

Monitoring the pandemic: A fractional filter for the COVID-19 contact rate

Tobias Hartl^{*a,b}

^aUniversity of Regensburg, 93053 Regensburg, Germany

^bInstitute for Employment Research, 90478 Nuremberg, Germany

February 2021

Abstract. This paper aims to provide reliable estimates for the COVID-19 contact rate of a Susceptible-Infected-Recovered (SIR) model. From observable data on confirmed, recovered, and deceased cases, a noisy measurement for the contact rate can be constructed. To filter out measurement errors and seasonality, a novel unobserved components (UC) model is set up. It specifies the log contact rate as a latent, fractionally integrated process of unknown integration order. The fractional specification reflects key characteristics of aggregate social behavior such as strong persistence and gradual adjustments to new information. A computationally simple modification of the Kalman filter is introduced and is termed the fractional filter. It allows to estimate UC models with richer long-run dynamics, and provides a closed-form expression for the prediction error of UC models. Based on the latter, a conditional-sum-of-squares (CSS) estimator for the model parameters is set up that is shown to be consistent and asymptotically normally distributed. The resulting contact rate estimates for several countries are well in line with the chronology of the pandemic, and allow to identify different contact regimes generated by policy interventions. As the fractional filter is shown to provide precise contact rate estimates at the end of the sample, it bears great potential for monitoring the pandemic in real time.

Keywords. COVID-19, filtering, long memory, SIR model, unobserved components.

JEL-Classification. C22, C51, C52.

*Corresponding author: Department of Economics and Econometrics, University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany, email: tobias1.hartl@ur.de

1 Introduction

Since the outbreak of COVID-19 reducing social contacts is widely viewed as the key way to contain the spread of the virus. In terms of the Susceptible-Infected-Recovered (SIR) model¹, this relates to the contact rate, defined as the average number of contacts per person per time unit multiplied by the probability of disease transmission between a susceptible and an infected individual (Hethcote; 2000). The probability of disease transmission should only depend on characteristics that are specific to the virus. Therefore, the contact rate can be interpreted as a proxy for aggregate social behavior and is the key variable addressed by social distancing measures. Knowing the trajectory of the contact rate would allow to draw inference on the impact of policy measures on contact reduction, to real-time monitor the dynamics of virus dispersion, and to design policy rules based on the current pandemic situation. Since the contact rate itself is unobservable, appropriate methods to estimate the contact rate are required, and will be considered in this paper.

At the early stage of the pandemic, first estimates for the natural logarithm of the contact rate were obtained by fitting a deterministic, linear time trend with structural breaks to transformations of data on confirmed, recovered, and deceased cases (Hartl, Wälde and Weber; 2020; Lee et al.; 2021; Liu et al.; 2021). Modeling the log contact rate by a piece-wise linear time trend was a reasonable and pragmatic approximation given the short time series on case numbers available at that time. However, it implies that contact rate growth evolves deterministically as a straight line with jumps at the break dates. This assumption is likely to be violated by the behavior of individuals. While structural breaks may be suitable to identify turning points of the contact rate, they are inappropriate for monitoring the current pandemic situation, as breaks require at least some post-break observations to be well identified.

This paper aims to improve estimates for the contact rate of COVID-19 by taking into account key features of aggregate social behavior. In detail, the log contact rate, as denoted by $\log \beta_t$, is modeled as an unobserved, fractionally integrated process of (unknown) order $d \in \mathbb{R}_+$, generated by stochastic shocks $\{\eta_i\}_{i=1}^t$.² The stochastic specification of the contact rate is motivated by the consideration that social decisions, e.g. on whether to meet, are made conditional on the information available at that time, e.g. on current social distancing measures or the state of the pandemic. As information does not evolve deterministically but appears as stochastic shocks, this suggests to treat $\log \beta_t$ as a stochastic process generated by the information shocks $\{\eta_i\}_{i=1}^t$. Specifying $\log \beta_t$ as a fractionally integrated process accounts for strong persistence and nonstationarity (in short: long memory) of social behavior. In contrast to structural breaks but also to random walks, the fractional specification allows social behavior to gradually adjust to new

¹The SIR model – in its various variants – has recently become a popular tool to study the economic impact of the pandemic and for policy simulations, see Acemoglu et al. (2020); Avery et al. (2020); Korolev (2021); Liu et al. (2021) among others.

²Fractional integration techniques have been found useful for describing the aggregate behavior of individuals in a variety of applications, e.g. for explaining the Deaton paradox (Diebold and Rudebusch; 1991) and for the estimation of the business cycle (Hartl, Tschernig and Weber; 2020).

information both at the individual and at the aggregate level. Individually, this reflects a gradual reduction or increase of contacts as new information becomes available (e.g. as new contact restrictions are imposed), while on aggregate it allows individuals to react heterogeneously both in terms of speed and intensity to novel information. As the persistence of the log contact rate is unknown, the integration order d is treated as an unknown parameter to be estimated.

Methodologically, this paper contributes to the literature on time series filtering by setting up a novel unobserved components (UC) model that does not require prior knowledge about the integration order of the variable under study. Current UC models and related filtering techniques rely heavily on prior assumptions about the integration order d and typically assume $d = 1$ (e.g. Harvey; 1985; Morley et al.; 2003; Chang et al.; 2009) or $d = 2$ (e.g. Clark; 1987; Hodrick and Prescott; 1997; Oh et al.; 2008) to be known. In contrast, the novel UC model reflects that the degree of persistence of the log contact rate is unknown. It allows to decompose a noisy measurement for the log contact rate that is based on a transformation of data on confirmed, recovered, and deceased cases, into measurement errors, seasonal components, and the unobserved log contact rate itself. As the latter is modeled by a fractionally integrated process, the model is called the fractional UC model.

The second methodological contribution of this paper is to derive a computationally much simpler estimator for the model parameters and the unobserved components compared to current state space methods. Current methods typically rely on the Kalman filter to set up a conditional (quasi-)likelihood function for the estimation of the model parameters. Given the parameter estimates, a time-varying signal for the unobserved components is then obtained from the Kalman smoother. Both the Kalman filter and smoother become computationally infeasible when the dimension of the state vector of UC models is high, as for fractionally integrated processes. To address this problem, this paper proposes a computationally simple modification of the Kalman filter and smoother that is termed the fractional filter. While filtered and smoothed estimates from the fractional filter are identical to the Kalman filter and smoother, the fractional filter avoids the computationally intensive recursions for the conditional variance. The fractional filter provides a closed-form expression for the prediction error of UC models, based on which a conditional-sum-of-squares (CSS) estimator for the fractional integration order and other model parameters is set up. While the CSS estimator has been found useful for the estimation of ARFIMA models, see Hualde and Robinson (2011) and Nielsen (2015), it has not been considered in the UC literature so far. The CSS estimator minimizes the sum of squared prediction errors that is proportional to the exponent in the conditional (quasi-)likelihood function based on the Kalman filter. Due to the computational gains from the fractional filter, the CSS estimator allows to estimate UC models with richer long-run dynamics. The paper provides the asymptotic theory for the CSS estimator, showing it to be consistent and asymptotically normally distributed, while the finite sample properties are assessed by a Monte Carlo study.

Using data from the Johns Hopkins University Center for Systems Science and Engineering (Dong et al.; 2020, JHU CSSE), estimates for contact and reproduction rate are presented for

Canada, Germany, Italy, and the United States, where benefits from the new methods directly become apparent: First, estimation results are not only well in line with the chronology of the pandemic, but also allow to identify different contact regimes generated by the strengthening and easing of contact restrictions. Second, a recursive window evaluation shows contact rate estimates at the end of a truncated sample to largely overlap with those based on the full sample information. This makes the fractional filter a suitable candidate for monitoring outbreaks at the current frontier of the data. And third, the proposed estimation and filtering techniques are shown to be fairly robust to under-reporting of recovered cases, which is of particular importance for the US, as several states do not report data on recovered individuals. While under-reporting heavily downward-biases contact and reproduction rate estimates in Lee et al. (2021), this is shown not to be the case for the fractional filter.

The remaining paper is organized as follows: Section 2 motivates the specification of the contact rate and sets up the fractional UC model. Section 3 introduces the fractional filter for $\log \beta_t$, covers parameter estimation via the CSS estimator and presents the asymptotic theory. Section 4 contains empirical results for Canada, Germany, Italy, and the United States, while section 5 concludes. The appendices include proofs for consistency and asymptotic normality of the CSS estimator as well as a Monte Carlo study on the finite sample properties.

2 A fractional unobserved components model for the contact rate

To motivate the estimation of the contact rate, consider the discrete SIR model, augmented to include deaths, which also forms the starting point of Pindyck (2020, eqn. 1–4) and Lee et al. (2021, eqn. 2.1)

$$1 = S_t + I_t + D_t + R_t, \quad (1)$$

$$\Delta I_t = \beta_t S_{t-1} I_{t-1} - \gamma I_{t-1}, \quad (2)$$

$$\Delta D_t = \gamma_d I_{t-1}, \quad (3)$$

$$\Delta R_t = \gamma_r I_{t-1}. \quad (4)$$

In (1), the (initial) population size, normalized to be one, is decomposed into S_t , the proportion of the population susceptible in t , I_t , the fraction of the population infected in t , D_t , the fraction that has died until t , and R_t , the proportion that has recovered until t . In (2), $\gamma = \gamma_d + \gamma_r$ denotes the rate at which infected either die, see (3), or recover, see (4), and obviously $\gamma_d, \gamma_r \geq 0$. Thus, γI_{t-1} denotes the fraction of outflows of infected at t . The fraction of new infections at t is captured by $\beta_t S_{t-1} I_{t-1}$, where $S_{t-1} I_{t-1}$ can be interpreted as the average probability of a contact being between a susceptible subject and an infected subject. $\beta_t > 0$ is called the contact (or transmission) rate. It equals the average number of contacts per person per time unit multiplied by the probability of disease transmission between a susceptible and an infectious

person (Hethcote; 2000). As in Lee et al. (2021), the contact rate is allowed to be time-varying. This reflects social behavior to change over time, e.g. in response to policy changes or to novel information on the pandemic. Since the contact rate determines inflows into infected, see (2), it is the key variable tackled by social distancing policies.

Based on the contact rate, the reproduction rate $\mathcal{R}_t = \beta_t/\gamma$ can be derived. It is the average number of infections caused by an infected subject during the infectious period $1/\gamma$ at the early stage of the pandemic (where $S_{t-1} \approx 1$). \mathcal{R}_t is an indicator for the current dynamics of the pandemic, as for $\mathcal{R}_t < 1$ outflows from infected exceed inflows, causing ΔI_t to converge, see (2) where $0 \leq S_{t-1} \leq 1$. Thus, if policy seeks to contain the spread of COVID-19, then it must control the contact rate, which controls the reproduction rate \mathcal{R}_t .

As shown by Lee et al. (2021), from (1) to (4) a measurement for the contact rate β_t can be obtained directly: Denote $C_t = I_t + R_t + D_t$ as the fraction of confirmed cases (consisting of infected, recovered, and deceased cases) and use $\Delta C_t = \Delta I_t + \Delta R_t + \Delta D_t$ together with (2) to (4) to obtain $\Delta C_t = \beta_t S_{t-1} I_{t-1} - \gamma I_{t-1} + (\gamma_d + \gamma_r) I_{t-1} = \beta_t S_{t-1} I_{t-1}$. Solving for β_t yields

$$\beta_t = \frac{\Delta C_t}{I_{t-1} S_{t-1}} =: Y_t, \quad (5)$$

see Lee et al. (2021, eqn. 2.2). As argued there, if for each t the data (C_t, R_t, D_t) can be observed, then the time-varying contact rate can be calculated straightforwardly via (5) using $S_t = 1 - C_t$, as well as $I_t = C_t - R_t - D_t$.

Unfortunately, reported case numbers for C_t , R_t , and D_t , such as the daily data from JHU CSSE used in the applications in section 4, suffer from measurement errors, see e.g. Hortaçsu et al. (2021). In addition, they display a strong weekly seasonal pattern that is likely to be driven by a varying number of tests conducted over the different days of the week (Bergman et al.; 2020). Under the assumption that Y_t is measured with a proportionally constant error variance resulting from seasonality and measurement errors, one has the following structure for the natural logarithm of the observable \tilde{Y}_t .

Assumption 1 (Multiplicative seasonal and measurement errors). *For each t , the observable \tilde{Y}_t satisfies*

$$\log \tilde{Y}_t = \log Y_t + \sum_{i=1}^7 \alpha_i s_{i,t} + u_t = \log \beta_t + \sum_{i=1}^7 \alpha_i s_{i,t} + u_t, \quad t = 1, \dots, n,$$

with Y_t as given in (5). $s_{i,t}$ are seasonal dummies for $i = 1, \dots, 7$, that capture the weekly patterns of reported case numbers, $\sum_{i=1}^7 \alpha_i = 0$, and the measurement error $u_t \sim WN(0, \sigma_u^2)$ is white noise.

Assumption 1 specifies an unobserved components (UC) model where the observable noisy measurement $\log \tilde{Y}_t$ is decomposed into an unobservable measurement error u_t , seasonal components $\sum_{i=1}^7 \alpha_i s_{i,t}$, and the log contact rate $\log \beta_t$. The log specification accounts for a proportional

impact of measurement errors and seasonality, and forces the contact rate to be strictly positive.

As the different components are not separately identified, an additional assumption on the dynamic structure of the contact rate is required. Empirical models of COVID-19 case numbers have so far assumed $\log \beta_t$ to follow a piece-wise linear time trend with structural breaks, see Hartl, Wälde and Weber (2020); Lee et al. (2021); Liu et al. (2021). As an alternative, the UC literature suggests to model time-varying coefficients as random walks (see Durbin and Koopman; 2012, for an overview). Both specifications assume contact rate growth $\Delta \log \beta_t$ only to be contemporaneously affected either by structural breaks or by stochastic shocks, an assumption that is likely to be violated. Reflecting that the persistence properties of social behavior, and thus of the contact rate, are unknown, assumption 2 specifies the log contact rate as a fractionally integrated process of unknown order d .

Assumption 2 (Specification of the contact rate). *The log contact rate follows a type II fractionally integrated process of order $d \in \mathbb{R}_+$, denoted as $\log \beta_t \sim I(d)$, where*

$$\log \beta_t = \mu + x_t, \quad \Delta_+^d x_t = \eta_t, \quad \eta_t \sim WN(0, \sigma_\eta^2), \quad t = 1, \dots, n,$$

μ is an intercept, and the η_t are white noise and are independent of the measurement error u_t .

Under assumption 2, the log contact rate $\log \beta_t$ is a stochastic long memory process generated by the shocks $\{\eta_i\}_{i=1}^t$. The shock η_t models the information new in t , such as news reports or policy announcements. Social decisions, reflected in $\log \beta_t$, however may additionally depend on past information $\eta_{t-1}, \dots, \eta_1$. Together, $\{\eta_i\}_{i=1}^t$ forms the information available at t , conditional on which social decisions, e.g. on whether to meet, are made. The specification takes into account that new information does not evolve deterministically, but appears as stochastic shocks, which cannot be captured by a deterministic specification as e.g. in Lee et al. (2021).

The degree of persistence of the log contact rate is determined by the integration order d , which controls for the persistent impact of past shocks via the fractional difference operator Δ_+^d . The latter exhibits a polynomial expansion in the lag operator L of order infinite

$$\Delta^d = (1 - L)^d = \sum_{i=0}^{\infty} \pi_i(d) L^i, \quad \pi_i(d) = \begin{cases} \frac{i-d-1}{i} \pi_{i-1}(d) & i = 1, 2, \dots, \\ 1 & i = 0. \end{cases} \quad (6)$$

The $+$ -subscript denotes a truncation of an operator at $t \leq 0$, $\Delta_+^d x_t = \sum_{i=0}^{t-1} \pi_i(d) x_{t-i}$, which reflects the type II definition of fractionally integrated processes (Marinucci and Robinson; 1999). For $d = 1$ the log contact rate is a random walk, which follows from plugging $d = 1$ into (6). Consequently, assumption 2 encompasses the predominant specification in the UC literature. However, assumption 2 allows for a far more general dynamic impact of past shocks η_1, \dots, η_t on

$\log \beta_t$, as can be seen by plugging $x_t = \Delta_+^{-d} \eta_t$ into $\log \beta_t = \mu + x_t$, which gives

$$\log \beta_t = \mu + \Delta_+^{-d} \eta_t = \mu + \sum_{i=0}^{t-1} \pi_i(-d) \eta_{t-i}. \quad (7)$$

While a random walk is an unweighted sum of past shocks η_1, \dots, η_t , so that $\pi_i(-1) = 1$ for all $i = 1, \dots, t-1$, allowing $d \neq 1$ yields non-uniform weights of past shocks in the impulse response function of $\log \beta_t$ and thus a gradual adjustment of the log contact rate to new information. This reflects that social behavior adjusts step-wise to new information both at the individual and the aggregate level. As processing new information on the Coronavirus and revising individual decisions (e.g. meeting friends, traveling, working from home) takes time and evolves gradually, individuals can be expected to step-wise adjust their contacts in response to new information. Overall, individuals will react heterogeneously both in terms of speed and intensity to novel information: Some will anticipate new information faster than others, and the extent of reaction will depend on individual characteristics such as risk awareness and attitudes. Such gradual adjustments are well captured by assumption 2, in particular when $1 < d < 2$: In that case, contact rate growth $\Delta \log \beta_t \sim I(d-1)$ is strongly persistent and mean-reverting, as will become apparent in the applications in section 4. Strong persistence reflects the gradual adjustment of social behavior to new information, while mean-reversion ensures an asymptotically declining impact of past information to today's contact rate growth.

