The fractional unobserved components model: a generalization of trend-cycle decompositions to data of unknown persistence

Tobias Hartl*[a,b,c]

[a]University of Regensburg, 93053 Regensburg, Germany
[b]Institute for Employment Research, 90478 Nuremberg, Germany
[c]Princeton University, 08544 Princeton, New Jersey, United States

October 2023

Abstract. This paper provides a data-driven solution to the specification of long-run dynamics in trend-cycle decompositions by introducing a state space model of form $y_t = x_t + c_t$, where the trend $x_t \sim I(d)$ is fractionally integrated of order d, whereas c_t represents a stationary cyclical component. The model encompasses the literature that typically assumes $x_t \sim I(1)$, or $x_t \sim I(2)$, but also allows for intermediate solutions between integer-integrated specifications and thus for richer long-run dynamics. Trend and cycle are estimated via the Kalman filter, for which a closed-form solution is provided. The integration order d is treated as unknown and is estimated jointly with the other model parameters. The paper derives the asymptotic theory for parameter estimation under relatively mild assumptions. While the proofs are carried out for a prototypical model, the asymptotic theory carries over to generalizations allowing for deterministic terms and correlated innovations. An application to monthly sea surface temperature anomalies reveals a smooth, diverging trend component, together with a cyclical component that is closely coupled to the Oceanic Niño Index.

Keywords. Unobserved components, trend-cycle decomposition, state space models, Kalman filter, long memory

JEL-Classification. C32, C51, Q54

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

1 Introduction

The decomposition of time series into trend and cycle plays a key role in applied research. In modern trend-cycle models, the long-run dynamics, particularly the integration order of the trend, must be specified prior to estimation, which opens the door to model specification errors. This paper introduces an encompassing trend-cycle model that treats the integration order as unknown. It offers a flexible, robust, and data-driven approach to decomposing time series into trend and cycle, and is termed the fractional unobserved components model.[1]

The literature on trend-cycle decompositions has been shaped by the seminal works of Beveridge and Nelson (1981), Harvey (1985), Clark (1987), and Hodrick and Prescott (1997). Since then, a variety of unobserved components (UC) models have been proposed, and often the integration order of the trend was subject to debate. The field is divided into two main groups, one assuming the trend to be integrated of order one in the spirit of Beveridge and Nelson (1981) and Harvey (1985), the other group preferring an integration order of two as suggested by Clark (1987) and Hodrick and Prescott (1997). Since empirical results are sensitive to the choice of the integration order, a data-driven model selection procedure would clearly be beneficial. However, the literature to date lacks an encompassing model allowing for trends of different memory. Thus, model specification is left open to the applied researcher, who often faces a trade-off between the economic plausibility of the model specification and the economic plausibility of the resulting decomposition. Little is known about the consequences of model misspecification on the estimates of the unobserved components. In addition, the asymptotic estimation theory is not fully developed for UC models, particularly when shocks are not necessarily Gaussian.

This paper aims to bridge these gaps by introducing a novel UC model that specifies the stochastic trend component x_t as a fractionally integrated process of order $d \in \mathbb{R}_+$, denoted as $x_t \sim I(d)$. It allows for random walk trend components (as suggested among others by Beveridge and Nelson; 1981; Harvey; 1985; Morley et al.; 2003) for $d = 1$, but also includes quadratic stochastic trend specifications (e.g. those of Clark; 1987; Hodrick and Prescott; 1997; Oh et al.; 2008) for $d = 2$. Since the integration order d can take any value on the positive real line and enters the model as an unknown parameter to be estimated, the model seamlessly links integer-integrated specifications. By including non-integer d, it allows for even more general patterns of persistence between the integer cases. Besides the fractional trend, the fractional UC model includes a cyclical component that encompasses the ARMA specifications common in the UC literature, but also allows for a broader class of processes such as e.g. the exponential model of Bloomfield (1973). Long- and short-run innovations are assumed to be martingale difference sequences, which is somewhat more general than the usual Gaussian white noise assumption.

While the UC literature has mostly considered integer-integrated specifications, there are some generalizations to non-integer integration orders in the state space literature: For asymptotically stationary processes (i.e. $d < 1/2$) Chan and Palma (1998, 2006), Palma (2007) and Grassi and

[1] Note that the literature has come up with a variety of names for unobserved components models, such as structural time series models and trend-cycle models among others. To avoid confusion, the term unobserved components model will be used for any model that specifies one or more time series as a function of latent components and assigns an interpretation to these components by imposing assumptions on their spectra.

de Magistris (2014) consider approximations to long memory processes in state space form by truncating either the autoregressive or the moving average representation of the fractional differencing polynomial. Their models have been found valuable for realized volatility modeling (see Ray and Tsay; 2000; Chen and Hurvich; 2006; Harvey; 2007; Varneskov and Perron; 2018) but exclude non-stationary stochastic trends that are indispensable for general UC models. Recently, Hartl and Jucknewitz (2022) studied ARMA approximations to fractionally integrated processes in state space form, also including the non-stationary domain. So far, the literature has focused on approximate representations of fractionally integrated processes to reduce the computational burdens of the Kalman filter. In contrast, this paper suggests an exact state space representation and provides a closed-form solution to the Kalman filter, thereby avoiding the computationally costly Kalman recursions.

To also assess the theoretical properties of parameter estimation, this paper derives the estimation theory for both the unobserved components and the model parameters. In line with the UC literature, the unobserved components are estimated by minimizing the objective function of the Kalman filter. While the literature typically relies on iterative estimates for trend and cycle via the Kalman recursions, I derive an analytical solution to the optimization problem of the Kalman filter.[2] Since iterative and analytical solution differ only in the way they are computed, both approaches yield the minimum variance linear unbiased estimator for trend and cycle (Durbin and Koopman; 2012, lemma 2). However, using the analytical solution is computationally less expensive for the fractional UC model. As an additional advantage, it provides a closed-form solution to the objective function of the conditional sum-of-squares (CSS) estimator, which is used to estimate the model parameters. Under the assumption that long- and short-run shocks are stationary martingale difference sequences, the CSS estimator is shown to be consistent. Under the somewhat stronger assumption that the prediction error of the Kalman filter is also a martingale difference sequence, the CSS estimator is shown to be asymptotically normally distributed.

The proofs are complicated by non-ergodicity of the prediction errors and non-uniform convergence of the objective function. The latter is caused by a prediction error that is stationary when the estimate for d is close to the true value, while it becomes non-stationary when the estimate is too far off. While all proofs are carried out for the conditional sum-of-squares (CSS) estimator, they are shown to extend seamlessly to the quasi-maximum likelihood (QML) estimator that is typically used in the UC literature. Furthermore, estimation results are shown to also hold for models with deterministic terms and correlated trend and cycle innovations (as e.g. in Balke and Wohar; 2002; Morley et al.; 2003). The finite sample properties of the CSS and QML estimators are evaluated in a Monte Carlo study, which supports the results on consistency for both estimators. In addition, the parameter estimates for the integration order outperform the exact local Whittle estimator of Shimotsu and Phillips (2005), which is biased by the cyclical fluctuations.

An application to monthly sea surface temperature anomalies illustrates the benefits from the fractional UC model: Temperature anomalies are estimated to be integrated of order around 1.75, and the hypothesis of an integer integration order is rejected. The resulting trend-cycle decompo-

[2] Analytical solutions to the Kalman filter have been derived for trend plus noise models by Burman and Shumway (2009) and Chang et al. (2009), where the trend is a random walk and the cycle is white noise.

sition finds trend temperature anomalies to be increasing since the mid of the 20th century, while cyclical temperature anomalies closely match the Oceanic Niño Index.

The rest of the paper is organized as follows: Section 2 introduces the fractional UC model and discusses the underlying assumptions. Section 3 discusses trend and cycle estimation, while section 4 details parameter estimation. Generalizations of the fractional UC model are discussed in section 5. Section 6 examines the finite sample properties of the proposed methods in a Monte Carlo study, while section 7 applies the fractional UC model to sea surface temperature anomalies. Section 8 concludes. The proofs for consistency and asymptotic normality are contained in the appendix. The code for this paper, as well as a computationally efficient R package containing all necessary functions for fractional UC models, is available at `https://github.com/tobiashartl/fracUCM`.

2 Model

While the literature on unobserved components (UC) models is vast, it builds on a simple model that decomposes an observable time series $\{y_t\}_{t=1}^n$ into unobserved trend x_t and cycle c_t

$$y_t = x_t + c_t. \tag{1}$$

c_t and x_t are distinguished by their different spectral densities: The cycle (or short-run component) c_t is assumed to follow a mean zero stationary process to capture the transitory features of y_t. The trend (or long-run component) x_t is characterized by an autocovariance function that decays more slowly than with an exponential rate. It models the persistent features of the observable series and is allowed to be non-stationary.

I generalize state-of-the-art UC models by modeling x_t as a fractionally integrated process of unknown memory $d \in \mathbb{R}_+$

$$\Delta_+^d x_t = \eta_t. \tag{2}$$

The fractional difference operator Δ_+^d depends only on the parameter d and controls the memory of x_t. Without subscript, it exhibits a polynomial expansion in the lag operator L of order infinite

$$\Delta^d = (1-L)^d = \sum_{j=0}^{\infty} \pi_j(d) L^j, \qquad \pi_j(d) = \begin{cases} \frac{j-d-1}{j}\pi_{j-1}(d) & j = 1, 2, ..., \\ 1 & j = 0, \end{cases} \tag{3}$$

where the weights $\pi_j(d)$ are determined recursively. The motivation behind (2) and (3) is that the higher d, the greater the effect of a past shock η_{t-j} on x_t, and the more differencing is required to eliminate the persistent impact of the past shock via (2). For this reason $x_t \sim I(d)$ is said to have long memory whenever $d > 0$ (see Hassler; 2019, for more details). The +-subscript in (2) denotes the truncation of an operator at $t \leq 0$, $\Delta_+^d x_t = \Delta^d x_t \mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} \pi_j(d) x_{t-j}$, where $\mathbb{1}(t \geq 1)$ is the indicator function that takes the value one for positive subscripts of x_{t-j}, otherwise zero. The truncated fractional difference operator reflects the type II definition of fractionally integrated processes (Marinucci and Robinson; 1999) and is required to treat the asymptotically stationary case alongside the non-stationary case.

Equation (2) encompasses several trend specifications in the literature: For $d = 1$, it nests the random walk trend model as considered by Harvey (1985), Balke and Wohar (2002), and Morley et al. (2003) among others. For $d = 2$, one has the double-drift model of Clark (1987) and Oh et al. (2008), but also the filter of Hodrick and Prescott (1997, HP filter in what follows) as will become clear. For $d \in \mathbb{N}$, the model of Burman and Shumway (2009) is obtained. Allowing for $d \in \mathbb{R}_+$ seamlessly links these integer-integrated models and allows for far more general dynamics of the trend: For $0 < d < 1/2$, it covers stationary and strongly persistent processes as considered by Ray and Tsay (2000), Chen and Hurvich (2006), and Varneskov and Perron (2018) for realized volatility modeling. For $1/2 < d < 1$, it allows for non-stationary but mean-reverting processes, while $d \geq 1$ yields non-stationary non-mean-reverting processes that are indispensable for trend-cycle decompositions of macroeconomic variables among others. Since d enters the model as an unknown parameter to be estimated, the model allows for a data-driven choice of d and provides statistical inference on the appropriate specification of UC models.

Turning to the cyclical component, I treat c_t as any short memory process that is independent of x_t and may depend non-linearly on a parameter vector φ

$$c_t = a(L, \varphi)\epsilon_t = \sum_{j=0}^{\infty} a_j(\varphi)\epsilon_{t-j}. \tag{4}$$

The parametric form of $a(L, \varphi)$ is assumed to be known. For example, c_t may be an ARMA process as typically assumed in the UC literature, but the specification generally captures a broader class of processes, e.g. the exponential model of Bloomfield (1973).

In what follows, the model (1), (2), and (4) is analyzed under the following assumptions:

Assumption 1 (Errors). *The errors ϵ_t, η_t are stationary and ergodic with finite moments up to order four and absolutely summable autocovariance function. For the joint σ-algebra $\mathcal{F}_t = \sigma((\eta_s, \epsilon_s), s \leq t)$, it holds that $\mathrm{E}(\epsilon_t|\mathcal{F}_{t-1}) = 0$, $\mathrm{E}(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\epsilon^2$, and $\mathrm{E}(\eta_t|\mathcal{F}_{t-1}) = 0$, $\mathrm{E}(\eta_t^2|\mathcal{F}_{t-1}) = \sigma_\eta^2$. Furthermore, conditional on $\mathcal{F}_{t-1}$, the third and fourth moments of ϵ_t, η_t are finite and equal their unconditional moments. Finally, ϵ_t and η_t are independent.*

Assumption 2 (Parameters). *Collect all model parameters in $\psi = (d, \sigma_\eta^2, \sigma_\epsilon^2, \varphi')'$, and let $\Psi = D \times \Sigma_\eta \times \Sigma_\epsilon \times \Phi$ denote the parameter space of $\psi \in \Psi$, where $D = \{d \in \mathbb{R} | 0 < d_{min} \leq d \leq d_{max} < \infty\}$, $\Sigma_\eta = \{\sigma_\eta^2 \in \mathbb{R} | 0 < \sigma_{\eta,min}^2 \leq \sigma_\eta^2 \leq \sigma_{\eta,max}^2 < \infty\}$, $\Sigma_\epsilon = \{\sigma_\epsilon^2 \in \mathbb{R} | 0 < \sigma_{\epsilon,min}^2 \leq \sigma_\epsilon^2 \leq \sigma_{\epsilon,max}^2 < \infty\}$, and $\Phi \subseteq \mathbb{R}^q$ is convex and compact. Then for the true parameters $\psi_0 = (d_0, \sigma_{\eta,0}^2, \sigma_{\epsilon,0}^2, \varphi_0')'$ it holds that $\psi_0 \in \Psi$.*

Assumption 1 allows for conditionally homoscedastic martingale difference sequences (MDS) η_t and ϵ_t. This is somewhat more general than the UC literature, which typically assumes Gaussian white noise disturbances (e.g. in Morley et al.; 2003). The generalization is of great practical importance given the applications of UC models in macroeconomics and finance. Independence of the shocks is assumed to simplify the derivation of the asymptotic estimation theory in section 4, and can be relaxed to allow for correlated innovations, see subsection 5.2.

Assumption 2 allows for both, stationary and non-stationary fractionally integrated trend components, and for an arbitrarily large interval $d \in D$. Positive integration orders guarantee that x_t is a long-run component, and that it can be distinguished from c_t based on its spectrum.

Assumption 3 (Stability of $a(L, \varphi)$)**.** *For all $\varphi \in \Phi$ and all z in the complex unit disc $\{z \in \mathbb{C} : |z| \leq 1\}$ it holds that*

(i) $a_0(\varphi) = 1$, and $\sum_{j=0}^{\infty} |a_j(\varphi)|$ is bounded and bounded away from zero,

(ii) each element of $a(e^{i\lambda}, \varphi)$ is differentiable in λ with derivative in $\mathrm{Lip}(\zeta)$ for any $\zeta > 1/2$,

(iii) $a(z, \varphi) = \sum_{j=0}^{\infty} a_j(\varphi) z^j$ is continuously differentiable in φ, and the partial derivatives $\dot{a}(z, \varphi) = \sum_{j=1}^{\infty} \frac{\partial a_j(\varphi)}{\partial \varphi} z^j = \sum_{j=1}^{\infty} \dot{a}_j(\varphi) z^j$ satisfy $\dot{a}_j(\varphi) = O(j^{-1-\zeta})$, and $\frac{\partial a_0(\varphi)}{\partial \varphi} = 0$.

Under assumption 3, $a(L, \varphi)^{-1} = b(L, \varphi) = \sum_{j=0}^{\infty} b_j(\varphi) L^j$ exists, is well defined, and the sum $\sum_{j=0}^{\infty} |b_j(\varphi)|$ is bounded and bounded away from zero. By the Lipschitz condition it holds that

$$a_j(\varphi) = O(j^{-1-\zeta}), \qquad b_j(\varphi) = O(j^{-1-\zeta}), \qquad \text{uniformly in } \varphi \in \Phi.$$

The rate for $a_j(\varphi)$ follows directly from assumption 3(ii), while that for $b_j(\varphi)$ follows from Zygmund (2002, pp. 46 and 71). The convergence rate for the partial derivative $\dot{a}_j(\varphi)$ is a direct consequence of compactness of Φ and continuity of $\partial a_j(\varphi)/\partial \varphi'$. Assumption 3 imposes some smoothness on the linear coefficients in $a(L, \varphi)$, and thus also on $b(L, \varphi)$. It is satisfied by any stationary and invertible ARMA process. For ARFIMA models, the asymptotic estimation theory is well established under assumptions similar to 1, 2, and 3, see Hualde and Robinson (2011) and Nielsen (2015).

3 Filtering and smoothing

The system introduced in (1), (2), and (4) forms a state space model, where (1) is the measurement equation and (2), (4) are the state equations for trend and cycle.[3] This opens the way to the Kalman filter, a powerful set of algorithms for filtering, predicting, and smoothing the latent components x_t and c_t, but also for parameter estimation. In this section, I derive an analytical solution to the optimization problem of the Kalman filter and smoother. As will become clear at the end of this section, the analytical solution has two decisive advantages over the usual recursive algorithm: it is computationally more efficient, and it greatly simplifies the asymptotic analysis of the objective function for parameter estimation. In addition, it encompasses the HP filter.

Note that y_t is only observable for $t \geq 1$. Thus, trend, cycle, and parameters can only be estimated based on a truncated representation of the cyclical lag polynomial. To arrive at a feasible representation, define the truncated polynomial $b_+(L, \varphi)$ via $b_+(L, \varphi) c_t = b(L, \varphi) c_t \mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} b_j(\varphi) c_{t-j}$. Furthermore, collect $x_{t:1} = (x_t, ..., x_1)'$ and $c_{t:1} = (c_t, ..., c_1)'$, and define the

[3] Section 5 outlines the state space representation and illustrates the dimensions of the system matrices. For further details on state space models and the Kalman filter, see Harvey (1989, ch. 3).

$t \times t$ differencing matrix $S_{d,t}$ and the $t \times t$ coefficient matrix $B_{\varphi,t}$

$$S_{d,t} = \begin{bmatrix} \pi_0(d) & \pi_1(d) & \cdots & \pi_{t-1}(d) \\ 0 & \pi_0(d) & \cdots & \pi_{t-2}(d) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_0(d) \end{bmatrix}, \qquad B_{\varphi,t} = \begin{bmatrix} b_0(\varphi) & b_1(\varphi) & \cdots & b_{t-1}(\varphi) \\ 0 & b_0(\varphi) & \cdots & b_{t-2}(\varphi) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_0(\varphi) \end{bmatrix}, \tag{5}$$

such that $S_{d,t}x_{t:1} = (\Delta_+^d x_t, ..., \Delta_+^d x_1)'$ and $B_{\varphi,t}c_{t:1} = (b_+(L,\varphi)c_t, ..., b_+(L,\varphi)c_1)'$. $S_{d,t}$ is defined analogously to the integer-integrated differencing matrix of Burman and Shumway (2009), and it holds that $S_{d,t}S_{-d,t} = I$, and $S_{0,t} = I$. In the following, I show the closed-form solutions for the updating step of the Kalman filter to be given by

$$\hat{x}_{t:1}(y_{t:1},\psi) = \left(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t}\right)^{-1} B'_{\varphi,t}B_{\varphi,t}y_{t:1} = \hat{x}_{t:1}(y_{t:1},\theta), \tag{6}$$

$$\hat{c}_{t:1}(y_{t:1},\psi) = \nu\left(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t}\right)^{-1} S'_{d,t}S_{d,t}y_{t:1} = \hat{c}_{t:1}(y_{t:1},\theta), \tag{7}$$

where the fraction $\nu = \sigma_\epsilon^2/\sigma_\eta^2$ controls for the variance ratio of the innovations, $\hat{x}_{t:1}(y_{t:1},\psi) = (\hat{x}_t(y_{t:1},\psi), ..., \hat{x}_1(y_{t:1},\psi))'$, $\hat{c}_{t:1}(y_{t:1},\psi) = (\hat{c}_t(y_{t:1},\psi), ..., \hat{c}_1(y_{t:1},\psi))'$ collect the filtered trend and cycle, and $\theta = (d,\nu,\varphi')'$. (6) and (7) are identical to the recursive solutions from the updating equation of the Kalman filter. The one-step ahead predictions for x_{t+1} and c_{t+1} are obtained by plugging (6) and (7) into the state equations (2) and (4)

$$\hat{x}_{t+1}(y_{t:1},\theta) = -\begin{pmatrix}\pi_1(d) & \cdots & \pi_t(d)\end{pmatrix}\hat{x}_{t:1}(y_{t:1},\theta), \tag{8}$$

$$\hat{c}_{t+1}(y_{t:1},\theta) = -\begin{pmatrix}b_1(\varphi) & \cdots & b_t(\varphi)\end{pmatrix}\hat{c}_{t:1}(y_{t:1},\theta). \tag{9}$$

Together, the updating equations (6), (7) and the prediction equations (8), (9) form the Kalman filter, see Harvey (1989, ch. 3.2) for details. Finally, smoothed estimates for x_t and c_t can be obtained from (6), (7) by setting $t = n$. They are identical to those obtained by the Kalman smoother.

