces to the logistic growth model. When $\alpha \neq 1$, the growing or decreasing dynamics of the epidemic is larger ($\alpha > 1$) or lower ($\alpha < 1$) than the logistic decay. These configurations are likely to be observed as we approach the inflection point of an epidemic.

3.2 Nonlinear least squares estimator

Let the set of parameters be denoted by vector $\Theta = (r, P, \alpha, K)$. In our analysis, the components of $\hat{\Theta}$ are estimated using nonlinear least squares (NLS):

$$\hat{\Theta} = \text{argmin } S(\Theta/C_{t+1}, C_t), \tag{2}$$

and

$$S(\Theta/C_{t+1}, C_t) = \Sigma_{i=1}^{N} \left[F(C_t, \Theta) - C_{t+1}\right]^2, \tag{3}$$

where N is the number of observations and $F(C_t; \Theta)$ is defined by

$$C_{t+1} = F(C_t; \Theta) = C_t + rC_t^P \left[1 - \left(\frac{C_t}{K}\right)^{\alpha}\right] + \varepsilon_t, \quad P \in [0, 1]. \tag{4}$$

Denoting $\boldsymbol{\varepsilon}$ the vector of residuals, we have

$$\hat{\Theta} \longrightarrow_a \mathcal{N}\left(\Theta, V(\Theta)\right), \text{ with } V(\Theta) = \sigma_{\varepsilon}^2 \left[(\partial\boldsymbol{\varepsilon}/\partial\Theta)'(\partial\boldsymbol{\varepsilon}/\partial\Theta)\right]^{-1}. \tag{5}$$

σ_{ε}^2 can be estimated by $\hat{\sigma}^2 = \hat{\varepsilon}'\varepsilon/N$. The vector of parameters $\hat{\Theta}$ is obtained by numerically solving the system of equations given by the first-order conditions of the minimization program

$$\partial S/\partial\Theta = 2\varepsilon'(\partial\boldsymbol{\varepsilon}/\partial\Theta = -2\boldsymbol{\varepsilon}'(\partial F/\partial\Theta) = \mathbf{0}. \tag{6}$$

The 95% confidence interval for each component of vector $\hat{\Theta}$ is obtained by computing the lowest and highest bounds as follows:

$$L(\hat{\Theta}_j) = \hat{\Theta}_j - 1.96\,\hat{\sigma}(\hat{\Theta}_j),\ U(\hat{\Theta}_j) = \hat{\Theta}_j + 1.96\,\hat{\sigma}(\hat{\Theta}_j). \tag{7}$$

3.3 Evidence of heterogeneous epidemic behaviours across regions

Our study focuses on the COVID-19 pandemic in recent years from January 22, 2020, to July 07, 2021. We consider both industrialised and emerging/developing

countries. The sample comprises 115 countries. The complete list of nations and their corresponding regional categorisation are outlined in Table A.1 in the Online Appendix A.

This period corresponds to almost a year and a half of the epidemic follow-up, a period during which all countries had passed through the first epidemic wave. At the end of this period, at the beginning of July 2021, some countries were still in a period of epidemic growth, whereas others saw the start of a new rebound in the number of cases, heralding the appearance of new virus variants. Moreover, the end of the period also corresponds to the period in which the discovery of several vaccines against COVID-19 has taken place, and many countries were well advanced in their vaccination campaigns. In our paper, the generalised Richards model does not allow us to model epidemic rebounds perfectly, as they are traditionally used in the case of seasonal infectious diseases (*e.g.*, malaria). It also makes it difficult to consider vaccination and the phenomenon of variants, which is why we have limited the study period to mid-2021. Regarding the choice of geographical area, we tried to be exhaustive by including as many countries as possible. Data collection was limited by the availability of both epidemic (cumulative number of cases) and socioeconomic data (variables used to construct the socioeconomic factors).

To estimate the parameters of Equation 1, we rely on daily data that counts new cases by country, provided by the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.[1]

The Equation 1 model is estimated for all 115 countries, and we also consider subsamples comprising regional subgroups. The interquartile interval, minimum, and maximum values of the estimated parameters for the individual countries, as well as the mean and standard deviation of our estimates, are reported in Table 1. The means are the averages of the estimated parameters for countries within each region. The standard deviation is computed in a similar way by considering the estimates within the regions. For the first line of each parameter entitled 'All countries,' calculations were made considering all countries, regardless of the region to which they belong. Of particular interest in our analysis are the estimations of two key parameters, $\hat{\alpha}$ and $\hat{P}$. Examining these estimations reveals several noteworthy findings. Sub-Saharan Africa is an outlier. The median P value for this sub-region is almost twice that of the median for all countries. The mean is also higher than that in other regions. With regard to the α parameter, sub-Saharan Africa stands out even more from other countries in terms of median and mean. The variability within this sub-region is also high compared to the others (see standard deviation), which is understandable given the larger number of countries in the sample.
If we leave out sub-Saharan Africa, for the α parameter, the across-region standard deviation (2.84) is greater than the within-region standard deviation for all

[1] https://github.com/CSSEGISandData/COVID-19 (Last accessed 9 January 2024).

