USFQ School of Economics
Working Paper Series

The analysis of covid-19 surveillance data: What can we learn from limited information?

Carlos Uribe-Teran[(1)], Santiago J. Gangotena[(2)]

[(1)] School of Economics, Universidad San Francisco de Quito, cauribe@usfq.edu.ec
[(2)] School of Economics, Universidad San Francisco de Quito, sjgangotena@usfq.edu.ec

Working Paper No. 0001 Issue Date: July 8, 2021

Abstract

What can we learn about the transmission process of SARS-CoV2 from daily counts of confirmed cases and tests as the pandemic unfolds? In this paper we use classic econometric techniques to filter away stochastic innovations that occur in epidemiological surveillance data that are orthogonal to the infection process. We then compute the effective reproduction number R_t and the test positivity rate ρ_t, and propose an analysis of joint trajectories over the (R_t, ρ_t) space to have a more precise assessment of the current status of the infection. We test our method using an agent-based model with an underlying SIERD infection process and a testing layer that produces *confirmed* positives. We find that epidemiological indicators estimates are systematically biased, but our method allows us to reduce the root mean squared error and the probability of type II errors, particularly when testing is concentrated among symptomatic patients. The joint analysis of R_t and ρ_t manages to reduce the probability of type I and type II classification errors to below 0.5%. Our country analysis shows that daily counts of cases and tests exhibit strong seasonal and atypical components which, if left untreated, can produce spurious dynamics of epidemiological indicators.

Keywords: Covid-19, Statistical Simulation, Time Series Analysis, Agent-Based Modeling, Health.
JEL Codes: C15, C22, C69, I10.

Universidad San Francisco de Quito

SCHOOL OF ECONOMICS USFQ

The Analysis of Covid-19 Surveillance Data: What can we Learn from Limited Information?*

Carlos Uribe-Teran
School of Economics
Universidad San Francisco de Quito
cauribe@usfq.edu.ec

Santiago J. Gangotena
School of Economics
Universidad San Francisco de Quito
sjgangotena@usfq.edu.ec

First Version: April 20, 2020
This Version: July 8, 2021

Abstract

What can we learn about the transmission process of SARS-CoV2 from daily counts of confirmed cases and tests as the pandemic unfolds? In this paper we use classic econometric techniques to filter away stochastic innovations that occur in epidemiological surveillance data that are orthogonal to the infection process. We then compute the effective reproduction number R_t and the test positivity rate ρ_t, and propose an analysis of joint trajectories over the (R_t, ρ_t) space to have a more precise assessment of the current status of the infection. We test our method using an agent-based model with an underlying SIERD infection process and a testing layer that produces *confirmed* positives. We find that epidemiological indicators estimates are systematically biased, but our method allows us to reduce the root mean squared error and the probability of type II errors, particularly when testing is concentrated among symptomatic patients. The joint analysis of R_t and ρ_t manages to reduce the probability of type I and type II classification errors to below 0.5%. Our country analysis shows that daily counts of cases and tests exhibit strong seasonal and atypical components which, if left untreated, can produce spurious dynamics of epidemiological indicators.

Keywords: Covid-19, Statistical Simulation, Time Series Analysis, Agent-Based Modeling, Health.
JEL Codes: C15, C22, C69, I10.

*We are extremely grateful to Julio Acuña, Ivan Sisa, Fadya Orozco, Enrique Teran, Pablo Endara, Andrea Gomez and participants at the USFQ-School of Economics Brown Bag seminar for useful comments on an early version of this paper. We are also thankful to Sebastian Jimenez for outstanding and expedite research assistance.

1. Introduction

The production of epidemiological surveillance statistics relies, among other things, on countries testing capabilities and strategies. Testing, thus, works like a survey with many limitations in its sample design. For example, while a survey can be designed using clustered random sampling to be representative of the population and can have clear protocols in case of no responses, testing has to focus primarily on serving as a diagnostic tool for people that present symptoms of being infected with a given pathogen, and health workers that are in the front line. If resources are limited (as it is the case in most countries around the world), most of the tests being performed during the course of a pandemic would be used for diagnosis, causing severe biases in the different measures that can be computed to monitor the status of the infection process. In light of these limitations, what can we learn about the transmission process of SARS-CoV2 from daily counts of confirmed cases and tests as the pandemic unfolds?

In this paper we concentrate on time-series issues that surveillance data exhibits. For this, we design a method that yields a clearer assessment of the status of the transmission process of a pathogen when the quality of data is poor. To do this we apply classic time-series treatments to filter away seasonal and atypical components, two types of stochastic innovations that can produce severe bias in any type of analysis. Then, we use the trend and cycle components of the resulting time series to uncover the underlying infection process (the data-generating process) that gave rise to the noisy signal we observe in the data.

During the seasonal adjustment stage, we apply the Hodrick-Prescott filter using a smoothing parameter of 1600 to obtain the stationary error term in each series. We believe that this initial calibration gives the filter enough flexibility to detect the seasonal and atypical components in the series. We take advantage of this first stage to endogenize the value of the smoothing parameter as a function of the variance of the seasonally adjusted error term. Specifically, we allow lower values of the smoothing parameter for error terms that exhibit lower variances. We use these smoothing parameters to estimate the trends of the test positivity rate and the daily count of tests.

Then, we combine the estimated trend of the test positivity rate with the trend of testing to emulate the infection process, and build confidence intervals around this trend by block bootstrapping. Using this result and external estimations of the generation time, we produce a simple estimate of the effective reproduction number R_t and compute its confidence interval using the bootstrapped versions of the emulated infection process and log-normally distributed random numbers calibrated to replicate the mean and standard deviation of the generation time.

Due to data limitations (we do everything only with daily counts of confirmed cases and tests), the value that we find for R_t is downward biased when $R_t \geq 1$ and upward biased when $R_t < 1$. We are also worried about the possibility of having type II errors, that is, estimating $R_t < 1$ when the real $R_t \geq 1$. Because of this, we complement this estimation with a joint analysis of R_t and the test positivity rate ρ_t. We divide (R_t, ρ_t) space into four quadrants that allow us to classify trajectories in terms of adequacy of testing (threshold at $\rho_t = 0.10$) and dynamics of the infection process (threshold at $R_t = 1$).

To check the robustness of our estimation procedure and assess how well it is able to reproduce the true dynamics of the infection, we apply our method to simulated data. For this we build an Agent Based Model (ABM) that in addition to the infection process, explicitly models data collection under different testing strategies. This way we can assess the performance of our estimation strategy and investigate how the parameters that govern testing (scale, testing growth rate, testing noise, and the probability of testing symptomatic patients) affect the quality of data gathered from tests, and the results we obtain with our method.

Once we understand how our methodology performs in a controlled environment, we apply it to Covid-19 data for 35 countries obtained from Our World in Data dataset (Max Roser and Hasell, 2020) and official reports of Ecuador's Ministry of Public Health. The result is a sample that exhibits significant heterogeneity regarding data quality, something that is evident when we study the dynamic consistency of our estimates. Although we apply the methodology to the entire sample, we show intermediate results for a sub-sample of 15 countries. The last update for this data occurred on August 13th.

In our simulated environments we find that, for a given growth rate in the number of daily tests, the scale of testing and the strategy used to administer tests have a significant impact on the information about the dynamics of infection that can be estimated using testing data. As can be expected, we find that the effective reproductive number calculated using testing data is systematically biased, with the bias falling when the real effective reproductive number is in the vicinity of one. Increasing the scale of testing significantly reduces this bias as does random testing. Classification of simulated trajectories in (R_t, ρ_t) space allows us to reduce classification errors with respect to the true trajectory to 0.35%. This is orders of magnitude less than the classification error of 6.1% and 23.6% we respectively obtain when looking at the reproduction number or test positivity rate separately.

Our country analysis shows that Covid-19 surveillance data exhibit strong seasonal components. Moreover, most countries atypical values are concentrated at the initial stages of the transmission process. This is expected, since countries experience an adjustment period (import more tests, adapt laboratories to process them, etc.). However, there are countries for which atypical values occur at more recent dates. If these sources of variation are not controlled for, then daily count data might exhibit periodic or unexpected falls (or peaks), showing variations that are completely orthogonal to the infection process.

We also find that the dynamic consistency of our estimation heavily relies on the quality of data, something that is true even for more complicated Bayesian methods. The reason for this is that our estimation might confuse new atypical innovations with changes in the trend, especially if these atypical innovations do not meet the cut to be eliminated in the cleaning phase of our method.

In our joint analysis of the effective reproduction number and the test positivity rate we find that among our sample of observations, this quadrant is characterized by case fatality and mortality rates that are at least 1.5 times the average that we find in other quadrants. What is interesting is that in our simulations, which have an underlying transmission process without any intervention, this quadrant is much more likely to occur. We also find that the odds of observing data in this

quadrant fall as the testing scale increases, and increases as the probability of testing symptomatic patients increase.

We organize this paper as follows. In the next section we look at the literature on the economics of the pandemic and present our contribution. In section 3 we describe the econometric procedure and the Agent-Based-Model that we use to assess the performance of our method. In section 4 present our results. Section 5 concludes.

2. Economics and Epidemiology

The unfolding of the Covid-19 pandemic has awaken the interest of many economists in modeling and understanding the details of disease transmission. Contributions in this line of research can be classified in, at least, two different groups. First, there are papers that are worried about the economic effects of the pandemic. The second group includes papers written by economists that attempt to contribute directly to the epidemiology literature.

Within the first group, there are two different strategies. Early on the onset of the pandemic, Correia et al. (2020) studied how public health interventions might affect the economy. Using data from the 1918 Influenza Pandemic in the U.S., they show that areas that were more exposed to the disease experienced sharper and more persistent declines in economic activity. These results have been tested again by Lilley et al. (2020), who find that such effects are driven by population growth and once differential trends are considered, positive effects of NPIs on economic activity are non-significant. Moreover Barro (2020) shows that NPIs did not have significant effects on curtailing mortality either, and explain that the most likely reason is that NPIs were not in place long enough.

The second strategy consists on using standard macroeconomic models augmented with SIR frameworks to analyze optimal policy responses to the pandemic. In this regard, one of the first papers to appear was Eichenbaum et al. (2020a), where the authors find that the endogenous households' response to cut back on consumption and work reduces the impact of the epidemic in terms of its death toll. They also find that although containment policy increases the severity of the recession, it saves about half a million lives in the U.S. In another effort, Eichenbaum et al. (2020b) show that testing without quarantining might have negative effects, both economic and related to public health. Moreover, a policy that optimally combines testing with quarantining infected individuals reduces significantly the trade-off between declines in economic activity and health outcomes that is triggered by general lockdowns. In the same line, Arellano et al. (2020) study how the pandemic might affect emerging markets, in particular, how lockdown policies might trigger prolonged debt crises.

Other papers look at endogenous reallocation of economic activity and the distributional effects triggered by the containment policy applied to control the pandemic. Again, incorporating SIR frameworks within macroeconomic models, Krueger et al. (2020) show that rational decisions to cut back on consumption and implement additional hygiene measures might allow infections to decline entirely on their own. More on the redistributive part of the story, Glover et al. (2020)

study optimal lockdown policies in an environment where the redistributive effects triggered by the intervention between young and old cohorts is explicitly modeled.

The most salient example for papers in the second group is Manski and Molinari (2020), where the authors study the measurement errors involved in estimating the Covid-19 infection rate, and propose a methodology to reduce this bias. Another attempt is made by Fernández-Villaverde and Jones (2020) that put forward economists' worry regarding the way the transmission process is modeled under the SIR framework. In particular, the assumption that the rate of transmission is an structural parameter. In this line, the authors start from the observation that the rate of transmission is actually endogenous, since it depends on consumption and work decisions made by households. They use information on death counts to recover the rate of transmission and, from here, estimate the effective reproduction number. In line with this contribution, to the best of our knowledge (and to our surprise), there is only one paper in the epidemiology literature that considers how agent's behavior might affect the transmission of disease (Eksin et al., 2019).

We contribute to this branch of the literature, by proposing an estimation framework that relies on daily counts of confirmed cases and tests (widely available) to monitor public policy in environments where access to more complete and perfect information is lacking. We also propose a variation to the SIERD model in which we add a layer with explicit modeling of the data-gathering process, that allows us to have a clearer view on how different testing strategies might trigger bias in the analysis of the transmission process when such analysis is based on surveillance data.

We also contribute to the epidemiology literature that deals with the measurement of the effective reproduction number. In this regard, our methodology relies on simple time-series analysis, contrary to most recent epidemiological papers that rely on more complex Bayesian methods to estimate the real-time dynamics of the effective reproduction number (Wallinga and Teunis, 2004; Bettencourt and Ribeiro, 2008; Obadia et al., 2012; Thompson et al., 2019; Abbott et al., 2020; Kubinec, 2020). These efforts make important contributions in terms of the epidemiological management of the data and real-time estimation of the generation time, which is fundamental for estimating R_t. However, to the best of our knowledge, no effort has been made in the direction of taking into consideration the data-gathering process that is involved in producing incidence data.[1]

3. Methods

We propose a method that, besides estimating the effective reproduction number and the test positivity rate, proposes a classification mechanism based on exogenous thresholds. The method can be thought of as a two-steps procedure. The first step consists on applying typical time series analysis to extract the underlying trend of confirmed positives and the positivity rate. The purpose

[1]Other approaches heavily rely on the SIR framework to produce estimates for R_t. Shim et al. (2020) compute a generalized growth model based on the equations of the SIR model and estimate the effective reproduction rate for South Korea. In a similar vein, Kucharski et al. (2020) use a stochastic transmission dynamic model to estimate early dynamics of the transmission process in Wuhan. They find an effective reproduction number that moves between 2.35 and 1.05.

of this step is to produce smooth infection and positivity rate curves across time to minimize the variation in the data that is orthogonal to the infection process.

In the second step, we use the estimated trends for the reproduction number and the test positivity rate to classify the trajectory followed by a population on the joint space defined by these two indicators. The goal is to use exogenous thresholds defined over the two indicators and use the joint analysis to minimize type I and II errors in assessing the current situation of the transmission process.

Once our method is in place, we apply it to a controlled environment. This setting consists on an agent-based-model with an embedded SIERD infection framework and an exogenously-determined testing layer (the data-gathering process). We use this controlled environment to study the limitations of the econometric analysis and our classification mechanism.

In this section we first describe in detail the econometric procedure to compute the effective reproduction number and the test positivity rate, and show how the joint trajectory can be used as a classification mechanism. Then we describe the agent-based environment and the main characteristics of the simulation.

3.1. Estimating Trends

The idea is simple: We are dealing with a dynamic process that, due to data gathering and the data generating process itself, is subject to several sources of noise. Before going into the specifics of the problem, consider any finite time series y_t. In economics usually y_t has three components: A trend τ_t, a cycle ε_t (or stochastic deviations from the trend), and a seasonal component η_t (which is stochastic and periodic). However, since typical Covid-19 data is reported daily, we need to account for two additional problems: There might be large atypical values ν_t, and there might be days when there are no official reports, so the series have missing values. Thus, y_t can be written as

$$y_t = \mathrm{N}_t(\tau_t + \varepsilon_t + \eta_t + \nu_t), \tag{1}$$

where N_t is a missing value operator that takes a value of 1 if the data point exists, and it is the empty set $\varnothing$ if data is not available. Notice that we have assumed that the components are additive (this can be achieved by working with logarithms). The purpose of the method is to filter away the atypical and seasonal components, clear all missing values, estimate trends to extract the underlying dynamic process and use bootstrapping on these trends to build confidence intervals.

To accomplish this, the first step is to eliminate missing values, and we do so by linear interpolation. This process can be applied as long as there are no long black-outs because, if this is the case, linear interpolation might cause severe bias in the behavior of the underlying dynamics of y_t. Thus, suppose there is a small, finite time period $n \in \mathrm{N}_t$ in an otherwise long time series. Also, let $L(\cdot)$ be a linear function that takes as input two consecutive data points to build a linear function. Then, we can write

$$\tilde{y}_n = L(y_n | n \in \mathrm{N}_t, y_{n-1}, y_{n+1}), \tag{2}$$

where y_{n-1} and y_{n+1} are two consecutive non-missing data points in y_t. If y_n occurs at any of the extremes of y_t (that is at y_0 or at y_T), then we use linear extrapolation based on the linear function defined by y_1 and y_2 or y_{T-2} and y_{T-1} respectively. For the sake of precision, we only extrapolate one period, if necessary. With this operation we can define

$$\hat{y}_t = \mathrm{N}_t^0 \, y_t + (1 - \mathrm{N}_t^0)\tilde{y}_n = \hat{\tau}_t + \hat{\varepsilon}_t + \hat{\eta}_t + \hat{\nu}_t, \tag{3}$$

where N_t^0 is the same missing operator with zeroes instead of the empty set and $\hat{x}_t$ is x_t after linear interpolation.