To prove (6) and (7), I first consider the objective function of the Kalman filter, which follows from maximizing the quasi-log likelihood of (1), (2), and (4) with respect to $x_{t:1} = (x_t, ..., x_1)'$, $c_{t:1} = (c_t, ..., c_1)'$ given $y_{t:1} = (y_t, ..., y_1)'$ and $\psi = (d, \sigma_\eta^2, \sigma_\epsilon^2, \varphi')'$. This is the same as minimizing

$$\hat{x}_{t:1}(y_{t:1},\psi) = \arg\min_{x_{t:1}} \frac{1}{t}\sum_{j=1}^{t}\left\{\frac{1}{\sigma_\epsilon^2}\left[b_+(L,\varphi)(y_j - x_j)\right]^2 + \frac{1}{\sigma_\eta^2}\left(\Delta_+^d x_j\right)^2\right\}, \tag{10}$$

$$\hat{c}_{t:1}(y_{t:1},\psi) = \arg\min_{c_{t:1}} \frac{1}{t}\sum_{j=1}^{t}\left\{\frac{1}{\sigma_\eta^2}\left[\Delta_+^d(y_j - c_j)\right]^2 + \frac{1}{\sigma_\epsilon^2}\left(b_+(L,\varphi)c_j\right)^2\right\}. \tag{11}$$

Here, the first residual in (10) stems from plugging (4) into the measurement equation and solving for ϵ_j, while the second is from (2). Analogously, the first term in (11) follows from inserting (2) into (1) and solving for η_j, while the second follows from solving (4) for ϵ_j. Constant terms are omitted. As x_t and c_t are estimated based on all observations until period t, it holds that $\hat{x}_{t:1}(y_{t:1},\psi) =$

$y_{t:1} - \hat{c}_{t:1}(y_{t:1}, \psi)$. If η_t and ϵ_t are assumed to be Gaussian, the optimization problems in (10) and (11) yield the conditional expectations $\hat{x}_{t:1}(y_{t:1}, \psi) = \mathrm{E}_\psi(x_{t:1}|y_{t:1})$ and $\hat{c}_{t:1}(y_{t:1}, \psi) = \mathrm{E}_\psi(c_{t:1}|y_{t:1})$, see Durbin and Koopman (2012, lemma 1), where the expected value operator $\mathrm{E}_\psi(z_t)$ of an arbitrary random variable z_t denotes that expectation is taken with respect to the distribution of z_t given ψ. If η_t, ϵ_t are not normally distributed, the optimization problems (10) and (11) remain valid. The filtered $\hat{x}_{t:1}(y_{t:1}, \psi)$, $\hat{c}_{t:1}(y_{t:1}, \psi)$ are the projections of $x_{t:1}$ and $c_{t:1}$ on the span of $y_{t:1}$, and are the minimum variance linear unbiased estimators for $x_{t:1}$ and $c_{t:1}$ given the observable information $y_1, ..., y_t$ (Durbin and Koopman; 2012, lemma 2). For $t = n$, $d = 2$, $b(L, \varphi) = 1$, $\nu = \sigma_\epsilon^2/\sigma_\eta^2$, (10) becomes the HP filter with ν being the tuning parameter. Thus, the HP filter constitutes a special case of the fractional UC model.

From (5), a matrix representation of (10) and (11) follows

$$\hat{x}_{t:1}(y_{t:1}, \psi) = \arg\min_{x_{t:1}} \frac{1}{t} \left\{ \frac{1}{\sigma_\epsilon^2} \left\| B_{\varphi,t}(y_{t:1} - x_{t:1}) \right\|^2 + \frac{1}{\sigma_\eta^2} x_{t:1}' S_{d,t}' S_{d,t} x_{t:1} \right\}, \tag{12}$$

$$\hat{c}_{t:1}(y_{t:1}, \psi) = \arg\min_{c_{t:1}} \frac{1}{t} \left\{ \frac{1}{\sigma_\eta^2} \left\| S_{d,t}(y_{t:1} - c_{t:1}) \right\|^2 + \frac{1}{\sigma_\epsilon^2} c_{t:1}' B_{\varphi,t}' B_{\varphi,t} c_{t:1} \right\}, \tag{13}$$

where $\|\cdot\|$ denotes the Euclidean norm. Calculating the derivative of (12) and (13) and solving for x_t and c_t yields (6) and (7). Note that (6) and (7) do not depend on the exact magnitudes of σ_η^2 and σ_ϵ^2, but only on their ratio ν, $0 < \nu < \infty$. Thus, for any positive constant $K > 0$, the parameter vector $\psi^* = (d, K\sigma_\eta^2, K\sigma_\epsilon^2, \varphi')'$ yields the same estimates $\hat{x}_{t:1}(y_{t:1}, \psi^*)$, $\hat{c}_{t:1}(y_{t:1}, \psi^*)$ as (6) and (7). By defining the parameter vector $\theta = (d, \nu, \varphi')'$, one has $\hat{x}_{t:1}(y_{t:1}, \psi) = \hat{x}_{t:1}(y_{t:1}, \theta)$ and $\hat{c}_{t:1}(y_{t:1}, \psi) = \hat{c}_{t:1}(y_{t:1}, \theta)$. This will be helpful for parameter estimation in section 4, since the conditional sum-of-squares estimator is not identified for ψ. Also, using θ reduces the dimension of the parameter vector, which speeds up the optimization. However, ψ can also be estimated directly by maximum likelihood as will be shown in subsection 5.3.

From the filtered latent components in (6) and (7), the one-step ahead predictions for x_{t+1} and c_{t+1} follow immediately by plugging (6) and (7) into the state equations (2) and (4). This yields (8) and (9). While (6), (7), (8), and (9) are required for parameter estimation, as discussed in the next section, estimates for x_t and c_t typically reported are the projections of x_t and c_t on the span of $y_1, ..., y_n$, i.e. on the full sample information. They follow immediately from (6) and (7) by setting $t = n$, and are identical to the Kalman smoother.

Note that the filtered, predicted and smoothed x_t and c_t can be computed either via the analytical solution above or recursively by executing the Kalman recursions (see Harvey; 1989, ch. 3, for the latter). Both approaches yield identical results and only differ in the way they are computed. However, the analytical solution has two decisive advantages over the traditional recursions: (i) It is computationally superior for fractional trends. As the state vector of the fractional trend in (2) is of dimension $n-1$, the dimension of the state vector for both trend and cycle is of dimension $m \geq n-1$. Thus, each recursion of the Kalman filter involves multiple multiplications of $(m \times m)$-dimensional covariance and system matrices, and each multiplication requires $2m^3 - m^2$ flops (Hunger; 2007). The analytical solution also requires the expensive computation of an $(n \times n)$ inverse, however the underlying matrix is symmetric, positive definite, and thus the Cholesky decomposition can be used to reduce the complexity to $n^3 + n^2 + n$ flops per iteration (Hunger; 2007). Since $m \geq n-1$, the analytical solution speeds up the computation considerably. This allows to run the Monte Carlo studies in section 6, which would otherwise be computationally infeasible. (ii) The solution allows to derive an objective function for parameter estimation that does not depend on the Kalman recursions and is thus easier to analyze. As usual, the objective function for parameter estimation is set up based on the one-step ahead prediction error, that is obtained by plugging (8) and (9) into the measurement equation (1). Since (8) and (9) depend only on the observable $y_1, ..., y_t$ as well as on the model parameters, the objective function does not depend on a recursive solution for the filtered trend and cycle. This greatly simplifies the asymptotic theory for parameter estimation in section 4, since the convergence rates of all coefficients are either known, or can be derived immediately.

4 Parameter estimation

To estimate $\theta_0 = (d_0, \nu_0, \varphi_0')'$, denote $\Theta = D \times \Sigma_\nu \times \Phi$ the respective parameter space, where $\Sigma_\nu = \{\nu \in \mathbb{R} | 0 < \nu_{min} \leq \nu \leq \nu_{max} < \infty\}$, and D, Φ as defined in assumption 2. By assumption 2, Θ is convex and compact. As usual in the state space literature, I set up the objective function for parameter estimation based on the one-step ahead forecast error for y_{t+1}, denoted as $v_{t+1}(\theta) = y_{t+1} - \hat{x}_{t+1}(y_{t:1}, \theta) - \hat{c}_{t+1}(y_{t:1}, \theta)$. By plugging in (8) and (9), $v_{t+1}(\theta)$ can be represented as

$$v_{t+1}(\theta) = \Delta_+^d y_{t+1} + \nu \left(b_1(\varphi) - \pi_1(d) \ \cdots \ b_t(\varphi) - \pi_t(d)\right) (B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t})^{-1} S_{d,t}' S_{d,t} y_{t:1}. \quad (14)$$

$v_{t+1}(\theta)$ depends on the fractionally differenced observable y_{t+1}, as well as on past $S_{d,t} y_{t:1} = (\Delta_+^d y_t, ..., \Delta_+^d y_1)'$, weighted by the $1 \times t$ coefficient vector on the right-hand side of (14) that fully depends on θ. Let $\xi_{t+1}(d) = \Delta_+^d y_{t+1} = \Delta_+^{d-d_0} \eta_{t+1} + \Delta_+^d c_{t+1}$ and $\xi_{t:1}(d) = (\xi_t(d) \cdots \xi_1(d))' = S_{d,t} y_{t:1}$ denote the fractionally differenced y_{t+1} and $y_{t:1}$ respectively. Then, (14) can be written as

$$v_{t+1}(\theta) = \xi_{t+1}(d) + \sum_{j=1}^{t} \tau_j(\theta, t) \xi_{t+1-j}(d) = \sum_{j=0}^{t} \tau_j(\theta, t) \xi_{t+1-j}(d), \quad (15)$$

where $\tau_0(\theta, t) = 1$, and $(\tau_1(\theta, t) \cdots \tau_t(\theta, t)) = \nu(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))(B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t})^{-1} S_{d,t}'$ collects the t coefficients belonging to $\xi_t(d), ..., \xi_1(d)$ in (15). The conditional sum-of-squares (CSS) estimator for θ_0 follows from minimizing the sum of squared forecast errors

$$\hat{\theta} = \arg\min_{\theta \in \Theta} Q(y, \theta), \qquad Q(y, \theta) = \frac{1}{n} \sum_{t=1}^{n} v_t^2(\theta). \quad (16)$$

Since the objective function is proportional to the exponent in the quasi-likelihood function, (16) is similar to the quasi-maximum likelihood estimator that is typically used in the state space literature, see e.g. Durbin and Koopman (2012, ch. 7). While the latter allows for a time-varying variance of the prediction error, (16) implicitly assumes a constant variance of the prediction error. However, as subsection 5.3 discusses in greater detail, the filtered prediction error variance of the fractional UC model converges to its steady state solution at an exponential rate. Thus, (16) and quasi-maximum likelihood estimation are asymptotically equivalent. Differences arise only due to a different weighting of prediction errors at the very beginning of the sample. However, (16) is computationally much simpler, because it avoids the Kalman recursions for the prediction error variance. Furthermore, parameter estimation via the steady-state Kalman filter is identical to (16) after some burn-in period, see Harvey (1989, ch. 4.2.2).

While the asymptotic theory for CSS estimation is well established for autoregressive fractionally integrated moving average (ARFIMA) models, see Hualde and Robinson (2011) and Nielsen (2015), only little is known about the asymptotic theory for unobserved components models of such generality. For the sub-class of $I(1)$ UC models with Gaussian white noise shocks η_t and ϵ_t, the asymptotic theory can be inferred from the ARIMA literature (Harvey and Peters; 1990; Morley et al.; 2003). Unfortunately, no such results are available for UC models with fractional trends, so the asymptotic theory for parameter estimation of fractional UC models must be derived from scratch. While the proofs in this section are given for the (simpler) CSS estimator, it is shown in subsection 5.3 that they also apply to the traditional quasi-maximum likelihood estimator. Due to the encompassing nature of the fractional UC model, the results below also hold for CSS and quasi-maximum likelihood estimation of all sub-classes of UC models such as e.g. integer-integrated models with MDS shocks.

Theorem 4.1. *For the model in* (1)*,* (2)*, and* (4)*, and under assumptions 1 to 3, the estimator* $\hat{\theta}$ *as defined via* (16) *is consistent, i.e.* $\hat{\theta} \xrightarrow{p} \theta_0$ *as* $n \to \infty$*.*

The proof is contained in Appendix B. While consistency ultimately follows from a uniform weak law of large numbers (UWLLN), showing that the UWLLN holds is complicated by the non-uniform convergence of the objective function within Θ, as well as by the non-ergodicity of the prediction errors in (14): First, as can be seen from (14), the prediction errors are $I(d_0 - d)$, and thus are asymptotically stationary for $d_0 - d < 1/2$, and otherwise non-stationary. In the former case, a UWLLN can be shown to hold for the objective function, while in the latter case a functional central limit theorem holds under some additional assumptions. Consequently, uniform convergence of the objective function fails around the point $d = d_0 - 1/2$. Following the idea of Nielsen (2015), I partition the parameter space D into three compact subsets, one where $v_t(\theta)$ is asymptotically non-stationary, one for stationary $v_t(\theta)$, and an overlapping subset. Next, whenever θ is not contained in the stationary region of the parameter space, I show that the objective function approaches infinity with probability converging to 1 as $n \to \infty$. Thus, the relevant region of the parameter space reduces asymptotically to the region where $d_0 - d < 1/2$ holds, and where uniform convergence of the objective function is not hindered.

Second, even within the asymptotically stationary region of the parameter space, the forecast errors are non-ergodic, as can be seen from (14) and (15): The truncated fractional differencing polynomial Δ_+^d includes more lags as t increases, and thus $\xi_t(d) = \Delta_+^{d-d_0}\eta_t + \Delta_+^d c_t$ is non-ergodic. In addition, $\tau_j(\theta, t)$ in (15) depends on t. Consequently, even for $d_0 - d < 1/2$, a law of large numbers for stationary and ergodic series does not apply directly to $v_t(\theta)$. I tackle this problem by showing that the difference between the prediction error in (14), and the untruncated and ergodic

$\tilde{v}_t(\theta) = \sum_{j=0}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d)$, is asymptotically negligible in probability, where $\tilde{\xi}_t(d) = \Delta^{d-d_0}\eta_t + \Delta^d c_t$ is the untruncated residual, while the coefficients $\tau_j(\theta)$ stem from the ∞-vector $(\tau_1(\theta), \tau_2(\theta)\cdots) = \nu(b_1(\varphi) - \pi_1(d), b_2(\varphi) - \pi_2(d), \cdots)(B'_{\varphi,\infty}B_{\varphi,\infty} + \nu S'_{d,\infty}S_{d,\infty})^{-1}S'_{d,\infty}$, and $\tau_0(\theta) = 1$. Since $\tilde{v}_t(\theta)$ is stationary and ergodic within the stationary region of the parameter space, it follows that a weak law of large numbers applies to the objective function. The final part of the proof is to strengthen pointwise convergence in probability to weak convergence, which yields the desired result of theorem 4.1.

With a consistent parameter estimator at hand, I next derive the asymptotic distribution of the CSS estimator. For this purpose, assumption 3 needs to be strengthened.

Assumption 4. *For all z in the complex unit disc $\{z \in \mathbb{C} : |z| \leq 1\}$, it holds that $a(z, \varphi)$ is three times continuously differentiable in φ on the closed neighborhood $N_\delta(\varphi_0) = \{\varphi \in \Phi : |\varphi - \varphi_0| \leq \delta\}$ for some $\delta > 0$, and the derivatives satisfy $\frac{\partial^2 a_j(\varphi)}{\partial\varphi_{(k)}\partial\varphi_{(l)}} = O(j^{-1-\zeta})$, and $\frac{\partial^3 a_j(\varphi)}{\partial\varphi_{(k)}\partial\varphi_{(l)}\partial\varphi_{(m)}} = O(j^{-1-\zeta})$, for all entries $\varphi_{(k)}, \varphi_{(l)}, \varphi_{(m)}$ of φ.*

Assumption 4 is similar to assumption E of Nielsen (2015), and strengthens the smoothness conditions of the linear coefficients in $a(L, \varphi)$. It ensures absolute summability of the partial derivatives, which is used to prove uniform convergence of the Hessian matrix and thus to evaluate the Hessian matrix at θ_0 in the Taylor expansion of the score. The convergence rates of the (second and third) partial derivatives are a direct consequence of compactness of $N_\delta(\varphi_0)$ together with continuity of the partial derivatives. Assumption 4 still includes the class of stationary ARMA processes, and even allows for a slower rate of decay of the autocovariance function.

Assumption 5. *The true prediction error of the untruncated process $\tilde{v}_t(\theta_0)$ is a MDS when adapted to the filtration $\mathcal{F}_t^{\tilde{\xi}} = \sigma(\tilde{\xi}_s, s \leq t)$, where $\tilde{\xi}_s = \tilde{\xi}_s(d_0)$.*

Assumption 5 can be motivated as follows: As shown in the proof of theorem 4.1, the prediction error of the Kalman filter converges to the untruncated, stationary and ergodic $\tilde{v}_t(\theta_0) = v_t(\theta_0) + o_p(1)$ as $t \to \infty$, while $\Delta_+^{d_0} y_t = \xi_t(d_0) = \tilde{\xi}_t + o_p(1)$ as $t \to \infty$, and thus the (relevant fraction) of the filtration $\mathcal{F}_t^{\tilde{\xi}}$ asymptotically equals the filtration generated by the $\Delta_+^{d_0} y_s$, $1 \leq s \leq t$. Consequently, assumption 5 requires the prediction error of the Kalman filter to converge to a MDS when adapted to a filtration that asymptotically is equal to the filtration generated by the differenced, observable variables. For assumption 5 to be satisfied, the one-step ahead forecasts for trend and cycle in (6) and (7) must converge to their expectations conditional on $\mathcal{F}_t^{\tilde{\xi}}$. Since $\tilde{v}_t(\theta_0)$ plays the role of the (asymptotic) residual for fractional UC models, assumption 5 fits well to the usual assumption of MDS residuals for CSS estimation, see e.g. Hualde and Robinson (2011), Nielsen (2015), and Hualde and Nielsen (2020). In the UC literature, Dunsmuir (1979, ass. C2.3) imposes the same assumption for his stationary signal plus noise model, but also discusses the possibility of relaxing the assumption (see Dunsmuir; 1979, pp. 502f). Trivially, assumption 5 is satisfied if long- and short-run innovations are Gaussian.

Theorem 4.2. *For the model in* (1), (2), *and* (4), *under assumptions 1 to 5, the estimator $\hat{\theta}$ as defined via* (16) *is asymptotically normally distributed, i.e. $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathrm{N}(0, \sigma_{v,0}^2\Omega_0^{-1})$ as*

$n \to \infty$, *with* $\sigma^2_{v,0} = \lim_{t\to\infty} \mathrm{Var}(v_t(\theta_0)) = \mathrm{Var}(\tilde{v}_t(\theta_0))$, *and* Ω_0 *has the* (i,j)*-th entry* $\Omega_{0_{(i,j)}} = \mathrm{E}\left(\frac{\partial \tilde{v}_t(\theta)}{\partial \theta_{(i)}}\big|_{\theta=\theta_0} \frac{\partial \tilde{v}_t(\theta)}{\partial \theta_{(j)}}\big|_{\theta=\theta_0}\right)$, $i, j = 1, ..., q+2$.

The proof of theorem 4.2 is contained in Appendix C. As usual, the asymptotic distribution of the CSS estimator is inferred from a Taylor expansion of the score function around θ_0. Analogous to Robinson (2006) and Hualde and Robinson (2011), it is first shown that the normalized score at θ_0 is asymptotically equivalent to the score function of the untruncated, stationary and ergodic residual $\sqrt{n}(\partial \tilde{Q}(y,\theta)/\partial\theta)\big|_{\theta=\theta_0} = (2/\sqrt{n})\sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\partial \tilde{v}_t(\theta)/\partial\theta)\big|_{\theta=\theta_0}$. Next, a UWLLN is shown to hold for the Hessian matrix, so that it can be evaluated at θ_0 in the Taylor expansion, and the difference between the truncated and untruncated Hessian matrix is shown to be asymptotically negligible in probability. Therefore, both the score and the Hessian matrix in the Taylor expansion can be replaced by their untruncated counterparts. While a weak law of large numbers applies to the untruncated Hessian matrix, under assumption 5 a central limit theorem for martingale difference sequences applies to the score and yields the asymptotic distribution. Finally, while theorem 4.2 does not give an analytical expression for the covariance matrix of the CSS estimator, it shows that Ω_0^{-1} can by estimated via the numerical Hessian matrix.

5 Generalizations

One key advantage of the fractional UC model is its state space representation: It makes the Kalman filter and smoother applicable, enables quasi-maximum likelihood estimation of the model parameters, allows to diffusely initialize the filter, and to seamlessly add additional structural components to the model. In addition, several useful methods and generalizations become available that are beyond the scope of this paper, such as frequency-domain optimization, additional observable explanatory variables, time-varying and nonlinear models, and mixed-frequency models among others; see Harvey (1989) for an overview. In this section, I outline some generalizations of the fractional UC model that are of immediate applied relevance: Subsection 5.1 introduces deterministic components to the model, while subsection 5.2 allows for correlated trend and cycle innovations. Subsection 5.3 generalizes parameter estimation to the quasi-maximum likelihood estimator. For all three modifications, the asymptotic results of section 4 are shown to remain valid. However, before turning to the three generalizations, I first introduce the state space representation of the fractional UC model.