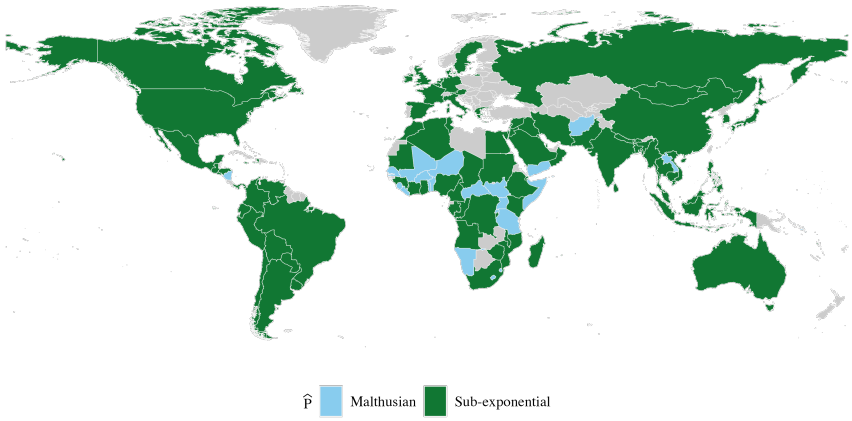
regions, except for the MENA group. Therefore, the table underscores the presence of substantial heterogeneity in the turning point of the epidemic curves across regions. However, industrialised countries stand out from the other regions, in that the mean and standard deviation are much lower than in the other regions. At the inflection point of the epidemic episode, the rate of increase or decrease in the number of cases is much slower than elsewhere, with more homogeneous behaviour between countries (the mean of α of 0.32 is almost five times lower than that of all countries outside sub-Saharan Africa, and the standard deviation of 0.83 is almost 3.5 times lower than that of all countries outside sub-Saharan Africa). Heterogeneity between sub-regions is less pronounced with respect to growth dynamics in the early stages of an epidemic episode. Indeed, the standard deviation of the parameter P estimated for the whole sample excluding sub-Saharan Africa (0.22) is close to the estimated standard deviation of the sub-regions, except for industrialised countries, whose behaviour is more homogeneous between them than elsewhere.

To summarise our findings, the parameter associated with the deceleration of growth, $\hat{P}$, reveals distinct patterns between industrialised and emerging/developing countries during the initial stages of the pandemic. Notably, in industrialised nations, the early onset of the pandemic exhibits characteristics of sub-exponential growth, as evidenced by the estimated values of $0 < \hat{P} < 1$, all well below 1. These countries are shown in blue in Figure 1. Conversely, emerging/developing countries display a more rapid and, importantly, exponential growth profile, as evidenced by $\hat{P}$ estimates close to 1. These nations are illustrated in orange in Figure 1.

The parameter $\hat{\alpha}$ indicates the timing of the inflection point on the epidemic curve. The higher the value of $\hat{\alpha}$, the faster the inflection point will occur. Figure 2 categorises countries into two groups based on $\hat{\alpha}$. Countries where $\hat{\alpha}$ is less than or equal to 1 are shown in pink, whereas those where $\hat{\alpha}$ exceeds 1 are shown in purple. This map facilitates the identification of countries characterised by shorter epidemic duration, *i.e.*, those where the peak is reached swiftly.

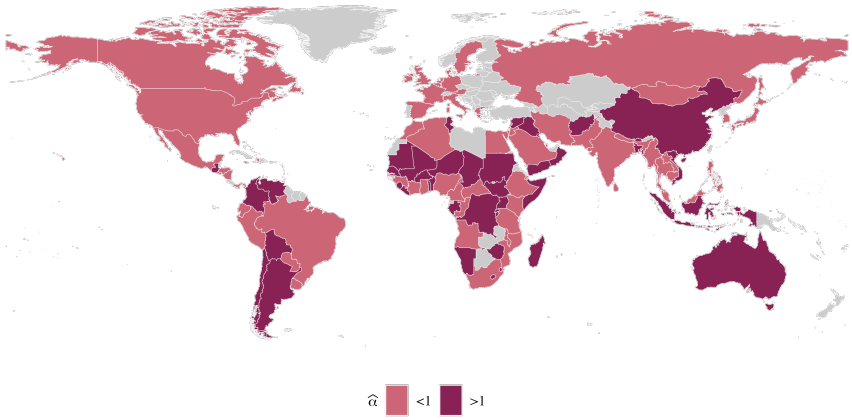
Interestingly, the map shows that there is no clear demarcation between industrialised and emerging/developing countries in terms of $\hat{\alpha}$. The latter group includes many developing countries with epidemic patterns similar to those of developed countries, including Algeria, Morocco, Oman, Cameroon, Colombia, and Brazil. Additionally, China and Australia deviate from the industrialised country trend, exhibiting a delay in the occurrence of the epidemic peak compared to their counterparts.

Figure 1: Geographical distribution of estimated values for $\hat{P}$ by countries



Source: Authors' calculations.

Figure 2: Geographical distribution of estimated values for $\hat{\alpha}$ by countries



Source: Authors' calculations.

4 Measuring the influence of socioeconomic variables on $\hat{P}_i$ and $\hat{\alpha}_i$

This section analyses the impact of socioeconomic variables on $\hat{P}_i$ and $\hat{\alpha}_i$, exploring their role in epidemic onset speed and dynamics at the peak. We employ six synthetic variables derived from principal component analysis, shedding light on key determinants in different global contexts.

Table 1: Descriptive statistics of the estimates of the Parameters of Generalised Richards model

