

## Article

# Modeling COVID-19 Infection Rates by Regime-Switching Unobserved Components Models

Paul Haimerl<sup>1,2,†</sup>  and Tobias Hartl<sup>2,3,\*,†</sup> 

<sup>1</sup> School of Business and Economics, Maastricht University, 6200 MD Maastricht, The Netherlands; p.haimerl@student.maastrichtuniversity.nl

<sup>2</sup> Department of Economics and Econometrics, University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany

<sup>3</sup> Institute for Employment Research, Regensburger Str. 104, 90478 Nuremberg, Germany

\* Correspondence: tobias1.hartl@ur.de; Tel.: +49-941-943-2739

† These authors contributed equally to this work.

**Abstract:** The COVID-19 pandemic is characterized by a recurring sequence of peaks and troughs. This article proposes a regime-switching unobserved components (UC) approach to model the trend of COVID-19 infections as a function of this ebb and flow pattern. Estimated regime probabilities indicate the prevalence of either an infection up- or down-turning regime for every day of the observational period. This method provides an intuitive real-time analysis of the state of the pandemic as well as a tool for identifying structural changes ex post. We find that when applied to U.S. data, the model closely tracks regime changes caused by viral mutations, policy interventions, and public behavior.

**Keywords:** COVID-19; regime-switching; unobserved components; Kim filter; Gibbs sampling

**JEL Classification:** C32; C51; I10



**Citation:** Haimerl, Paul, and Tobias Hartl. 2023. Modeling COVID-19 Infection Rates by Regime-Switching Unobserved Components Models. *Econometrics* 11: 10. <https://doi.org/10.3390/econometrics11020010>

Academic Editors: Benedikt Poetscher, Leopold Sögner and Martin Wagner

Received: 30 September 2022

Revised: 9 March 2023

Accepted: 28 March 2023

Published: 3 April 2023

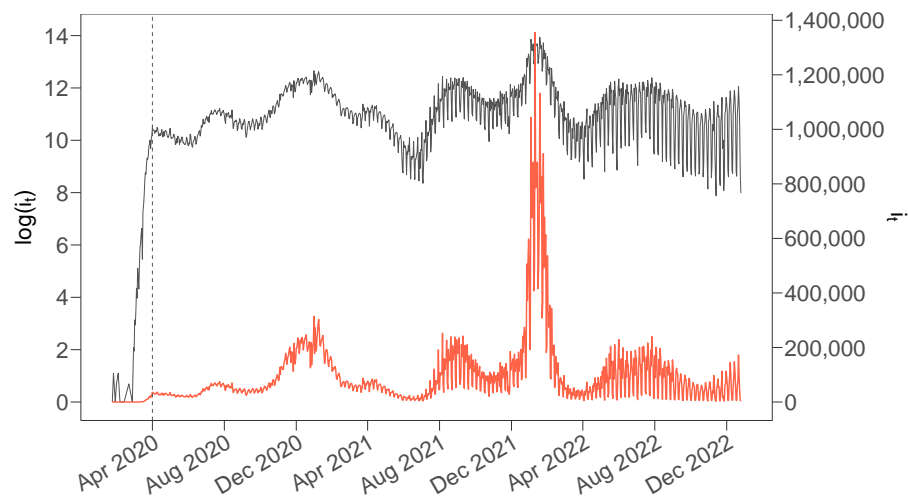


**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Since its onset in late 2019, the COVID-19 pandemic has permeated next to all facets of public life. Even in the transition of the COVID-19 pandemic to an endemic state, the by now well-known routine of alternating infection peaks and troughs will demand close observation for the foreseeable future (Telenti et al. 2021). However, analyzing and monitoring the state and development of the pandemic is complicated by the nature of the data on COVID-19 case numbers: a glimpse at Figure 1 reveals key characteristics include the strong persistence and nonstationarity of case numbers (Dolton 2021), as well as alternating regimes of increasing and decreasing infections caused by policy interventions, medical innovations, seasonal climate conditions, and the evolution of the virus itself (Doornik et al. 2022; Fison et al. 2021). In addition, infection dynamics are overlapped by a seasonal pattern of increasing volatility, generated by a varying number of tests over the days of the week (Bergman et al. 2020), as well as by measurement errors (Hortaçsu et al. 2021).

While a variety of econometric tools have been utilized to study the dynamics of COVID-19 case numbers, in particular, unobserved components (UC) models have been found successful in capturing the aforementioned characteristics of COVID-19-related data.<sup>1</sup> In the early stages of the pandemic, UC models were used to fit linear deterministic trends to COVID-19 case numbers and to identify structural breaks (Hartl et al. 2020; Lee et al. 2021; Liu et al. 2021). With an increasing number of observations available, UC models with stochastic trends have been considered by Moosa (2020) and Doornik et al. (2022), while (stationary) seasonal components were added by Navas Thorakkattle et al. (2022) and Xie (2022), among others. However, appropriately accounting for the alternating peaks and troughs in the trend of COVID-19 case numbers remains an open challenge.



**Figure 1.** Daily U.S. COVID-19 infections  $i_t$  (orange, right scale) and logarithm of daily infections  $\log(i_t)$  (gray, left scale). The vertical line (gray, dashed) indicates the 1st of April 2020, the start of the observational period.

We contribute to the UC literature by explicitly modeling the peak and trough pattern of infection numbers, which has emerged as one of the defining features of the COVID-19 pandemic. For this purpose, we introduce a regime-switching UC model that decomposes log daily COVID-19 infections into trend, seasonal, and cyclical components. While the trend is formulated as a random walk (RW) with drift to capture the long-run dynamics of log new infections, the novelty of our model is that the drift term is made regime-dependent to account for alternating periods of increasing and decreasing infections, as well as for the strong persistence of the respective regime. A seasonal component is added to model the weekly recurring pattern of case numbers, while an autoregressive (AR) term accounts for short-term dependencies in the data.

Regime-switching and mixture models have been studied extensively in the past. We refer to Frühwirth-Schnatter (2006); Kim and Nelson (2017) for an overview. As proposed in Kim (1994); Kim and Nelson (2017, ch. 5), we estimate the trend, seasonal, and cyclical components via the Kim filter, which is an extension of the Kalman filter to regime-switching models: between the prediction and updating step of the Kalman filter, it executes the recursions of Hamilton (1989) to estimate the regime probabilities and thus allows for regime switching in a state-space framework. Parameter estimation is carried out by numerical optimization of the likelihood function, where we make use of an extensive grid search to be robust against local optima. As an alternative, one could also utilize the Expectation–Maximization (EM) algorithm, which has recently been derived for the Kim filter by Degras et al. (2022). As an alternative estimation strategy, we also consider a Bayesian–Gibbs sampling approach. Markov chain Monte Carlo (MCMC) techniques have been widely employed for state-space and especially regime-switching models (see Frühwirth-Schnatter 2006). In contrast to the Kim filter, inference is based on the joint distribution of the state vector, regime probabilities, and additional parameters, rather than on conditional distributions (Frühwirth-Schnatter 2006, ch. 13).