The next item we tackle is the seasonal component. For this, we apply an $S(m, n)$ seasonal filter. Then, we sort the cycle and the atypical values and eliminate the top and bottom quintiles. This step allows us to obtain ε_t^*, which corresponds to a seasonally adjusted-atypical free error term.

Even though we already have the adjusted cycle component, we still need to compute the trend, which is the hart of our procedure. To do this, we compute the seasonally adjusted-atypical corrected series as

$$y_t^* = \hat{\tau}_t + \varepsilon_t^*, \tag{4}$$

where $\hat{\tau}_t$ is still the trend component that was calculated when the time series was not adjusted nor corrected. To solve this issue, we filter y_t^* again. We use ε_t^* to compute endogenous smoothing parameters. In particular, we set $\lambda^* = \zeta\sigma_\varepsilon^2$, where σ_ε^2 is the variance of the adjusted and corrected error term, and $\zeta \geq 1$ is a scaling parameter. Notice that the smoothing parameter implies a more rigid filter as σ_ε^2 increases.[2] In this case, by applying the HP Filter with smoothing parameter λ^* we obtain τ_t^*. Since both τ_t^* and ε_t^* are random variables, we follow Gallego and Johnson (2005) to build confidence intervals around both components using block bootstrapping.

The use of the HP-filter deserve a few words. By design, the filter is built having in mind business cycles and macroeconomic time series (Hodrick and Prescott, 1997). One particular process that the authors have in mind is GDP deviations from long-term growth, so they interpret the trend as the long term component, and the deviations from trend as short-term business cycles. Thus, the authors recommend different values for λ depending on the periodicity of the data (100 for yearly data, 1600 for quarterly data and 14400 for monthly data). That is, lambda should increase substantially with the periodicity of data, so for high frequency data (as the one for Covid), λ should approach infinity.

However, the problem with this approach is that as $\lambda \to \infty$, the HP filter converges to a linear trend, and this is something that we want to avoid at all costs, specially for countries with stable data. The reason for this is that, despite its high frequency, the series of daily positives and test positivity rate are significantly non-linear processes (it follows the infection process, which exhibits Gaussian-like behavior). Another alternative would be to use ad-hoc values for the smoothing

[2] This approach clearly differs from the original definition for the smoothing parameter, where $\lambda = \sigma_\tau^2/\sigma_\varepsilon^2$.

parameter, as is done by Fernández-Villaverde and Jones (2020). However, this might be problematic, since the quality of surveillance data is very heterogeneous.

Now we apply this procedure to Covid-19 data. Let p_t and s_t denote the series of daily reported positives and tests performed, both measured in logarithms. We apply linear interpolation, seasonally adjustment and atypical correction to both series. Thus, after this step we obtain

$$\begin{aligned} p_t^* &= \hat{\tau}_{pt} + \varepsilon_{pt}^*, \\ s_t^* &= \hat{\tau}_{st} + \varepsilon_{st}^*. \end{aligned}$$

With Covid-19 data, p_t^* and s_t^* are highly correlated, and p_t^* is downward biased because it is unfeasible to test the entire population. Moreover, $\hat{\tau}_{pt}$ is not a good estimate of the dynamics of the infection process, since it was computed with the data before any correction was made; something similar happens with $\hat{\tau}_{st}$. To tackle the first issue, we compute the test positivity rate. We do this because this indicator is more stable than p_t^* or s_t^* on their own. Remembering that both series are in logs, we can write the test positivity rate ρ_t, as

$$\begin{aligned} \rho_t^* &= p_t^* - s_t^*, \\ &= \hat{\tau}_{pt} - \hat{\tau}_{st} + \varepsilon_{pt}^* - \varepsilon_{st}^*, \\ &= \hat{\tau}_{\rho t} + \varepsilon_{\rho t}^*, \end{aligned}$$

where $\hat{\tau}_{\rho t}$ is the trend of the test positivity rate, and $\varepsilon_{\rho t}^*$ is the error term. We use the latter to compute the endogenous smoothing parameter to estimate $\tau_{\rho t}^*$. In the same way, and to tackle the second issue regarding the slope of the testing procedure, we use ε_{st}^* to compute the endogenous smoothing parameter to estimate τ_{st}^*.

However, we are not done yet. Given that ρ_t^* is computed with data based on testing, we need to be aware that testing cannot be performed over the entire population and that some countries experience severe delays in processing samples, so ρ_t^* is a biased estimator of the true infection probability. Following Manski and Molinari (2020), we can reduce this bias by computing the appropriate adjustment for each period. In particular, we could obtain

$$\rho_{at}^* = \omega_t \rho_t^* = \omega_t \tau_{\rho t}^* + \omega_t \varepsilon_{\rho t}^*. \tag{5}$$

Then, we use ρ_t^{a*} to compute bias-corrected daily positives as

$$\begin{aligned} p_{at}^* &= \rho_t^{a*} + s_t^*, \\ &= \omega_t(\tau_{\rho t}^* + \varepsilon_{\rho t}^*) + s_t^*, \\ &= \omega_t \tau_{\rho t}^* + \tau_{st}^* + \omega_t \varepsilon_{pt}^* + (1-\omega_t)\varepsilon_{st}^*, \end{aligned} \tag{6}$$

In a first approach, we set $\omega_t = 1$ for all t, so no adjustment is made. Of course, all our results are still biased, but at least we can use the limited data that we have at hand to present the method. Moreover, in spite of the limitations, we believe that this exercise provides valuable

lessons regarding the limitations that many countries have in terms of data management. With richer data, we could make further adjustments to correct these biases, similar to what is done by Abbott et al. (2020).

We use p^*_{at} to compute the effective reproduction number R_t applying the Lotka-Euler equation (Dublin and Lotka, 1925; Feller, 2015; Metz and Diekmann, 2014; Keyfitz and Caswell, 2005)[3]. For this, we consider only the trend component of p^{a*}_t which is given by

$$\tau^*_{pat} = \omega_t \tau^*_{\rho t} + \tau^*_{st},$$

and compute the exponential growth rate of positive cases γ_t by means of a linear regression which can be written as

$$\tau^*_{pat} = \beta_0 + \gamma_t t + \epsilon_t,$$

for all $t \in T_w = [\underline{t}, \overline{t}]$, where T_w is a sliding time window of n_w periods. Form here, we compute R_t as

$$R_t = \exp\{\theta \gamma_t\}, \tag{7}$$

where θ denote the time between successive cases of infection (generation time) estimated by (Abbott et al., 2020). Again, we apply block bootstrapping and assume that θ is log-normally distributed to estimate confidence intervals for R_t. As before, if $\omega_t = 1$ then the level of R_t is downward biased when $R_t \geq 1$ and upward biased when $R_t < 1$. However, what is important to check is how likely it is for R_t to be below 1.

3.2. Trajectories Over the (R_t, ρ_t) Space

As the pandemic unfolds, the data-gathering process produces volatile information that is subject to selection bias (testing patients with symptoms with higher probability) and measurement error (systematic bias in the measure of Rt). Because of this, it is important to look at a combination of indicators to assess the current situation in specific territories. In this section we argue that looking at the effective reproduction number is not enough to assess whether or not the infection process is contained.

The reason for this is simple. In spite of all data corrections that can potentially be made (ours and delay corrections such as in Abbott et al., 2020), the value and dynamics of the effective reproduction number depend on testing and the test positivity rate. Thus, there could be cases in which $R_t < 1$ because there is an premature reduction in the testing scale (daily tests fall on average) combined with constant (or just marginal reductions) in the test positivity rate. If this

[3] We opt for using this equation. However, since the time period that we are considering involves days, weeks and months, we assume total population remains constant, so the results are approximately equal to what would be obtained using the equation proposed by Anderson et al. (1992); Pybus et al. (2001); Ferguson et al. (2005); Wallinga and Lipsitch (2007) where $R_t = 1 + \gamma_t \tau$.

occurs, then $R_t < 1$ does not imply that transmission is under control, but that the testing scale is not appropriate given the size of the outbreak.

To rule out this wrong assessment, we propose to look at the joint trajectory of the effective reproduction number and the test positivity rate, and argue that there is evidence of a contained infection process only when both indicators are sufficiently low. In fact, while R_t measures how fast the pathogen is spreading, the test positivity rate provides a measure of the level of infection at a given point in time.

[Figure 1 about here.]

For the design of this assessment scheme we consider two thresholds. For the effective reproduction number we take $R_t = 1$. The reason is simple: When $R_t < 1$ the exponential growth rate of daily new infections is negative, so daily counts of new cases are falling over time. For the test positivity rate, we set the threshold to $\rho_t = 0.10$, or a test positivity rate of 10%, which is a goal set by the World Health Organization (public statement by Dr. Michael Ryan, Executive Director of the WHO, cited by Bult, 2020). The reason for this is that, as we mentioned before, the test positivity rate is highly correlated to the true infection process, so as the disease spreads over susceptible population at positive exponential growth rates, test positivity rates are expected to be high. As the infection process is contained with effective interventions, and if testing is done correctly (i.e. the scale of the testing scheme is enough to cover the size of the outbreak), the test positivity rate should be low.[4]

As shown in Figure 1, the classification scheme defines four quadrants. Quandrant I corresponds to a process that is active with a testing scale that is not enough for the scale of the infection process. Quadrant II corresponds to process that might seem to be under control (R_t is lower than one), but the testing scale is insufficient. This implies that what we observe in the dynamics of the effective reproduction number might be just the consequence of a poorly designed testing strategy.

Quadrant II is the only section of this space where we can be sure that, at least temporarily, the infection process is under control. In this case one have to be cautious with respect to trajectories that move towards quadrant IV. There might be two reasons for this type of displacement. The first reason is that, by construction, R_t tends to 1 as the exponential growth rate of confirmed cases converges to zero. Thus, if R_t is converging to 1 from below but the test positivity rate is close to zero, then the situation is under control and there is nothing to worry about. The second reason would take the form of a trajectory that moves to the north-east; that is, we have a simultaneous increase in the effective reproduction number and the test positivity rate. In this case there are reasons to believe that the population is facing a new wave of infections.

Finally, in quadrant III we have two types of populations. There are those that are at the beginning of the infection process, where positivity rates are low and infection reproductive numbers are high; and those that are facing new waves of infection as explained in the previous paragraph.

[4] In fact, according to the Johns Hopkins University Coronavirus Resource Center, the WHO advice for governments is to make sure that testing positivity remains below 5% during, at least, 14 weeks before ending confinement measures (https://coronavirus.jhu.edu/testing/testing-positivity).

3.3. A Controlled Environment

To assess the dynamics of an unfolding epidemic we need to understand how the noisy data collected from testing relates to the true infection curve under different testing strategies. To do this we construct an Agent Based Model (ABM) that explicitly simulates both the infection process and noisy data collected under different testing schemes applied to our simulated population. Daily testing data and the true epidemiological curves from our ABM allows us to assess the performance of our model for estimating the dynamics of an unfolding epidemic. In addition, it allows us to imf investigate how the parameters that govern testing affect the quality of data gathered from tests, and the limitations this data may have in providing information about the true dynamics of an unfolding epidemic for any model that attempts to estimate its dynamics.

While we use a SEIRD setup for our simulation, we are agnostic as to whether compartmental differential equation models give an accurate representation of the spread of an epidemic when compared to models that incorporate behavioral responses (see Eksin et al., 2019; Fernández-Villaverde and Jones, 2020, for example). Unlike other equation and agent based models of contagion, the purpose of our model is not to represent contagion in a real population in order to predict the true trajectory of an epidemic. The purpose of our model is to generate noisy data that plausibly arise from real testing strategies in order to compare the data generated by those tests with the true infection curve given an underlying epidemiological process.

3.3.1. Simulation Setup

Our simulation is implemented in the NetLogo platform (Wilensky, 1999), and can be divided into two layers. The first layer simulates the epidemiological process, and the second layer simulates testing. The epidemiological layer is a straightforward agent-based implementation of the equation based SEIRD model in Meidan et al. (2020) and uses similar parameters values. In our simulation, an initial population of exposed individuals moves about randomly among a population of susceptible individuals. The epidemic spreads through a population of N agents as E_0 exposed individuals make contact with those who are susceptible $S_0 = N - E_0$. Once agents become exposed they undergo a progression of disease that evolves stochastically. We simulate the progression of disease for symptomatic as well as asymptomatic individuals as this will be necessary when we apply testing strategies to a sample of the population. Figure 2 presents all the compartments used in the simulation and the progression of the disease as it unfolds for each individual.

In the epidemiological layer agents walk randomly on the surface of a toriod with 10,000 patches. Agents become exposed with probability, $P_{exposed}$, when a susceptible and an infectious agent share the same patch. The probability of contagion $P_{exposed}$ is calibrated to match the β parameter of the SEIRD model of Meidan et al. (2020). Recall that in a standard SIR model $\beta = \bar{\rho} * P_{exposed}$, where $\bar{\rho}$ is the average population density. Once an agent becomes exposed the agent transitions between stages of disease following the probabilities in figure 2 and in table 1. We model the onset of symptoms and the onset of infectiousness separately with their respective incubation and latency periods. In order to focus our attention on the effect of testing strategies on

the quality of data, we run the epidemiological layer with the same parameters and same random number generator seed to keep the true underlying epidemiological process identical between runs.

[Table 1 about here.]

[Figure 2 about here.]

The focus of our simulation is not the epidemiological layer but testing. Each time period a number of tests are performed on a sample of the population. Three parameters, the scale of testing $scale$, the growth rate of testing g_{tests}, and a shock to the growth rate σ_u determine the number of tests. One parameter, P_{tests}, determines how the sample that undergoes testing is selected. These four parameters define the space that we explore to see the effect of different testing strategies in the simulation and how this relates to the performance of our model. The number of tests in each period of the simulation follows a function with exponential growth trend with a normal i.i.d. stochastic shock with mean $\mu_u = 0$, and standard deviation = σ_u. Formally,

$$tests_t = scale * N \exp\{g_{tests} t + u_t\}. \tag{8}$$

The way the sample of the population that is tested is selected is very important because testing strategies can bias what we can learn about the true epidemiological curves from testing data. While others have looked at the important social benefits of testing as an integral part of mitigation strategies (Eichenbaum et al., 2020b, for example), there is relatively little work on the differences between testing strategies when there are binding constraints on the number of tests that can be deployed each day. Testing of symptomatic infections for diagnostic purposes is very important, but the data gathered from diagnostic tests is probably biased due to the higher probability of a positive result if one presents symptoms. Many countries, especially those with limited testing capacity, have prioritized testing of symptomatic individuals. On the other hand, random samples of the population may give unbiased data, but certainly sacrifice resources that could be used for diagnostic purposes. Thus countries with limited testing scale and constrained growth rate at which this scale can be expanded face this trade-off when it comes to selecting a testing strategy.

In our simulation the sample of the population that is selected for testing is as follows. Each period the pool of individuals that can be tested is composed of those individuals that have not been previously tested or, if they had, the have not had a positive result. The parameter P_{tests} determines the proportion of tests that are performed on individuals that are infected and show symptoms. We explore two testing strategies, random testing ($P_{tests} = 0$) and testing where symptomatic individuals have higher priority to be tested. Thus for $P_{tests} > 0$, at each time period t we test $tests_t^{symptoms} = \min\{P_{tests} * tests_t, I_t^{symptomsc}\}$ symptomatic individuals, where $I_t^{symptomsc}$ is the total number of symptomatic individuals in the population. The remainder of tests given by $tests_t - tests_t^{symptoms}$, corresponds to a random sample from the pool of individuals that can be tested.

We do not account for background rates of symptoms that are present in SARS-CoV 2 and other diseases. A more realistic model should include this aspect. Furthermore we assume testing

is perfect such that there are zero false positives and false negatives. We take this approach to focus on the effect of testing strategies and not on the accuracy of the tests being used. We also assume sampling and testing is carried out and processed instantaneously to focus on the effects of testing strategy and abstract away from processing capacity, although the testing and test processing delay distribution probably varies widely by country.