The remaining assumptions are imposed mainly for technical reasons. The type II definition of fractional integration assumes zero starting values for the fractionally integrated process by truncating the polynomial expansion of the fractional difference operator, $\Delta_+^d x_t = \sum_{i=0}^{t-1} \pi_i(d) x_{t-i}$. It is required to treat the asymptotically stationary ($d < 1/2$, from now on 'stationary' for brevity) and the asymptotically nonstationary case ($d > 1/2$, from now on 'nonstationary') alongside each other. While the type II definition may be a strong assumption for some time series, it is plausible for the contact rate, as we have data covering roughly the whole pandemic. Thus, the pre-sample shocks η_i , $i \leq 0$, should be zero. Independence of u_t and η_t follows from the characterization of u_t as a measurement error that should not influence the contact rate. In general, the assumption can be relaxed to allow for $\text{Corr}(\eta_t, u_t) \neq 0$, as for instance in correlated UC models (Morley et al.; 2003), and will not affect the asymptotic results in section 3. The distributional assumptions on η_t and u_t are somewhat weaker than the assumption of Gaussian white noise on which UC models typically rely (Morley et al.; 2003). They will be shown to be largely satisfied in the applications of section 4. Finally, $d > 0$ is required to separately identify $\log \beta_t$ and u_t .

3 Parameter estimation, filtering and smoothing

In this section, the fractional filter is derived. It is a computationally simple modification of the Kalman filter that avoids the Kalman recursions for the conditional variance. The modification is necessary, as the Kalman filter becomes computationally infeasible for UC models when the

dimension of the state vector is high, as for fractionally integrated processes. The fractional filter provides a closed-form expression for the prediction error of the UC model. Based on that, a conditional-sum-of-squares (CSS) estimator for the model parameters is set up. It minimizes the sum of squared prediction errors obtained from the fractional filter. The CSS estimator is shown to be consistent and asymptotically normally distributed. Given the CSS parameter estimates, the log contact rate can be estimated by the fractional filter given the full sample information. Finally, estimation of the mean and seasonal components is considered.

Under assumptions 1 and 2, the fractional UC model is given by

$$\log \tilde{Y}_t = \log \beta_t + \sum_{i=1}^7 \alpha_i s_{i,t} + u_t, \quad \log \beta_t = \mu + x_t, \quad x_t = \Delta_+^{-d} \eta_t, \quad t = 1, \dots, n. \quad (8)$$

Denote $\mu_0, \alpha_{1,0}, \dots, \alpha_{7,0}, d_0, \sigma_{\eta,0}^2, \sigma_{u,0}^2$ as the true parameters of the data-generating mechanism. Leaving aside the deterministic terms for the moment, by defining $y_t = \log \tilde{Y}_t - \mu - \sum_{i=1}^7 \alpha_i s_{i,t}$, the stochastic part of the fractional UC model (8) is

$$y_t = x_t + u_t, \quad x_t = \Delta_+^{-d} \eta_t, \quad t = 1, \dots, n. \quad (9)$$

In the following, let $\theta = (d, \sigma_\eta^2, \sigma_u^2)' \in \Theta$ denote the vector holding the parameters of (9), and let $\theta_0 = (d_0, \sigma_{\eta,0}^2, \sigma_{u,0}^2)' \in \Theta$, where $\Theta = D \times \Omega_\eta \times \Omega_u$ denotes the parameter space with $D = \{d \in \mathbb{R} | 0 < d \leq d_{max}\}$ and $\Omega_i = \{\sigma_i^2 \in \mathbb{R} | 0 < \sigma_i^2 < \infty\}$, $i = \eta, u$. Define \mathcal{F}_t as the σ -algebra generated by y_1, \dots, y_t , and let the expected value operator $E_\theta(z_t)$ of an arbitrary random variable z_t denote that expectation is taken with respect to the distribution of z_t given θ , so that $E_{\theta_0}(z_t) = E(z_t)$. Furthermore, let $\Sigma^{(i,j)}$ denote the (i, j) -th entry of an arbitrary matrix Σ .

Estimation of the parameters θ_0 is carried out by the CSS estimator that minimizes the sum of squared prediction errors of model (9). The prediction error is defined as the one-step ahead forecast error of y_{t+1} given \mathcal{F}_t

$$v_{t+1}(\theta) = y_{t+1} - E_\theta(y_{t+1} | \mathcal{F}_t) = y_{t+1} - E_\theta(x_{t+1} | \mathcal{F}_t). \quad (10)$$

It depends on $E_\theta(x_{t+1} | \mathcal{F}_t)$, for which the fractional filter provides an analytical solution. The filter is introduced in the following lemma.

Lemma 3.1 (Fractional filter for x_{t+1} given \mathcal{F}_t). *Under assumptions 1 and 2*

$$E_\theta(x_{t+1} | \mathcal{F}_t) = \sum_{i=1}^t \pi_i(-d) \Sigma_{\eta_{t:1} y_{t:1}}^{(i,\cdot)} \Sigma_{y_{t:1}}^{-1} y_{t:1},$$

where $y_{t:1} = (y_t, \dots, y_1)'$, $\eta_{t:1} = (\eta_t, \dots, \eta_1)'$, $\Sigma_{\eta_{t:1} y_{t:1}} = \text{Cov}_\theta(\eta_{t:1}, y_{t:1})$, and $\Sigma_{y_{t:1}} = \text{Var}_\theta(y_{t:1})$. The

superscript in $\Sigma_{\eta_{t:1}y_{t:1}}^{(i,\cdot)}$ denotes the i -th row of the matrix, and

$$\Sigma_{\eta_{t:1}y_{t:1}}^{(i,j)} = \begin{cases} \pi_{i-j}(-d)\sigma_{\eta}^2 & \text{if } i \geq j, \\ 0 & \text{else,} \end{cases} \quad \Sigma_{y_{t:1}}^{(i,j)} = \begin{cases} \sigma_u^2 + \sigma_{\eta}^2 \sum_{k=0}^{t-i} \pi_k^2(-d) & \text{if } i = j, \\ \sigma_{\eta}^2 \sum_{k=0}^{t-\max(i,j)} \pi_k(-d)\pi_{k+|i-j|}(-d) & \text{else.} \end{cases}$$

The proof is contained in appendix C. As can be seen from lemma 3.1, the fractional filter provides a solution for $E_{\theta}(x_{t+1}|\mathcal{F}_t)$ that only depends on θ and y_1, \dots, y_t . By plugging it into (10), one has the closed-form expression for the prediction error

$$v_{t+1}(\theta) = y_{t+1} - \sum_{i=1}^t \pi_i(-d) \Sigma_{\eta_{t:1}y_{t:1}}^{(i,\cdot)} \Sigma_{y_{t:1}}^{-1} y_{t:1}. \quad (11)$$

Based on (11) the objective function of the CSS estimator for θ_0 is set up

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n v_t^2(\theta). \quad (12)$$

Note that estimating the parameters of the fractional UC model via the CSS estimator (12) in combination with the fractional filter deviates from the methodological state space literature: There, expectation and variance of x_{t+1} conditional on \mathcal{F}_t are typically obtained from the Kalman filter recursions (see e.g. Durbin and Koopman; 2012, ch. 4.3). The resulting prediction error and its conditional variance then enter the Gaussian (quasi-)likelihood function that is maximized to estimate θ_0 . However, the Kalman filter becomes computationally infeasible when the dimension of the state vector is high, as for fractionally integrated processes. Thus, a computationally simpler filter is required. The fractional filter, as defined in lemma 3.1, is a modification of the Kalman filter: Its solution for $E_{\theta}(x_{t+1}|\mathcal{F}_t)$ is identical to the Kalman filter (see Durbin and Koopman; 2012, ch. 4.2), but it avoids the Kalman recursions for the conditional variance of x_{t+1} . While the conditional variance is necessary for (quasi-)maximum likelihood estimation, the CSS estimator only requires a closed-form expression for the prediction error, for which the fractional filter is sufficient. The objective function of the CSS estimator in (12) is of course proportional to the exponent in the conditional Gaussian (quasi-)likelihood function. However, CSS estimation is computationally much simpler due to the fractional filter. Together, the fractional filter and the CSS estimator provide a computationally feasible alternative to the Kalman filter and the (quasi-)maximum likelihood estimator, particularly for UC models with richer long-run dynamics.

While the asymptotic theory of the CSS estimator is well established for ARFIMA models, see Hualde and Robinson (2011) and Nielsen (2015), it has not yet been derived for structural UC models. To fill this gap, theorems 3.2 and 3.3 summarize the asymptotic estimation theory for the CSS estimator for fractional UC models. In addition, the finite sample properties are addressed by a Monte Carlo study in appendix B. For consistency and asymptotic normality of the CSS estimator, the moment assumptions on the shocks η_t, u_t need to be strengthened.

Assumption 3 (Higher moments of η_t, u_t). *The conditional moments of η_t, u_t (conditional on past $\eta_{t-1}, \eta_{t-2}, \dots$, and u_{t-1}, u_{t-2}, \dots) are finite up to order four and equal the unconditional moments.*

Theorem 3.2 (Consistency). *Under assumptions 1 to 3 the CSS estimator $\hat{\theta}$ is consistent, $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \rightarrow \infty$.*

The proof of theorem 3.2 is given in appendix C and is carried out as follows: First, the model in (9) is shown to be identified. Next, $v_t(\theta)$, as given in (11), is shown to be integrated of order $d_0 - d$, and thus is stationary for $d_0 - d < 1/2$ and nonstationary for $d_0 - d > 1/2$. As the asymptotic behavior of the objective function changes around the point $d_0 - d = 1/2$, the objective function does not uniformly converge in probability on Θ . Adopting the results of Hualde and Robinson (2011) and Nielsen (2015), who show for ARFIMA models encompassing the reduced form of (9) that the probability of the CSS estimator to stay in the region of the parameter space where $v_t(\theta)$ is nonstationary is asymptotically zero, the relevant region of Θ asymptotically reduces to the region where $d_0 - d < 1/2$ holds. Within the relevant region of the parameter space this paper then proves weak convergence of the objective function by showing the objective function to satisfy a uniform weak law of large numbers. This yields consistency of the CSS estimator, see Wooldridge (1994, thm. 4.3).

Theorem 3.3 (Asymptotic normality). *Under assumptions 1 to 3 the CSS estimator $\hat{\theta}$ is asymptotically normally distributed, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega_0^{-1})$ as $n \rightarrow \infty$.*

The proof of theorem 3.3 is again contained in appendix C. Since the CSS estimator is consistent, the asymptotic distribution theory is inferred from a Taylor expansion of the score function about θ_0 . A central limit theorem is shown to hold for the score function at θ_0 , together with a uniform weak law of large numbers for the Hessian matrix. The latter allows to evaluate the Hessian matrix in the Taylor expansion of the score function at θ_0 . Thus, the asymptotic distribution of the CSS estimator, as given in theorem 3.3, can be inferred from solving the Taylor expansion for $\sqrt{n}(\hat{\theta} - \theta_0)$. As usual in the state space literature, no analytical solution to the asymptotic variance of the CSS estimator can be provided. The parameters of the reduced form depend non-trivially on θ , so that the partial derivatives of the reduced form cannot be analytically derived. However, from theorem 3.3 it follows that an estimate for the parameter covariance matrix can be obtained from the negative inverse of the Hessian matrix computed in the numerical optimization.

Estimation of the latent component x_t in (9) is considered next. In line with the methodological literature on state space models, x_t is estimated by plugging the CSS estimates $\hat{\theta}$ into the projection

$$x_{t|n}(\theta) = \text{Cov}_{\theta}(x_t, y_{n:1}) \text{Var}_{\theta}(y_{n:1})^{-1} y_{n:1} = \sum_{i=0}^{t-1} \pi_i(-d) \Sigma_{\eta_{t:1} y_{n:1}}^{(i, \cdot)} \Sigma_{y_{n:1}}^{-1} y_{n:1}, \quad (13)$$

where $y_{n:1} = (y_n, \dots, y_1)'$, $\Sigma_{\eta_{t:1}y_{n:1}} = \text{Cov}_\theta(\eta_{t:1}, y_{n:1})$, and $\Sigma_{y_{n:1}} = \text{Var}_\theta(y_{n:1})$. The superscript in $\Sigma_{\eta_{t:1}y_{n:1}}^{(i,\cdot)}$ denotes the i -th row of the matrix, and

$$\Sigma_{\eta_{t:1}y_{n:1}}^{(i,j)} = \begin{cases} \pi_{n-t+i-j}(-d)\sigma_\eta^2 & \text{if } n-j \geq t-i, \\ 0 & \text{else,} \end{cases}$$

while the entries of $\Sigma_{y_{n:1}}$ follow from lemma 3.1 by setting $t = n$. For $\theta = \theta_0$, $x_{t|n}(\theta_0)$ is the minimum variance linear unbiased estimator given y_1, \dots, y_n , see Durbin and Koopman (2012, lemma 2). Due to theorem 3.2, this property holds asymptotically for $x_{t|n}(\hat{\theta})$. Note that (13) is identical to the Kalman smoother, see Durbin and Koopman (2012, ch. 4.4). However, (13) is computationally much simpler, as it avoids the computationally intensive Kalman recursions for the conditional variance. In line with lemma 3.1, (13) is the fractional filter for x_t given \mathcal{F}_n .

Finally, estimation of the seasonal components $\alpha_{i,0}$, $i = 1, \dots, 7$, and μ_0 is considered. Theoretically, all parameters of the model in (8) could be estimated jointly by the CSS estimator. But as Tschernig et al. (2013) explain, including deterministic terms in the optimization can lead to poor results in finite samples for fractionally integrated processes, particularly when d_0 is close to unity, as the deterministic terms suffer from poor identification. They provide simulation evidence and a line of reasoning explaining why the following two-step estimator is more robust: In the first step, the integration order d_0 is estimated using the exact local Whittle estimator of Shimotsu (2010), which allows for unknown deterministic terms and yields \hat{d}_{EW} . Based on \hat{d}_{EW} , the deterministic terms $\mu_0, \alpha_{0,1}, \dots, \alpha_{0,7}$ in

$$\Delta_+^{\hat{d}_{EW}} \log \tilde{Y}_t = \Delta_+^{\hat{d}_{EW}} \mu + \sum_{i=1}^7 \alpha_i \Delta_+^{\hat{d}_{EW}} s_{i,t} + \text{error}_t, \quad (14)$$

are estimated by ordinary least squares. In the second step, the objective function of the CSS estimator in (12) is minimized for the adjusted $\log \tilde{Y}_t - \hat{\mu} - \sum_{i=1}^7 \hat{\alpha}_i s_{i,t}$.

As an alternative to (14), one could also eliminate the seasonal components by averaging over seven neighboring observations, as $\sum_{i=1}^7 \alpha_{i,0} s_{i,t} = 0$. The intercept in

$$\Delta_+^{\hat{d}_{EW}} \frac{1}{7} \sum_{i=0}^6 \log \tilde{Y}_{t-q+i} = \Delta_+^{\hat{d}_{EW}} \mu + \text{error}_t, \quad 0 \leq q \leq 6, \quad (15)$$

could then be estimated by ordinary least squares. q determines whether averages are calculated solely based on past data ($q = 6$), based on centered data around t ($q = 3$), or based on future data ($q = 0$). While the second approach does not require to estimate $\alpha_{1,0}, \dots, \alpha_{7,0}$, averaging over seven days smooths out potential kinks in the contact rate which is problematic. Furthermore, averaging may pollute the estimates of x_t and induce spurious long memory. Finally, the choice of q is not trivial: While for forecasting purposes $q = 6$ is adequate, choosing $q = 0$ is likely to account best for the delay in reporting of case numbers, and obviously $q = 3$ may be a good

compromise between the two options. In the applications $\mu_0, \alpha_{0,1}, \dots, \alpha_{0,7}$ will be estimated via (14).

4 Empirical results

In this section, estimation results for the time-varying contact rate β_t are presented for Canada, Germany, Italy, and the United States. The underlying data on confirmed, recovered, and deceased cases stems from the JHU CSSE. As in Lee et al. (2021) and Liu et al. (2021), $t = 1$ is set once the number of cumulative cases reaches 100. Prior smoothing as suggested by Lee et al. (2021) and Liu et al. (2021), who use one-sided three-day rolling averages to smooth the data, is avoided, as this likely pollutes the kinks in the contact rate that occur due to containment measures.³

Instead of smoothing out seasonality, the data is adjusted for weekly seasonal patterns as described at the end of section 3 using (14). The bandwidth for the exact local Whittle estimator in (14) is set to $m = \lfloor n^{0.65} \rfloor$, which is justified by the Monte Carlo study in appendix B. Based on the seasonally adjusted data, the parameters θ_0 are estimated via the CSS estimator (12), where 100 combinations of starting values for θ_0 are drawn from uniform distributions with appropriate support (in particular $d \in [0.5, 2]$) to avoid convergence to a local optimum. However, due to the parsimonious parametrization of the model, all combinations of starting values converged to virtually identical optima, implying that the procedure is robust to the choice of starting values. Plugging the CSS estimates into (13), together with $\hat{\mu}$ in (14), yields the log contact rate estimate $\log \hat{\beta}_t$.

The average infected period is required for \mathcal{R}_t and is estimated by solving (2) for γ and taking the average

$$\hat{\gamma} = \frac{1}{n-1} \sum_{t=2}^n \left[\hat{\beta}_t S_{t-1} - \frac{\Delta I_t}{I_{t-1}} \right]. \quad (16)$$

This reflects that the definition of recovered varies over the countries under study, particularly as non-hospitalized persons are typically assumed to have recovered h days after they tested positive, and h varies over the countries under study. The choice of h proportionally affects the number of currently infected I_t , and thus β_t is inversely proportional to h by (5). Consequently, only the dynamics of β_t should be compared over the different countries, not the absolute numbers. In contrast, the reproduction rate $\mathcal{R}_t = \beta_t/\gamma$ accounts for the different h when γ is estimated via (16). If instead $\gamma = 1/18$ is fixed as in Lee et al. (2021), the dependence on h is not resolved and countries with a higher h will exhibit a smaller reproduction rate by construction. This is precisely the reason for the implausible estimates for \mathcal{R}_t in Lee et al. (2021), and is solved by

³Liu et al. (2021) argue that one-sided three-day rolling averages smooth out noise generated by the timing of the reporting. However, the opposite should be the case, as a one-sided smoothing shifts case numbers from past to present, while a delay in reporting shifts case numbers from present to the future. To fix the latter, a forward-looking filter is required, not a backward-looking one, see the discussion at the end of section 3.

accounting for different h via (16).

