NBER WORKING PAPER SERIES

BAYESIAN ESTIMATION OF EPIDEMIOLOGICAL MODELS: METHODS, CAUSALITY, AND POLICY TRADE-OFFS

Jonas E. Arias
Jesús Fernández-Villaverde
Juan Rubio Ramírez
Minchul Shin

Working Paper 28617
http://www.nber.org/papers/w28617

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2021

We thank Frank Diebold for useful comments. Some of the ideas in this paper came out of previous work with Chad Jones. John Cochrane was the first to point out to us the potential key importance of time-varying parameters in epidemiological models. Disclaimer: The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of Philadelphia, or the Federal Reserve System. Any errors or omissions are the responsibility of the authors. No statements here should be treated as legal advice. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Bayesian Estimation of Epidemiological Models: Methods, Causality, and Policy Trade-Offs
Jonas E. Arias, Jesús Fernández-Villaverde, Juan Rubio Ramírez, and Minchul Shin
NBER Working Paper No. 28617
March 2021
JEL No. C1,C5,I1

ABSTRACT

We present a general framework for Bayesian estimation and causality assessment in epidemiological models. The key to our approach is the use of sequential Monte Carlo methods to evaluate the likelihood of a generic epidemiological model. Once we have the likelihood, we specify priors and rely on a Markov chain Monte Carlo to sample from the posterior distribution. We show how to use the posterior simulation outputs as inputs for exercises in causality assessment. We apply our approach to Belgian data for the COVID-19 epidemic during 2020. Our estimated time-varying-parameters SIRD model captures the data dynamics very well, including the three waves of infections. We use the estimated (true) number of new cases and the time-varying effective reproduction number from the epidemiological model as information for structural vector autoregressions and local projections. We document how additional government-mandated mobility curtailments would have reduced deaths at zero cost or a very small cost in terms of output.

Jonas E. Arias
Federal Reserve Bank of Philadelphia
10 N Independence Mall W
Philadelphia, PA 19106
United States of America
jonas.arias@phil.frb.org

Jesús Fernández-Villaverde
Department of Economics
University of Pennsylvania
The Ronald O. Perelman Center
for Political Science and Economics
133 South 36th Street Suite 150
Philadelphia, PA 19104
and CEPR
and also NBER
jesusfv@econ.upenn.edu

Juan Rubio Ramírez
Emory University
Rich Memorial Building
Atlanta, GA 30322-2240
and Atlanta Federal Reserve Bank
and Fulcrum Asset Management
juan.rubio-ramirez@emory.edu

Minchul Shin
Research Department
Federal Reserve Bank of Philadelphia
Ten Independence Mall
Philadelphia, PA 19106
visiblehand@gmail.com

1 Introduction

We present a general framework for the Bayesian estimation of epidemiological models and their use for causality assessment and policy evaluation. Also, we show how to measure the trade-off between containing the spread of an epidemic and maintaining economic activity induced by non-pharmaceutical interventions, such as "shelter-in-place" orders. We do so in four steps. First, we write a generic epidemiological model using a state-space representation. This representation provides a familiar and intuitive notation to encompass a wide range of models, independently of the number of compartments, reinfection probabilities, and other features.

State-space representations deal efficiently with epidemiological models with time variation in the parameters controlling an infectious disease's dynamics. Time variation is critical: i) to incorporate changes in the behavior of individuals as they respond to the public health conditions (either voluntarily due to precautionary behavior or forced by government mandates); and ii) to include shifts in the transmission and clinical outcomes of the epidemic, such as virus variants, new medical treatments, or better logistical organization of social distancing. Thanks to points i) and ii), epidemiological models with time-varying parameters can account for the successive waves often observed in epidemic data. For example, as individuals observe rising death rates, they reduce their mobility, which lowers the number of effective contacts. After some periods, fewer contacts lead to falling death rates, and individuals revert to higher mobility, thus increasing effective contacts and death rates after a lag. Without time variation in parameters, one would be forced, in an attempt to fit the data, to depart from the class of standard epidemiological models (see Ho et al., 2021, for a similar argument).[1]

Another strength of the state-space representation is that it easily incorporates variables observed at different frequencies, from high-frequency and regular data on hospitalizations, deaths, and reported new cases to low-frequency and irregular data from seroprevalence studies. As our empirical application shows, this flexibility is vital while estimating epidemiological models as researchers might need to scramble disparate sources of data.

Second, and as the key step in our paper, we use sequential Monte Carlo techniques to evaluate the likelihood of the model while fully accounting for the essential nonlinearities and possible non-normalities present in the dynamics of epidemiological models. Doing so is necessary to capture correctly the fast variations in the spread of infectious diseases and the turning points of the different waves of an epidemic. Sequential Monte Carlos enjoy many advantages. More concretely, we will implement a particle filter that handles the nonlinearities and non-normalities

[1] We do not claim that these time-varying parameters are structural in the sense of being invariant to policy interventions à la Hurwicz (1966). We consider them only as behavioral parameters, which might be complex functions of structural parameters describing preferences, technology, and information processes. See Fernández-Villaverde and Rubio-Ramírez (2007) for a more detailed discussion of the distinction between structural and behavioral parameters. Alternatively, we could write a constant-parameters decision theory model, but that would move us away from the typical epidemiological models we consider.

in the model much more efficiently than older alternatives such as the extended Kalman filter, whose behavior deteriorates quickly when the nonlinearities of the model are severe, a common feature among many epidemiological models. The particle filter is easy to code and well-suited for implementation in massively parallel computational environments.

Third, we use Markov chain Monte Carlo techniques to sample from the posterior of the epidemiological model. In that way, we can implement a fully Bayesian approach. In our context, the Bayesian approach allows incorporating information in the prior of the parameters from laboratory results, clinical studies, seroprevalence surveys, and the experience of other regions or countries. For example, the average duration of a spell at a recovering compartment of an inhabitant of Belgium (the country we will use for our empirical application) is likely to be close to the average duration of a spell at a recovering compartment of an inhabitant of neighboring countries. If we have high-quality estimates of the latter, we can build an informative prior for the former. The Bayesian choice is also attractive because it quickly gauges the range of uncertainty existing in the data by looking at the whole posterior, instead of focusing on the maximum likelihood point estimate or relying on simulation errors as in Li et al. (2020). Furthermore, by evaluating the marginal likelihood, the Bayesian approach allows for efficient model comparison.[2]

Fourth, we show how to use outputs from the epidemiological model to answer critical policy questions involving causality in a time-series environment. Let us illustrate this point with an example. A key element in measuring the effect of a government mandate curtailing mobility, such as "shelter-in-place" orders, is to have access to a reliable sample of new infectious cases. However, such data might not exist or be subject to large and biased measurement error due to problems like testing constraints, unwillingness to test, etc. (this point was recognized long ago in the epidemiological literature; for example, see O'Neill and Roberts, 1999). To complicate matters, the biased error in the measured new cases might not be constant over time (for instance, as testing becomes more accessible), and simple corrections (such as scaling up the number of recorded cases by some factor) would produce flawed results.[3] Similarly, recorded death due to the disease might be different from the true deaths, as miscounting and classification errors are likely, especially at the start of an epidemic or during peak moments.

An epidemiological model can take observations of recorded new cases, recorded hospitalizations, recorded new deaths in hospitals and at home, and from irregular seroprevalence studies and impose the discipline brought by the cross-equation restrictions determined by the model's

[2]Also, a Bayesian approach can help with weak identification, a common problem in epidemiological models (Alahmadi et al., 2020; Korolev, 2021). This advantage is not so much a consequence of the informativeness of the prior (we could always rely on flat priors), but because Bayesian inference can integrate over the whole posterior to obtain important outputs for policy analysis, such as a smoothed estimate of the effective reproduction number, even if portions of the parameters' posterior are flat.

[3]Nonetheless, see the more sophisticated approaches in Manski and Molinari (2021) and Toulis (2021) to estimate COVID-19 prevalence based on partial identification.

dynamics. In such a way, we can transform, for example, a series of recorded new cases into a series of estimated true new cases and use the latter as an input to ascertain the effects of a "shelter-in-place" order using standard techniques such as structural vector autoregressions (SVARs) and local projections (LPs). In comparison, a reduced-form model –such as those described in Gostic et al. (2020)– that does not use the structure of an epidemiological model cannot build on the strengths of the cross-equation restrictions and the biological and clinical information they bring to the estimation.[4] Interestingly, this methodology goes well beyond epidemiological models: it can be applied to any causal inference problem that relies on obtaining good estimates of unobserved variables.[5]

We illustrate these four steps using data from the COVID-19 epidemic in Belgium during 2020. While this application is of interest in itself, we must emphasize that our methods are much more general than the details of the model we specify or the data we use. As we mentioned above, they are applicable to a wide range of epidemiological models, other diseases beyond COVID-19, and alternative techniques to ascertain causality in time series. In that sense, our application should be read as being representative of the range of exercises that can be performed.

We pick Belgium for five reasons. First, it is one of the countries that has suffered the most from the COVID-19 epidemic. As of March 13, 2021, it is the fourth territory globally with the highest COVID-19 deaths per capita, behind only tiny Gibraltar and San Marino and the larger Czechia. Second, Belgium is a small country, 30,689 km^2 (roughly 20% larger than Maryland), and geographically and climatically quite homogeneous in the area where most of the population is concentrated (that is, in the coastal plain and central plateau outside the Ardennes). That allows us to consider Belgium as a single unit for our analysis.[6] Third, as mentioned above, we will use the cross-equation restrictions to discipline the estimated true new cases and deaths. Belgium has high-quality national data, including reported new cases, hospitalizations, deaths in hospitals and at home, and several national seroprevalence studies that allow us to efficiently implement those cross-equation restrictions. Fourth, Belgium experienced, within our sample, three waves of the epidemic, which would demonstrate how our methods can handle intricate data patterns. Fifth, the Belgian government has passed different mandates curtailing mobility on several occasions, which will give us the identification of the effects of these mandates.

We postulate a SIRD model of COVID-19 for Belgium with time variation in the effective

[4]The cross-equations discipline could come at a cost if the model is misspecified: the resulting smoothed estimates might be biased. Since our framework is general enough to encompass many epidemiological models, we could detect misspecification by comparing our SIRD model against flexible reduced-form time-series models such as the one proposed by Ho et al. (2021).

[5]In a different environment, Gilchrist and Zakrajšek (2012) follow a similar approach by using a pricing model to recover a measure of sentiment in the corporate credit market. They show, using reduced-form regressions and SVARs, how such a measure can affect the economy and asset prices.

[6]The incidence of COVID-19 has been around 10% higher in Wallonia than in Flanders in per capita terms (with Flanders having a worst first wave and Wallonia a more damaging third wave) and about average in Brussels. To keep our analysis as transparent as possible, we can ignore these relatively small regional differences.

contact rate among individuals in the model, the hospitalization rates, and death probabilities. The first variation embodies the idea that individuals react to the prevailing health conditions and "shelter-in-place" orders. The second and third variations incorporate changes in hospital capacity and improvements in medical protocols (and possible changes in the epidemic's demographics not fully added to the model). We let this variation evolve as random walks, with the likelihood function telling us about the most likely innovations to them.

We estimate the model using data from Sciensano, a public institution in Belgium, of COVID-19 per capita deaths in hospitals and at home, per capita hospitalizations due to COVID-19, the observed per capita new cases, and the seroprevalence rates. The sample starts on March 15, 2020, the first day with available data on COVID-19 hospitalizations, and ends on November 30, 2020, the latest available data when we estimated the model. We show how the data are informative about the parameters of the model, how the point estimates are in line with other evidence, and how the model fits the data very well (including tracking the three waves of the epidemic). Also, we recover the smoothed states such as time-varying reproduction numbers, time-varying death rates, and new cases.

Next, we illustrate how we can use these outcomes from the estimated epidemiological model to answer policy-relevant questions, such as the effectiveness of mobility-curtailing policies in controlling the epidemic's spread and its lethality and in measuring their cost in terms of output. Identifying the causal effects of these policies is challenging because mobility, as measured by indexes such as the Google COVID-19 Community Mobility Reports, changes due to both voluntary behavior and government orders.

To get around this problem, we rely on the two most salient methods for assessing causality in a time-series context: SVARs and LPs. We use an SVAR to identify a government stringency shock (such as a "shelter-in-place" order). We use the LP approach to analyze how a reproduction shock (such as the spread of a new, more contagious variant of the SARS-CoV-2 virus) affects the rest of the variables depending on the level of government mobility limitations. However, for SVARs and LPs to deliver reliable answers, we need accurate inputs for their estimation. Some of those inputs, such as the effective reproduction number, are not directly observed. Others, such as new cases, are observed subject to large, biased, and time-varying measurement errors. As highlighted above, our estimated epidemiological model can deliver those inputs and, therefore, be a vital piece for SVARs and LPs and other causality-assessment exercises.

Using the SVAR, we estimate that a positive stringency shock —normalized such that, upon impact, the posterior median increase is equivalent to a one-unit increase in one of the ordinal components of the Oxford Stringency Index of mobility curtailments— leads to roughly $1,000$ fewer deaths in Belgium after 2 months, or around 6% fewer deaths in the sample at a negligible cost in terms of output. In fact, our point estimate implies a small economic gain of €4.2 per capita (although it is hard to distinguish it from zero cost or a small cost). The intuition is that,

by controlling the epidemic, a positive stringency shock brings higher economic activity after a few weeks that more than compensate for the very short-run losses.

Using the LP, we find that a high government stringency —i.e., a level of stringency above the median in our sample— could save up to about 250 deaths in the first two weeks after the reproduction shock, or around 1.5% of the deaths in our sample, compared with a low government stringency, at the small cost of between €2 and €4 per capita.

The rest of the paper is organized as follows. Section 2 frames our paper within the literature. Section 3 presents a general framework for the Bayesian estimation of epidemiological models. Section 4 presents a time-varying model of the COVID-19 epidemic. Section 5 introduces the data from Belgium for our application. Section 6 reports our results and Section 7 shows how to exploit these results for causality and policy trafe-off analysis with SVARs and LPs. Section 8 concludes.

2 Literature Review

Our paper makes a contribution to the literature estimating epidemiological models and to the emerging literature studying the causal effects of health policy.

Bayesian approach The use of a Bayesian approach in epidemiological models became popular after the contributions of O'Neill and Roberts (1999). The state-of-the-art methods for Bayesian analysis of infectious diseases are reviewed by Alahmadi et al. (2020) and Broemeling (2021). New research in the area includes approximate Bayesian Computation for epidemic models by Kypraios et al. (2017) and Minter and Retkute (2019), importance sampling-based Bayesian inference by Black (2019) and Li et al. (2020). That last paper, the closest to our approach, uses an ensemble adjustment Kalman filter to estimate with maximum likelihood an epidemiological model for 375 Chinese cities. We improve on this approach along two key dimensions. First, the particle filter is more powerful than variations of the Kalman filter when dealing with highly nonlinear models, like the ones in epidemiology. Second, we can assess the whole uncertainty on the posterior instead of relying in simulation errors to build credible intervals for the maximum likelihood estimates as Li et al. (2020) do. This will be important when we evaluate the infectious period (a central component of the model), which we conclude has a much wider range of likely values than their tight estimate.

Black (2019) and Walker, Black, and Ross (2019) propose sampling algorithms to perform Bayesian inference when one wants to exactly match epidemiological observations available to state variables. While these are useful techniques, the COVID-19 epidemic is characterized by the misreporting issue mentioned above. Walker, Ross, and Black (2017) propose two methods —exact Bayesian inference using data-augmentation and approximate Bayesian inference— for

Bayesian inference of within-household transmission, recovery, and between-household transmission using data from the first few hundred studies (a data collection process that occurs at the onset of a pandemic influenza outbreak). Their approach is valuable for computing the effective reproductive number in the early stages of an epidemic, but as acknowledged by the authors, it relies on the assumption of perfect detection of infectious cases.

More recently, Atkeson et al. (2020) use a Bayesian approach to estimate the trend growth of daily deaths by assuming that it follows a mixture of Weibull density functions. Then, they recover the time-varying effective reproduction number combining their estimated reduced-form model for deaths and the restrictions implied by a SIR model. Arnon et al. (2020) estimate the time-varying effective reproduction number by using the Bayesian method developed by Cori et al. (2013), but they do not evaluate the likelihood or estimate the behavorial parameters of the model. Bognanni et al. (2020) estimate the location-specific parameters of infection rates per excursion and present the cost of infection of a spatial economic-SAIRD model using a Bayesian approach similar to ours, but they do not formulate a general framework to encompass abstract epidemiological models. Dehning et al. (2020) estimate a Bayesian SIR model to infer change points in the spread of COVID-19, but they can only evaluate a very simple likelihood.

Frequentist approach Some papers estimate a similar compartmental model under the frequentist paradigm. For example, Toda (2020) and Korolev (2021) estimate a compartmental model by minimizing a loss function measured by the distance between the model-implied share of new cases and the reported new cases. Hortaçsu et al. (2021) estimate the fraction of unreported infections in epidemics with a known epicenter by utilizing the covariation in initial reported infections across regions and the number of travelers to these regions from the epicenter. Our approach requires much less granular information. Renne et al. (2020) estimate a SIRD model using a quasi-likelihood methodology and the extended Kalman filter. Arroyo-Marioli et al. (2021), Fernández-Villaverde and Jones (2020), Pesaran and Yang (2020), and Lee et al. (2021) estimate time-varying contact rates using the restrictions implied by the SIR model. Avery et al. (2020, Table 1) list a dozen COVID-19 predictive epidemiological models, but most of them either rely on simple curve fitting or the formal estimation takes a secondary role to granularity in location details.

Our main methodological innovation relative to the previous papers is to show how to combine sequential Monte Carlo methods with a general epidemiological framework that simultaneously deals with under-reporting cases and includes data on new cases, hospitalizations, deaths in hospitals and at home, as well as serological studies and time-varying parameters. Importantly, in addition to its usefulness for establishing causal inference (to be discussed below), our paper is potentially useful for policymaking. For example, our smoothed estimates about the share of individuals recovering in hospitals and at home can be applied to planning purposes such as

new construction of hospitals or the optimal design of lockdowns, both in terms of stringency and duration (as in Acemoglu et al., 2020).

Causal effects of policy on health outcomes and behavior Chernozhukov, Kasahara, and Schrimpf (2021) quantify the causal effects of masks, policies, and behavior on new cases and death outcomes using a structural equations model based on causal diagrams. Since this approach requires the use of the true cases, it dovetails perfectly with the outputs of our estimated SIRD model. The same can be said of the growing literature on the effects of policy on behavior as measured by Google COVID-19 Community Mobility Reports, e.g., Abouk and Heydari (2021), Maloney and Taskin (2020), Andersen (2020), and Wilson (2020). Since all of these papers rely on reported cases, the outputs of our estimated SIRD model can be most valuable for their estimations by correcting the biases in those reported cases.

Hsiang et al. (2020) and Courtemanche et al. (2020) use a reduced-form model to gauge the effects of policies on the growth rate of infections. While reduced-form approaches help in understanding properties of the data, the lessons from modern empirical studies on the propagation of structural shocks indicate that a structural model is essential to establish causal inference.

In summary, one of our paper's main contributions is to combine the output from our estimated model with state-of-the-art causal inference methods to establish how mobility curtailments affect the transmission rates and deaths of COVID-19. More in general, we offer an innovative approach for conducting causal inference in situations where obtaining good estimates of unobserved variables is critical for establishing causality.

3 A General Framework

We use the well-known state-space representation of a dynamic model to encompass a large class of epidemiological models (SIRD, SIS, etc.). We rely on this representation because it is convenient for filtering, smoothing, forecasting, and estimation. Section 4 will show how the abstract notation on the next pages works in an application dealing with COVID-19.

The state-space representation consists of a transition equation, governing the evolution of the states of the model, and a measurement equation, linking those states with the observed variables. The transition equation of a general epidemiological model takes the form:

$$\boldsymbol{X}_t = f(\boldsymbol{X}_{t-1}, \boldsymbol{\mathcal{E}}_t; \boldsymbol{\Theta}) \tag{1}$$

where the vector $\boldsymbol{X}_t$ stacks all the states of the model (share of the population in each of the different compartments of the model in period t, hospital and ICU bed occupancy rates,

government policy, the season of the year, etc.), $\boldsymbol{\mathcal{E}}_t$ is a vector of innovations, and $\boldsymbol{\Theta}$ is a vector of parameters indexing the function $f(\cdot)$ and the distributions of $\boldsymbol{\mathcal{E}}_t$ and $\boldsymbol{V}_t$ (to be defined below). We use capital letters to denote random variables and lower case letters for the realizations of those random variables, a straight bold font for the vector of parameters, and, for all other vectors, an italic bold font. Vectors can always be interpreted, if needed, as being 1×1.[7]

We highlight five points regarding the transition equation. First, the Markov structure of Equation (1) is not a tight restriction because we can redefine the vector $\boldsymbol{X}_t$ to include additional lags of the relevant state variables. Second, the vector $\boldsymbol{X}_t$ can include state variables indexed by location (i.e., susceptible population in location j) and individual characteristics (i.e., susceptible population over 65) or any other combination of interest. Third, we are not forcing any restriction in the function $f(\cdot)$. In general, this function will be nonlinear and may include threshold effects or jumps. Fourth, at this level of abstraction, we are not imposing any distributional assumption on $\boldsymbol{\mathcal{E}}_t$, and it might include non-Gaussian innovations. Fifth, we will include in $\boldsymbol{X}_t$ all the time-varying parameters of the model that capture the notion that the behavior of individuals, government policies, and medical technology (broadly construed) can change over time. We will reserve $\boldsymbol{\Theta}$ for the parameters that are time-invariant. That is, if the mortality rate changes over time (for example, due to better clinical protocols), the mortality rate at period t will be a state of the model, while the parameters of the law of motion of the mortality rate will belong to $\boldsymbol{\Theta}$.