The basic state space representation has the form

$$y_t = Z\alpha_t + u_t, \tag{17}$$

$$\alpha_t = T\alpha_{t-1} + R\zeta_t, \tag{18}$$

where the states may be partitioned into $\alpha_t = (\alpha_t^{(x)\prime}, \alpha_t^{(c)\prime}, \alpha_t^{(r)\prime})'$, with $(n-1)$-vectors for trend $\alpha_t^{(x)} = (x_t, x_{t-1}, ..., x_{t-n+2})'$, and cycle $\alpha_t^{(c)} = (c_t, c_{t-1}, ..., c_{t-n+2})'$. The observation matrix is $Z = (Z^{(x)}, Z^{(c)}, Z^{(r)})$, where $Z^{(x)} = (1, 0, ..., 0)$, $Z^{(c)} = (1, 0, ..., 0)$ are $(n-1)$-dimensional row vectors picking the first entry of $\alpha_t^{(x)}$ and $\alpha_t^{(c)}$. For the transition equation (18), one has $T =$

$\mathrm{diag}(T^{(x)}, T^{(c)}, T^{(r)})$, $R = \mathrm{diag}(R^{(x)}, R^{(c)}, R^{(r)})$,

$$T^{(x)} = \begin{bmatrix} -\pi_1(d) & -\pi_2(d) & \cdots & -\pi_{n-1}(d) \\ 1 & & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \qquad T^{(c)} = \begin{bmatrix} -b_1(\varphi) & -b_2(\varphi) & \cdots & -b_{n-1}(\varphi) \\ 1 & & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix},$$

and $R^{(x)} = (1, 0, ..., 0)'$, $R^{(c)} = (1, 0, ..., 0)'$ are $(n-1)$-vectors picking the respective entries of $\zeta_t = (\eta_t, \epsilon_t, \zeta_t^{(r)'})'$. Finally, the components $\alpha_t^{(r)}$, $\zeta_t^{(r)}$ allow for general specifications with $\alpha_t^{(r)} = T^{(r)}\alpha_{t-1}^{(r)} + R^{(r)}\zeta_t^{(r)}$ that load on y_t via $Z^{(r)}\alpha_t^{(r)}$. They may capture additional stochastic trends (possibly of different memory) and seasonal components among others. Furthermore, u_t may account for additional terms in the measurement equation, such as measurement errors, deterministic terms, or observable explanatory variables. While both, $\alpha_t^{(r)}$ and u_t are implicitly set to zero in section 4, their specification in practice is left open to the applied researcher. Finally, $\mathrm{Var}(\zeta_t) = Q$.

5.1 Deterministic components

In practice, deterministic components often need to be considered. As will become clear, such terms can be straightforwardly added to the state space framework, and their estimation can be carried out efficiently by a combination of the Kalman filter, the GLS estimator, and the CSS estimator. For the GLS estimator to be a consistent estimator for the coefficients of the deterministic components, the deterministic terms must diverge at a rate similar to the divergence rate of the stochastic trend.

Deterministic components can be taken into account either by detrending the data prior to estimating the fractional UC model, or by adding the components to the state space model. However, prior detrending biases the estimates for both deterministic and stochastic trends whenever the data are non-stationary, and thus should be avoided (Harvey; 1989, ch. 6.1.3). An alternative is to include the deterministic terms into the state vector and to explicitly model their dynamics via the state equation (18). However, state space models with deterministic components in the state vector are not stabilisable, so the Kalman filter does not converge to its steady state solution and the CSS estimator is not applicable, see Harvey (1989, ch. 4.2.5). Following the suggestion there, I place the deterministic terms directly in the measurement equation (17). This allows to estimate the deterministic components by the GLS estimator and does not interfere with the steady state convergence of the Kalman filter. The remaining parameters θ_0 can be estimated via CSS as described in section 4, with the asymptotic theory being unaffected.

To model the deterministic terms, I set $u_t = \mu' w_t$ in the measurement equation (17), where w_t is a non-stochastic k-vector holding k deterministic components, and μ is a k-vector of unknown parameters to be estimated. The modified measurement equation is then $y_t = \mu' w_t + Z\alpha_t$. Letting $W = (w_1, ..., w_n)'$ denote the $n \times k$ matrix collecting all w_t, and $V = \mathrm{Var}(x_{1:n} + c_{1:n})$ denote the variance-covariance matrix of $x_{1:n} + c_{1:n}$, the GLS estimator for μ is given by $\tilde{\mu} = (W'V^{-1}W)^{-1}W'V^{-1}y_{1:n}$, see Harvey (1989, ch. 3.4.2). As also shown there, it is not necessary to compute V^{-1}. To see this, assume for the moment that $y_t - \mu' w_t$ was observable. The Kalman filter, when applied to $y_t - \mu' w_t$, yields the filtered values for trend and cycle in (6) to (9), together with the prediction errors as denoted by $v_t^*(\theta)$ in the following for the modified model. These prediction errors correspond to the linear filtering $F(\theta)(y_{1:n} - W\mu)$, where $F(\theta)$ from the Cholesky decomposition $V^{-1}(\psi) = F(\theta)' D^{-1}(\psi) F(\theta)$ is a p.d. lower triangular matrix with ones on the leading diagonal, $D(\psi)$ is a diagonal p.d. matrix, and $V(\psi)$ is the covariance matrix of $x_{1:n} + c_{1:n}$ conditional on ψ. Since the Kalman filter is linear, it can be applied separately to the observable y_t and w_t, yielding $F(\theta) y_{1:n} = y^*(\theta)$ and $F(\theta)W = W^*(\theta)$ as prediction errors. The GLS estimator $\tilde{\mu}$ then follows from regressing $y^*(\theta) = (y_1^*(\theta), ..., y_n^*(\theta))'$ on $W^*(\theta) = (w_1^*(\theta), ..., w_n^*(\theta))'$, see Harvey (1989, ch. 3.4.2). The concentrated CSS estimator $\tilde{\theta} = (\tilde{d}, \tilde{\nu}, \tilde{\varphi}')'$ follows from minimizing the modified sum of squared prediction errors

$$\tilde{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{t=1}^{n} v_t^*(\theta)^2, \tag{19}$$

and $v_t^*(\theta) = y_t^*(\theta) - \tilde{\mu}' w_t^*(\theta)$ is the GLS residual. Asymptotic standard errors can be obtained from the Fisher information matrix (Harvey; 1989, ch. 4.5.3 and ch. 7.3).

To derive the asymptotic properties of both the GLS estimator $\tilde{\mu}$ and the concentrated CSS estimator (19), let the j-th term in w_t be $w_{j,t} = O(t^{\beta_j})$, $t \geq 1$, $\beta_j \in \mathbb{R}$, such that $w_{j,t}$ is a polynomial trend. I will only consider $-1 < \beta_j \leq d_0$ for all j, as the lower bound is required for $\Delta_+^{d_0} t^{\beta_j} = O(t^{\beta_j - d_0})$ to hold, see Robinson (2005), while the upper bound ensures that the fractional stochastic trend is not drowned by the deterministic terms. This guarantees that the results on consistency and asymptotic normality of the CSS estimator in theorems 4.1 and 4.2 remain valid. However, at least for CSS estimation of ARFIMA models, Hualde and Nielsen (2020) recently derived the asymptotic theory where they also allowed for deterministic trends of higher power, $\beta_j > d_0$. As the focus of this paper is not on the deterministic components, showing their results to carry over is left open for future research.

Note that within $-1 < \beta_j \leq d_0$, the arguments for consistency of the CSS estimator of θ_0 remain unchanged: $y^*(\theta) = F(\theta) y_{1:n}$ is $I(d_0 - d)$ and precisely equals the initial prediction error (14) in section 3 if y_t contains no deterministic terms, since $F(\theta) y_{1:n}$ is the residual from applying the Kalman filter as defined in section 3 to $y_{1:n}$ given the parameters θ. If deterministic terms are present in y_t, then $y^*(\theta) = F(\theta) y_{1:n}$ equals the prediction error (14) shifted either by a constant, or by an $o(1)$ term (depending on how close β_j is to d_0, as will become clear). Therefore, also the prediction error $v_t^*(\theta) = [y^*(\theta) - W^*(\theta)(W^{*'}(\theta) W^*(\theta))^{-1} W^{*'}(\theta) y^*(\theta)]_{(t)}$ is $I(d_0 - d)$. Thus, both $y_t^*(\theta)$ and $v_t^*(\theta)$ are asymptotically stationary for $d_0 - d < 1/2$, otherwise non-stationary. By the same proof as for (B.1), the objective function (19) can be shown to converge in probability whenever $d_0 - d > -1/2$, and to diverge in the opposite case. Therefore, the probability of the CSS estimator to converge within the non-stationary region of the parameter space is asymptotically zero. Thus, it is sufficient to consider the region of the parameter space where $v_t^*(\theta)$ is asymptotically stationary. Within this region, the same proof as for theorem 4.1 applies, showing that a UWLLN holds for the objective function. Thus, $\tilde{\theta}$ is consistent. This result is somewhat obvious, as the assumption on β_j ensures that the filtered $y_t^*(\theta_0)$ contains at most deterministic terms of order $O(1)$.

For the GLS estimator, define $u^*(\theta) = (u_1^*, ..., u_n^*)' = F(\theta)(x_{1:n} + c_{1:n})$ as the residual from applying the Kalman filter to the true $x_{1:n}$ and $c_{1:n}$. $u_t^*(\theta)$ would equal the prediction error $v_t^*(\theta)$ if there were no deterministic terms. The GLS estimates $\tilde{\mu}$ are thus

$$
\begin{aligned}
\tilde{\mu} &= (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})F(\tilde{\theta})y_{n:1} \\
&= (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})F(\tilde{\theta})\left[W\mu_0 + x_{1:n} + c_{1:n}\right] \\
&= \mu_0 + (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta}),
\end{aligned}
\tag{20}
$$

where μ_0 denotes the true coefficients to be estimated. $\tilde{\mu}$ is consistent if and only if the latter term in (20) is $o_p(1)$, i.e. the bias converges to zero as $n \to \infty$. For the purpose of illustration, I will focus only on a single deterministic term, such that $W^*(\tilde{\theta}) = (w_1^*(\tilde{\theta}), ..., w_n^*(\tilde{\theta}))'$. However, the results carry over directly to several deterministic components. First, note that by the fractional differencing via $F(\tilde{\theta})$, $w_t^*(\tilde{\theta}) = O(t^{\beta-\tilde{d}})$, while $u_t^*(\tilde{\theta}) \sim I(d_0 - \tilde{d})$. By consistency of the concentrated CSS estimator, $u_t^*(\tilde{\theta})$ is asymptotically $I(0)$, while $w_t^*(\tilde{\theta}) = O(t^{\beta-d_0})$, and thus $\sum_{t=1}^n w_t^{*^2}(\tilde{\theta}) = \sum_{t=1}^n O(t^{2(\beta-d_0)})$, see Hualde and Nielsen (2020, lemma S.10). Hence, for a single deterministic component, the bias term in (20) can be written as

$$
(W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta}) = \left(\frac{\sum_{t=1}^n w_t^{*^2}(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}\right)^{-1} \frac{\sum_{t=1}^n w_t^*(\tilde{\theta})u_t^*(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}, \tag{21}
$$

where $n^{-1-2(\beta-\tilde{d})}\sum_{t=1}^n w_t^{*^2}(\tilde{\theta})$ is bounded from above and below as $n \to \infty$. In contrast, by Hualde and Nielsen (2020, eqn. (S.88)), $n^{-1-2(\beta-\tilde{d})}\sum_{t=1}^n w_t^* u_t^*(\tilde{\theta}) = o_p(1)$ if and only if $d_0 - 1/2 < \beta$. Thus, the GLS estimator for the deterministic terms is consistent only if the deterministic and stochastic trends diverge at similar rates. As also can be seen from (21), the power of the deterministic term affects the rate of convergence of the GLS estimator: Since $n^{-1/2-(\beta-\tilde{d})}\sum_{t=1}^n w_t^*(\tilde{\theta})u_t^*(\tilde{\theta})$ converges in distribution when $n \to \infty$, see Hualde and Nielsen (2020, proof of cor. 1), it follows that the GLS estimator converges at the rate $n^{1/2+(\beta-d_0)}$ as $n \to \infty$, and thus the rate is slower than the standard $\sqrt{n}$-convergence whenever the deterministic terms are dominated by the stochastic trend.

In summary, any trend of order $d_0 - 1/2 < \beta_j \leq d_0$ can be estimated consistently, and the convergence rate of the GLS estimator will be faster the closer β_j is to d_0. This is in line with the well-established finding in the literature, that an intercept (i.e. $\beta_j = 0$) cannot be estimated consistently for time series with unit roots ($d_0 = 1$), whereas a linear trend ($\beta_j = 1$) can be estimated consistently. Moreover, the convergence rate matches the findings of Robinson (2005) for semiparametric long memory models with deterministic components, of Hualde and Nielsen (2020) for parametric ARFIMA models with deterministic components, and the general literature on the estimation of the sample mean for fractionally integrated processes, see e.g. Hassler (2019, ch. 7).

5.2 Correlated trend and cycle innovations

As shown by Morley et al. (2003), at least for integer-integrated structural time series models of log US real GDP, correlation between permanent and transitory shocks is found to be highly significant. Therefore, this subsection generalizes the fractional UC model to account for correlated innovations

$$\operatorname{Var}\begin{pmatrix}\eta_t\\ \epsilon_t\end{pmatrix} = \begin{bmatrix}\sigma_\eta^2 & \sigma_{\eta\epsilon}\\ \sigma_{\eta\epsilon} & \sigma_\epsilon^2\end{bmatrix} = \varSigma.$$

The new optimization problem of the Kalman filter is then

$$\begin{aligned}\hat{x}_{t:1}(y_{t:1},\tilde{\psi}) &= \arg\min_{x_{t:1}} \frac{1}{t}\sum_{j=1}^{t}\left[\begin{pmatrix}\eta_j & \epsilon_j\end{pmatrix}\varSigma^{-1}\begin{pmatrix}\eta_j\\ \epsilon_j\end{pmatrix}\right]\\ &= \arg\min_{x_{t:1}} \frac{1}{t}\frac{1}{\sigma_\eta^2\sigma_\epsilon^2-\sigma_{\eta\epsilon}^2}\sum_{j=1}^{t}\left[\sigma_\epsilon^2\eta_j^2 - 2\sigma_{\eta\epsilon}\eta_j\epsilon_j + \sigma_\eta^2\epsilon_j^2\right],\end{aligned}$$

where $\tilde{\psi} = (d, \sigma_\eta^2, \sigma_{\eta\epsilon}, \sigma_\epsilon^2, \varphi')'$ denotes the new parameter vector that now also includes the covariance $\sigma_{\eta\epsilon}$. By dropping the determinant and plugging in $\eta_j = \Delta_+^d x_j$ as well as $\epsilon_j = b_+(L,\varphi)(y_j - x_j)$, the optimization problem can be written as

$$\begin{aligned}\hat{x}_{t:1}(y_{t:1},\tilde{\psi}) &= \arg\min_{x_{t:1}} \frac{1}{t}\sum_{j=1}^{t}\left[\sigma_\epsilon^2(\Delta_+^d x_j)^2 - 2\sigma_{\eta\epsilon}\Delta_+^d x_j b_+(L,\varphi)(y_j-x_j) + \sigma_\eta^2\left(b_+(L,\varphi)(y_j-x_j)\right)^2\right]\\ &= \arg\min_{x_{t:1}} \frac{1}{t}\left[\sigma_\eta^2\|B_{\varphi,t}(y_{t:1}-x_{t:1})\|^2 - 2\sigma_{\eta\epsilon}(y_{t:1}-x_{t:1})'B_{\varphi,t}'S_{d,t}x_{t:1} + \sigma_\epsilon^2 x_{t:1}'S_{d,t}'S_{d,t}x_{t:1}\right],\end{aligned}$$

where the matrix representation in the last step is derived analogously to (12). The solution to the optimization problem is then

$$\begin{aligned}\hat{x}_{t:1}(y_{t:1},\tilde{\psi}) &= \left[\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}'B_{\varphi,t} + B_{\varphi,t}'S_{d,t}) + \sigma_\epsilon^2 S_{d,t}'S_{d,t}\right]^{-1}\\ &\quad\times\left(\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}S_{d,t}'B_{\varphi,t}\right)y_{t:1},\end{aligned}\tag{22}$$

and, either by solving the same optimization steps for $\hat{c}_{t:1}(y_{t:1},\tilde{\psi})$, or by using $y_{t:1} = \hat{x}_{t:1}(y_{t:1},\tilde{\psi}) + \hat{c}_{t:1}(y_{t:1},\tilde{\psi})$

$$\begin{aligned}\hat{c}_{t:1}(y_{t:1},\tilde{\psi}) &= \left[\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}'B_{\varphi,t} + B_{\varphi,t}'S_{d,t}) + \sigma_\epsilon^2 S_{d,t}'S_{d,t}\right]^{-1}\\ &\quad\times\left(\sigma_\epsilon^2 S_{d,t}'S_{d,t} + \sigma_{\eta\epsilon}B_{\varphi,t}'S_{d,t}\right)y_{t:1}.\end{aligned}\tag{23}$$

Obviously, (22) and (23) equal (6) and (7) for $\sigma_{\eta\epsilon} = 0$. As before, the number of parameters in the optimization may be reduced by dividing the first and second parenthesis in (22) and (23) by σ_η^2, defining $\nu = \sigma_\epsilon^2/\sigma_\eta^2$ as well as $\nu_2 = \sigma_{\eta\epsilon}/\sigma_\eta^2$, and replacing $\tilde{\psi}$ by $\bar{\theta} = (d, \nu, \nu_2, \varphi')'$. This is necessary for the CSS estimator to be identified, however the quasi-maximum likelihood estimator derived in subsection 5.3 can be used to estimate $\tilde{\psi}_0 = (d_0, \sigma_{\eta,0}^2, \sigma_{\eta\epsilon,0}, \sigma_{\epsilon,0}^2, \varphi_0')$, the true parameters, directly.

The objective function for the CSS estimator can be constructed analogously to section 4: First, the one-step ahead predictions for x_{t+1} and c_{t+1} are obtained as in (8) and (9). Next, they are subtracted from y_{t+1}, which gives the prediction error

$$\begin{aligned}v_{t+1}(\tilde{\psi}) &= \Delta_+^d y_{t+1} + (b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\\ &\quad\times\left[\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}'B_{\varphi,t} + B_{\varphi,t}'S_{d,t}) + \sigma_\epsilon^2 S_{d,t}'S_{d,t}\right]^{-1}\left(\sigma_\epsilon^2 S_{d,t}' + \sigma_{\eta\epsilon}B_{\varphi,t}'\right)S_{d,t}y_{t:1}.\end{aligned}\tag{24}$$

Based on (24), a CSS estimator for the true parameters $\bar{\theta}_0 = (d_0, \nu_0, \nu_{2,0}, \varphi_0')$ can be set up. Note that y_{t+1} enters (24) in fractional differences, and also note that all terms in (24) have the same convergence rates as for the case with uncorrelated errors. Thus, the CSS estimator with correlated innovations can be shown to be consistent and asymptotically normally distributed by carrying out the same proofs as summarized in section 4. Finally, as noted by Morley et al. (2003), for the integer-integrated case $d_0 = 1$, the model is not identified if c_t follows an AR(p) with $p < 2$, since the autocovariance function of Δy_t dies out after lag one. For non-integer integration orders, identification is not a problem, as the autocovariance function of $\Delta_+^d y_t$ dies out only at lag t.

5.3 Maximum likelihood estimation

Since the vast majority of state space models are estimated by quasi-maximum likelihood (QML), this subsection relates the CSS estimator to the QML estimator. For this purpose, denote $\psi = (d, \sigma_\eta^2, \sigma_\epsilon^2, \varphi)'$ the vector holding the model parameters of the fractional UC model. Furthermore, let $\text{Var}_\psi\left(v_t(\psi)|y_1, ..., y_{t-1}\right) = \sigma_{v_t}^2$ denote the (hypothetical) variance of $v_t(\psi)$ that is obtained when evaluating the conditional distribution of $v_t(\psi)$ at ψ. While the CSS estimator allowed to concentrate out the variance parameters $\sigma_\eta^2, \sigma_\epsilon^2$ and model only their variance ratio $\nu = \sigma_\epsilon^2/\sigma_\eta^2$, this is not possible for the QML estimator, since the levels of $\sigma_\eta^2, \sigma_\epsilon^2$ determine $\sigma_{v_t}^2$. Thus, optimization is conducted over ψ. Note further that ψ can be extended to account for correlated innovations, as described in subsection 5.2. A recursive solution for $\sigma_{v_t}^2$ is typically obtained from the Kalman filter, see Durbin and Koopman (2012, ch. 4.3). The quasi-log likelihood is then set up based on the conditional distribution of $v_t(\psi)$ and is given by

$$\log L(\psi) = -\frac{1}{2}\sum_{t=1}^{n} \log \sigma_{v_t}^2 - \frac{1}{2}\sum_{t=1}^{n} \frac{v_t^2(\psi)}{\sigma_{v_t}^2},$$

see Harvey (1989, ch. 3.4). Now, if the Kalman filter converges to its steady state solution at an exponential rate, the QML estimator is asymptotically independent of the initialization of the Kalman filter, see Harvey (1989, ch. 3.4.2), and $\sigma_{v_t}^2$ converges to a constant. Thus, neither initialization of the Kalman filter, nor time-dependence of $\sigma_{v_t}^2$ matter asymptotically, and therefore the CSS estimator in (16) has the same asymptotic distribution as the QML estimator, see Harvey (1989, p. 129).