Results are reported for Canada, Germany, and Italy in subsection 4.1. They are selected as they are all members of the G7 and have implemented containment measures of different strength, duration, and at different points in time, thus making a comparison interesting. For the selected countries, there exist reliable data on confirmed, recovered, and deceased cases provided by the JHU CSSE. As will be shown, the latter is not the case for the US, where data on recovered subjects suffers heavily from under-reporting, yielding a severe downward-bias for the estimated contact rate and the resulting reproduction rate \mathcal{R}_t as reported by Lee et al. (2021). The problem is fixed by an assumption on the average duration of an infection, and results for the US are presented in subsection 4.2. As will become apparent there, the fractional filter is quite robust to under-reporting of recovered cases. To monitor the pandemic in real-time, subsection 4.3 examines the precision of the fractional filter at the end of the sample.

4.1 Canada, Germany, and Italy

For Canada, figure 1 sketches the estimated log contact rate and the resulting reproduction rate $\hat{\mathcal{R}}_t = \hat{\beta}_t/\hat{\gamma}$ in the first row. The average duration of an infection is estimated to be $1/\hat{\gamma} = 18.29$ days. The second row of figure 1 displays the estimated prediction error $v_t(\hat{\theta})$ and its estimated autocorrelation function.

Based on the top-left panel of figure 1, several turning points of the contact rate can be identified using a simple algorithm that defines a minimum (maximum) whenever the contact rate β_t at t is smaller (greater) than all $\beta_{t+1}, \dots, \beta_{t+10}$, the contact rates of the next ten days. These periods correspond to several policy regimes characterized by the strengthening and easing of containment measures. While a small selection of policy measures is presented below, a detailed overview is given by McCoy et al. (2020).

1. **March 13 – March 21:** The contact rate increases and peaks on March 21. As a reaction, several provinces and territories declare the state of emergency between March 13 and March 22, impose gathering bans, close schools, universities, and businesses, and cancel mass events, among others.
2. **March 22 – July 5:** After the implementation of containment measures the contact rate decreases continuously. While additional containment measures such as travel restrictions are implemented in April, several provinces and territories start to step-wise relax their restrictions in May. On July 1, Canada’s national summer holidays begin.
3. **July 6 – July 24:** During the first half of the summer holidays the contact rate increases sharply. The reproduction rate increases above unity on July 20.
4. **July 25 – August 3:** The contact rate slightly decreases, while the reproduction rate remains above unity.
5. **August 4 – September 30:** After a short phase of reduction, contact and reproduction rate start to increase again, while schools re-open on September 8. Canada’s prime minister Trudeau says the second wave of COVID-19 is already underway.

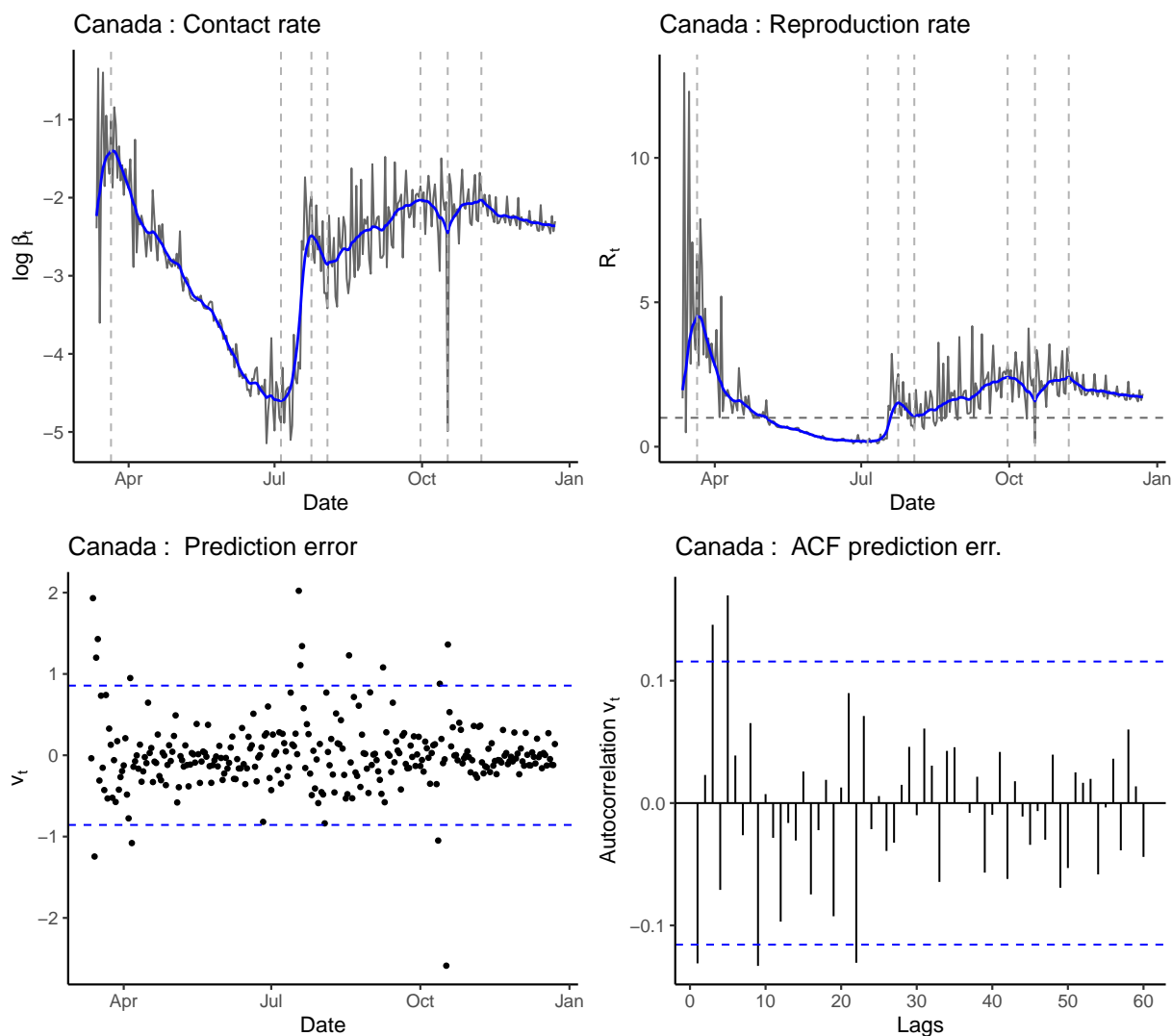


Figure 1: Estimation results for Canada. The top-left panel displays the estimated contact rate $\log \hat{\beta}_t$ in blue together with the observable $\log \hat{Y}_t$ in gray. The top-right panel shows the estimated reproduction rate $\hat{\mathcal{R}}_t = \hat{\beta}_t / \hat{\gamma}$ in blue together with $\hat{Y}_t / \hat{\gamma}$ in gray. The dashed horizontal line corresponds to $\mathcal{R} = 1$. The dashed vertical lines correspond to turning points of the contact rate. The bottom-left panel shows the estimated prediction error $v_t(\hat{\theta})$ in (11) together with two standard deviations in blue, dashed. The bottom-right panel sketches the estimated autocorrelation function of the prediction error $v_t(\hat{\theta})$ together with a 95% confidence interval.

6. **October 1 – October 17:** Contact and reproduction rate slightly decrease. Thanksgiving takes place on October 12.
7. **October 18 – November 7:** After Thanksgiving, contact and reproduction rate increase slightly. On November 3rd, Ontario introduces an incidence-based system for when to tighten containment measures.
8. **November 8 – December 23:** Contact and reproduction rate exhibit a slight but steady decrease. Reproduction rate remains above unity.

The estimated contact rate and the resulting reproduction rate are well in line with the chronology of policy interventions. In particular, the fractional filter allows to identify turning points of the contact rate that are not visible from the raw data that is plotted in gray color in figure 1.

The two graphs at the bottom of figure 1 illustrate how well the Canadian data fits the model assumptions. Assumptions 1 and 2 assume the measurement error u_t and the log contact rate shock η_t to be homoscedastic white noise processes. Since $\hat{\theta}$ is consistent, see theorem 3.2, by (C.1) the prediction error $v_t(\hat{\theta})$ becomes a white noise process as $n \rightarrow \infty$ if assumptions 1 and 2 hold. While some outliers exist, about 95% of the prediction errors lie within two standard deviations, as the bottom-left panel illustrates. The bottom-right panel shows that there is not much autocorrelation left in the prediction error. Given the parsimonious parametrization of the fractional UC model, this is surprising. The two panels at the bottom of figure 1 thus substantiate that the dynamics of the log contact rate are well captured by a fractionally integrated process.

The estimated integration order is $\hat{d} = 1.2166$, which implies that a unit shock on the contact rate growth $\Delta \log \beta_t$ retains 21.66% of its impact in $t + 1$, 13.17% in $t + 2$, and 9.73% in $t + 3$. After one week, the impact is still 5.10%, after two weeks 2.98%, and after three weeks 2.17%, which is due to the strong persistence of fractionally integrated processes, see assumption 2 for the formula for $\pi_i(d - 1)$. The slow decay may very well describe the persistent impact of past information shocks on today's social behavior.

For Germany, figure 2 plots the empirical results. The average infected period is estimated to be $1/\hat{\gamma} = 21.27$ days, which is slightly greater compared to Canada and likely results from different algorithms to estimate the number of recovered individuals. The integration order estimate is of similar size as for Canada ($\hat{d} = 1.2693$, see table A.1). From the top-left panel of figure 2 the following contact regimes can be identified:

1. **March 2 – March 5:** The contact rate starts at a comparably high level on March 2, likely caused by Carnival celebrations and ski tourism during Germany's winter holidays (Felbermayr et al.; 2020). It peaks on March 5.
2. **March 6 – May 2:** A slight decrease at the beginning of March turns into a sharp decrease around March 16. Measures to contain the spread of the virus, such as school closings and event cancellations, are implemented from March 13 on. From March 22 on, gatherings of more than two people are prohibited and several businesses are closed (Hartl, Wälde and Weber; 2020). At the end of April, schools and businesses partly re-opened.

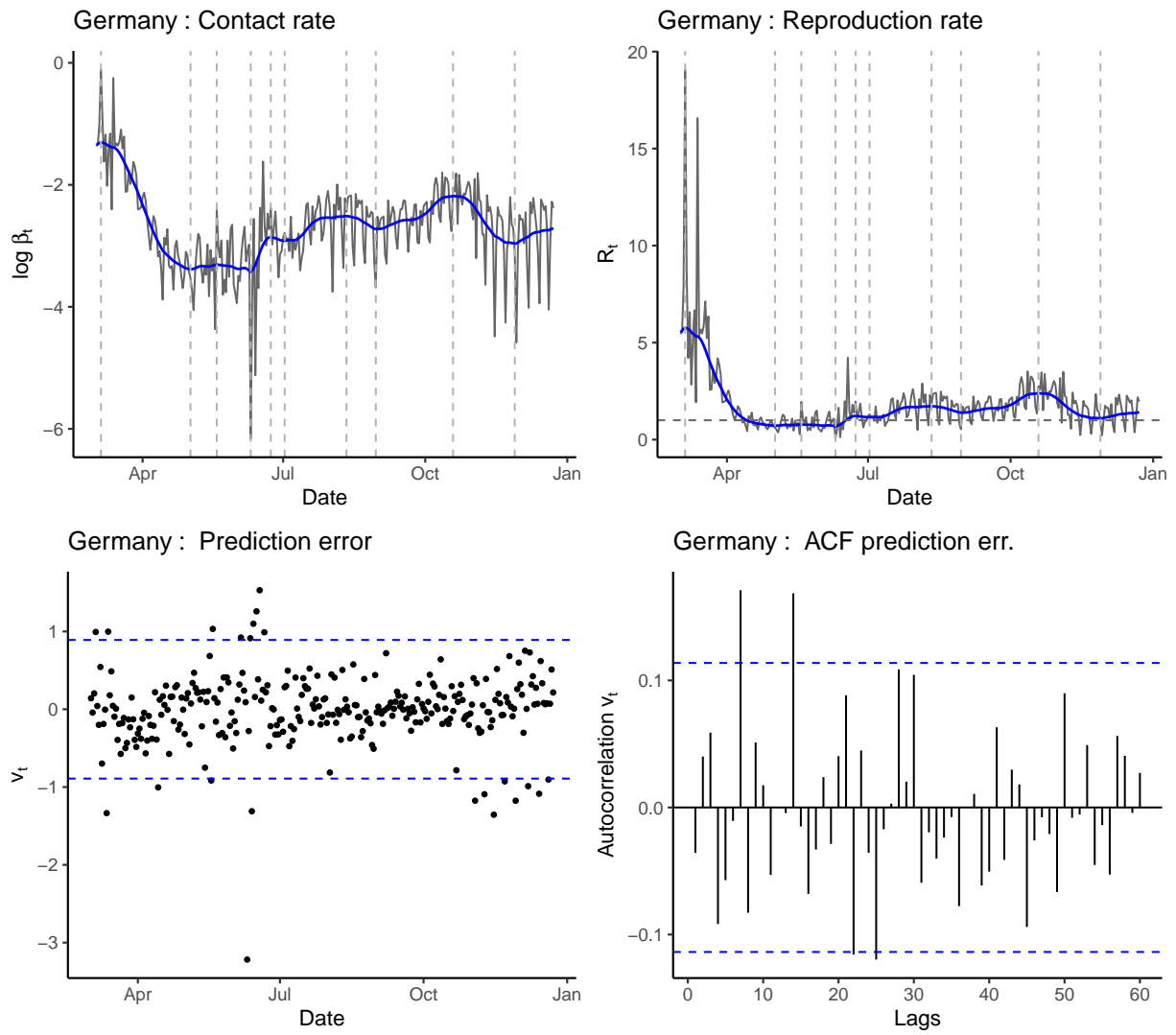


Figure 2: Estimation results for Germany. For a description see figure 1.

3. **May 3 – May 19:** Contact and reproduction rate slightly increase. Gathering restrictions are relaxed and most businesses are allowed to re-open on May 6.
4. **May 20 – June 10:** Contact and reproduction rate slightly decrease.
5. **June 11 – June 23:** A short but strong increase is caused, among others, by a massive outbreak in a meat factory (BBC News; 2020d).
6. **June 24 – July 2:** A slight decline follows. On June 29, summer holidays begin in Germany’s largest state.
7. **July 3 – August 11:** During the summer holidays, Germany experiences a further increase in its contact and reproduction rate.
8. **August 12 – August 30:** Contact and reproduction rate decrease.
9. **August 31 – October 19:** A slight increase of the contact rate is followed by a strong increase at the end of September. On October 15 stricter rules for hotspots are implemented, including mask obligations, contact restrictions, and curfews (Deutsche Welle; 2020b).
10. **October 20 – November 28:** Contact and reproduction rate decrease, but the latter remains above unity. On October 28 a ‘lockdown light’ is announced for Germany, inducing gathering restrictions and business closings among others (BBC News; 2020a).
11. **November 29 – December 23:** While contact and reproduction rate slightly increase, the government announces a tightening of its lockdown measures on December 13 (BBC News; 2020b).

Similar to Canada, the two panels at the bottom of figure 2 indicate that the model assumptions are largely satisfied, despite little remaining autocorrelation in the prediction error.

For Italy, a slight adjustment of the data is required, as $\Delta C_t = 0$ on June 19 and thus $\log \tilde{Y}_t$ is not defined, see (5). To adjust the single observation, averages from the neighboring observations are used (i.e. $\Delta C_t = 1/3(\Delta C_{t-1} + \Delta C_{t+1})$), while the adjusted cases in $t - 1$ and $t + 1$ will equal $2/3$ of the reported cases. Thus, the cumulated number of reported cases is unaffected by the adjustment.

The empirical results are displayed in figure 3. An estimate for the average infected period is $1/\hat{\gamma} = 35.92$ days, which is significantly higher than for Canada and Germany. The discussion below (16) gives an explanation for the high variation in γ over the different countries. The estimated integration order is $\hat{d} = 1.4304$ (see table A.1), which is somewhat greater than the estimates for Germany and Canada, implying that a shock η_t in Italy will yield a more persistent effect on the contact rate. This may be explained by the severity of the pandemic in Italy in spring 2020, which likely had a long-lasting impact on social behavior. Based on the top-left panel of figure 3 the following regimes and turning points are visible:

1. **February 24 – February 25:** Due to several clusters in northern Italy more than 100 cumulative cases are counted on February 23. While the contact rate peaks on February 25, epicenters in northern Italy effectively went into lockdown on February 22 (Signorelli et al.; 2020).