The model is applied to daily U.S. infection data provided by the Johns Hopkins University Center for Systems Science and Engineering (JH/CSSE) (Dong et al. 2020). We found that the estimated regime probabilities closely track the pattern of infection waves throughout the whole chronology of the pandemic. This allows for easy ex post evaluation of the effectiveness of public health interventions or the severity of viral mutations: whereas significant changes tip the system into the opposite regime, harmless mutations or insignificant interventions do not trigger a switch. Moreover, we introduce a nowcasting application that provides an easy-to-understand and concise monitoring tool for the current state of the pandemic.

The plan of this paper is as follows. Section 2 describes the data and methodology. Section 3 presents empirical findings and examines the nowcasting application. Section 4 discusses alternative model specifications, contrasts the Kim filter with a Bayesian–Gibbs sampling approach, and presents a Monte Carlo study to evaluate the reliability of parameter estimates for the preferred specification. Section 5 concludes. All R code to replicate this paper is available at the Github repository <https://github.com/Paul-Haimerl/Regime-SW-UC-COVID-19> Repository accessed on 2 April 2023.

## 2. Data and Methodology

To analyze the state and dynamics of the COVID-19 pandemic, we consider data from the JH/CSSE, which can be downloaded from <https://github.com/CSSEGISandData/COVID-19> (accessed on 14 March 2023). Figure 1 sketches the data on reported daily COVID-19 infections from the 22nd of January 2020 to the 25th of December 2022 for the U.S. in both levels and logarithms.

As can be seen, data on daily infections in the early phase of the pandemic are downward-biased due to limited test capacities. As test availability increases, one sees a rapid increase in case numbers in March 2020.<sup>2</sup>

Accounting for this bias, we select the 1st of April 2020 as the start date for our study. Tracking the pandemic until the end of 2022 provides a total of  $T = 1005$  daily observations.<sup>3</sup>

To set up our model, let  $i_t$  denote new daily U.S. COVID-19 infections and define  $y_t = \log(i_t)$ . The measurement equation of the unobserved components model is then given by

$$y_t = \mu_t + \gamma_t + c_t, \quad t = 1, \dots, T, \quad (1)$$

where  $\mu_t$  denotes the trend,  $\gamma_t$  models the seasonal component, and  $c_t$  is a stationary cyclical component. As Figure 1 suggests, the model is formulated in terms of logarithmic case numbers to reflect the exponential growth of COVID-19. Furthermore, taking logs allows for proportional seasonal effects and measurement errors rather than additive effects (see e.g., Doornik et al. 2021; Lee et al. 2021; Liu et al. 2021).

As typical in the literature, the seasonal component  $\gamma_t$  is modeled as a mean-zero deterministic process (Durbin and Koopman 2012; Navas Thorakkattle et al. 2022, ch. 3.2.2)

$$\gamma_t = - \sum_{j=1}^6 \gamma_{t-j}. \quad (2)$$

The cyclical component  $c_t$  is to capture short-run dynamics in the data. Therefore, we define  $c_t$  as a stationary AR(2) process

$$(1 - L\phi_1 - L^2\phi_2)c_t = \eta_t \quad \eta_t \sim i.i.d.N(0, \sigma_\eta^2). \quad (3)$$

$L$  denotes the lag operator  $Lx_t = x_{t-1}$ . The motivation behind (3) is to allow for some additional short-term dependencies and thus for autocorrelated measurement errors instead of a purely unsystematic additive component with no temporal correlation. Such short-term fluctuations can arise, e.g., due to testing bottlenecks, large-scale public events, the gradual introduction of new policies, or time lags in the reporting of new cases.

Turning to the trend, we specify  $\mu_t$  as a random walk with drift to capture the permanent, smooth low-frequency dynamics, as displayed in Figure 1.

$$\begin{aligned} \mu_t &= \mu_{t-1} + v_t + \zeta_t & \zeta_t &\sim i.i.d.N(0, \sigma_\zeta^2), \\ v_t &= v_0 + S_t v_1. \end{aligned} \quad (4)$$

The drift term  $v_t$  depends on the current regime indicator  $S_t \in \{0, 1\}$ . In order to uniquely identify the system, we impose  $v_1 < 0$ , which declares  $S_t = 1$  as the infection down-turning regime and  $S_t = 0$  as the infection up-turning regime. Consequently, (4)

allows for alternating regimes of increasing and decreasing infections, as well as for different rates of growth and decay. Specification (4) can be justified by visual inspection of Figure 1, which shows that log infections are driven by persistent but alternating phases of increasing and decreasing case numbers. Regime switches may be triggered by major epidemic events that have a persistent impact, such as virus mutations, policy changes, or medical innovations, among others.

The switching behavior between the two states is modeled by a first-order stationary Markov process (Hamilton 1989, ch. 2), with the transition probabilities

$$\begin{aligned} \Pr(S_t = 1|S_{t-1} = 1) &= p, & \Pr(S_t = 0|S_{t-1} = 1) &= 1 - p, \\ \Pr(S_t = 0|S_{t-1} = 0) &= q, & \Pr(S_t = 1|S_{t-1} = 0) &= 1 - q. \end{aligned} \quad (5)$$

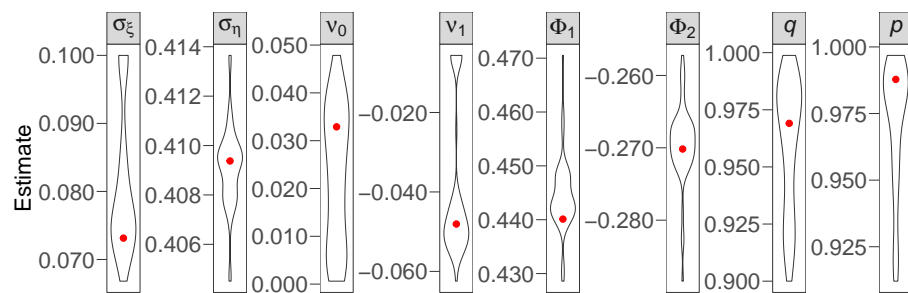
The transition probabilities are constrained to satisfy  $p, q > 90\%$ , which resembles the path-dependent behavior of infection-increasing and infection-decreasing regimes.<sup>4</sup>

In the following, we refer to the specification in (1) to (5) as the *D.Seas.C* model. To estimate the parameters  $\theta = (\sigma_{\xi}, \sigma_{\eta}, \nu_1, \nu_0, \phi_1, \phi_2, p, q)'$  we proceed similarly to the UC literature: First, the model is cast in state-space form. Based on the Kim filter, we obtain a conditional likelihood function that combines the prediction and updating steps of the Kalman filter with the regime-switching recursions of Hamilton (1989) (see Kim 1994, ch. 2 for details). Maximizing the conditional likelihood yields the desired parameter estimates and is analogous to the usual prediction error decomposition in the UC literature, which involves the Kalman filter instead of the Kim filter (see Harvey 1989, ch. 4).<sup>5</sup> While the estimation results for the *D.Seas.C* model are presented in the next section, we also discuss generalizations, including a stochastic specification of the seasonal term and the inclusion of a third state in Section 4.