The measurement equation of a general epidemiological model can be written as:

$$\boldsymbol{Y}_t = g(\boldsymbol{X}_t, \boldsymbol{V}_t; \boldsymbol{\Theta}) \tag{2}$$

where $\boldsymbol{Y}_t$ is a vector of observables (cases, deaths, hospitalizations, ...) and $\boldsymbol{V}_t$ is a vector of innovations to observables. As before, we are not imposing a functional form on $g(\cdot)$ or a distribution assumption on $\boldsymbol{V}_t$. In particular $g(\cdot)$ can be, along some dimensions, an identity (i.e., we observe one or several states of the model).

Two aspects of Equation (2) deserve further discussion. First, the observable variables can come at different frequencies, with some components of $\boldsymbol{Y}_t$ being the empty set for some periods. This feature is desirable as some data may come daily (e.g., new hospitalizations and deaths), others may be more informative weekly (e.g., reported new cases), or some might be gathered only sporadically (e.g., seroprevalence surveys). The latter would be the case in our application in Section 4. Second, $\boldsymbol{V}_t$ might be a shock (e.g., an unexpectedly high level of deaths given the number of recovering individuals) or a measurement error (e.g., cases are under-reported due to

[7] In this paper, we will deal with epidemiological models in discrete time. Since we will be taking these models to the data, which comes in discrete units of time, this is a more natural framework than continuous time, which is often more convenient for mathematical analysis. We could rework our framework into continuous time with some extra notation.

an insufficient testing).

Combining Equations (1) and (2), we get:

$$\mathbf{Y}_t = g(f(\boldsymbol{X}_{t-1}, \boldsymbol{\mathcal{E}}_t; \boldsymbol{\Theta}), \boldsymbol{V}_t; \boldsymbol{\Theta}) = h(\boldsymbol{X}_{t-1}, \boldsymbol{\mathcal{E}}_t, \boldsymbol{V}_t; \boldsymbol{\Theta}),$$

which shows that, conditional on the states $\boldsymbol{X}_{t-1}$, $\boldsymbol{Y}_t$ is a change of variables of $\boldsymbol{\mathcal{E}}_t$ and $\boldsymbol{V}_t$ and that the distributions of these innovations induce a distribution for $\boldsymbol{Y}_t$: $p\left(\boldsymbol{Y}_t|\boldsymbol{X}_{t-1}; \boldsymbol{\Theta}\right)$.

Unfortunately, since the functions $f(\cdot)$ and $g(\cdot)$ are, in standard epidemiological models, nonlinear, even if the distributions of $\boldsymbol{\mathcal{E}}_t$ and $\boldsymbol{V}_t$ were to belong to well-known parametric families (e.g., Gaussian), $\boldsymbol{Y}_t$ would not follow any known distribution (except in nongeneric situations).

Given a sample of observables over T periods, $\boldsymbol{y}^T = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T\}$, our goal is to evaluate the likelihood function of $\boldsymbol{y}^T$ given $\boldsymbol{\Theta}$, i.e., $p\left(\boldsymbol{y}^T; \boldsymbol{\Theta}\right)$. We can write:

$$p\left(\boldsymbol{y}^T; \boldsymbol{\Theta}\right) = p\left(\boldsymbol{y}_1|\boldsymbol{\Theta}\right) \prod_{t=2}^{T} p\left(\boldsymbol{y}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right).$$

Then:

$$p\left(\boldsymbol{y}^T; \boldsymbol{\Theta}\right) = \int p\left(\boldsymbol{y}_1|\boldsymbol{X}_1; \boldsymbol{\Theta}\right) p(\boldsymbol{X}_1; \boldsymbol{\Theta}) d\boldsymbol{X}_1 \prod_{t=2}^{T} \int p\left(\boldsymbol{y}_t|\boldsymbol{X}_t; \boldsymbol{\Theta}\right) p\left(\boldsymbol{X}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right) d\boldsymbol{X}_t. \tag{3}$$

In other words, knowledge of the sequence $\left\{p\left(\boldsymbol{X}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right)\right\}_{t=1}^{T}$ allows the evaluation of the likelihood in Equation (3) of the model (given the ability to solve or approximate the corresponding integrals). Also, notice that $\boldsymbol{y}^0$ is the empty set and, therefore, $p(\boldsymbol{X}_1; \boldsymbol{\Theta}) = p\left(\boldsymbol{X}_1|\boldsymbol{y}^0; \boldsymbol{\Theta}\right)$.

The sequence $\left\{p\left(\boldsymbol{X}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right)\right\}_{t=1}^{T}$ can be found with filtering, by solving the Chapman-Kolmogorov equation:

$$p\left(\boldsymbol{X}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right) = \int p\left(\boldsymbol{X}_t|\boldsymbol{X}_{t-1}; \boldsymbol{\Theta}\right) p\left(\boldsymbol{X}_{t-1}|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right) d\boldsymbol{X}_{t-1} \tag{4}$$

and applying Bayes' theorem:

$$p\left(\boldsymbol{X}_t|\boldsymbol{y}^{t}; \boldsymbol{\Theta}\right) = \frac{p\left(\boldsymbol{y}_t|\boldsymbol{X}_t; \boldsymbol{\Theta}\right) p\left(\boldsymbol{X}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{y}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right)} \tag{5}$$

where:

$$p\left(\boldsymbol{y}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right) = \int p\left(\boldsymbol{y}_t|\boldsymbol{X}_t; \boldsymbol{\Theta}\right) p\left(\boldsymbol{X}_t|\boldsymbol{y}^{t-1}; \boldsymbol{\Theta}\right) d\boldsymbol{X}_t \tag{6}$$

is the marginal distribution of $\boldsymbol{y}_t$ given $\boldsymbol{y}^{t-1}$. We assume that $\boldsymbol{\Theta}$ includes $\boldsymbol{X}_0$. Hence, we are computing the likelihood conditional on those initial states.

Since this filtering problem is standard, we will not spend much time addressing it. Fernández-

Villaverde et al. (2016) review this topic in detail. Suffice it to say that we can use the particle filter or other sequential Monte Carlo methods to efficiently simulate swarms of particles representing possible alternative values of $\boldsymbol{X}_t$. Relying on a law of large numbers, we can substitute the integrals in Equations (3), (4), (5), and (6) by averages of the corresponding probabilities evaluated at the simulation. Beyond their applicability to an extensive class of filtering problems, sequential Monte Carlo methods have the additional advantages of being easy to code and well suited for massive parallelization. Thus, in this paper, we will use the particle filter to evaluate the likelihood function.[8]

Alternatives to sequential Monte Carlo methods such as the extended Kalman filter or the unscented Kalman filter are unlikely to work with high enough accuracy in epidemiological models because of the strong nonlinearities present in these models. Examples, in other fields, where a well-designed particle filter clearly outperforms variations of the Kalman filter in head-to-head comparisons include Chatzi and Smyth (2009) and Lee et al. (2010). Also, Smyth et al. (1999) document the poor performance of the extended Kalman filter in highly nonlinear setups.

Once we have the likelihood function, $p\left(\boldsymbol{y}^T;\boldsymbol{\Theta}\right)$, we can either maximize it or combine it with a prior $p\left(\boldsymbol{\Theta}\right)$ to obtain a posterior $p\left(\boldsymbol{\Theta}|\boldsymbol{y}^T\right) \propto p\left(\boldsymbol{y}^T;\boldsymbol{\Theta}\right)p\left(\boldsymbol{\Theta}\right)$ and perform a Bayesian analysis. We can rely on clinical and experimental evidence to build informative priors.[9]

The posterior $p\left(\boldsymbol{\Theta}|\boldsymbol{y}^T\right)$ can be sampled using a Markov chain Monte Carlo such as the Metropolis-Hastings or the Hamiltonian Monte Carlo. When we combine the particle filter to evaluate the likelihood with a Metropolis-Hastings to draw from the posterior, the joint algorithm is often known as the PFMH. We follow this approach in this paper. One central object of interest in our exercise will be the smoothed sequence of states conditional on the posterior mode, since it will inform us about the epidemiological situation at any given moment, a key input for policymakers and our causal assessment of mobility curtailments.

With the output of the Markov Chain Monte Carlo, we can integrate the likelihood function $p\left(\boldsymbol{y}^T;\boldsymbol{\Theta}\right)$ with respect to the parameters $\boldsymbol{\Theta}$. The resulting marginal likelihood $p\left(\boldsymbol{y}^T\right)$ can be used to build Bayes factors to compare different epidemiological models (for example, to determine the best number of compartments). See Llorente et al. (2020) for an updated review of methods for computing the marginal likelihood.

[8]There is an associated smoothing problem where we compute the sequence $\left\{p\left(\boldsymbol{X}_t|\boldsymbol{y}^T;\boldsymbol{\Theta}\right)\right\}_{t=1}^{T}$, that is, the probability distribution of states conditional on *all* the data. In general, future observations are informative about current states (as we will show in our application).

[9]If we adopt a uniform prior $p\left(\boldsymbol{\Theta}\right)$, we can draw from a posterior that is proportional to the likelihood. By looking at the parameter values $\boldsymbol{\Theta}^*$ that deliver the maximum value of $p\left(\boldsymbol{y}^T;\boldsymbol{\Theta}\right)$ in the Markov chain, we would have obtained a maximum likelihood estimate that is often more reliable than the one that comes from traditional optimization algorithms. We can also use $\boldsymbol{\Theta}^*$ as an (excellent) initial guess for an optimization algorithms.

4 Our Application

As our application, we select a SIRD compartmental model of COVID-19 to be estimated with data from Belgium. The application will highlight the importance of allowing for time variation in i) the effective contact rate among individuals; ii) the hospitalization rates; and iii) death probabilities.

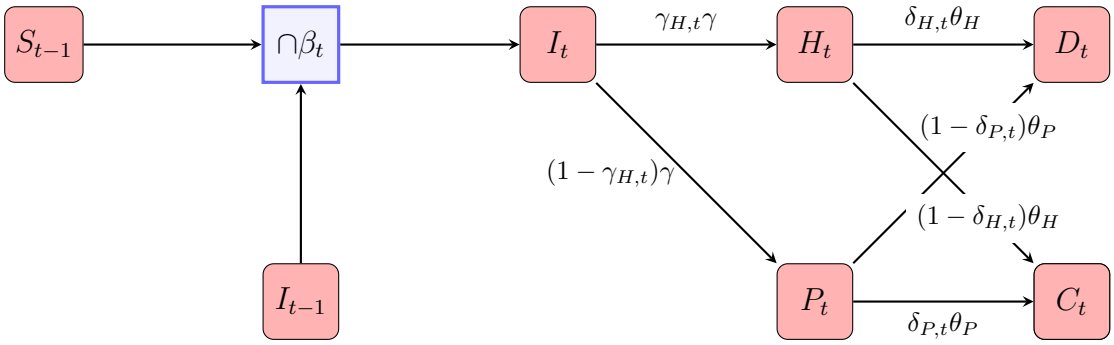


Figure 1: Outline of the model.

Figure 1 outlines the compartments' structure. All individuals start as susceptible (S) except for a small fraction of infectious individuals (I). Susceptible individuals meet with infectious individuals and become, with some probability, infectious themselves. This probability is controlled by β_t, the time-varying effective contact rate. Every day, a share γ of infectious individuals can recover in a hospital (H) or at home (P). Here, "home" means all dwellings, including private residences and retirement communities, outside of hospitals (unfortunately, our data set does not distinguish between private residences and retirement communities, a potentially important distinction). Of those, the share of individuals recovering in hospitals is time varying and equal to $\gamma_{H,t}$. Hence, the share $1-\gamma_{H,t}$ recovers at home. Recovery ends with death (D) or a return to a healthy status (C). Every day a share θ_H (θ_P) of individuals in a hospital (at home) recover. Of those, $\delta_{H,t}$ ($\delta_{P,t}$) die, while $1-\delta_{H,t}$ ($1-\delta_{P,t}$) return to healthy status. These death rates are also time varying.

We assume that, once cured, an individual cannot become infectious again. We pick this specification because the evidence is that re-infections with SARS-CoV-2 during 2020 (our data sample) were possible but unlikely. However, it would be easy to extend the model to allow for re-infections, perhaps after the (stochastic) waning of immunity or the arrival of new varieties of the virus, a common concern later in the epidemic. Indeed, our causal analysis in Section 7 treats the emergence of new variants as a possible interpretation of a reproduction shock.

We move now to the detailed presentation of the model.

4.1 The Transition Equation

The transition equation of the model we will estimate is:

$$\begin{pmatrix} S_t \\ I_t \\ H_t \\ P_t \\ b_t \\ g_{H,t} \\ d_{H,t} \\ d_{P,t} \\ n_t \end{pmatrix} = \begin{pmatrix} S_{t-1} \\ I_{t-1} \\ H_{t-1} \\ P_{t-1} \\ b_{t-1} \\ g_{H,t-1} \\ d_{H,t-1} \\ d_{P,t-1} \\ n_{t-1} \end{pmatrix} + \begin{pmatrix} -\beta_t S_{t-1} I_{t-1} \\ \beta_t S_{t-1} I_{t-1} - \gamma I_{t-1} \\ \gamma_{H,t} \gamma I_{t-1} - \theta_H H_{t-1} \\ (1-\gamma_{H,t})\gamma I_{t-1} - \theta_P P_{t-1} \\ \sigma_b \varepsilon_{b,t} \\ \sigma_g \varepsilon_{g,t} \\ \sigma_h \varepsilon_{h,t} \\ \sigma_p \varepsilon_{p,t} \\ \sigma_n \varepsilon_{n,t} \end{pmatrix}, \tag{7}$$

where we have a 9×1 vector of states $\boldsymbol{X}_t = \{S_t, I_t, H_t, P_t, b_t, g_{H,t}, d_{H,t}, d_{P,t}, n_t\}$ and a 5×1 vector of innovations $\boldsymbol{\mathcal{E}}_t = \{\varepsilon_{b,t}, \varepsilon_{g,t}, \varepsilon_{h,t}, \varepsilon_{p,t}, \varepsilon_{n,t}\}$. Let us describe each of these variables.

Row one of Equation (7) tells us that the share of the population that is susceptible on day t, S_t, is equal to the share of the susceptible population the day before, S_{t-1}, minus the new infections as a share of the population, given by a matching function of the share of susceptibles and infectious, I_{t-1}, yesterday: $\beta_t S_{t-1} I_{t-1}$.

The most relevant feature of row one, and a central aspect of our model, is that the effective contact rate, β_t, in the matching function is time varying. In many textbook epidemiological models, the effective contact rate is a constant parameter. However, individuals respond to the epidemiological situation for two reasons. First, individuals take voluntary precautionary measures (lower mobility, wearing personal protection equipment, changed business protocols). For this point to hold, we do not need to assume full agent rationality; we only require some degree of endogenous reaction. Second, individuals change how often they effectively interact with each other in any given period because governments impose non-pharmaceutical interventions (NPIs) in response to the health situation, such as curtailments on businesses and travel or mandatory mask-wearing. By letting the effective contact rate be time varying, the model can capture these two mechanisms.

At this point, we could take two routes. First, we could introduce an explicit decision-theory model of the individual and policy responses. This route would force us to impose a considerable degree of additional structure. Alternatively, we can specify a flexible form for the evolution of β_t and let the data inform us about it.

We chose this second route by assuming that $b_t = \log(\beta_t)$ follows a random walk. More concretely, in row five of Equation (7), $b_t = b_{t-1} + \sigma_b \varepsilon_{b,t}$, where the innovation is a truncated

standard normal distribution:

$$\varepsilon_{b,t} \mid \beta_{t-1}, I_{t-1} \sim \mathcal{N}(0, 1, \text{lb}(\boldsymbol{X}_t, \boldsymbol{y}_t), -\sigma_b^{-1} (\log(I_{t-1}) + \log(\beta_{t-1}))).$$

The lower bound of the truncation, $\text{lb}(\boldsymbol{X}_t, y_t)$, is imposed to avoid having more new cases in the model $(S_{t-1} - S_t)$ than in the data (notice that we are defining the function $\text{lb}(\boldsymbol{X}_t, y_t)$ implicitly). Later, we will argue that there will be a percentage of true cases of COVID-19 that are not reported (for instance, because the cases are asymptomatic or due to insufficient testing), but that the situation where there are *more* reported cases than true cases is not relevant empirically. The number of "false positives" is most likely trivially small and swamped, by at least an order of magnitude, by the under-reporting of "false negatives."[10] The upper bound of the truncation, $-\sigma_b^{-1} (\log(I_{t-1}) + \log(\beta_{t-1}))$, prevents S_t from becoming negative.

Row two of Equation (7) describes how the share of infectious, I_t, evolves. The new share is equal to the share yesterday, I_t, plus the new infections $\beta_t S_{t-1} I_{t-1}$, minus the share of infectious that move to the next compartment, γI_{t-1}.

Row three of Equation (7) determines the evolution of H_t, the share of the population hospitalized on day t. The share evolves through inflows from the compartment of infectious at a rate $\gamma_{H,t}\gamma I_{t-1}$ and outflows at a rate θ_H. We allow $\gamma_{H,t}$ to move to reflect the changing availability of hospital beds and shifting decisions by patients (e.g., should I go to a hospital?) and clinical protocols (e.g., should this patient be hospitalized or sent back home?).

As we did with the effective contact rate, we specify a flexible law of motion for $\gamma_{H,t}$. Specially, in row six of Equation (7), we define a random walk for $g_{H,t}$, $g_{H,t} = g_{H,t-1} + \sigma_g \varepsilon_{g,t}$, where $\varepsilon_{g,t} \sim \mathcal{N}(0, 1)$ and $\gamma_{H,t} = \frac{e^{g_{H,t}}}{e^{g_{H,t}}+1}$. Hence, the mapping from $\gamma_{H,t}$ to $g_{H,t}$ is:

$$g_{H,t} = \log\left(\frac{\gamma_{H,t}}{1 - \gamma_{H,t}}\right).$$

This mapping will be useful for interpreting $g_{H,0}$ later on.

Row four of Equation (7) governs the evolution of the share of individuals recovering at home, P_t. In parallel to H_t, the share evolves through inflows from the compartment of infectious at a rate $(1 - \gamma_{H,t})\gamma I_{t-1}$ and outflows at a rate θ_P. Rows three and four implicitly assume that individuals recovering in a hospital or at home do not switch between these two compartments. This assumption is due to data limitations (as we do not observe how many patients previously in a hospital die while recovering at home). Fortunately, this assumption does not cause too many problems. If a patient is at home but later hospitalized, we can consider her as still

[10] Many clinical tests are designed to minimize "false positives" at the cost of "false negatives." Nonetheless, we could change the lower bound to allow for more reported cases than true cases (this might be helpful in other applications). In a robustness analysis, we checked for the possibility of more cases reported than true ones, and we only got a worse fit of the model.

being in the infectious compartment. Conversely, individuals formerly hospitalized and currently recovering at home are unlikely to be infectious and, hence, they could be considered as being in the recovered compartment. Also, the measurement error will help us tackle, empirically, deviations from this assumption in the data.

Rows seven to nine of Equation (7) describe the evolution of three random walks that will be used in the measurement equation below, with innovations $\varepsilon_{i,t} \sim \mathcal{N}(0,1)$ for $i \in \{h, p, n\}$.

Finally, notice that the rows of Equation (7) define an implicit law of motion of C_t because:

$$S_t + I_t + H_t + P_t + D_{H,t} + D_{P,t} + C_t = 1, \text{ with } C_0 \geq 0. \tag{8}$$

Hence, $C_t = C_{t-1} + (1 - \delta_{H,t})\theta_H H_{t-1} + (1 - \delta_{P,t})\theta_p P_{t-1}$.

4.2 The Measurement Equation

In our data set, we observe deaths in a hospital ($D^{obs}_{H,t}$), deaths at home ($D^{obs}_{P,t}$), and hospitalized patients (H^{obs}_t) at a daily frequency; new cases (G^{obs}_t) at a weekly frequency; and the point estimates of S_t from the seroprevalence surveys for some periods (see Section 5 for details). We assume that the log of each of these five variables is measured with a normally distributed error $u_{i,t} \sim \mathcal{N}(0,1)$ for $i \in \{D_H, D_P, H, G, S\}$. The log form ensures that the model only predicts positive values for the five observables. The measurement error arises for many reasons, from administrative mistakes and delays in file keeping to the under-reporting of cases due to testing bottlenecks. We will let the data tell us about the standard deviation of this measurement error.

Thus, the measurement equation is:

$$\begin{pmatrix} \log\left(\frac{1}{\text{Population}} + \Delta D^{obs}_{H,t}\right) \\ \log\left(\frac{1}{\text{Population}} + \Delta D^{obs}_{P,t}\right) \\ \log\left(H^{obs}_t\right) \\ \log\left(G^{obs}_{t+5}\right) \\ \log\left(S^{obs}_t\right) \end{pmatrix} = \begin{pmatrix} \log\left(\frac{1}{\text{Population}} + \delta_{H,t}\theta_H H_{t-1}\right) \\ \log\left(\frac{1}{\text{Population}} + \delta_{P,t}\theta_P P_{t-1}\right) \\ \log\left(H_t\right) \\ \log\left((S_{t-1} - S_t)\gamma_{n,t}\right) \\ \log(S_t) \end{pmatrix} + \begin{pmatrix} \sigma_{D_H} u_{D_H,t} \\ \sigma_{D_P} u_{D_P,t} \\ \sigma_H u_{H,t} \\ \sigma_G u_{G,t} \\ \sigma_S u_{S,t} \end{pmatrix}. \tag{9}$$

Row one of Equation (9) links the (log) first difference of the observed share of deaths in hospitals, $\Delta D^{obs}_{H,t}$, with the share of individuals in hospitals H_{t-1} and the time-varying death rate $\delta_{H,t} = \frac{e^{d_{H,t}}}{1+e^{d_{H,t}}}$, where $d_{H,t}$ comes from row seven of Equation (7). For the first two observed series in Equation (9) we add the term 1/Population because there are few days with zero deaths.