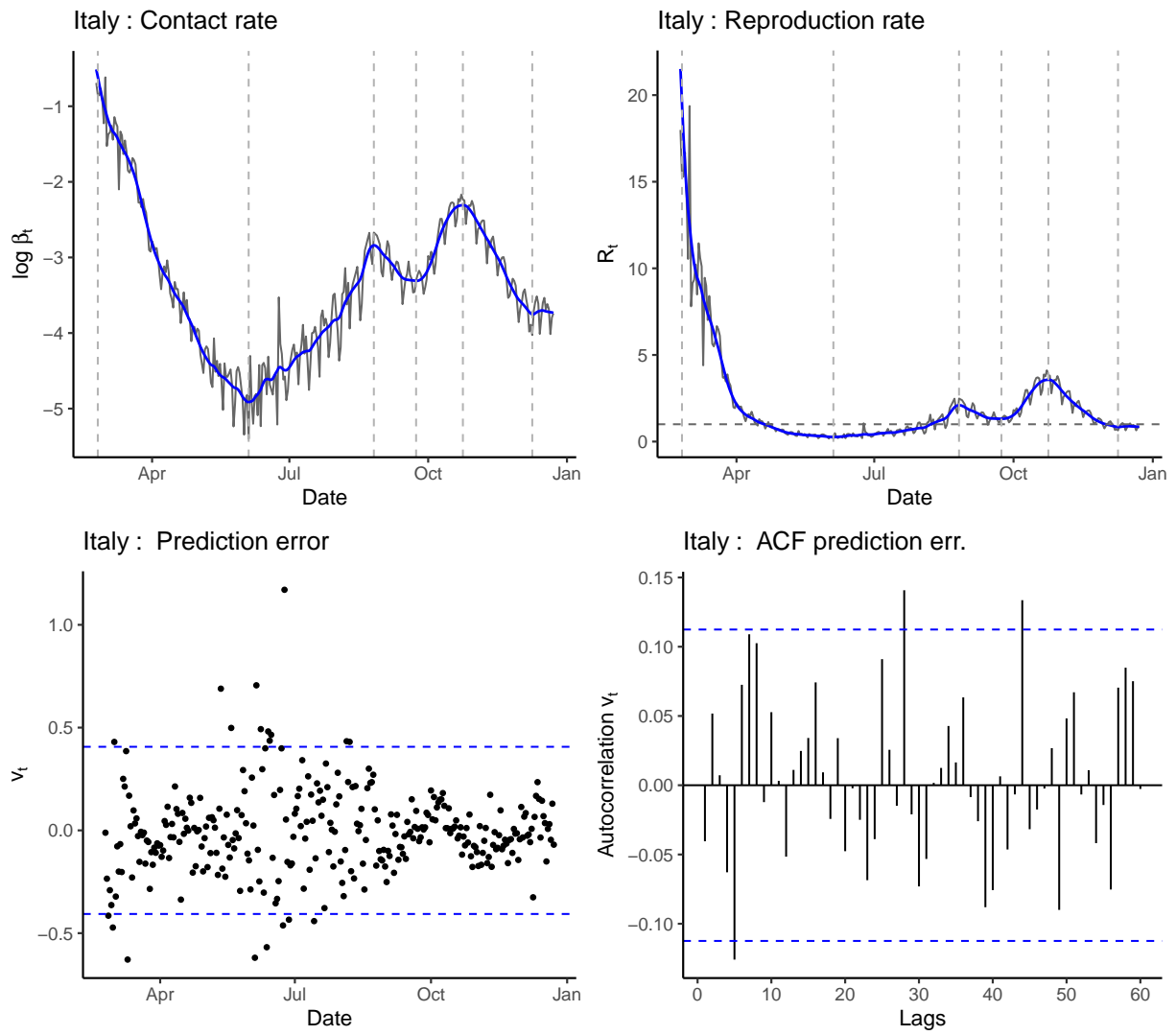


Figure 3: Estimation results for Italy. For a description see figure 1.

2. **February 26 – June 4:** The contact rate exhibits the strongest decline among all countries under study. During March, the government announces several containment measures such as school closings, halting all non-essential businesses, and tight regulations on free movement (Signorelli et al.; 2020). During May, the lockdown is lifted gradually. It effectively ends on June 3 (BBC News; 2020c).
3. **June 5 – August 26:** The contact rate exhibits a long and steep increase.
4. **August 27 – September 23:** A short decrease of the contact rate follows.
5. **September 24 – October 24:** The contact rate again increases and reaches a level as high as at the end of March. Gatherings, restaurants, sports and school activities are again restricted (Deutsche Welle; 2020a).
6. **October 25 – December 9:** The contact rate strongly declines, while additional restrictions on bars and restaurants are implemented (Deutsche Welle; 2020c).
7. **December 10 – December 23:** A minor increase in the contact rate is visible the days before Christmas.

The two panels at the bottom of figure 3 are similar to Canada and Germany, and indicate that the prediction errors are rather homoscedastic, although outliers exist, and little autocorrelation is left.

4.2 United States

The US is treated separately, since data on recovered cases reported by the JHU CSSE seem heavily downward-biased. To see this, consider the difference between lagged cumulative confirmed, cumulative recovered, and cumulative deceased cases for different lags h

$$C_{t-h} - R_t - D_t. \tag{17}$$

For $h = 0$, (17) measures the number of currently infected subjects. For small h , (17) should be positive, as it takes some time for the infected subjects to either recover or die. As h increases, (17) should turn negative, as an increasing number of subjects contained in the cumulative cases C_{t-h} and subjects infected between $t - h$ and t (and thus contained in $C_t - C_{t-h}$) either recover or die. The turning point, denoted by \bar{h} , should be close to the average infected period $1/\gamma$, as long as new confirmed cases between $t - h$ and t , i.e. $C_t - C_{t-h}$, do not explode. If they do, then \bar{h} should be smaller than $1/\gamma$, as outflows from $C_t - C_{t-h}$ disproportionately increase R_t and D_t .

Figure 4 plots (17) in case numbers for lags $h = 15, 18, \dots, 42, 45$. As can be seen, even after 45 days the difference between lagged cumulative confirmed, cumulative recovered, and cumulative deceased cases is predominantly positive. This is at odds with the average infected periods for Canada, Germany, and Italy as found in subsection 4.1, and indicates that data on R_t , D_t may suffer from under-reporting. As stated by the JHU CSSE, data on recovered cases are based on local media reports as well as on state reporting when available. US state-level recovered cases stem from the COVID Tracking Project (www.covidtracking.com). The COVID Tracking Project

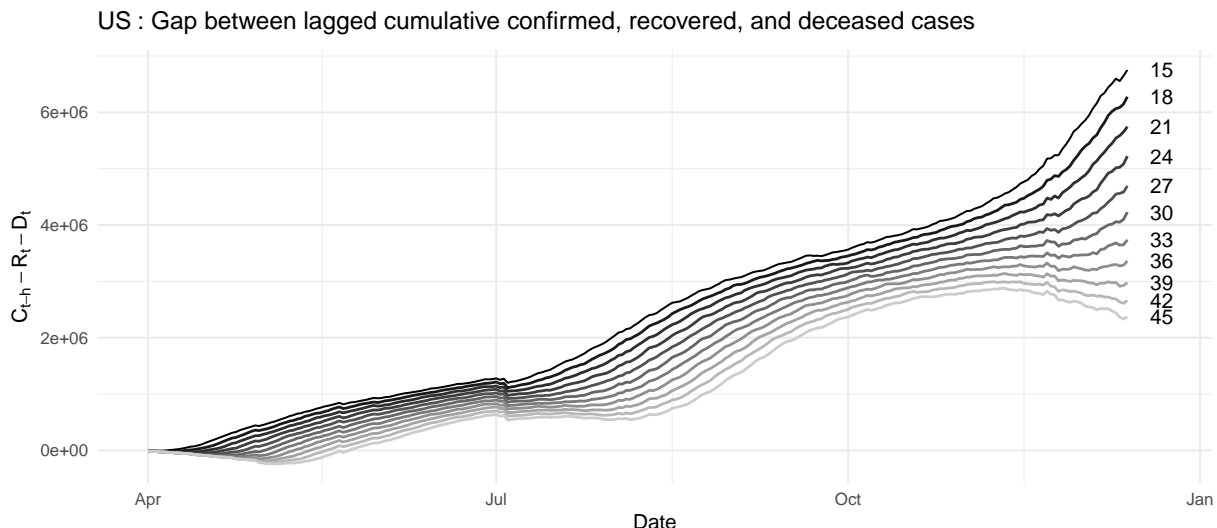


Figure 4: Difference between lagged cumulative confirmed, recovered, and deceased cases $C_{t-h} - R_t - D_t$ (in case numbers) for $h = 15, 18, \dots, 42, 45$.

(2021) recently pointed out that several states and territories, including California and Florida, do not report data on recovered cases, which may explain the downward-bias. In addition, definitions of recovered cases differ considerably across states, and the majority of the states consider a case as recovered if a certain number of days (generally between 10 and 30) after a positive test result or symptom onset have passed and the patient has not died.

Since no reliable data on recovered cases is available for the US, an approximation is required. In the following, it will be assumed that individuals either recover or die $\bar{h} = 21$ days after they tested positive, $C_{t-21} = D_t + R_t$. The assumption is justified as follows: First, it is similar to the average infected period estimated for Germany and more conservative than the estimate for Canada. And second, it is centered in the range of definitions for recovered individuals by the federal states. In addition, estimates for $\bar{h} = 18$ and $\bar{h} = 24$ days are presented, which gives a reasonable interval for the contact rate.

Under the assumption of $\bar{h} = 21$, the estimated integration order equals $\hat{d} = 1.2499$ (see table A.1) and is very similar to the results in subsection 4.1. Estimates for contact and reproduction rate are visualized in figure 5 and again allow to decompose the chronology of the pandemic into different regimes:

1. **March 5 – March 6:** Contact and reproduction rate peak at the initial stage of the epidemic.
2. **March 7 – May 11:** Contact and reproduction rate steadily decrease, despite a small blip on March 19. The national state of emergency is declared on March 13, and several containment measures such as business closings and stay-at-home orders are implemented mainly during the second half of March. Depending on the state, businesses re-open from April 20 on. More than half of the states have opened businesses on May 7 (Chernozhukov

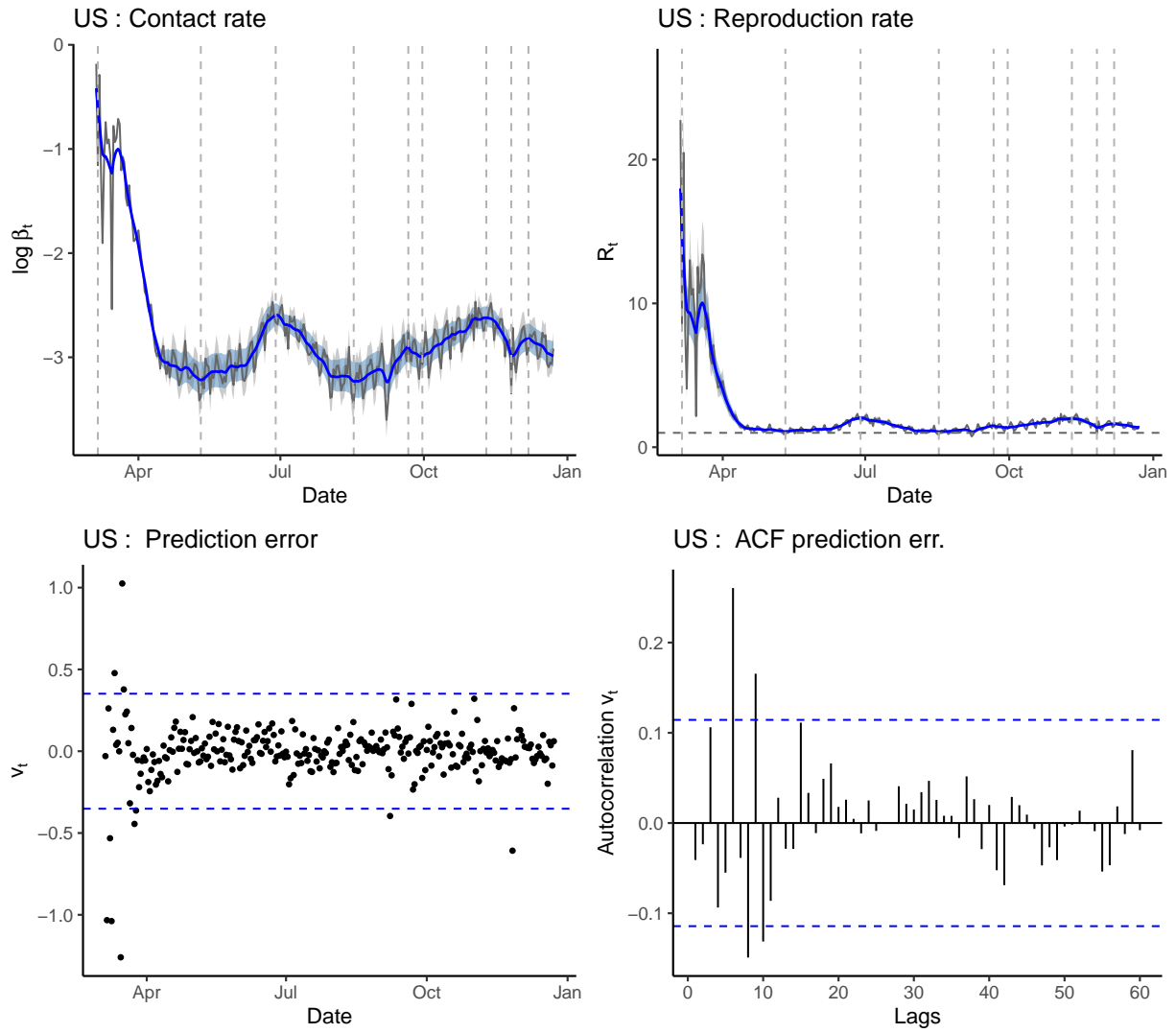


Figure 5: Estimation results for the United States for $\bar{h} = 21$. Shaded areas correspond to $\bar{h} = 18$ and $\bar{h} = 24$. For a description see figure 1.

et al.; 2021).

3. **May 12 – June 28:** As containment measures are relaxed, contact and reproduction rate increase slightly during the second half of May and experience a strong increase during June.
4. **June 29 – August 17:** Contact and reproduction rate decline, reaching the level of May.
5. **August 18 – September 21:** A slight increase is visible.
6. **September 22 – September 30:** A short decrease follows.
7. **October 1 – November 10:** Contact and reproduction rate exhibit a steady increase, reaching the June peak. Regional containment measures are implemented at the beginning of November.
8. **November 11 – November 26:** Contact and reproduction rate decrease.
9. **November 27 – December 7:** After Thanksgiving, contact and reproduction rate experience a short increase.
10. **December 8 – December 23:** A decrease is visible from December 8 on.

As can be seen from figure 5, estimates for the contact rate are rather robust to the choice of \bar{h} . They are slightly greater for $\bar{h} = 18$, as the number of currently infected I_t is smaller by construction and thus additional contacts are required to explain new confirmed cases, while they are slightly smaller for $\bar{h} = 24$ exactly for the opposite reason. The estimated reproduction rate is virtually identical among the three scenarios, as it is normalized by the average infected period $\hat{\mathcal{R}}_t = \hat{\beta}_t/\hat{\gamma}$. For $\bar{h} = 21$, the two panels at the bottom of figure 5 indicate that the model assumptions are largely satisfied, despite some weak correlation in the prediction errors. The plots are very similar for $\bar{h} = 18$ and $\bar{h} = 24$ and thus not shown. The reproduction rate $\hat{\mathcal{R}}_t$ is greater than unity during the whole sample, which contradicts the results of Lee et al. (2021) who rely on the downward-biased data on recovered subjects from the JHU CSSE while fixing the average infected period to be $1/\gamma = 18$.

4.3 Monitoring the current state of the pandemic

This subsection investigates the end-of-sample properties of the fractional filter for real-time estimation of the contact rate. Reliable contact rate estimates at the current frontier of the data would allow to real-time monitor the state of the pandemic and can serve as a surveillance measure for future outbreaks. Based on reliable real-time estimates for the contact rate, policy rules can be implemented to prevent an exponential growth of case numbers. Acting early reduces economic and social costs of containment measures, and consequently a well-designed policy rule will be beneficial, given that the fractional filter yields a reliable estimate for the current level of the contact rate. Drawing inference on the latter is the focus of this subsection.

In detail, real-time monitoring is simulated by truncating the sample at a certain point t , $r \leq t \leq n$, where r is the minimum sample size for the CSS estimator to produce reasonable estimates. The parameters $\theta_0, \mu_0, \alpha_{1,0}, \dots, \alpha_{7,0}$ of (8) are then estimated as described in section 3 using the information available at time t , \mathcal{F}_t , and the resulting parameter estimates are denoted

as $\hat{\theta}^{(t)}$, $\hat{\mu}^{(t)}$, etc. To take into account reporting lags, and to be robust against outliers at the end of the sample, a little backward-smoothing is allowed by reporting the smoothed estimate for the log contact rate at period $t - 3$ given the information available at period t . From (13), the smoothed estimates are

$$\log \hat{\beta}_{t-3|t} = \hat{\mu}^{(t)} + x_{t-3|t}(\hat{\theta}^{(t)}). \quad (18)$$

As (18) only depends on information available at t , it mimics the situation of a policy maker at t and can be used to draw inference on the monitoring properties of the fractional filter at time t . Based on $\hat{\beta}_{t-3|t}$, policy rules to prevent an exponential spread of the virus can be designed. Such rules could, for instance, define a threshold for $\hat{\mathcal{R}}_{t-3}$ at which additional containment measures are implemented. As the threshold should naturally depend on the number of currently infected, current hospital capacities, and other parameters, the precise design of such a policy rule is left to the experts, and only a primitive policy rule will be introduced later for illustrative purposes.

The reliability of real-time estimates for the contact rate is assessed by the following experiment: First, an estimation sample that consists of information available until May 31 is defined, for which $\theta_0, \mu_0, \alpha_{1,0}, \dots, \alpha_{7,0}$ are estimated. It consists of at least 80 observations, which is considered as a reasonable sample size for the estimation sample. Based on these estimates, $\log \hat{\beta}_{r-3|r}$ is obtained as described above. In a second step, information available on June 1 is added to the sample and parameter estimates are updated using the $\hat{\theta}^{(r)}$ from the estimation sample as starting values for the CSS estimator, which gives $\hat{\theta}^{(r+1)}$. As before, the estimate for $\log \hat{\beta}_{r-2|r+1}$ is stored. The procedure repeats for all t , $r < t \leq n$, where in every step t the CSS estimator is initialized by $\hat{\theta}^{(t-1)}$. The resulting real-time estimates for the contact rate are then compared to those of subsections 4.1 and 4.2 to draw inference on their reliability. In addition, a primitive policy rule is introduced. It assumes governments to take action as soon as $\hat{\mathcal{R}}_{t-3} = \hat{\beta}_{t-3|t}/\hat{\gamma} > 1.2$. The latter is motivated by the observation that preventing an exponential propagation (i.e. $\mathcal{R}_{t-3} > 1$) is desirable, and a margin of 0.2 is included to be robust against outliers. Finally, the real-time contact rate estimates are compared to a rolling seven-day average

$$\log \hat{\beta}_{t-3|t}^{benchmark} = \frac{1}{n} \sum_{i=0}^6 \log \tilde{Y}_{t-i}, \quad (19)$$

which includes three forward-looking observations and should smooth out the seasonality.

The real-time experiment considered in this paper deviates from Lee et al. (2021), who suggest to monitor the current state of the pandemic by fitting a linear time trend with structural breaks to $\log \tilde{Y}_t$. Lee et al. (2021) evaluate the monitoring properties of their contact rate estimate ex-post, using all information available in their sample. Consequently, their estimates at point t depend on information that was not available to policy makers at period t whenever $t < n$. As structural breaks are not well identified at the end of the sample, recent changes in the contact rate cannot be expected to be found by the estimator of Lee et al. (2021).