For the Kalman filter to converge to its steady state solution at an exponential rate, it is sufficient that the state space model is detectable and stabilizable (Harvey; 1989, ch. 3.3.3). Detectability is implied by observability, while stabilizability is implied by controllability (Harvey; 1989, ch. 3.3.1). The state space model as introduced at the beginning of this section is controllable if $\text{Rank}(G, TG, ..., T^{m-1}G) = m$, where m is the dimension of α_t, and $G = RS'$ where S is the upper-triangular matrix from the Cholesky decomposition of the covariance matrix $Q = S'S$ (Harvey; 1989, ch. 3.3.1). The rank condition can be verified by simple algebra, and depends crucially on Q having full rank. Controllability means that given a realization of α_t at some period t, the innovations ζ_{t+j}, $j = 1, ..., m$, can be chosen such that an arbitrarily prescribed value α_{t+m}^* is obtained. Since in each period a new innovation enters (18) for both x_t and c_t, their states in

α_{t+m} can be controlled by controlling ζ_{t+j}. Thus, the state space model is controllable. Similarly, the state space model is observable if $\text{Rank}(Z', T'Z', ..., (T')^{m-1}Z') = m$ (Harvey; 1989, ch. 3.3.1), which again can be verified algebraically. The idea of observability is that α_t can be uniquely determined if $y_t, ..., y_{t+m-1}$, as well as $\zeta_t, ..., \zeta_{t+m-1}$ are known. This is easy to see: Suppose y_{t+j} is known for some $j > 0$. Then $\Delta_+^d y_{t+j} = \eta_{t+j} + \Delta_+^d c_{t+j}$ can be calculated. With η_{t+j} at hand, we can directly calculate c_{t+j}, and thus also x_{t+j}. It follows that the system is observable. Thus, as $n \to \infty$, the CSS estimator and the QML estimator become identical, which was also pointed out by Harvey (1989, p. 187) for integer-integrated models. Consequently, the results in section 4 also hold for the QML estimator.

Finally, while computational efficiency clearly favors the CSS estimator, which avoids the Kalman recursions for the conditional variance of the state vector, the QML estimator may be advantageous in finite samples where the initialization of the Kalman filter plays a non-negligible role. In particular, a combination of the QML estimator, for an initial burn-in period, and the CSS estimator, once the filtered prediction error variance has sufficiently converged, seems promising: It combines the possibility of diffuse initialization and thus assigns a lower weight to initial prediction errors, but switches to the computationally efficient CSS estimator once the benefits of the QML estimator have vanished. The performance of this estimator, typically called the steady-state filter (Harvey; 1989, p. 185f), is also examined in a Monte Carlo study in section 6 and compared to the CSS estimator.

6 Simulations

By the means of a Monte Carlo study, this section examines the finite sample estimation properties for the latent components and parameters of the fractional UC model as introduced in section 2. By considering both the CSS estimator of section 4 and the QML estimator of subsection 5.3, the study demonstrates the loss of estimation accuracy of the computationally simpler CSS estimator by treating the filtered prediction error variance to be constant. Thus, the study puts a price tag on the computational efficiency gains and provides empirical researchers with guidance on when to use the CSS estimator. Furthermore, the parameter estimates for the integration order are compared to the exact local Whittle estimator of Shimotsu and Phillips (2005) for various choices of tuning parameters as a prominent benchmark. To see whether allowing for fractional trends matters, I also present results for the integer-integrated UC models in the spirit of Harvey (1985) and Morley et al. (2003). Doing so, I examine whether fractional trends are well approximated by integer-integrated models, or whether the estimates for x_t and c_t are significantly biased. Furthermore, I investigate whether misspecifying d to be one biases the parameter estimates.

Two different data-generating mechanisms are considered: Subsection 6.1 simulates data based on the fractionally integrated UC model with uncorrelated trend and cycle innovations as introduced in section 2, while subsection 6.2 in addition allows for correlated innovations as discussed in subsection 5.2. Both studies vary over the sample size $n \in \{100, 200, 300\}$, the integration order $d_0 \in \{0.75, 1.00, 1.25, 1.75\}$, and the variance ratio of trend and cycle innovations $\nu_0 = \frac{\sigma^2_{\epsilon,0}}{\sigma^2_{\eta,0}} \in \{1, 5, 10\}$. Thus, they capture small to medium sized samples as typical in empirical applications of UC models, allow for non-stationary mean-reverting trends as well as for non-mean-reverting trends, and reflect situations where short- and long-run shocks are of equal magnitude as well as situations where the long-run shocks are drowned by the short-run dynamics. Each simulation consists of $R = 1000$ replications.

Unlike the CSS estimator, the QML estimator uses the Kalman iterations for the variance of the prediction error, thereby allowing it to be time-dependent: In the Kalman filter, the trend is initialized with variance zero, as implied by the type II definition of fractional integration in (2), whereas the cycle is initialized with its long-run variance as typical in the UC literature. Next, in a burn-in period, the QML estimator takes into account the exponential convergence of the prediction error variance by allowing it to converge to its steady-state value. Once the prediction error variance has converged sufficiently, i.e. it satisfies $\left|\frac{\mathrm{Var}_\psi(v_{t+1}(\psi)|y_1,\ldots,y_t) - \mathrm{Var}_\psi(v_t(\psi)|y_1,\ldots,y_{t-1})}{\mathrm{Var}_\psi(v_t(\psi)|y_1,\ldots,y_{t-1})}\right| < 0.01$, the optimization switches to the steady state Kalman filter, which assumes the prediction error variance to be constant from that point on. This avoids further iterations of the Kalman filter for the prediction error variance, speeds up the computation, and has a negligible impact on the estimation accuracy.

Both the CSS and the QML estimator are initialized by first evaluating the objective functions at a large, equally-spaced grid for the model parameters, and the grid point referring to the lowest value of (16) for the CSS estimator or the lowest negative likelihood is chosen as the starting point for numerical optimization. As a benchmark, the exact local Whittle estimator of Shimotsu and Phillips (2005) is introduced, using $m = \lfloor n^j \rfloor$ Fourier frequencies, $j \in \{.50, .55, .60, .65, .70\}$.

Parameter estimates are compared by the root mean squared error (RMSE), as well as by the bias. To assess how well trend and cycle are estimated, the coefficients of determination R_x^2 and R_c^2 from regressing x_t and c_t on their respective estimates from the Kalman smoother are reported for both CSS and QML estimates.

6.1 Fractional UC model with uncorrelated innovations

In this subsection, I study the finite sample properties of the CSS and QML estimator for the simple fractional UC model

$$y_t = x_t + c_t, \qquad \Delta_+^d x_t = \eta_t, \qquad c_t - b_1 c_{t-1} - b_2 c_{t-2} = \epsilon_t, \tag{25}$$

where $\eta_t \sim \mathrm{NID}(0,1)$, $\epsilon_t \sim \mathrm{NID}(0,\nu)$ are uncorrelated. The cyclical coefficients are set to $b_{1,0} = 1.6$, $b_{2,0} = -0.8$ to reflect strong cyclical patterns. To allow for a better comparison of the CSS and the QML estimator, $\sigma_{\eta,0}^2 = 1$ is fixed and is assumed to be known in the QML optimization, such that estimation is carried out over θ for both the CSS and the QML estimator.

Table A.1 shows the RMSE and the bias for the estimated integration orders for the CSS estimator, the QML estimator, and the exact local Whittle estimator. As can be seen, both RMSE and bias decrease as n increases, which is in line with the theoretical results on consistency. As can be expected from the parametric nature, the fractional UC model yields a much smaller RMSE as compared to the nonparametric Whittle estimator. The differences are particularly striking for high ν_0, where the signal of the fractional trend is drowned by a strong cyclical variation, and for high n. In a direct comparison, the QML estimator slightly outperforms the CSS estimator for the estimation of the integration order, but except for $d_0 = 1.75$, the differences are rather small. Both the CSS and the QML estimator appear to have little or no bias for d_0, while the cyclical dynamics induce a strong negative bias on the exact local Whittle estimates.

Tables A.2 and A.3 contain the RMSE and the bias for ν_0 and the autoregressive parameters, for both the CSS and the QML estimates. In addition to the fractional UC model, the table also displays the estimation results for an $I(1)$ UC benchmark that sets $d = 1$, both for the CSS and the QML estimator. While for $b_{1,0}$ and $b_{2,0}$, the CSS estimator and the QML estimator show a similar performance, major differences occur for the estimate of ν_0, where both the bias and the RMSE are significantly smaller for the QML estimator. In particular, the CSS estimate for ν_0 is always upward-biased, while no such bias is visible for the QML estimator. While the CSS estimator, when compared to the QML estimator, showed little to no disadvantages for the estimation of d_0, $b_{1,0}$, and $b_{2,0}$, the price for the computational simplicity is obviously a biased, imprecise estimate for ν_0. The direct comparison with the $I(1)$ benchmark reveals a slightly smaller RMSE for the fractional UC model for the estimation of $b_{1,0}$ and $b_{2,0}$, while ν_0 is estimated with a significantly higher precision via the fractional UC model whenever $d_0 \neq 1$. Interestingly, for $d_0 = 1.75$ the QML estimate of the $I(1)$ UC model for ν_0 is strongly upward-biased, while no bias is visible for the QML estimate of the fractional UC model.

Table A.4 compares the estimates for x_t and c_t for the fractional UC model and the $I(1)$ UC benchmark (which sets $d = 1$). As before, it contains the results for both the CSS estimator and the QML estimator. As can be seen, differences between the coefficients of determination are almost negligible for the CSS and the QML estimator of the fractional UC model, with the latter exhibiting slightly larger coefficients of determination. Strikingly, for $d_0 = 1$ the fractional UC model shows no loss in efficiency compared to the $I(1)$ UC model. For non-integer d_0, the fractional model clearly outperforms the benchmark model, especially when ν_0 is small. However, for $d_0 \leq 1.25$, the coefficients of determination are still relatively high for the $I(1)$ benchmark, so that, at least for integration orders close to unity, integer-integrated UC models appear to be able to approximate the fractionally integrated trend well, while for $d_0 = 1.75$ integer-integrated UC models clearly fail to resemble the dynamics of the two latent components.

6.2 Fractional UC model with correlated innovations

To examine the estimation properties for the latent components and parameters of the fractional UC model when the long- and short-run innovations are allowed to be correlated, I modify (25) by allowing for a non-diagonal Q in

$$\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim \mathrm{NID}(0, Q). \tag{26}$$

As before, the cyclical coefficients are set to $b_{1,0} = 1.6$, $b_{2,0} = -0.8$. Q_0 is parameterized as $\sigma^2_{\eta,0} = 1$, $\sigma^2_{\epsilon,0} = \nu_0 \in \{1, 5, 10\}$, which yields medium to strong cyclical fluctuations. To mimic strong (but not perfect) correlation between long- and short-run innovations, I set $\sigma_{\eta\epsilon,0} = \rho_0\sqrt{\nu_0}$ with $\rho_0 = -0.8$.

Note that while optimization is carried out over $\bar{\theta} = (d, \nu, \nu_2, \varphi')'$ for the CSS estimator, and over $\tilde{\psi} = (d, \sigma_\eta^2, \sigma_{\eta\epsilon}, \sigma_\epsilon^2, \varphi')'$ for the QML estimator, to simplify the interpretation results are reported for the transformed $\rho = \nu_2/\sqrt{\nu} = \sigma_{\eta\epsilon}^2/(\sigma_\eta \sigma_\epsilon)$ instead of reporting ν_2 or $\sigma_{\eta\epsilon}$.

For the correlated fractional UC model, table A.5 shows RMSE and bias for the estimated integration orders via CSS, QML, and the exact local Whittle estimator. As before, RMSE and bias are similar for CSS and QML, and decrease in n. While the fractional UC model outperforms most of the Whittle estimates, the latter performs surprisingly well for a bandwidth choice of $\alpha = 0.65$ for $n = 100$, and $\alpha = 0.70$ for $n = 200$. As before, estimates for the fractional UC model show little bias for d_0, while the benchmarks are significantly perturbed by the cyclical dynamics.

For the CSS estimator, table A.6 shows RMSE and bias for ν_0, ρ_0, and the autoregressive parameters both for the fractional UC model and the integer-integrated UC model, while those for the QML estimator are contained in table A.7. As in the uncorrelated case, CSS estimates for ν_0 exhibit a large RMSE. For $\nu_0 \leq 5$, the CSS estimator is typically upward-biased, whereas it is downward-biased for $\nu_0 = 10$. As can be expected, the bias is more pronounced for the $I(1)$ benchmark, where the RMSE is also higher. More interestingly, the benchmark estimates for ν_0 are typically upward-biased whenever $d_0 < 1$, and downward-biased whenever $d_0 > 1$. Since $\nu_0 = \sigma_{\epsilon,0}^2/\sigma_{\eta,0}^2$ is the variance ratio of the innovations, this is natural: Whenever $d_0 < 1$, the random walk for a fixed σ_η^2 has a faster diverging variance than the $I(d_0)$ process. To compensate for the slower rate of divergence of the $I(d_0)$ process, $\hat{\nu}$ must be upward-biased in the $I(1)$ model, and vice versa for $d_0 > 1$. For ρ_0, note that a similar pattern is visible whenever $\nu_0 = 1$: For $d_0 < 1$, estimates for the correlation between long- and short-run shocks are upward-biased, and sometimes even positive. This is due to the upward-biased $\hat{\nu}$, which yields an estimate for the trend that is smoother than the true one. Thus, the cycle needs to account for the additional long-run fluctuations that are not captured by the smooth trend, which can be achieved by a positive estimate for the correlation coefficient. For $d_0 > 1$, the smoothed trend of the $I(1)$ model is more volatile than the true one, and the $I(1)$ UC model re-adjusts by estimating a downward-biased correlation coefficient, resulting in a more negative relation between trend and cycle than in the data-generating mechanism. Note that the potential for adjustment of the $I(1)$ model to fractionally integrated trends via the correlation parameter estimate is limited by the nature of the correlation $\rho \in [-1; 1]$, and thus corner solutions with $\hat{\rho} = -1$ can be expected when d_0 is greater than one, and with $\hat{\rho} = 1$ whenever d_0 is smaller than one. As before, there are only little differences for the estimates of the autoregressive coefficients between the fractional model and the $I(1)$ model, except for $d_0 = 1.75$, where the estimates of the $I(1)$ UC model are heavily biased by the misspecification of the integration order.

From the QML results of the fractional UC model in table A.7, it becomes apparent that $\hat{\sigma}_{\eta QML}^2$, $\hat{\sigma}_{\epsilon QML}^2$ exhibit some bias and a higher RMSE, particularly when d_0 and ν_0 are high and n is small. Fortunately, both RMSE and bias decrease as the sample size increases, however the level of precision with which the variance parameters are estimated appears to be lower compared to the other parameters. In line with the CSS results, table A.7 shows a high RMSE for the estimate of $\sigma_{\eta,0}^2$ from the integer-integrated UC model whenever $d_0 = 1.75$, together with strong, positive bias. This is natural, as the higher variance parameter is required to capture the additional variation that is induced by the strong persistence and not captured by the $I(1)$ trend specification. A similar bias is visible for the estimate of $\sigma^2_{\epsilon,0}$ in the integer-integrated setup, indicating that also the cyclical component is perturbed by the integration order exceeding unity. As for the CSS estimator, for $\nu_0 = 1$ the correlation estimate $\hat{\rho}^{I(1)}_{QML}$ is upward-biased whenever $d_0 < 1$, and downward-biased whenever $d_0 > 1$, while no such bias is detected for the fractional UC model. Moreover, for $d_0 \leq 1.25$ the autoregressive parameters are estimated with great precision for both, fractional and $I(1)$ UC model, with both bias and RMSE slightly favoring the fractional model whenever $d_0 \neq 1$. Whenever $d_0 = 1.75$, estimates for the AR coefficients from the integer-integrated models are biased, as for the uncorrelated scenario.

Table A.8 compares the coefficients of determination for the smoothed trend and cycle components of the fractional and the $I(1)$ UC model. For the fractional UC model, the QML estimator typically has a minor advantage over the CSS estimator in terms of the coefficients of determination. Moreover, for $d_0 = 1$ the fractional UC model shows no efficiency loss compared to the $I(1)$ UC models. For $d \neq 1$, the fractional UC model outperforms the integer-integrated models, where the difference is particularly striking for $d_0 = 1.75$.

7 Application

In this section, I apply the fractional UC model to monthly global sea surface temperature anomalies. Trends and cycles of climate time series have recently attracted attention in the econometric literature, see Chang et al. (2020), Gadea Rivas and Gonzalo (2020), and Proietti and Maddanu (2022), however fractional trends have not played a role so far. Beyond estimating the memory parameter, which may be of interest in its own right, the fractional UC model allows to draw inference on trending and cyclical temperature phenomena, as well as on their interaction once correlation is allowed for. On the one hand, the estimate for d_0 allows to test for mean reversion of the trend. If rejected, the smoothed trend component reveals the extent of permanent temperature rise. On the other hand, the cyclical component of monthly global sea surface temperature can be matched with well-understood cyclical climate phenomena, such as El Niño and La Niña. Estimation results from the fractional UC model can be compared against those of $I(1)$ and $I(2)$ UC models. In particular, the hypothesis of an integer integration order is testable, and, if rejected, the fractional UC model sheds light on the extent to which trend and cycle estimates are perturbed when the trend memory is misspecified in traditional UC models.

Data on monthly global sea surface temperature anomalies stem from the National Centers for Environmental Information and are calculated based on the extended reconstructed sea surface data of Huang et al. (2017).[4] The series spans from January 1850 to July 2023, thus consists of 2083 observations, and is measured as the deviation from the 1901 – 2000 average in degrees Celsius.

[4]Data were accessed on 2023/09/12 and can be downloaded from `https://www.ncei.noaa.gov/access/monitoring/global-temperature-anomalies/anomalies`

To decompose temperature anomalies into trend and cycle, I specify the fractional UC model

$$y_t = x_t + c_t, \qquad \Delta_+^d x_t = \mu + \eta_t, \qquad \sum_{j=0}^{p} b_j c_{t-j} = \epsilon_t, \tag{27}$$

where $b_0 = 1$, and thus c_t is an autoregressive process of order p with all roots of $b(L) = \sum_{j=0}^{p} b_j L^j$ outside the unit circle, as typical in the UC literature. The specification of the trend allows for a non-zero mean in $\Delta_+^d x_t$, generating a deterministic trend of order d in y_t. This is a generalization of integer-integrated UC models, that allow either for a linear deterministic trend whenever $x_t \sim I(1)$ (see e.g. Harvey; 1985; Morley et al.; 2003) or for a quadratic one whenever $x_t \sim I(2)$ (see e.g. Clark; 1987; Oh et al.; 2008). Moreover, $\mathrm{Var}(\eta_t, \epsilon_t)' = Q$ is allowed to be non-diagonal.

Estimation of the fractional UC model is carried out by the QML estimator as described in subsection 5.3, as the QML estimator was found to be more accurate for the covariance parameters of trend and cycle innovations in the simulation studies in section 6 than the CSS estimator. To estimate the fractional UC model, I draw 100 combinations of starting values from uniform distributions with appropriate support.[5] As numerical optimization of the quasi-likelihood of the fractional UC model is computationally intensive for $n = 2083$ observations, I use ARMA(3, 3) approximations for the fractional differencing operator as suggested by Hartl and Jucknewitz (2022) to speed up the grid search: As they describe in great detail, a continuous function that maps from d onto the six ARMA(3, 3) coefficients is obtained first by choosing those six ARMA coefficients that minimize the Euclidean distance between the Wold representation of the fractional differencing polynomial and the Wold representation of the ARMA polynomials for a sequence of d (here: $d \in [0; 2.5]$). Next, the mapping is made continuous by smoothing over the sequence of d, as well as the ARMA coefficients, using splines. Consequently, optimization is carried out over d, however the use of ARMA(3, 3) approximations yields a low-dimensional state space representation of the (approximate) fractional UC model and thus greatly speeds up the computations. Finally, the estimate that maximizes the likelihood of the (approximate) fractional UC model is taken as starting value for the numerical likelihood maximization of the (exact) fractional UC model. Estimation is carried out for $p \in [1; 3; ...; 12]$ autoregressive lags, and $p = 4$ is selected as this minimizes both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for the (exact) fractional UC model. In addition to the QML estimates of the fractional UC model, I also present estimation results for an $I(1)$ and an $I(2)$ UC model that set $d = 1$ and $d = 2$ in (27) respectively.[6]

Table A.9 contains the estimation results for the fractional UC model and the two integer-integrated benchmarks. All models allow for $p = 4$ autoregressive lags in (27), as suggested by the AIC for all models.[7] The QML estimator for the fractional UC model yields $\hat{d}_{QML} = 1.753$, together with a 95% confidence interval $[1.634; 1.872]$, and a 99% confidence interval $[1.596; 1.909]$. Consequently, both hypotheses that $d_0 = 1$ and $d_0 = 2$ are rejected, supporting a specification of the

[5]More precisely, d is drawn from $[1/2; 2]$, Q is drawn from reasonable combinations of σ_η^2, σ_ϵ^2, and $\sigma_{\eta\epsilon}$ that can generate the realized variation in the observable y_t, and autoregressive parameters are drawn randomly from the set of coefficients that ensure the cyclical AR polynomial to be stable.

[6]Estimation for the benchmark models is carried out as for the fractional UC model, i.e. via the QML estimator where starting values are chosen via a grid search with 100 grid points.