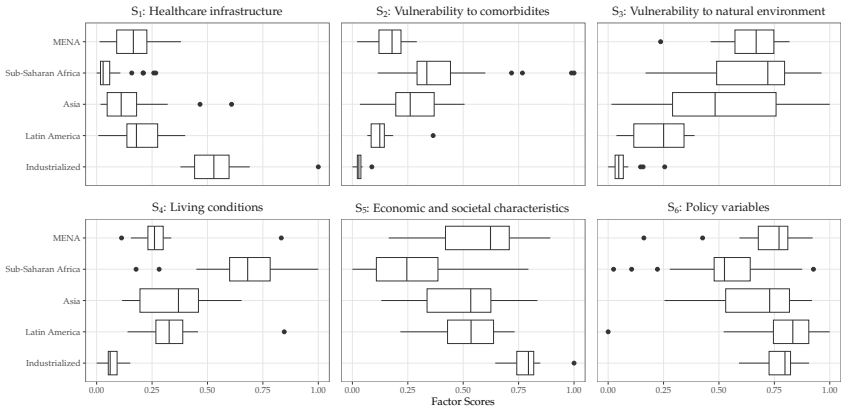
	Min	Max	Q1	Median	Q3	Mean	SD	n
r								
All countries	0.00	100.00	0.06	19.58	46.41	26.23	27.90	115
All countries (exc. SSA)	0.00	100.00	6.08	36.40	53.47	34.37	28.78	73
Asia	0.01	100.00	0.93	8.22	32.88	19.86	26.11	23
Industrialised	0.12	100.00	45.53	56.17	79.68	59.33	24.44	16
Latin America	0.00	100.00	12.09	47.37	55.99	38.84	27.79	18
MENA	0.01	53.47	8.38	23.01	42.51	25.23	20.04	16
Sub-Saharan Africa	0.00	86.57	0.01	0.04	22.27	12.10	19.70	42
MENA	0.01	53.47	8.38	23.01	42.51	25.23	20.04	16
Sub-Saharan Africa	0.00	86.57	0.01	0.04	20.63	11.34	19.24	45
P								
All countries	0.00	1.00	0.36	0.46	0.90	0.57	0.29	115
All countries (exc. SSA)	0.00	1.00	0.37	0.44	0.62	0.50	0.22	73
Asia	0.00	1.00	0.44	0.61	0.74	0.60	0.26	23
Industrialised	0.37	0.82	0.39	0.42	0.45	0.45	0.10	16
Latin America	0.21	1.00	0.31	0.40	0.53	0.45	0.19	18
MENA	0.29	1.00	0.35	0.37	0.46	0.49	0.25	16
Sub-Saharan Africa	0.16	1.00	0.30	0.95	1.00	0.70	0.34	42
α								
All countries	0.00	26.82	0.09	0.15	3.48	2.50	4.49	115
All countries (exc. SSA)	0.00	13.92	0.06	0.13	1.98	1.55	2.84	73
Asia	0.00	7.25	0.02	0.06	3.23	1.52	2.42	23
Industrialised	0.04	3.45	0.11	0.12	0.14	0.32	0.83	16
Latin America	0.00	9.17	0.04	0.14	2.20	1.50	2.55	18
MENA	0.02	13.92	0.14	0.16	6.08	2.87	4.31	16
Sub-Saharan Africa	0.00	26.82	0.15	2.44	5.52	4.16	6.13	42
K ($\times 10^6$)								
All countries	0.00	1,394.87	0.70	10.17	42.59	57.01	184.88	115
All countries (exc. SSA)	0.00	1,394.87	4.17	16.74	52.27	79.40	227.92	73
Asia	0.00	1,394.87	2.17	35.70	125.54	176.10	385.11	23
Industrialised	0.03	324.82	9.73	26.87	66.25	54.98	80.35	16
Latin America	0.06	208.50	4.66	13.76	31.30	32.29	52.69	18
MENA	0.01	84.36	2.35	7.41	34.56	17.82	23.28	16
Sub-Saharan Africa	0.00	193.57	0.02	2.16	17.11	18.07	36.98	42

Notes: The number of observations in each geographical group is denoted by n. Values for parameter K have been scaled by a factor of 10^6. 'SD' stands for 'Standard Deviation,' '$Q1$' and '$Q3$' denote first and third sample quartiles, n is the number of observations, 'All countries (exc. SSA)' stands for 'All countries except Sub-Saharan Africa'.
Source: Authors' calculations.

4.1 Unpacking socioeconomic influences: Principal components

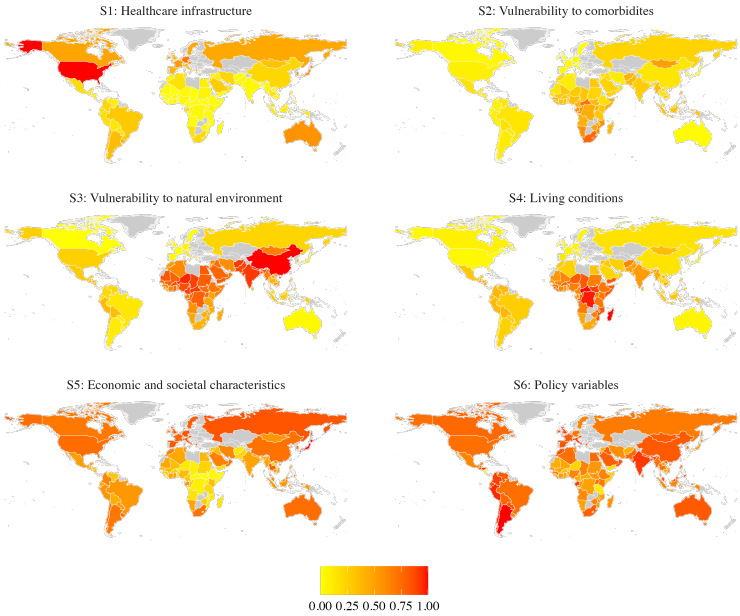
To capture multiple dimensions underlying our analysis, we employ Principal Component Analysis (PCA) to derive six synthetic variables: Healthcare infrastructure (S_1), Vulnerability to comorbidities (S_2), Vulnerability to the natural environment (S_3), Living conditions (S_4), Economic and societal characteristics (S_5), and Policy variables (S_6). Each of these synthetic variables, denoted as S_k, is derived through PCA, based on a set of socioeconomic variables $\mathcal{S}_k$ that pertain to the respective synthetic factor. For a comprehensive list of factors and socioeconomic variables, please refer to Online Appendix A. For each synthetic variable S_k, we initiate PCA while retaining as many principal components as there are variables within the corresponding set. The synthetic variable is then obtained by computing the weighted average of the individual coordinates within the PCA-transformed space. The weights assigned to each coordinate correspond to the proportion of the explained variance associated with the respective PCA components. The synthetic factors are time invariant and change only across countries. They are normalised between 0 and 1 to make their effects comparable.

Figure 3: Distribution of factor scores estimated with the Principal Component Analysis in each region.