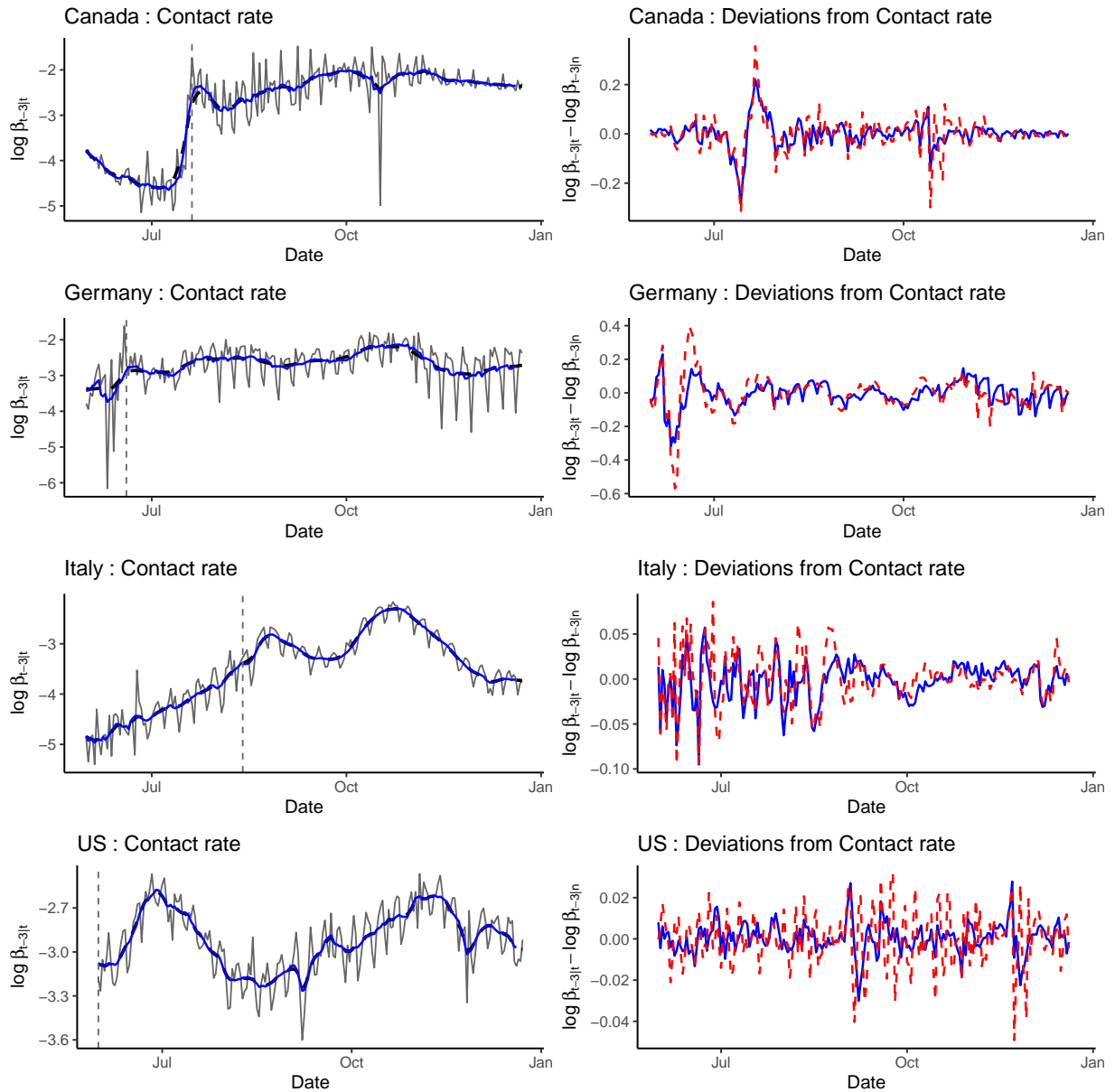


Figure 6: Real-time estimates for the log contact rate. The left panels display real-time contact rate estimates $\log \hat{\beta}_{t-3|t}$ (blue, solid), full sample estimates $\log \hat{\beta}_{t-3|n}$ (black, dashed), and the observable data $\log \tilde{Y}_{t-3}$ (gray) for Canada, Germany, Italy, and the United States. The dashed vertical line corresponds to the date where the real-time estimate for the reproduction rate exceeds 1.2. The right panels show deviations from the full sample contact rate estimates for the real-time estimates $\log \hat{\beta}_{t-3|t} - \log \hat{\beta}_{t-3|n}$ (blue, solid) and the benchmark $\log \hat{\beta}_{t-3|t}^{benchmark} - \log \hat{\beta}_{t-3|n}$ (red, dashed).

The results of the real-time experiment are visualized in figure 6. The four panels on the left side sketch the resulting real-time estimates for the contact rate for the four countries under study, together with the (full sample) results of subsections 4.1 and 4.2. As can be seen, the real-time estimates almost perfectly overlap with estimates using the full sample information. This implies that the real-time estimates are well suited for monitoring the current state of the contact rate. Minor deviations are visible for Canada at the end of July, and for Germany at the middle of June, while no greater deviations are visible for Italy and the US. The real-time estimates exceed the threshold $\hat{\mathcal{R}}_{t-3} = \hat{\beta}_{t-3|t}/\hat{\gamma} > 1.2$ on the same day as estimates based on the full sample for Italy (August 13) and for the US (May 31), where May 31 is the first observation for the real-time estimates. For Canada, the reproduction rate based on the full sample exceeds the threshold on July 21, one day after the real-time estimate. For Germany, the real-time estimates exceed the threshold for the reproduction rate on June 19, two days before those based on the full sample. These findings again substantiate the usefulness of the fractional filter as a surveillance measure to monitor the current state of the pandemic. From the results in figure 6, it follows that the primitive policy rule would have required governments to take action during the summer where case numbers were comparably low across the four countries under study. Such an early intervention would have likely reduced the economic and social costs compared to the lockdown measures as implemented at the end of year 2020.

The four panels on the right side of figure 6 display the deviations of $\log \hat{\beta}_{t-3|t}$ and $\log \hat{\beta}_{t-3|t}^{benchmark}$ from the log contact rate estimates based on the full sample information $\log \hat{\beta}_{t-3|n}$. Thus, they shed light on whether the fractional filter improves estimates for the contact rate compared to a rolling seven-day average that uses three forward-looking observations. For Italy and the US, the advantages of the fractional filter directly become apparent, as the benchmark exhibits greater deviations. For Canada and Germany, the fractional filter performs comparably well when large outliers occur, e.g. around July 20 for Canada and around June 20 for Germany.

5 Conclusion

To extract a time-varying signal for the COVID-19 contact rate from daily data on confirmed, recovered, and deceased cases, this paper introduces a novel unobserved components model. It models the log contact rate as a fractionally integrated process of unknown integration order. A computationally simple modification of the Kalman filter is introduced and is termed the fractional filter. It provides a closed-form expression for the prediction error that allows to estimate the model parameters by a conditional-sum-of-squares (CSS) estimator. The asymptotic theory for the CSS estimator is provided. For the countries under study, estimation results are well in line with the chronology of the pandemic. They allow to draw inference on the impact of policy measures such as contact restrictions. The new filtering method bears great potential as a monitoring device for the current state of the pandemic, as it yields reliable contact rate estimates at the current frontier of the data.

As vaccines become more and more available, future research can generalize the model to include the number of vaccinated. For instance, this can be done by decomposing $1 = S_t + I_t + R_t + D_t + V_t$, where V_t is the fraction of vaccinated. The states R_t and V_t should be non-overlapping as long as vaccines are not rolled out to recovered subjects. While vaccine recommendations vary over the different countries, some assign a lower priority to recovered subjects, so that R_t and V_t are non-overlapping at the early stage of the vaccine roll-out. Furthermore, mutations of the Coronavirus can be taken into account e.g. by allowing for a smooth transition between a contact rate with a low probability of virus transmission and one with a high probability.

For applications beyond COVID-19 related data, the fractional filter offers a robust, flexible, and data-driven way for signal extraction of data of unknown persistence. It requires no prior assumptions on the integration order of a process, and thus provides a solution to model specification in the unobserved components literature. Due to its computational advantages compared to the classic Kalman filter, it allows to estimate unobserved components models with richer dynamics.

Acknowledgments

The author thanks Nicolas Apfel, Uwe Hassler, Timon Hellwagner, Roland Jucknewitz, Alina Prechtel, Veronika Püschel, Lars Schlereth, Rolf Tschernig, Enzo Weber, and the participants of the Department Seminar at the University of Regensburg for very helpful comments.

A Estimation results

	Canada	Germany	Italy	United States
d	1.2166 (0.3271)	1.2693 (0.1989)	1.4304 (0.1242)	1.2499 (0.2638)
σ_η^2	0.0133 (0.1105)	0.0107 (0.1188)	0.0149 (0.2472)	0.0117 (0.2190)
σ_u^2	0.2018 (0.6128)	0.7991 (0.7636)	0.3067 (0.5920)	0.0764 (0.3791)

Table A.1: Estimation results $\hat{\theta}$ from the CSS estimator as described in section 3 for Canada, Germany, Italy, and the United States. Standard errors are denoted in parentheses and were calculated based on the inverse of the numeric Hessian matrix, see theorem 3.3.

B Monte Carlo evidence

The finite sample performance of the CSS estimator is assessed in a Monte Carlo study, where, to be in line with (9), the data-generating mechanism is given by

$$y_t = x_t + u_t, \quad \Delta_+^{d_0} x_t = \eta_t, \quad t = 1, \dots, n. \quad (\text{B.1})$$

$u_t \sim NID(0, \sigma_{u,0}^2)$, $\eta_t \sim NID(0, \sigma_{\eta,0}^2)$, u_t, η_t are uncorrelated, and $\sigma_{\eta,0}^2 = \rho \sigma_{u,0}^2$ so that ρ controls the signal-to-noise ratio. The integration orders $d_0 \in \{0.75, 1.25, 1.75\}$ cover the relevant interval for the applications in section 4, while $\rho \in \{0.5, 1, 2\}$ captures high and low signal-to-noise ratios. The variance parameter is set to $\sigma_{u,0}^2 = 1$. Different sample sizes $n \in \{100, 200, 300\}$ covering the relevant regions for the applications in section 4 are considered. The parameters $\theta_0 = (d_0, \sigma_{\eta,0}^2, \sigma_{u,0}^2)$ are estimated via the CSS estimator as described in section 3. For each specification, 1000 replications are simulated, and starting values are set to $\theta_{start} = (1, 1, 1)$.

In addition to the CSS estimates, estimation results for d_0 from the exact local Whittle estimator of Shimotsu (2010) are reported as benchmarks for $m = \lfloor n^j \rfloor$ Fourier frequencies, $j \in \{0.45, 0.50, 0.55, 0.60, 0.65, 0.70\}$. Finally, the mean squared error MSE_x and the coefficient of determination R_x^2 for the estimation of x_t , that are calculated via

$$MSE_x = \frac{1}{n} \sum_{t=1}^n (x_t - x_{t|n}(\hat{\theta}))^2, \quad R_x^2 = 1 - \frac{\sum_{t=1}^n (x_t - x_{t|n}(\hat{\theta}))^2}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad (\text{B.2})$$

are reported, and indicate how well x_t is estimated by the fractional filter (13).

The results for the Monte Carlo study are contained in table B.1. Not surprisingly, the parametric CSS estimator outperforms the exact local Whittle estimator. However, gains are quite large in terms of the MSE for the integration order for all $n \in \{100, 200, 300\}$ and all combinations of ρ and d_0 . The mean squared error of the integration order estimate becomes

ρ	d_0	\hat{d}	$\hat{d}_{EW}^{.45}$	$\hat{d}_{EW}^{.50}$	$\hat{d}_{EW}^{.55}$	$\hat{d}_{EW}^{.60}$	$\hat{d}_{EW}^{.65}$	$\hat{d}_{EW}^{.70}$	MSE_x	R_x^2
$n = 100$										
.5	0.75	0.0641	0.1021	0.0804	0.0762	0.0736	0.0728	0.0775	0.4786	0.6747
.5	1.25	0.0387	0.1011	0.0721	0.0664	0.0694	0.0789	0.1054	0.3719	0.9796
.5	1.75	0.0285	0.0876	0.0620	0.0576	0.0637	0.0809	0.1293	0.3418	0.9992
1	0.75	0.0409	0.0943	0.0673	0.0585	0.0505	0.0446	0.0433	0.6245	0.7914
1	1.25	0.0299	0.0978	0.0644	0.0535	0.0484	0.0465	0.0570	0.4880	0.9867
1	1.75	0.0239	0.0851	0.0539	0.0470	0.0453	0.0475	0.0710	0.4258	0.9995
2	0.75	0.0277	0.0919	0.0615	0.0504	0.0407	0.0318	0.0264	0.7861	0.8711
2	1.25	0.0231	0.0977	0.0601	0.0489	0.0393	0.0323	0.0319	0.6282	0.9915
2	1.75	0.0204	0.0830	0.0511	0.0422	0.0372	0.0325	0.0384	0.5306	0.9997
$n = 200$										
.5	0.75	0.0232	0.0662	0.0446	0.0396	0.0393	0.0427	0.0488	0.3985	0.8158
.5	1.25	0.0154	0.0615	0.0394	0.0320	0.0313	0.0381	0.0541	0.3432	0.9934
.5	1.75	0.0128	0.0519	0.0348	0.0281	0.0260	0.0328	0.0561	0.3287	0.9999
1	0.75	0.0171	0.0641	0.0390	0.0307	0.0256	0.0238	0.0248	0.5479	0.8742
1	1.25	0.0124	0.0614	0.0378	0.0280	0.0225	0.0214	0.0267	0.4502	0.9956
1	1.75	0.0106	0.0513	0.0335	0.0260	0.0206	0.0192	0.0276	0.4085	0.9999
2	0.75	0.0128	0.0622	0.0372	0.0275	0.0202	0.0159	0.0140	0.7169	0.9185
2	1.25	0.0104	0.0620	0.0370	0.0268	0.0193	0.0149	0.0145	0.5815	0.9972
2	1.75	0.0091	0.0510	0.0331	0.0252	0.0185	0.0141	0.0148	0.5068	0.9999
$n = 300$										
.5	0.75	0.0157	0.0448	0.0339	0.0284	0.0274	0.0311	0.0386	0.3770	0.8594
.5	1.25	0.0106	0.0404	0.0288	0.0218	0.0199	0.0242	0.0386	0.3394	0.9964
.5	1.75	0.0089	0.0361	0.0265	0.0195	0.0164	0.0189	0.0364	0.3278	0.9999
1	0.75	0.0120	0.0423	0.0301	0.0224	0.0185	0.0173	0.0190	0.5227	0.9031
1	1.25	0.0086	0.0402	0.0282	0.0199	0.0157	0.0145	0.0187	0.4442	0.9976
1	1.75	0.0075	0.0361	0.0258	0.0187	0.0145	0.0125	0.0177	0.4069	1.0000
2	0.75	0.0093	0.0412	0.0288	0.0203	0.0153	0.0118	0.0105	0.6895	0.9366
2	1.25	0.0072	0.0401	0.0276	0.0193	0.0144	0.0109	0.0103	0.5712	0.9985
2	1.75	0.0064	0.0360	0.0257	0.0186	0.0140	0.0102	0.0099	0.5037	1.0000

Table B.1: Mean squared error (MSE) and R_x^2 for d_0 and x_t in (B.1). The columns \hat{d} and \hat{d}_{EW}^j show the MSE for the CSS estimator of d_0 as well as for the exact local Whittle estimator of Shimotsu (2010) for $m = \lfloor n^j \rfloor$ Fourier frequencies, $j \in \{0.45, 0.50, 0.55, 0.60, 0.65, 0.70\}$. MSE_x displays the mean squared error for x_t , while R_x^2 is the coefficient of determination, see (B.2).

smaller for higher d_0 , which is plausible as the fraction of total variation of y_t generated by x_t increases with d_0 . For the same reason, it decreases with increasing ρ . The same conclusions on the precision with which d_0 is estimated hold for the mean squared error of x_t , which decreases as n , d , and ρ increase. The proportion of explained variation of x_t , measured by R_x^2 , is high and thus x_t is estimated well via (13). Particularly for $d = 1.25$, which is the relevant case for the applications in section 4, the R_x^2 is close to unity for all n .

C Proofs

Proof of Lemma 3.1. First, note that $E_\theta(x_{t+1}|\mathcal{F}_t) = E_\theta(y_{t+1}|\mathcal{F}_t)$, so that it is sufficient to derive the latter expression. For this, consider the reduced form of (9), which follows from taking fractional differences and utilizing the aggregation properties of MA processes, see Granger and Morris (1976), so that

$$\Delta_+^d y_t = \eta_t + \Delta_+^d u_t = \eta_t + \sum_{i=0}^{t-1} \pi_i(d) u_{t-i} = \sum_{i=0}^{t-1} \phi_i(\theta) \varepsilon_{t-i} = \phi(L, \theta) \varepsilon_t, \quad (\text{C.1})$$

with $\phi_0(\theta) = 1$, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, and $\phi(L, \theta)$ is invertible. σ_ε^2 and the coefficients in $\phi(L, \theta)$ can be derived by matching the autocovariance functions of (C.1), see Watson (1986, eqn. 2.6), and depend non-linearly on θ . However, they are not required for the proof. Solving for ε_t yields

$$\varepsilon_t = \phi(L, \theta)^{-1} \Delta_+^d y_t = y_t - \sum_{i=1}^{\infty} A_i(\theta) y_{t-i}.$$

From the type II definition of fractional integration, see assumption 2, it follows that $\text{Cov}_\theta(y_t, y_j) = 0$ for all $j \leq 0$, $t > 0$, and thus $E_\theta(y_{t+1}|\mathcal{F}_t) = \sum_{i=1}^t A_i(\theta) y_{t+1-i}$. The (yet unknown) coefficients $A_i(\theta)$ follow from the Yule-Walker equations

$$\begin{pmatrix} \text{Cov}_\theta(y_{t+1}, y_t) \\ \text{Cov}_\theta(y_{t+1}, y_{t-1}) \\ \vdots \\ \text{Cov}_\theta(y_{t+1}, y_1) \end{pmatrix} = \begin{bmatrix} \text{Var}_\theta(y_t) & \text{Cov}_\theta(y_{t-1}, y_t) & \cdots & \text{Cov}_\theta(y_1, y_t) \\ \text{Cov}_\theta(y_t, y_{t-1}) & \text{Var}_\theta(y_{t-1}) & \cdots & \text{Cov}_\theta(y_1, y_{t-1}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}_\theta(y_t, y_1) & \text{Cov}_\theta(y_{t-1}, y_1) & \cdots & \text{Var}_\theta(y_1) \end{bmatrix} \begin{pmatrix} A_1(\theta) \\ A_2(\theta) \\ \vdots \\ A_t(\theta) \end{pmatrix},$$

so that by defining the vectors $A(\theta) = (A_1(\theta), \dots, A_t(\theta))$, $y_{t:1} = (y_t, \dots, y_1)'$, and solving the Yule-Walker equations for $A(\theta)$, one has $A(\theta) = \text{Cov}_\theta(y_{t+1}, y_{t:1}) \text{Var}_\theta(y_{t:1})^{-1}$, which implies

$$E_\theta(y_{t+1}|\mathcal{F}_t) = \sum_{i=1}^t A_i(\theta) y_{t-i} = \text{Cov}_\theta(y_{t+1}, y_{t:1}) \text{Var}_\theta(y_{t:1})^{-1} y_{t:1}.$$

From $\text{Cov}_\theta(y_{t+1}, y_{t:1}) = \text{Cov}_\theta(x_{t+1}, y_{t:1}) = \sum_{i=1}^t \pi_i(-d) \text{Cov}_\theta(\eta_{t+1-i}, y_{t:1})$, see assumption 2, lemma 3.1 follows.