[7]The BIC suggests two autoregressive lags for the benchmarks.

trend component with a longer memory than a random walk, but a shorter memory than a quadratic trend. The estimated variance ratio of short- and long-run innovations $\hat{\nu}_{QML} = \hat{\sigma}^2_{\epsilon_{QML}}/\hat{\sigma}^2_{\eta_{QML}} = 146621$ reveals a very smooth trend component and leaves rich variation to the cycle. Although the estimate for $\sigma^2_{\eta,0}$ is small, the hypothesis that the long-run component is purely deterministic (i.e. $\sigma^2_{\eta,0} = \sigma_{\eta\epsilon,0} = 0$) is rejected on all conventional levels of significance, as the log likelihood of the restricted model is 5420.9, such that the test statistic of the likelihood ratio test for the respective hypothesis is 31.4. Estimates for the autoregressive coefficients suggest a persistent cyclical pattern, with the greatest eigenvalue of the AR polynomial being 0.92. Moreover, long- and short-run innovations are found to be mildly negatively correlated.

In line with simulation results in section 6, estimates for the autoregressive coefficients are very similar for the fractional UC model and the two benchmarks, while the variance-covariance estimates for long- and short-run innovations are strongly biased for the integer-integrated models: As also noted in section 6, if in integer-integrated models the integration order of the trend is assumed lower than in the data-generating mechanism, the additional long-run variation not captured by the trend specification upward-biases the estimate for the variance of the long-run innovations. Vice versa, if an integration order higher than in the data-generating mechanism is assumed, the estimate for $\sigma^2_{\eta,0}$ will be downward-biased. Consequently, the estimate for $\sigma^2_{\eta,0}$ from the $I(2)$ benchmark is smaller than the one from the fractional UC model, while the estimate from the $I(1)$ benchmark is greater. Moreover, both benchmarks converge towards the corner solution of (almost) perfectly correlated long- and short-run innovations. This behavior is again in line with the simulation results in table A.7 for integration orders 0.75 and 1.75, and a variance ratio $\nu > 1$.

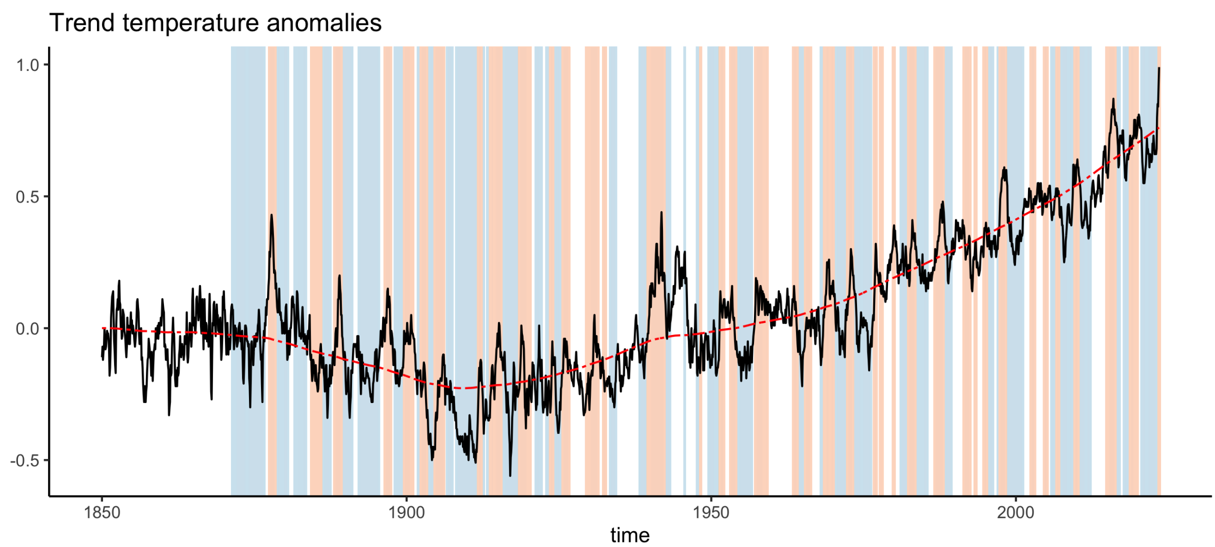

Figure 1: Trend temperature anomalies: The plot shows monthly global sea surface temperature anomalies (black) together with the estimated trend $\hat{x}_t(y_{n:1}, \hat{\psi}_{QML})$ (red, dashed) from the fractional UC model. Shaded areas correspond to warm (red) and cold (blue) periods based on a threshold of $\pm 1/2$ degree Celsius for the Oceanic Niño Index (ONI).[9]

[9]From 1950 on, the ONI is reported by the Climate Prediction Center of the National Weather Service and can be downloaded from `https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_`

Figure 1 plots the smoothed trend estimate $\hat{x}_t(y_{n:1}, \hat{\psi}_{QML})$, together with the series for monthly global sea surface temperature anomalies. The smooth nature of the estimated trend component follows directly from the high estimate of the integration order and the low estimate for the variance of the long-run innovations. While the first half of the sample does not clearly point towards a decreasing or increasing nature of the trend component, at least since the mid 20th century trend temperature anomalies are strictly increasing. In July 2023, the last observational period, the estimated trend component equals +0.76 degrees Celsius.

Figure A.1 allows to compare the trend estimate from the fractional UC model to those of the $I(1)$ UC model, the $I(2)$ UC model, and the HP filter with tuning parameter $\lambda = 14,400$ as typical for monthly data. Contrary to the fractional model, the benchmarks attribute significant short-run variation to the trend component: Clearly, the $I(1)$ UC model yields a much more erratic trend that behaves countercyclical, i.e. it increases during the cold Niña periods and decreases during the warm Niño periods. HP filter and the $I(2)$ benchmark attribute more of the overall variation to the trend component, as their estimates for the trend match the observable series much more closely compared to the smoothed trend component of the fractional UC model. Obviously, the additional short-run dynamics in the benchmark models are generated by the (almost) perfect negative correlation coefficient that ties trend and cycle component together, generating (spurious) cyclical dynamics in the trend component.

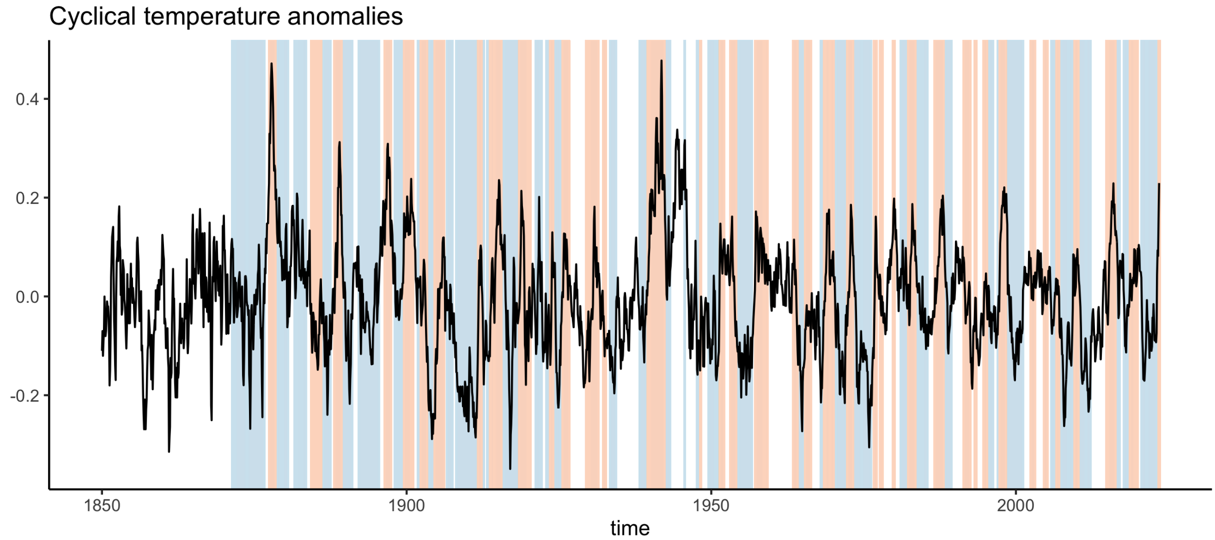

Figure 2: Cyclical temperature anomalies: The plot shows estimated cyclical sea surface temperature anomalies $\hat{c}_t(y_{n:1}, \hat{\theta})$ from the fractional UC model. Shaded areas correspond to warm (red) and cold (blue) periods according to the Oceanic Niño Index (ONI), see figure 1 for details.

Figure 2 shows the smoothed cyclical component $\hat{c}_t(y_{n:1}, \hat{\psi}_{QML})$ for the fractional UC model. As already noted above, the estimates for the autoregressive parameters as well as for the variance-

`v5.php`. As the ONI is not available for the years prior to 1950, I use the extended multivariate ENSO index (MEI.ext) of Wolter and Timlin (2011) that starts in 1871 and can be downloaded from `https://psl.noaa.gov/enso/mei.ext/`. The latter is scaled to arrive at the same standard deviation as the ONI. Since the MEI.ext is a bi-monthly rolling average, a month is considered a cold (warm) month once the bi-monthly rolling average of the current and the following month crosses the threshold.

ratio of short- and long-run innovations attribute rich variation to the cyclical component and generate a persistent series. Clearly, $\hat{c}_t(y_{n:1}, \hat{\psi}_{QML})$ evolves along the Oceanic Niño index, as peaks typically occur during El Niño phases and are followed by troughs during La Niña.

Figure A.2 highlights the differences between the smoothed cyclical component of the fractional UC model and those of the three benchmarks. Setting the integration order to unity attributes additional pro-cyclical variation (in terms of the ONI) to the smoothed cycle. This is straightforward, as the smoothed trend component of the $I(1)$ UC model was found to behave anti-cyclical. HP filter and the $I(2)$ UC model yield similar deviations from the cyclical component of the fractional UC model. They dampen the cyclical variation, because their respective trend components follow the observable series more closely, leaving fewer variation to be captured by the cycle.

Finally, figure A.3 plots the estimated autocorrelation function up to 48 lags for the one-step ahead forecast errors of the fractional UC model and the two integer-integrated benchmarks. As can be seen, misspecifying the integration order to either one or two generates spurious, strongly persistent autocorrelation in the prediction errors, thus violating the MDS assumption. In contrast, little to no autocorrelation is left in the prediction errors of the fractional UC models.

8 Conclusion

This paper introduces a novel unobserved components model in which the trend component is specified as a type II fractionally integrated process. The model encompasses the bulk of unobserved components models in the literature, allows for richer long-run dynamics beyond integer-integrated specifications, and for a data-dependent specification of the trend. Trend and cycle are estimated via the analytical solution to the optimization problem of the Kalman filter. The model allows for a joint estimation of the integration order and the other model parameters via the conditional sum-of-squares estimator, which is shown to be consistent and asymptotically normally distributed. While the asymptotic estimation theory is derived for a prototypical model, it is shown to carry over to models with deterministic components, correlated long- and short-run innovations, and quasi-maximum likelihood estimation. For monthly global sea surface temperature anomalies, the fractional unobserved components model reveals a smooth trend component that is increasing since the mid of the 20th century, together with a rich cyclical component that matches the Oceanic Niño index.

To applied researchers, the fractional unobserved components model offers a robust, flexible, and data-driven method for signal extraction of data of unknown persistence. It does not require prior assumptions about the integration order, nor the choice of any tuning parameter. Therefore, it provides a solution to the model specification problem in the unobserved components literature, and calls for further applications beyond temperature anomalies.

A Additional figures and tables

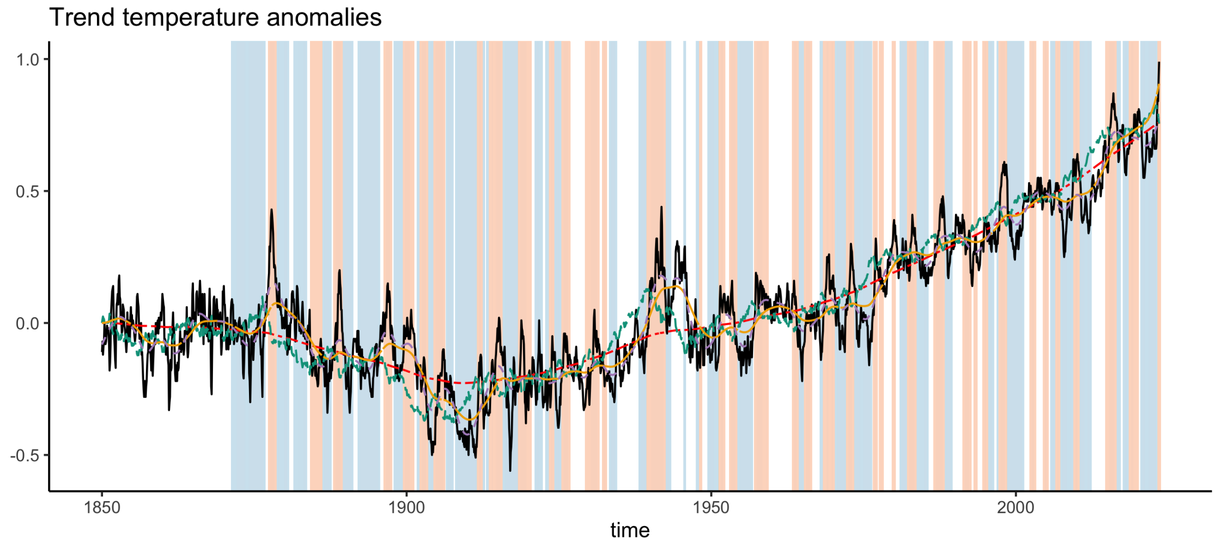

Figure A.1: Smoothed trend component of monthly global sea surface temperature anomalies (relative to 1900-2000 average in degrees Celsius) via the fractional UC model (red), the $I(1)$ UC model (green), the $I(2)$ UC model (yellow), and the HP filter with $\lambda = 14,400$ (purple). The original series is plotted in black. Shaded areas correspond to warm (red) and cold (blue) periods according to the Oceanic Niño Index (ONI), see figure 1 for details

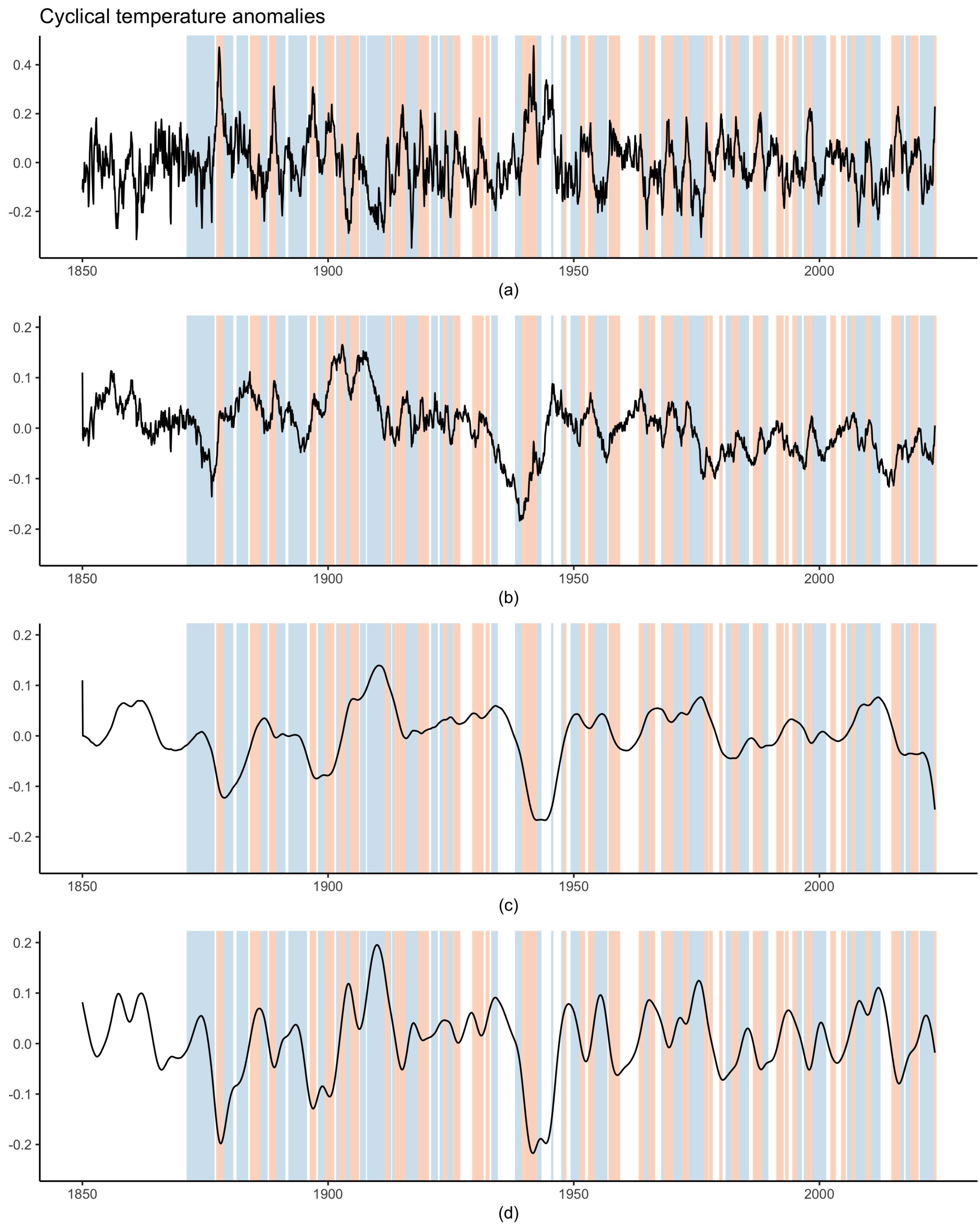

Figure A.2: Deviations from smoothed cyclical component of monthly global sea surface temperature anomalies (relative to 1900-2000 average in degrees Celsius): Figure (a) shows the smoothed cyclical component of the fractional UC model, while all other plots show the deviations of the respective smoothed cyclical component from the fractional UC model for (b) the $I(1)$ UC model, (c) the $I(2)$ UC model, and (d) the HP filter with $\lambda = 14,400$ (purple). Consequently, smoothed cyclical components of the integer-integrated models are obtained by adding (a) to the second, third, and fourth figure respectively. Shaded areas correspond to warm (red) and cold (blue) periods according to the Oceanic Niño Index (ONI), see figure 1 for details

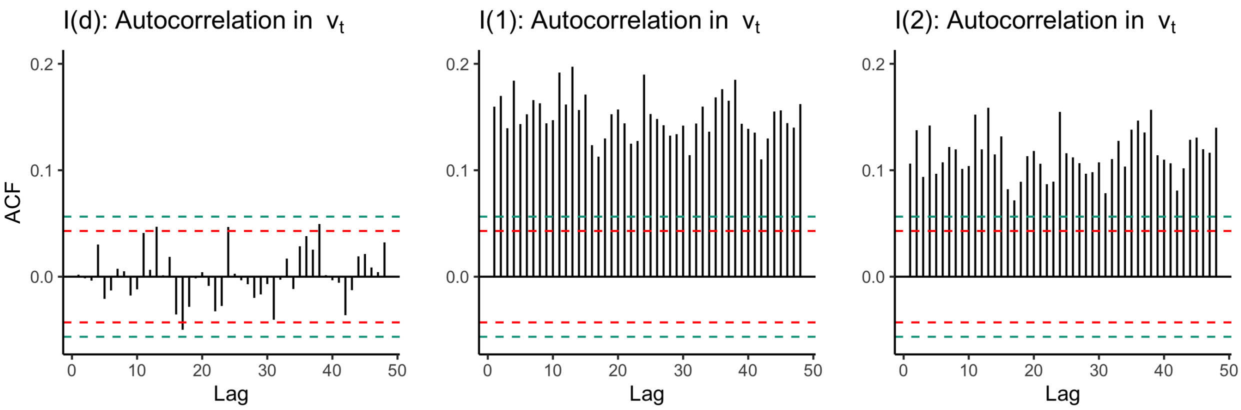

Figure A.3: Estimated autocorrelation function of the prediction errors for the fractional UC model (left), the I(1) UC model (center), and the I(2) UC model (right), together with 5% (red) and 1% (blue) confidence bands.