Row two of (9) links the (log) first difference of the observed share of deaths at home, $\Delta D^{obs}_{P,t}$, with the share of individuals recovering at home P_{t-1} and the time-varying death rate $\delta_{P,t} = \frac{e^{d_{P,t}}}{1+e^{d_{P,t}}}$, where $d_{P,t}$ is a random walk that comes from row eight of Equation (7).

Time variation of the death rates in a hospital and at home captures improvements in clinical

protocols that raise the survival of patients, variation in hospital congestion (which may facilitate or complicate the treatment of patients), and changes in the mix of recovering individuals across different demographic groups.[11]

Row three of (9) tells us that the observed share of the population in hospitals is measured with error. An interpretation of this measurement error (beyond administrative mistakes in recording hospitalizations and discharges) is that some patients in hospitals are still infectious because isolation measures have not been fully implemented. Thus, while measured as being in a hospital, the individuals are still in the infectious compartment for the purposes of the model.

Row four of (9) gives us the observed share of cases, equal to the difference of the share of susceptible individuals times a factor $\gamma_{n,t} = \frac{e^{n_t}}{e^{n_t}+1} e^{-\mu} < 1$ that determines the percentage of cases that are reported. The rate of under-reporting depends on n_t, which follows the random walk defined in row nine of Equation (7).[12] The parameter μ ensures that under-reporting never goes to zero, for example, due to false negatives. As we discussed above, we let cases be under-reported by the $\gamma_{n,t}$ factor because this was a key factor at the start of the COVID-19 epidemic and we want to recover an estimate of the rate of reporting. At the same time, we assume that cases cannot be over-reported. As we will explain in more detail in Section 5, we only use row four every Friday by matching it to the new cases reported the following Wednesday.[13]

Row five of (9) links the point estimates of the susceptible share of the population from Herzog et al. (2020) to the corresponding share from our model. As we will describe below in more detail, this study is a prospective serial cross-sectional nationwide seroprevalence evaluation conducted in Belgium using blood samples collected during five different periods. We only use this equation for the dates for which the study is available. This row forces our smoothed states to incorporate the high-quality information from seroprevalence surveys, but allowing for differences due to sampling uncertainty and possible measurement errors.

Rows four and five of Equation (9) illustrate how state-space representations parsimoniously incorporate observables at different frequencies.

4.3 The Time-Varying Reproduction Numbers

A key state in our model is the effective contact rate β_t, the time-varying parameter that determines the speed of contagion in the matching function of susceptible and infectious

[11] In a model with several compartments for individuals indexed by age, this last effect would disappear. However, in our model, given the data available, we have only one group of individuals.

[12] The degree of under-reporting is sometimes called in the epidemiological literature the multiplication factor (MF). See Gibbons et al. (2014) for a survey of the somewhat *ad hoc* methods used to estimate the MF. A related approach to our measurement of under-reporting, also using an epidemiological model, is in Chudik et al. (2020).

[13] We could generalize this assumption. For example, Li et al. (2020) use a separate observational delay model and Arnon et al. (2020) assume 3 days from exposure to the onset of symptoms and 7 days from the onset of symptoms to a positive result.

individuals. When β_t is high, the infection spreads quickly. When β_t is low, the infection spreads slowly and might eventually abate.

The basic reproduction number, $\mathcal{R}_{0,t}$, is a closely related concept. The popularity of this measure is due to its straightforward interpretation: the basic reproduction number is the expected number of cases generated by one case in a population where all individuals are susceptible. In our model, $\mathcal{R}_{0,t} = \frac{\beta_t}{\gamma}$. Thus, the basic reproduction number inherits the same time variability encoded in β_t.

A similar measure to $\mathcal{R}_{0,t}$ is the effective reproduction number, $\mathcal{R}_{e,t}$, which considers the changes in the share of the susceptible population over time:

$$\mathcal{R}_{e,t} = \mathcal{R}_{0,t} S_t = \frac{\beta_t S_t}{\gamma}$$

By doing so, $\mathcal{R}_{e,t}$ offers a better measure of the instantaneous speed of the spreading of the infection. In Section 6, we will report the estimated evolution of $\mathcal{R}_{0,t}$ and $\mathcal{R}_{e,t}$.

4.4 Summary

We summarize all the variables and parameters of the model in Table 1. For ease of exposition, we partition Table 1 into five parts. The first part outlines the model compartments, the second part summarizes the model time-varying parameters, the third part presents observed variables, the fourth part shows the constant parameters of the model, and the fifth part lists the initial states. We assume that the initial states' distribution, $\boldsymbol{X}_0$, is degenerate. As mentioned above, we will treat these initial states as additional parameters and set priors over them.

5 The Data

We build daily frequency data, including data on deaths in hospitals, total deaths, hospitalizations, and new cases from Sciensano, a public institution recognized as a research institution by the Belgian Science Policy Office. All the data except for deaths in hospitals, which will be explained below, were downloaded from Sciensano on December 17, 2020. Our data span the period from March 15, 2020, to November 30, 2020. The starting date corresponds to the first day for which we have data on COVID-19 hospitalizations, and the ending date is the latest day for which we have data on deaths in hospitals. Deaths at home are computed as the difference between total deaths and deaths in hospitals. The latter was obtained upon request from Sciensano and it is dated as of December 15, 2020. In a few days, the number of total deaths in hospitals was above the number of total deaths. In such cases, we input zero deaths at home.

The raw data on new cases exhibit weekend and holiday effects. Hence, we use the data from

Table 1: Model Road Map

Model Compartments	
S_t	Share of the population that is susceptible at time t.
I_t	Share of the population that is infectious at time t.
H_t	Share of the population that is hospitalized at time t.
P_t	Share of the population that is recovering at home (outside the hospital) at time t.
$D_{H,t}$	Share of the population that has died in a hospital as of time t.
$D_{P,t}$	Share of the population that has died at home as of time t.
C_t	Share of the population that has recovered as of time t.
Model Time-Varying Parameters	
β_t	Effective contact rate at time t.
b_t	Natural logarithm of the effective contact rate, i.e., $b_t = \log(\beta_t)$.
$\gamma_{H,t}$	Share of the population no longer infectious at time t because they are recovering in a hospital.
$g_{H,t}$	Inverse of the logit function mapping $g_{H,t}$ to $\gamma_{H,t}$, i.e., $g_{H,t} = \log(\gamma_{H,t}/(1-\gamma_{H,t}))$.
$\delta_{H,t}$	Share of those leaving the hospital at time t due to death.
$d_{H,t}$	Inverse of the logit function mapping $d_{H,t}$ to $\delta_{H,t}$, i.e., $d_{H,t} = \log(\delta_{H,t}/(1-\delta_{H,t}))$.
$\delta_{P,t}$	Share of those no longer recovering at home at time t due to death.
$d_{P,t}$	Inverse of the logit function mapping $d_{P,t}$ to $\delta_{P,t}$, i.e., $d_{P,t} = \log(\delta_{P,t}/(1-\delta_{P,t}))$.
$\gamma_{n,t}$	Share of new cases at time t detected in the data.
n_t	Inverse of the logit function mapping n_t to $\gamma_{n,t}$, i.e., $n_t = \log(\gamma_{n,t}/(1-\gamma_{n,t}))$.
Observed Variables	
$D_{H,t}^{obs}$	Observed share of the population that has died at a hospital as of time t.
$D_{P,t}^{obs}$	Observed share of the population that has died at home as of time t.
H_t^{obs}	Observed share of the population that is hospitalized at time t.
G_t^{obs}	Observed new cases at time t as a share of the population.
S_t^{obs}	Observed share of the population that is susceptible at time t.
Model Constant Parameters	
γ	Share of the population that is no longer infectious at time t.
θ_H	Share of the population that leaves the hospital at time t.
θ_P	Share of the population that is no longer recovering at home at time t.
σ_b	Standard deviation of the innovation to b_t.
σ_g	Standard deviation of the innovation to $g_{H,t}$.
σ_h	Standard deviation of the innovation to $d_{H,t}$.
σ_p	Standard deviation of the innovation to $d_{P,t}$.
σ_n	Standard deviation of the innovation to n_t.
σ_{D_H}	Standard deviation of the innovation to the measurement equation of $\Delta D_{H,t}$.
σ_{D_P}	Standard deviation of the innovation to the measurement equation of $\Delta D_{P,t}$.
σ_H	Standard deviation of the innovation to the measurement equation of H_t.
σ_G	Standard deviation of the innovation to the measurement equation of $S_{t-1} - S_t$.
σ_S	Standard deviation of the innovation to the measurement equation of S_t.
μ	Upper bound of the share of detected cases.
Initial Values	
S_0, I_0, H_0, P_0, b_0, $g_{H,0}$, $d_{H,0}$, $d_{P,0}$, n_0	

new cases reported every Wednesday save for Wednesday, November 11, 2020, a public holiday in Belgium (Remembrance Day). For this date, we use the reported cases on Tuesday, November 10, 2020. We consider that those cases were infected the previous Friday. Since Sciensano assigns new cases according to the date on which the sample was taken, we impute new cases in the model to 5-day-ahead reported new cases.[14] Thus, we match an average incubation period of 5 days. As the CDC reports, symptoms may appear 2-14 days after exposure to the virus and some studies consider day 5 as the typical day of the onset of symptoms (Kucirka et al., 2020).

We picked this specification after extensive testing of alternatives. In those preliminary tests, we found that alternative approaches that seasonally adjusted for nonbusiness days not only distorted the number of reported cases but, more importantly, required potentially dubious assumptions to capture the nonlinear nature of the time series.[15] Hence, we use the data for new cases at a weekly frequency; this is the simplest and most transparent solution that we found.

Concerning the share of the susceptible population, we use the point estimates from Herzog et al. (2020). This study is a prospective serial cross-sectional nationwide seroprevalence evaluation conducted in Belgium using blood samples collected during the following five collection periods: March 30-April 5, April 20-April 26, May 18-May 25, June 8-June 13, and June 29-July 3. In total, the collection periods include 33 days, i.e., about 15% of our sample. The population data used to express variables in per capita terms correspond to 2019 and were obtained from the World Bank's website.

6 Results

We now present our results. First, we will describe the algorithm we employ for estimation. Second, we will present our priors and report the posterior moments. Third, we will show that the model fits the data extremely well. Fourth, we will report the time-varying reproduction numbers implied by the model. Fifth, we will discuss the estimated time-varying death probabilities (i.e., the time-varying death rates divided by θ_H and θ_P, respectively). Sixth, we will back up a measure of which percentage of cases were reported. We will conclude with a brief discussion of the findings regarding the other states of the model.

6.1 Algorithm

As described in Section 3, we implement a PFMH algorithm. To do so, we first need to evaluate the likelihood function $p\left(\boldsymbol{y}^T; \boldsymbol{\Theta}\right)$ given a set of parameters $\boldsymbol{\Theta}$ (we include $\boldsymbol{X_0}$ in $\boldsymbol{\Theta}$ since we

[14] According to the data documentation from Sciensano, the diagnostics include molecular techniques (i.e., polymerase chain reaction or PCR) and rapid antigen tests.

[15] Similarly, the nonlinearity of the model made the use of cumulators to aggregate cases over the week challenging to implement and generated results that were not transparent.

take the initial states $\boldsymbol{X_0}$ as given) using the particle filter with the following pseudo-code:

Step 0, Initialization: Set $t \rightsquigarrow 1$.

Step 1, Prediction: Sample $\{\tilde{\boldsymbol{X}}_t^i\}_{i=1}^N$ from the conditional density $p(\boldsymbol{X}_t^i|\boldsymbol{\Theta})$ if $t = 1$, and $p(\boldsymbol{X}_t^i|\boldsymbol{X}_{t-1}^i;\boldsymbol{\Theta})$ otherwise.

Step 2, Filtering: Let the weights $\omega_t^i = p(\boldsymbol{y}_t|\boldsymbol{X}_t^i;\boldsymbol{\Theta})$ for $i = 1,\ldots,N$.

Step 3, Sampling: Sample with replacement $\{\boldsymbol{X}_t^i\}_{i=1}^N$ from $\{\tilde{\boldsymbol{X}}_t^i\}_{i=1}^N$ using weights ω_t^i.

Step 4, Recursion: If $t < T$ set $t \rightsquigarrow t+1$ and go to step 1. Otherwise move to next step.

Step 5, Evaluation: Use $\{\{\boldsymbol{X}_t^i\}_{i=1}^N\}_{t=1}^T$ to estimate the likelihood function as:

$$\widehat{p}(\boldsymbol{y}^T;\boldsymbol{\Theta}) = \prod_{t=1}^{T}\left(\sum_{i=1}^{N}\frac{\omega_t^i}{N}\right)$$

Step 1 is implemented by sampling from the distribution of innovations (which, given their normality, is straightforward) and, given the swarm of simulated states at $t-1$, using Equation (7) to get the swarm of simulated states at t. Notice how simple it would be to introduce non-normal innovations; we would just need to switch the sampler of the innovations. Step 2 is also straightforward. From Equation (9), we can plug in the observations and the states and back up the measurement errors that make both the left- and right-hand side equal. Then, we evaluate the probability of those measurement errors given a normal distribution. Because of the sampling in Step 3, the estimated likelihood $\widehat{p}(\boldsymbol{y}^T;\boldsymbol{\Theta})$ in Step 5 is a random variable and not differentiable with respect to the parameter values.

We nest the particle filter with a random walk Metropolis-Hastings to approximate the posterior distribution of interest. The following algorithm generates a Markov chain that converges to the posterior distribution of $\boldsymbol{\Theta}$.

Step 0, Initialization: Set $\boldsymbol{\Theta}^{(0)} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ and $g \rightsquigarrow 1$

Step 1, Proposal: Draw $\boldsymbol{\Theta}'$ from a multivariate normal with mean $\boldsymbol{\Theta}^{(g-1)}$ and covariance matrix $\boldsymbol{\Sigma}_\Theta$.

Step 2, Acceptance/Rejection: Set $\boldsymbol{\Theta}^{(g)} = \boldsymbol{\Theta}'$ with probability

$$\alpha(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(g-1)}) = \min\left\{1,\ \frac{\widehat{p}(\boldsymbol{y}^T;\boldsymbol{\Theta}')p(\boldsymbol{\Theta}')}{\widehat{p}(\boldsymbol{y}^T;\boldsymbol{\Theta}^{(g-1)})p(\boldsymbol{\Theta}^{(g-1)})}\right\}$$

and $\Theta^{(g)} = \Theta^{(g-1)}$ otherwise.

Step 3, Recursion: If $g < G$, set $g \rightsquigarrow g+1$ and go to step 1. Otherwise exit

the algorithm with $\{\boldsymbol{\Theta}^{(g)}\}_{g=1}^{G}$.

In Step 1, we set $\boldsymbol{\mu}_0$ as the posterior mode, and $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_\Theta$. Step 2 is based on the random-walk proposal with covariance matrix $\boldsymbol{\Sigma}_\Theta$. In accordance with standard practice (Roberts and Rosenthal, 2001), we select $\boldsymbol{\Sigma}_\Theta$ so that the resulting acceptance rate of the algorithm is close to 20%. We set $G = 100,000$ with a burn-in of $10,000$. Using the draws from the algorithm, we approximate the posterior moments of interest such as mean, 5%, and 95% quantiles of $\boldsymbol{\Theta}$. We approximate the posterior distribution of states by the marginalization

$$p(\boldsymbol{X}^T|\boldsymbol{y}^T) = \int p(\boldsymbol{X}^T|\boldsymbol{\Theta}, \boldsymbol{y}^T)p(\boldsymbol{\Theta}|\boldsymbol{y}^T)d\boldsymbol{\Theta}$$

where $p(\boldsymbol{\Theta}|\boldsymbol{y}^T)$ is given by the algorithm above and $p(\boldsymbol{X}^T|\boldsymbol{\Theta}, \boldsymbol{y}^T)$ can be approximated by either the particle smoother or the particle backward sampler (see Fernández-Villaverde et al., 2016).

6.2 Prior and Posterior Moments

We impose prior distributions for the parameters and initial states of the model that capture previous existing knowledge about the features of COVID-19. For transparency, we use well-known parametric families commonly applied in similar problems. These priors are described in Table 2. The table also describes the posterior moments. We use $N = 50,000$ to compute the particle filter and 30 repetitions of the filter to compute the mean and the standard deviation of the posterior mode estimates.

Prior Let us briefly discuss the rationale underlying the parametric choices for our priors, their central tendency, and their tightness. We pick beta priors for the three parameters determining transitions across compartments: γ, θ_H, and θ_p. The prior for γ is centered at 0.2 (implying that the median time a person remains infectious is about five days) consistent with the evidence in Bar-On et al. (2020). The standard deviation of γ is 0.05, which implies an interdecile range (IDR) of 4 to 7 days and attributes a 10% probability to average infectious periods beyond 7 days.[16] The prior mean and standard deviation for θ_H are set to 0.10 and 0.02, respectively, implying a median length of average stays in hospitals of about 10, an IDR of 6 to 13 days, and a 10% probability of average stays longer than 13 days, broadly in line with the average hospital stays in Belgium according to Catteau et al. (2020), who found a median stay of 9 days with a 6–to-15-day interquartile range. Even so, our prior rules out neither average hospital stays as short as five days nor average stays longer than 20 days. We set the prior mean for θ_P

[16] We are dealing with the average infectious period across the population, which may include longer and shorter infectious periods. The standard deviation of γ embodies uncertainty about the average duration of spells of infectiousness, not about their dispersion in the cross-section of individuals.

equal to 0.15, implying a median average home recovery of between 6 and 7 days and an IDR of about 4 to 11 days. While less is known about home recoveries, our prior assumes that, on average, individuals recovering outside of hospitals present less severe symptoms and hence are more likely to recover faster. In any case, the standard deviation of this prior allows the model to encompass a wide range of possibilities. Our choice of standard deviations for the priors of θ_H and θ_P highlights a quintessential benefit of following a Bayesian approach. We can take advantage of the nationwide scope of the data in Catteau et al. (2020) and the observation that most of the studies outside of China find an average stay in hospitals slightly below 10 days (e.g., Lavery et al. 2020, and Rees et al. 2020) to discipline the estimation, while remaining more agnostic about those parameters, such as the average duration of home recoveries, for which it is harder to gather external data.

We pick inverse gamma priors for the parameters governing the law of motion of the time-varying parameters σ_b, σ_g, σ_h, σ_p, and σ_n. Inverse gamma priors are popular choices for the step size changes across a variety of applications. We set the prior mean equal to 0.2 for σ_b, σ_g, σ_h, σ_p, and σ_n to allow for quick changes in behavior, policies, and medical treatments. In the case of the effective contact rate, our prior for σ_b implies that this rate varies, on average, by about 20% from one day to the next, in line with the fast changes recorded after stay-at-home orders. In the case of the law of motion governing the share of those infected recovering in a hospital, our prior implies (up to first order) that the share of those infected recovering in a hospital changes on average by about $20 \times (1 - \gamma_{H,t-1})\%$ from period $t-1$ to period t.[17] Thus, when the share of the population hospitalized is small, we expect larger changes than when this share is large. Similarly, the prior mean of σ_h to 0.2 implies (up to first order) that the death rate at a hospital changes on average by about $20 \times (1 - \delta_{H,t-1})\%$ from period $t-1$ to period t. This is a flexible specification that allows death death rates in hospitals to fluctuate more when the death rate is low, capturing sharp increases in death rates. The persistence of the death rate implied by the prior increases with its level, as it is likely to be subject to shocks that are proportionally smaller when the death rate level is high. We impose an identical prior for σ_p. With regard to σ_n, our prior implies that (up to first order) the share of detected cases changes on average by about $20 \times (1 - \gamma_{n,t-1})\%$ from period $t-1$ to period t. Thus, when the detection rate is low, the day-to-day changes are larger than when the detection rate is large. Importantly, the standard deviation of the prior for σ_b, σ_g, σ_h, σ_p, and σ_n is set to 0.2. Hence, our prior allows for either extremely slow or fast changes in the time-varying parameters.

Next, we discuss the prior for the initial value of the states. We set the prior mean for S_0 equal to 0.98. This is a value that is about one percentage point above the share of the population susceptible to the virus as of March 30, 2020, according to the point estimates in

[17] This follows from linearizing $\gamma_{h,t} = \frac{e^{g_{H,t}}}{e^{g_{H,t}}+1}$ around $\gamma_{h,t-1}$ and dividing the resulting expression by $\gamma_{h,t-1}$, that is, $\frac{\gamma_{H,t}-\gamma_{H,t-1}}{\gamma_{H,t-1}} = (1 - \gamma_{H,t-1})(g_{H,t} - g_{H,t-1})$.

Herzog et al. (2020). The standard deviation is set equal to 0.01, which is about two times the standard deviation associated with the 95% confidence interval for the seroprevalence of COVID-19 in Belgium as of March 30, 2020, reported in Herzog et al. (2020). The prior mean for I_0 is 0.001, which implies that 0.1% of the Belgian population was infectious on March 15, 2020. Notice that assuming a contagion rate of 0.6 (a value we will describe below), then the number of new cases on March 16, 2020 implied by this assumption is 6,752, which is broadly consistent with the order of magnitude implied by inflating the reported cases under our implied prior for the share of detected cases as will be discussed below. The standard deviation of I_0 is set to 0.001, which is also in line with our assumptions on the detection rate. Since the priors for S_0 and I_0 are independent, and hence to make sure that the constraints in equation (8) are satisfied, we discard any draw of S_0 and I_0 that violates such restrictions.