### 3. Empirical Results

To initiate the Kim filter, we initialize  $\mu_0$  with the number of reported COVID-19 cases on the 31st of May 2020, one day prior to the start of the observational period. The remaining entries to the state vector and the diagonal of the state covariance matrix are initialized diffusely, as is common in the UC literature (see Durbin and Koopman 2012, ch. 5).

To ensure stable estimates of all parameters in  $\theta$  and to avoid convergence to local optima, we employ an extensive three-step grid search that covers the relevant parameter space while remaining computationally feasible: In a first global grid search, 30,000 parameter combinations are randomly drawn from uniform distributions with sufficiently wide support to encompass the entire parameter space. To narrow down the locations of local and global optima, we pick the 50  $\theta$ -vectors corresponding to the greatest likelihood values. For each parameter, we store the minimum and maximum value of these 50 combinations, which gives us the range of the relevant parameter space. In the second step, these minimum and maximum values are employed as bounds to construct a finer parameter grid. After computing the likelihood of each grid point, we use the 50 grid points yielding the greatest likelihood as starting values for numerical optimization via the Nelder–Mead algorithm. The optimized parameters corresponding to the greatest log likelihood are then chosen as the final parameter estimate  $\hat{\theta}$ . Figure 2 outlines the parameter distributions as generated by these 50 optimization results.



**Figure 2.** Parameter-specific kernel densities of the final-step grid search results. The respective best estimate is marked in red. Note that  $p$  and  $q$  are floored at 90% (see Section 2).

As all parameters of our model (1) to (5) feature a narrow distribution around their best estimate, they appear accurately identified.

Table 1 presents the estimation results and corresponding standard errors.

**Table 1.** Maximum likelihood estimation result for the *D.Seas.C* specification.

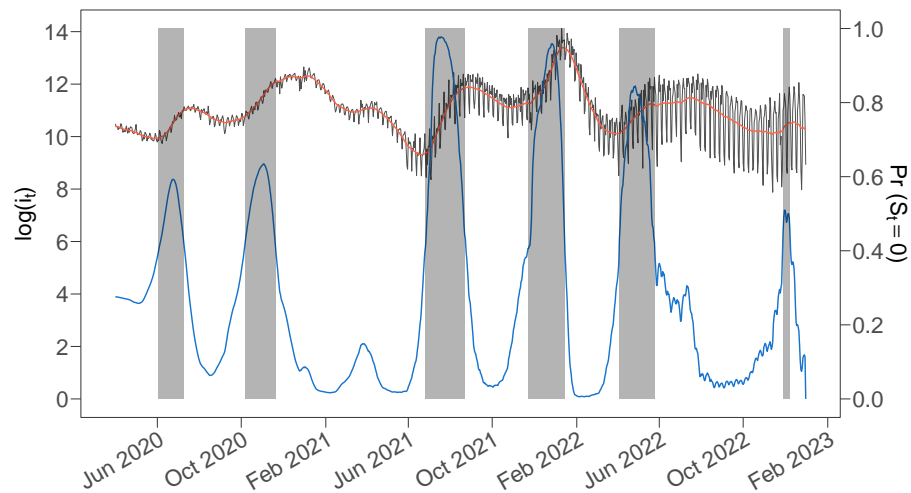
Parameter	Estimate	Standard Error
$\sigma_{\xi}$	0.073	0.008
$\sigma_{\eta}$	0.409	0.010
$\nu_0$	0.033	0.004
$\nu_1$	−0.048	0.010
$\phi_1$	0.440	0.033
$\phi_2$	−0.270	0.032
$q$	0.969	0.017
$p$	0.988	0.010

**Log L:** −677.783 **AIC:** 1.379 **BIC:** 1.452 **HQ:** 1.407

*Notes:* The maximum likelihood estimates are based on the grid search, as described in Section 3. Standard errors are obtained from the inverted numerical Hessian matrix. Information criteria are adjusted to reflect the diffuse initialization of the state vector (Durbin and Koopman 2012, ch. 7.4).

Based on the transition probabilities  $\hat{p}$  and  $\hat{q}$ , it is straightforward to calculate the regime durations. The model estimates an expected duration of  $(1 - \hat{p})^{-1} \approx 83$  days for the down-turning  $S_t = 1$  regime and  $(1 - \hat{q})^{-1} \approx 32$  days for the up-turning  $S_t = 0$  state. The drift estimates  $\hat{\nu}_1$  and  $\hat{\nu}_0$  are interpreted as the average day-to-day change in the trend in log COVID-19 cases. Transforming back into levels gives an average daily decrease of  $1 - \exp^{\hat{\nu}_0 + \hat{\nu}_1} \approx 1.49\%$  for the down-turning regime on the one hand, and a daily increase of  $\exp^{\hat{\nu}_0} - 1 \approx 3.36\%$  in periods of the up-turning regime on the other. Whereas the COVID-19 infections double roughly every  $\hat{\nu}_0^{-1} \log(2) \approx 21$  days in an up-turning state, approximately  $(\hat{\nu}_0 + \hat{\nu}_1)^{-1} \log(\frac{1}{2}) \approx 46$  days are required to halve these case numbers again in a subsequent down-turning regime.

Executing the Kim filter and smoothing recursions yields estimates for the state vector. Figure 3 displays the smoothed trend  $\hat{\mu}_t$  and regime probabilities  $\hat{\Pr}(S_t = 0 | y_T, \dots, y_1)$ , together with the noisy measurement of log COVID-19 infections.



**Figure 3.** Smoothed trend estimates  $\hat{\mu}_t$  (orange, left scale), smoothed regime probabilities  $\hat{\Pr}(S_t = 0|y_T, \dots, y_1)$  (blue, right scale), and log COVID-19 cases  $\log(i_t)$  (gray, left scale). The smoothed trend is a probability-weighted average of the two regime-specific trend estimates. Infection waves (up-turning regime) as identified by  $\hat{\Pr}(S_t = 0|y_T, \dots, y_1) > 40\%$  are shaded.

As Figure 3 shows, the model allows for the identification of episodes of up- and down-turning regimes. In particular, it assigns a strong path dependence to the infection regimes, such that long episodes of containment are followed by new infection waves. The increasing proportion of seasonal distortions is attributed to the seasonal and short-run component, and thus does not deteriorate the smoothed trend nor the smoothed regime probabilities. It is only towards the end of the observational horizon that lateral dynamics, strong seasonal fluctuations, and the limited number of future observations complicate the separation of trend, seasonal, and cyclical dynamics. As a consequence, the smoothed trend and regime probabilities appear more erratic.