□

Proof of Theorem 3.2. First, the model in (9) is shown to be identified. Identification follows if the parameters σ_η^2 , σ_u^2 can be recovered from the autocovariance function of the reduced form $\phi(L, \theta)\varepsilon_t$ in (C.1). To see this, consider the covariances $\text{Var}_\theta(\phi(L, \theta)\varepsilon_t) = \sigma_\varepsilon^2 \sum_{i=0}^{t-1} \phi_i^2(\theta) = \sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d)$, and $\text{Cov}_\theta(\phi(L, \theta)\varepsilon_t, \phi(L, \theta)\varepsilon_{t-1}) = \sigma_\varepsilon^2 \sum_{i=0}^{t-2} \phi_i(\theta)\phi_{i+1}(\theta) = \sigma_u^2 \sum_{i=0}^{t-2} \pi_i(d)\pi_{i+1}(d)$. In matrix form this gives

$$\sigma_\varepsilon^2 \begin{pmatrix} \sum_{i=0}^{t-1} \phi_i^2(\theta) \\ \sum_{i=0}^{t-2} \phi_i(\theta)\phi_{i+1}(\theta) \end{pmatrix} = \begin{bmatrix} 1 & \sum_{i=0}^{t-1} \pi_i^2(d) \\ 0 & \sum_{i=0}^{t-2} \pi_i(d)\pi_{i+1}(d) \end{bmatrix} \begin{pmatrix} \sigma_\eta^2 \\ \sigma_u^2 \end{pmatrix}, \quad (\text{C.2})$$

so that solving for $(\sigma_\eta^2, \sigma_u^2)'$ yields

$$\begin{pmatrix} \sigma_\eta^2 \\ \sigma_u^2 \end{pmatrix} = \frac{1}{\sum_{i=0}^{t-2} \pi_i(d)\pi_{i+1}(d)} \begin{bmatrix} \sum_{i=0}^{t-2} \pi_i(d)\pi_{i+1}(d) & -\sum_{i=0}^{t-1} \pi_i^2(d) \\ 0 & 1 \end{bmatrix} \begin{pmatrix} \sum_{i=0}^{t-1} \phi_i^2(\theta) \\ \sum_{i=0}^{t-2} \phi_i(\theta)\phi_{i+1}(\theta) \end{pmatrix} \sigma_\varepsilon^2,$$

and thus $(\sigma_\eta^2, \sigma_u^2)'$ can be uniquely recovered from the reduced form. The assumption that $d > 0$ is crucial, as it guarantees $\sum_{i=0}^{t-2} \pi_i(d)\pi_{i+1}(d) \neq 0$, so that the matrix in (C.2) has full rank.

Next, the CSS estimator based on the reduced form (C.1) is derived and is shown to be identical to (12). Multiplying (C.1) by $\phi(L, \theta)^{-1}$ yields $\varepsilon_t = \phi(L, \theta)^{-1} \Delta_+^d y_t$, based on which a reduced form CSS estimator can be constructed. Define $\psi_+(L, \theta) = [\phi(L, \theta)^{-1}]_+ = [1 - \sum_{i=1}^{\infty} \psi_i(\theta)L^i]_+$, as the (truncated) inverse of $\phi(L, \theta)$, and denote $\varepsilon_t(\theta) = [\phi(L, \theta)^{-1}]_+ \Delta_+^d y_t = \psi_+(L, \theta) \Delta_+^d y_t$ as the reduced form residual given the observable variables y_1, \dots, y_n and θ . From $\varepsilon_t(\theta)$, the reduced form CSS estimator is

$$\hat{\theta} = \arg \min_{\theta \in \Theta} R(\theta), \quad R(\theta) = \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2(\theta), \quad (\text{C.3})$$

and equals the CSS estimator in (12). To see this, add and subtract y_t from $\varepsilon_t(\theta) = \psi_+(L, \theta) \Delta_+^d y_t$, so that $y_t = (1 - \psi_+(L, \theta) \Delta_+^d) y_t + \varepsilon_t(\theta)$, and plug y_t into the conditional expectation in (10)

$$v_t(\theta) = y_t - \text{E}_\theta(y_t | \mathcal{F}_{t-1}) = y_t - \text{E}_\theta \left[(1 - \psi_+(L, \theta) \Delta_+^d) y_t | \mathcal{F}_{t-1} \right] = \psi_+(L, \theta) \Delta_+^d y_t = \varepsilon_t(\theta).$$

The third equality follows from $(1 - \psi_+(L, \theta) \Delta_+^d) y_t$ being \mathcal{F}_{t-1} -measurable, since $\psi(L, \theta) = 1 - \sum_{i=1}^{\infty} \psi_i(\theta)L^i$ and $\pi_0(d) = 1$. Thus, the contemporaneous y_t cancel in the expectation operator and the whole term can be taken out of the expectation operator. From $v_t(\theta) = \varepsilon_t(\theta)$ it follows that the optimization problems in (12) and (C.3) are identical.

Next, the integration order of the residuals is assessed. Since $y_t \sim I(d_0)$, the residuals satisfy

$$\varepsilon_t(\theta) = \psi_+(L, \theta) \Delta_+^d y_t = \psi_+(L, \theta) \Delta_+^{d-d_0} \eta_t + \psi_+(L, \theta) \Delta_+^d u_t \sim I(d_0 - d). \quad (\text{C.4})$$

For $d_0 - d < 1/2$ the residuals are stationary, while for $d_0 - d > 1/2$ they are nonstationary. As the

asymptotic behavior of the objective function changes around $d_0 - d = 1/2$, the objective function does not uniformly converge on the set of admissible values for d . The same problem is addressed by Hualde and Robinson (2011) and by Nielsen (2015) for ARFIMA models encompassing (C.1). Nielsen (2015, eqn. 8) shows that a weak law of large numbers (WLLN) applies to the sum of squared residuals whenever $d_0 - d < 1/2$, while the sum of squared residuals diverges in probability whenever $d_0 - d \geq 1/2$, which translates into

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2(\theta) = \begin{cases} \text{E}[\tilde{\varepsilon}_t^2(\theta)] & \text{if } d_0 - d < 1/2, \\ \infty & \text{else.} \end{cases} \quad (\text{C.5})$$

$\tilde{\varepsilon}_t(\theta) = \psi(L, \theta)\phi(L, \theta_0)\Delta^{d-d_0}\varepsilon_t$ is the untruncated residual generated by the untruncated Δ^d and $\psi(L, \theta)$. In addition, letting $D^*(\kappa) = D \cap \{d : d_0 - d \leq 1/2 - \kappa\}$, $0 < \kappa < 1/2$ denote the region of the parameter space where $\varepsilon_t(\theta)$ is stationary, Nielsen (2015, eqn. 13) shows that for any constant $K > 0$ there exists a fixed $\bar{\kappa} > 0$ such that

$$\Pr \left(\inf_{d \in D \setminus D^*(\bar{\kappa}) \cap \theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \varepsilon_t^2(\theta) > K \right) \rightarrow 1, \quad \text{as } n \rightarrow \infty, \quad (\text{C.6})$$

implying that $\Pr(\hat{d} \in D^*(\bar{\kappa}) \cap \theta \in \Theta) \rightarrow 1$ as $n \rightarrow \infty$. From (C.6) it follows that the relevant parameter space asymptotically reduces to the stationary region $\Theta^*(\bar{\kappa}) = \{\theta | \theta \in \Theta, d \in D^*(\bar{\kappa})\}$.

Since the model is identified, for consistency it remains to be shown that a uniform weak law of large numbers (UWLLN) holds for the objective function within the stationary region of the parameter space. A UWLLN holds if both, the objective function and the supremum of the gradient, satisfy a WLLN, see Wooldridge (1994, thm. 4.2 and eqn. 4.4) and Newey (1991, cor. 2.2). While a WLLN for the objective function follows directly from (C.5), it remains to be shown that

$$\sup_{\theta \in \Theta^*(\kappa)} \left| \frac{\partial R(\theta)}{\partial \theta} \right| = O_p(1), \quad (\text{C.7})$$

for any fixed $0 < \kappa < 1/2$.

To prove (C.7), it will be helpful to note that for a white noise process ε_t , MA weights $\sum_{i=0}^{\infty} |m_{h,i}(\theta)| < \infty$, $h = 1, 2$, and the set $\tilde{\Theta} = \{\theta | \theta \in \Theta, d_0 - d < 1/2\}$, it holds that

$$\sup_{\theta \in \tilde{\Theta}} \left| \frac{1}{n} \sum_{t=1}^n \left[\frac{\partial^j \Delta_+^{d-d_0}}{\partial d^j} \sum_{i=0}^{\infty} m_{1,i}(\theta) \varepsilon_{t-i} \right] \left[\frac{\partial^k \Delta_+^{d-d_0}}{\partial d^k} \sum_{i=0}^{\infty} m_{2,i}(\theta) \varepsilon_{t-i} \right] \right| = O_p(1), \quad (\text{C.8})$$

for $j, k \geq 0$ as shown by Nielsen (2015, lemma B.3).

Now, consider the partial derivatives of (C.3)

$$\frac{\partial R(\theta)}{\partial \theta} = \frac{2}{n} \sum_{t=1}^n \varepsilon_t(\theta) \frac{\partial \varepsilon_t(\theta)}{\partial \theta}, \quad \frac{\partial \varepsilon_t(\theta)}{\partial \theta} = \frac{\partial \psi_+(L, \theta)}{\partial \theta} \Delta_+^d y_t + \psi_+(L, \theta) \frac{\partial \Delta_+^{d-d_0}}{\partial \theta} \Delta_+^{d_0} y_t. \quad (\text{C.9})$$

Since $\psi_+(L, \theta)$ satisfies the absolute summability condition for (C.8), it follows that

$$\sup_{\theta \in \Theta^*(\kappa)} \left| \frac{1}{n} \sum_{t=1}^n \varepsilon_t(\theta) \psi_+(L, \theta) \frac{\partial \Delta_+^{d-d_0}}{\partial d} \Delta_+^{d_0} y_t \right| = O_p(1), \quad (\text{C.10})$$

while the partial derivatives of $\Delta_+^{d-d_0}$ w.r.t. $\sigma_\eta^2, \sigma_u^2$ are zero.

For the remaining term in (C.9), note that the sum of absolute coefficients of the truncated polynomial $\psi_+(L, \theta)$ is bounded by the sum of absolute coefficients of the untruncated polynomial $\psi(L, \theta) = \phi(L, \theta)^{-1}$. Thus, it is sufficient to prove absolute summability of the coefficients in $\partial\psi(L, \theta)/\partial\theta = -\psi(L, \theta)^2(\partial\phi(L, \theta)/\partial\theta)$. Absolute summability of the coefficients in $\partial\phi(L, \theta)/\partial\theta$ is shown in lemma D.1 in appendix D. Since $\psi(L, \theta)$ is stable, $\partial\psi(L, \theta)/\partial\theta$ satisfies the absolute summability condition for (C.8) and thus

$$\sup_{\theta \in \Theta^*(\kappa)} \left| \frac{1}{n} \sum_{t=1}^n \varepsilon_t(\theta) \frac{\partial \psi_+(L, \theta)}{\partial \theta} \Delta_+^d y_t \right| = O_p(1). \quad (\text{C.11})$$

From (C.10) and (C.11) it follows that (C.7) holds. Consequently, the supremum of the gradient satisfies a WLLN for $\theta \in \Theta^*(\kappa)$, which generalizes the pointwise convergence of the objective function to weak convergence, implying that a UWLLN holds for the objective function. Since the model is identified, consistency of the CSS estimator follows from the UWLLN together with (C.6), and thus $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \rightarrow \infty$, see Wooldridge (1994, thm. 4.3). \square

Proof of Theorem 3.3. Since the CSS estimator is consistent, see theorem 3.2, the asymptotic distribution theory can be inferred from a Taylor expansion of the score function about θ_0

$$0 = \sqrt{n} \frac{\partial R(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}} = \sqrt{n} \frac{\partial R(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} + \sqrt{n} \frac{\partial^2 R(\theta)}{\partial \theta \partial \theta'} \Big|_{\theta=\bar{\theta}} (\hat{\theta} - \theta_0), \quad (\text{C.12})$$

where the entries in $\bar{\theta}$ satisfy $|\bar{\theta}_i - \theta_{0,i}| \leq |\hat{\theta}_i - \theta_{0,i}|$ for all $i = 1, 2, 3$, and θ_i denotes the i -th entry of $\theta = (d, \sigma_\eta^2, \sigma_u^2)'$, $i = 1, 2, 3$. The score function at θ_0 follows from (C.9)

$$\sqrt{n} \frac{\partial R(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} = \frac{2}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t(\theta_0) \frac{\partial \varepsilon_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} = S_n + o_p(1), \quad (\text{C.13})$$

where

$$S_n = \frac{2}{\sqrt{n}} \sum_{t=1}^n \varepsilon_t \frac{\partial \tilde{\varepsilon}_t(\theta)}{\partial \theta} \Big|_{\theta=\theta_0}. \quad (\text{C.14})$$

$\tilde{\varepsilon}_t(\theta) = \psi(L, \theta) \phi(L, \theta_0) \Delta^{d-d_0} \varepsilon_t$ is the untruncated residual generated by the untruncated Δ^d and $\psi(L, \theta) = 1 - \sum_{i=1}^{\infty} \psi_i(\theta) L^i$, and the second equality in (C.13) is shown to hold by Robinson (2006, pp. 135-136). In the following, let $S_n^{(j)}$ denote the j -th entry of S_n holding the partial derivative

w.r.t. θ_j , $j = 1, 2, 3$, and let $C_{1,j}(L, \theta) = \sum_{i=1}^{\infty} C_{1,j,i}(\theta)L^i = \phi(L, \theta_0)(\partial/\partial\theta_j)[\psi(L, \theta)\Delta^{d-d_0}]$ denote the coefficients of the partial derivative of $\tilde{\varepsilon}_t(\theta)$ w.r.t. θ_j .

To derive the asymptotic distribution theory for the CSS estimator, a central limit theorem (CLT) is shown to hold for the score function at θ_0 . Next, it is proven that a UWLLN holds for the Hessian matrix by showing that the Hessian matrix and its first partial derivatives satisfy a WLLN (Wooldridge; 1994, thm. 4.2). The UWLLN allows to evaluate the Hessian matrix in (C.12) at θ_0 and yields the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$. As the reduced form coefficients $\phi(L)$ depend non-trivially on θ , no analytical expression for the asymptotic variance of the CSS estimator is provided. Instead, it will be shown that the CSS estimator is asymptotically normally distributed, and its asymptotic variance is shown to exist. This allows to estimate $\text{Var}(\hat{\theta})$ e.g. via the inverse of the numerical Hessian matrix.