The prior mean for b_0 is set equal to $\log(0.6)$, so that prior belief about $\mathcal{R}_{0,t}$ is around 3 on March 15, 2020, a value consistent with other assumptions in the literature, e.g., D'Arienzo and Coniglio (2020). The standard deviation for b_0 is set to 0.5, so that if we were to use values of b_0 one standard deviation below and above the mean, we would obtain values of R_0 in the interval [1.8, 5.9]. The prior mean for $g_{H,0}$ is set so that when mapped to $\gamma_{H,0}$ it implies that 1% of those recovering from COVID-19 were doing so in a hospital on March 15, 2020. This number is computed as follows. We sum the "true COVID-19" cases embedded in our prior as implied by a detection rate of 5.5% (a plausible value according to our prior) from March 11 until March 15. The result is 27,054. Since 266 individuals were hospitalized on March 15, we have that $266/27,054 \approx 0.01$. The standard deviation for $g_{H,0}$ is set equal to 2, which is large enough to cover an interval of hospitalized individuals between $[15; 800]$, which is quite wide given the reported 266. The prior mean for $d_{H,0}$ is set equal to $\log(0.15/(1\text{-}0.15))$. This number is obtained by the measurement equation assuming that deaths in hospitals are measured without error on March 15, 2020; thus $\delta_{H,0} = \frac{\Delta D_{H,0}}{H_0 \theta_H} = \frac{4}{266 \times 0.1} = 0.15$. The standard deviation for $d_{H,0}$ is set equal to 2 so that the prior for the probability of those leaving the hospital because of death is concentrated in the interval $[0.02, 0.56]$. The prior mean and standard deviation for $d_{P,0}$ is set equal to the ones for $d_{H,0}$.

The prior mean for n_0 is set equal to -1.00 and the standard deviation is set equal to 1.00. This choice, together with the prior over μ (described below), implies a fairly flat prior over the share of detected cases on March 15, 2020 as shown in Appendix A.1.

We have degenerate priors for H_0 and P_0. In particular, we set $H_0 = 2.31 \times 10^{-5}$ (that is, 266 hospitalized expressed as a share of the Belgian population the day before the start of our sample) and $P_0 = 0$. The first is given to us by the data and the second is just for convenience, since its effects on the estimation are trivially small.

Next, we justify our choice of prior for μ, i.e., the parameter controlling the share of false negatives. The prior mean of μ is 0.175, the mid-point of the 2 to 33% interval of false

negatives in PCR tests reported by Arevalo-Rodriguez et al. (2020) when the type of specimen is nasopharyngeal or oropharyngeal. In our model, this implies that as the permanent component of the share of detected cases approaches one, a false negative rate of 17.5% is the most likely outcome. The standard deviation for μ is set equal to 0.05. This choice for the standard deviation of μ implies an IDR range for the false negative rate of 11 to 24%. Although rare, our prior does not completely rule out false negative rates below 5% or above 35%. We allow for this wide range of possible values because there can be differences between the accuracy of PCR tests as measured in a lab and when applied widely in a myriad of different situations.

Finally, we impose degenerate priors over the parameters governing the standard deviation of the measurement errors of observables, σ_{D_H}, σ_{D_P}, σ_H, σ_G, and σ_S. We use the data to calibrate σ_{D_H}, σ_{D_P}, σ_H, and σ_G by applying a Hodrick-Prescott filter to the observed variables. We tune the frequency of the filter so that the cyclical component is nearly serially uncorrelated. Then, we set the mean of the parameters in question equal to the standard deviation of the serially uncorrelated cyclical component. This procedure results in $\sigma_{D_H} = 0.29$, $\sigma_{D_P} = 0.32$, $\sigma_H = 0.017$, and $\sigma_G = 0.16$.

The prior mean for σ_S is set equal to 0.02, which is roughly about 5 times the average standard deviation across collection periods implied by Herzog et al. (2020). Serological studies are a useful guide, but likely to be subject to large measurement error, as is evident in the variation in the share of individuals infected by the virus throughout the collection periods.

Columns two to four of Table 2 summarize the priors of each parameter.

Posterior We now discuss the posterior moments reported in Table 2 (mode, mean, and the 90% posterior probability interval).[18] For ease of exposition, we focus on the posterior mean, save for a few parameters where the 90% posterior probability interval is of particular interest.

The posterior mean and 90% posterior probability interval for γ are 0.070 and $[0.055; 0.105]$, respectively. This suggests that the implied posterior mean estimate of the average length a person remains infectious is between 13 and 14 days and the 90% posterior probability interval is between 9 and 19 days (we compute the statistics regarding days by inverting every draw of γ, not by inverting the mean γ; the same will apply to all other statistics regarding time). Hence, according to our estimates, an infected person remains contagious for a longer period than in the mean of our prior beliefs. The posterior mean and 90% posterior probability interval for θ_H are 0.193 and $[0.161; 0.229]$, respectively. Hence, the posterior mean of the average length of stay in hospitals is between 5 and 6 days and the 90% posterior probability interval is between 4 and 7 days, which is a shorter duration than our prior. The posterior mean and 90% posterior probability interval for θ_P are 0.144 and $[0.084; 0.215]$. The implied posterior mean of the average

[18] Appendix A.2 shows the prior and posterior distribution for all the estimated parameters. While some parameters are more sharply identified than others, the contrast between the prior and posterior distribution shows that the data are informative about the estimated parameters.

Table 2: Prior and Posterior Moments

Parameter/	Prior			Posterior			
Initial State	Dist.	Mean	Std	Mode	Mean	5%	95%
γ	B	0.20	0.05	0.070	0.075	0.055	0.105
θ_H	B	0.10	0.02	0.208	0.193	0.161	0.229
θ_P	B	0.15	0.05	0.122	0.144	0.084	0.215
σ_b	IG	0.20	0.20	0.161	0.186	0.126	0.263
σ_g	IG	0.20	0.20	0.153	0.175	0.139	0.221
σ_h	IG	0.20	0.20	0.078	0.091	0.065	0.122
σ_p	IG	0.20	0.20	0.152	0.149	0.116	0.188
σ_n	IG	0.20	0.20	0.224	0.307	0.168	0.485
S_0	B	0.98	0.01	0.973	0.976	0.959	0.99
I_0	B	0.001	0.001	0.0009	0.0013	0.0006	0.0024
b_0	N	log(0.60)	0.50	log(0.357)	log(0.362)	log(0.232)	log(0.562)
$g_{H,0}$	N	logit(0.01)	2.00	logit(0.225)	logit(0.149)	logit(0.068)	logit(0.325)
$d_{H,0}$	N	logit(0.15)	2.00	logit(0.121)	logit(0.138)	logit(0.101)	logit(0.192)
$d_{P,0}$	N	logit(0.15)	2.00	logit(0.041)	logit(0.024)	logit(0.009)	logit(0.072)
n_0	N	-1.00	1.00	-0.954	-1.478	-2.611	-0.220
μ	G	0.175	0.05	0.121	0.124	0.082	0.170
Mean $\widehat{p}(\boldsymbol{\Theta}\|\boldsymbol{y}^T)$				281.91	281.27		
Std $\widehat{p}(\boldsymbol{\Theta}\|\boldsymbol{y}^T)$				1.35	1.04		

Note: This table is based on an MCMC chain with 90,000 posterior draws obtained after a burn-in period of 10,000 draws. The number of particles used in the estimation is 50,000. The acceptance rate is about 16%. The mean and standard deviation of the log posterior density are based on 500 evaluations at the posterior mean and mode, respectively.

length of stay at home is between 7 and 8 days, and the 90% posterior probability interval is between 4 and 12 days. The likelihood overturns the prior for θ_H and θ_P: our point estimates suggest a longer recovery at home, perhaps due to higher-quality treatments given in hospitals.

Now we turn to the parameters controlling the step size of changes in behavior, policies, and medical treatments. The posterior distribution for σ_b is centered around the prior mean, 0.20, but it is more concentrated than our prior, ruling out small and abrupt changes in daily behavior. A similar conclusion emerges for the parameter controlling the share of infected individuals recovering in hospitals, σ_g. In contrast, the posterior mean for σ_h, 0.091, is much lower than the prior mean of 0.20. Hence, according to our model, moderate-to-large step sizes (such as a value of about 0.2) in the transmission rate are more plausible than in the transmission rate or in the share of infected individuals recovering in hospitals than in the share of mortality rates in hospitals, σ_h. In the case of the mortality rates at home, σ_p, the posterior mean is 0.149, which is a step-size value somewhat higher than in the case of mortality rates in hospitals. The posterior mean for changes in the share of detected cases is 0.307, which is larger than our prior mean. This is consistent with expedited increases in testing capacity.

Finally, we describe the posterior mean for the initial conditions, which in our estimation

corresponds to March 15, 2020, as well as the posterior mean of the share of false negatives. The posterior mean and 90% posterior probability bands for the initial share of the population susceptible to the virus are 0.976 and $[0.959; 0.99]$. Consequently, our posterior estimates indicate that, with high probability, about 2% of the Belgian population had come in contact with the virus by mid-March 2020. The posterior mean and posterior probability bands for the initial share of the population that was infectious are 0.0013 and $[0.0006; 0.0024]$. Hence, according to our estimates, about 0.1% of the Belgian population was infectious with high probability by around March 15. The posterior mean of the initial log of the transmission rate is $\log(0.362)$, implying a basic reproduction number, $\mathcal{R}_{0,t}$, of around 4.8.

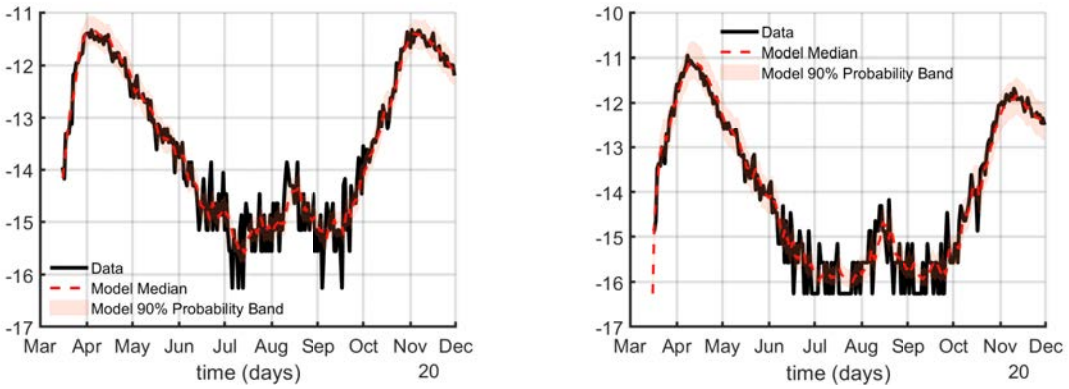
The posterior mean of the process driving the initial share of the population hospitalized is $\text{logit}(0.149)$ so that 14.9% of those infected with COVID-19 were recovering in hospitals at the beginning of our estimation sample. The posterior mean of the share of those no longer recovering in hospitals due to deaths is 13.8% and the posterior mean of the share of those no longer recovering at home due to deaths is 2.4%. These numbers imply an initial death probability of 2.6% conditional on being hospitalized and an initial death probability of 0.35% conditional on recovering at home. These findings suggest that, by mid-March 2020, very sick patients were being admitted to hospitals for COVID-19 treatment. As documented in Appendix A.2, the posterior distribution n_0 is very similar to the prior distribution, with roughly the same mean. However, our choice of prior for n_0 was an educated guess: since this parameter was hard to identify, we conducted multiple optimizations of n_0 starting our Monte Carlo simulation at different initial values. The optimizer consistently delivered values concentrated around the mean of our prior.

The posterior mean of the share of false negatives implied by our model is 12.4%, which is smaller than our prior. The posterior mean of parameters μ and n_0 imply a 90% posterior probability interval for the initial share of detected cases of between 9 and 31%. Thus, according to our estimates, as of March 15, 7 out of 10 cases were being undetected in Belgium.

6.3 Data Fit

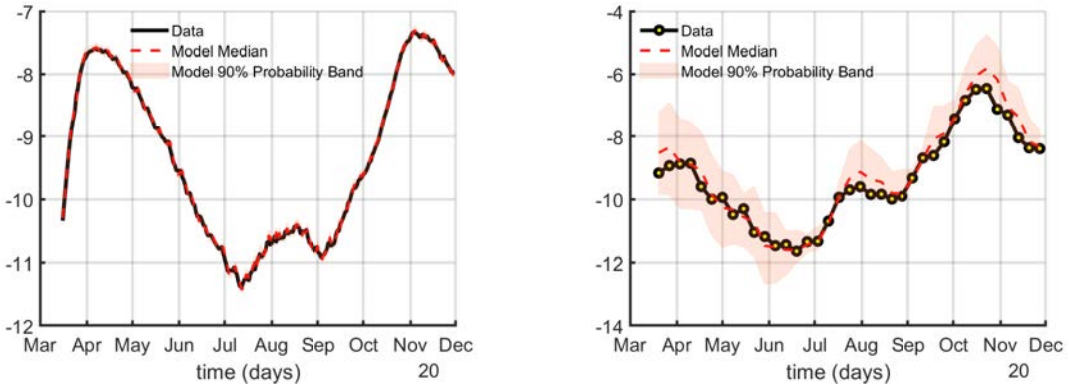
Figures 2 and 3 show the fit of the model to the data at the posterior mean. The left panel of Figure 2 plots, in a continuous black line, the (log) first difference of the share of (daily) deaths in hospitals in Belgium from mid-March until late November (i.e., the share of new deaths). We can see two large peaks, one in early April and one in early November, corresponding to the first and third waves of COVID-19 in Belgium, plus a smaller wave in mid-August. In red, we plot the median of the one-step-ahead forecast of the model, evaluated at the posterior mean, together with a 90% probability band. The right panel plots, following the same formatting as the left panel, the (log) first difference of the share of the deaths outside of hospitals, the

one-step-ahead forecast, and the 90% probability band.



(a) Death in hospitals (b) Deaths at home

Figure 2: Deaths: One-step-ahead forecast and data, Belgium



(a) Population in hospitals (b) New reported cases

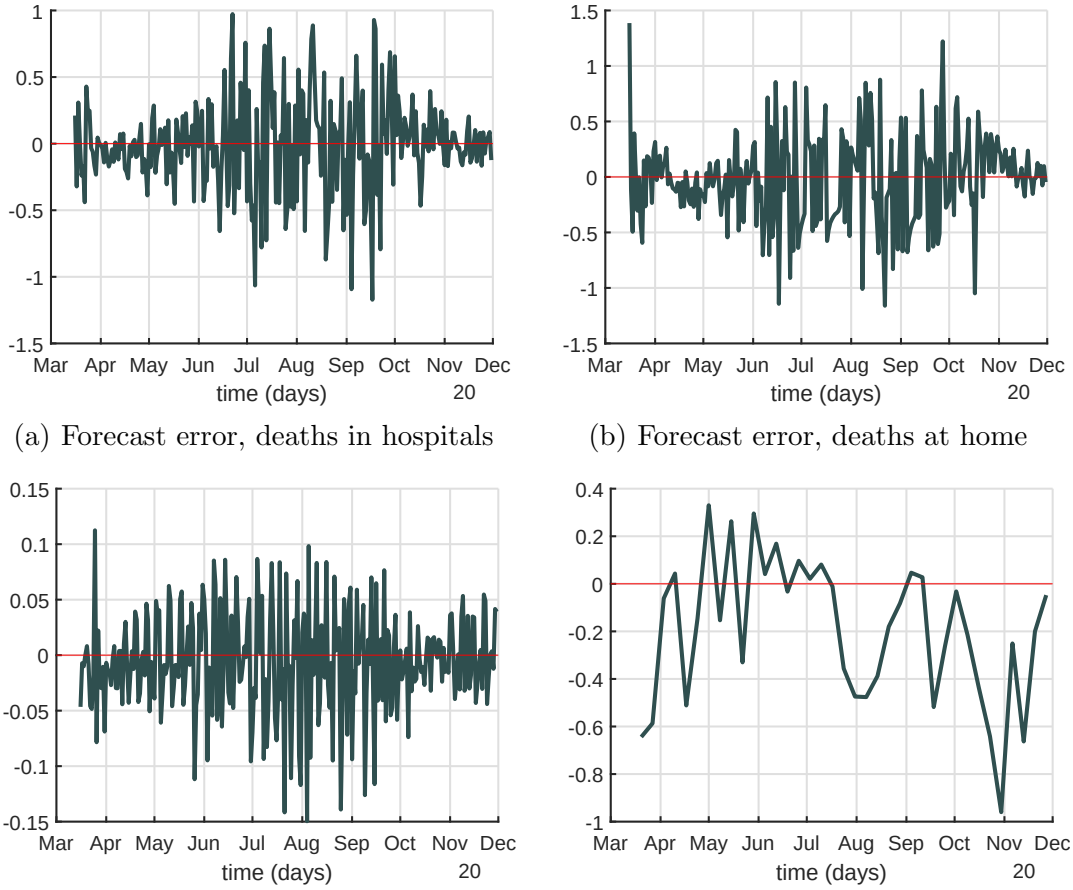
Figure 3: Hospitalized population and new cases: One-step-ahead forecast and data, Belgium

The main lesson from Figure 2 is that the model fits the data remarkably well, capturing the three waves of deaths. The observations fall within the 90% probability band on most days. The only misses are during the summer of 2020, when deaths were fewer than ten a day. During this time, small random differences between one period to the next are extremely difficult to forecast. The model also accounts for the observation that deaths in hospitals had reached, by early November, the same level as in early April, but deaths at home had not. This might reflects Belgian hospitals's better ability to cope with severe COVID-19 patients due to additional available beds: hospitals can admit more patients and, therefore, more deaths occur there.[19]

This hypothesis is supported by the left panel of Figure 3, where we plot the (log) share of hospitalizations following the same formatting as in Figure 2. As can be seen, hospitals had more

[19]We could also plot the one-step-ahead forecast integrating over the whole posterior. This alternative exercise makes little difference in practice, but complicates the interpretation of the figures.

patients at the end of the sample. This left panel also shows that the model can account for hospital utilization. Finally, the right panel of Figure 3 draws the (log) share of newly reported cases. The data indicate many more cases in the last peak than in the first. However, recall that this panel plots reported cases and our forecast of those recorded cases, not actual cases. We will come back to this point below.



(a) Forecast error, deaths in hospitals

(b) Forecast error, deaths at home

(c) Forecast error, population in hospitals

(d) Forecast error, new reported cases

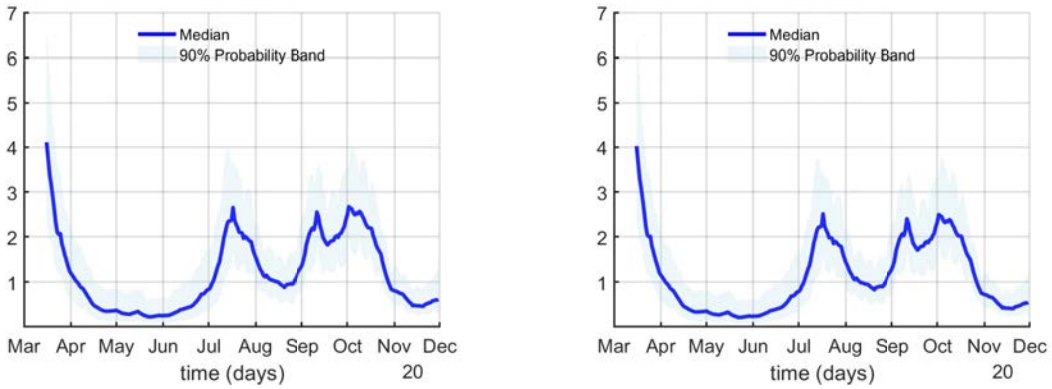
Figure 4: One-step-ahead forecast errors, Belgium

An alternative way to assess the fit of the model is Figure 4, where we plot the one-step-ahead forecast errors for our four observables. The errors are well grouped around zero (except, partially, reported cases), although they present heteroscedasticity. The clustering of forecast errors around zero is another indication that the model fits the data well. The heteroscedasticity is not a problem for our Bayesian approach. Still, it suggests an important avenue for a future extension of the model to account for this time-varying volatility.

6.4 The Estimated Time-Varying Reproduction Numbers

From our posterior distribution, we can recover the smoothed value of the unobserved states (or related variables) and their probability distribution. These are, often, the most relevant objects for policymaking, as they describe the epidemiological situation of a given area (country/region/...) under study and can be used, as inputs, in a loss function to pick an optimal policy.

A key variable to monitor in an epidemic is the basic reproduction number $\mathcal{R}_{0,t}$. The left panel on Figure 5 plots the smoothed $\mathcal{R}_{0,t}$ and the 90% smoothed band. We see that the basic reproduction number started in mid-March slightly above 4, around the values that several clinical studies have suggested. For example, see Table 1 in Katul et al. (2020) for a list of estimates (the authors conclude that their best estimate of an unmitigated $\mathcal{R}_{0,t}$ for COVID-19 is 4.5, quite close to our result). However, $\mathcal{R}_{0,t}$ fell very rapidly and, by mid-April, it was well below 1. In May and June, $\mathcal{R}_{0,t}$ stabilized around 0.5 and increased back to 2 during the summer of 2020, fluctuating until late October between 2 and 1. After November, the basic reproduction number is again at a much lower level, with the whole 90% smoothed band below 1.



(a) $\mathcal{R}_{0,t} = \beta_t/\gamma$ (b) $\mathcal{R}_{e,t} = \mathcal{R}_{0,t}S_t$

Figure 5: Reproduction numbers, Belgium

The right panel of Figure 5 reports our smoothed median and 90% probability estimates of $\mathcal{R}_{e,t}$, the effective reproduction number. While $\mathcal{R}_{e,t}$ starts at the same level as $\mathcal{R}_{0,t}$, as time passes by and the share of the suspectible populations shrinks, $\mathcal{R}_{e,t}$ moves toward lower values.

The difference, however, between $\mathcal{R}_{0,t}$ and $\mathcal{R}_{e,t}$ is not large. Figure 6, which draws our smoothed estimate of the share of susceptibles (and the 90% probability band), tells us why. Even as late as November, our model estimates that only around 12% of the population had ever been infected. By early December, our estimates suggest that Belgium was far away from herd immunity.