To illustrate the benefits from our UC model for identifying structural changes ex post, we select a simple threshold of the smoothed regime probability  $\hat{\Pr}(S_t = 0|y_T, \dots, y_1) > 40\%$  in order to indicate up-turning states. While we leave the choice of a policy rule up to the experts in the field, our arbitrarily chosen threshold is rather cautious (i.e., the probability threshold is set below 1/2) and allows to identify several infectious waves that are shaded in Figure 3. A chosen list of policy measures and events, as reported by the U.S. Centers for Disease Control and Prevention in proximity of the six identified infection waves, is presented below.<sup>6</sup>

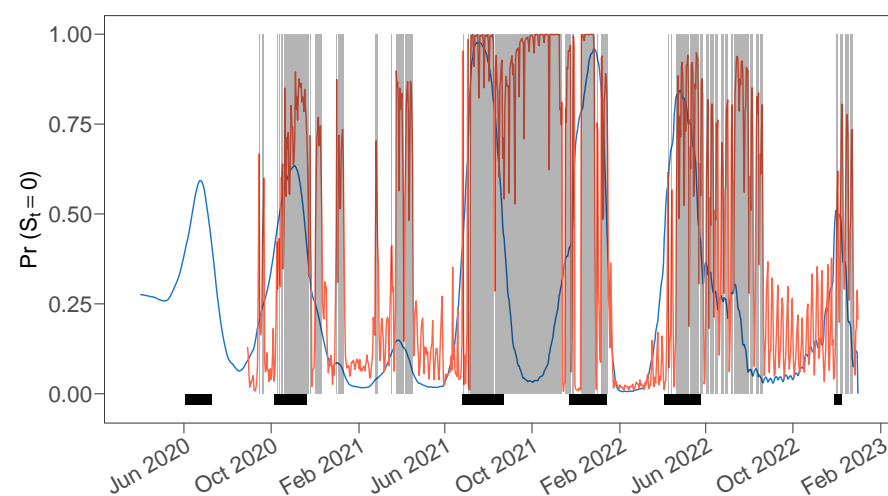
1. **3 June 2020–10 July 2020** 10.04.: The U.S. is the country with the most reported COVID-19 cases and deaths worldwide. 13.04.: Most states in the U.S. report widespread cases of COVID-19. 13.06.: CDC releases consolidated COVID-19 testing guidelines. 22.06.: The U.S. President extends the temporary suspension on new immigrant visas through the end of the year. 30.06.: Dr. Anthony Fauci warns of new infections overwhelming the healthcare system. 01.07.: The U.S. has more than 50 K new daily COVID-19 cases. 14.07.: The CDC again calls on all people to wear cloth face masks when leaving their homes.
2. **6 October 2020–20 November 2020** 04.11.: New U.S. COVID-19 cases surpass 100 K in a day. 10.11.: Total cases of COVID-19 in the U.S. surge past 10 M. 13.11.: COVID-19 case numbers spike across the U.S. after Halloween celebrations. 20.11.: The CDC recommends to stay home for Thanksgiving and to avoid contact as case numbers surge.
3. **26 June 2021–23 August 2021** 27.07.: Amid a Delta variant surge, the CDC releases updated guidance recommending that everyone in areas with high transmission wears a mask. 30.07.: The CDC releases data suggesting that vaccinated people

- infected with Delta can transmit the virus to others. 23.08.: The FDA fully approves the Pfizer–BioNTech COVID-19 vaccine for all people ages 18 years and older.
4. **22 November 2021–14 January 2022** 26.11.: The WHO designates the COVID-19 Omicron variant as a “variant of concern”. 20.12.: The CDC releases data estimating that the Omicron variant is around 1.6 times more transmissible than the Delta variant. 27.12.: The CDC shortens the recommended isolation period for people with COVID-19 to six days.
  5. **4 April 2022–25 May 2022** 13.04.: The Omicron subvariant BA.2 makes up more than 85% of all new COVID-19 infections in the U.S. 18.04.: The CDC’s mask mandate for indoor public transportation is struck down in court. 21.04.: The DHS extends the COVID-19 vaccine requirement for all noncitizens entering the U.S.
  6. **28 November 2022–8 December 2022** 08.12.: The FDA authorizes bivalent COVID-19 vaccines for children as young as 6 months of age. 15.12.: The Biden administration announces the COVID-19 Winter Preparedness plan.

In addition to an ex post analysis, we also propose the use of the prediction step of the Kim filter as an easy real-time monitoring device for the pandemic. In case the one-step-ahead prediction of entering the up-turning regime exceeds the aforementioned threshold of 40%, policies may be triggered to quickly shorten the duration of the incoming up-turning regime.

In the following, we evaluate the appropriateness of this monitoring application based on its past propensity for type I (false-positive) and type II (false-negative) errors, as well as on the time lag of filtered predictions relative to the smoothed estimates. To mimic the situation of a decision maker in real time, we start with 150 observations and estimate the parameters  $\theta$ , as described at the beginning of this section. We then obtain one-step-ahead predictions for the conditional probability for the up-turning state from the Kim filter. Next, an additional observation is added to the sample, and the procedure repeats. To speed up the procedure, we update the parameter estimates only every two iterations and use the estimates from the previous iteration as starting values for the numerical optimization. Every 500 iterations, an additional grid search, as described at the beginning of this section, is performed to robustify the procedure.

Figure 4 compares the smoothed estimates and the filtered one-step-ahead regime probabilities. Time periods where the one-step-ahead prediction exceeds the 40% threshold are shaded.



**Figure 4.** Smoothed regime probabilities  $\hat{\Pr}(S_t = 0|y_T, \dots, y_1)$  (blue) and filtered regime probabilities  $\hat{\Pr}(S_t = 0|y_{t-1}, \dots, y_1)$  (orange) for the up-turning regime. Days on which the filtered predictions  $\hat{\Pr}(S_t = 0|y_{t-1}, \dots, y_1)$  exceed a threshold of 40% are shaded. Horizontal black bars denote infection waves based on the smoothed estimates from Figure 3.

As expected, the filtered regime probabilities are significantly more jagged and slightly lag the smoothed estimates due to their smaller information set. Nonetheless, all of the previously established up-turning periods that fall within the scope of the nowcasting exercise are identified by the monitoring device. A feature that underpins the rather conservative nature of the monitoring device is that it detects the onset of an up-turning regime almost in real time. However, this comes at the cost of a higher type I error, i.e., the detection of an up-turning regime that contradicts the smoothed estimates, e.g., at the end of 2021. During the summer of 2022, the monitoring device detected several additional up-turning regimes that are classified as down-turning ones based on the smoothed regime probabilities. However, as Figure 3 indicates, the smoothed trend is actually increasing during that period; however, the magnitude of the increase is small compared with other up-turning periods.

Table 2 contrasts the infection waves, as identified by the smoothed regime probabilities (shaded periods in Figure 3), with those derived from the filtered one-step-ahead predictions (shaded periods in Figure 4).

**Table 2.** Periods of the up-turning  $S_t = 0$  regime as identified by smoothed estimates and corresponding nowcasting one-step-ahead predictions.