To assess the performance of our model we construct measures of its performance for our simulated data. First we find the true effective reproduction number for our simulation,

$$R_t^{sim} = R_0^{sim}\frac{S_t}{S_0} = \beta^{sim}\theta^{sim}\frac{S_t}{S_0} \tag{9}$$

to use as our benchmark (Nishiura and Chowell, 2009)[5].

The parameter $\beta^{sim} = 1.25$ is the simulation parameter that controls how fast the infection spreads in its initial phase, and $\theta^{sim} = 5.05$ is the serial interval for the epidemiological parameters of the model. Then we calculate a *naive* effective reproduction number that is based solely on observed daily positive tests R_t^{naive}, and the effective reproduction number, R_t^{model} obtained applying the estimation model from section 3. Both R_t^{naive} and R_t^{model} are found using equation (7).

We first define two performance measures that compare R_t^{naive} and R_t^{model} to the true effective reproduction number from our simulation, R_t^{sim}. Our first measure η_j captures the probability of type II error when using R_t^j to asses when the true reproductive number is smaller than one. In the language of hypothesis tests let $H_0 : R_t^{sim} \geq 1$. Then we compute the percentage of time periods in the simulation when $R_t^j < 1$ and $R_t^{sim} \geq 1$. Formally,

$$\eta_j = \mathrm{P}[R_t^j < 1 \wedge R_t^{sim} \geq 1], \tag{10}$$

for $j = \{naive, model\}$, where the probability is found from the observed values for each simulation run. This is our most important performance measure because in dealing with noisy data, we would like our model to do as little harm as possible and thus we seek to have η_2 be as close to 0 as possible. This performance measure allows us to assess how well alternative estimations of R_t^j capture the dynamic behavior of the infection process.

Our second performance measure is the root mean squared error between R_t^j and R_t^{sim} for each model run, which can be written as

$$RMSE_j = \sqrt{\frac{1}{T}\sum_{t=0}^{T}(R_t^j - R_t^{sim})^2}. \tag{11}$$

As noted before we don't expect either the *naive* or the *model* reproductive numbers to equal the true value due to testing bias. However the $RMSE_j$ allows us to see how the *naive* and *model*

[5] We find the true effective reproduction number for the simulation using its definition given in Nishiura and Chowell (2009) because have time invariant epidemiological parameters in our simulation. We could also estimate the true effective reproduction number using the infection curve and the Euler-Lotcka equation as in (7), but this change plays in our favor

reproductive values deviate from the true reproductive number in the simulation as a function of the testing parameters. With this measure we assess both the dynamic behavior and how well the estimation alternatives fit the true value. Thus $RMSE_j$ is the average measurement error of the estimation of R_t^j based on testing data.

4. Results and Discussion

We begin this section by applying our econometric procedure to the controlled environment developed in the previous section. The goal is to understand how the testing process affects the signal of the true infection curve that we observe, and how our methodology allows us to improve such observation. Once this is well understood, we apply the method to country data and evaluate these results through the lens of the theoretical assessment obtained in the agent-based framework.

4.1. Simulation Results

We run the agent-based environment for $T = 60$ days, $N = 100,000$ agents, and an initial exposed population of $E_0 = 50$ agents. We run each of the 40 parameter combinations 100 times giving us a total of 4,000 observations. The parameter space is given by all the combinations of parameters in table 2. We do not vary the growth rate in testing to focus on situations where the testing rate cannot be increased by a large amount due to binding capacity constraints. As a benchmark, we compare our estimation of the effective reproduction number with an *naive* estimation where no time-series corrections are performed (i.e. we take the testing data *as it is*).

[Table 2 about here.]

To obtain an overall picture of our simulations we first graph the evolution of R_t^j and R_t^{sim} over the course of the simulation. Figures 3a and 3b show the evolution of the effective reproduction number with respect to time and with respect to the true reproduction number. As is immediately apparent, both model and naive reproductive numbers are systematically biased, underestimating the true reproductive number before $R_t^{sim} = 1$, and overestimating after this point. The bias can be seen to be larger for smaller testing scales ($scale$) and for testing strategies that prioritize testing symptomatic individuals (i.e. larger values of P_{tests}), while the size of the disturbance to the testing error trend (σ_u) seems to play a very small, if any role.

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

To quantify the effects of testing scale and testing strategy on the systematic bias in the effective reproduction number we find the conditional mean and standard error of $RMSE_j$ and η_j[6]. The conditional probability distribution for $RMSE_j$ and η_j can be respectively seen in

[6]This is equivalent to regressing $RMSE_j$ and β_j with respect to all the the parameters and their interactions.

figure 4 and figure 6. The conditional means for each parameter combination can be seen in figure 5 and figure 7.

Both performance measures exhibit very similar behavior with respect to the testing parameters. As can be seen in figure 5 for the lowest testing scale, the *naive* model outperforms our model, but this gap falls sharply as the testing strategy switches from random testing to prioritize symptomatic patients. As expected, increasing testing scale reduces the $RMSE$ for both the *naive* and the *model*. Under random testing the fall in $RMSE_j$ due to increasing testing scale is greater for the naive calculation of $R^j(t)$ than for the model. However as soon as testing strategies become non-random ($P_{tests} > 0$) the fall in $RMSE_j$ from increased testing scale is much greater for the *model* than for the *naive* effective reproduction number. This is significant because most countries with binding testing constraints probably prioritize symptomatic patients. While this is good in terms of diagnostics, there is a trade-off involved in the precision of the effective reproductive number.

[Figure 6 about here.]

[Figure 7 about here.]

Our analysis of the effect of testing parameters on the probability of type II error follows a similar pattern to the RMSE. Figure 7 shows the conditional mean of the probability of type II error. As can be seen under random testing ($P_{tests} = 0$) we have the lowest rate of type II error. Under non-random testing, increasing the testing scale reduces the probability of type II error more for the *model* than for the *naive* estimation of $R^j(t)$. For example the average rate of type II error at the largest testing scale ($scale = 0.01$, one percent of the testable population is tested each day) is 3% for the model while it is more than 6 times larger, 18.9%, for the naive calculation of $R^j(t)$. This is significant if the effective reproduction number is to be used as an indication that the epidemiological process is entering its waning phase.

While these results show that both the *naive*, and *model* estimations of the effective reproduction number are biased, the size and direction of this bias is sensible to the parameters that govern the data gathering process. As we have taken the epidemiological process as given, this bias is independent of the data generating process (i.e. the epidemiological process) and due solely to the data gathering process (i.e. testing).

Next we apply the trajectory classification scheme to our simulated data to study how this methodology allows us to reduce the probability of type I and type II errors that can be triggered by the data gathering process. We present the density plot of trajectories in Figure 8. Lighter colors represent a higher concentration of points in a given area. For reference, we also present the true trajectory followed by the simulated infection process. For this, we use the R_t derived directly from the simulated process, and we compute the true positivity rate as the ratio of infected individuals with respect to total population at a given point in time. Moreover, black dotted lines delimit the four quadrants that are defined by the combination of thresholds.

[Figure 8 about here.]

This plot shows that, even in the presence of significant variation, it seems like trajectories based on testing data follow the true trajectory and coincide when they pass through the four possible quadrants. This result is encouraging because it implies that, even in the presence measurement error and selection bias in the estimates of R_t and ρ_t, looking a its joint distribution could be a reasonable approach to minimize the probability of wrongly confirming that the transmission process is under control.

To formalize this evidence, we define three errors that occur due to measurement error and selection bias. The first error term aims to analyzing deviations in R_t. Differently to the previous section were we only considered type II errors, in this section we count situations were R_t is below (above) 1 while R_t^{sim} is above (below) 1. Formally, we have

$$\delta_R = \mathbf{1}[(R_t < 1 \wedge R_t^{sim} \geq 1) \vee (R_t \geq 1 \wedge R_t^{sim} < 1)],$$

where $\mathbf{1}$ is an indicator function that takes the value of 1 if the condition in square brackets holds and 0 otherwise, so δ_R is a dummy variable. Following the same idea, we count deviations from the true positivity rate considering the 10% threshold, so we can write

$$\delta_\rho = \mathbf{1}[(\rho_t < 1 \wedge \rho_t^{sim} \geq 1) \vee (\rho_t \geq 1 \wedge \rho_t^{sim} < 1)].$$

Finally, we define a fail when $\delta_R = \delta_\rho = 1$ which corresponds to situations in which both estimates (R_t and ρ_t) are deviating from their true values at the same time. Formally, we have

$$\delta_{(R,\rho)} = \mathbf{1}[\delta_R = 1 \wedge \delta_\rho = 1].$$

We compute these three dummy variables for $t \in (5, 60)$ for the 4,000 runs that we simulated using our ABM, giving us a total of 215,933 observations. To see how the joint analysis of R_t and ρ_t reduces the probability of type I and type II errors, we compute the probability of error in R_t, ρ_t and (R_t, ρ_t). We show the results of these calculations in Table 3 and present standard deviations in parenthesis.

[Table 3 about here.]

In all three cases, the percentage of classification errors is statistically different from zero, and much higher for the test positivity rate (23.6%) than for the effective reproduction number (6.06%). However, when we look at the joint analysis, classification errors fall drastically to 0.34%.

How do testing affect these errors? To answer this question we take advantage of the exogeneity provided by our simulations, and estimate the following regression,

$$\delta_{itj} = \beta_j + \boldsymbol{\gamma}_j \mathbf{D}_i^{sc} + \boldsymbol{\alpha}_j \mathbf{D}_i^{symp} + \boldsymbol{\lambda}_j (\mathbf{D}_i^{sc})' \mathbf{D}_i^{symp} + \kappa_j D_i^{\epsilon} + \mu_{itj}, \tag{12}$$

where i corresponds to each simulation, t denotes the time period, $j = \{R, \rho, (R, \rho)\}$, β_j is the constant term, $\mathbf{D}_i^{sc}$ are dummy variables for the four possible values that testing scale can take

in the simulations, $\mathbf{D}_i^{symp}$ are dummies for the probability testing symptomatic, $\mathbf{D}_i^{\epsilon}$ is a dummy variable that takes the value of one when the scale of the innovation in the testing process is high and $(\mathbf{D}_i^{sc})'\mathbf{D}_i^{symp}$ is a set of interactions which is included as long as the model can be estimated.

In principle, we estimate equation (12) by OLS with standard errors clustered at the simulation level. Notice that, since parameters do not vary within simulations, we are automatically accounting for fixed effects in our regression. Moreover, since we have complete control over the parameters in each simulation, the correlation between the error term and the regressors is zero, so we are in fact estimating causal effects. The only limitation with our strategy is that we are assuming linearity. To see how this assumption might affect our results, we estimate (12) assuming cumulative logistic (logit) and normal (probit) distributions, and estimate the parameters by maximum likelihood. We present the marginal effects of these estimations in Table 4.

[Table 4 about here.]

Our results are sensible to the linearity assumption, particularly for classification errors in R_t and ρ_t although we do not find significant difference when we look at the joint analysis. Nonetheless, we choose non-linear estimations over the linear model and report the results of all the estimations. Moreover, due to the small number of observations in which the joint analysis produces classification errors, we are not able to include interaction terms. In every case we estimate our models with 215,933 observations and 4,000 clusters.

The effects of the scale of testing over wrong classifications regarding the effective reproduction number are monotonic, while the effects of the probability of testing symptomatic seem to have a non-linearity when testing is closer to being random. In particular, increasing the scale of testing reduces the probability of classification error according to the R_t threshold; while testing some proportion of symptomatic patients reduces the probability of this classification error in at least 2.6% with respect to totally random testing. Although this last result might seem counter intuitive, remember that the measurement in R_t originates in attenuation bias that we generate when we estimate the exponential growth rate using confirmed cases. As we test more symptomatic, this growth rate increases (other things equal), alleviating the attenuation bias triggered by limited testing.

The effects of the testing strategy over classification errors regarding the postivity rate are completely monotonic. In this case, increasing the testing scale reduces this probability in up to 13%, while increasing the probability of testing symptomatics increases the probability of type I and type II error in this classification by as much as 22%.

Finally, when we look at classification errors using the joint analysis it is interesting to see that, although all coefficients are significant at, at least, 95% confidence level, the magnitudes are really low. This result is important, because it provides us with evidence that although there might be systematic bias in R_t and ρ_t separately, when analyzed jointly the effects of different testing schemes over the probability of wrong classification are negligible.

The results in this section are relevant for the analysis of the current state of the transmission process in environments with limited information and poor testing strategies. What we have shown

is that the joint analysis of the test positivity rate and the effective reproduction number provides a safer assessment.

4.2. Application to Country Data

Once we have explored our methodology's limitations in controlled environments, in this section we apply it to Covid-19 data to determine the current status of the infection process for 35 different countries.[7] From this sample, we choose 15 countries for which we show explicitly how each step of our model is applied.

The 15 countries that we consider include 6 developed economies and 9 emerging markets: United States, Canada, Italy, Switzerland, United Kingdom, Mexico, Chile, Colombia, New Zealand, Japan, India, Bolivia, Ecuador, Uruguay and Israel. The quality of data in this group is very heterogeneous and ranges from high quality data based on massive testing, proper data registries and short sample processing delays (as is the case for United States or Israel), to extremely bad quality that reflects poor data gathering practices, long delays in sample processing and very limited testing, usually clustered among people presenting symptoms (as is the case of Ecuador or Mexico).

Another feature that makes these event studies interesting is that there is heterogeneity regarding the stage of the infection process at which each country is at the last date that we updated the model. We also look at the cases of Uruguay or New Zealand, countries that were very successful at tackling the pandemic at very early stages, so they never experienced exponential growth in the number of confirmed positives.

We use data available from the Our World in Data Covid-19 Dataset (OWD) Max Roser and Hasell (2020), except for Ecuador. For the latter we build the time series by hand using the information that Ecuador's Health Ministry (MSP) publishes daily in the presidency's Risk Management Commission web page. The latest update for each country included in this paper was August 13th.

[Figure 9 about here.]

In general, data that is widely available for analysis is subject to several reporting issues that trigger variations that are not necessarily related to the infection process. In Figure 9 we show infection trajectories according to reported data for the 35 countries included in our sample. Horizontal axis measure total confirmed cases until the last update, while vertical axis measure daily counts of confirmed cases. Both axis are in logarithmic scales, and countries are classified according to the quadrant they occupy in the (R_t, ρ_t) space considering its most updated observations.

This type of trajectories are quite informative. First, it is possible to identify countries that have left exponential growth trends, as the group of countries we observe in Figure 9a, countries

[7] In an earlier version of the paper we were working with a sample of 56 countries. However, in many cases the test registry have changed, causing irreparable breaks in the series and comparability issues with the number of confirmed positives. The main issue is that, while the number of tests include only molecular tests (PCR), the total number of confirmed positives include both molecular and antigen tests. In some countries the number of antigen tests is very small, but in others it can produce severe bias when calculating the positivity rate.

that continue to expand on exponential trends like in Figure 9c, or countries with serious data issues, like what we observe in Figure 9d for the case of Mexico and Bolivia. In every case, missing values occur either because data was not available for given dates, there are dates where daily positives were zero, or dates where the evolution of confirmed cases imply negative values[8] Something similar occurs with daily counts on tests.

With this in mind, in the sections to follow we illustrate how we apply our method to the data we have at hand and being aware that we are not making necessary adjustments that are complementary to our method. This process is divided in four stages: First we extract the seasonal and atypical components, and then use the adjusted series to compute the smoothing parameter of the HP filter as a function of the standard deviation of the cycle component. Second, we rebuild the series of daily positives using the estimated trends and cycles for daily positives and tests. Third, we estimate R_t and ρ_t using smoothed series. Finally, we classify each countries' trajectories in the (R_t, ρ_t) space.

4.2.1. Extracting the Seasonal and Atypical Components

To be able to work with additive components we transform everything to logarithms. Since there are days when neither positives are detected nor tests are applied, we use linear interpolation to fill these gaps. Then we detrend all series using the HP filter setting $\lambda = 1600$ for positives (we change this later) and a linear trend for tests (testing monotonicity assumption, see Manski and Molinari, 2020, for details).

The next step consists on cleaning the error terms from the seasonal component; for this, we apply an $S(3,3)$ filter. Figure 10 shows the estimated seasonal components for daily positives (blue) and daily tests (red) for each country. We measure these in percent deviations from trend since the first day for which we have information on tests being performed.

[Figure 10 about here.]

To our surprise, we find strong seasonal components for all countries in our reduced sample. It is clear that this type of noise can be extremely misleading for the analysis of the infection dynamics, since seasonality generates peaks and valleys that are solely explained by factors that are completely exogenous, like the fact that testing processing and sampling can be limited to hospitals during the weekend.