Figure 6 also plots, in red squares, the point estimates of the seroprevalence studies reported by Herzog et al. (2020) and, in black crosses, the 95% confidence interval of the studies. Figure

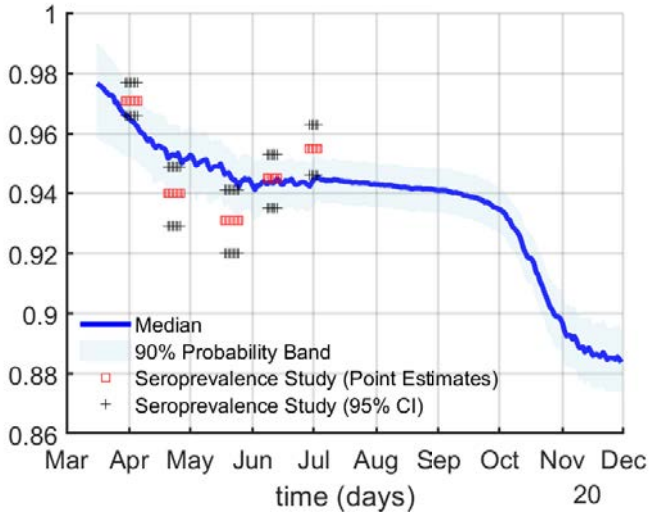


Figure 6: Share of susceptible population, Belgium

6 illustrates our argument regarding the power of state-space representations to incorporate disparate data sources. Our model can account for those seroprevalence studies: in all cases, our 90% probability band and the 95% confidence interval of the studies overlap.

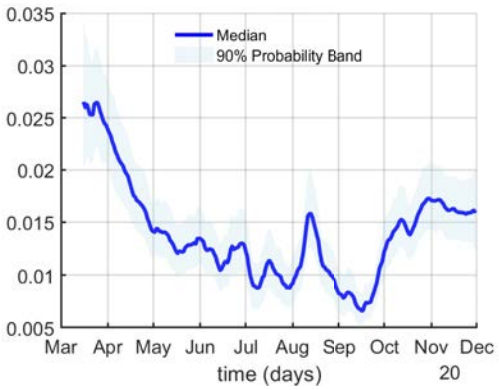
6.5 Time-Varying Death Probabilities

Another central state variable in our model is the time-varying death probabilities. Those probabilities can change for many reasons. We can enumerate a few. First, medical protocols vary. As health workers learn more about an infection, they can handle patients better, even in the absence of effective treatments. Second, hospitals experience different occupancy rates, with variations in the inflows and total capacity, as the supply of beds and ICU units responds to the crisis. Third, the demographics of patients can change, by varying either in terms of age or in terms of comorbidity levels.[20]

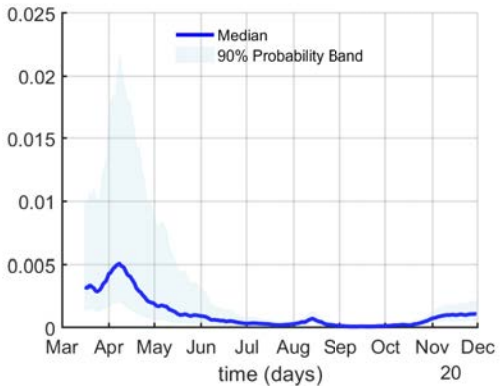
The left panel of Figure 7 shows the in-hospital death probability (median smoothed, plus the 90% probability band), which went down from over 2.5% in March to less than 1% by early July. The sizeable third peak of COVID infections in the fall of 2020 increased that probability only around 1.5%, suggesting a considerable degree of improvement in clinical outcomes.

The right panel of Figure 7 shows the at-home death probability, which fell from around 0.5% in March to less than 0.1% by early July. Here the changing conditions at retirement communities, which were unprepared for the virulence of COVID-19 in the late winter of 2020, are probably at the core of the estimated variation in death probabilities.

[20]Imagine, for example, that individuals with a high probability of infection (e.g., due to their social networks) and high fatality rate (e.g., smokers) got infected in the first wave. As there are fewer of these individuals in the population when the second wave arrives, the measured death rates will mechanically fall.



(a) In-hospital death probabilities, Belgium



(b) At-home death probabilities, Belgium

Figure 7: Death probabilities, Belgium

6.6 New Cases

Figure 8 plots, in blue, the smoothed median and 90% probability bands of the true new cases, and, in black, the reported new cases in Belgium. We can clearly see that the first peak is much smaller, with either measure, than the third one (with a minor second peak in the middle). The fall in death probabilities explain why deaths did not reach the first wave levels.

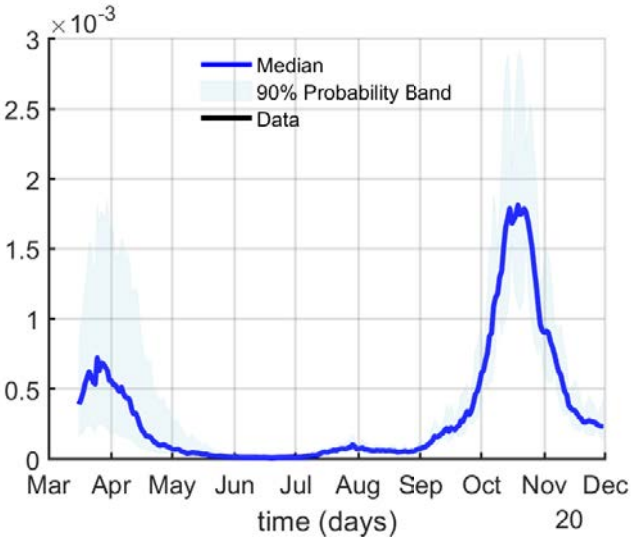
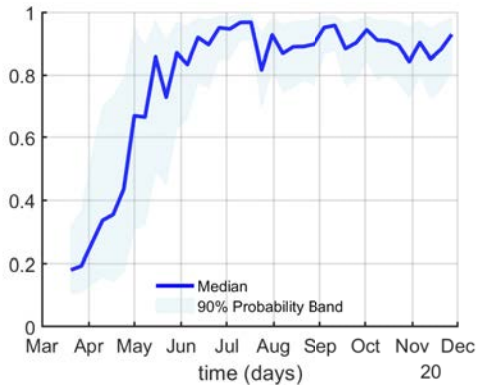


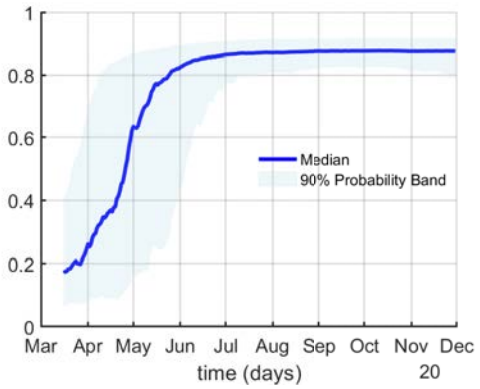
Figure 8: New cases, Belgium

In addition, the reported cases and the estimated cases were very different during the first wave. Figure 9 makes this point clearly by showing, in the left panel, our median smoothed estimate of the share of reported cases and, in the right panel, the permanent component of this share. By late March, less than 20% of all cases were being reported, while by the summer of 2020, after testing became more prevalent, around 88% of cases were being reported. The

permanent component of this share suggests that this increase in reported cases is very persistent.



(a) Share of reported cases, Belgium



(b) Permanent component of the share of reported cases, $\gamma_{n,t}$, Belgium

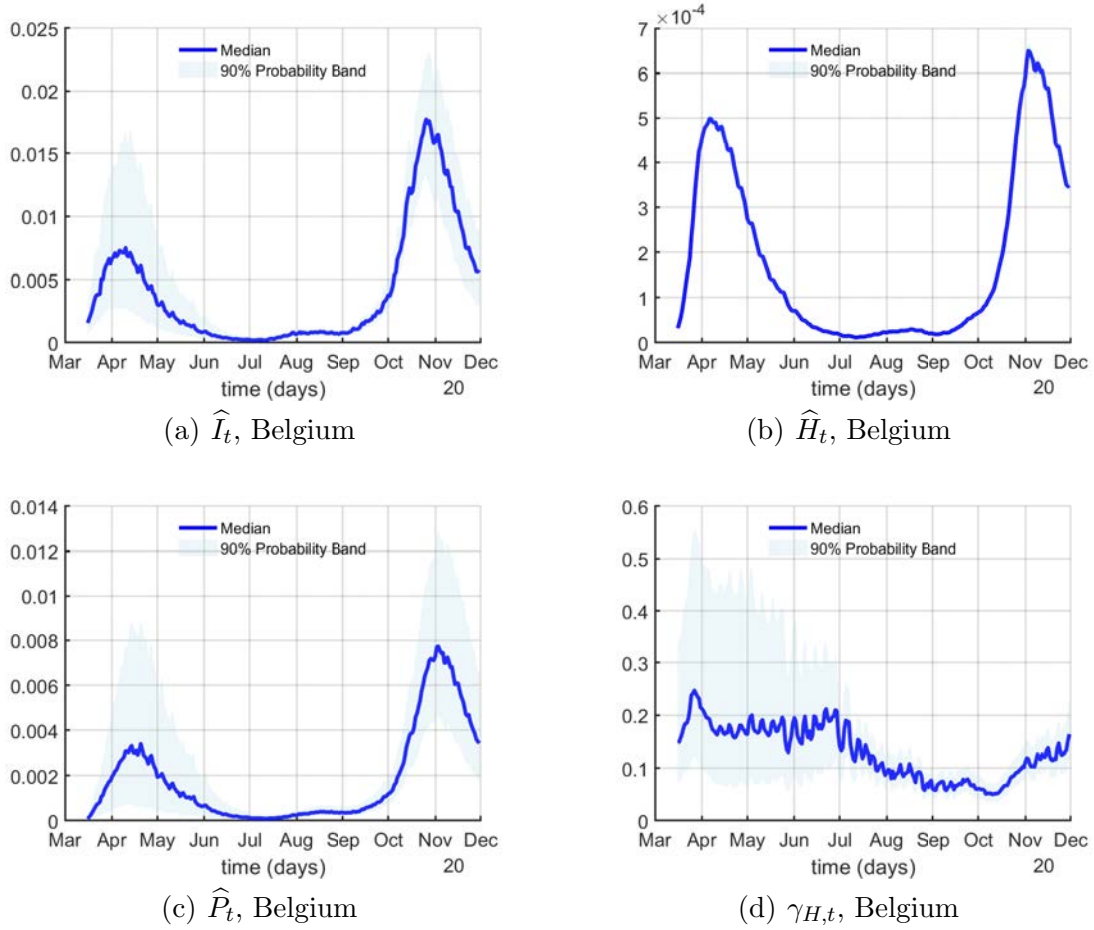
Figure 9: Share of reported cases, Belgium

6.7 Other States

Finally, we present the smoothed estimates of the share of the population that is infectious, the share of the population that is recovering (in hospitals and at home), and the inflow of hospitalizations as a share of those that are no longer infectious.

Figure 10 shows the time series of these smoothed variables throughout our sample. The share of infectious increases to 1.8% by the end of March: more than twice as large as the March 15 estimate of 0.7%. This sharp increase in the share of infectious may provide a rationale for the first nationwide lockdown imposed by the Belgian government on March 18, 2020. These measures appear to have had an effect as the number of infectious dropped at the beginning of April and continued to decline, reaching a trough at the end of June 2020. After that, the share of infectious began to increase but at a moderate pace up until the first week of September, when we observe a second exponential increase in the share of infectious, leading to a reintroduction of lockdown measures on November 2, 2020.

The shares of the population recovering in hospitals and at home broadly track the contour of the share of infectious with a lag. For example, the first peak of infections occurs on March 30, and the first peak of hospitalizations occurs on April 7. A similar pattern emerges for the share of the population recovering outside of hospitals. Notice that while the peak of the third wave of infections and the one of those recovering at home is more than twice as high as the peak of the first wave, during the second wave, hospitalizations peak at a level only marginally higher than the one of the first wave. This is consistent with the decline in the share of those recovering from COVID-19 in hospitals during the third wave relative to the first one.



(a) $\widehat{I}_t$, Belgium (b) $\widehat{H}_t$, Belgium

(c) $\widehat{P}_t$, Belgium (d) $\gamma_{H,t}$, Belgium

Figure 10: Other state variables, Belgium

7 Causality and Policy Trade-offs

This section illustrates how to exploit the outcomes from the estimated epidemiological model in Section 4 for causality and policy trade-off analysis. We measure, with two different approaches, the effects of mobility curtailment policies on the spread of the virus, the death toll, and economic activity using the smoothed estimates of variables such as the effective reproduction number, new cases, deaths in hospitals, and deaths at home. These exercises gauge the trade-off between slowing down the spread of the virus and decreasing economic activity that policymakers face when enacting shelter-in-place and/or compulsory business closure orders.

Why do we need a causality assessment? Starting in March 2020, individuals' mobility in most countries plummeted due to COVID-19, slowing down the virus's spread at the cost of lower economic activity. Some of the reductions in mobility were voluntary, as individuals took extra precautions to avoid getting infected (or were affected by other individuals taking such measures; for instance, a household canceling its home cleaning services to avoid having third

parties come inside its dwelling reduces the mobility of the workers of the cleaning service).[21] Some of the reductions in mobility were triggered by government mandates, such as orders to shelter-in-place or compulsory business closures.

Since both mechanisms coincided in time, in order to ascertain the causality effect of these government mandates on deaths and economic activity, we need to disentangle voluntary and policy-induced changes in mobility, the virus's spread, the death toll, and economic activity. Knowing by how much governments can affect current and future epidemiological conditions and gauging their cost in terms of economic activity are key factors when designing the length and severity of mobility curtailments.

Why are the outputs from the estimated epidemiological model useful? One of the challenges when performing the above described exercises is that some of the relevant variables needed, such as the effective reproduction number, are not directly observable. Furthermore, other variables, such as the number of reported cases, are subject to large, persistent, biased, and time-varying measurement errors. These problems either preclude the application of standard techniques to assert causality in time series or generate flawed results.

The structure imposed by an epidemiological model allows us to tackle these challenges. The dynamics of an epidemiological model discipline the data by enforcing the cross-equation restrictions among the model's states implied by the transitions among compartments dictated by the disease's biological and clinical properties. The likelihood of the model tells us, for example, that in the case of COVID-19, relatively high seroprevalence rates and low reported new cases in Belgium during the first half of 2020 can only be reconciled with large under-reporting of cases. Likewise, the parallel evolution of hospitalizations, deaths in hospitals vs. deaths at home, and the reported new cases during 2020 informs us of the evolution of under-reporting and changes in mortality probabilities. Thanks to our use of an epidemiological model, we transform noisy and biased observations of new cases, new deaths in hospitals and at home, and seroprevalence surveys into useful unobserved outputs, such as smoothed estimates of the effective reproduction number or the (true) number of new cases.

In contrast, one could try to estimate the effective reproduction number using purely statistical methods (e.g., comparing newly reported cases along a moving window), but such an exercise could not correct the changing share of unreported cases. Another option would be to perform the causality and policy analysis exercises without the effective reproduction number, but such a policy analysis would suffer from an omitted variable bias, which we know from other environments (like studying the effects of monetary policy) can be a fatal flaw.

[21] Arnon et al. (2020) estimate that the bulk of reductions in U.S. contact rates and employment came from voluntary changes in behavior. Gupta et al. (2020), Maloney and Taskin (2020), and Andersen (2020) report similar results.

Two procedures for causality assessment We address the effects of government-mandated mobility curtailments on both the spread of the virus and economic activity using the two most salient methods for assessing causality in time series: structural vector autoregressions (SVARs) and local projections (LPs).

First, we build on the tradition of SVARs and identify a government stringency shock by restricting the systematic component of the government stringency policies rule that maps health and economic conditions into mobility curtailments. A stringency shock should be thought as an *unexpected* change in the mobility curtailment policy. The identification assumptions are motivated by the observation that health policymakers have continuously emphasized that their mandates follow a data-driven approach when imposing mobility curtailments. In situations where we can impose credible restrictions on the systematic component of policy, SVARs have been shown to offer reliable answers (Wolf, 2020).

Second, we work with LPs to identify a reproduction shock and analyze how it affects the rest of the variables (including deaths in hospital and at home and economic conditions) depending on the level of government mobility curtailments. A reproduction shock can be thought of as a shock that changes the contagious properties of the virus. This can be due to biological factors (e.g., the spreading of a variant of the virus) or social mechanisms (e.g., the distribution of better facial masks, improvements in ventilation in public spaces, better organization of social distancing). More concretely, we want to measure whether government mobility curtailments affect the transmission of the reproduction shock. LPs are a very flexible approach that allows us to address state dependencies without making strong parametric assumptions.

While our selection of methods is illustrative, it is not exhaustive. Many other methods for assessing causality can use the output from our estimated epidemiological model, such as a vector error correction model, regression discontinuity designs that exploit local variations in government mandates (e.g., Goolsbee and Syverson, 2020), synthetic controls (e.g., Cho, 2020), event studies (e.g., Gupta et al., 2020), or many of the other ideas in Imbens and Rubin (2015). We could also generalize our SVAR analysis to a Markov-switching SVAR à la Sims and Zha (2006). We skip all those additional experiments to keep our study tightly focused.

7.1 Government Stringency Shock

We write our SVAR as:

$$\boldsymbol{y}_t'\boldsymbol{A}_0 = \boldsymbol{x}_t'\boldsymbol{A}_+ + \boldsymbol{\varepsilon}_t' \quad \text{for } 1 \leq t \leq T,$$

where $\boldsymbol{y}_t$ is an $n \times 1$ vector of endogenous variables, $\boldsymbol{x}_t' = \begin{bmatrix} \boldsymbol{y}_{t-1}' & \cdots & \boldsymbol{y}_{t-p}' & \boldsymbol{z}_t' & 1 \end{bmatrix}$, $\boldsymbol{z}_t$ is a $z \times 1$ vector of exogenous variables, $\boldsymbol{\varepsilon}_t$ is an $n \times 1$ vector of structural shocks, $\boldsymbol{A}_0$ is an $n \times n$ invertible matrix of parameters, $\boldsymbol{A}_+$ is a $(np + z + 1) \times n$ matrix of parameters, p is the lag length, and T is the sample size. The vector $\boldsymbol{\varepsilon}_t$, conditional on past information and the initial

conditions $\boldsymbol{y}_0, ..., \boldsymbol{y}_{1-p}$, is Gaussian with mean zero and covariance matrix $\boldsymbol{I}_n$ (the $n \times n$ identity matrix). The matrices $\boldsymbol{A}_0$ and $\boldsymbol{A}_+$ are the structural parameters.

One of the equations in the SVAR characterizes the policymaker's behavior when imposing mobility curtailments aimed at slowing the transmission of the virus. Consequently, we call such an equation the government stringency policy rule.[22]

We summarize the decisions of the policymakers using a government stringency policies indicator that we will describe below. We will call this indicator the policy instrument. We assume the indicator reacts to other variables in the system such as new cases and the effective reproduction number. Without loss of generality, we assume that the first equation of the SVAR characterizes the policy rule. This implies that:

$$\boldsymbol{y}_t' \boldsymbol{a}_{0,1} = \boldsymbol{x}_t' \boldsymbol{a}_{+,1} + \boldsymbol{z}_t' \boldsymbol{g} + \boldsymbol{\varepsilon}_{1t} \quad \text{for } 1 \leq t \leq T,$$

is the policy equation, where $\boldsymbol{\varepsilon}_{1t}$ denotes the first entry of $\boldsymbol{\varepsilon}_t$, $\boldsymbol{a}_{+,1}$ denotes the first column of $\boldsymbol{A}_+$ for $0 \leq \ell \leq p$, and $a_{s,ij}$ for $s \in \{0, +\}$ denotes the (i, j) entry of $\boldsymbol{A}_s$ and describes the systematic component of the policy rule. Thus, restricting the systematic component of the policy rule is equivalent to restricting $a_{s,ij}$ for $s \in \{0, +\}$ and identifies a policy shock that we call the stringency shock.

Next, we describe the SVAR specification in more detail and how we restrict the policy rule. Our baseline SVAR sample runs from March 21 through November 30 and contains seven endogenous variables. Three of these seven endogenous variables come from outside sources. First, we use the Oxford Stringency (OS) index for Belgium as our mobility curtailment policies indicator (Hale et al., 2020). The authors of the OS index compile information on when and which measures governments take. The particular index we use is a simple average of nine individual component indicators. Each component is an ordinal measure of closings of schools and universities, closings of workplaces, canceling of public events, limits on gatherings, the closing of public transportation, orders to shelter-in-place, curtailments on internal movement between cities/regions, prohibitions on international travel for non-citizens, and the presence of public information campaigns. Due to data limitations, we do not consider non-mobility-related NPIs such as face-masks or improved ventilation.

Second, we use a mobility index for Belgium from the Google COVID-19 Community Mobility Reports, available at https://www.google.com/covid19/mobility/. This measure of mobility is a simple average of the measures of nonresidential mobility categories in the Google Mobility Reports (excluding parks): i) Retail and recreation; ii) Grocery and pharmacy; iii) Transit stations; and iv) Workplaces. The mobility measure is expressed in percentage points and it

[22] We could have several policy equations, each capturing one different containment policy as a function of public health variables, as in Chernozhukov, Kasahara, and Schrimpf (2021). We keep our analysis to one policy equation to avoid overparameterizing the SVAR.

corresponds to daily changes in mobility relative to a baseline value for that day of the week, which is the median value observed during the 5-week period Jan 3–Feb 6, 2020.