Infection Wave	Smoothed Estimates		Filtered One-Step-Ahead Predictions	
	Beginning	End	Beginning	End
1	3 June 2020	10 July 2020	-	-
2	6 Oct 2020	20 Nov 2020	9 Oct 2020	10 Oct 2020
			12 Oct 2020	13 Oct 2020
			15 Oct 2020	17 Oct 2020
			19 Oct 2020	22 Nov 2020
3	26 June 2021	23 Aug 2021	24 Nov 2020	25 Nov 2020
			26 June 2021	27 June 2021
4	22 Nov 2021	14 Jan 2022	3 July 2021	4 July 2021
			7 July 2021	9 Aug 2021
			17 Nov 2021	24 Nov 2021
			27 Nov 2021	29 Nov 2021
5	04 Apr 2022	25 May 2022	9 Dec 2021	28 Dec 2021
			31 Dec 2021	1 Jan 2022
			06 Jan 2021	15 Jan 2022
			9 Apr 2022	10 Apr 2022
			9 Apr 2022	10 Apr 2022
6	28 Nov 2022	8 Dec 2022	14 Apr 2022	15 Apr 2022
			21 Apr 2022	8 May 2022
			10 May 2022	22 May 2022
			24 May 2022	28 May 2022
6	28 Nov 2022	8 Dec 2022	1 Dec 2022	3 Dec 2022
			6 Dec 2022	10 Dec 2022

Notes: False-positive periods of the up-turning regime are omitted (see Figure 4). The first infection wave is not covered due to an initialization period of 150 days for the nowcasting application.

Across the whole covered period, the nowcasting tool lags only slightly behind the smoothed estimates. These findings suggest the suitability of the model in Equations (1)–(5) for a real-time monitoring tool to provide an intuitive and concise state of the pandemic.

#### 4. Robustness and Discussion

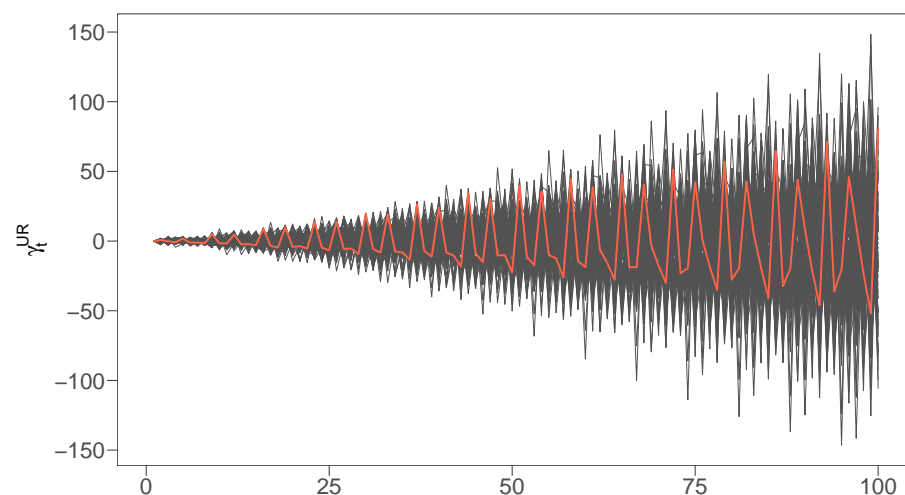
The daily observations of COVID-19 case numbers in Figure 1 display an increasing seasonal variation over time. The deterministic seasonality of the *D.Seas.C.* model in Equation (2) is not equipped to handle such dynamic behavior. As a result, some of the increasing seasonality may be captured by the estimated cyclical and trend component, which



subsequently distorts the regime-switching behavior of the system. A straightforward extension of the model would therefore be to substitute (2) with a stochastic specification capable of assuming seasonal behavior with a progressively increasing variance. One process with such properties is a seasonal unit root (see Bauer and Wagner 2012 for details). It can be added to the state-space model by specifying the new seasonal component

$$\gamma_t^{UR} = - \sum_{j=1}^6 \gamma_{t-j}^{UR} + x_t, \quad (1 - L^7)x_t = \omega_t, \quad \omega_t \sim i.i.d.N(0, \sigma_\omega^2). \quad (6)$$

Similar to the traditional UC literature,  $\gamma_t^{UR}$  is set to be a stochastic process with expected average equal to zero (see e.g., Durbin and Koopman 2012, ch. 3.2.2). However, contrary to the deterministic seasonality in (2), the stochastic specification (6) introduces nonstationary dynamics at lag seven. By plugging in  $x_t$  and rearranging, (6) can be expressed as  $\sum_{j=0}^6 \gamma_{t-j}^{UR} = \sum_{j=7}^{13} \gamma_{t-j}^{UR} + \omega_t$ , and thus  $(1 - L^7) \sum_{j=0}^6 \gamma_{t-j}^{UR} = \omega_t$ . Therefore, the unit root at lag seven links the current seasonal dynamics to those of the previous week, which generates a repeating seasonal pattern. In each period, a shock is added to  $\sum_{j=0}^6 \gamma_{t-j}^{UR}$ , which yields an ever-increasing volatility over time. Figure 5 sketches 100 simulations of  $\gamma_t^{UR}$  and illustrates the oscillating property and nonstationary nature of this seasonal component. Note that (6) differs from unit root seasonal components in the spirit of Harrison and Stevens (1976), which are rather to allow for time-variant seasonal dummies instead of generating an oscillating, diverging behavior.



**Figure 5.** One hundred simulated paths of the seasonal component  $\gamma_t^{UR}$ , as given in (6) (gray). The orange sample path depicts a single exemplary trajectory.

Table 3 presents the estimation results for substituting the deterministic seasonal term (2) with the unit root specification (6), where the parameter estimates are obtained as before. Furthermore, we provide estimates for specifications that exclude the cyclical component (3) in favor of an unsystematic error term  $\epsilon_t$  for both the deterministic and stochastic seasonality, respectively.

Another point of concern is the lateral behavior of infection numbers towards the end of the observational horizon. Ambiguous time periods that cannot be fully attributed to either the up-turning or the down-turning regime may lead to incoherent regime probabilities and thus bias the regime estimates at an earlier stage of the pandemic. As a potential remedy, we introduce a specification involving a third state  $S_t = 2$ , during which the drift term  $\nu_t$  of the trend component in (4) is set to zero. Thus, the third state is intended to reflect periods that feature neither a strong upward nor downward trend in the reported case numbers. Table 3 provides estimates of the proposed extensions together with the preferred *D.Seas.C.* specification displayed in Section 3.

**Table 3.** Maximum likelihood parameter estimates for different model specifications.

	<i>D.Seas.C.</i>	<i>D.Seas.</i>	<i>UR.Seas.C.</i>	<i>UR.Seas.</i>	<i>D.Seas. 3 St.</i>
$\sigma_{\xi}$	0.073 (0.008)	0.081 (0.010)	0.075 (0.006)	0.075 (0.006)	0.042 (0.012)
$\sigma_{\omega}$	-	-	0.005 (0.001)	0.005 (0.001)	-
$\sigma_{\varepsilon}$	-	0.445 (0.011)	-	0.188 (0.006)	-
$\sigma_{\eta}$	0.409 (0.010)	-	0.188 (0.006)	-	0.322 (0.008)
$\nu_0$	0.033 (0.004)	0.034 (0.004)	0.004 (0.003)	0.042 (0.004)	0.035 (0.006)
$\nu_1$	-0.048 (0.010)	-0.047 (0.012)	-0.055 (0.013)	-0.055 (0.013)	-0.257 (0.027)
$\phi_1$	0.440 (0.033)	-	0.007 (0.767)	-	0.272 (0.038)
$\phi_2$	-0.270 (0.032)	-	0 (0.009)	-	-0.185 (0.033)
$q, P_{00}$	0.969 (0.017)	0.971 (0.017)	0.972 (0.014)	0.973 (0.013)	0.900 (0.001)
$P_{01}$	-	-	-	-	0.092 (0.019)
$P_{10}$	-	-	-	-	0 (0.001)
$p, P_{11}$	0.988 (0.010)	0.990 (0.008)	0.976 (0.007)	0.991 (0.007)	0.947 (0.019)
$P_{20}$	-	-	-	-	0.018 (0.010)
$P_{21}$	-	-	-	-	0.007 (0.003)
<b>Log L</b>	-677.783	-768.670	-157.759	-161.242	-662.714
<b>AIC</b>	1.379	1.552	0.348	0.347	1.376
<b>BIC</b>	1.452	1.605	0.431	0.410	1.518
<b>HQ</b>	1.407	1.572	0.379	0.371	1.430