Source: Authors' calculations.

Figure 4: Normalised factor scores by countries



Source: Authors' calculations.

Figure 3 shows boxplots for each country group and indicator category. Industrialised countries have higher values for the health infrastructure scores (S_1) than the other groups. Sub-Saharan African countries have the lowest values among the groups. With regard to vulnerability to comorbidities (S_2), the countries of sub-Saharan Africa and Asia have higher scores overall than those of the other zones. Asian countries all have a score lower than or roughly equal to 1; sub-Saharan African countries have scores higher than 2. Moreover, the dispersion of scores for these two regions is greater than that for the other three zones. In terms of vulnerability to natural environment scores (S_3), MENA, sub-Saharan Africa, and Asian countries have higher scores than Latin America and industrialised countries, and Asia has a higher variation in these scores. With respect to the living conditions score (S_4), industrialised countries have the lowest scores and sub-Saharan African countries the highest scores. This result may seem ambiguous, but it should be noted that the indicators considered in this score are generally poor development indicators, such as the prevalence of undernourishment or the poverty rate. Thus, countries with lower scores have better living conditions. In terms of economic and societal characteristics (S_5), the scores of industrialised countries are higher than those of the other groups. In addition, they have a lower dispersion than the other groups. Finally, for the Policy variables (S_6), MENA, Latin American, and industrialised countries have the highest scores with a fairly low dispersion. The dispersion of scores is high within the

Asia and sub-Saharan Africa groups.

Figure 4 displays the normalised factor scores for each dimension of countries around the world. It shows that there is heterogeneity across countries. For example, we can see a relatively low level of health infrastructure on the African continent and in certain countries in Southeast Asia. Africa and parts of Asia show a high vulnerability to comorbidities, as well as to negative environmental externalities. On the other hand, we observe greater homogeneity among countries in terms of the adoption of anti-COVID policies.

Rather than examining the effects of socioeconomic variables on the cumulative number of cases, we look at the effects during the different phases of the incidence curve, *i.e*, on the parameters that identify the severity of the ascending and declining phases of the crises (slow or fast) and the dynamics of the disease in the early stage of a pandemic. Our two variables of interest are the parameters $\hat{P}_i$ and $\hat{\alpha}_i$ from Equation 1, where sub-index i refers to a country.

4.2 Epidemic onset speed

Parameter $\hat{P}_i$ is used to describe the dynamics during the initial stage of an epidemic. We define this parameter using a three-level multidimensional ordered variable:

$$V_{1i} = \begin{cases} 1, & \text{if } 0 \leq \hat{P}_i \leq P_{.25} \\ 2, & \text{if } P_{.25} < \hat{P}_i \leq P_{.75} \\ 3, & \text{if } P_{.75} < \hat{P}_i \leq 1 \end{cases}, \tag{8}$$

where $P_{.25}$ and $P_{.75}$ are chosen as the first and third empirical quartiles of the vector $(\hat{P}_{i=1,...,100})$. The values 1,2,3 mean 'slow,' 'medium,' and 'fast,' respectively.

We estimate an ordered multinomial Probit model, using the {`oglmx`} `R` package (Carroll, 2018). For the endogenous variable V_{1i}, we consider the following latent model:

$$V_{1i}^* = \sum_{k=1}^{6} \beta_k \, S_{ki} + \sigma_i \, \varepsilon_i, \qquad \varepsilon_{\kappa t} \approx \mathcal{N}(0,1). \tag{9}$$

The latent variable allows to define a partition of the endogenous variables as follows:

$$\hat{V}_{1i} = \begin{cases} 1, & \text{if } V_{1i}^* \in (-\infty, \lambda_1] \\ 2, & \text{if } V_{1i}^* \in (\lambda_1, \lambda_2] \\ 3, & \text{if } V_{1i}^* \in (\lambda_2, \infty) \end{cases}. \tag{10}$$

The coefficients $\lambda_1 > \lambda_2$ are unknown parameters.

Define $\boldsymbol{S}_i = (S_{1i}, ..., S_{6i})$ as the vector of explanatory variables and let $\boldsymbol{\beta}$ be the vector of parameters to be estimated. The probability of an outcome conditional

on the explanatory variables writes:

$$\mathbb{P}\left(\hat{V}_{1i} = \nu \mid \boldsymbol{S}_i,\ \beta\right) = \zeta\left(\boldsymbol{\lambda}_{\nu+1} - \boldsymbol{S}_i'\,\beta\right) - \zeta\left(\lambda_\nu - \boldsymbol{S}_i'\,\boldsymbol{\beta}\right), \tag{11}$$

where $\nu \in \{1, 2, 3\}$, $\lambda \in \{\lambda_1, \lambda_2\}$, and $\zeta(.)$ is the cumulative distribution function of a Probit model.

Let each element of the vector of synthetic variables S_i be denoted by S_{ki}, where $k = 1, ..., 6$. The objective is to compute marginal effects, which quantify how a one-unit change in variable S_{ki} impacts the associated probability of a specific outcome ν, where $\nu = 1, 2, 3$. The estimated coefficients of S_{ki} in the latent regression are denoted by $\hat{\beta}_k$.

We calculate the marginal effect of each synthetic variable on λ_1 and λ_2 and average the marginal effects across the sample:

$$\text{AME}(\nu) = \frac{1}{N}\sum_i \text{ME}(\nu, S_{ki}'\hat{\beta}_k), \tag{12}$$

where

$$\text{ME}(\nu, S_{ki}'\hat{\beta}_k) = \frac{\partial \mathbb{P}(V_{1i} = \nu)}{\partial S_{ki}}, \tag{13}$$

and where $\hat{\beta}_k$ are the estimated coefficients from Equation 9.