Third, we use a daily economic news sentiment (ENS) indicator for Belgium constructed by Algaba et al. (2021) using natural language processing as a daily index of economic activity. This indicator is based on the media archive of the national Belgian News Agency. When aggregated at a monthly frequency, the index is positively correlated with the National Bank of Belgium's monthly consumer confidence survey and other measures of economic activity such as construction, manufacturing, business-related services, and industrial confidence in the euro area, among others. The daily ENS indicator was obtained in two formats: a latent daily series and a 14-day moving average of the series. We use the latter because the former is too noisy.

The other four endogenous variables come from the estimated epidemiological model: i) the model-implied point-wise median smoothed estimate of the effective reproduction number; ii) the model-implied point-wise median smoothed estimate of new cases (as reported in Figure 5); iii) the model-implied point-wise median smoothed estimate of daily deaths per capita in hospitals; iv) and the model-implied point-wise median smoothed estimate daily deaths per capita at home (the two estimates of deaths per capita as reported in Section 6).

The stringency index, the effective reproduction number, and new cases are expressed in log percent. Daily deaths enter in levels so that we can compute the impulse response function (IRF) of cumulative deaths to a stringency shock. The SVAR includes 14 lags, a constant term, and, as an exogenous variable, the average daily temperature to control for the effect that weather conditions might have on the variables of interest. The observable corresponds to the temperature measured at the Brussels Airport Station and it was downloaded from https://www.wunderground.com/. While in general weather conditions are heterogeneous within a country, our measure is representative of the weather conditions for Belgium as a whole given that this nation's territory is only about 30,689 km^2 and most of the population is concentrated in the flat coastal plain and central plateau.

The SVAR is estimated with a Bayesian approach following Arias, Rubio-Ramírez, and Waggoner (2018). We impose a normal-generalized-normal (NGN) prior distribution over the structural parameters $(\boldsymbol{A}_0, \boldsymbol{A}_+)$. The NGN prior is a conjugate prior characterized by four parameters $(\nu, \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{\Omega})$. The parameters ν and $\boldsymbol{\Phi}$ govern the marginal prior distribution of $vec(\boldsymbol{A}_0)$: if $\nu = n$ –as will be the case in our application– $vec(\boldsymbol{A}_0)$ is normally distributed with mean zero and variance $\boldsymbol{\Phi}^{-1}$. The remaining parameters $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ govern the prior distribution of $vec(\boldsymbol{A}_+)$, conditional on $\boldsymbol{A}_0$. Such a distribution is normal with mean $\boldsymbol{\Psi} vec(\boldsymbol{A}_0)$ and variance $\boldsymbol{\Omega}$. We set $\nu = n$, $\boldsymbol{\Phi} = \boldsymbol{0}_{n,n}$, $\boldsymbol{\Psi} = \boldsymbol{0}_{mn,n^2}$, and $\boldsymbol{\Omega}^{-1} = \boldsymbol{0}_{mn,mn}$.

The policy equation is identified with sign restrictions on the systematic component of the policy actions in line with Arias, Caldara, and Rubio-Ramírez (2019) and the SVAR tradition of Leeper, Sims, and Zha (1996). The identification restrictions are:

Restriction 1. *The stringency index is the mobility curtailment policies indicator, and it reacts contemporaneously and positively to the effective reproduction number, mobility, new cases, deaths, and the index of economic activity.*

Restriction 1 embodies the idea that policymakers react to current public health conditions and economic activity. For example, on March 17, 2020, the Belgian government announced the first nationwide lockdown arguing that: "The situation has evolved and forced us to take severe measures to stem the spread of the virus."[23] On April 15, the government relaxed some measures since: "We are aware that the measures taken will have serious long-term consequences, both psychologically and economically. We plead that the measures last as long as necessary."[24] On October 16, the Belgian government ordered bars and restaurants to close, justifying the measure with sentences such as: "The number of confirmed cases is rising, every day, and not just by a few percentage points." Other sentences such as: "This virus is affecting our country in a very hard way," "Thirty-five people died yesterday from the effects of COVID-19," and "In the days to come, the news will be bad" make clear that policymakers were reacting to public health conditions in real time.[25]

Figure 11 plots (in brown lines) the IRFs to a positive (increase) stringency shock and (in yellow bands) the 68% point-wise posterior probability bands. The stringency shock is normalized such that, upon impact, the posterior median increase in the stringency index equals 7.4%. This is equivalent to a one-unit increase in one of the ordinal measures that composed the OS index. For instance, this change corresponds to a nationwide measure that unexpectedly increases the "shelter-in-place" component of the index from 1 (which recommends not leaving the house) to 2 (which requires not leaving the house, with exceptions for daily exercise, grocery shopping, and essential trips). Other examples include an unexpected nationwide increase in the limits on gatherings from 2 (curtailments on gatherings between 101-1000 people) to 3 (curtailments on gatherings between 11-100 people).

A positive stringency shock leads to a drop in the effective reproduction number for more than a month. This drop is *beyond* the fall triggered by individuals' endogenous responses due to changed risk conditions (which are controlled for by the coefficients in the SVAR). Mobility, new cases, and deaths also decline, albeit with different degrees of persistence. In the case of mobility, the decline is more transient than in the case of the effective reproduction number, suggesting an exhaustion effect among individuals: after around three weeks, mobility is back to its baseline value without a stringency shock. The decline in deaths in hospitals and at home

[23] See the March 17, 2020 Reuters article "Belgium to impose coronavirus lockdown from Wednesday."

[24] See the April 15, 2020 EURACTIC article "Belgium extends COVID-19 lockdown until 3 May, but relaxes some measures."

[25] See the October 16, 2020 Reuters article "Belgium to close all bars and restaurants for a month, imposes night curfew," and the AP article "Belgium imposes Covid curfew, closes bars and restaurants," on the same day. Notice that, in comparison with Chernozhukov et al. (2021), we allow for policy to react to behavior.

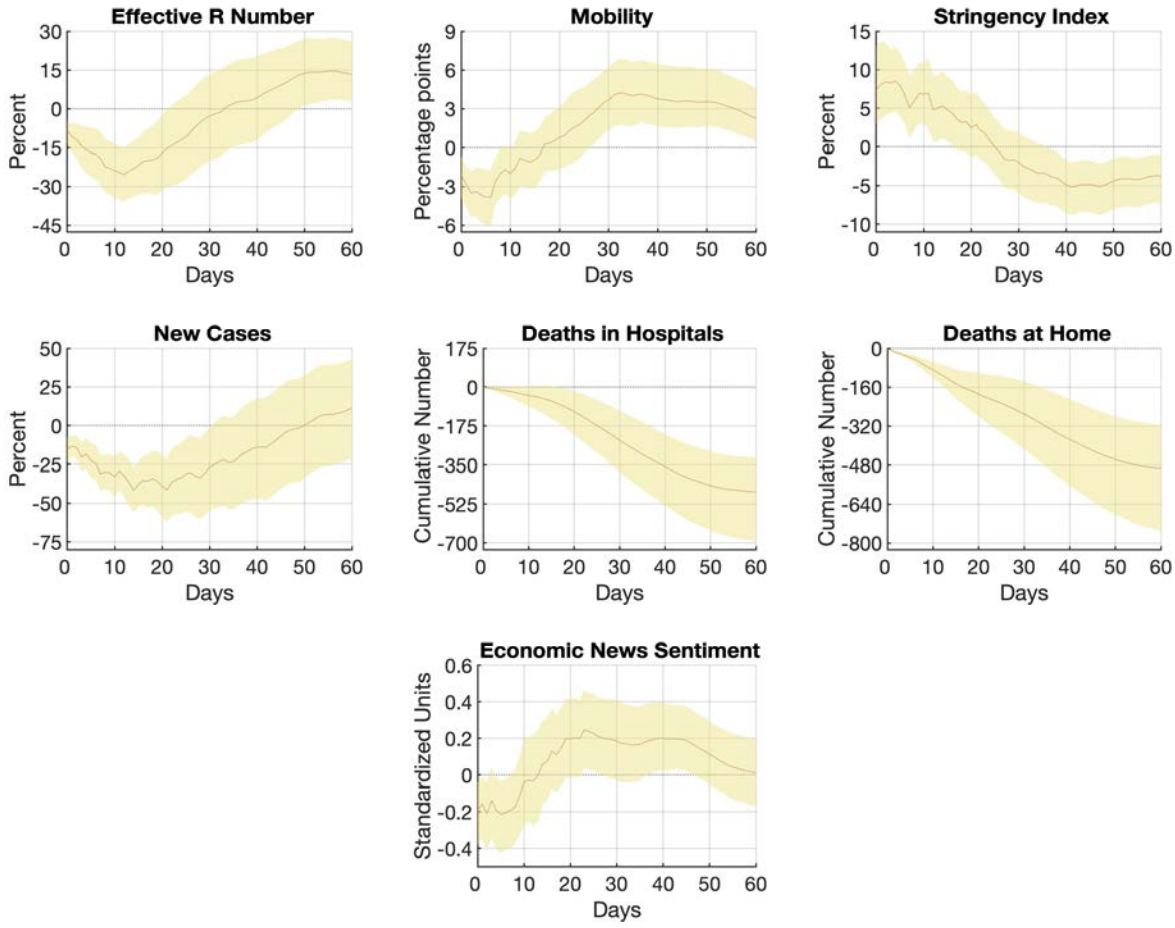


Figure 11: IRFs of a one standard deviation positive stringency shock. The solid curves represent the point-wise posterior medians, and the shaded areas represent the 68% equal-tailed point-wise probability bands. The figure is based on 10,000 independent draws.

lasts much longer, reflecting the illness's lag effects. We measure that a positive stringency shock leads to roughly $1,000$ fewer deaths (500 in hospitals and 500 at home) after 2 months. This represents about 6% of the total number of deaths in our sample, $16,840$.

To interpret the effects of a positive stringency shock in terms of output, we scale the ENS indicator following Lewis et al. (2020). We aggregate the daily ENS indicator to quarterly frequency, denote the resulting series by ENS_q, and run a regression of four-quarter real GDP growth 1 quarter ahead on a constant and ENS_q over the sample 2000Q1-2019Q4, i.e.,

$$\text{RGDPq4q4}_{q+1} = \alpha + \gamma \text{ENS}_q + u_t, \text{ where } q \in \{2000\text{Q1}, \ldots, 2019\text{Q3}\}.$$

We estimate this regression by OLS and obtain $\hat{\alpha} = 1.6$ and $\hat{\gamma} = 1.2$. Based on these point estimates, we can express the news sentiment indicator in terms of real GDP: A decline of 0.2 standardized units in sentiment implies that, if such a value were to persist for an entire quarter,

we would expect (on average) real GDP in the next quarter to be about 0.25% lower than in the absence of the shock.

The pointwise posterior median IRF of the ENS indicator to a stringency shock is on average about 0.04 during the first 90 days, with a 68% probability interval of $[-0.1459, 0.2298]$. If we take the posterior median 0.04 and compute the equivalent in terms of the Belgian GDP per capita, we find that a stringency shock brings a net per capita gain of €4.2. Repeating the same exercise for the two extremes of the probability interval, we obtain the result that the per capita cost(-)/benefit(+) of a stringency shock is between −€14.7 and +€23.1.[26]

This finding suggests that the effect of a positive stringency shock on economic activity is negligible (given the uncertainty, it is hard to tell whether there is a net cost or benefit). In other words: in our sample, and given the systematic component of the health policy of the Belgian government and the voluntary changes in behavior, a marginal and unexpected tightening of mobility curtailments would have saved many lives with close to a zero impact on income per capita. By controlling the virus's spread, a positive stringency shock has a short-term output cost for 15 days, but increases economic activity later. As the stringency measures improve public health conditions, the policy's systematic component drives the stringency index below zero 25 days after the initial increase. In comparison, the reproduction number starts to increase only 40 days after the shock. This difference in timing is consistent with the presence of transient precautionary behavior by the public.

The systematic component of the health policy rule Beyond the analysis of IRFs, SVARs provide useful information regarding the systematic component of the government-mandates policy implied by our identification scheme.[27] More specifically, the contemporaneous coefficients correspond to ratios of entries in the vector $\boldsymbol{a}_{0,1}$. Thus, abstracting from lags and the constant term, the health policy equation can be written as:

$$\mathrm{str}_t = \psi_R \mathrm{re}_t + \psi_M \mathrm{m}_t + \psi_{-\Delta S} \mathrm{nc}_t + \psi_{D_H} \mathrm{ndh}_t + \psi_{D_P} \mathrm{ndp}_t + \psi_{\mathrm{ns}} \mathrm{ns}_t + \sigma \varepsilon_{1,t},$$

where str denotes the stringency index in log-percent, r_t denotes the effective reproduction number expressed in log-percent, m_t denotes mobility in percentage points, nc_t denotes the log of new cases in log-percent, ndh_t denotes the number of new deaths in hospitals, ndp_t denotes the number of new deaths at home, and ns_t denotes the ENS indicator. Accordingly, the coefficient

[26] See Table A.1 in Appendix A.3 for the regression results and see Figure A.5 in Appendix A.3, which plots the ENS indicator in real GDP units along with one-quarter-ahead four-quarter real GDP growth. Given the short time span (90 days) and the low real interest rates prevailing at the moment, discounting the GDP flows to put them in present terms does not make any quantitative difference. Notice that we cannot measure the welfare effect of the positive stringency shock.

[27] This is a common practice when identifying monetary or fiscal policy equations in SVARs: e.g., Leeper and Zha (2003), Sims and Zha (2006), and Caldara and Kamps (2017).

$\psi_R = -\boldsymbol{a}_{0,11}/\boldsymbol{a}_{0,31}$ denotes the contemporaneous response of the stringency index to the effective reproduction number, $\psi_M = -\boldsymbol{a}_{0,21}/\boldsymbol{a}_{0,31}$ denotes the contemporaneous response of the stringency index to mobility, $\psi_{-\Delta S} = -\boldsymbol{a}_{0,41}/\boldsymbol{a}_{0,31}$ denotes the contemporaneous response of the stringency index to new cases, $\psi_{D_H} = -\boldsymbol{a}_{0,51}/\boldsymbol{a}_{0,31}$ denotes the contemporaneous response of the stringency index to new deaths in hospitals, $\psi_{D_P} = -\boldsymbol{a}_{0,61}/\boldsymbol{a}_{0,31}$ denotes the contemporaneous response of the stringency index to new deaths at home, $\psi_{\text{ns}} = -\boldsymbol{a}_{0,71}/\boldsymbol{a}_{0,31}$ denotes the contemporaneous response of the stringency index to economic conditions as measured by the ENS indicator, and $\sigma = 1/\boldsymbol{a}_{0,31}$ is the standard deviation of the health policy shock.

Table 3: Contemporaneous Coefficients in the Health Policy Equation

Coefficient	ψ_R	ψ_M	$\psi_{-\Delta S}$	ψ_{D_H}	ψ_{D_P}	ψ_{ns}
Median	0.86	3.32	0.49	5.37	1.83	34.05
68% Prob. Int.	[0.21;3.37]	[0.78;13.06]	[0.12;1.68]	[1.43;18.78]	[0.43;6.73]	[9.24;115.54]
90% Prob. Int.	[0.06;10.06]	[0.24;42.53]	[0.04;4.45]	[0.45;51.47]	[0.12;19.56]	[2.85;321.06]

Note: The table's entries denote the posterior median estimates of the contemporaneous coefficients in the health policy equation under our identification. The 68% and 90% equal-tailed posterior probability intervals are reported in brackets. The table is based on 10,000 independent draws.

Table 3 reports the posterior distribution of the contemporaneous coefficients in the health policy equation. The posterior median of ψ_R equals 0.93, which implies that the stringency index increases by about 9.3% in response to a 10% increase in the reproduction number. As a reference point, recall that a 7.4% rise in the index is equivalent to a one-unit increase in one of the ordinal measures that composed the OS index. The posterior median of ψ_M equals 3.3, indicating that the stringency index increases by about 16.5% (where 16.5=3.3 $\times$ 5) in response to a 5 percentage point increase in daily mobility. The posterior median of $\psi_{-\Delta S}$ equals 0.5, that is, the stringency index increases by about 5% in response to a 10% rise in the daily new cases. The posterior median of ψ_{D_H}, 5.73, suggests that the stringency index increases by about 57.3% in response to 10 new daily deaths in hospitals. The posterior median of ψ_{D_P}, 1.94, means that the stringency index increases by about 10.94% in response to 10 new daily deaths at home. The posterior median of ψ_{ns}, 34.05, means that the stringency index decreases by about 17.02% in response to a half-standard-deviation decrease in the ENS indicator. In terms of output, this means stringency increases 17.02% (i.e., about a one-unit increase in two of the ordinal measures that composed the OS index) in response to a drop of around 0.6% in next quarter real GDP. Overall, the coefficients are plausible given the behavior of health authorities in 2020 across the advanced economies. The 68% and 90% probability intervals are wide, but one could easily extend our analysis to impose bounds on coefficients based on external evidence or judgment; see Arias, Caldara, and Rubio-Ramírez (2019). We skipped this venue because we want to present our result as free of prior effects as possible.

7.2 A Reproduction Shock and Government Stringency Level

We next identify a reproduction shock and analyze whether its effects on the virus's spread, deaths, and economic activity depend on the level of government stringency. To answer this state-dependency question we use the LP approach proposed by Jordà (2005) and further developed by Ramey and Zubairy (2018), Stock and Watson (2018), and Plagborg-Møller and Wolf (2021), among others. In particular, we use LPs with interaction terms as in Ramey and Zubairy (2018) to study how a reproduction shock propagates depending on the level of government stringency. Our identification scheme consists of sign and zero restrictions implemented as described in Plagborg-Møller and Wolf (2021). Consider the following LP specification:

$$\begin{aligned} \boldsymbol{w}_{i,t+h} = I_{t-1} & \left(\boldsymbol{\mu}_{\mathbb{H},i,h} + \boldsymbol{\beta}'_{\mathbb{H},i,h} \boldsymbol{w}_t + \sum_{\ell=1}^{\nu} \boldsymbol{\delta}_{\mathbb{H},i,h,\ell} \boldsymbol{w}_{t-\ell} + \boldsymbol{\gamma}'_{\mathbb{H},i,h} \boldsymbol{s}_t \right) \\ + (1 - I_{t-1}) & \left(\boldsymbol{\mu}_{\mathbb{L},i,h} + \boldsymbol{\beta}'_{\mathbb{L},i,h} \boldsymbol{w}_t + \sum_{\ell=1}^{\nu} \boldsymbol{\delta}_{\mathbb{L},i,h,\ell} \boldsymbol{w}_{t-\ell} + \boldsymbol{\gamma}'_{\mathbb{L},i,h} \boldsymbol{s}_t \right) + \boldsymbol{\xi}_{i,h,t} \end{aligned} \tag{10}$$

for $i = 1, \ldots, n$, and $h = 0, \ldots, H$, where $\boldsymbol{w}_t$ is the $w \times 1$ vector of endogenous variables equal to $\boldsymbol{y}_t$ defined in Section 7.1 save for the OS index, which is excluded from $\boldsymbol{w}_t$, $\boldsymbol{w}_{i,t+h}$ denotes the value of the i-th variable in $\boldsymbol{w}_{t+h}$, and $\boldsymbol{s}_t$ is an $s \times 1$ vector of exogenous variables.

The exogenous variables include $\boldsymbol{z}_t$ (defined in Section 7.1), the point-wise median smoothed death probability in hospitals, and the point-wise median smoothed death probability at home. I_{t-1} is a dummy variable that indicates whether Belgium is in a high government stringency regime. The high government stringency regime is determined based on whether the OS index is above its median level over the sample (58.33). The parameters $\boldsymbol{\mu}_{\mathbb{H},i,h}$, $\boldsymbol{\beta}'_{\mathbb{H},i,h}$, $\boldsymbol{\delta}_{\mathbb{H},i,h,\ell}$, and $\boldsymbol{\gamma}'_{\mathbb{H},i,h}$ correspond to the high government stringency regime and the parameters $\boldsymbol{\mu}_{\mathbb{L},i,h}$, $\boldsymbol{\beta}'_{\mathbb{L},i,h}$, $\boldsymbol{\delta}_{\mathbb{L},i,h,\ell}$, and $\boldsymbol{\gamma}'_{\mathbb{L},i,h}$ correspond to the low government stringency regime, i.e., the stringency index is below its sample median. The innovation term for $h = 1$, $\boldsymbol{\xi}_{1,t} = (\boldsymbol{\xi}_{1,1,t}, \ldots, \boldsymbol{\xi}_{w,1,t})'$, is assumed to be mean zero with covariance matrix $\mathbb{E}_t(\boldsymbol{\xi}_{1,t}\boldsymbol{\xi}'_{1,t}) = \boldsymbol{\Sigma}$.[28]

All told, the LP specification is similar to the SVAR specification used above with a few modifications tailored to the question at hand. First, we do not include the stringency index because it is the variable we use to split the sample. Second, we use 3 lags to reduce parameter uncertainty. Third, we add death probabilities as exogenous variables to control for any effect related to the sample split. In particular, it could be the case that high (low) government stringency episodes are correlated with high (low) mortality rates. In fact, the death probability was higher at the beginning of our sample as better treatments and refined clinical protocols for the disease were not ready yet. The restrictions to identify the reproduction shock are:

[28] We have also experimented with the case in which the covariance matrix is regime specific. The main conclusions of this section remained unchanged.

Restriction 2. *The reproduction number increases for at least three days in response to a positive reproduction shock. Mobility decreases for at least three days in response to a positive reproduction shock. The impact response of mobility in percentage points is bounded to be smaller than the percentage point increase in the reproduction number. In addition, the reproduction shock does not affect deaths, new cases, and the ENS indicator contemporaneously.*

Restriction 2 identifies what we call a reproduction shock, that is, an exogenous variation in the transmission rate of COVID-19. As we explained above, such an exogenous variation could occur when people relax their compliance with social distancing measures or when a more contagious variant of the virus emerges. The positive sign restrictions on the reproduction number's impact response is a normalization; the positive sign restrictions on the subsequent days are imposed to sharpen identification. We impose just three days to be cautious and let the data dictate the response's shape as much as possible while keeping identification. The negative sign restriction on mobility is imposed based on the notion that, on average, people will stay at home in response to an unexpected increase in reproduction numbers.