Starting with the score function, similar to Nielsen (2015, p. 175) a CLT can be inferred from the Cramér-Wold device by showing that for any 3-dimensional vector $\mu = (\mu_1, \mu_2, \mu_3)'$, it holds that $\mu' S_n = \sum_{j=1}^3 \mu_j S_n^{(j)} \xrightarrow{d} N(0, \mu' \Omega_0 \mu)$. To see this, define the σ -algebra $\tilde{\mathcal{F}}_t = \sigma(\{\varepsilon_s, s \leq t\})$ generated by the white noise ε_t and its lags. Next, note that in (C.14) the term $\varepsilon_t[\partial\tilde{\varepsilon}_t(\theta)/\partial\theta]_{\theta=\theta_0}$ adapted to $\tilde{\mathcal{F}}_t$ is a stationary MDS, since ε_t is white noise, the partial derivatives are $\tilde{\mathcal{F}}_{t-1}$ -measurable, and the coefficients of the partial derivatives are absolutely summable, as shown in the proof of theorem 3.2. It follows for $\mu' S_n = 2n^{-1/2} \sum_{t=1}^n \nu_t$ with

$$\nu_t = \sum_{j=1}^3 \nu_{j,t} \quad \nu_{j,t} = \mu_j \varepsilon_t \left. \frac{\partial \tilde{\varepsilon}_t(\theta)}{\partial \theta_j} \right|_{\theta=\theta_0},$$

that ν_t adapted to $\tilde{\mathcal{F}}_t$ is a stationary MDS. Similar to Nielsen (2015, p. 175), by the law of large numbers for stationary and ergodic processes, the sum of conditional variances for $\mu' S_n$ with S_n as given in (C.14) is then

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \text{E}(\nu_t^2 | \tilde{\mathcal{F}}_{t-1}) &= \frac{1}{n} \sum_{t=1}^n \sum_{j,k=1}^3 \text{E} \left[\nu_{j,t} \nu_{k,t} | \tilde{\mathcal{F}}_{t-1} \right] = \sum_{j,k=1}^3 \mu_j \mu_k \sigma_{\varepsilon,0}^2 \frac{1}{n} \sum_{t=1}^n \left. \frac{\partial \tilde{\varepsilon}_t(\theta)}{\partial \theta_j} \right|_{\theta=\theta_0} \left. \frac{\partial \tilde{\varepsilon}_t(\theta)}{\partial \theta_k} \right|_{\theta=\theta_0} \\ &\xrightarrow{p} \sum_{j,k=1}^3 \mu_j \mu_k \sigma_{\varepsilon,0}^4 \sum_{i=1}^{\infty} C_{1,j,i}(\theta_0) C_{1,k,i}(\theta_0) = \sum_{j,k=1}^3 \mu_j \mu_k \Omega_0^{(j,k)}. \end{aligned} \quad (\text{C.15})$$

In $C_{1,j}(\theta_0) = \phi(L, \theta_0)(\partial/\partial\theta_j)[\psi(L, \theta)\Delta^{d-d_0}]_{\theta=\theta_0}$, the partial derivatives of the first polynomial $\partial\psi(L, \theta)/\partial\theta_j = -2\psi(L, \theta)(\partial\phi(L, \theta)/\partial\theta_j)$ are absolutely summable for all $j = 1, 2, 3$, as $\psi(L, \theta)$ and $\partial\phi(L, \theta)/\partial\theta_j$ are absolutely summable, see lemma D.1 in appendix D. Furthermore, $(\partial/\partial d)\Delta^{d-d_0}|_{\theta=\theta_0} = \sum_{j=1}^{\infty} j^{-1}L^j$ (Nielsen; 2015, p. 175), so that $\sum_{i=1}^{\infty} C_{1,j,i}(\theta_0)C_{1,k,i}(\theta_0) = O(1)$. Consequently, by the CLT for stationary MDS (see e.g. Davidson; 2000, thm. 6.2.3) $S_n \xrightarrow{d} N(0, 4\Omega_0)$.

To evaluate the Hessian matrix in (C.12) at θ_0 , it remains to be shown that a UWLLN applies to the Hessian matrix (Wooldridge; 1994, thm. 4.4), for which it is sufficient to show that a WLLN

holds for the Hessian matrix and for the supremum of its first partial derivatives

$$\sup_{\theta \in \Theta^*(\kappa)} \left| \frac{\partial^3 R_t(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| = O_p(1), \quad j, k, l = 1, 2, 3, \quad (\text{C.16})$$

for any fixed $\kappa \in (0, 1/2)$, see Newey (1991, cor. 2.2) and Wooldridge (1994, thm. 4.2).

The Hessian matrix can be derived from (C.9) and is given by

$$H(\theta) = \frac{\partial^2 R(\theta)}{\partial \theta \partial \theta'} = \frac{2}{n} \sum_{t=1}^n \left[\frac{\partial \varepsilon_t(\theta)}{\partial \theta} \frac{\partial \varepsilon_t(\theta)}{\partial \theta'} + \varepsilon_t(\theta) \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta \partial \theta'} \right], \quad (\text{C.17})$$

and a WLLN holds for the Hessian matrix if the absolute summability condition for (C.8) is satisfied by the two different terms of the Hessian matrix. Since the coefficients of the first partial derivatives of $\varepsilon_t(\theta)$ were shown to be absolutely summable in the proof of theorem 3.2 for $\theta \in \Theta^*(\kappa)$, the first term in (C.17) directly satisfies the condition for (C.8) and thus is bounded in probability. It remains to be shown that absolute summability holds for the coefficients of $\partial^2 \varepsilon_t(\theta)/(\partial \theta \partial \theta')$. From (C.9)

$$\begin{aligned} \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta_j \partial \theta_k} &= \frac{\partial \psi_+(L, \theta)}{\partial \theta_j} \frac{\partial \Delta_+^{d-d_0}}{\partial \theta_k} \Delta_+^{d_0} y_t + \frac{\partial \psi_+(L, \theta)}{\partial \theta_k} \frac{\partial \Delta_+^{d-d_0}}{\partial \theta_j} \Delta_+^{d_0} y_t \\ &\quad + \psi_+(L, \theta) \frac{\partial^2 \Delta_+^{d-d_0}}{\partial \theta_j \partial \theta_k} \Delta_+^{d_0} y_t + \frac{\partial^2 \psi_+(L, \theta)}{\partial \theta_j \partial \theta_k} \Delta_+^d y_t, \end{aligned} \quad (\text{C.18})$$

for $j, k = 1, 2, 3$. The coefficients in $\partial \psi_+(L, \theta)/\partial \theta_j$ were already shown to be absolutely summable in the proof of theorem 3.2, and thus the first and second term in (C.18) satisfy the absolute summability condition for (C.8). As the coefficients in $\psi(L, \theta)$ are absolutely summable, the third term in (C.18) is also bounded by (C.8), so that only the coefficients of the second partial derivatives of $\psi_+(L, \theta)$ need to be shown to be absolutely summable. As their sum is bounded by the sum of absolute coefficients of the untruncated polynomial $\psi(L, \theta) = \phi(L, \theta)^{-1}$, it is sufficient to prove absolute summability for the latter. For this, consider

$$\frac{\partial^2 \psi(L, \theta)}{\partial \theta_j \partial \theta_k} = 2\psi(L, \theta)^3 \frac{\partial \phi(L, \theta)}{\partial \theta_j} \frac{\partial \phi(L, \theta)}{\partial \theta_k} - \psi(L, \theta)^2 \frac{\partial^2 \phi(L, \theta)}{\partial \theta_j \partial \theta_k}, \quad j, k = 1, 2, 3, \quad (\text{C.19})$$

where the coefficients of first and second partial derivatives of $\phi(L, \theta)$ are shown to be absolutely summable in lemma D.1 in appendix D. Thus, (C.18) satisfies the absolute summability condition for (C.8), so that the Hessian matrix (C.17) satisfies a WLLN.

To prove (C.16), consider

$$\frac{\partial^3 R(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} = \frac{2}{n} \sum_{t=1}^n \left[\frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta_j \partial \theta_k} \frac{\partial \varepsilon_t(\theta)}{\partial \theta_l} + \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta_j \partial \theta_l} \frac{\partial \varepsilon_t(\theta)}{\partial \theta_k} + \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta_k \partial \theta_l} \frac{\partial \varepsilon_t(\theta)}{\partial \theta_j} + \varepsilon_t(\theta) \frac{\partial^3 \varepsilon_t(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right],$$

$j, k, l = 1, 2, 3$, where absolute summability of the coefficients of the first three terms was already

shown. Consequently, for the last term to also satisfy the condition for (C.8), the coefficients of the third partial derivatives of $\varepsilon_t(\theta)$ need to be shown to be absolutely summable. The derivatives are

$$\frac{\partial^3 \varepsilon_t(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} = \frac{\partial^3 \psi_+(L, \theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \Delta_+^d y_t + \psi_+(L, \theta) \frac{\partial^3 \Delta_+^{d-d_0}}{\partial \theta_j \partial \theta_k \partial \theta_l} \Delta_+^{d_0} y_t + r_t(\theta), \quad (\text{C.20})$$

and $r_t(\theta)$ holds the products of first and second partial derivatives of $\psi(L, \theta)$ and $\Delta_+^{d-d_0}$ that have already been shown to satisfy the absolute summability condition for (C.8). The second term in (C.20) directly satisfies the condition for (C.8), so that only the first term remains to be checked. As before, the partial derivatives of the untruncated polynomial are considered, as they are an upper bound for the sum of absolute coefficients of the truncated polynomial. From (C.19)

$$\begin{aligned} \frac{\partial^3 \psi(L, \theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} = & 2\psi(L, \theta)^3 \left[\frac{\partial^2 \phi(L, \theta)}{\partial \theta_j \partial \theta_k} \frac{\partial \phi(L, \theta)}{\partial \theta_l} + \frac{\partial^2 \phi(L, \theta)}{\partial \theta_j \partial \theta_l} \frac{\partial \phi(L, \theta)}{\partial \theta_k} + \frac{\partial^2 \phi(L, \theta)}{\partial \theta_k \partial \theta_l} \frac{\partial \phi(L, \theta)}{\partial \theta_j} \right] \\ & - 6\psi(L, \theta)^4 \frac{\partial \phi(L, \theta)}{\partial \theta_j} \frac{\partial \phi(L, \theta)}{\partial \theta_k} \frac{\partial \phi(L, \theta)}{\partial \theta_l} - \psi(L, \theta)^2 \frac{\partial^3 \phi(L, \theta)}{\partial \theta_j \partial \theta_k \partial \theta_l}, \quad j, k, l = 1, 2, 3. \end{aligned}$$

Absolute summability of the coefficients of the partial derivatives of $\phi(L, \theta)$ up to order three is shown in lemma D.1 in appendix D. Consequently, (C.20) satisfies the absolute summability condition for (C.8), so that (C.16) holds. Thus, a UWLLN holds for the Hessian matrix, so that pointwise convergence generalizes to weak convergence. This, together with consistency of $\hat{\theta}$ (see theorem 3.2) allows to evaluate the Hessian matrix in (C.12) at θ_0 . Analogously to (C.13), it follows from the argument of Robinson (2006, pp. 135-136) that the partial derivatives of $\varepsilon_t(\theta)$ in (C.17) can be replaced by those of $\tilde{\varepsilon}_t(\theta)$ as $n \rightarrow \infty$, and $\varepsilon_t(\theta_0)$ can be replaced by ε_t , which yields

$$\frac{\partial^2 R_t(\theta)}{\partial \theta_j \partial \theta_k} \Big|_{\theta=\theta_0} = \frac{2}{n} \sum_{t=1}^n \left[\frac{\partial \varepsilon_t(\theta)}{\partial \theta_j} \Big|_{\theta=\theta_0} \frac{\partial \varepsilon_t(\theta)}{\partial \theta_k} \Big|_{\theta=\theta_0} + \varepsilon_t(\theta_0) \frac{\partial^2 \varepsilon_t(\theta)}{\partial \theta_j \partial \theta_k} \Big|_{\theta=\theta_0} \right] \xrightarrow{p} 2\Omega_0^{(j,k)}, \quad (\text{C.21})$$

as $n \rightarrow \infty$. The second term converges to zero in probability, as the second partial derivatives are $\tilde{\mathcal{F}}_{t-1}$ -measurable, and thus the second term adapted to $\tilde{\mathcal{F}}_{t-1}$ is a stationary MDS.

Solving (C.12) for $\sqrt{n}(\hat{\theta} - \theta_0)$ and plugging in the limits for first and second partial derivatives yields

$$\sqrt{n}(\hat{\theta} - \theta_0) = H_t(\bar{\theta})^{-1} \frac{1}{\sqrt{n}} \frac{\partial R(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} \xrightarrow{d} \text{N}(0, \Omega_0^{-1}), \quad (\text{C.22})$$

as $n \rightarrow \infty$, which completes the proof. \square

D Partial derivatives of $\phi(L, \theta)$

Lemma D.1 (Absolute summability of partial derivatives). *For $\phi(L, \theta)$ in*

$$\phi(L, \theta) \sigma_\varepsilon \varepsilon_t^* = \sigma_\varepsilon \sum_{i=0}^{t-1} \phi_i(\theta) \varepsilon_{t-i}^* = \sigma_\eta \eta_t^* + \Delta_+^d \sigma_u u_t^* = \sigma_\eta \eta_t^* + \sigma_u \sum_{i=0}^{t-1} \pi_i(d) u_{t-i}^*, \quad (\text{D.1})$$

with $\varepsilon_t^* \sim \text{WN}(0, 1)$, $u_t^* \sim \text{WN}(0, 1)$, $\eta_t^* \sim \text{WN}(0, 1)$, $\phi_0(\theta) = 1$, it holds that

$$\lim_{t \rightarrow \infty} \sum_{i=1}^{t-1} \left| \frac{\partial \phi_i(\theta)}{\partial \theta_j} \right| < \infty, \quad (\text{D.2})$$

$$\lim_{t \rightarrow \infty} \sum_{i=1}^{t-1} \left| \frac{\partial^2 \phi_i(\theta)}{\partial \theta_j \partial \theta_k} \right| < \infty, \quad (\text{D.3})$$

$$\lim_{t \rightarrow \infty} \sum_{i=1}^{t-1} \left| \frac{\partial^3 \phi_i(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| < \infty, \quad (\text{D.4})$$

for all $j, k, l = 1, 2, 3$, and all $\theta \in \Theta$, where θ_j denotes the j -th entry of $\theta = (d, \sigma_\eta^2, \sigma_u^2)'$.

Proof of lemma D.1. The following results are required to prove (D.2) to (D.4). For σ_ε^2 , note that by solving the variance of (D.1) for σ_ε^2

$$\sigma_\varepsilon^2 = \frac{\sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d)}{\sum_{i=0}^{t-1} \phi_i^2(\theta)}. \quad (\text{D.5})$$

Since $\partial^j \pi_i(d) / \partial d^j = O(i^{-d-1}(1 + \log i)^j)$ for all $i \geq 1$, $j \geq 0$, see Johansen and Nielsen (2010, lemma B.3), and thus $\lim_{t \rightarrow \infty} \sum_{i=1}^{t-1} |\partial^j \pi_i(d) / \partial d^j| < \infty$ for all $j \geq 0$, it follows that

$$\frac{\partial}{\partial \theta_j} \left[\sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d) \right] = O(1), \quad (\text{D.6})$$

$$\frac{\partial^2}{\partial \theta_j \partial \theta_k} \left[\sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d) \right] = O(1), \quad (\text{D.7})$$

$$\frac{\partial^3}{\partial \theta_j \partial \theta_k \partial \theta_l} \left[\sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d) \right] = O(1), \quad (\text{D.8})$$

for all $j, k, l = 1, 2, 3$. For the same reason, it follows from (D.1) that

$$\frac{\partial \phi(L, \theta) \sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j} = O(1) \varepsilon_t^* + \sum_{i=1}^{t-1} O(i^{-d-1}(1 + \log i)) \varepsilon_{t-i}^*, \quad (\text{D.9})$$

$$\frac{\partial^2 \phi(L, \theta) \sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j \partial \theta_k} = O(1) \varepsilon_t^* + \sum_{i=1}^{t-1} O(i^{-d-1}(1 + \log i)^2) \varepsilon_{t-i}^*, \quad (\text{D.10})$$

$$\frac{\partial^3 \phi(L, \theta) \sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j \partial \theta_k \partial \theta_l} = O(1) \varepsilon_t^* + \sum_{i=1}^{t-1} O(i^{-d-1} (1 + \log i)^3) \varepsilon_{t-i}^*, \quad (\text{D.11})$$

and the limits stem from the first, second and third partial derivatives of $\sigma_\eta \eta_t^* + \sigma_u \sum_{i=0}^{t-1} \pi_i(d) u_{t-i}^*$ w.r.t. d , while all coefficients of the other partial derivatives are bounded below. Consequently, (D.9) to (D.11) are MA processes with absolutely summable coefficients. Note that this is not sufficient for absolute summability of the partial derivatives of $\phi(L, \theta)$, as σ_ε in the numerators of (D.9) to (D.11) also depends on θ .

For (D.2), consider $\partial \sigma_\varepsilon^2 / \partial \theta_j = c_1(\theta, \theta_j) - c_2(\theta, \theta_j)$, where

$$c_1(\theta, \theta_j) = \frac{\frac{\partial}{\partial \theta_j} \left[\sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d) \right]}{\sum_{i=0}^{t-1} \phi_i^2(\theta)} = O(1), \quad c_2(\theta, \theta_j) = \frac{2\sigma_\varepsilon^2 \sum_{i=1}^{t-1} \phi_i(\theta) \frac{\partial \phi_i(\theta)}{\partial \theta_j}}{\sum_{i=0}^{t-1} \phi_i^2(\theta)}, \quad (\text{D.12})$$

and the first term is $O(1)$ due to (D.6). For the partial derivative of $\phi(L, \theta) \sigma_\varepsilon \varepsilon_t^*$ one then has

$$\begin{aligned} \frac{\partial \phi(L, \theta) \sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j} &= \sigma_\varepsilon \sum_{i=1}^{t-1} \frac{\partial \phi_i(\theta)}{\partial \theta_j} \varepsilon_{t-i}^* + \frac{1}{2\sigma_\varepsilon} \frac{\partial \sigma_\varepsilon^2}{\partial \theta_j} \phi(L, \theta) \varepsilon_t^* \\ &= \frac{c_1(\theta, \theta_j)}{2\sigma_\varepsilon} \phi(L, \theta) \varepsilon_t^* - \frac{c_2(\theta, \theta_j)}{2\sigma_\varepsilon} \phi(L, \theta) \varepsilon_t^* + \sigma_\varepsilon \sum_{i=1}^{t-1} \frac{\partial \phi_i(\theta)}{\partial \theta_j} \varepsilon_{t-i}^*. \end{aligned} \quad (\text{D.13})$$

From (D.9) it follows that the term on the left hand side (LHS) is a MA process with absolutely summable coefficients for any t . Since the same holds for $\phi(L, \theta) \varepsilon_t^*$, by (D.12) the first term on the right hand side (RHS) is also a MA process with absolutely summable coefficients. Consequently, the difference of the latter two terms on the RHS

$$\sigma_\varepsilon \sum_{i=1}^{t-1} \frac{\partial \phi_i(\theta)}{\partial \theta_j} \varepsilon_{t-i}^* - \frac{c_2(\theta, \theta_j)}{2\sigma_\varepsilon} \phi(L, \theta) \varepsilon_t^* = \frac{-c_2(\theta, \theta_j)}{2\sigma_\varepsilon} \varepsilon_t^* + \sum_{i=1}^{t-1} \left[\sigma_\varepsilon \frac{\partial \phi_i(\theta)}{\partial \theta_j} - \frac{c_2(\theta, \theta_j) \phi_i(\theta)}{2\sigma_\varepsilon} \right] \varepsilon_{t-i}^*,$$

is also a MA process with absolutely summable coefficients. As the contemporaneous impact of ε_t^* cannot cancel, it follows that $c_2(\theta, \theta_j) = O(1)$ is bounded, and thus the second term on the RHS of (D.13) is a MA process with absolutely summable coefficients. For the equality in (D.13) to hold, it must thus hold that $\sigma_\varepsilon \sum_{i=1}^{t-1} (\partial \phi_i(\theta) / \partial \theta_j) \varepsilon_{t-i}^*$ is also a MA process with absolutely summable coefficients for any t , which proves (D.2).