*Notes:* The maximum likelihood estimates are derived via the grid search proposed in Section 3. Standard errors are reported in parenthesis. The columns refer to model specifications with deterministic seasonality including (*D.Seas.C.*), as well as excluding (*D.Seas.*), the AR(2) cyclical component (3), a model accounting for a seasonal unit root process with (*UR.Seas.C.*) and without (*UR.Seas.*), the cyclical component, and a specification with a third state reflecting time periods of neither falling nor rising infection numbers (*D.Seas. 3 St.*).

All information criteria clearly favor the more general specifications that include a seasonal unit root process over the *D.Seas.C.* model presented in Section 3. This is due to the seasonal unit root component, which better grasps the oscillating seasonal pattern along with its increasing volatility over time, as compared with the deterministic seasonal specification. The latter attributes the strong short-run fluctuations to the cyclical component. Since  $c_t$  cannot adequately capture the strong seasonal patterns due to its stationary nature, the smoothed cyclical shocks appear to be autocorrelated, which reduces the likelihood and thus increases the information criteria.<sup>7</sup> However, even though the observed series can be fitted more accurately overall when the seasonal component is generalized to include a seasonal unit root, it is striking that the trend and regime-switching characteristics vary little across specifications: as illustrated by Figure 6, the smoothed regime probabilities of the different models almost coincide, and only minimal differences

between the smoothed trends can be spotted. Therefore, smoothed trend and regime probabilities appear robust to the specification of seasonal and short-run components.

The introduction of a third state, reflecting episodes of stable case numbers, is able to improve upon the fit of the two-state *D.Seas.C.* model, but this refinement is not sufficient enough as to be favored by the BIC and HQ information criteria. In addition, the constraint  $q > 90\%$  is binding, indicating that the model is not suitable to depict the regime path persistence that is inherent to the pandemic.<sup>8</sup> Consequently, we choose the simpler *D.Seas.C.* model for the remainder of our analysis.

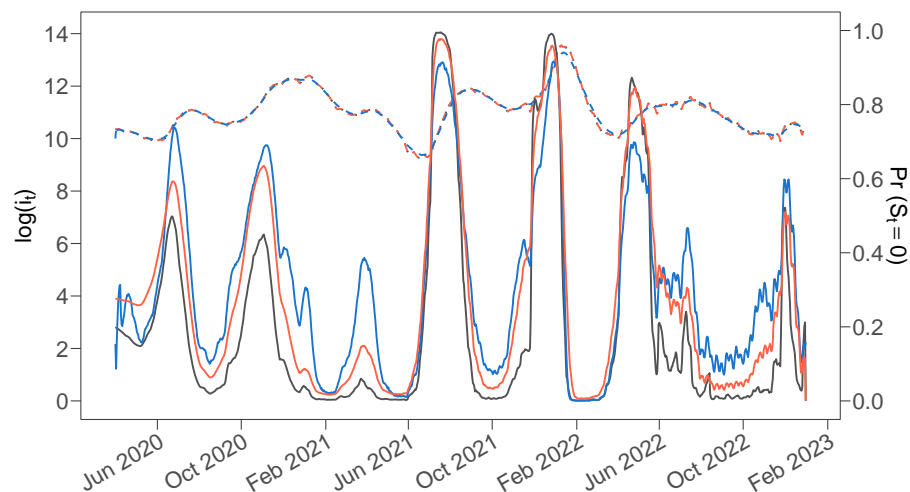
Additional consideration is given to the general estimation strategy. Numerical optimization of a likelihood function, as employed by the Kim filter, has been traditionally regarded as the standard technique for state-space models (Durbin and Koopman 2012, ch. 7). However, Bayesian approaches, especially MCMC techniques, are also a well-developed field in the literature (Frühwirth-Schnatter 2006; Kim and Kang 2019; Luginbuhl and de Vos 1999). In the Bayesian estimation of UC models, the state vector and additional parameters are drawn from a joint distribution, rather than treating the unknown parameters as fixed when maximizing the likelihood numerically. As a consequence, the Bayesian approach bears the advantage that the posteriors incorporate any underlying uncertainty regarding the additional parameters (Luginbuhl and de Vos 1999). Moreover, in order for the Kim filter to account for all possible regime permutations,  $M^2$  individual predictions need to be produced in a single time period for a model with  $M$  states. Since the path dependencies grow at a rate of  $O(M^{2T})$ , the Kim filter exploits regime probabilities to collapse  $M^2$  state vector posteriors into  $M$  posteriors at the end of each time period in order to remain computationally feasible (Kim and Nelson 2017, ch. 5.2). Although previous studies have shown that this approximation entails only a small effect on the final estimates, some bias may be introduced by using weighted averages to collapse the posteriors (Kim and Kang 2019). The Bayesian approach, on the other hand, relies on draws from the joint distribution of the state vector, regime probabilities, and additional parameters. The need for a similar approximation step in an effort to track all conditional distributions is eliminated. Inference only requires the convergence of the Markov chain (Frühwirth-Schnatter 2006, ch. 13).

To robustify our findings, we therefore also estimate the preferred *D.Seas.C.* model using a Gibbs sampling approach. Table 4 portrays the results. Furthermore, Figure 6 includes the averaged trend and regime probability posterior draws.

**Table 4.** Gibbs sampling posterior draws for the preferred *D.Seas.C.* model specification.

	Mean	Std.Dev	Median	2.5%	97.5%
$\sigma_{\xi}$	0.033	0.009	0.031	0.023	0.060
$\sigma_{\eta}$	0.424	0.010	0.424	0.404	0.445
$\nu_0$	0.065	0.037	0.054	0.025	0.174
$\nu_1$	−0.089	0.030	−0.080	−0.178	−0.058
$\phi_1$	0.433	0.035	0.433	0.364	0.500
$\phi_2$	−0.254	0.034	−0.254	−0.319	−0.186
$q$	0.902	0.002	0.902	0.900	0.908
$p$	0.911	0.015	0.904	0.900	0.955

*Notes:* The estimation result refers to the preferred model in Equations (1)–(5). The inference is based on 30,000 draws after an initial burn-in period of 5000 iterations. For robustness, we only store and evaluate every third draw. As in the frequentist estimation technique, the transition probabilities are constrained to be  $>90\%$ , and  $\nu_1$  is bounded to be negative. Apart from the transition probabilities and the trend estimate, all priors are set diffusely. The trend is initialized with the number of COVID-19 infections one day prior to the start of the analysis. The initial transition probabilities are set to 0.95. Splitting the final chain of stored parameter draws in the middle yields a Gelman–Rubin criterion of 1.014, indicating convergence.