The elasticity bound is imposed to discipline the identified set of mobility. In the absence of such a bound, the identified set would include a decline in mobility of 100 percentage points as equally likely to no decline in mobility following an unexpected 10% increase in the effective reproduction number. Such a result is implausible. Hence, we use a bound to rule out dubious IRFs as in Kilian and Murphy (2012) and Arias, Caldara, and Rubio-Ramírez (2019). The zero restrictions on deaths and new cases are predicated on the CDC's reports that it takes more than one day for the symptoms to develop, and on evidence showing that it takes more than one day to die from COVID-19. For example, Wortham, Lee, Althomsons, et al. (2020) report a median clinical course of the disease of 10 days with an interquartile range of 6 to 15 days. The zero restriction on the impact response of the ENS indicator is that it takes at least one day to have broad coverage of the mobility restrictions following the shock and for the audience to process it.

Given the sign and zero restrictions described above, we compute the identified set by numerically solving the quadratic program described in the supplement to Plagborg-Møller and Wolf (2021). In particular, let $\boldsymbol{S}_1$ be a 7xw matrix that selects the IRFs that we restrict to be either positive or negative, and let $\boldsymbol{Z}_1$ be a 4xw matrix that selects the IRFs that we restrict to zero. Then, for each regime r $\in \{H, L\}$, we draw 1 million orthogonal matrices $\boldsymbol{Q}_r$ that satisfy:

$$\begin{bmatrix} \boldsymbol{S}_1 \hat{\boldsymbol{C}}_{0,r} \hat{\boldsymbol{B}} \mathbf{Q}_r \boldsymbol{e}_1 \\ \left(\boldsymbol{e}_2' \hat{\boldsymbol{C}}_{0,r} \hat{\boldsymbol{B}} \mathbf{Q}_r \boldsymbol{e}_1 / \boldsymbol{e}_1' \hat{\boldsymbol{C}}_{0,r} \hat{\boldsymbol{B}} \mathbf{Q}_r \boldsymbol{e}_1 \right) - 1 \end{bmatrix} \geq \mathbf{0} \text{ and } \boldsymbol{Z}_1 \hat{\boldsymbol{C}}_{0,r} \hat{\boldsymbol{B}} \mathbf{Q}_r \boldsymbol{e}_1 = \mathbf{0} \tag{11}$$

where $\hat{\boldsymbol{C}}_{0,r} = \left(\hat{\boldsymbol{\beta}}_{r,1,0}, \ldots, \hat{\boldsymbol{\beta}}_{r,w,0} \right)'$, $\hat{\boldsymbol{\beta}}_{r,j,0}$ denotes the OLS estimate of $\boldsymbol{\beta}_{r,j,0}$ for $j = 1, \ldots, w$, $\hat{\boldsymbol{B}} = \text{chol}(\hat{\boldsymbol{\Sigma}})'$, chol is the upper triangular Cholesky decomposition of $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Sigma}}$ is the OLS estimate

of $\boldsymbol{\Sigma}$, $\boldsymbol{e}_2'\hat{\boldsymbol{C}}_{0,r}\hat{\boldsymbol{B}}\boldsymbol{Q}_r\boldsymbol{e}_1/\boldsymbol{e}_1'\hat{\boldsymbol{C}}_{0,r}\hat{\boldsymbol{B}}\boldsymbol{Q}_r\boldsymbol{e}_1$ denotes the ratio between the impact IRF of mobility and the reproduction number, and $\boldsymbol{e}_i$ denotes the i-th column of the identity matrix.

The numerical computation of the identified set is similar to Algorithm 2 of Giacomini and Kitagawa (2018): given $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{C}}_{0,r}$, we draw $\boldsymbol{Q}_r$ K times and let $\{\boldsymbol{Q}_{r,k} : k = 1, \ldots, K\}$ be the draws that satisfy the sign and zero restrictions in Equation (11). Then, letting $\boldsymbol{q}_{r,k,1}$ denote the first column of $\boldsymbol{Q}_{r,k}$ the identified set for variable i at horizon h is given by

$$\left[\min_k \boldsymbol{e}_i'\hat{\boldsymbol{C}}_{0,r}\hat{\boldsymbol{B}}\boldsymbol{q}_{r,k,1}, \max_k \boldsymbol{e}_i'\hat{\boldsymbol{C}}_{0,r}\hat{\boldsymbol{B}}\boldsymbol{q}_{r,k,1}\right].$$

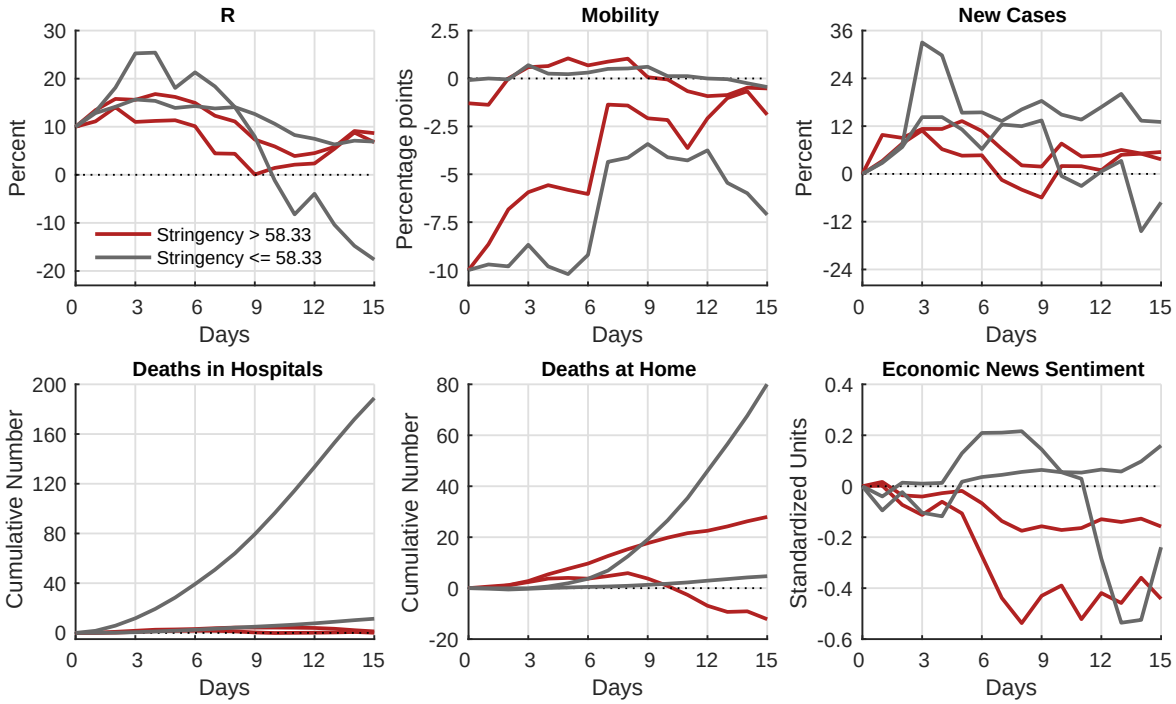


Figure 12: IRFs of a reproduction shock

Figure 12 reports the identified set following a reproduction shock. We show the IRFs from horizon 0 up to horizon 15. The shorter horizon relative to the horizon of the IRFs shown in the SVAR is due to the length of our sample and the parameter uncertainty associated with LPs. Notice, in particular, that the LP looks at the variation in the systematic regime, while the SVAR focuses on the effect of a shock within the regime. Figure 12 truncates the horizon at 15 days, which encompasses the illness duration of patients who died from COVID-19 reported in Wortham, Lee, Althomsons, et al. (2020), who find a median duration of 10 days and an interquartile range of 6 to 15 days, and it is a tad below the findings of Sousa et al. (2020), who report a median illness duration of 19 days and an interquartile range of 12 to 23 days.

In response to a reproduction shock that increases the reproduction number by 10%, the average decrease in the mobility index in the high government stringency regime, −41.3 percentage

points, is larger than the average mobility index in the low government stringency regime, −17.0 percentage points. Consequently, new cases increase by less in the high government stringency regime, leading to fewer deaths (both in hospitals and at home). Another factor explaining the higher level of new cases and deaths in the low government stringency regime is that the increase in the effective reproduction number is much more persistent in such a regime. Our results imply that high government stringency could save up to about 250 deaths in the first two weeks after the reproduction shock. This is around 1.5% of the deaths in our sample. Let us now assess the effects on economic activity of the reproduction number shock under the high and low stringency regimes. In response to the shock, the ENS indicator decreases by about 0.2 to 0.4 standardized units after 7 days in the high stringency regime and increases by about 0 to 0.2 standardized units in the low stringency regime.

As in Subsection 7.1, it is useful to map the standardized units to real per capita GDP terms. After a reproduction shock, a high stringency regime saves 250 lives at a cost of between €2 and €4 per capita. The cost of preventing a death is higher in the case of LPs, but note that the horizon of the IRFs in the case of LPs is 15 days, while in the case of the SVARs it is 60 days, when there is more room for a recovery of economic activity. In any case, the economic cost of a high stringency regime is negligible.

8 Conclusion

This paper has presented methods to estimate epidemiological models using Bayesian methods and applied the results to address causality and policy questions. There are many venues in which to generalize our results. First, we could implement a formal identification analysis along the lines suggested by Poyiadjis et al. (2011). Second, we could introduce a rich network structure between different compartments (for example, reflecting heterogeneity by age, gender, occupation, and location) and estimate how the parameters governing the movements among those compartments evolve (do effective contact rates between regions drop more persistently than effective contact rates within regions?). Examples of these richer data structures appear in Acemoglu et al. (2020) and Aguirregabiria et al. (2020). Third, we could integrate our epidemiological model within a more economic model and estimate how individuals make decisions regarding mobility instead of imposing a random walk variation. Fourth, we could consider how one infectious disease interacts with other contagious diseases. For instance, it has been noted that changes in behavior due to COVID-19 have also caused a dramatic drop in flu contagion; see, e.g., Hills et al. (2020) and Olsen et al. (2020). Fifth, we could study how the introduction of new medical treatments, like a vaccine, affects the spread of the disease, both in terms of changed biological considerations (vaccinated individuals are less likely to get infected) and in terms of endogenous decisions (vaccinated individuals are less likely to follow precautionary behavior).

References

Abouk, R. and B. Heydari (2021). The Immediate Effect of COVID-19 Policies on Social-Distancing Behavior in the United States. *Public Health Reports 136*(2), 245–252.

Acemoglu, D., V. Chernozhukov, I. Werning, and M. D. Whinston (2020). A Multi-Risk SIR Model with Optimally Targeted Lockdown. Technical report, MIT.

Aguirregabiria, V., J. Gu, Y. Luo, and P. Mira (2020, May). A Dynamic Structural Model of Virus Diffusion and Network Production: A First Report. Technical Report 14750, C.E.P.R. Discussion Papers.

Alahmadi, A., S. Belet, A. Black, D. Cromer, J. Flegg, T. House, P. Jayasundara, J. Keith, J. McCaw, R. Moss, J. Ross, F. Shearer, S. Tun, J. Walker, L. White, J. Whyte, A. Yan, and A. Zarebski (2020). Influencing Public Health Policy with Data-Informed Mathematical Models of Infectious Diseases: Recent Developments and New Challenges. *Epidemics 32*, 100393.

Algaba, A., S. Borms, K. Boudt, and B. Verbeken (2021). Daily News Sentiment and Monthly Surveys: A Mixed–Frequency Dynamic Factor Model for Nowcasting Consumer Confidence.

Andersen, M. (2020). Early Evidence on Social Distancing in Response to COVID-19 in the United States. *Available at SSRN 3569368*.

Arevalo-Rodriguez, I., D. Buitrago-Garcia, D. Simancas-Racines, P. Zambrano-Achig, R. del Campo, A. Ciapponi, O. Sued, L. Martinez-Garcia, A. Rutjes, N. Low, P. M. Bossuyt, J. A. Perez-Molina, and J. Zamora (2020). False-Negative Results of Initial RT-PCR Assays for COVID-19: A Systematic Review. *medRxiv*.

Arias, J. E., D. Caldara, and J. F. Rubio-Ramírez (2019). The Systematic Component of Monetary Policy in SVARs: An Agnostic Identification Procedure. *Journal of Monetary Economics 101*, 1–13.

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018). Inference Based on Structural Vector Autoregressions Identified with Sign and Zero Restrictions: Theory and Applications. *Econometrica 86*(2), 685–720.

Arnon, A., J. Ricco, and K. Smetters (2020). Epidemiological and Economic Effects of Lockdown.

Arroyo-Marioli, F., F. Bullano, S. Kucinskas, and C. Rondón-Moreno (2021). Tracking R of COVID-19: A New Real-Time Estimation using the Kalman Filter. *PLOS One 16*(1), e0244474.

Atkeson, A., K. Kopecky, and T. Zha (2020). Four Stylized Facts about COVID-19.

Avery, C., W. Bossert, A. Clark, G. Ellison, and S. F. Ellison (2020, November). An Economist's Guide to Epidemiology Models of Infectious Disease. *Journal of Economic Perspectives 34*(4), 79–104.

Bar-On, Y. M., A. Flamholz, R. Phillips, and R. Milo (2020). Science Forum: SARS-CoV-2 (COVID-19) by the Numbers. *eLife 9*, e57309.

Black, A. J. (2019). Importance Sampling for Partially Observed Temporal Epidemic Models. *Statistics and Computing 29*(4), 617–630.

Bognanni, M., D. Hanley, D. Kolliner, and K. Mitman (2020). Economics and Epidemics: Evidence from an Estimated Spatial Econ-SIR Model. IZA Discussion Paper 13797, Institute of Labor Economics (IZA), Bonn.

Broemeling, L. D. (2021). *Bayesian Analysis of Infectious Diseases: COVID-19 and Beyond*. CRC Press.

Caldara, D. and C. Kamps (2017). The Analytics of SVARs: A Unified Framework to Measure Fiscal Multipliers. *The Review of Economic Studies 84*(3), 1015–1040.

Catteau, L., N. Dauby, M. Montourcy, B. E., J. Hautekiet, E. Goetghebeur, S. van Ierssel, E. Duysburgh, H. Van Oyen, C. Wyndham-Thomas, D. Van Beckhoven, and H. S. Belgian Collaborative Group on COVID-19 (2020). Low-dose Hydroxychloroquine Therapy and Mortality in Hospitalised Patients with COVID-19: A Nationwide Observational Study of 8075 Participants. *International Journal of Antimicrobial Agents 56*(4), 106144.

Chatzi, E. N. and A. W. Smyth (2009). The Unscented Kalman Filter and Particle Filter Methods for Nonlinear Structural System Identification with Non-collocated Heterogeneous Sensing. *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures 16*(1), 99–123.

Chernozhukov, V., H. Kasahara, and P. Schrimpf (2021). Causal Impact of Masks, Policies, Behavior on Early COVID-19 Pandemic in the U.S. *Journal of Econometrics 220*(1), 23–62. Pandemic Econometrics.

Cho, S.-W. S. (2020, 08). Quantifying the Impact of Non-Pharmaceutical Interventions during the COVID-19 Outbreak – The Case of Sweden. *The Econometrics Journal*.

Chudik, A., M. H. Pesaran, and A. Rebucci (2020). Voluntary and Mandatory Social Distancing: Evidence on COVID-19 Exposure Rates from Chinese Provinces and Selected Countries. Working Paper 27039, National Bureau of Economic Research.

Cori, A., N. Ferguson, C. Fraser, and S. Cauchemez (2013). A New Framework and Software to Estimate Time-Varying Reproduction Numbers during Epidemics. *American Journal of Epidemiology 178*(9), 1505–1512.

Courtemanche, C., J. Garuccio, A. Le, J. Pinkston, and A. Yelowitz (2020). Strong Social Distancing Measures in the United States Reduced the COVID-19 Growth Rate. *Health Affairs 39*(7), 1237–1246.

D'Arienzo, M. and A. Coniglio (2020). Assessment of the SARS-CoV-2 Basic Reproduction Number, R0, Based on the Early Phase of COVID-19 Outbreak in Italy. *Biosafety and Health 2*(2), 57 – 59.

Dehning, J., J. Zierenberg, F. P. Spitzner, M. Wibral, J. P. Neto, M. Wilczek, and V. Priesemann (2020). Inferring Change Points in the Spread of COVID-19 Reveals the Effectiveness of Interventions. *Science 369*(6500).

Fernández-Villaverde, J. and J. Rubio-Ramírez (2007). How Structural Are Structural Parameters? *NBER Macroeconomics Annual 22*, 83–137.

Fernández-Villaverde, J., J. Rubio-Ramírez, and F. Schorfheide (2016). Solution and Estimation Methods for DSGE models. In J. B. Taylor and H. Uhlig (Eds.), *Handbook of Macroeconomics*, Volume 2, pp. 527–724. Elsevier.

Fernández-Villaverde, J. and C. I. Jones (2020, May). Estimating and simulating a sird model of covid-19 for many countries, states, and cities. Working Paper 27128, National Bureau of Economic Research.

Giacomini, R. and T. Kitagawa (2018, November). Robust Bayesian Inference for Set-identified Models. (CWP61/18).

Gibbons, C. L., M.-J. Mange, D. Plass, A. H. Havelaar, R. J. Brooke, P. Kramarz, K. L. Peterson, A. L. Stuurman, A. Cassini, E. M. Fèvre, and M. E. Kretzschmar (2014). Measuring Underreporting and Under-Ascertainment in Infectious Disease Datasets: A Comparison of Methods. *BMC Public Health 14*. Article number: 147.

Gilchrist, S. and E. Zakrajšek (2012, June). Credit spreads and business cycle fluctuations. *American Economic Review 102*(4), 1692–1720.

Goolsbee, A. and C. Syverson (2020, June). Fear, Lockdown, and Diversion: Comparing Drivers of Pandemic Economic Decline 2020. Working Paper 27432, National Bureau of Economic Research.

Gostic, K. M., L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn, R. Niehus, J. Hay, P. M. De Salazar, J. Hellewell, S. Meakin, J. Munday, N. I. Bosse, K. Sherrat, R. N. Thompson, L. F. White, J. S. Huisman, J. Scire, S. Bonhoeffer, T. Stadler, J. Wallinga, S. Funk, M. Lipsitch, and S. Cobey (2020). Practical Considerations for Measuring the Effective Reproductive Number, R_t. *PLOS Computational Biology 16*(12).

Gupta, S., T. D. Nguyen, F. L. Rojas, S. Raman, B. Lee, A. Bento, K. I. Simon, and C. Wing (2020). Tracking Public and Private Responses to the COVID-19 Epidemic: Evidence from State and Local Government Actions.

Hale, T., S. Webster, A. Petherick, T. Phillips, and B. Kira (2020). Oxford COVID-19 Government Response Tracker. Available at https://covidtracker.bsg.ox.ac.uk/.

Herzog, S., J. De Bie, S. Abrams, I. Wouters, E. Ekinci, L. Patteet, A. Coppens, S. De Spiegeleer, P. Beutels, P. Van Damme, N. Hens, and H. Theeten (2020). Seroprevalence of Igg Antibodies against SARS Coronavirus 2 in Belgium – A Serial Prospective Cross-Sectional Nationwide Study of Residual Samples. medRxiv.

Hills, T., N. Kearns, C. Kearns, and R. Beasley (2020). Influenza Control during the COVID-19 Pandemic. *The Lancet 396*(10263), 1633–1634.

Ho, P., T. A. Lubik, and C. Matthes (2021). How to Go Viral: A COVID-19 Model with Endogenously Time-Varying Parameters. *Journal of Econometrics*.

Hortaçsu, A., J. Liu, and T. Schwieg (2021). Estimating the fraction of unreported infections in epidemics with a known epicenter: An application to covid-19. *Journal of Econometrics 220*(1), 106–129.

Hsiang, S., D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L. Y. Huang, A. Hultgren, E. Krasovich, et al. (2020). The Effect of Large-Scale Anti-Contagion Policies on the COVID-19 pandemic. *Nature 584*(7820), 262–267.

Hurwicz, L. (1966). On the Structural Form of Interdependent Systems. In *Studies in Logic and the Foundations of Mathematics*, Volume 44, pp. 232–239. Elsevier.

Imbens, G. W. and D. B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review 95*(1), 161–182.

Katul, G. G., A. Mrad, S. Bonetti, G. Manoli, and A. J. Parolari (2020). Global Convergence of COVID-19 Basic Reproduction Number and Estimation from Early-time SIR Dynamics. *PLOS One 16*(4), 351–367.

Kilian, L. and D. P. Murphy (2012). Why Agnostic Sign Restrictions Are Not Enough: Understanding the Dynamics of Oil Market VAR Models. *Journal of the European Economic Association 10*(5), 1166–1188.

Korolev, I. (2021). Identification and Estimation of the SEIRD Epidemic Model for COVID-19. *Journal of Econometrics 220*(1), 63–85.

Kucirka, L. M., S. A. Lauer, O. Laeyendecker, D. Boon, and J. Lessler (2020). Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Annals of Internal Medicine*.

Kypraios, T., P. Neal, and D. Prangle (2017). A Tutorial Introduction to Bayesian Inference for Stochastic Epidemic Models using Approximate Bayesian Computation. *Mathematical Biosciences 287*, 42–53.

Lavery, A. M., L. E. Preston, J. Y. Ko, J. R. Chevinsky, C. L. DeSisto, A. F. Pennington, L. Kompaniyets, S. D. Datta, E. S. Click, T. Golden, et al. (2020). Characteristics of Hospitalized COVID-19 Patients Discharged and Experiencing Same-Hospital Readmission—United States, March–August 2020. *Morbidity and Mortality Weekly Report 69*(45), 1695.