For (D.3) one has $\partial^2 \sigma_\varepsilon^2 / (\partial \theta_j \partial \theta_k) = c_3(\theta, \theta_j, \theta_k) - c_4(\theta, \theta_j, \theta_k)$ with

$$\begin{aligned} c_3(\theta, \theta_j, \theta_k) &= \left\{ \frac{\partial^2}{\partial \theta_j \partial \theta_k} \left[\sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d) \right] - 2 \frac{\partial \sigma_\varepsilon^2}{\partial \theta_k} \sum_{i=1}^{t-1} \phi_i(\theta) \frac{\partial \phi_i(\theta)}{\partial \theta_j} \right\} \left[\sum_{i=0}^{t-1} \phi_i^2(\theta) \right]^{-1} \\ &\quad - \left\{ 2 \frac{\partial \sigma_\varepsilon^2}{\partial \theta_j} \sum_{i=1}^{t-1} \phi_i(\theta) \frac{\partial \phi_i(\theta)}{\partial \theta_k} + 2\sigma_\varepsilon^2 \sum_{i=1}^{t-1} \frac{\partial \phi_i(\theta)}{\partial \theta_j} \frac{\partial \phi_i(\theta)}{\partial \theta_k} \right\} \left[\sum_{i=0}^{t-1} \phi_i^2(\theta) \right]^{-1}, \end{aligned} \quad (\text{D.14})$$

$$c_4(\theta, \theta_j, \theta_k) = \left[2\sigma_\varepsilon^2 \sum_{i=1}^{t-1} \phi_i(\theta) \frac{\partial^2 \phi_i(\theta)}{\partial \theta_j \partial \theta_k} \right] \left[\sum_{i=0}^{t-1} \phi_i^2(\theta) \right]^{-1}, \quad (\text{D.15})$$

and $c_3(\theta, \theta_j, \theta_k) = O(1)$ is bounded due to (D.2) and (D.7). The second partial derivatives of $\phi(L, \theta)\sigma_\varepsilon \varepsilon_t^*$ are

$$\begin{aligned} \frac{\partial^2 \phi(L, \theta)\sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j \partial \theta_k} &= z_1(\theta, \theta_j, \theta_k) + \frac{1}{2\sigma_\varepsilon} \frac{\partial^2 \sigma_\varepsilon^2}{\partial \theta_j \partial \theta_k} \phi(L, \theta)\varepsilon_t^* + \sigma_\varepsilon \sum_{i=1}^{t-1} \frac{\partial^2 \phi_i(\theta)}{\partial \theta_j \partial \theta_k} \varepsilon_{t-i}^*, \\ z_1(\theta, \theta_j, \theta_k) &= \frac{1}{2\sigma_\varepsilon} \frac{\partial \sigma_\varepsilon^2}{\partial \theta_k} \sum_{i=1}^{t-1} \frac{\partial \phi_i(\theta)}{\partial \theta_j} \varepsilon_{t-i}^* + \frac{1}{2\sigma_\varepsilon} \frac{\partial \sigma_\varepsilon^2}{\partial \theta_j} \sum_{i=1}^{t-1} \frac{\partial \phi_i(\theta)}{\partial \theta_k} \varepsilon_{t-i}^* - \frac{1}{4\sigma_\varepsilon^3} \frac{\partial \sigma_\varepsilon^2}{\partial \theta_j} \frac{\partial \sigma_\varepsilon^2}{\partial \theta_k} \phi(L, \theta)\varepsilon_t^*, \end{aligned} \quad (\text{D.16})$$

and $z_1(\theta, \theta_j, \theta_k)$ is a MA process with absolutely summable coefficients due to (D.2). Plugging in $\partial^2 \sigma_\varepsilon^2 / (\partial \theta_j \partial \theta_k) = c_3(\theta, \theta_j, \theta_k) - c_4(\theta, \theta_j, \theta_k)$ and rearranging terms yields

$$\begin{aligned} \frac{\partial^2 \phi(L, \theta)\sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j \partial \theta_k} - \frac{c_3(\theta, \theta_j, \theta_k)}{2\sigma_\varepsilon} \phi(L, \theta)\varepsilon_t^* - z_1(\theta, \theta_j, \theta_k) &= -\frac{c_4(\theta, \theta_j, \theta_k)}{2\sigma_\varepsilon} \varepsilon_t^* \\ &+ \sum_{i=1}^{t-1} \left[\sigma_\varepsilon \frac{\partial^2 \phi_i(\theta)}{\partial \theta_j \partial \theta_k} - \frac{c_4(\theta, \theta_j, \theta_k)}{2\sigma_\varepsilon} \phi_i(\theta) \right] \varepsilon_{t-i}^*, \end{aligned}$$

where the LHS is a MA process with absolutely summable coefficients for any t due to (D.10) and (D.14). Again, as the contemporaneous ε_t^* cannot cancel out, $c_4(\theta, \theta_j, \theta_k) = O(1)$ is bounded. Therefore, $c_4(\theta, \theta_j, \theta_k)/(2\sigma_\varepsilon)\phi(L, \theta)\varepsilon_t^*$ is a MA process with absolutely summable weights, so that for the equality above to hold, $\sum_{i=1}^{t-1} \partial^2 \phi_i(\theta)/(\partial \theta_j \partial \theta_k)\varepsilon_{t-i}^*$ must also be a MA process with absolutely summable weights for any t , which proves (D.3).

Turning to (D.4), the third partial derivatives of the variance parameter σ_ε^2 can be represented as $\partial^3 \sigma_\varepsilon^2 / (\partial \theta_j \partial \theta_k \partial \theta_l) = c_5(\theta, \theta_j, \theta_k, \theta_l) - c_6(\theta, \theta_j, \theta_k, \theta_l)$ with

$$c_6(\theta, \theta_j, \theta_k, \theta_l) = \left[2\sigma_\varepsilon^2 \sum_{i=1}^{t-1} \phi_i(\theta) \frac{\partial^3 \phi_i(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right] \left[\sum_{i=0}^{t-1} \phi_i^2(\theta) \right]^{-1}. \quad (\text{D.17})$$

$c_5(\theta, \theta_j, \theta_k, \theta_l)$ holds the products of first and second partial derivatives of σ_ε^2 and $\phi(1, \theta)$ that have already been shown to be $O(1)$, as well as $\partial^3 / (\partial \theta_j \partial \theta_k \partial \theta_l) \left[\sigma_\eta^2 + \sigma_u^2 \sum_{i=0}^{t-1} \pi_i^2(d) \right]$ that is $O(1)$ as shown in (D.8). Consequently $c_5(\theta, \theta_j, \theta_k, \theta_l) = O(1)$, and the exact expression is omitted for brevity. The third partial derivatives of $\phi(L, \theta)\sigma_\varepsilon \varepsilon_t^*$ follow from (D.16) and equal

$$\frac{\partial^3 \phi(L)\sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j \partial \theta_k \partial \theta_l} = z_2(\theta, \theta_j, \theta_k, \theta_l) + \frac{1}{2\sigma_\varepsilon} \frac{\partial^3 \sigma_\varepsilon^2}{\partial \theta_j \partial \theta_k \partial \theta_l} \phi(L, \theta)\varepsilon_t^* + \sigma_\varepsilon \sum_{i=1}^{t-1} \frac{\partial^3 \phi_i(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \varepsilon_{t-i}^*, \quad (\text{D.18})$$

where $z_2(\theta, \theta_j, \theta_k, \theta_l)$ holds the products of the first and second partial derivatives of σ_ε^2 and $\phi(L, \theta)$ for which absolute summability was shown above. Therefore, $z_2(\theta, \theta_j, \theta_k, \theta_l)$ is a MA process with absolutely summable coefficients. Plugging in $\partial^3 \sigma_\varepsilon^2 / (\partial \theta_j \partial \theta_k \partial \theta_l) = c_5(\theta, \theta_j, \theta_k, \theta_l) - c_6(\theta, \theta_j, \theta_k, \theta_l)$

and rearranging gives

$$\begin{aligned} \frac{\partial^3 \phi(L, \theta) \sigma_\varepsilon \varepsilon_t^*}{\partial \theta_j \partial \theta_k \partial \theta_l} - \frac{c_5(\theta, \theta_j, \theta_k, \theta_l)}{2\sigma_\varepsilon} \phi(L, \theta) \varepsilon_t^* - z_2(\theta, \theta_j, \theta_k, \theta_l) &= -\frac{c_6(\theta, \theta_j, \theta_k, \theta_l)}{2\sigma_\varepsilon} \varepsilon_t^* \\ &+ \sum_{i=1}^{t-1} \left[\sigma_\varepsilon \frac{\partial^3 \phi_i(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} - \frac{c_6(\theta, \theta_j, \theta_k, \theta_l)}{2\sigma_\varepsilon} \phi_i(\theta) \right] \varepsilon_{t-i}^*, \end{aligned}$$

where the LHS is a MA process with absolutely summable coefficients for any t by (D.11). As for the first and second partial derivatives, $c_6(\theta, \theta_j, \theta_k, \theta_l) = O(1)$ holds, as the contemporaneous ε_t^* do not cancel on the RHS. Due to boundedness of $c_6(\theta, \theta_j, \theta_k, \theta_l)$, the term $c_6(\theta, \theta_j, \theta_k, \theta_l) = O(1) \phi(L, \theta) \varepsilon_t^*$ is a MA process with absolutely summable weights. Since all other terms are MA processes with absolutely summable weights, $\sum_{i=1}^{t-1} \partial^3 \phi_i(\theta) / (\partial \theta_j \partial \theta_k \partial \theta_l) \varepsilon_{t-i}^*$ must also be a MA process with absolutely summable coefficients for the above equality to hold. This proves (D.4). \square

References

- Acemoglu, D., Chernozhukov, V., Werning, I. and Whinston, M. D. (2020). Optimal targeted lockdowns in a multi-group SIR model, *NBER Working Paper 27102*, National Bureau of Economic Research.
URL: <https://ideas.repec.org/p/nbr/nberwo/27102.html>
- Avery, C., Bossert, W., Clark, A., Ellison, G. and Ellison, S. F. (2020). An economist’s guide to epidemiology models of infectious disease, *Journal of Economic Perspectives* **34**(4): 79–104.
- BBC News (2020a). Coronavirus: Germany restricts social life in ‘lockdown light’. 2 November 2020. (Accessed 15 January 2021).
URL: <https://www.bbc.com/news/world-europe-54778345>
- BBC News (2020b). Coronavirus: Germany to go into lockdown over Christmas. 13 December 2020. (Accessed 15 January 2021).
URL: <https://www.bbc.com/news/world-europe-55292614>
- BBC News (2020c). Coronavirus: Italy’s Conte offers hope as travel restrictions end. 3 June 2020. (Accessed 15 January 2021).
URL: <https://www.bbc.com/news/world-europe-52914089>
- BBC News (2020d). Coronavirus: What went wrong at Germany’s Gütersloh meat factory. 25 June 2020. (Accessed 15 January 2021).
URL: <https://www.bbc.com/news/world-europe-53177628>
- Bergman, A., Sella, Y., Agre, P. and Casadevall, A. (2020). Oscillations in U.S. COVID-19 incidence and mortality data reflect diagnostic and reporting factors, *mSystems* **5**(4).
- Chang, Y., Miller, J. I. and Park, J. Y. (2009). Extracting a common stochastic trend: Theory with some applications, *Journal of Econometrics* **150**(2): 231–247.
- Chernozhukov, V., Kasahara, H. and Schrimpf, P. (2021). Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S., *Journal of Econometrics* **220**(1): 23–62.
- Clark, P. K. (1987). The cyclical component of U.S. economic activity, *The Quarterly Journal of Economics* **102**(4): 797–814.
- Davidson, J. (2000). *Econometric Theory*, Blackwell Publishers.
- Deutsche Welle (2020a). Coronavirus digest: Europe toughens restrictions as cases rise. 13 October 2020. (Accessed 15 January 2021).
URL: <https://www.dw.com/en/coronavirus-digest-europe-tightens-restrictions-as-cases-rise/a-55268548>
- Deutsche Welle (2020b). Coronavirus: Germany toughens restrictions as it enters ‘decisive’ phase. 14 October 2020. (Accessed 15 January 2021).
URL: <https://www.dw.com/en/coronavirus-germany-toughens-restrictions-as-it-enters-decisive-phase/a-55273942>
- Deutsche Welle (2020c). Italy toughens coronavirus measures amid second wave surge. 25 October 2020. (Accessed 15 January 2021).

URL: <https://www.dw.com/en/italy-toughens-coronavirus-measures-amid-second-wave-surge/a-55392933>

- Diebold, F. X. and Rudebusch, G. D. (1991). Is consumption too smooth? Long memory and the Deaton paradox, *The Review of Economics and Statistics* **73**(1): 1–9.
- Dong, E., Du, H. and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time, *The Lancet Infectious Diseases* **20**(5): 533–534.
- Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*, Oxford University Press, Oxford.
- Felbermayr, G., Hinz, J. and Chowdhry, S. (2020). Après-ski: The spread of Coronavirus from Ischgl through Germany, *Covid Economics: Vetted and Real-Time Papers* **22**: 177–204.
- Granger, C. W. J. and Morris, M. J. (1976). Time series modelling and interpretation, *Journal of the Royal Statistical Society. Series A (General)* **139**(2): 246–257.
- Hartl, T., Tschernig, R. and Weber, E. (2020). Fractional trends and cycles in macroeconomic time series, *arXiv:2005.05266v2* .
URL: <https://arxiv.org/abs/2005.05266v2>
- Hartl, T., Wälde, K. and Weber, E. (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, A. C. (1985). Trends and cycles in macroeconomic time series, *Journal of Business & Economic Statistics* **3**(3): 216–227.
- Hethcote, H. W. (2000). The mathematics of infectious diseases, *SIAM Review* **42**(4): 599–653.
- Hodrick, R. J. and Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation, *Journal of Money, Credit and Banking* **29**(1): 1–16.
- Hortaçsu, A., Liu, J. and Schwieg, T. (2021). Estimating the fraction of unreported infections in epidemics with a known epicenter: An application to COVID-19, *Journal of Econometrics* **220**(1): 106–129.
- Hualde, J. and Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models, *The Annals of Statistics* **39**(6): 3152–3181.
- Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model, *Journal of Econometrics* **158**(1): 51–66.
- Korolev, I. (2021). Identification and estimation of the SEIRD epidemic model for COVID-19, *Journal of Econometrics* **220**(1): 63–85.
- Lee, S., Liao, Y., Seo, M. H. and Shin, Y. (2021). Sparse HP filter: Finding kinks in the COVID-19 contact rate, *Journal of Econometrics* **220**(1): 158–180.
- Liu, L., Moon, H. R. and Schorfheide, F. (2021). Panel forecasts of country-level Covid-19 infections, *Journal of Econometrics* **220**(1): 2–22.
- Marinucci, D. and Robinson, P. M. (1999). Alternative forms of fractional Brownian motion, *Journal of Statistical Planning and Inference* **80**(1–2): 111–122.

- McCoy, L. G., Smith, J., Anchuri, K., Berry, I., Pineda, J., Harish, V., Lam, A. T., Yi, S. E., Hu, S., COVID-19 Canada Open Data Working Group: Non-Pharmaceutical Interventions and Fine, B. (2020). CAN-NPI: A curated open dataset of Canadian non-pharmaceutical interventions in response to the global COVID-19 pandemic, *Working paper*, medRxiv.
URL: <https://doi.org/10.1101/2020.04.17.20068460>
- Morley, J. C., Nelson, C. R. and Zivot, E. (2003). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?, *The Review of Economics and Statistics* **85**(2): 235–243.
- Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity, *Econometrica* **59**(4): 1161–1167.
- Nielsen, M. Ø. (2015). Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time-series models, *Journal of Time Series Analysis* **36**(2): 154–188.
- Oh, K. H., Zivot, E. and Creal, D. (2008). The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics, *Journal of Econometrics* **146**(2): 207–219.
- Pindyck, R. S. (2020). COVID-19 and the welfare effects of reducing contagion, *NBER Working Paper 27121*, National Bureau of Economic Research.
URL: <https://ideas.repec.org/p/nbr/nberwo/27121.html>
- Robinson, P. M. (2006). Conditional-sum-of-squares estimation of models for stationary time series with long memory, in H.-C. Ho, C.-K. Ing and T. L. Lai (eds), *Time Series and Related Topics: In Memory of Ching-Zong Wei*, Vol. 52 of *IMS Lecture Notes-Monograph Series*, Institute of Mathematical Statistics, Beachwood, Ohio, pp. 130–137.
- Shimotsu, K. (2010). Exact local Whittle estimation of fractional integration with unknown mean and time trend, *Econometric Theory* **26**(2): 501–540.
- Signorelli, C., Scognamiglio, T. and Odone, A. (2020). Covid-19 in Italy: impact of containment measures and prevalence estimates of infection in the general population, *Acta Biomedica* **91**(3): 175–179.
- The COVID Tracking Project (2021). The “good” metric is pretty bad: Why it’s hard to count the people who have recovered from COVID-19. 13 January 2021. (Accessed 16 January 2021).
URL: <https://covidtracking.com/analysis-updates/why-its-hard-to-count-recovered>
- Tschernig, R., Weber, E. and Weigand, R. (2013). Fractionally integrated VAR models with a fractional lag operator and deterministic trends: Finite sample identification and two-step estimation, *Working Paper 471*, University of Regensburg, Regensburg.
URL: <https://epub.uni-regensburg.de/27269/>
- Watson, M. W. (1986). Univariate detrending methods with stochastic trends, *Journal of Monetary Economics* **18**(1): 49–75.
- Wooldridge, J. M. (1994). Estimation and inference for dependent processes, in R. F. Engle and D. McFadden (eds), *Handbook of Econometrics*, Vol. 4, Elsevier, Amsterdam, pp. 2639–2738.