**Figure 6.** Smoothed trend  $\hat{\mu}$  (dashed, left scale) and regime probability  $\hat{\Pr}(S_t = 0|y_T, \dots, y_1)$  (right scale) estimates for the preferred *D.Seas.C.* specification (orange, see Table 1), as well as for the seasonal unit root *UR.Seas.* model (grey, see Table 3) using the Kim filter. Averaged posterior draws of the trend and regime probabilities for the *D.Seas.C.* model as derived by the Gibbs sampler are shown in blue (see Table 4).

Comparing the Kim filter estimates in Table 1 with the Bayesian inference in Table 4, the cyclical dynamics almost perfectly match. However, when analyzing the trend component, the Gibbs sampler provides a lower estimate of the innovation variance  $\sigma_\zeta^2$  while, in parallel, specifying the regimes as less persistent and more dissimilar. Therefore, more variance of the data-generating process is attributed to the regime-switching behavior, driving more frequent regime switches and a greater regime-dependent effect. Nevertheless, even with more volatile regime probability estimates, inference based on the Gibbs sampler exhibits only marginal differences, as Figure 6 reveals.

Regime-switching models are notoriously difficult to estimate. In particular, inference regarding the regime transitions is often based on only a small number of observed switches; thus, it presents a substantial challenge to the practitioner. Therefore, it is a sensible exercise to evaluate and test the performance of the specification at hand via a Monte Carlo simulation study. We simulate a sample path over 1000 time periods for 1000 iterations, where the parameter values as well as the length of the data-generating process are chosen to mimic the reported number of COVID-19 infections. The results of the Monte Carlo study are shown in Table 5.

**Table 5.** Result from 1000 Monte Carlo simulations.

<b>T = 1000</b>	<b>Mean</b>	<b>Std.Dev</b>	<b>Median</b>	<b>95% CI</b>
$\sigma_\zeta = 0.050$	0.046	0.025	0.044	[0.045; 0.048]
$\sigma_\eta = 0.500$	0.499	0.013	0.499	[0.499; 0.500]
$\nu_0 = 0.040$	0.037	0.035	0.040	[0.036; 0.039]
$\nu_1 = -0.060$	-0.357	8.903	-0.060	[-0.819; 0.104]
$\phi_1 = 0.500$	0.497	0.036	0.497	[0.495; 0.499]
$\phi_2 = -0.200$	-0.194	0.036	-0.194	[-0.196; -0.192]
$q = 0.970$	0.973	0.018	0.976	[0.972; 0.974]
$p = 0.990$	0.986	0.015	0.990	[0.986; 0.987]

*Notes:* For each of the 1000 iterations, we simulate a sample path of 1000 time periods that corresponds to the *D.Seas.C.* model specification. Confidence intervals are asymptotic. Parameter values are chosen so as to resemble the data-generating process of the COVID-19 pandemic. Estimates are derived via the Kim filter.

Considering the confounding factors that complicate the estimation of regime-switching state-space models, the *D.Seas.C.* specification in (1) to (5) seems to accurately identify the data-generating process. Only the mean of the estimates of the additional drift parameter  $\nu_1$  deviates from its true value, together with a very high standard deviation. The drift parameter  $\nu_1$  materializes only during episodes where the down-turning regime  $S_t = 1$  is turned on (see Equation (4)). As a consequence, in iterations where the randomly generated sample path contains only very few and brief periods of the down-turning regime, any parameter estimate will be imprecise and subject to a high variance. This is also evident from the median of point estimates in Table 5, which is robust against such outlier cases.

Regarding the robustness of the *D.Seas.C.* model, it can be inferred that a sufficient data history, as well as a sufficient exposure to both regimes, are required in order to accurately estimate a regime-switching model. Examining Figure 1, these conditions appear to be satisfied.

## 5. Conclusions

In this article, we employ a regime-switching UC model to decompose log daily COVID-19 infections and estimate the probabilities of alternating regimes of up- and down-turning case numbers over an extensive period, spanning from the 1st of April 2020 to the 25th of December 2022.

Our findings indicate that a regime-switching UC model is capable of capturing many characteristics of the COVID-19 pandemic that more inflexible approaches cannot absorb: a regime-dependent drift assumes persistent long-run dynamics; the weekly patterns of reported COVID-19 infections are modeled by a seasonal component; and a stationary autoregressive component captures short-run dynamics and measurement errors.

The results show that: (i) our approach is well-suited to assess ex post the severity and or efficacy of structural changes, such as viral mutations, regulatory loosening and tightening, behavioral changes, and novel therapeutic approaches ex post; (ii) the model can be applied as a policy tool to monitor the state of the pandemic by nowcasting the current propensity of either the up- or down-turning regime being switched on.

Remaining issues arising from highly inconsistent and erratic seasonality could be overcome by extending the model to allow for fractionally integrated components. In particular, this would allow for gradual adjustments of the trend to structural changes (see e.g., Hartl and Jucknewitz 2022). Another possible extension could be the use of time-inhomogeneous transition probabilities, as proposed in Kaufmann (2015). There, an exogenous variable drives the transition matrix over time, allowing for changes in the frequency of regime switches over the course of the pandemic. However, identifying a candidate exogenous variable is difficult and remains an open question for future research.

**Author Contributions:** Conceptualization, P.H. and T.H.; methodology, P.H. and T.H.; software, P.H.; validation, T.H.; writing—original draft preparation, P.H. and T.H.; writing—review and editing, P.H. and T.H.; visualization, P.H.; supervision, T.H.; funding acquisition, T.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** The APC was funded by the Institute for Employment Research, 90478 Nuremberg, Germany. Tobias Hartl gratefully acknowledges support through the project 356439312 financed by the German Research Foundation (DFG).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. These data can be found at the Github COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University <https://github.com/CSSEGISandData/COVID-19> (accessed on 14 March 2023).

**Acknowledgments:** We wish to thank the editor and the anonymous referees for helpful comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

AR	Autoregressive
EM	Expectation-Maximization
JH/CSSE	Johns Hopkins University Center for Systems Science and Engineering
RW	Random Walk
UC	Unobserved Components