Once we have the error terms clear of seasonality, we move towards identifying atypical values. These events are characterized by sudden increases (or decreases) that occur very limited times during the time we observe data. In the case of Ecuador, for example, we have such events explained by unexpected increases in test processing capacities, or data revisions performed by the authorities.

To detect these events in a systematic manner, we take the seasonally adjusted error terms for positives and tests, order them, and eliminate the top and bottom 5%. To fill the gaps generated

[8] This occurs in the case of Ecuador due to massive data revisions that have been occurring since May 4th.

by this procedure, we use linear interpolation. Figure 11 shows the identified atypical values for daily positives (blue) and daily tests (red).[9]

[Figure 11 about here.]

In general we find that atypical values regarding testing occur at the beginning of the outbreak or a second wave (as the case of Israel). This result makes sense, since it is expected for countries to experience an adjustment period regarding testing capacities, which can be quite erratic. Moreover, investment on testing infrastructure made during early stages of the infection process payoff at the moment that politicians need to start making decisions about confinement measures. Atypical confirmed positives, on the other hand, are scattered over the entire time window, and respond to periods in which delayed samples are processed in small time frames.

However, there are cases like Ecuador, Canada or Japan where atypical values occur in the second half of the period that we observe. This should be a source of concern, since this occurs 40 or 50 days into the infection process. At this point, usually there is economic, political, and social pressure triggered by confinement measures and its secondary effects that pushes politicians towards hasty decision making. If this process comes in hand with increased levels of noise in the data that are solely explained by events that are completely uncorrelated to the infection process, then wrong decisions can easily trigger new outbreaks.

Once we have the seasonally adjusted error terms without extreme values, we rebuild the series for positives and tests, and use these series for the dynamics estimations that we describe in the next section. We present the results in Figure 12, Panel 12a for daily positives and Panel 12b for daily testing. Blue dots correspond to original data, red lines are the adjusted versions.

[Figure 12 about here.]

The final step in this stage is related to the estimation of the trend of the probability of detecting positives among people tested and testing itself. In both cases we use the HP filter calibrated with endogenous smoothing parameters that are directly correlated with the level of volatility in the error term of each series. The corresponding values of λ^* for the probability of positive results and daily tested are presented in Table 5. Notice that these parameters change every time the model is updated.

[Table 5 about here.]

4.2.2. Estimating the Infection Dynamics

We use the adjusted time series of daily positives and tests performed to compute the probability of a positive result among those tested (all in logarithms). We treat this as a new time series that only has a trend and the cycle component, since both positives and tests are already the adjusted versions. Then, we filter the trend component for each variable using the endogenous smoothing parameters that we found in the previous section.

[9] Keep in mind that, since we are working with logarithms, registries with zeroes or negatives are eliminated at the beginning and replaced with linear interpolations.

With these elements, we reconstruct the trend and cycle components of daily positives from the trend and cycles that we obtained for the probability of positive results and daily testing. Notice that these do not coincide with the trend and cycle that we obtained in the cleaning stage. The reason for this is that we are not using the same smoothing parameters. Then, we use block bootstraping to build confidence intervals for the estimated trend.

[Figure 13 about here.]

We show the estimated trend and confidence intervals for each country in Figure 14. The estimated trend corresponds to the blue line, with 95% confidence intervals in dotted red lines. For comparison, we also include the adjusted daily positives time series (in gray bars). These results show that our method does a good job *sorting* the data.

The usual practice (see Abbott et al., 2020, for example) consists on assuming delay distributions to adjust reported positives according to onset of symptoms and actual date of infection. Given the data that we have at hand, we are not able to perform these adjustments, so our estimates are lagged behind. Nonetheless, a feature that makes our approach appealing is that our infection curves are continuous and smooth (something that is not obtained just by adjusting the temporal framework).

This is important, because the computation of the effective reproduction number R_t (i.e. the rate at which the disease spreads among the population) is based on the exponential growth rate of daily positives. Even if this curve is subject to delay adjustments (the usual practice), it still exhibits sudden variations, and this affects significantly the dynamics of R_t: You can easily have periods of $R_t < 1$ that do not respond the the infection process itself, but to other exogenous factors.

[Figure 14 about here.]

To compute R_t we fit a linear trend to the logarithm of the trend of adjusted daily positives in sliding windows of 10 days. We do this for each of the bootstrapped trend, so we can account for this variation the moment that we build the confidence interval.

But this is not the only source of uncertainty. As we mentioned before, the generation time is also a random variable. In this case, we take Abbott et al. (2020) estimate (mean of 3.6 days and a standard deviation of 3.0) and assume that the generation number is log-normally distributed. We simulate 1000 different samples with these characteristics so, together with the 1000 bootstrapped trends, gives a grand total of 1,000,000 simulated values for R_t at each point in time. We sort these simulations and compute the 95% confidence interval. We present the results of our estimations in Figure 14, where the blue solid line represents the mean, and the red dotted lines each of the 95% confidence bounds.

A valid worry with our methodology is how new data might affect past and current trends. In particular, our model could easily confuse atypical values with a change in the underlying trend, specially in countries where data is very noisy. Its clear that this will have a significant effect in the trend itself, but how much does new data affect the dynamics and the value of the effective reproduction number?

To answer this question, we apply our methodology to data 25 days ago, and then add more recent observations, one at the time. We present the results of this exercise in Figure 15. In red we show the behavior of R_t in the oldest data set, while the blue line corresponds to the newest. Then, gray lines show intermediate cases, from lighter (older) to darker (newer).

[Figure 15 about here.]

This exercise shows how new information is incorporated to the model with each new observation we obtain from the data, and reflects the quality of data in each country. For example, for cases like United States, Italy, Switzerland or United Kingdom, the quality of the data that is produced is good enough, so new observations maintain the trend implied by previous data. Thus, each new observation provides valuable information.

In countries like Canada, Colombia, Mexico, Chile or Israel, data contains more noise. However, in spite of being subject to significant corrections, the adjustment of the curve over time is monotonic. In these countries it is necessary to have more data (i.e. wait for longer) until one can assess the current situation.

This is not the case with Uruguay or Ecuador. In both cases information is so noisy (even after all the cleaning process that it is subject to in our methodology), that each new release has the potential to completely change the main trend. Thus, results with these countries are very volatile, and takes much longer to be sure about what is going on.

This level of variation can produce analysis that have the potential to misinform the decision making process, and can put population under serious risk.[10] The reason for this is that due to data quality, it is possible to have a second outbreak alarm on one day followed by a controlled infection picture the following day. Thus, the estimation is not stable over time.

[Figure 16 about here.]

Because of this, we also analyze the dynamics of the test positivity rate which is more stable than the effective reproduction number and can provide useful complementary information to have a better assessment of the current state of the infection process. In Figure 16 we present the dynamics of the test positivity rate. The blue line represents the estimated trend and the dotted red lines the 95% confidence interval.

The test positivity rate measures the proportion of confirmed positives out of the daily batch of processed tests. The World Health Organization set a target test positivity rate of 0.10 (black lines in Figure 16), believing that at this rate the testing scale was in line with the scale of the infection process. But, even if the test positivity rate is below this threshold, it tends to increase during infection waves, and it tends to zero as the infection process is contained.

It is worth noticing the relation that exists between the test positivity rate and the effective reproduction number. Comparing Figures 14 and 16 we see that there are cases (such as Israel, for example) where there is a clear positive correlation between the effective reproduction number and

[10] It is worth noticing that this is an important limitation of the model we propose, but it also affects significantly any analysis based on Bayesian methods, since the prior distribution might be very volatile from day to day, making Bayesian updating quite challenging.

the test positivity rate. However, in cases such as Ecuador the test positivity rate is significantly above the 0.10 threshold in spite of the effective reproduction number being very close to (or even below) 1. In this cases in particular is when the joint analysis of the two indicators is important, since assessing only the effective reproduction number could easily generate a type II error (i.e. believing the transmission process is under control when, in reality, is not).

4.2.3. Countries Trajectories Over the (R_t, ρ_t) Space

For this analysis we use the filtered trend of the test positivity rate and the effective reproduction number. Then, we compute the joint density of both indicators over a two-dimensional grid with 101 points. For this we use the information for all countries and all time periods. We present the results using two-dimensional contour plots and heat maps instead of using a three-dimensional plot to have a better view of the implied dynamics. Figure 17 show the results.

There is one obvious difference between the data-generating process in our country data and our simulations: Most countries in our sample have applied different measures to contain the spread of the virus; some have been successful, some have been ineffective. But a first conclusion rises from this difference: Countries with ineffective contention measures are more likely to end up with trajectories that look like our simulations. This can be seen in Figure 17, which is similar to Figure 8, but in this case we consider trajectories defined by the 35 countries that we include in our sample.

[Figure 17 about here.]

The analysis with the joint density of the effective reproduction number and the test positivity rate provides information about the current state of the countries we have in the sample, as well as showing the most common paths followed during the pandemic. However, for this analysis to be complete we need to look at the joint trajectories that countries have followed in the (R_t, ρ_t) space. Figure 18 show the trajectories followed by each country according to the Cartesian order. Countries are classified according to their most recent state, which is marked with countries' labels and hollow circles. Full circles mark the starting point for each country. As in the heat maps that we presented before, we mark thresholds for the effective reproduction number and the test positivity rate are in dotted red lines for reference.

Trajectories carry a lot of information. Panel 18c show the trajectories followed by countries tha currently are in a low R_t, low ρ_t state. Within this group of countries, we can distinguish two sub groups. There are countries, like Italy, that where in the bottom-right corner of the plot at the beginning of the pandemic, and then passed through the top quadrants before controlling the infection process. Part of the reason why these countries registered high positivity rates could be related either to wrong testing strategies or a strong infection process which affected most of their population within a very small time frame.

[Figure 18 about here.]

For countries that currently are located in the bottom-right quadrant we have, again, three types of trajectories. First, there are countries like India or Polonia that are moving to the bottom-left quadrant; that is, the effective reproduction number is still high, but is falling, and the test positivity rate is under control. A second subgroup are those that started in the bottom-left quadrant (controlled infection process) and are experiencing a second outbreak. An example in this group is South Korea.

Countries in the top-right quadrant are experiencing exponential growth infection processes. The most worrying case is Bolivia, a country that is going through an explosive infection process with an estimated effective reproduction number that is above 1.3 (and growing), and a test positivity rate over 50% (and growing). In similar positions are Peru and Mexico that, although in these cases the effective reproduction number has been falling, it is still significantly over 1 and register very large test positivity rates, especially in the case of Peru.

Finally, we have the top-left quadrant which currently includes only Ecuador, but there are some other countries with trajectories that passed through this quadrant, which is characterized by low R_t but test positivity rates that are still above 10%. One thing that all these countries have in common (Ecuador and countries that went through this quadrant such as the United States, United Kingdom or Italy), is that non-pharmaceutical interventions where either set too late, or they were just not effective. Moreover, in the case of Ecuador, there is evidence of a poorly designed testing strategy that targeted people with severe symptoms and workers in the front line; we can infer this from Ecuador's trajectory, that started with a test positivity rate of about 50% and in 149 days since case 100 have not been able to reduce this rate below 30%.

To close this section, in Table 6 we present the effective reproduction number, the test positivity rate, the case fatality rate (CFR), the mortality rate and the testing scale using the most recent data available for the 35 countries in our sample. Countries are classified into the four quadrants in the (R_t, ρ_t) space. We added an additional category, *Controlled Process*, for those countries that have effective reproduction numbers equal or below 1 and a test positivity rate that is virtually zero. This category is necessary because, due to the formula we are using to compute it, the effective reproduction number converges to 1 as the exponential growth rate converges to 0.

[Table 6 about here.]

Based on the most recent data included in this paper, we find that there are only four countries in our sample with the infection process under control: Taiwan, South Korea, Myanmar and New Zealand. What is interesting to notice is that among these four countries there is no much variation in fatality and mortality rates, but there is a wide range of variation regarding the testing scale (from 23 tests per 10000 inhabitants in Myanmar to 1042 tests in New Zealand). This wide variation is present within all groups in our classification mechanism.

Due to the characteristics of countries that go through the top-left quadrant, it turns out that the case fatality rate and the mortality rate are significantly higher (at least 1.5 times higher) than in other quadrants. In Table 6 we present the case fatality and the mortality rates computed for each quadrant. For this, we consider countries for which their more recent data place them in each Quadrant. We find that in the top-left quadrant (Q4: $Rt < 1$, $\rho_t >= 0.10$) the case fatality and

mortality rates are 7.5% and 2.6 per 10,000 inhabitants, respectively, which are higher than what is found in other quadrants.

5. Conclusion

Management of the Covid-19 pandemic around the world have not been free from political calculations, which in most countries implied changes in the way surveillance data is released to the public, methodological changes improvised in the unfold of the pandemic, and information gaps that can last for several days. This lack of information hinders the ability of agents external to the government (such as academics) to fulfill their role as monitoring agencies. In this paper we have attempted to reduce this information gap by proposing a simple method that allows us to form a clear picture of the infection process when the only data available corresponds to daily counts of confirmed cases and tests.

Our method consisted in subjecting the available data to *old school* time series filtering processes that allows us to abstract the underlying infection process from stochastic innovations that are orthogonal to the transmission process. In particular, before computing any epidemiological indicator, we introduce a data cleaning process in which we eliminate the seasonal and atypical components that affect the time series of daily counts. Then, using the filtered trend of daily positives, which we calculate imputing endogenously calibrated smoothing parameters to the HP filter, we compute the effective reproduction number R_t and the test postivity rate ρ_t.

We applied our method to simulated data before using it on real Covid-19 surveillance data. From this exercise we learned that any epidemiological indicator computed from daily counts of confirmed cases and tests alone will produce systematically biased indicators. However, our data treatment methodology managed to reduce the root mean squared error between the true effective and the estimated reproduction number, and the probability of type II error (that is, having an estimation of the effective reproduction number that is below 1, when the real process implies an R_t that is still above 1). In other words our simulations show that, although there is uncertainty about the true level of the infection process, there is much that we can learn just by studying the dynamics of the process. This information is highly valuable particularly for public policy.

Further, we argued that the joint analysis of these two indicators allowed us to control for the systematic bias present in R_t and to control for cases in which type II errors are more likely. Moreover, studying the trajectories followed by countries over the (R_t, ρ_t) space we were able to find dangerous situations in which $R_t < 1$ in spite of having test postivity rates that are well above 10%. The reason for our worry is that our country analysis showed higher mortality and case fatality rates compared to the other quadrants. Moreover, this analysis also allowed us to pin down trajectories that might be related to outbreaks in countries that already had the situation under control in previous weeks.

Among the limitations of our analysis, the most salient one is related to the method we use to estimate R_t. In particular, we do not make use of Bayesian statistics for this task, which is the most common method in the epidemiological literature for the estimation of this indicator. However, since these methods use the same surveillance data to build prior distributions, we strongly

believe that the data-cleaning process that we propose would improve its performance. In fact, our recommendation is to add our data treatment to the corrections that are already being applied by, for example, Abbott et al. (2020).

References

ABBOTT, S., J. HELLEWELL, J. MUNDAY, J. YOUNG CHUN, R. THOMPSON, N. BOSSE, Y.-W. DESMOND, T. RUSSELL, C. JARVIS, S. FLASCHE, A. KUCHARSKI, R. EGGO, AND S. FUNK (2020): "Temporal variation in transmission during the Covid-19 outbreak," *Centre for Mathematical Modelling of Infectious Diseases, London School of Hygine and Tropial Medicine*.

ANDERSON, R. M., B. ANDERSON, AND R. M. MAY (1992): *Infectious diseases of humans: dynamics and control*, Oxford university press.

ARELLANO, C., Y. BAI, AND G. MIHALACHE (2020): "Deadly Debt Crises: COVID-19 in Emerging Markets," *Mimeo*.

BARRO, R. (2020): "Non-Pharmaceutical Interventions and Mortality in U.S. Cities during the Great Influenza Pandemic, 1918-1919," *NBER Working Paper*.

BETTENCOURT, L. M. AND R. M. RIBEIRO (2008): "Real time bayesian estimation of the epidemic potential of emerging infectious diseases," *PLoS One*, 3.

BULT, L. (2020): "The US Tested the Wrong People for the Coronavirus," *Vox Explanatory Journalism*.