Lee, K., A. Oka, E. Pollakis, and L. Lampe (2010). A Comparison between Unscented Kalman Filtering and Particle Filtering for RSSI-based Tracking. In *2010 7th Workshop on Positioning, Navigation and Communication*, pp. 157–163.

Lee, S., Y. Liao, M. Seo, and Y. Shin (2021). Sparse HP Filter: Finding Kinks in the COVID-19 Contact Rate. *Journal of Econometrics 220*(1), 158–180.

Leeper, E. M., C. A. Sims, and T. Zha (1996). What Does Monetary Policy Do? *Brookings Papers on Economic Activity 1996*(2), 1–78.

Leeper, E. M. and T. Zha (2003). Modest Policy Interventions. *Journal of Monetary Economics 50*(8), 1673–1700.

Lewis, D. J., K. Mertens, J. H. Stock, and M. Trivedi (2020). Measuring Real Activity Using a Weekly Economic Index. *Federal Reserve Bank of New York Staff Reports* (920).

Li, R., S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman (2020). Substantial Undocumented Infection Facilitates the Rapid Dissemination of Novel Coronavirus (SARS-CoV-2). *Science 368*(6490), 489–493.

Llorente, F., L. Martino, D. Delgado, and J. Lopez-Santiago (2020). Marginal Likelihood Computation for Model Selection and Hypothesis Testing: An Extensive Review. *arXiv preprint arXiv:2005.08334*.

Maloney, W. and T. Taskin (2020). *Determinants of Social Distancing and Economic Activity during COVID-19: A Global View*. The World Bank.

Manski, C. F. and F. Molinari (2021). Estimating the COVID-19 Infection Rate: Anatomy of an Inference Problem. *Journal of Econometrics 220*(1), 181–192. Pandemic Econometrics.

Minter, A. and R. Retkute (2019). Approximate Bayesian Computation for Infectious Disease Modelling. *Epidemics 29*, 100368.

Olsen, S. J., E. Azziz-Baumgartner, A. P. Budd, L. Brammer, S. Sullivan, R. F. Pineda, C. Cohen, and A. M. Fry (2020). Decreased Influenza Activity during the COVID-19 Pandemic—United States, Australia, Chile, and South Africa, 2020. *American Journal of Transplantation 20*(12), 3681–3685.

O'Neill, P. D. and G. O. Roberts (1999). Bayesian Inference for Partially Observed Stochastic Epidemics. *Journal of the Royal Statistical Society. Series A (Statistics in Society) 162*(1), 121–129.

Pesaran, M. H. and C. F. Yang (2020). Matching Theory and Evidence on Covid-19 using a Stochastic Network SIR Model.

Plagborg-Møller, M. and C. K. Wolf (2021). Local Projections and VARs Estimate the Same Impulse Responses. *Econometrica (Forthcoming)*.

Poyiadjis, G., A. Doucet, and S. S. Singh (2011). Particle Approximations of the Score and Observed Information Matrix in State Space Models with Application to Parameter Estimation. *Biometrika 98*(1), 65–80.

Ramey, V. A. and S. Zubairy (2018). Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data. *Journal of Political Economy 126*(2), 850–901.

Rees, E. M., E. S. Nightingale, Y. Jafari, N. R. Waterlow, S. Clifford, C. A. Pearson, T. Jombart, S. R. Procter, G. M. Knight, C. W. Group, et al. (2020). COVID-19 Length of Hospital Stay: A Systematic Review and Data Synthesis.

Renne, J.-P., G. Roussellet, and G. Schwenkler (2020). Preventing COVID-19 Fatalities: State versus Federal Policies.

Roberts, G. O. and J. S. Rosenthal (2001). Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science 16*(4), 351–367.

Sims, C. A. and T. Zha (2006). Were There Regime Switches in US Monetary Policy? *American Economic Review*, 54–81.

Smyth, A. W., S. F. Masri, A. G. Chassiakos, and T. K. Caughey (1999). On-line Parametric Identification of MDOF Nonlinear Hysteretic Systems. *Journal of Engineering Mechanics 125*(2), 133–142.

Sousa, G., T. Garces, V. Cestari, R. Florêncio, T. Moreira, and M. Pereira (2020). Mortality and Survival of COVID-19. *Epidemiology & Infection 148*.

Stock, J. H. and M. W. Watson (2018). Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments. *The Economic Journal 128*(610), 917–948.

Toda, A. A. (2020). Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact. U.C.S.D. manuscript.

Toulis, P. (2021). Estimation of Covid-19 Prevalence from Serology Tests: A Partial Identification Approach. *Journal of Econometrics 220*(1), 193–213. Pandemic Econometrics.

Walker, J. N., A. J. Black, and J. V. Ross (2019). Bayesian Model Discrimination for Partially-Observed Epidemic Models. *Mathematical Biosciences 317*, 108266.

Walker, J. N., J. V. Ross, and A. J. Black (2017). Inference of Epidemiological Parameters from Household Stratified Data. *PLOS One 12*(10), e0185910.

Wilson, D. J. (2020). Weather, Mobility, and COVID-19: A Panel Local Projections Estimator for Understanding and Forecasting Infectious Disease Spread. Technical Report Working Paper 2020-23, Federal Reserve Bank of San Francisco.

Wolf, C. K. (2020, October). SVAR (Mis)identification and the Real Effects of Monetary Policy Shocks. *American Economic Journal: Macroeconomics 12*(4), 1–32.

Wortham, J., J. Lee, S. Althomsons, et al. (2020). Characteristics of Persons Who Died with COVID-19 — United States. *Morbidity and Mortality Weekly Report 69(45)*, 923–929.

A Online Appendix

A.1 Implied Density over the Initial Share of Detected Cases

Figure A.1 plots the prior and posterior distributions of the initial share of detected cases, $\gamma_{n,0}$, implied by our model. While the prior assigns substantial probability mass to values as high as 0.6, the posterior distribution is concentrated in values below 0.4, indicating that, according to our model, the share of detected cases on March 15, 2020, was below 40% with high probability.

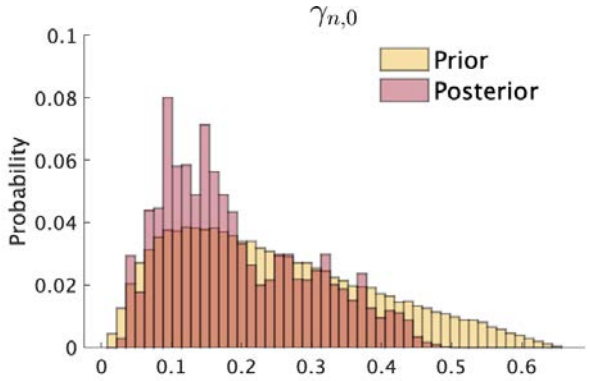


Figure A.1: Prior histograms (yellow) are based on 100,000 independent draws from the prior distribution presented in Table 2. Posterior histograms (red) are based on the MCMC chain with 90,000 posterior draws obtained after a burn-in period of 10,000 draws.

A.2 Prior versus Posterior

Figure A.2 shows the prior and posterior distributions for γ, θ_H, and θ_P. These parameters are inversely related to the average duration a person remains infectious, the average duration of stay in hospitals, and the average duration of stay at home while recovering from COVID-19. Figure A.2 reveals that the data are very informative about γ and θ_H and less informative about θ_P. Figure A.3 shows the prior and posterior distributions for σ_b, σ_h, σ_p, σ_g, and σ_n. These parameters govern the step size of the time-varying parameters of our model. Clearly, the data are informative about them. Finally, Figure A.4 shows the prior and posterior distributions for b_0, d_{H_0}, d_{P_0}, g_{H_0}, S_0, I_0, n_0, and μ. These parameters are the initial-value parameters and the share of false negatives parameter in the case of μ. Figure A.4 documents that data are informative about these 8 parameters as well. The prior for n_0 is truncated at 1 to rule out large values of the permanent component of detected cases. Even so, the posterior indicates that there is not much probability mass near the truncation, suggesting that the upper bound for the prior could be relaxed without affecting our conclusions.

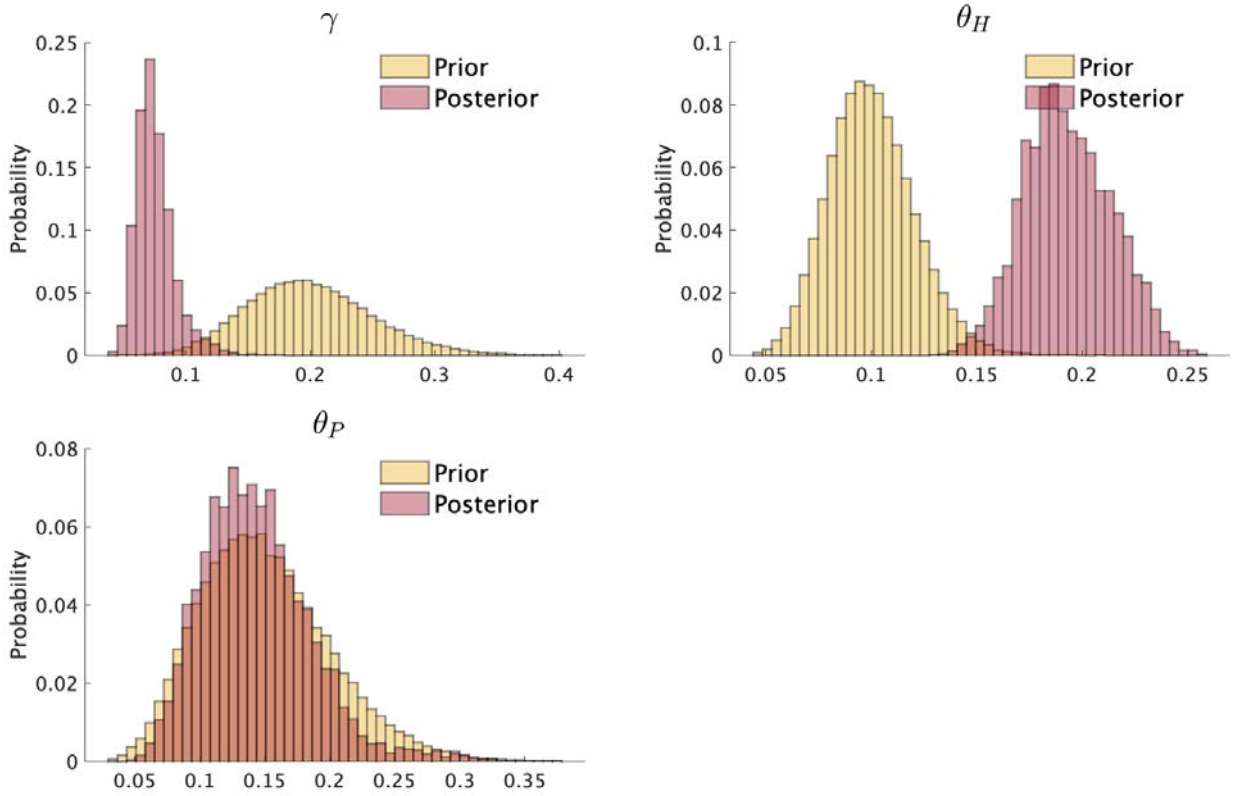


Figure A.2: Duration Parameters. Table 1 presents definitions of these parameters. Prior histograms (yellow) are based on 100,000 independent draws from the prior distribution presented in Table 2. Posterior histograms (red) are based on the MCMC chain with 90,000 posterior draws obtained after a burn-in period of 10,000 draws.

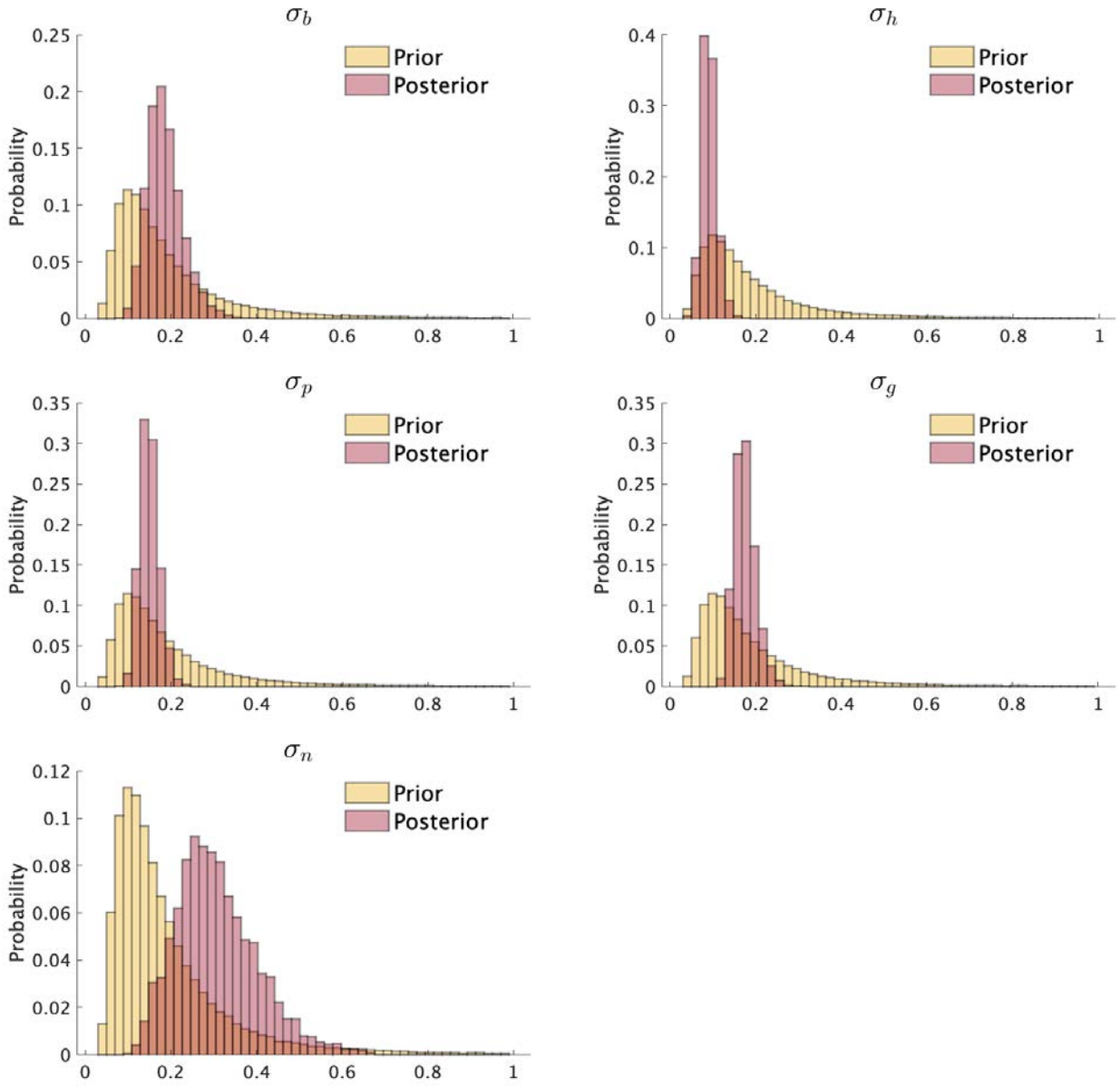


Figure A.3: Step-size Parameters. Table 1 presents definitions of these parameters. Prior histograms (yellow) are based on 100,000 independent draws from the prior distribution presented in Table 2. Posterior histograms (red) are based on the MCMC chain with 90,000 posterior draws obtained after a burn-in period of 10,000 draws.

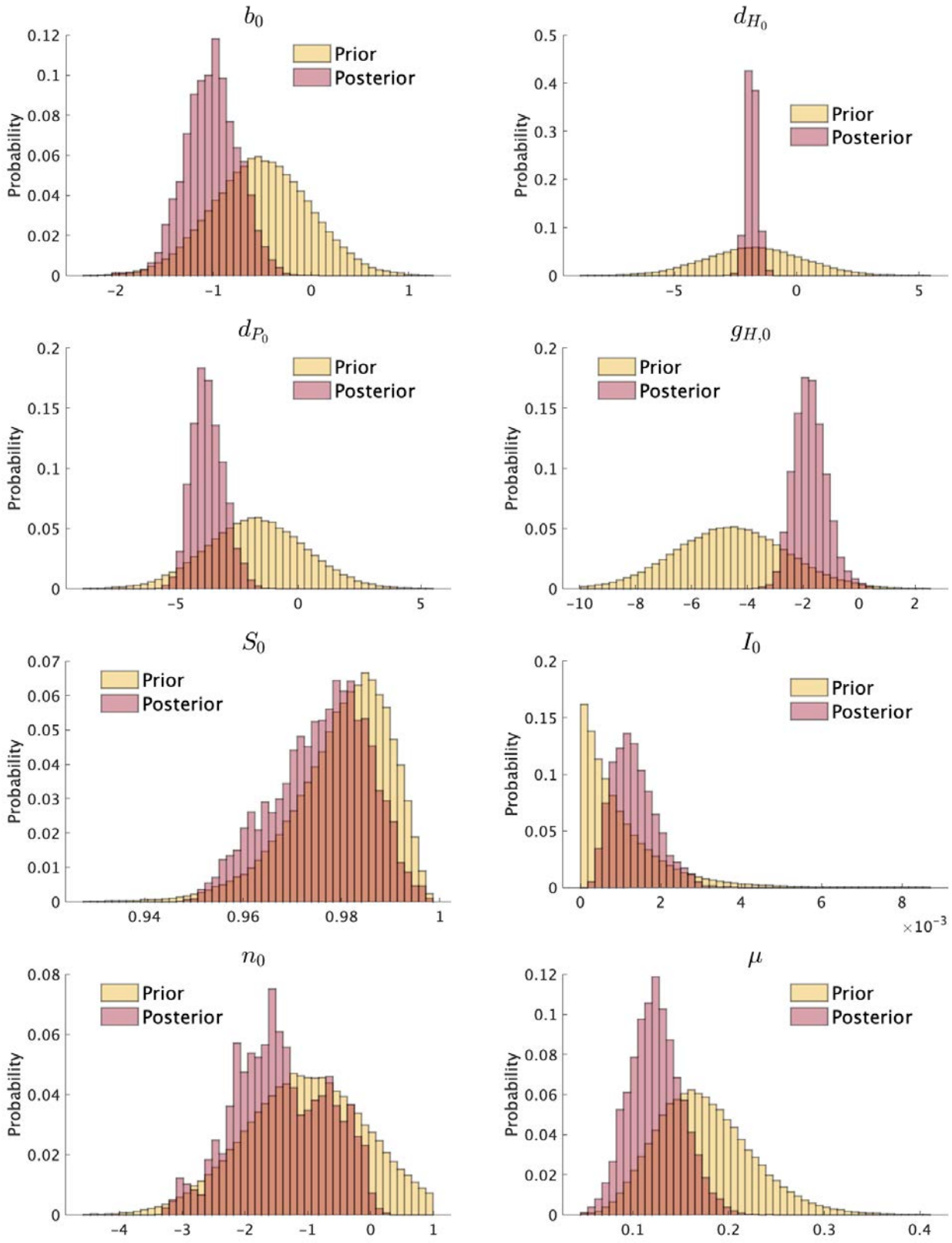


Figure A.4: Initial-value Parameters and Share of False Negatives Parameter. Table 1 presents definitions of those parameters. Prior histograms (yellow) are based on 100,000 independent draws from the prior distribution presented in Table 2. Posterior histograms (red) are based on the MCMC chain with 90,000 posterior draws obtained after a burn-in period of 10,000 draws.

A.3 Sentiment and Real GDP Growth Units

Table A.1 shows the OLS coefficients, the 95% confidence intervals for the coefficient estimates (in brackets), and the R^2 of a regression of four-quarter real GDP growth 1 quarter ahead on a constant and ENS_q over the sample 2000Q1-2019Q4, i.e., $\text{RGDPq4q4}_{q+1} = \alpha + \gamma \text{ENS}_q + u_t$, where $q \in \{2000\text{Q1}, \ldots, 2019\text{Q3}\}$.

Table A.1

Regressors	Coefficients
Constant	1.61 [1.36;1.87]
Economic News Sentiment	1.17 [0.84;1.51]
$R^2 = 0.39$	

Figure A.5 plots the quarterly average of the daily economic news sentiment index expressed in GDP units along with one-quarter-ahead four-quarter real GDP growth.

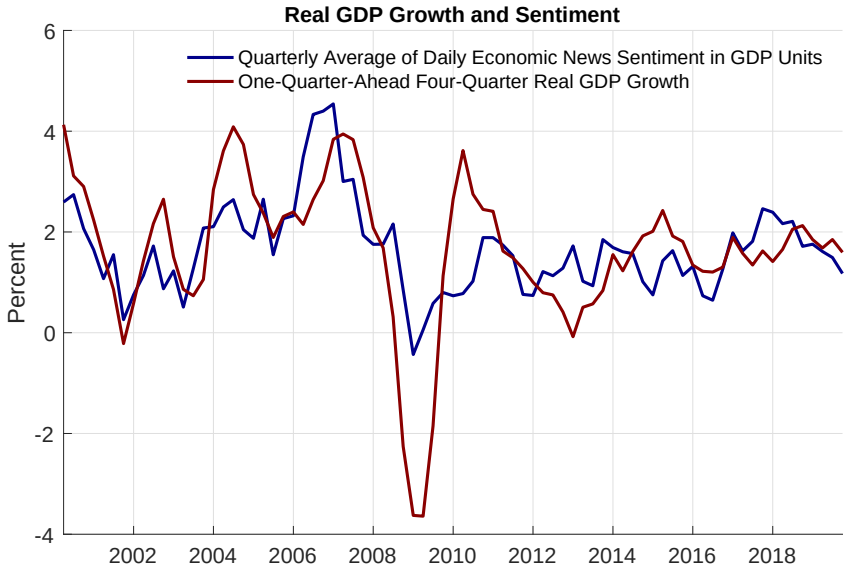


Figure A.5