			RMSE							bias						
n	ν_0	d_0	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$
100	1	.75	.154	.131	.638	.579	.410	.228	.574	-.049	-.032	-.621	-.554	-.361	.043	.521
		1.00	.183	.132	.681	.614	.460	.222	.397	-.038	-.035	-.650	-.581	-.419	-.099	.327
		1.25	.186	.130	.651	.591	.464	.258	.258	-.007	-.032	-.612	-.554	-.425	-.185	.163
		1.75	.166	.119	.546	.507	.425	.298	.158	-.011	-.033	-.496	-.463	-.383	-.252	-.062
	5	.75	.189	.168	.714	.673	.507	.268	.743	-.052	-.051	-.709	-.662	-.466	.054	.695
		1.00	.224	.195	.871	.810	.638	.289	.526	-.058	-.074	-.858	-.793	-.608	-.156	.459
		1.25	.205	.179	.903	.842	.694	.382	.338	-.036	-.060	-.880	-.818	-.668	-.317	.233
		1.75	.205	.144	.820	.779	.685	.505	.234	-.057	-.042	-.789	-.752	-.660	-.477	-.143
	10	.75	.197	.188	.726	.690	.527	.276	.773	-.062	-.062	-.722	-.681	-.487	.055	.726
		1.00	.247	.235	.919	.866	.692	.309	.549	-.102	-.103	-.911	-.854	-.664	-.171	.483
		1.25	.239	.217	.995	.934	.779	.426	.354	-.093	-.079	-.978	-.915	-.755	-.359	.245
		1.75	.234	.153	.938	.894	.795	.593	.268	-.109	-.042	-.910	-.869	-.774	-.568	-.173
200	1	.75	.108	.092	.618	.642	.568	.389	.139	-.033	-.022	-.603	-.633	-.555	-.369	.030
		1.00	.106	.087	.598	.637	.563	.415	.153	-.016	-.020	-.571	-.619	-.546	-.397	-.098
		1.25	.110	.082	.530	.584	.526	.407	.200	.004	-.017	-.498	-.563	-.508	-.389	-.169
		1.75	.121	.076	.390	.465	.436	.358	.241	.010	-.012	-.343	-.438	-.414	-.339	-.220
	5	.75	.152	.133	.722	.732	.697	.521	.164	-.049	-.045	-.719	-.731	-.692	-.506	.036
		1.00	.151	.127	.821	.852	.784	.615	.221	-.022	-.040	-.808	-.843	-.774	-.603	-.167
		1.25	.141	.108	.786	.835	.773	.640	.335	.001	-.029	-.765	-.820	-.760	-.628	-.312
		1.75	.158	.086	.642	.724	.694	.609	.448	-.020	-.015	-.615	-.707	-.679	-.596	-.436
	10	.75	.164	.153	.736	.743	.719	.553	.169	-.060	-.060	-.735	-.743	-.717	-.539	.036
		1.00	.175	.161	.890	.914	.857	.683	.241	-.053	-.059	-.882	-.908	-.850	-.672	-.186
		1.25	.149	.124	.890	.934	.870	.729	.384	-.037	-.036	-.872	-.920	-.859	-.719	-.361
		1.75	.171	.085	.753	.833	.800	.713	.537	-.056	-.013	-.731	-.817	-.787	-.702	-.527
300	1	.75	.091	.076	.508	.607	.603	.494	.216	-.022	-.013	-.487	-.596	-.594	-.484	-.194
		1.00	.078	.071	.448	.577	.581	.487	.272	-.011	-.011	-.417	-.559	-.569	-.476	-.257
		1.25	.081	.067	.369	.515	.534	.457	.290	.002	-.011	-.330	-.495	-.520	-.445	-.277
		1.75	.104	.062	.242	.385	.428	.381	.280	.012	-.007	-.178	-.359	-.412	-.368	-.266
	5	.75	.125	.112	.671	.724	.723	.650	.305	-.030	-.028	-.664	-.722	-.721	-.644	-.285
		1.00	.118	.099	.682	.795	.796	.701	.431	-.007	-.023	-.665	-.785	-.788	-.693	-.421
		1.25	.119	.086	.611	.754	.769	.691	.491	.009	-.018	-.589	-.740	-.758	-.681	-.482
		1.75	.140	.067	.439	.629	.673	.625	.507	-.008	-.008	-.407	-.612	-.661	-.615	-.498
	10	.75	.143	.132	.707	.739	.739	.687	.326	-.042	-.041	-.703	-.739	-.738	-.683	-.306
		1.00	.136	.121	.771	.869	.870	.778	.484	-.027	-.033	-.758	-.862	-.864	-.771	-.474
		1.25	.121	.098	.715	.850	.862	.782	.568	-.021	-.021	-.697	-.838	-.852	-.773	-.561
		1.75	.147	.065	.545	.732	.773	.725	.603	-.039	-.005	-.520	-.717	-.762	-.715	-.595

Table A.1: Root mean squared errors (RMSE) and bias for the integration order estimates of the fractional UC model with uncorrelated innovations in subsection 6.1. The columns indicate the integration order estimates via the CSS estimator ($\hat{d}_{CSS}$), the QML estimator ($\hat{d}_{QML}$), and the Whittle estimator of Shimotsu and Phillips (2005) with tuning parameter α ($\hat{d}^{EW}_{\alpha}$).

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1CSS}$	$\hat{b}_{1QML}$	$\hat{b}^{I(1)}_{1CSS}$	$\hat{b}^{I(1)}_{1QML}$	$\hat{b}_{2CSS}$	$\hat{b}_{2QML}$	$\hat{b}^{I(1)}_{2CSS}$	$\hat{b}^{I(1)}_{2QML}$
100	1	.75	3.874	.316	10.518	.414	.108	.092	.178	.137	.107	.085	.171	.127
		1.00	5.588	.315	5.969	.311	.110	.090	.115	.095	.112	.086	.119	.092
		1.25	7.149	.300	7.728	.395	.106	.098	.087	.079	.110	.085	.119	.091
		1.75	6.085	.281	9.608	5.011	.110	.109	.233	.233	.105	.080	.137	.081
	5	.75	8.863	.960	15.505	1.035	.085	.078	.107	.090	.081	.073	.101	.084
		1.00	10.246	.947	10.367	.948	.083	.076	.085	.079	.080	.072	.084	.076
		1.25	10.441	.937	6.254	1.160	.081	.077	.074	.070	.081	.074	.079	.081
		1.75	9.132	.997	9.996	3.332	.091	.080	.171	.182	.090	.073	.106	.071
	10	.75	9.084	1.759	13.285	1.774	.080	.077	.090	.083	.077	.073	.084	.078
		1.00	9.886	1.721	10.307	1.707	.079	.075	.079	.077	.077	.071	.077	.074
		1.25	9.864	1.705	8.095	2.062	.080	.076	.072	.071	.080	.072	.074	.079
		1.75	9.366	1.778	9.773	4.683	.090	.078	.132	.150	.089	.072	.102	.086
200	1	.75	1.105	.263	9.095	.319	.068	.061	.132	.091	.070	.060	.131	.086
		1.00	3.246	.244	2.889	.219	.071	.060	.070	.058	.075	.058	.074	.058
		1.25	4.710	.207	4.855	.283	.072	.059	.064	.056	.079	.057	.084	.065
		1.75	4.811	.200	9.324	5.716	.075	.069	.274	.245	.078	.056	.136	.087
	5	.75	7.282	.726	16.253	.761	.054	.049	.076	.057	.055	.048	.075	.054
		1.00	8.618	.690	7.711	.669	.052	.049	.053	.048	.054	.048	.055	.048
		1.25	9.429	.650	3.774	.786	.054	.049	.054	.045	.055	.048	.055	.053
		1.75	8.494	.684	9.327	3.417	.056	.050	.206	.202	.058	.047	.105	.066
	10	.75	8.572	1.301	13.867	1.270	.050	.047	.058	.051	.050	.047	.057	.049
		1.00	9.362	1.252	9.181	1.199	.050	.047	.048	.046	.050	.046	.049	.047
		1.25	9.485	1.206	7.078	1.405	.052	.047	.050	.043	.052	.047	.051	.051
		1.75	9.206	1.342	9.550	3.460	.056	.048	.169	.173	.057	.046	.094	.066
300	1	.75	.654	.210	7.623	.277	.054	.051	.108	.077	.055	.049	.107	.071
		1.00	1.346	.219	1.712	.179	.055	.050	.056	.048	.058	.048	.059	.047
		1.25	3.091	.181	3.216	.228	.057	.049	.058	.048	.061	.047	.068	.052
		1.75	3.385	.166	9.118	6.339	.063	.054	.278	.243	.066	.046	.143	.091
	5	.75	5.918	.574	16.197	.624	.045	.042	.067	.047	.045	.040	.066	.044
		1.00	7.560	.557	6.448	.535	.043	.041	.043	.039	.045	.040	.044	.039
		1.25	8.615	.546	3.254	.603	.043	.041	.051	.039	.044	.040	.051	.041
		1.75	7.814	.543	9.614	3.540	.048	.041	.227	.214	.048	.039	.111	.064
	10	.75	8.028	1.035	14.114	1.014	.042	.040	.049	.042	.042	.039	.048	.040
		1.00	8.932	.998	8.131	.952	.041	.039	.040	.038	.041	.039	.041	.038
		1.25	8.939	.969	6.972	1.081	.043	.039	.047	.037	.044	.039	.048	.040
		1.75	8.858	1.095	9.575	3.165	.048	.040	.181	.185	.049	.038	.106	.059

Table A.2: Root mean squared errors (RMSE) for the other parameter estimates of the fractional UC model with uncorrelated innovations in subsection 6.1. The different columns indicate the parameter estimates via the CSS estimator (subscript CSS) and the QML estimator (subscript QML) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1CSS}$	$\hat{b}_{1QML}$	$\hat{b}^{I(1)}_{1CSS}$	$\hat{b}^{I(1)}_{1QML}$	$\hat{b}_{2CSS}$	$\hat{b}_{2QML}$	$\hat{b}^{I(1)}_{2CSS}$	$\hat{b}^{I(1)}_{2QML}$
100	1	.75	.981	-.007	4.832	.188	-.022	-.019	-.062	-.077	.014	.011	.044	.062
		1.00	1.721	.026	1.786	.017	-.021	-.022	-.022	-.028	.014	.013	.015	.020
		1.25	2.587	-.003	2.490	.068	-.016	-.022	.027	.021	.010	.011	.003	.017
		1.75	2.130	.015	5.096	4.181	.018	-.015	.188	.220	-.014	.006	.005	-.024
	5	.75	3.959	-.191	11.594	.322	-.015	-.018	-.048	-.042	.006	.010	.037	.032
		1.00	4.886	-.184	4.828	.069	-.010	-.017	-.018	-.025	.003	.010	.011	.020
		1.25	4.699	-.117	-.715	.234	-.005	-.018	.022	.000	-.003	.012	-.024	.023
		1.75	2.833	.103	3.024	3.001	.007	-.017	.141	.167	-.023	.012	.005	.023
	10	.75	2.824	-.341	9.959	.389	-.006	-.019	-.034	-.035	-.002	.011	.024	.026
		1.00	3.220	-.339	4.040	.133	-.004	-.017	-.016	-.024	-.004	.011	.008	.019
		1.25	2.429	-.175	-3.712	.473	-.001	-.019	.015	-.006	-.008	.013	-.020	.024
		1.75	.214	.208	-4.299	2.860	.006	-.019	.107	.131	-.023	.014	-.017	.054
200	1	.75	.240	.028	3.636	.184	-.009	-.010	-.045	-.056	.009	.009	.036	.049
		1.00	.682	.081	.496	.014	-.010	-.014	-.008	-.012	.010	.013	.008	.012
		1.25	1.265	.010	.827	.060	-.008	-.009	.039	.027	.009	.008	-.013	.012
		1.75	1.395	.027	5.405	5.169	.011	-.010	.241	.231	-.008	.008	-.043	-.033
	5	.75	3.038	-.080	13.024	.305	-.007	-.006	-.042	-.027	.006	.005	.038	.022
		1.00	3.714	-.070	3.002	.044	-.005	-.007	-.008	-.011	.005	.006	.007	.011
		1.25	4.146	-.061	-2.433	.146	-.002	-.007	.034	.010	.001	.007	-.032	.012
		1.75	2.841	.056	2.886	3.199	.005	-.009	.178	.191	-.010	.008	-.014	.006
	10	.75	3.068	-.143	11.996	.339	-.002	-.006	-.027	-.021	.000	.005	.024	.016
		1.00	3.423	-.126	3.582	.071	-.000	-.007	-.007	-.011	-.000	.006	.005	.010
		1.25	2.802	-.102	-5.817	.320	.002	-.008	.027	.005	-.004	.007	-.028	.014
		1.75	1.444	.196	-3.418	1.438	.006	-.009	.145	.163	-.010	.009	-.046	.033
300	1	.75	.148	.028	2.494	.174	-.006	-.005	-.034	-.049	.005	.005	.025	.042
		1.00	.237	.107	.236	.008	-.006	-.012	-.005	-.007	.005	.012	.004	.006
		1.25	.643	-.001	.195	.040	-.005	-.005	.042	.028	.005	.003	-.024	.005
		1.75	.939	.017	5.161	5.917	.007	-.007	.242	.227	-.008	.004	-.043	-.029
	5	.75	2.161	-.037	13.312	.274	-.004	-.003	-.041	-.022	.004	.002	.038	.017
		1.00	2.977	-.051	2.107	.014	-.003	-.003	-.004	-.006	.002	.002	.003	.005
		1.25	3.576	-.065	-2.804	.087	-.001	-.003	.037	.014	-.000	.003	-.036	.006
		1.75	2.696	.003	3.670	3.359	.006	-.005	.198	.204	-.009	.004	-.024	-.006
	10	.75	2.894	-.068	12.927	.284	-.000	-.003	-.025	-.015	-.001	.002	.022	.011
		1.00	3.169	-.070	2.731	.019	.002	-.003	-.003	-.005	-.002	.003	.002	.005
		1.25	2.540	-.090	-6.350	.235	.004	-.004	.031	.009	-.005	.003	-.032	.008
		1.75	1.684	.116	-3.530	.941	.006	-.006	.152	.176	-.010	.005	-.052	.021

Table A.3: Bias for the other parameter estimates of the fractional UC model with uncorrelated innovations in subsection 6.1. The different columns indicate the parameter estimates via the CSS estimator (subscript CSS) and the QML estimator (subscript QML) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

n	ν_0	d_0	Trend R^2_{CSS}	R^2_{QML}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$	Cycle R^2_{CSS}	R^2_{QML}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$
100	1	.75	.506	.528	.484	.523	.839	.849	.814	.841
		1.00	.751	.781	.762	.786	.771	.789	.776	.793
		1.25	.901	.922	.865	.885	.702	.725	.621	.618
		1.75	.984	.993	.679	.735	.536	.594	.045	.039
	5	.75	.294	.306	.323	.329	.944	.948	.938	.943
		1.00	.592	.609	.617	.633	.905	.911	.907	.918
		1.25	.830	.842	.828	.818	.861	.870	.855	.799
		1.75	.981	.983	.778	.717	.760	.781	.226	.084
	10	.75	.229	.235	.278	.279	.965	.969	.961	.966
		1.00	.511	.522	.550	.565	.935	.939	.938	.946
		1.25	.780	.788	.791	.774	.897	.905	.899	.852
		1.75	.975	.975	.859	.722	.816	.832	.440	.124
200	1	.75	.625	.637	.597	.628	.850	.857	.829	.849
		1.00	.868	.877	.871	.879	.793	.802	.797	.805
		1.25	.967	.971	.933	.943	.735	.746	.667	.644
		1.75	.998	.999	.798	.831	.588	.626	.013	.013
	5	.75	.394	.408	.405	.408	.945	.948	.942	.941
		1.00	.743	.755	.748	.763	.909	.913	.911	.917
		1.25	.929	.932	.925	.913	.872	.876	.867	.817
		1.75	.997	.997	.847	.835	.788	.797	.149	.024
	10	.75	.311	.320	.338	.330	.967	.968	.965	.963
		1.00	.671	.681	.684	.697	.937	.939	.940	.944
		1.25	.901	.903	.900	.883	.906	.908	.904	.857
		1.75	.995	.996	.901	.830	.835	.841	.404	.037
300	1	.75	.689	.697	.664	.686	.856	.860	.835	.849
		1.00	.909	.914	.912	.915	.801	.806	.804	.808
		1.25	.982	.984	.964	.967	.744	.750	.703	.675
		1.75	1.000	1.000	.826	.834	.610	.635	.008	.008
	5	.75	.482	.488	.480	.477	.947	.948	.943	.940
		1.00	.815	.823	.818	.828	.913	.915	.914	.917
		1.25	.959	.961	.959	.949	.875	.878	.873	.833
		1.75	.999	.999	.851	.839	.793	.800	.102	.013
	10	.75	.394	.399	.404	.390	.967	.967	.965	.961
		1.00	.759	.765	.766	.774	.939	.941	.940	.943
		1.25	.941	.943	.941	.929	.908	.910	.907	.869
		1.75	.998	.998	.919	.843	.838	.843	.388	.018

Table A.4: Coefficient of determination from regressing true trend and cycle x_t and c_t on their respective estimates from the Kalman smoother for the uncorrelated UC models.

			RMSE							bias						
n	ν_0	d_0	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$
100	1	.75	.149	.131	.639	.581	.415	.247	.681	-.018	-.033	-.624	-.560	-.367	.071	.636
		1.00	.204	.174	.618	.532	.357	.235	.616	.025	-.014	-.582	-.491	-.293	.089	.574
		1.25	.219	.191	.550	.458	.286	.255	.566	.030	-.001	-.501	-.403	-.201	.149	.537
		1.75	.237	.219	.463	.369	.218	.201	.245	-.051	-.106	-.397	-.291	-.087	.155	.243
	5	.75	.171	.178	.726	.692	.528	.285	.851	-.035	-.035	-.723	-.684	-.489	.078	.809
		1.00	.209	.220	.853	.784	.592	.262	.681	-.043	-.031	-.839	-.766	-.556	-.055	.633
		1.25	.224	.231	.854	.772	.589	.269	.516	-.052	-.034	-.830	-.746	-.556	-.137	.463
		1.75	.297	.264	.770	.700	.557	.305	.191	-.136	-.150	-.739	-.670	-.524	-.239	.123
	10	.75	.173	.204	.733	.705	.546	.287	.856	-.072	-.005	-.731	-.698	-.507	.078	.814
		1.00	.236	.252	.911	.851	.659	.281	.663	-.131	-.018	-.902	-.838	-.627	-.096	.610
		1.25	.269	.245	.960	.882	.693	.328	.479	-.149	-.039	-.940	-.860	-.665	-.223	.413
		1.75	.338	.260	.896	.829	.692	.429	.188	-.193	-.155	-.869	-.805	-.667	-.384	.038
200	1	.75	.123	.102	.609	.636	.569	.396	.159	-.005	-.017	-.594	-.627	-.557	-.378	.047
		1.00	.160	.136	.554	.585	.495	.323	.154	.024	.009	-.526	-.566	-.477	-.299	.069
		1.25	.165	.151	.495	.517	.420	.248	.184	.020	.012	-.461	-.495	-.399	-.215	.127
		1.75	.178	.158	.422	.432	.326	.164	.191	-.031	-.057	-.382	-.404	-.296	-.107	.168
	5	.75	.143	.147	.730	.738	.713	.551	.191	-.019	-.024	-.728	-.737	-.710	-.538	.045
		1.00	.152	.153	.804	.834	.761	.580	.188	-.024	-.019	-.791	-.824	-.750	-.567	-.082
		1.25	.174	.174	.761	.800	.725	.563	.215	-.039	-.034	-.741	-.785	-.712	-.550	-.155
		1.75	.237	.191	.658	.709	.645	.515	.267	-.106	-.085	-.633	-.691	-.630	-.500	-.240
	10	.75	.153	.165	.742	.746	.732	.581	.194	-.069	-.013	-.742	-.746	-.731	-.568	.042
		1.00	.187	.175	.880	.904	.841	.659	.217	-.097	-.012	-.872	-.898	-.834	-.647	-.126
		1.25	.216	.182	.863	.902	.830	.669	.289	-.113	-.026	-.847	-.889	-.818	-.657	-.243
		1.75	.263	.188	.760	.817	.763	.643	.397	-.138	-.086	-.739	-.801	-.750	-.631	-.379
300	1	.75	.104	.078	.491	.603	.599	.498	.223	.001	-.012	-.469	-.592	-.590	-.488	-.201
		1.00	.131	.104	.412	.538	.530	.420	.169	.023	.012	-.380	-.521	-.516	-.407	-.140
		1.25	.134	.122	.364	.478	.463	.347	.114	.008	.008	-.325	-.458	-.448	-.331	-.064
		1.75	.146	.134	.310	.404	.379	.253	.095	-.020	-.039	-.263	-.380	-.360	-.231	.029
	5	.75	.120	.123	.688	.733	.733	.676	.334	-.017	-.017	-.682	-.732	-.732	-.672	-.312
		1.00	.130	.125	.669	.787	.782	.681	.388	-.019	-.013	-.652	-.777	-.774	-.672	-.374
		1.25	.158	.153	.600	.737	.738	.644	.396	-.038	-.036	-.579	-.724	-.728	-.634	-.385
		1.75	.197	.156	.489	.638	.650	.572	.386	-.076	-.056	-.462	-.623	-.639	-.561	-.376
	10	.75	.136	.140	.722	.745	.745	.709	.355	-.059	-.023	-.719	-.745	-.745	-.707	-.332
		1.00	.163	.145	.765	.869	.864	.767	.456	-.081	-.016	-.752	-.861	-.858	-.760	-.443
		1.25	.196	.165	.703	.837	.838	.746	.497	-.099	-.039	-.686	-.825	-.828	-.738	-.487
		1.75	.215	.158	.581	.738	.756	.686	.515	-.101	-.059	-.558	-.725	-.746	-.677	-.507

Table A.5: Root mean squared errors (RMSE) and bias for the integration order estimates of the fractional UC model with correlated innovations in subsection 6.2. The different columns indicate the integration order estimates via the CSS estimator ($\hat{d}_{CSS}$), the QML estimator ($\hat{d}_{QML}$), and the Whittle estimator of Shimotsu and Phillips (2005) with tuning parameter α ($\hat{d}^{EW}_{\alpha}$).