CORREIA, S., L. STEPHAN, AND E. VERNER (2020): "Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu," *SSRN Working Paper*.

DUBLIN, L. I. AND A. J. LOTKA (1925): "On the true rate of natural increase: As exemplified by the population of the United States, 1920," *Journal of the American statistical association*, 20, 305–339.

EICHENBAUM, M., M. TRABANDT, AND S. REBELO (2020a): "The Macroeconomics of Epidemics," *Mimeo*.

——— (2020b): "The Macroeconomics of Testing and Quarantining," *Mimeo*.

EKSIN, C., K. PAARPORN, AND J. S. WEITZ (2019): "Systematic biases in disease forecasting–The role of behavior change," *Epidemics*, 27, 96–105.

FELLER, W. (2015): "On the integral equation of renewal theory," in *Selected Papers I*, Springer, 567–591.

FERGUSON, N. M., D. A. CUMMINGS, S. CAUCHEMEZ, C. FRASER, S. RILEY, A. MEEYAI, S. IAMSIRITHAWORN, AND D. S. BURKE (2005): "Strategies for containing an emerging influenza pandemic in Southeast Asia," *Nature*, 437, 209–214.

FERNÁNDEZ-VILLAVERDE, J. AND C. JONES (2020): "Estimating and Simulating a SIRD Model of COVID-19 for Many Countries, States and Cities," *Mimeo*.

GALLEGO, F. A. AND C. A. JOHNSON (2005): "Building confidence intervals for band-pass and Hodrick–Prescott filters: an application using bootstrapping," *Applied Economics*, 37, 741–749.

GLOVER, A., J. HEATHCOTE, D. KRUEGER, AND V. RÍOS-RULL (2020): "Health versus Wealth: On the Distributional Effects of Controlling a Pandemic," *CEPR DP 14606 Working Paper*.

HODRICK, R. J. AND E. C. PRESCOTT (1997): "Postwar US business cycles: an empirical investigation," *Journal of Money, credit, and Banking*, 1–16.

KEYFITZ, N. AND H. CASWELL (2005): *Applied mathematical demography*, vol. 47, Springer.

KRUEGER, D., H. UHLIG, AND T. XIE (2020): "Macroeconomic Dynamics and Reallocation in an Epidemic," *CEPR DP 14607 Working Paper*.

KUBINEC, R. (2020): "A Retrospective Bayesian Model for Measuring Covariate Effects on Observed COVID-19 Test and Case Counts," *SocArXiv. April*, 1.

KUCHARSKI, A. J., T. W. RUSSELL, C. DIAMOND, Y. LIU, J. EDMUNDS, S. FUNK, R. M. EGGO, F. SUN, M. JIT, J. D. MUNDAY, ET AL. (2020): "Early dynamics of transmission and control of COVID-19: a mathematical modelling study," *The lancet infectious diseases*.

LILLEY, A., M. LILLEY, AND G. RINALDI (2020): "Public Health Interventions and Economic Growth: Revisiting The Spanish Flu Evidence," *SSRN Working Paper*.

MANSKI, C. F. AND F. MOLINARI (2020): "Estimating the covid-19 infection rate: Anatomy of an inference problem," *Journal of Econometrics*.

MAX ROSER, HANNAH RITCHIE, E. O.-O. AND J. HASELL (2020): "Coronavirus Pandemic (COVID-19)," *Our World in Data*, https://ourworldindata.org/coronavirus.

MEIDAN, D., R. COHEN, S. HABER, AND B. BARZEL (2020): "An alternating lock-down strategy for sustainable mitigation of COVID-19," *arXiv preprint arXiv:2004.01453*.

METZ, J. A. AND O. DIEKMANN (2014): *The dynamics of physiologically structured populations*, vol. 68, Springer.

NISHIURA, H. AND G. CHOWELL (2009): "The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends," in *Mathematical and statistical estimation approaches in epidemiology*, Springer, 103–121.

OBADIA, T., R. HANEEF, AND P.-Y. BOËLLE (2012): "The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks," *BMC medical informatics and decision making*, 12, 147.

PYBUS, O. G., M. A. CHARLESTON, S. GUPTA, A. RAMBAUT, E. C. HOLMES, AND P. H. HARVEY (2001): "The epidemic behavior of the hepatitis C virus," *Science*, 292, 2323–2325.

SHIM, E., A. TARIQ, W. CHOI, Y. LEE, AND G. CHOWELL (2020): "Transmission potential and severity of COVID-19 in South Korea," *International Journal of Infectious Diseases*.

THOMPSON, R., J. STOCKWIN, R. [VAN GAALEN], J. POLONSKY, Z. KAMVAR, P. DEMARSH, E. DAHLQWIST, S. LI, E. MIGUEL, T. JOMBART, J. LESSLER, S. CAUCHEMEZ, AND A. CORI (2019): "Improved inference of time-varying reproduction numbers during infectious disease outbreaks," *Epidemics*, 29, 100356.

WALLINGA, J. AND M. LIPSITCH (2007): "How generation intervals shape the relationship between growth rates and reproductive numbers," *Proceedings of the Royal Society B: Biological Sciences*, 274, 599–604.

WALLINGA, J. AND P. TEUNIS (2004): "Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures," *American Journal of epidemiology*, 160, 509–516.

WILENSKY, U. (1999): "NetLogo," http://ccl.northwestern.edu/netlogo/, Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

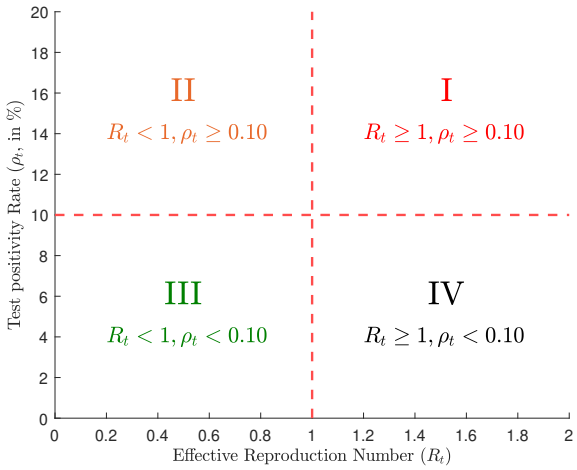


Figure 1: **Join Trajectories Classification Scheme.** We divide the (R_t, ρ_t) space in four quadrants following the $R_t = 1$ threshold that determines if a transmission process is under control and $\rho_t = 0.10$ which implies that the testing scale is appropriate according to the WHO.

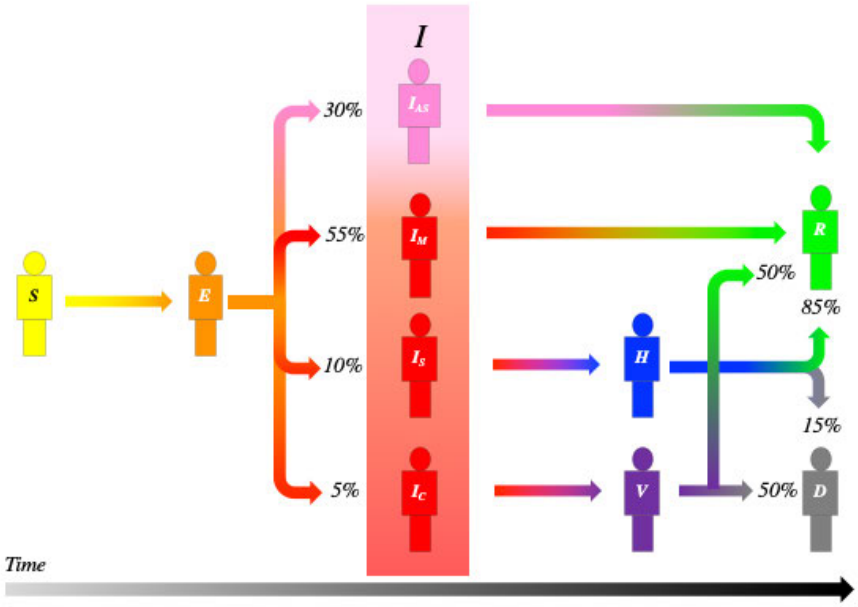
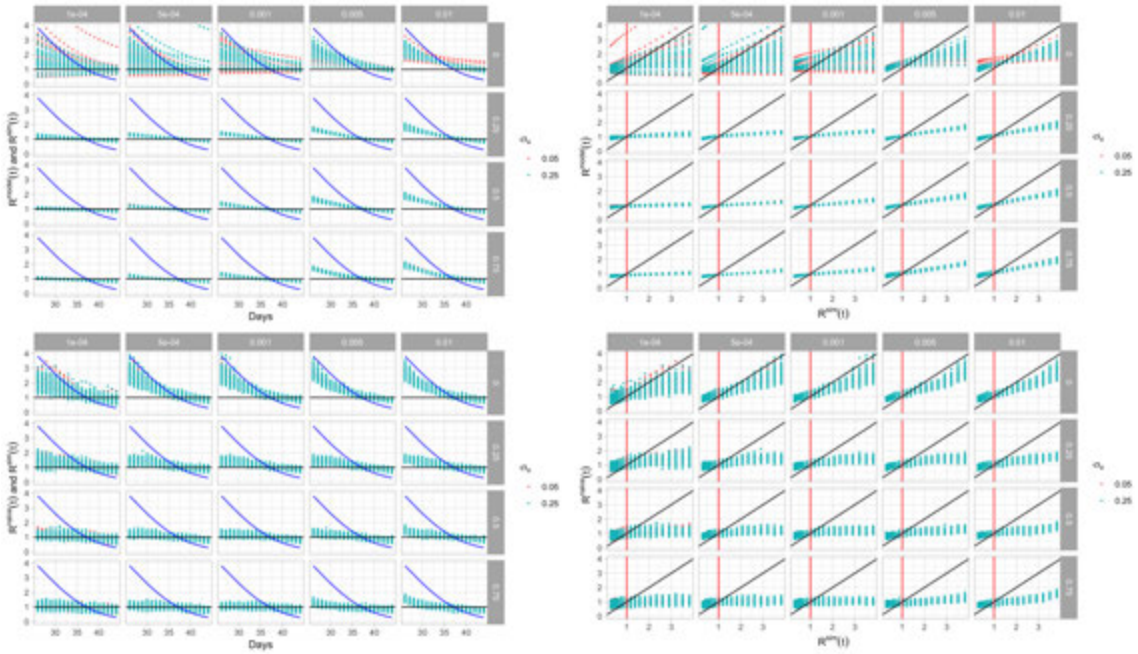


Figure 2: **SEIRD compartments for the evolution of infections in our simulation.** Percentages are transition probabilities between compartments.



(a) (b)

Figure 3: **(a) The time evolution of R_t^j.** Columns represent testing $scale$ and rows represent P_{tests}. The blue line is the true effective reproduction number for the simulation R_t^{sim}. **(b) R_t^j vs. R_t^{sim}.** The black line is at 45 degrees. Columns represent testing $scale$ and rows represent P_{tests}.

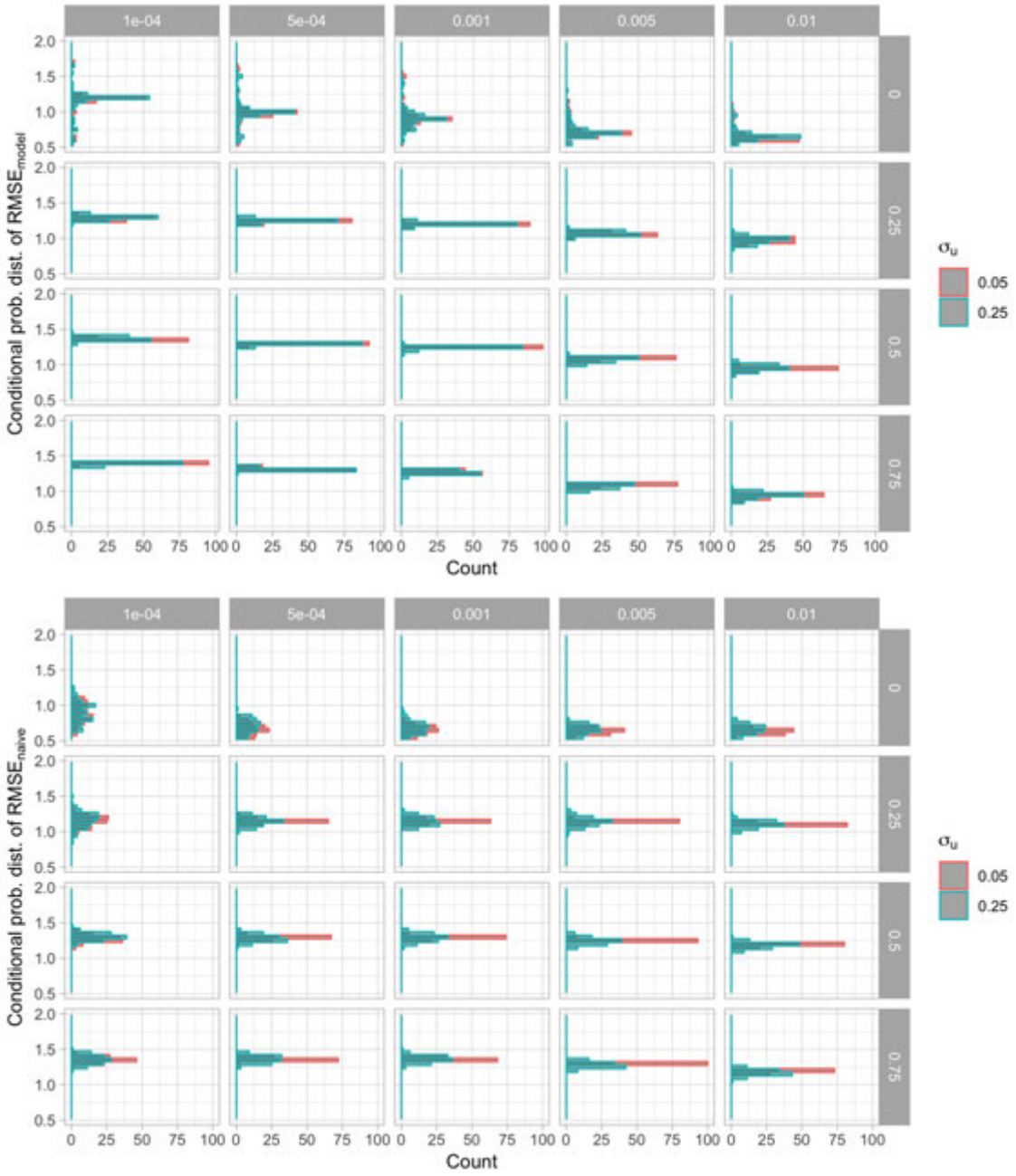


Figure 4: **Conditional probability distribution for $RMSE_j$ where $j = \{model, naive\}$.**

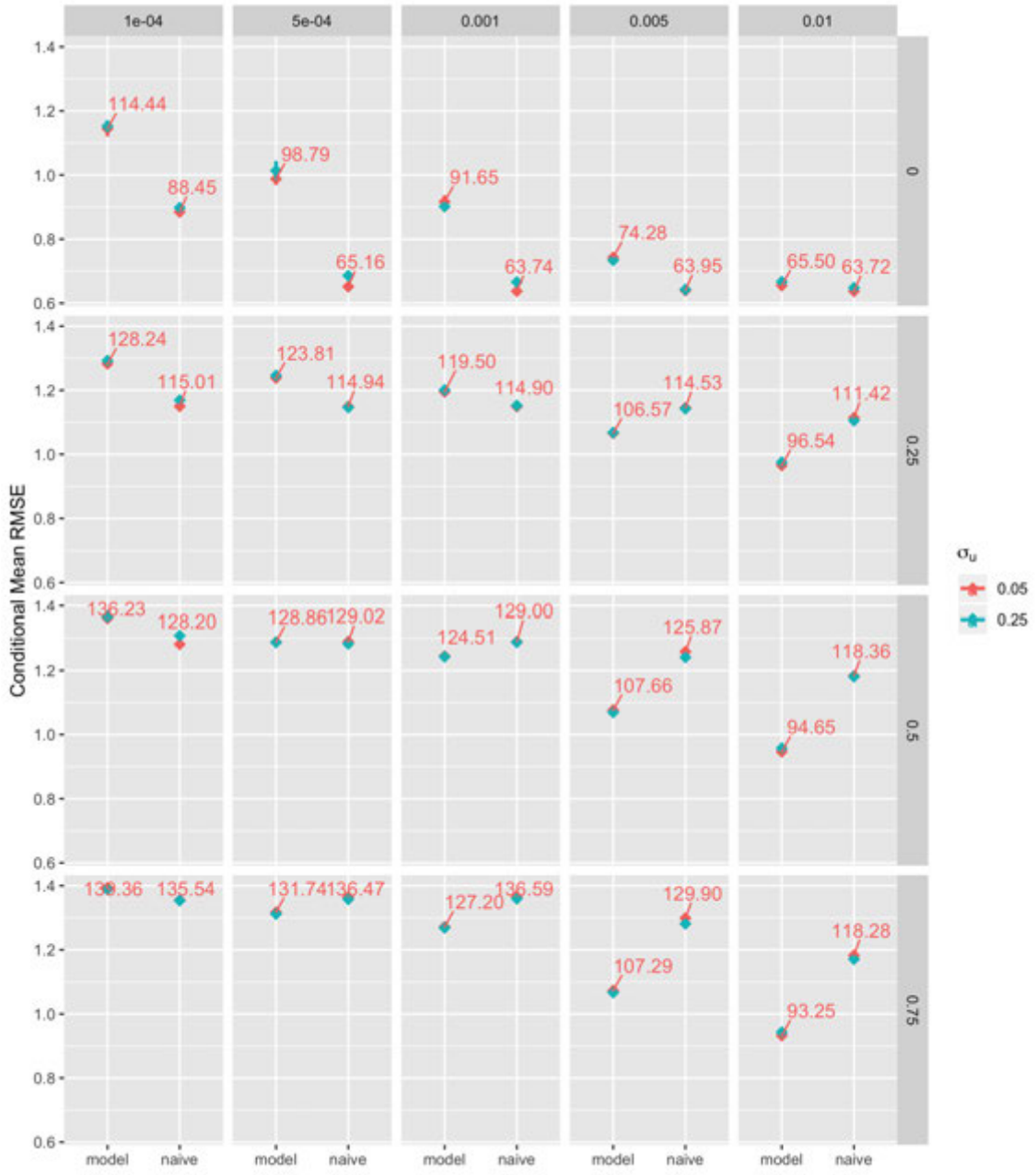


Figure 5: **Conditional Means of RMSE for all testing parameter combinations.** Only the value of the conditional mean for $\sigma_u = 0.05$ is shown as it is very similar to that for $\sigma_u = 0.025$. The standard error shown as a vertical line is barely visible due to its small size.