We begin our analysis by estimating the coefficients $\hat{\beta}_k$, using an ordered probit model characterised by heteroskedastic variance.[2] We then proceed to calculate the marginal effects, and the resulting estimates are presented in Table 2. It is noteworthy to reiterate that in our classification, the onset speed of an epidemic is stratified into three different modalities, specifically labelled as '1' to denote a slow start, '2' for a moderate start, and '3' for a rapid start.

We first consider the full sample of countries, and then a sample without sub-Saharan African countries. Indeed, as we saw previously, this sub-region differs from the others in terms of both mean and median values of the P and alpha parameters and shows greater heterogeneity between countries than elsewhere. It is therefore interesting to test the robustness of the results by taking a sub-sample free of this sub-group of countries, which we consider outliers.

The coefficients of Regression (9), which measure the effects of variations in the independent variables on an unobservable latent variable, cannot be interpreted directly. However, this variable provides information on a partition of the data between different outcomes. Therefore, the information to be looked at concerns marginal effects. This informs us about the effects of a small variation in the explanatory variables on the probability of being in a given outcome. This is what

[2] The complete set of estimated coefficients can be found in the supplementary material provided online.

the line in the table with $\mathbb{P}(V_1 = 1)$, $\mathbb{P}(V_1 = 2)$, etc, means. As the outcomes are complementary, the sum of the coefficients is necessarily equal to zero. A negative sign suggests that a small variation in the independent variable reduces the probability of being in a given outcome. By contrast, a positive sign indicates that this probability increases.

Table 2: Marginal effects of the synthetic variables on the probability of being in each regime for V1

	All countries			Without sub-Saharan Africa		
	$\mathbb{P}(V_1 = 1)$ (Slow)	$\mathbb{P}(V_1 = 2)$ (Medium)	$\mathbb{P}(V_1 = 3)$ (Fast)	$\mathbb{P}(V_1 = 1)$ (Slow)	$\mathbb{P}(V_1 = 2)$ (Medium)	$\mathbb{P}(V_1 = 3)$ (Fast)
S_1: Healthcare infrastructure	-0.36**	-0.05	0.41*	-0.54	0.33	0.2
	(0.17)	(0.1)	(0.23)	(0.37)	(0.27)	(0.17)
S_2: Vulnerability to comorbidities	-0.66	0.64	0.02	-0.27	-0.19	0.46
	(0.46)	(0.64)	(0.32)	(0.87)	(1.09)	(0.52)
S_3: Vulnerability to natural environment	0.14	0.02	-0.16	0.19	-0.12	-0.07
	(0.1)	(0.04)	(0.11)	(0.19)	(0.12)	(0.08)
S_4: Living conditions	0.42	-1.99***	1.57***	-0.18	-0.53	0.71*
	(0.39)	(0.52)	(0.47)	(0.84)	(0.91)	(0.42)
S_5: Econ. and societal characteristics	-0.2	-0.03	0.23	-0.43	0.26	0.16
	(0.23)	(0.06)	(0.26)	(0.4)	(0.28)	(0.16)
S_6: Policy variables	0.33*	0.05	-0.37*	0.56*	-0.35	-0.21
	(0.18)	(0.08)	(0.21)	(0.34)	(0.26)	(0.15)

Notes: This table provides estimates of the coefficients from Equation 12. Positive marginal effects are depicted in green, whereas negative effects are depicted in red. Standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. 'Slow,' 'Medium,' and 'Fast' correspond to the propagation speeds at the beginning of the epidemics. S_1: Healthcare infrastructure (higher values depict better healthcare infrastructure), S_2: Vulnerability to comorbidities (higher values depict more vulnerability to comorbidities), S_3: Vulnerability to natural environment (higher values depict more vulnerability to natural environment), S_4: Living conditions (higher values depict more vulnerability to living conditions), S_5: Economic and societal characteristics (higher values depict better living conditions), S_6: Policy variables (higher values depict stricter anti-COVID policies). Source: Authors' calculations.

Our initial regression encompasses all countries from our sample. The results shown in the left section of Table 2 reveal several main interesting facts.

First, the variable representing healthcare infrastructure appears to play a role in the speed of the epidemic onset, with its marginal effect coefficient being statistically significant at the 5% level. The negative sign in outcome $V_1 = 1$ (-0.36) suggests that more developed healthcare infrastructure delays the spread of an epidemic at the very start, when it has just begun.

Second, better living conditions ($S4$ decreases) slow the spread of an epidemic when it starts strongly. Indeed, the sum of significant coefficients for outcomes $V_1 = 2$ and $V_1 = 3$ is negative).

Third, when an epidemic starts aggressively (*i.e.*, very rapidly), anti-COVID policies are effective in slowing down this rapid progression. Indeed, the only significant coefficient for this variable indicates a negative marginal effect on the probability of being in outcome $V_1 = 3$.

We perform the same regression excluding sub-Saharan African countries to assess if the results change. The results are shown in the right section of Table 2. They differ in several respects.

The new regressions show that the speed of epidemic start-up no longer depends on health infrastructure, and that, this time, deteriorated living conditions favour a rapid epidemic start-up (positive marginal effects of S4 for outcome $V_1 = 3$. An interesting result is that COVID control policies are even more effective when sub-Saharan Africa is removed from the sample. When they become more restrictive, the epidemic is twice as likely to start slowly (positive and significant marginal effect of 0.56 for outcome $V_1 = 3$), whereas the probability is only 0.33 for the full sample.

The coefficients that are not statistically significant (whether or not sub-Saharan Africa is included in the sample) also provide us with interesting information.