## Notes

- <sup>1</sup> Harvey (1989) refers to these models as structural time series models. To avoid confusion, the term UC model is used for any state-space model that specifies one or multiple time series as a function of latent components and assigns an interpretation to these components by imposing assumptions on their spectra.
- <sup>2</sup> Carvalho et al. (2021) and the U.S. Centers for Disease Control and Prevention (<https://www.cdc.gov/museum/timeline/covid19.html>, accessed on 14 March 2023), provide an exhaustive timeline regarding the development of the COVID-19 pandemic. Additional data on the number of performed tests, as well as on a wide range of other indicators, can be found at <https://ourworldindata.org/coronavirus> (accessed on 14 March 2023) (Ritchie et al. 2020).
- <sup>3</sup> It is reasonable to assume that the number of reported COVID-19 cases in the last week of the year is again biased downward due to the holiday period. We therefore omit the last week of 2022 and end the observational horizon of our analysis on the 25th of December.
- <sup>4</sup> Constraining the parameter space to  $p, q > 90\%$  implies a maximum expected value of three regime switches per 30 days. This is a reasonable constraint given the realized dynamics of the COVID-19 pandemic and, furthermore, speeds up the parameter optimization.
- <sup>5</sup> Note that we estimate  $\nu_0$  as part of the state vector. However, the state-space representation of the model in (1) to (5) is not uniquely defined. Different, albeit observationally identical approaches, such as, e.g., the estimation of  $\nu_0$  via ML or constraining  $\nu_1$  to be  $>0$  and flipping labels with  $\nu_0$ , are possible.
- <sup>6</sup> A more detailed and extensive overview can be seen at <https://www.cdc.gov/museum/timeline/covid19.html> (accessed on 14 March 2023) as well as <https://www.defense.gov/Spotlights/Coronavirus-DOD-Response/Timeline/> (accessed on 14 March 2023).
- <sup>7</sup> Another way to get rid of the strong seasonal pattern would be to take seven-day averages of the log case numbers before estimating the model, which would yield a smooth series and eliminate the seasonal variation.
- <sup>8</sup> Since the focus of the analysis lies on identifying coherent periods of up- or down-turning infection regimes, we do not constrain  $P_{22}$  to be  $>90\%$ .

## References

- Bauer, Dietmar, and Martin Wagner. 2012. A state space canonical form for unit root processes. *Econometric Theory* 28: 1313–49. [[CrossRef](#)]
- Bergman, Aviv, Yehonatan Sella, Peter Agre, and Arturo Casadevall. 2020. Oscillations in U.S. COVID-19 incidence and mortality data reflect diagnostic and reporting factors. *mSystems* 5: e00544–20. [[CrossRef](#)]
- Carvalho, Thiago, Florian Krammer, and Akiko Iwasaki. 2021. The first 12 months of COVID-19: A timeline of immunological insights. *Nature Reviews Immunology* 21: 245–56. [[CrossRef](#)] [[PubMed](#)]
- Degras, David, Chee-Ming Ting, and Hernando Ombao. 2022. Markov-switching state-space models with applications to neuroimaging. *Computational Statistics & Data Analysis* 174: 107525. [[CrossRef](#)]
- Dolton, Peter. 2021. The statistical challenges of modelling COVID-19. *National Institute Economic Review* 257: 46–82. [[CrossRef](#)]
- Dong, Ensheng, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 20: 533–34. [[CrossRef](#)] [[PubMed](#)]
- Doornik, Jurgen A., Jennifer L. Castle, and David F. Hendry. 2021. Modeling and forecasting the COVID-19 pandemic time-series data. *Social Science Quarterly* 102: 2070–87. [[CrossRef](#)]
- Doornik, Jurgen A., Jennifer L. Castle, and David F. Hendry. 2022. Short-term forecasting of the Coronavirus pandemic. *International Journal of Forecasting* 38: 453–66. [[CrossRef](#)]
- Durbin, James, and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods: Second Edition*. Oxford: Oxford University Press. [[CrossRef](#)]
- Fiscon, Giulia, Francesco Salvatore, Valerio Guarrasi, Anna Rosa Garbuglia, and Paola Paci. 2021. Assessing the impact of data-driven limitations on tracing and forecasting the outbreak dynamics of COVID-19. *Computers in Biology and Medicine* 135: 104657. [[CrossRef](#)]
- Frühwirth-Schnatter, Sylvia. 2006. *Finite Mixture and Markov Switching Models*. New York: Springer. [[CrossRef](#)]

- Hamilton, James D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57: 357–84. [CrossRef]
- Harrison, Peter J., and C. F. Stevens. 1976. Bayesian forecasting. *Journal of the Royal Statistical Society: Series B* 38: 205–28. [CrossRef]
- Hartl, Tobias, and Roland Jucknewitz. 2022. Approximate state space modelling of unobserved fractional components. *Econometric Reviews* 41: 75–98. [CrossRef]
- Hartl, Tobias, Klaus Wälde, and Enzo Weber. 2020. Measuring the impact of the German public shutdown on the spread of COVID-19. *Covid Economics: Vetted and Real-Time Papers* 1: 25–32.
- Harvey, Andrew C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Hortaçsu, Ali, Jiarui Liu, and Timothy Schwieg. 2021. Estimating the fraction of unreported infections in epidemics with a known epicenter: An application to COVID-19. *Journal of Econometrics* 220: 106–29. [CrossRef] [PubMed]
- Kaufmann, Sylvia. 2015. K-state switching models with time-varying transition distributions—Does loan growth signal stronger effects of variables on inflation? *Journal of Econometrics* 187: 82–94. [CrossRef]
- Kim, Chang-Jin. 1994. Dynamic linear models with Markov-switching. *Journal of Econometrics* 60: 1–22. [CrossRef]
- Kim, Chang-Jin, and Charles R. Nelson. 2017. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge: MIT Press. [CrossRef]
- Kim, Young Min, and Kyu Ho Kang. 2019. Likelihood inference for dynamic linear models with Markov switching parameters: On the efficiency of the Kim filter. *Econometric Reviews* 38: 1109–30. [CrossRef]
- Lee, Sokbae, Yuan Liao, Myung Hwan Seo, and Youngki Shin. 2021. Sparse HP filter: Finding kinks in the COVID-19 contact rate. *Journal of Econometrics* 220: 158–80. [CrossRef] [PubMed]
- Liu, Laura, Hyungsik Roger Moon, and Frank Schorfheide. 2021. Panel forecasts of country-level COVID-19 infections. *Journal of Econometrics* 220: 2–22. [CrossRef] [PubMed]
- Luginbuhl, Rob, and Aart de Vos. 1999. Bayesian analysis of an unobserved-component time series model of GDP with Markov-switching and time-varying growths. *Journal of Business & Economic Statistics* 17: 456–65. [CrossRef]
- Moosa, Imad A. 2020. The effectiveness of social distancing in containing COVID-19. *Applied Economics* 52: 6292–05. [CrossRef]
- Navas Thorakkattle, Muhammed, Shazia Farhin, and Athar Ali Khan. 2022. Forecasting the trends of COVID-19 and causal impact of vaccines using Bayesian structural time series and ARIMA. *Annals of Data Science* 9: 1025–47. [CrossRef]
- Ritchie, Hannah, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, and Max Roser. 2020. Coronavirus pandemic (COVID-19). *Our World in Data*. Available online: <https://ourworldindata.org/coronavirus> (accessed on 14 March 2023).
- Telenti, Amalio, Ann Arvin, Lawrence Corey, Davide Corti, Michael S. Diamond, Adolfo García-Sastre, Robert F. Garry, Edward C. Holmes, Phillip S. Pang, and Herbert W. Virgin. 2021. After the pandemic: Perspectives on the future trajectory of COVID-19. *Nature* 596: 495–504. [CrossRef]
- Xie, Liming. 2022. The analysis and forecasting COVID-19 cases in the United States using Bayesian structural time series models. *Biostatistics & Epidemiology* 6: 1–15. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.