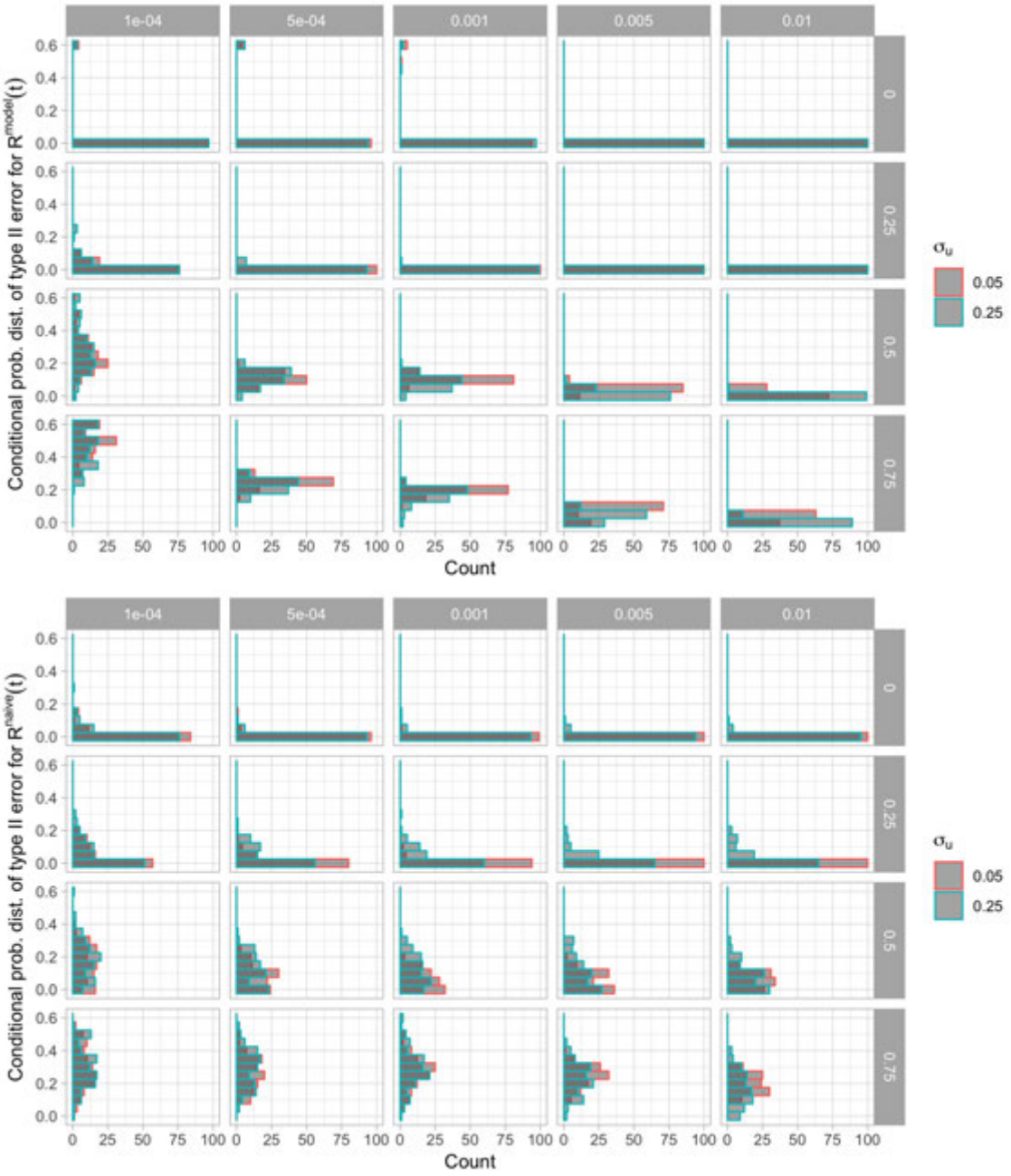


Figure 6: **Conditional probability distribution of type II error, β_j, where $j = \{model, naive\}$.**

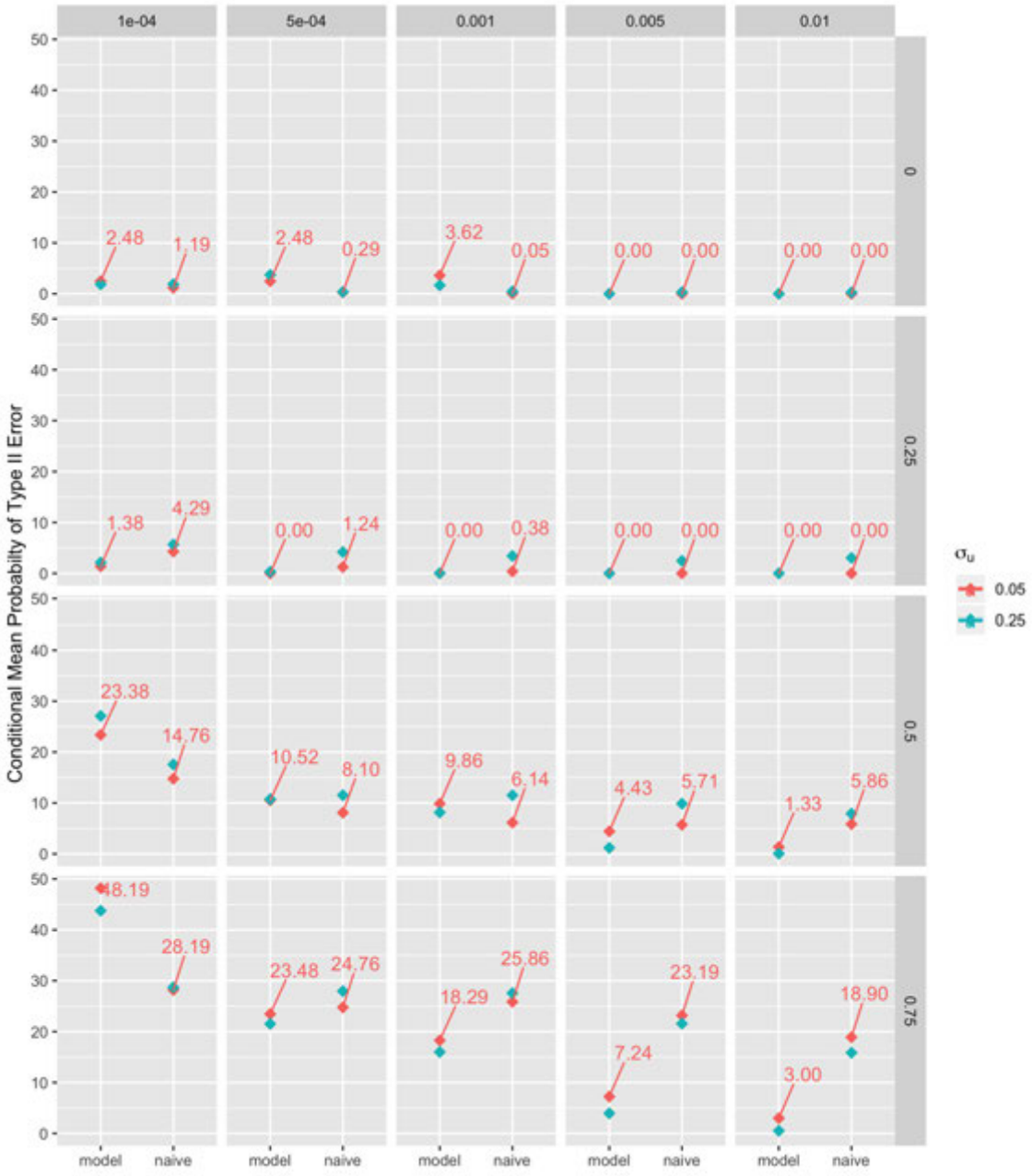


Figure 7: **Conditional Means for the probability of the type II error for all testing parameter combinations.** Only the value of the conditional mean for $\sigma_u = 0.05$ is shown as it is very similar to that for $\sigma_u = 0.025$. The standard error shown as a vertical line is barely visible due to its small size. The best performance is found at high testing scales and random testing.

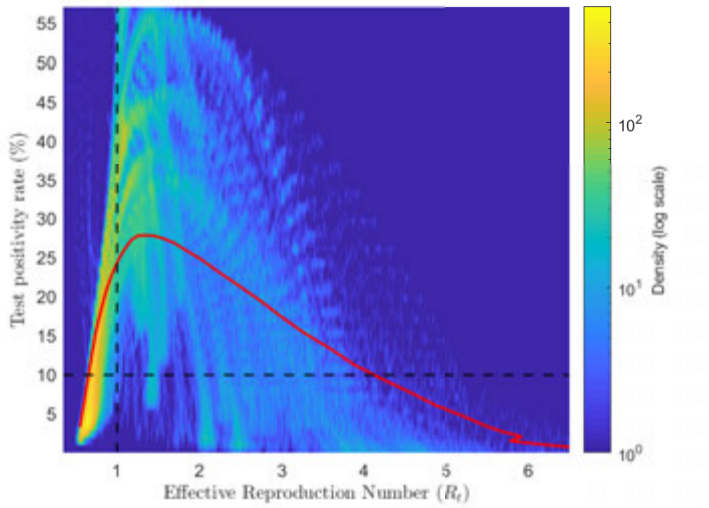
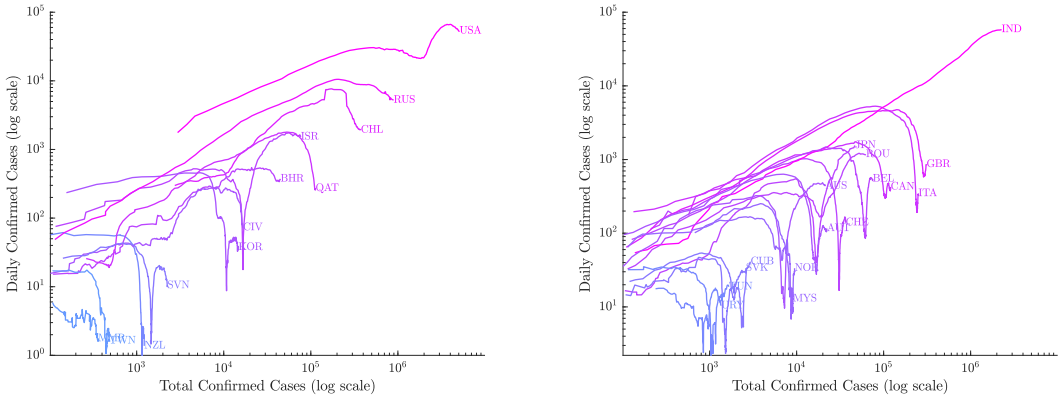


Figure 8: **Density of the joint dynamics of the effective reproduction number and the test positivity rate.** Warmer colors correspond to areas with a higher concentration of simulated trajectories.



(a) Low R_t, low ρ_t

(b) High R_t, low ρ_t

(c) High R_t, high ρ_t

(d) Low R_t, high ρ_t

Figure 9: **Trajectories of countries included in the sample.** Horizontal axis measures cumulative confirmed cases, vertical axis measures daily counts; both axis are measured in logarithmic scales. Dotted lines correspond to corrected data. We organize countries according to the quadrant they occupy in the (R_t, ρ_t) space.

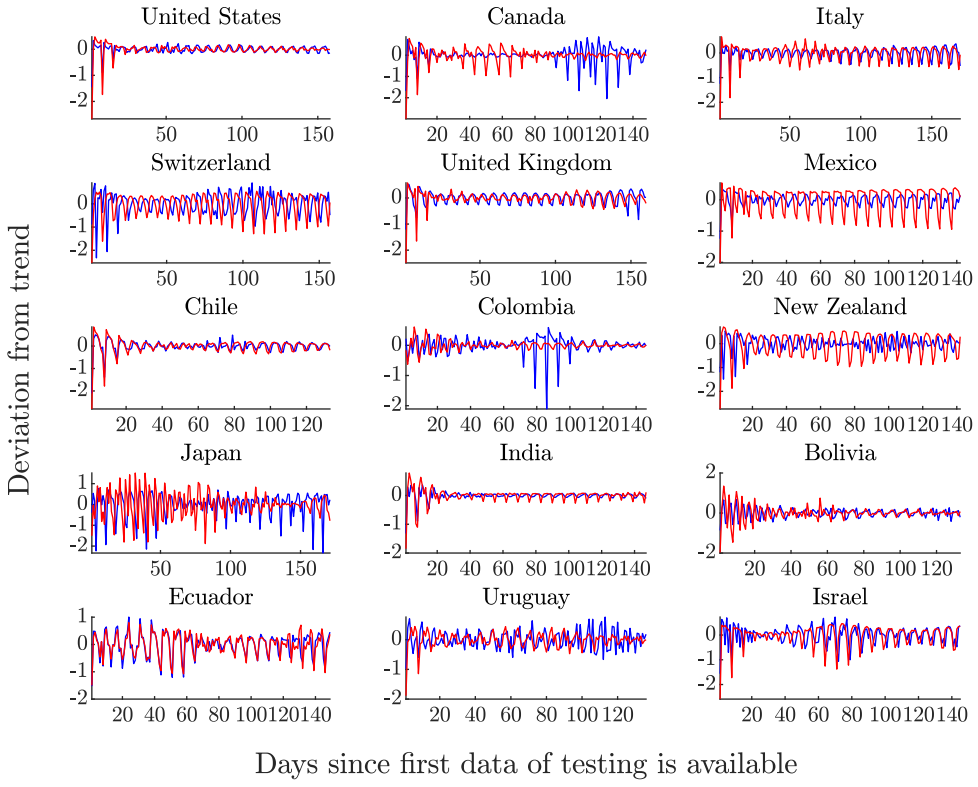


Figure 10: **Seasonal components for daily positives (in blue) and daily tests (in red).**