First, whether an epidemic starts quickly or very slowly is independent of countries' general health of population (which is captured here by the variable measuring vulnerability to comorbidities). Indeed, we find no significant marginal effect of S_2. Recall that this variable synthesises not only the diseases of the richest countries, including diabetes, obesity, and cancer, but also those of the poorest, including tuberculosis, malaria, and HIV. Our results imply that none of these factors was decisive in the rate at which a pandemic like COVID-19 spread. Even if an individual was more likely to be contaminated by the epidemic if he or she was suffering from hypertension, cardiovascular, complicated diabetes or related diseases, at the level of an entire population (epidemiological level), causal links cannot be clearly established. Our results confirm some findings in the medical literature, showing little evidence of co-existing diseases and COVID-19. Only one in five people worldwide with chronic comorbidities is susceptible of being at higher risk of COVID-19. In addition, there is an issue of reverse causality: chronic diseases can worsen after acute COVID-19 (see, for instance Ranard et al., 2021, Adab et al., 2022, and Bigdelou et al., 2022)

Second, the occurrence of the COVID-19 pandemic has raised questions about the need to adopt 'One health' type control strategies of such a disease (in the sense that these strategies involve several components (medical, combating global warming, nutrition policy, reducing globalisation, etc.). As far as the environment is concerned, there seems to be a wide spread view in the epidemiology and ecological literature that viruses emerged and spread as a result of disturbances to wildlife reservoirs, themselves created by a deterioration in environmental conditions (climate change, fast development of international transport, high tem-

peratures, flooding, etc.). These factors are captured here by the synthetic factor S_3, which measures a country's vulnerability to the natural environment. While the medical literature shows that environmental disturbances create favourable conditions for the emergence of re-emerging diseases such as COVID-19 (see, for example, McNeely, 2021), our work suggests that they do not, however, explain the degree of severity of a pandemic at its outset.

Third, our estimates also show that not all economic and social characteristics explain the strength of the early stages of epidemics. The variable S_5 captures some factors around the world that are likely to foster, or on the contrary to limit, contact between people, and therefore potentially effective in explaining cases of COVID-19: population density linked to the scale of the informal economy in poor countries, the frequency of use of tools for communicating at a distance such as cell phones or the Internet, the fact of being a very open country, or on the contrary closed to international migration, and so on. Taken as a whole, our regressions suggest that they play no role in growth dynamics during the initial phases of the epidemic.

4.3 Dynamics at the epidemic peak

We now turn our attention to the variable V_2, following the same principles. This variable characterises the dynamics when the peak of the epidemic is reached. Deceleration before or after the epidemic peak can occur rapidly or at a 'normal' pace, considered here as following a logistic decay.

Parameter $\hat{\alpha}_i$ in Equation 1 serves to characterise the rate at which an epidemic either intensifies or subsides. We define a qualitatively ordered variable with two distinct attributes: the rate of epidemic intensification or subsidence can either be slower than the logistic decay ($\hat{\alpha}_i \leq 1$) or faster than the logistic decay ($\hat{\alpha}_i > 1$):

$$V_{2i} = \begin{cases} 1, & \text{if } \hat{\alpha}_i \leq 1, \\ 2, & \text{if } \hat{\alpha}_i > 1, \end{cases} \tag{14}$$

where 1 means 'lower' and 2 means 'higher.'

Characterising the dynamics of a disease around the epidemic peak, *i.e.*, what happens during the pre-peak and post-peak periods, is a key challenge.

Indeed, when $\alpha_i \leq 1$ (outcome $V_{2i} = 1$), the disease dynamics around the epidemic peak are asymmetrical. When $\alpha_i \leq 1$, during the pre-peak phase, the disease takes time to spread. However, once the peak is reached, the number of new cases falls rapidly. When $\alpha_i > 1$, the proliferation in the number of new cases leads to the rapid emergence of an epidemic peak. However, once the latter is reached, the disease takes time to decelerate. The first scenario is preferred by health authorities, whereas the second scenario is more feared.

As in Section 4.2, we assume the existence of a latent model:

$$V_{2i}^{*} = \sum_{k=1}^{6} \beta_k \, S_{ki} + \sigma_i \, \varepsilon_i, \qquad \varepsilon_{\kappa t} \approx \mathcal{N}(0,1). \tag{15}$$

The latent variable V_{2i}^{*} allows to define a partition of the endogenous variables:

$$\hat{V}_{2i} = \begin{cases} 1, & \text{if } V_{2i}^{*} \in (-\infty, \gamma], \\ 2, & \text{if } V_{2i}^{*} \in (\gamma, \infty). \end{cases} \tag{16}$$

The coefficient γ is an unknown parameter.

Note that because there are only two levels for $\hat{V}_{2i}$ where there were three for $\hat{V}_{1i}$, Equation 11 becomes:

$$\mathbb{P}\left(\hat{V}_{2i} = \nu \mid \boldsymbol{S}_i, \, \beta\right) = \zeta\left(\boldsymbol{\gamma}_{\nu+1} - \boldsymbol{S}_i^{'} \, \beta\right) - \zeta\left(\gamma_{\nu} - \boldsymbol{S}_i^{'} \, \boldsymbol{\beta}\right), \tag{17}$$

where $\nu \in \{1, 2\}$.

The marginal effects of the synthetic variables $\boldsymbol{S}_i$ are estimated in a manner similar to that described in Section 4.2. The results are reported in Table 3.

Table 3: Marginal effects of the synthetic variables on the probability of being in regime 1 for V_2

	All countries		Without sub-Saharan Africa	
	$\mathbb{P}(V_1 = 1)$ (Slower)	$\mathbb{P}(V_1 = 2)$ (Faster)	$\mathbb{P}(V_1 = 1)$ (Slower)	$\mathbb{P}(V_1 = 2)$ (Faster)
S_1: Healthcare infrastructure	0.5	-0.5	1.46***	-1.46***
	(0.6)	(0.6)	(0.45)	(0.45)
S_2: Vulnerability to comorbidities	-0.13	0.13	1.71	-1.71
	(0.93)	(0.93)	(2.26)	(2.26)
S_3: Vulnerability to natural environment	-0.26	0.26	0.44	-0.44
	(0.29)	(0.29)	(0.31)	(0.31)
S_4: Living conditions	-0.18	0.18	-0.33	0.33
	(0.65)	(0.65)	(1.89)	(1.89)
S_5: Economic and societal characteristics	-0.37	0.37	-0.15	0.15
	(0.39)	(0.39)	(0.22)	(0.22)
S_6: Policy variables	-0.03	0.03	0.4**	-0.4**
	(0.19)	(0.19)	(0.19)	(0.19)