			RMSE								bias							
n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{CSS}^{I(1)}$	$\hat{\rho}_{CSS}$	$\hat{\rho}_{CSS}^{I(1)}$	$\hat{b}_{1CSS}$	$\hat{b}_{1CSS}^{I(1)}$	$\hat{b}_{2CSS}$	$\hat{b}_{2CSS}^{I(1)}$	$\hat{\nu}_{CSS}$	$\hat{\nu}_{CSS}^{I(1)}$	$\hat{\rho}_{CSS}$	$\hat{\rho}_{CSS}^{I(1)}$	$\hat{b}_{1CSS}$	$\hat{b}_{1CSS}^{I(1)}$	$\hat{b}_{2CSS}$	$\hat{b}_{2CSS}^{I(1)}$
100	1	.75	4.532	10.347	.603	1.427	.099	.117	.095	.111	1.314	7.715	.145	1.270	.016	-.042	-.013	.032
		1.00	5.927	5.172	.664	.498	.111	.120	.100	.106	2.155	1.722	.212	.186	.010	-.025	-.010	.022
		1.25	5.154	3.284	.547	.202	.153	.291	.115	.236	1.680	.642	.145	-.140	.011	-.132	-.007	.139
		1.75	2.283	1.887	.358	.196	.185	.507	.160	.539	.386	.381	.035	-.194	.025	-.411	-.018	.478
	5	.75	7.072	14.366	.206	.298	.087	.082	.082	.075	2.660	12.278	-.092	-.018	.016	-.026	-.011	.018
		1.00	6.941	7.278	.245	.209	.088	.089	.083	.085	2.413	2.977	-.097	-.026	.017	-.024	-.013	.018
		1.25	6.748	4.321	.311	.166	.097	.119	.083	.104	2.001	-2.024	-.058	-.104	.019	-.028	-.011	.028
		1.75	6.661	3.999	.447	.194	.152	.440	.115	.435	.502	-3.576	.079	-.190	.028	-.326	-.014	.352
	10	.75	7.169	11.771	.190	.306	.084	.079	.080	.073	-1.416	9.409	-.083	-.005	.011	-.019	-.008	.011
		1.00	7.536	8.062	.251	.286	.087	.087	.083	.084	-1.784	2.237	-.084	.010	.012	-.022	-.010	.019
		1.25	7.682	7.667	.345	.199	.096	.104	.086	.095	-1.975	-5.533	-.042	-.031	.017	-.024	-.012	.019
		1.75	8.303	8.787	.563	.189	.124	.388	.099	.359	-2.401	-8.567	.156	-.183	.018	-.259	-.009	.264
200	1	.75	4.140	9.250	.596	1.496	.062	.067	.063	.067	1.158	7.132	.165	1.426	.004	-.020	-.006	.019
		1.00	4.668	3.280	.535	.280	.067	.067	.064	.063	1.303	.738	.161	.093	-.002	-.007	-.003	.008
		1.25	4.013	2.055	.363	.178	.086	.262	.071	.216	1.004	.281	.079	-.165	-.000	-.128	-.001	.140
		1.75	1.418	1.442	.206	.197	.130	.499	.117	.551	.273	.268	-.003	-.195	.021	-.428	-.018	.510
	5	.75	7.995	15.475	.163	.172	.057	.052	.054	.048	3.383	14.188	-.113	-.077	.006	-.012	-.005	.005
		1.00	7.155	6.353	.181	.126	.056	.054	.054	.053	2.934	2.616	-.109	-.039	.006	-.008	-.006	.007
		1.25	6.291	3.664	.199	.137	.058	.067	.053	.059	2.038	-2.969	-.091	-.116	.006	-.010	-.005	.011
		1.75	5.342	3.942	.279	.197	.103	.489	.078	.495	.057	-3.773	.024	-.196	.018	-.411	-.008	.442
	10	.75	7.840	12.421	.159	.170	.054	.049	.052	.046	-.286	11.149	-.099	-.066	.000	-.006	-.001	-.000
		1.00	7.841	7.435	.178	.174	.056	.054	.054	.053	-.630	2.610	-.097	-.018	.003	-.008	-.004	.008
		1.25	7.899	7.763	.241	.120	.057	.057	.055	.053	-.540	-6.809	-.061	-.038	.005	-.007	-.005	.004
		1.75	7.757	8.855	.372	.193	.083	.437	.066	.411	-1.657	-8.798	.089	-.193	.010	-.347	-.004	.344
300	1	.75	3.403	8.272	.583	1.513	.048	.050	.048	.050	.962	6.563	.165	1.466	.002	-.013	-.004	.014
		1.00	3.611	1.965	.433	.189	.051	.051	.050	.048	.881	.357	.119	.060	-.003	-.004	-.001	.005
		1.25	2.842	1.313	.266	.177	.065	.239	.054	.201	.563	.157	.039	-.171	-.000	-.116	-.001	.133
		1.75	1.054	.238	.144	.198	.098	.493	.086	.554	.220	.109	-.007	-.197	.013	-.435	-.011	.519
	5	.75	8.363	15.872	.149	.154	.043	.040	.042	.037	3.571	14.879	-.110	-.094	.002	-.009	-.003	.002
		1.00	6.719	4.663	.152	.105	.042	.043	.042	.042	2.457	1.699	-.110	-.027	.003	-.006	-.003	.004
		1.25	5.507	3.430	.162	.138	.046	.052	.042	.047	1.504	-3.347	-.094	-.127	.004	-.007	-.003	.007
		1.75	4.870	3.898	.221	.198	.069	.499	.056	.507	.237	-3.886	.011	-.198	.010	-.436	-.004	.464
	10	.75	8.326	12.790	.150	.154	.042	.039	.041	.036	.450	11.783	-.104	-.083	-.002	-.005	.000	-.003
		1.00	8.176	6.455	.159	.149	.042	.043	.042	.041	-.295	1.942	-.094	-.014	.000	-.005	-.001	.004
		1.25	7.899	7.817	.203	.097	.045	.045	.043	.042	-.640	-7.579	-.064	-.044	.003	-.003	-.001	-.000
		1.75	7.584	8.916	.310	.196	.060	.474	.051	.444	-.954	-8.913	.071	-.196	.005	-.402	-.001	.390

Table A.6: Root mean squared errors (RMSE) and bias for the other parameter estimates of the fractional UC model with correlated innovations in subsection 6.2. The different columns indicate the parameter estimates via the CSS estimator (subscript CSS) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

n	ν_0	d_0	RMSE										bias									
			$\hat{\sigma}^2_{\eta_{QML}}$	$\hat{\sigma}^{2^{I(1)}}_{\eta_{QML}}$	$\hat{\sigma}^2_{\epsilon_{QML}}$	$\hat{\sigma}^{2^{I(1)}}_{\epsilon_{QML}}$	$\hat{\rho}_{QML}$	$\hat{\rho}^{I(1)}_{QML}$	$\hat{b}_{1_{QML}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{QML}}$	$\hat{b}^{I(1)}_{2_{QML}}$	$\hat{\sigma}^2_{\eta_{QML}}$	$\hat{\sigma}^{2^{I(1)}}_{\eta_{QML}}$	$\hat{\sigma}^2_{\epsilon_{QML}}$	$\hat{\sigma}^{2^{I(1)}}_{\epsilon_{QML}}$	$\hat{\rho}_{QML}$	$\hat{\rho}^{I(1)}_{QML}$	$\hat{b}_{1_{QML}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{QML}}$	$\hat{b}^{I(1)}_{2_{QML}}$
100	1	.75	.570	.821	.626	.477	.390	1.359	.101	.126	.095	.118	.102	-.746	.081	-.186	.055	1.147	.013	-.040	-.009	.027
		1.00	.868	.725	.919	.781	.444	.501	.122	.125	.106	.113	.162	-.093	.151	.074	.091	.199	.011	-.025	-.008	.023
		1.25	1.864	4.690	1.789	4.192	.434	.208	.158	.353	.123	.291	.327	2.037	.259	2.014	.106	-.097	.015	-.153	-.009	.169
		1.75	3.945	2509.261	3.601	2472.683	.215	.179	.189	.535	.170	.578	1.327	522.673	1.052	510.969	-.024	-.137	.043	-.274	-.036	.411
	5	.75	1.218	.823	1.438	1.355	.173	.366	.080	.079	.076	.068	.508	-.679	.017	-.233	-.089	-.051	.008	-.021	-.004	.006
		1.00	1.463	.934	1.616	1.652	.246	.258	.081	.087	.077	.081	.561	.025	.073	.131	-.065	-.017	.006	-.021	-.003	.015
		1.25	2.051	2.931	2.063	3.405	.299	.178	.096	.111	.084	.112	.767	1.983	.309	1.444	-.034	-.064	.012	-.026	-.007	.038
		1.75	3.603	1176.013	4.074	1112.683	.330	.184	.130	.500	.109	.488	1.679	193.079	1.503	183.935	.024	-.156	.026	-.278	-.023	.358
	10	.75	1.609	1.077	2.553	2.680	.191	.342	.082	.076	.078	.067	.434	-.499	-.018	-.607	-.088	-.089	.016	-.013	-.010	-.000
		1.00	1.937	1.139	2.855	3.061	.300	.351	.081	.083	.078	.077	.553	.132	.014	.142	-.052	.002	.011	-.019	-.007	.013
		1.25	2.640	4.319	3.644	4.989	.373	.238	.088	.101	.083	.103	.905	3.057	.560	1.806	.005	-.002	.012	-.022	-.010	.027
		1.75	4.494	527.802	5.505	543.198	.372	.199	.115	.447	.099	.423	2.088	131.662	2.286	134.202	.036	-.154	.035	-.259	-.030	.316
200	1	.75	.418	.845	.357	.365	.382	1.472	.061	.072	.061	.072	.052	-.836	.033	-.248	.060	1.385	.002	-.017	-.002	.015
		1.00	.476	.444	.392	.344	.394	.307	.070	.067	.066	.065	.009	-.133	.025	-.017	.098	.126	-.003	-.005	.001	.008
		1.25	.701	4.268	.571	3.502	.322	.153	.091	.325	.073	.268	.093	2.003	.065	1.934	.077	-.133	.001	-.147	-.001	.168
		1.75	2.058	2232.846	1.828	2251.348	.143	.169	.129	.514	.119	.577	.609	618.162	.429	618.671	-.021	-.131	.029	-.257	-.024	.409
	5	.75	.983	.792	.924	.951	.148	.201	.050	.051	.048	.045	.356	-.764	.081	-.354	-.090	-.113	-.002	-.004	.001	-.006
		1.00	1.132	.644	1.034	1.000	.151	.133	.051	.054	.048	.050	.383	-.086	.045	.018	-.071	-.040	-.001	-.007	.002	.006
		1.25	1.669	3.336	1.241	2.505	.191	.112	.057	.067	.052	.065	.631	2.181	.150	1.098	-.059	-.084	.003	-.012	-.002	.022
		1.75	2.978	545.783	2.557	982.156	.238	.173	.096	.540	.078	.523	1.170	197.162	.667	210.049	.009	-.154	.019	-.308	-.012	.392
	10	.75	1.529	.755	1.789	1.944	.164	.222	.051	.048	.049	.045	.401	-.656	-.023	-.930	-.100	-.124	.001	.004	.000	-.013
		1.00	1.728	.725	2.006	1.985	.189	.200	.050	.054	.049	.050	.491	-.015	.061	-.046	-.064	-.035	-.000	-.005	.001	.002
		1.25	2.455	4.390	2.453	2.800	.252	.114	.055	.056	.052	.055	.807	3.575	.266	1.225	-.025	-.013	.002	-.007	-.002	.009
		1.75	3.102	2821.943	3.347	2812.517	.287	.206	.071	.507	.062	.467	1.204	235.101	.761	226.891	.043	-.153	.014	-.302	-.010	.358
300	1	.75	.354	.852	.269	.341	.324	1.497	.046	.053	.045	.053	.042	-.847	.017	-.272	.038	1.444	-.000	-.010	-.000	.010
		1.00	.412	.370	.308	.264	.318	.209	.054	.051	.052	.049	-.032	-.113	-.004	-.019	.086	.088	-.006	-.003	.003	.006
		1.25	.597	3.616	.455	3.126	.249	.155	.065	.290	.055	.241	.068	1.719	.043	1.766	.049	-.146	-.000	-.130	-.000	.156
		1.75	1.339	9217.928	1.133	9346.245	.118	.183	.099	.512	.089	.584	.387	1066.647	.266	1060.340	-.016	-.132	.021	-.271	-.017	.425
	5	.75	.943	.798	.783	.805	.129	.171	.039	.040	.038	.036	.316	-.782	.137	-.377	-.084	-.119	-.003	-.001	.002	-.008
		1.00	1.121	.485	.905	.810	.123	.110	.040	.043	.038	.041	.401	-.088	.075	.001	-.065	-.028	-.001	-.005	.003	.004
		1.25	1.666	2.973	1.067	1.931	.143	.114	.044	.060	.041	.055	.705	2.399	.174	1.066	-.070	-.096	.003	-.011	-.001	.018
		1.75	2.137	635.502	1.960	685.964	.204	.187	.064	.542	.055	.527	.762	223.000	.369	220.305	.004	-.137	.010	-.291	-.006	.380
	10	.75	1.684	.744	1.444	1.709	.153	.187	.040	.040	.039	.038	.562	-.676	.041	-.959	-.102	-.134	-.003	.008	.002	-.016
		1.00	1.788	.568	1.682	1.623	.160	.155	.040	.042	.038	.039	.613	-.012	.148	-.067	-.057	-.026	-.001	-.003	.002	.001
		1.25	2.724	4.496	2.018	2.376	.181	.086	.044	.045	.041	.044	1.079	3.932	.281	1.144	-.039	-.018	.002	-.004	-.001	.005
		1.75	2.438	976.651	2.649	11833.501	.249	.194	.056	.531	.048	.486	.875	217.345	.435	574.983	.038	-.151	.010	-.324	-.007	.376

Table A.7: Root mean squared errors (RMSE) and bias for the other parameter estimates of the fractional UC model with correlated innovations in subsection 6.2. The different columns indicate the parameter estimates via the QML estimator (subscript QML) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

n	ν_0	d_0	Trend R^2_{CSS}	R^2_{QML}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$	Cycle R^2_{CSS}	R^2_{QML}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$
100	1	.75	.534	.573	.275	.291	.851	.868	.752	.751
		1.00	.750	.782	.774	.774	.794	.823	.828	.825
		1.25	.904	.911	.830	.799	.763	.777	.684	.608
		1.75	.987	.986	.861	.802	.711	.607	.300	.124
	5	.75	.426	.423	.422	.385	.949	.951	.948	.874
		1.00	.664	.681	.720	.654	.925	.930	.934	.867
		1.25	.861	.864	.848	.827	.900	.903	.885	.871
		1.75	.980	.975	.883	.801	.827	.676	.484	.282
	10	.75	.382	.385	.373	.350	.963	.970	.959	.903
		1.00	.575	.615	.649	.578	.939	.950	.948	.873
		1.25	.797	.823	.826	.786	.912	.921	.914	.885
		1.75	.968	.971	.892	.793	.841	.740	.576	.421
200	1	.75	.657	.703	.342	.348	.869	.890	.733	.735
		1.00	.883	.897	.903	.900	.840	.861	.875	.872
		1.25	.971	.974	.914	.887	.830	.835	.702	.622
		1.75	.998	.998	.910	.890	.791	.718	.234	.077
	5	.75	.541	.549	.574	.468	.956	.958	.964	.866
		1.00	.816	.829	.846	.817	.942	.946	.949	.926
		1.25	.946	.949	.941	.938	.926	.931	.910	.913
		1.75	.996	.997	.938	.896	.883	.788	.401	.154
	10	.75	.475	.488	.498	.405	.968	.973	.973	.871
		1.00	.752	.780	.810	.737	.952	.961	.965	.907
		1.25	.918	.932	.930	.922	.936	.947	.931	.933
		1.75	.995	.997	.955	.886	.898	.813	.506	.251
300	1	.75	.727	.772	.406	.412	.878	.900	.722	.725
		1.00	.933	.941	.943	.941	.861	.878	.889	.886
		1.25	.987	.988	.950	.932	.853	.858	.712	.648
		1.75	1.000	.999	.931	.890	.819	.760	.208	.066
	5	.75	.610	.620	.640	.552	.958	.961	.966	.900
		1.00	.881	.891	.900	.892	.947	.951	.955	.950
		1.25	.974	.976	.971	.970	.935	.938	.916	.920
		1.75	.999	.999	.963	.906	.904	.829	.369	.111
	10	.75	.539	.548	.554	.463	.969	.973	.974	.891
		1.00	.830	.849	.871	.849	.956	.964	.968	.948
		1.25	.958	.967	.964	.962	.943	.953	.935	.941
		1.75	.999	.999	.979	.923	.918	.846	.461	.192

Table A.8: Coefficient of determination from regressing true trend and cycle x_t and c_t on their respective estimates from the Kalman smoother for the correlated UC models.

	$I(d)$		$I(1)$		$I(2)$	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
d	1.753	0.061				
σ_η^2	1.351E-08	1.527E-08	1.032E-04	4.499E-05	6.179E-10	7.081E-10
$\sigma_{\eta\epsilon}$	-2.202E-06	2.620E-06	-5.465E-04	1.402E-04	-1.094E-06	6.279E-07
σ_ϵ^2	1.981E-03	6.171E-05	2.901E-03	2.313E-04	1.955E-03	6.103E-05
b_1	-1.024	0.022	-0.997	0.020	-1.033	0.019
b_2	0.101	0.031	0.094	0.024	0.137	0.014
b_3	0.064	0.031	0.027	0.007	0.018	0.000
b_4	-0.063	0.022	-0.033	0.012	-0.040	0.006
ν	1.466E+05		28.115		3.163E+06	
ν_2	-162.993		-5.296		-1.771E+03	
ρ	-0.426		-0.999		-0.996	
$\log L(\psi)$	5436.6		5428.1		5430.4	
$Q(y,\psi)$	4.1315		4.9048		4.6589	
AIC	-10855.1		-10840.2		-10844.8	
BIC	-10804.4		-10795.1		-10799.6	

Table A.9: Estimation results for monthly global temperature anomalies from the fractional UC model, the $I(1)$ UC model, and the $I(2)$ UC model via the QML estimator. All three models allow for correlated innovations. Optimization is carried out over $\psi = (d, \sigma_\eta^2, \sigma_{\eta\epsilon}, \sigma_\epsilon^2, b_1, ..., b_4)'$, and estimates for ν, ν_2, ρ are calculated based on the estimates of ψ. $\log L(\psi)$ denotes the log likelihood, $Q(y,\psi)$ denotes the conditional sum-of-squares, AIC is the Akaike Information Criterion, and BIC is the Bayesian Information Criterion. Standard errors are obtained from the numerical Hessian matrix.

B Proof of theorem 4.1

Proof of theorem 4.1. Theorem 4.1 holds if the objective function (16) satisfies a uniform weak law of large numbers (UWLLN), i.e. there exists a function $g_t(y_{t:1}) \geq 0$ such that for all $\theta_1, \theta_2 \in \Theta$, it holds that $|v_t^2(\theta_1) - v_t^2(\theta_2)| \leq g_t(y_{t:1})||\theta_1 - \theta_2||$, and both, $v_t(\theta)$ and $g_t(y_{t:1})$ satisfy a WLLN (Wooldridge; 1994, thm. 4.2). Since $v_t^2(\theta)$ is continuously differentiable, a natural choice for $g_t(y_{t:1})$ is the supremum of the absolute gradient, as follows from the mean value expansion of $v_t^2(\theta)$ about θ, see Newey (1991, cor. 2.2) and Wooldridge (1994, eqn. 4.4).

However, as can be seen from (15), uniform convergence of the objective function fails around the point $d = d_0 - 1/2$: Since y_t is $I(d_0)$, the d-th differences $\Delta_+^d y_{t+1} = \xi_{t+1}(d)$ as well as $S_d y_{t:1} = \xi_{t:1}(d)$ are $I(d_0 - d)$, and thus asymptotically stationary whenever $d > d_0 - 1/2$, otherwise non-stationary. Subsequently, I will show that the pointwise probability limit of $Q(y, \theta)$ is given by

$$\operatorname{plim}_{n\to\infty} Q(y,\theta) = \operatorname{plim}_{n\to\infty} \tilde{Q}(y,\theta) = \begin{cases} \mathrm{E}(\tilde{v}_t^2(\theta)) & \text{for } d - d_0 > -1/2, \\ \infty & \text{else,} \end{cases} \tag{B.1}$$

where $\tilde{v}_t(\theta)$ denotes the untruncated forecast error

$$\tilde{v}_t(\theta) = \tilde{\xi}_t(d) + \sum_{j=1}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d) = \sum_{j=0}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d), \tag{B.2}$$

generated by the untruncated fractional differencing polynomial Δ^d and the untruncated polynomial $b(L,\varphi) = \sum_{j=0}^{\infty} b_j(\varphi)L^j$. $\tilde{\xi}_t(d) = \Delta^{d-d_0}\eta_t + \Delta^d c_t$ is the untruncated residual, while the $\tau_j(\theta)$ stem from the ∞-vector $(\tau_1(\theta), \tau_2(\theta), \cdots) = \nu(b_1(\varphi) - \pi_1(d), b_2(\varphi) - \pi_2(d), \cdots)(B'_{\varphi,\infty}B_{\varphi,\infty} + \nu S'_{d,\infty}S_{d,\infty})^{-1}S'_{d,\infty}$, and $\tau_0(\theta) = 1$ as before. Note that the dependence of the $\tau_j(\theta)$ on t is resolved in (B.2) by letting the dimension of the t-dimensional coefficient vector go to infinity. Hence, while the truncated forecast errors in (15) are non-ergodic, the untruncated errors (B.2) are ergodic within the stationary region of the parameter space where $d - d_0 > -1/2$, as will become clear.