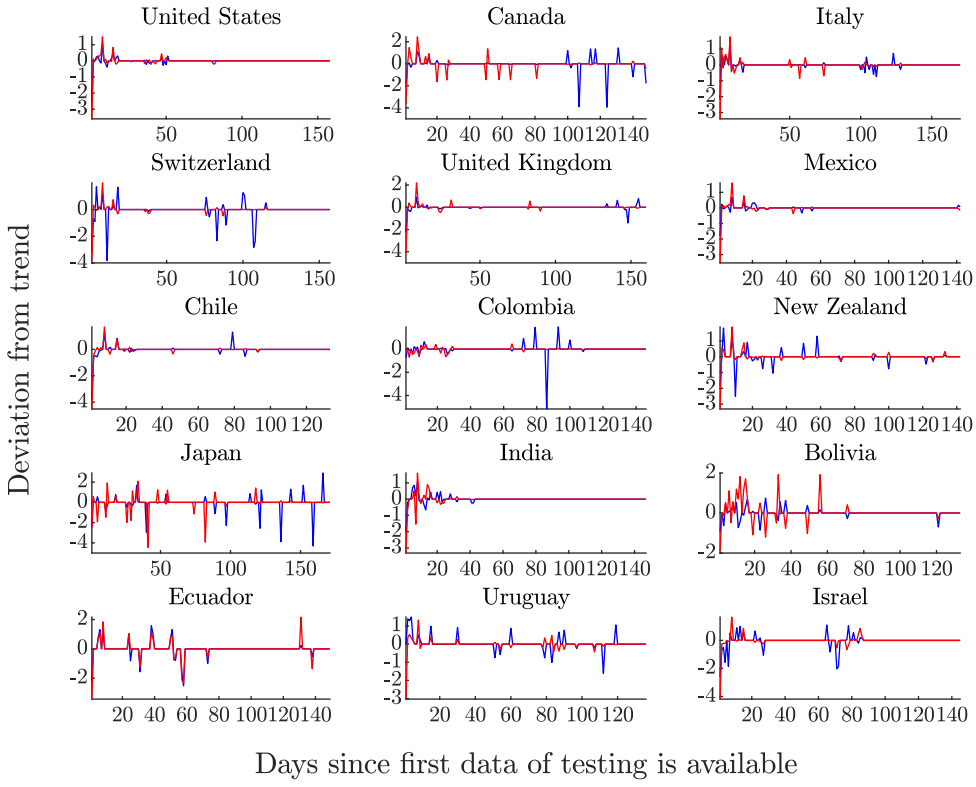
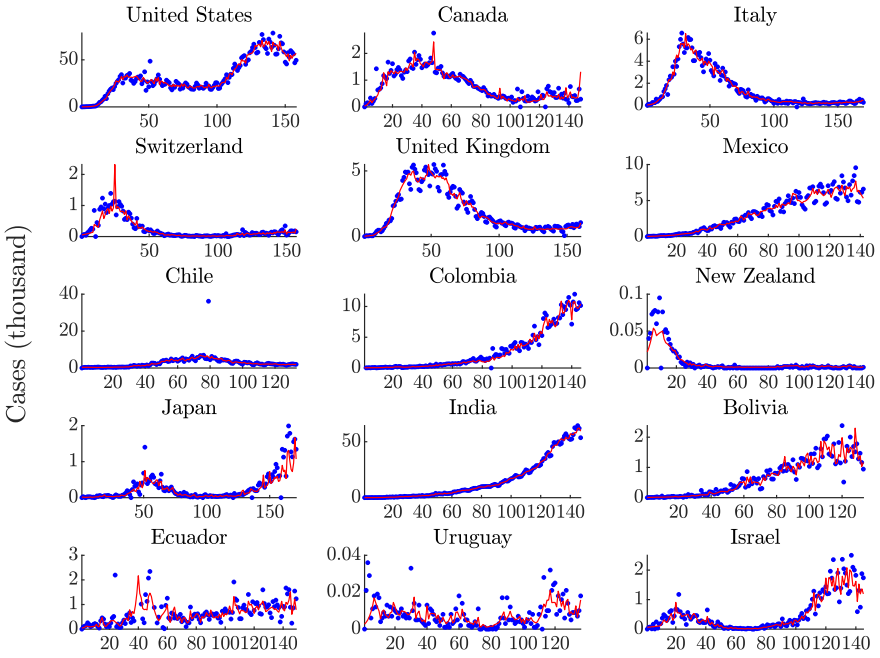
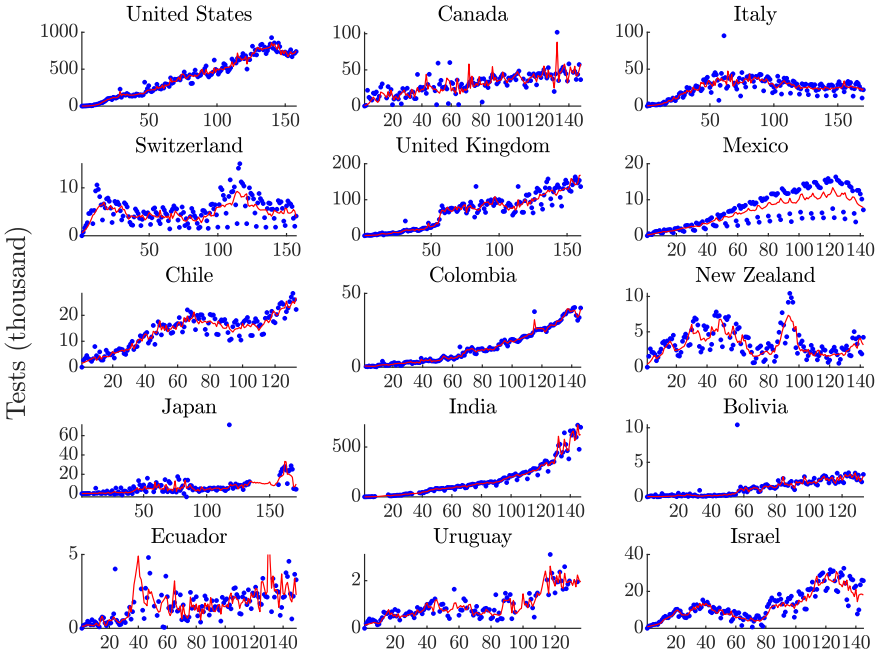


Figure 11: **Atypical values for daily positives (in blue) and daily tests (in red).**



(a) Daily Positives



(b) Daily Tests

Figure 12: **Original (in blue) and adjusted (in red) time series for daily positives and tests for each country.** Vertical axis are measured in thousand cases (tests).

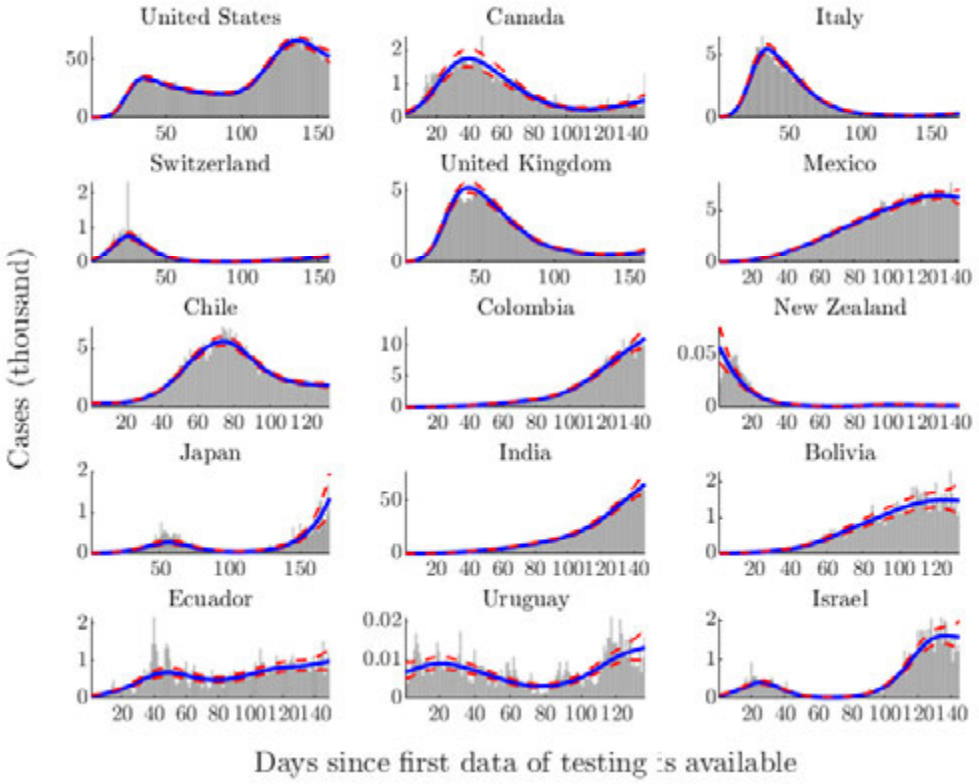


Figure 13: **Estimated trend and confidence intervals for daily positives.** Trend is presented in blue and 95% confidence intervals in dotted red lines. In gray bars we present adjusted daily positives.

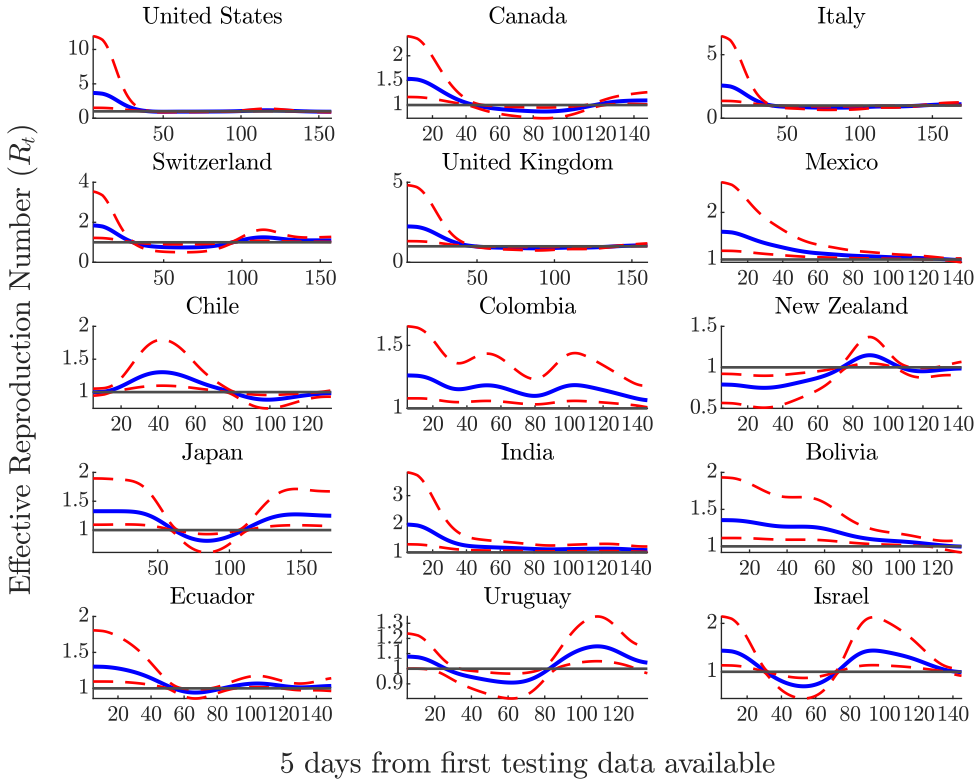


Figure 14: **Estimated effective reproduction number R_t.** The mean is presented in the blue line, and red dotted lines mark the 95% confidence interval. Due to significant variation among countries, we use different vertical axis for each case.

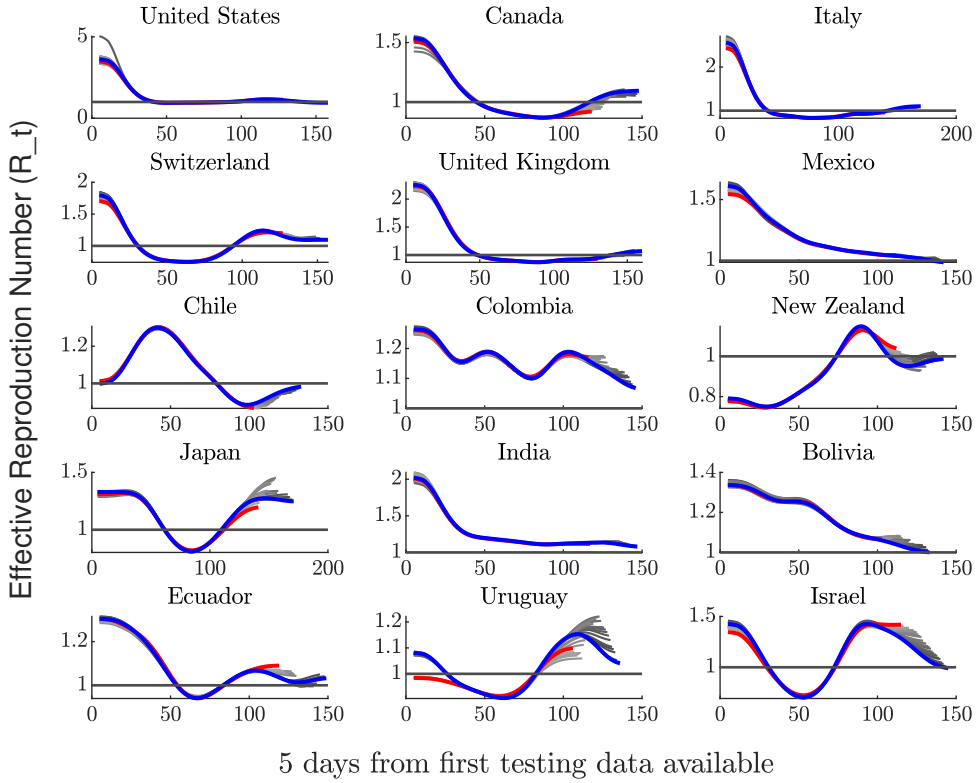


Figure 15: **Estimated effective reproduction number R_t for different sub-samples in each country**. We consider data as it was 25 days ago, and then add new information day by day. The oldest data set is presented in the red line, while the newest is in blue.

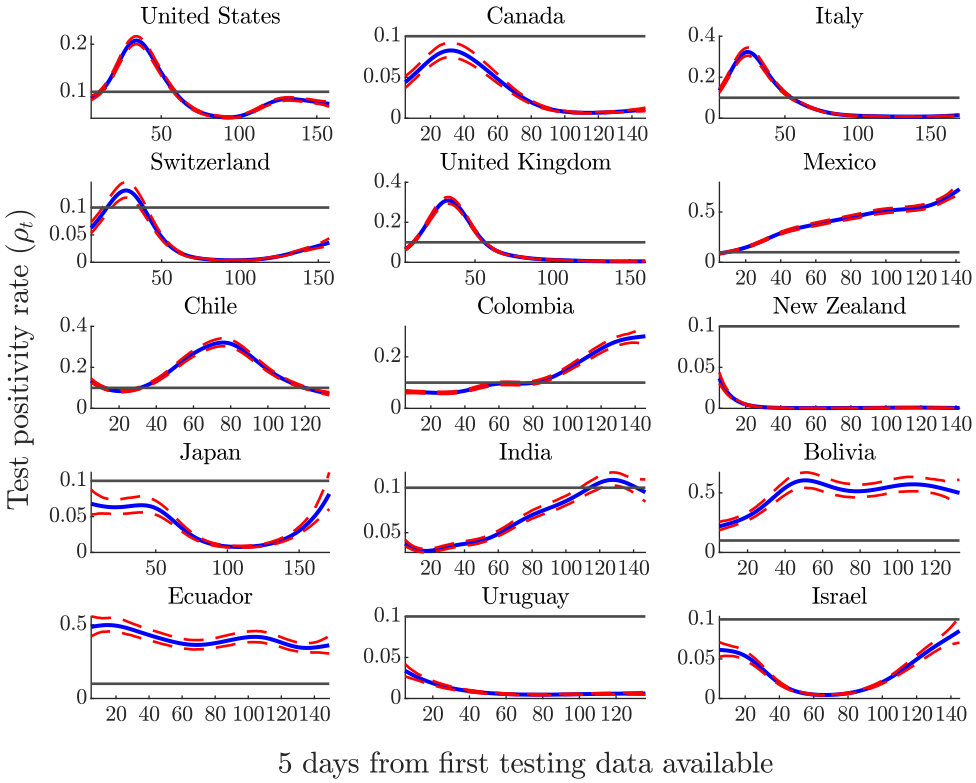
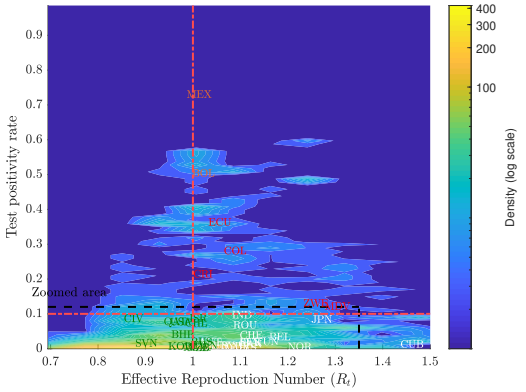
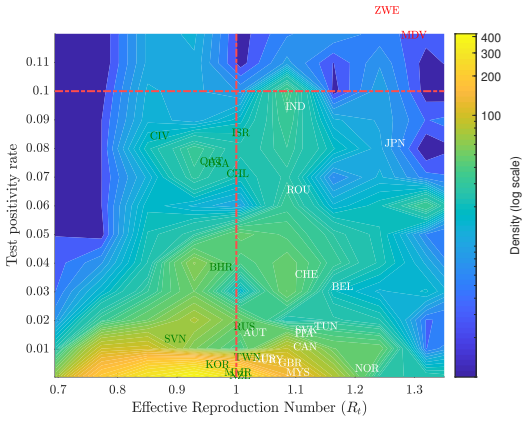


Figure 16: **Estimated test positivity rate ρ_t.** The mean is presented in the blue line, and red dotted lines mark the 95% confidence interval. Due to significant variation among countries, we use different vertical axis for each case.