Notes: This table provides estimates of the coefficients from Equation 12. Positive marginal effects are depicted in green, whereas negative effects are depicted in red. Standard errors are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. 'Slower' and 'Faster' denote speeds that are respectively slower and faster than the logistic decay. S_1: Healthcare infrastructure (higher values depict better healthcare infrastructure), S_2: Vulnerability to comorbidities (higher values depict more vulnerability to comorbidities), S_3: Vulnerability to natural environment (higher values depict more vulnerability to natural environment), S_4: Living conditions (higher values depict more vulnerability to living conditions), S_5: Economic and societal characteristics (higher values depict better living conditions), S_6: Policy variables (higher values depict stricter anti-COVID policies). Source: Authors' calculations.

We start with the right-hand side of the table when sub-Saharan Africa is excluded from the sample. We see that better infrastructure and more rigorous anti-COVID policies increase the probability of being in the most favourable scenario, *i.e.*, a relatively slow spread of the epidemic before the peak, and then a rapid disappearance afterwards. Indeed, the marginal effects of the variables S_1 and S_6 are significantly positive for Outcome 1 and negative for Outcome 2. It is interesting that these two variables are significant, and the others are not because they make it easier to control an epidemic. Indeed, when a disease such as COVID-19 spreads slowly in a population, there is a high risk that mutants of the initial virus will appear. Advanced medical technology coupled with restrictive policies reduces the likelihood of such a scenario occurring.

When Africa is included in the sample, no marginal effect is significant. This suggests that in this case, the dominant factors explaining epidemic peaks may be purely biochemical or biophysical, molecular, etc., in nature. Human behaviour (including political decisions) would have no role to play. Here, we see the importance of the 'Africa effect' in modifying the results.

5 Conclusion

In this study, we were interested in the influence of socioeconomic factors on two key variables of the curves describing the propagation dynamics of COVID-19. On one hand, a parameter captures the speed at the start of the epidemic, and on the other hand, a second parameter reflects the inertia during the rise and fall of the disease. This decomposition provides interesting information, because it allows us to investigate two important determinants of the evolution of the number of cumulative cases. First, we want to know how long it takes for the disease to start. Second, once the disease has spread in the population, we are interested in knowing how long it takes for the disease to dissipate.

Previous studies in the literature on the effect of government anti-COVID measures usually conclude that this variable has a significant effect on the number of new cases.

Our regressions are useful to explain the reasons for the success of such policies; if they are effective in reducing the overall number of new infections, it is because they influence not only the epidemic peaks, but also the speed of spread of the disease in its early stages. However, the effect on the peak is likely to be overlooked if the situation in sub-Saharan African countries is not addressed separately.

The second important result of our research concerns the role of healthcare infrastructure. They are just as effective as anti-COVID policies, not only in preventing an epidemic from spreading too quickly at the outset, but also in creating the desired dynamic around peaks: slow spreading, followed by rapid disappearance.

This work was done using cross-sectional data. A possible extension would be

to consider the case in which the alpha and P parameters are time-varying. This temporal variability would be interesting to consider in order to capture different phenomena: resilience to COVID-19 over time due, for example, to the acquisition of herd immunity or vaccination policies. Even if this complicates the estimation, the difficulty lies in the socioeconomic variables, which are by definition structural and whose changes have very high inertia. For example, to evaluate the impact of changes in vulnerabilities linked to comorbidities on our parameters, it would be necessary to have epidemiological data over long periods, beyond that covered by the duration of successive waves of COVID. For this reason, a comparison by country or region is more appropriate.

Supplementary material

Supplementary material is available on the OUP website. These are the data and replication files and the online appendix.

Acknowledgements

The authors would like to thank the two anonymous referees, and Francis Teal, the Managing Editor, for their fruitful comments.

References

Adab, P., Haroon, S., O'Hara, M. and Jordan, R. (2022). Comorbidities and covid-19: better understanding is essential for health system planning. *BMJ* 377, doi:10.1136/bmj.o1431.

Amdaoud, M., Arcuri, G. and Levratto, N. (2021a). Are regions equal in adversity? A spatial analysis of spread and dynamics of COVID-19 in Europe. *The European Journal of Health Economics* 22: 629–642, doi:10.1007/s10198-021-01280-6.

Amdaoud, M., Arcuri, G. and Levratto, N. (2021b). Healthcare system and social trust in the fight against COVID-19: the case of France. *European Journal of Public Health* 31: 895–900, doi:10.1093/eurpub/ckab112.

Andrade, M., Achcar, J., Conceição, K. and Ravishanker, N. (2021). Time series regression models for covid-19 deaths. *Journal of Data Science* 19(2021): 269–292, doi:10.6339/21-JDS991.

Ascani, A., Faggian, A. and Montresor, S. (2020). The geography of COVID-19 and the structure of local economies: The case of italy. *Journal of Regional Science* 61: 407–441, doi:10.1111/jors.12510.

Atkeson, A. (2020). On using sir models to model disease scenatios for covid-19. *Federal Reserve Bank of Minneapolis Quarterly Review* 41: 1–35, doi:10.21034/qr.4111.

Barrios, J. M., Benmelech, E., Hochberg, Y. V., Sapienza, P. and Zingales, L. (2021). Civic capital and social distancing during the covid-19 pandemic. *Journal of Public Economics* 193: 104310, doi:10.1016/j.jpubeco.2020.104310.

Bartscher, A. K., Seitz, S., Siegloch, S., Slotwinski, M. and Wehrhöfer, N. (2021). Social capital and the spread of covid-19: Insights from European countries. *Journal of Health Economics* 80: 102531, doi:10.1016/j.jhealeco.2021.102531.

Bernadi, F. and Manuchehr, A. (2021). Epidemiology and the sir model: Historical context to modern applications. *CODEE Journal* 14.

Bigdelou, B., Sepand, M. R., Najafikhoshnoo, S., Negrete, J., Sharaf, M., S., J. H., Sullivan, I., Chauhan, P., Etter, M., Shekarian, T., , Liang, O., Esfandiarpour, G. H. andRahim and Zanganeh, S. (2022). Covid-19 and preexisting comorbidities: Risks, synergies, and clinical outcomes. *Front. Immunol.* 13, doi:10.3389/fimmu.2022.890517.

Borjas, G. (2020). Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods. Tech. rep., IZA Institute of Labor Economics Working Paper.

Boucekkine, R., Carvajal, A., Chakraborty, S. and Goenka, A. (2021). The economics of epidemics and contagious diseases: An introduction. *Journal of Mathematical Economics* 93, doi:10.1016/j.jmateco.2021.102498.

Bourdin, S., Jeanne, L., Nadou, F. and Noiret, G. (2021). Does lockdown work? A spatial analysis of the spread and concentration of COVID-19 in Italy. *Regional Studies* 55(7): 1185–1193, doi:10.1080/00343404.2021.1887471.

Braveman, P. and Gottlieb, L. (2014). The social determinants of health: it's time to consider the causes of the causes. *Public Health Reports* 129: 19–31, doi:10.1177/00333549141291S206.

Brodeur, A., Grigoryeva, I. and Kattan, L. (2021). Stay-at-home orders, social distancing, and trust. *Journal of Population Economics* 34: 1321––1354.

Capasso, V. (2008). *Mathematical structures of epidemic systels (2nd ed.)*. Lecture notes in biomathematics. Berlin Heildelberg: Springer.

Carroll, N. (2018). oglmx: Estimation of Ordered Generalized Linear Models. R package version 3.0.0.0.

Chen, C., Frey, C. B. and Presidente, G. (2021). Culture and contagion: Individualism and compliance with COVID-19 policy. *Journal of Economic Behavior & Organization* 190: 191–200, doi:https://doi.org/10.1016/j.jebo.2021.07.026.

Chowell, G., Sattenspiel, L., Bansal, S. and Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of life reviews* 18: 66–97, doi:10.1016/j.plrev.2016.07.005.

Chyon, F., Suman, M., Fahim, M. and Ahmmed, M. (2022). Time series analysis and predicting covid-19 affected patients by arima model using machine learning. *Journal of Virological Methods* 301, doi:10.1016/j.jviromet.2021.114433.

Doornik, J., Castle, J. and Hendry, D. (2021). Modeling and forecasting the covid-19 pandemic time-series data. *Social Sciences Quarterly* 102: 2070–2087, doi:10.1111/ssqu.13008.

Gallic, E., Lubrano, M. and Michel, P. (2021). Optimal lockdowns for COVID-19 pandemics: Analyzing the efficiency of sanitary policies in Europe. *Journal of Public Economic Theory* doi:10.1111/jpet.12556.

Maleki, M., Mahmoudi, M., Wraith, D. and Pho, K. (2020). Time series modelling to forecast the confirmed and recovered cases of covid-19. *Travel Medicine and Infectious Disease* 37, doi:10.1016/j.tmaid.2020.101742.

McNeely, J. (2021). Nature and covid-19: The pandemic, the environment, and the way ahead. *Ambio* 50: 767–781, doi:10.1007/s13280-020-01447-0.

Mongey, S., Pilossoph, L. and Weinberg, A. (2020). Which workers bear the burden of social distancing policies. Tech. rep., National Bureau of Economic Research Working Paper 27085.

Pell, B., Kuang, Y., Viboud, C. and Chowell, G. (2018). Using phenomenological models for forecasting the 2015 Ebola challenge. *Epidemics* 22: 62–70, doi:10.1016/j.epidem.2016.11.002.

Porwal, P., Shrivastava, P. and Iwari, S. (2015). Study of simple sir epidemic model. *Advances in Applied Science Research* 6: 1–4.

Ranard, B., Megjhani, M., Terilli, K., Doyle, K., Claassen, J., Pinsky, M., Clermont, G., Vodovotz, Y., Asgari, S. and Park, S. (2021). Identification of endotypes of hospitalized covid-19 patients. *Ambio* 8, doi:10.3389/fmed.2021.770343.

Rodrigues, H. (2016). Application of sir epidemiological model: new trends. *International Journal of Applied Mathematics and Informatics* 10: 92–97.

Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S. and Molter, A. (2020). Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society* 62: 102418, doi:10.1016/j.scs.2020.102418.

Schmitt-Grohé, S., Teoh, K. and Uribe, M. (2020). Covid-19: Testing inequality in New York City. *Covid Economics* 8: 27–43.

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., Kitayama, S., Mobbs, D., Napper, L. E., Packer, D. J., Pennycook, G., Peters, E., Petty, R. E., Rand, D. G., Reicher, S. D., Schnall, S., Shariff, A., Skitka, L. J., Smith, S. S., Sunstein, C. R., Tabri, N., Tucker, J. A., Linden, S. v. d., Lange, P. v., Weeden, K. A., Wohl, M. J. A., Zaki, J., Zion, S. R. and Willer, R. (2020). Using social and behavioural science to support covid-19 pandemic response. *Nature Human Behaviour* 4: 460–471, doi:10.1038/s41562-020-0884-z.

Viboud, C., Simonsen, L. and Chowell, G. (2016). A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks. *Epidemics* 15: 27 – 37, doi:10.1016/j.epidem.2016.01.002.

Wang, X., Wu, J. and Yang, Y. (2012). Richards model revisited: Validation by and application to infection dynamics. *Journal of Theoretical Biology* 313: 12–19, doi:10.1016/j.jtbi.2012.07.024.