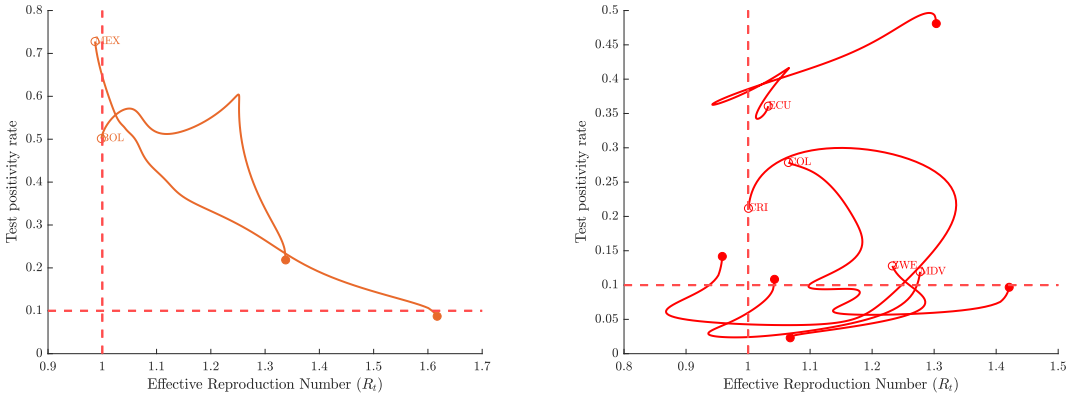


(a) All Countries. Black lines delimit the zoomed area that is presented in panel 17b.



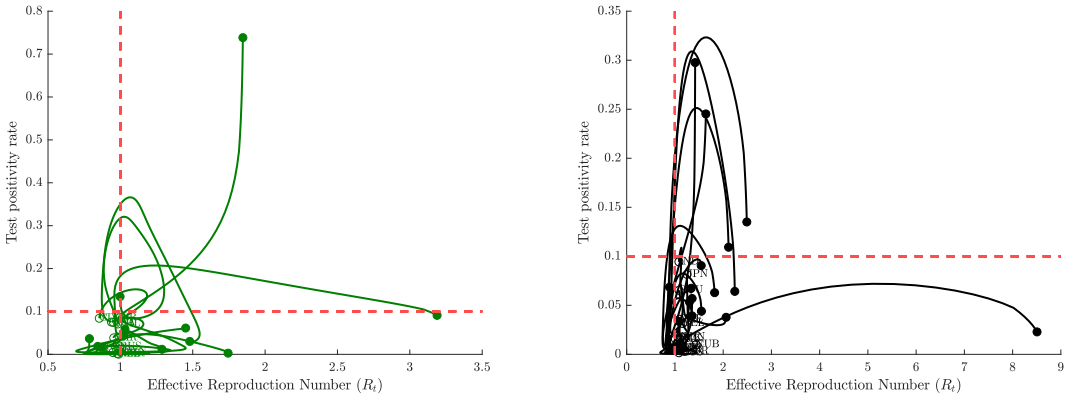
(b) Zoomed Section

Figure 17: **Density of the joint dynamics of the effective reproduction number and the test positivity rate.** Warmer colors correspond to areas with a higher concentration of countries' trajectories. Countries' labels are located at the most recent location.



(a) Low R_t, high ρ_t

(b) High R_t, high ρ_t



(c) Low R_t, low ρ_t

(d) High R_t, low ρ_t

Figure 18: **Trajectories of countries for which their most recent bundle of effective reproduction number and test positivity rate is located at each quadrant.** Full circles represent the starting point, while hollow circles and labels represent most recent data. Thresholds for the effective reproduction number and the test positivity rate are in dotted red lines for reference.

Parameter	Value	Description
N	100,000	Population
E_0	50	Initial number exposed
ρ	10	Average population density $= N/10,000$ patches
$P_{exposed}$	12.5%	Probability of contagion $= \beta/\rho$, assuming $\beta = 1.25$ as estimated in Meidan et al. (2020).
d_{inc}	$d_{inc} \sim f(d_{inc})$	Incubation period follows a Weibull distribution $f(d_{inc})$ with shape parameter $= 1.47$, and scale parameter $= 11.04$.
d_{lat}	$d_{lat} \sim f(d_{lat})$	Latency period follows a Weibull distribution $f(d_{lat})$ with shape parameter $= 1.47$, and scale parameter $= 4.42$.
$d_{I_{AS},R}$	$d_{I_{AS},R} \sim f(d_{I_{AS},R})$	Days to recovery for asymptomatic infections follow Weibull distribution $f(d_{I_{AS},R})$ with shape parameter $= 1.47$, and scale parameter $= 11.04$.
$\gamma_{I_M,R}$	$1/5$	Rate of recovery for I_M.
$\gamma_{I_S,H}$	$1/4$	Rate of transition from I_S to H.
$\gamma_{H,RD}$	$1/11$	Rate of transition from H to resolution.
$\gamma_{I_C,V}$	$1/3$	Rate of transition from I_C to V.
$\gamma_{V,RD}$	$1/13$	Rate of transition from V to resolution.
θ^{sim}	5.05 days	Average Serial interval is found from the average number of days between primary and secondary infections for the model after 60 time periods.
T	60 days	Maximum number of days.

Table 1: **Epidemiological parameters for SEIRD simulations.**

Parameter	Values
$scale$	0.0001, 0.0005, 0.001, 0.005, 0.01
P_{tests}	0, 0.25, 0.5, 0.75
σ_u	0.05, 0.25
g_{tests}	0.025

Table 2: **Testing parameter values used for the simulations.**

	Error in R_t	**Error in ρ_t**	**Error in (R_t, ρ_t)**
Mean	6.06	23.64	0.34
Standard Deviation	(0.239)	(0.425)	(0.058)

Table 3: **Errors in R_t, ρ_t and its joint analysis.** We present means and standard deviations measured in percentages.

	$y := \delta_R$			$y := \delta_\rho$			$y := \delta_{(R,\rho)}$		
	(1) OLS	(2) Logit	(3) Probit	(4) OLS	(5) Logit	(6) Probit	(7) OLS	(8) Logit	(9) Probit
Testing Scale (Tests per 10000 people)									
5	.0234**	–.0402***	–.0402***	–.1095***	–.0322***	–.0322***	–.0035*	–.0035*	–.0043**
	(.0102)	(.0033)	(.0033)	(.0064)	(.0018)	(.0018)	(.0019)	(.0019)	(.0018)
10	.0300***	–.0445***	–.0450***	–1194***	–.0343***	–.0342***	–.0036**	–.0036**	–.0044**
	(.0110)	(.0035)	(.0035)	(.0060)	(.0017)	(.0017)	(.0018)	(.0018)	(.0017)
50	–.0150**	–.0720***	–.0720***	–1540***	–.0885***	–.0885***	–.0074***	–.0074***	–.0079***
	(.0074)	(.0028)	(.0028)	(.0050)	(.0014)	(.0014)	(.0015)	(.0015)	(.0015)
100	–.0354***	–.0797***	–.0800***	–.1411***	–.1288***	–.1288***	–.0077***	–.0077***	–.0083***
	(.0070)	(.0027)	(.0027)	(.0045)	(.0014)	(.0014)	(.0014)	(.0015)	(.0015)
Probability of Testing Symptomatics (Percentage)									
25	–.0797***	–.0843***	–.0843***	.1300***	.1738***	.1738***	–.0087***	–.0087***	–.0090***
	(.0067)	(.0030)	(.0030)	(.0047)	(.0016)	(.0016)	(.0014)	(.0014)	(.0014)
50	.0059	–.0613***	–.0613***	.1341***	.1976***	.1976***	–.0071***	–.0071***	–.0074***
	(.0080)	(.0031)	(.0031)	(.0047)	(.0016)	(.0016)	(.0015)	(.0015)	(.0015)
75	.0942***	–.0260***	–.0259***	.1372***	.2221***	.2221***	–.0064***	–.0064***	–.0068***
	(.0088)	(.0032)	(.0032)	(.0048)	(.0016)	(.0016)	(.0016)	(.0016)	(.0015)
Error scale	.0008	.0008	.0010	.0033***	.0033***	.0033***	.0004	.0004	.0004
	(.0017)	(.0017)	(.0016)	(.0009)	(.0009)	(.0010)	(.0131)	(.0009)	(.0008)
Interactions	YES	YES	YES	YES	YES	YES	NO	NO	NO
Observations	215,933	215,933	215,933	215,933	215,933	215,933	215,933	215,933	215,933

Table 4: **Effects of testing parameters on classification errors.** We report marginal effects in all cases. (*) significant at the 90% level, (**) significant at the 95% level, (***) significant at the 99% level. Clustered standard errors (at the simulation level) in parenthesis. Ommitted categories for testing scale, probability of testing symptomatics and error scale are testing 1 per every 10,000 people, random testing and low error scale respectively.

Country	λ_ρ^*	λ_s^*	**Country**	λ_ρ^*	λ_s^*
Bolivia	5,039.38	3,330.30	Japan	12,210.44	6,908.89
Canada	5,084.30	3,302.42	Mexico	664.33	767.95
Chile	942.85	1,931.43	New Zealand	6,084.17	2,658.15
Colombia	1,390.30	787.28	Switzerland	4,050.83	1,049.76
Ecuador	2,813.13	6,387.85	United Kingdom	1,028.52	1,364.81
India	1,360.89	972.38	United States	585.29	311.90
Israel	3,386.80	2,099.95	Uruguay	10,292.96	2,238.21
Italy	1,364.23	623.95			

Table 5: **Estimated smoothing parameters for the probability of detecting positives λ_ρ^* and daily testing λ_s^* for each country.** The scaling parameter used to compute these numbers was $\zeta = 50,000$. Keep in mind that, since these parameters are endogenous, they change every time that data is updated.

Country	R_t	ρ_t	CFR	Mortality Rate	Test Scale
		Quadrant 1			
Ecuador	[0.96 , 1.03 , 1.14]	[0.30 , 0.36 , 0.43]	9.72	5.42	133.22
Colombia	[1.01 , 1.07 , 1.18]	[0.25 , 0.28 , 0.32]	3.31	2.59	389.69
Costa Rica	[0.92 , 1.00 , 1.08]	[0.17 , 0.21 , 0.27]	1.01	0.46	185.04
Zimbabwe	[1.07 , 1.23 , 1.59]	[0.09 , 0.13 , 0.18]	2.23	0.07	48.18
Maldives	[1.09 , 1.28 , 1.74]	[0.09 , 0.12 , 0.16]	0.38	0.35	1656.76
		Quadrant 2			
Mexico	[0.94 , 0.99 , 1.03]	[0.67 , 0.73 , 0.79]	10.92	3.92	79.61
Bolivia	[0.92 , 1.00 , 1.07]	[0.40 , 0.50 , 0.61]	4.04	3.12	153.82
		Quadrant 3			
Israel	[0.92 , 0.99 , 1.06]	[0.07 , 0.09 , 0.10]	0.73	0.65	2157.84
Cote d'Ivoire	[0.68 , 0.85 , 0.95]	[0.06 , 0.08 , 0.12]	0.63	0.04	41.58
Qatar	[0.83 , 0.94 , 0.99]	[0.07 , 0.08 , 0.09]	0.17	0.65	1837.13
United States	[0.88 , 0.95 , 0.99]	[0.07 , 0.07 , 0.08]	3.21	4.94	1910.94
Chile	[0.93 , 0.98 , 1.02]	[0.06 , 0.07 , 0.08]	2.70	5.30	988.49
Bahrain	[0.86 , 0.95 , 1.01]	[0.03 , 0.04 , 0.05]	0.37	0.97	5426.37
Russia	[0.96 , 1.00 , 1.03]	[0.02 , 0.02 , 0.02]	1.69	1.03	2145.84
Slovenia	[0.74 , 0.88 , 0.96]	[0.01 , 0.01 , 0.02]	5.32	0.58	667.26
		Quadrant 4			
India	[1.02 , 1.08 , 1.23]	[0.08 , 0.09 , 0.11]	1.99	0.33	183.20
Japan	[1.07 , 1.25 , 1.64]	[0.06 , 0.08 , 0.11]	2.22	0.08	101.88
Romania	[1.01 , 1.08 , 1.24]	[0.05 , 0.07 , 0.08]	4.36	1.42	730.26
Switzerland	[1.02 , 1.10 , 1.27]	[0.03 , 0.04 , 0.04]	4.69	1.98	979.04
Belgium	[1.05 , 1.16 , 1.39]	[0.03 , 0.03 , 0.03]	13.21	8.52	1629.10
Tunisia	[1.03 , 1.13 , 1.36]	[0.01 , 0.02 , 0.03]	3.08	0.04	86.33
Slovakia	[1.02 , 1.10 , 1.28]	[0.01 , 0.02 , 0.02]	1.19	0.06	520.38
Austria	[0.95 , 1.01 , 1.09]	[0.01 , 0.02 , 0.02]	3.25	0.80	1091.47
Italy	[1.03 , 1.10 , 1.25]	[0.01 , 0.02 , 0.02]	14.04	5.82	716.10
Canada	[1.01 , 1.10 , 1.28]	[0.01 , 0.01 , 0.01]	7.47	2.38	1203.36
Cuba	[1.12 , 1.44 , 2.20]	[0.01 , 0.01 , 0.01]	2.98	0.08	269.85
Australia	[0.94 , 1.03 , 1.16]	[0.00 , 0.01 , 0.01]	1.46	0.12	1953.02
Uruguay	[0.97 , 1.04 , 1.15]	[0.00 , 0.01 , 0.01]	2.73	0.11	383.56
United Kingdom	[1.02 , 1.07 , 1.18]	[0.00 , 0.00 , 0.01]	14.98	6.86	1619.41
Norway	[1.06 , 1.20 , 1.52]	[0.00 , 0.00 , 0.00]	2.70	0.47	885.79
Malaysia	[0.99 , 1.08 , 1.26]	[0.00 , 0.00 , 0.00]	1.37	0.04	328.53
		Controlled Process			
Taiwan	[0.93 , 1.00 , 1.06]	[0.01 , 0.01 , 0.01]	1.46	0.00	35.20
South Korea	[0.84 , 0.95 , 1.01]	[0.00 , 0.00 , 0.01]	2.07	0.06	317.79
Myanmar	[0.89 , 0.98 , 1.05]	[0.00 , 0.00 , 0.00]	1.67	0.00	23.31
New Zealand	[0.90 , 0.99 , 1.07]	[0.00 , 0.00 , 0.00]	1.80	0.05	1042.47

Table 6: **Classification in the (R_t, ρ_t) Space. Results are computed using the most recent information available for each country.** CFR refers to the Case Fatality Rate. The *Controlled Process* category includes countries for which the effective reproduction number is close to (or below) 1 and the test positivity rate is virtually zero. The mortality rate and the testing scale are both measured per 10000 inhabitants.