To deal with non-uniform convergence in (B.1), I adapt the proof strategy of Nielsen (2015) for CSS estimation of ARFIMA models: I partition the parameter space for d into three compact subsets $D_1 = D_1(\kappa_1) = D \cap \{d : d - d_0 \leq -1/2 - \kappa_1\}$, $D_2 = D_2(\kappa_2, \kappa_3) = D \cap \{d : -1/2 - \kappa_2 \leq d - d_0 \leq -1/2 + \kappa_3\}$, and $D_3 = D_3(\kappa_3) = D \cap \{d : -1/2 + \kappa_3 \leq d - d_0\}$, for some constants $0 < \kappa_1 < \kappa_2 < \kappa_3 < 1/2$ to be determined later. Note that $\cup_{i=1}^3 D_i = D$. Within D_1 and D_3 convergence is uniform, while within the overlapping D_2, which covers both stationary and non-stationary forecast errors, convergence is non-uniform. Denote the partitioned parameter spaces for θ as $\Theta_j = D_j \times \Sigma_\nu \times \Phi$, $j = 1, 2, 3$. Non-uniform convergence of (B.1) is then asymptotically ruled out by showing that for a given constant $K > 0$ there always exists a fixed $\bar{\kappa} > 0$ such that

$$\Pr\left(\inf_{d\in D\backslash D_3(\bar{\kappa}), \nu\in\Sigma_\nu, \varphi\in\Phi} Q(y,\theta) > K\right) \to 1 \quad \text{as } n \to \infty, \tag{B.3}$$

which implies $\Pr(\hat{\theta} \in D_3(\bar{\kappa}) \times \Sigma_\nu \times \Phi) \to 1$, i.e. the parameter space asymptotically reduces to the stationary region $\Theta_3(\bar{\kappa}) = D_3(\bar{\kappa}) \times \Sigma_\nu \times \Phi$. The second part of the proof shows that within $\Theta(\kappa_3)$, a UWLLN applies to the objective function, i.e. for any fixed $\kappa_3 \in (0, 1/2)$

$$\sup_{\theta \in D_3(\kappa_3) \times \Sigma_\nu \times \Phi} \left| Q(y, \theta) - \mathrm{E}(\tilde{v}_{t+1}^2(\theta)) \right| \xrightarrow{p} 0, \qquad \text{as } n \to \infty, \tag{B.4}$$

which holds if both the objective function and the supremum of its absolute gradient satisfy a WLLN (Wooldridge; 1994, thm. 4.2). While the results in (B.3) and (B.4) are well established for the CSS estimator in the ARFIMA literature, see Hualde and Robinson (2011) and Nielsen (2015), showing them to carry over to the fractional UC model requires some additional effort. Even within $\theta \in \Theta_3(\kappa_3)$, the forecast errors in (14) are not ergodic for two reasons: First, since the lag polynomial generated by the truncated fractional differencing polynomial Δ_+^d includes more lags as t increases, $\xi_t(d) = \Delta_+^{d-d_0}\eta_t + \Delta_+^d c_t$ are not ergodic. Second, the $\tau_j(\theta, t)$ in (15) depend on t. Consequently, also within $\Theta_3(\kappa_3)$ a WLLN for stationary and ergodic processes does not immediately apply. I tackle these problems by showing the expected difference between (15) and (B.2) to be

$$\mathrm{E}\left[(\tilde{v}_{t+1}(\theta) - v_{t+1}(\theta))^2\right] \to 0, \qquad \text{as } t \to \infty, \tag{B.5}$$

for all $\theta \in \Theta_3(\kappa_3)$ (pointwise). As within $\Theta_3(\kappa_3)$, $\tilde{v}_{t+1}(\theta)$ is stationary and ergodic, it follows by (B.5) that the WLLN for stationary and ergodic processes carries over from $\tilde{v}_{t+1}(\theta)$ to $v_{t+1}(\theta)$

$$Q(y, \theta) = \tilde{Q}(y, \theta) + o_p(1) \xrightarrow{p} \mathrm{E}(\tilde{v}_t^2(\theta)), \qquad \text{as } n \to \infty. \tag{B.6}$$

(B.6) can be generalized to uniform convergence by showing that a WLLN also holds for the supremum of the absolute gradient, which yields (B.4). From (B.3) and (B.4), theorem 4.1 follows. In the proofs, let $z_{(j)}$ denote the j-th entry of some vector z, and let $Z_{(i,j)}$ denote the (i, j)-th entry (i.e. the entry in row i and column j) for some matrix Z.

Convergence on $\Theta_3(\kappa_3)$ and proof of (B.4) **and** (B.6) I begin with the case $\theta \in \Theta_3(\kappa_3) = D_3(\kappa_3) \times \Sigma_\nu \times \Phi$ where $v_t(\theta)$ is asymptotically stationary. To prove (B.5), I first show that

$$\begin{aligned}
\tilde{v}_{t+1}(\theta) - v_{t+1}(\theta) = &\sum_{j=0}^{t} \tau_j(\theta, t) \left(\tilde{\xi}_{t+1-j}(d) - \xi_{t+1-j}(d)\right) \\
&+ \sum_{j=t+1}^{\infty} \tau_j(\theta)\tilde{\xi}_{t+1-j}(d) + \sum_{j=0}^{t} (\tau_j(\theta) - \tau_j(\theta, t))\, \tilde{\xi}_{t+1-j}(d) \\
= &\sum_{j=0}^{\infty} \phi_{\eta,j}(\theta, t)\eta_{t+1-j} + \sum_{j=0}^{\infty} \phi_{\epsilon,j}(\theta, t)\epsilon_{t+1-j},
\end{aligned} \tag{B.7}$$

where $\phi_{\eta,j}(\theta, t)$ is $O((1+\log(t+1))^2(t+1)^{\max(-d+d_0,-\zeta)-1})$ for $j \le t$, and $O((1+\log j)^3 j^{\max(-d+d_0,-\zeta)-1})$ for $j > t$, whereas $\phi_{\epsilon,j}(\theta, t)$ is $O((1 + \log(t + 1))^2(t + 1)^{\max(-d,-\zeta)-1})$ for $j \le t$, and $O((1 + \log j)^4 j^{\max(-d,-\zeta)-1})$ for $j > t$. This can be verified by considering the three different terms in (B.7) separately. For the first term, plugging in $\xi_t(d) = \Delta_+^{d-d_0}\eta_t + \Delta_+^d c_t$, $\tilde{\xi}_t(d) = \Delta^{d-d_0}\eta_t + \Delta^d c_t$

yields

$$\sum_{j=0}^{t}\tau_j(\theta,t)\left(\tilde{\xi}_{t+1-j}(d)-\xi_{t+1-j}(d)\right)=\sum_{j=t+1}^{\infty}\phi_{1,\eta,j}(\theta,t)\eta_{t+1-j}+\sum_{j=t+1}^{\infty}\phi_{1,\epsilon,j}(\theta,t)\epsilon_{t+1-j},\tag{B.8}$$

where $\phi_{1,\eta,j}(\theta,t)=\sum_{k=0}^{t}\tau_k(\theta,t)\pi_{j-k}(d-d_0)$, and $\phi_{1,\epsilon,j}(\theta,t)=\sum_{k=0}^{t}\tau_k(\theta,t)\sum_{l=0}^{j-t-1}a_l(\varphi_0)\pi_{j-k-l}(d)$. Using Johansen and Nielsen (2010, lemma B.4), who show $\sum_{k=1}^{j-1}k^{\max(-d,-\zeta)-1}(j-k)^{-d+d_0-1}\leq K(1+\log j)j^{\max(-d+d_0,-\zeta)-1}$ for some finite constant $K>0$, together with assumption 3, (D.1), lemma D.2, and $j>t$, the coefficients in (B.8) are $\phi_{1,\eta,t}=O((1+\log j)^2j^{\max(-d+d_0,-\zeta)-1})$, and $\phi_{1,\epsilon,t}=O((1+\log j)^3j^{\max(-d,-\zeta)-1})$.

Next, consider the second term in (B.7)

$$\sum_{j=t+1}^{\infty}\tau_j(\theta)\tilde{\xi}_{t+1-j}(d)=\sum_{j=t+1}^{\infty}\eta_{t+1-j}\phi_{2,\eta,j}(\theta,t)+\sum_{j=t+1}^{\infty}\epsilon_{t+1-j}\phi_{2,\epsilon,j}(\theta,t),\tag{B.9}$$

with $\phi_{2,\epsilon,j}(\theta,t)=\sum_{k=0}^{j-t-1}\tau_{t+1+k}(\theta)\sum_{l=0}^{j-t-1-k}a_l(\varphi_0)\pi_{j-t-1-k-l}(d)=O((1+\log j)^3j^{\max(-d,-\zeta)-1})$, and $\phi_{2,\eta,j}(\theta,t)=\sum_{k=0}^{j-t-1}\pi_k(d-d_0)\tau_{j-k}(\theta)=O((1+\log j)^2j^{\max(-d+d_0,-\zeta)-1})$ by assumption 3, lemma D.1 and lemma D.2.

For the third term in (B.7), by lemma D.3

$$\begin{aligned}\sum_{j=0}^{t}\left(\tau_j(\theta)-\tau_j(\theta,t)\right)\tilde{\xi}_{t+1-j}(d)=&-\sum_{j=0}^{\infty}\eta_{t+1-j}\sum_{k=0}^{\min(j,t)}\pi_{j-k}(d-d_0)\sum_{m=t+1}^{\infty}r_{\tau,k,m}(\theta)\\&-\sum_{j=0}^{\infty}\epsilon_{t+1-j}\sum_{k=0}^{\min(j,t)}\left(\sum_{m=t+1}^{\infty}r_{\tau,k,m}(\theta)\right)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d)\\=&\sum_{j=0}^{\infty}\phi_{3,\eta,j}(\theta,t)\eta_{t+1-j}+\sum_{j=0}^{\infty}\phi_{3,\epsilon,j}(\theta,t)\epsilon_{t+1-j}.\end{aligned}\tag{B.10}$$

By lemma D.3, $\sum_{m=t+1}^{\infty}r_{\tau,k,m}(\theta)=O((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1})$, while $\pi_j(d-d_0)=O(j^{-d+d_0-1})$ and $\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d)=O((1+\log(j-k))(j-k)^{\max(-d,-\zeta)-1})$, see lemma D.1 together with Johansen and Nielsen (2010, lemma B.4). Thus, for $j\leq t$, it holds that $\phi_{3,\eta,j}(\theta,t)=-\sum_{k=0}^{\min(j,t)}\left(\sum_{m=t+1}^{\infty}r_{\tau,k,m}(\theta)\right)\pi_{j-k}(d-d_0)$ is $O\left((1+\log(t+1))^2(t+1)^{\max(-d+d_0,-\zeta)-1}\right)$, whereas for $j>t$ it is $O\left((1+\log j)^3j^{\max(-d+d_0,-\zeta)-1}\right)$. Similarly, for $j\leq t$, the coefficient $\phi_{3,\epsilon,j}(\theta,t)=\sum_{k=0}^{\min(j,t)}\left(\sum_{m=t+1}^{\infty}r_{\tau,k,m}(\theta)\right)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d)$ is $O\left((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\right)$, and for $j>t$ it is $O\left((1+\log j)^4j^{\max(-d,-\zeta)-1}\right)$. Together, (B.8), (B.9), (B.10) and the rates established below prove (B.7).

(B.5) can be proven by noting that $\tilde{v}_{t+1}(\theta)$ is stationary and ergodic, so that a WLLN for stationary and ergodic processes applies. Thus, it is sufficient to consider

$$\mathrm{E}[(\tilde{v}_{t+1}(\theta)-v_{t+1}(\theta))^2]=\sum_{j=1}^{\infty}\left[\phi_{\eta,j}^2(\theta,t)\,\mathrm{E}(\eta_{t+1-j}^2)+\phi_{\epsilon,j}^2(\theta,t)\,\mathrm{E}(\epsilon_{t+1-j}^2)\right]$$

$$
\begin{aligned}
&= \sum_{j=1}^{t} O\left((1+\log(t+1))^4 (t+1)^{2\max(-d+d_0,-\zeta)-2}\right) \\
&\quad + \sum_{j=t+1}^{\infty} O\left((1+\log(t+1))^8 (t+1)^{2\max(-d+d_0,-\zeta)-2}\right) = o(1),
\end{aligned}
$$

where the first equality follows by assumption 1, while the second follows from the convergence rates of $\phi_{\eta,j}(\theta,t)$, $\phi_{\epsilon,j}(\theta,t)$ as derived above, and the third equality follows from $\zeta > 0$ and $d - d_0 + 1/2 > \kappa_3 > 0$ for all $\theta \in \Theta_3(\kappa_3)$. (B.5) follows directly. From the law of large numbers for stationary and ergodic processes, (B.6) follows immediately.

(B.6) can be generalized to uniform convergence in probability by showing the supremum of the absolute gradient to be bounded in probability for all $\theta \in \Theta(\kappa_3)$ and any κ_3, see Newey (1991, cor. 2.2) and Wooldridge (1994, th. 4.2). Then (B.4) holds, so that the objective function satisfies a UWLLN within the stationary region of the parameter space $\Theta_3(\kappa_3)$. The gradient of the objective function is given by

$$
\begin{aligned}
\frac{\partial Q(y,\theta)}{\partial \theta_{(l)}} &= \frac{2}{n}\sum_{t=1}^{n} v_t(\theta)\frac{\partial v_t(\theta)}{\partial \theta_{(l)}}, \\
\frac{\partial v_t(\theta)}{\partial \theta_{(l)}} &= \sum_{j=1}^{t-1} \frac{\partial \tau_j(\theta,t)}{\partial \theta_{(l)}} \xi_{t-j}(d) + \sum_{j=0}^{t-1} \tau_j(\theta,t)\frac{\partial \xi_{t-j}(d)}{\partial \theta_{(l)}},
\end{aligned}
\tag{B.11}
$$

where $\theta_{(l)}$ denotes the l-th parameter in θ. Now, denote $\tilde{\tau}_i(L,\theta) = \sum_{j=0}^{\infty} \tilde{\tau}_{i,j}(\theta)L^j$ as any polynomial satisfying $\sum_{j=0}^{\infty} |\tilde{\tau}_{i,j}(\theta)| < \infty$, $i = 1,2$, uniformly in $\theta \in \Theta$. Then, for $z_{1,t}(\theta) = \eta_t$, $z_{2,t}(\theta) = \epsilon_t$, and for the set $\tilde{\Theta}\{(d_1,d_2,\nu,\varphi) \in D \times D \times \Sigma_\nu \times \Phi : \min(d_1+1, d_2+1, d_1+d_2+1) \geq a\}$, it holds that

$$
\begin{aligned}
&\sup_{(d_1,d_2,\nu,\varphi)\in\tilde{\Theta}} \left| \frac{1}{n}\sum_{t=1}^{n} \left[\frac{\partial^k \Delta_+^{d_1}}{\partial d_1^k} \sum_{m=0}^{\infty} \tilde{\tau}_{i,m}(\theta) z_{i,t-m}(\theta)\right] \left[\frac{\partial^l \Delta_+^{d_2}}{\partial d_2^l} \sum_{m=0}^{\infty} \tilde{\tau}_{j,m}(\theta) z_{j,t-m}(\theta)\right] \right| \\
&\qquad = \begin{cases} O_p(1) & \text{for } a > 0, \\ O_p((\log n)^{1+k+l} n^{-a}) & \text{for } a \leq 0, \end{cases}
\end{aligned}
\tag{B.12}
$$

$i,j = 1,2$, $k,l = 1,2,...$, as shown by Nielsen (2015, lemma B.3). Now, note that by lemmas D.2 and D.4 both the coefficients $\tau_j(\theta,t)$ and their partial derivatives satisfy the absolute summability condition, i.e. $\sum_{j=0}^{t-1} |\tau_j(\theta,t)| < \infty$ and $\sum_{j=0}^{t-1} |\partial\tau_j(\theta,t)/\partial\theta_{(l)}| < \infty$ for all $\theta_{(l)}$ and uniformly in $\theta \in \Theta$. In addition, by assumption 3, the absolute summability condition also holds for the polynomials $\sum_{j=0}^{t-1} \tau_j(\theta,t)L^j a(L,\varphi_0)$ and $\sum_{j=0}^{t-1} \partial\tau_j(\theta,t)/(\partial\theta_{(l)})L^j a(L,\varphi_0)$. Furthermore, note that the (truncated) fractional difference operator and the (truncated) polynomials $\sum_{j=1}^{t-1} \tau_j(\theta,t)L^j$ as well as their partial derivatives can be interchanged, e.g. $\Delta_+^d \sum_{j=0}^{t-1} \tau_j(\theta,t)\eta_{t-j} = \sum_{j=0}^{t-1} \tau_j(\theta,t)\Delta_+^d \eta_{t-j}$, as the sum is bounded at $t-1$. Finally, for $\theta \in \Theta_3(\kappa_3)$, it holds that $d - d_0 > -1/2$, so that within $v_t(\theta)$ the term $\Delta_+^{d-d_0}\eta_t$ is integrated of order smaller 1/2, and the same holds for the partial derivative $\partial\xi_t(d)/\partial d = (\partial\Delta_+^{d-d_0}/\partial d)\eta_t + (\partial\Delta_+^d/\partial d)c_t$. Thus, all terms in (B.11) satisfy the conditions for (B.12) with $a > 0$. By (B.12), it follows that $\sup_{\theta\in\Theta_3(\kappa_3)} \left|\frac{\partial Q(y,\theta)}{\partial\theta_{(l)}}\right| = O_p(1)$ for all entries in θ. Hence, (B.6) holds uniformly in $\theta \in \Theta_3(\kappa_3)$. As this holds for any κ_3, this proves (B.4).

Convergence on $\Theta_2(\kappa_1,\kappa_2)$ Next, consider the case $\theta\in\Theta_2(\kappa_1,\kappa_2)=D_2(\kappa_1,\kappa_2)\times\Sigma_\nu\times\Phi$. Then for the objective function in (16), together with (15), it holds that

$$\begin{aligned}Q(y,\theta)=&\frac{1}{n}\sum_{t=1}^{n}\left[\sum_{j=0}^{t-1}\tau_j(\theta,t)\xi_{t-j}(d)\right]^2\geq\frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\Big)^2\\&+\frac{2}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\Big)\Big(\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\Big),\end{aligned}\tag{B.13}$$

where the fractional difference operator and the polynomial $\sum_{j=0}^{t-1}\tau_j(\theta,t)L^j$ can be interchanged as the latter is truncated at $t-1$.

For the second term in (B.13), by lemma D.2 $\sum_{j=0}^{t-1}|\tau_j(\theta,t)|<\infty$, and by assumption 3 and lemma D.2 $\sum_{j=0}^{\infty}\sum_{k=0}^{\min(j,t-1)}|\tau_j(\theta,t)a_{k-j}(\varphi_0)|<\infty$. Furthermore, as $d>0$, $d-d_0\geq-1/2-\kappa_2>-1$, it holds that $\min(1+d-d_0,1+d,1+2d-d_0)=1+d-d_0>0$, so that by (B.12)

$$\sup_{\theta\in\Theta_2(\kappa_2,\kappa_3)}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right]\left[\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\right]\right|=O_p(1).\tag{B.14}$$

Next, consider the first term in (B.13), for which one has by lemma D.3

$$\begin{aligned}\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}&=\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta)\eta_{t-j}+\Delta_+^{d-d_0}\sum_{j=1}^{t-1}\left(\sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta)\right)\eta_{t-j}\\&=\Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j}+r_{\eta,t}(\theta),\end{aligned}\tag{B.15}$$

where

$$r_{\eta,t}(\theta)=-\Delta_+^{d-d_0}\sum_{j=t}^{\infty}\tau_j(\theta)\eta_{t-j}+\Delta_+^{d-d_0}\sum_{j=1}^{t-1}\eta_{t-j}\sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta)=\Delta_+^{d-d_0}\sum_{j=1}^{\infty}\alpha_j\eta_{t-j},\tag{B.16}$$

and $\alpha_j=\sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta)$ for $j<t$ and $\alpha_j=-\tau_j(\theta)$ for $j\geq t$. By lemmas D.2 and D.3, $\tau_j(\theta)=O\big((1+\log j)j^{\max(-d,-\zeta)-1}\big)$ and $\sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta)=O\big((1+\log t)^2t^{\max(-d,-\zeta)-1}\big)$, so that $\alpha_j=O\big((1+\log t)^2t^{\max(-d,-\zeta)-1}\big)$ for $j<t$ and $\alpha_j=O\big((1+\log j)j^{\max(-d,-\zeta)-1}\big)$ for $j\geq t$. Apply the Beveridge-Nelson decomposition to $r_{\eta,t}(\theta)$

$$r_{\eta,t}(\theta)=\Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j+\Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j},\qquad\alpha_j^*=-\sum_{i=j+1}^{\infty}\alpha_i,\tag{B.17}$$

where $\sum_{j=1}^{\infty}\alpha_j=O((1+\log t)^2t^{\max(-d,-\zeta)})$. Again, by the Beveridge-Nelson decomposition for $\Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j}$ in (B.15)

$$\Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j}=\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)+\Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j},\tag{B.18}$$

where $\tau_j^*(\theta) = -\sum_{i=j+1}^{\infty} \tau_i(\theta)$, and $\sum_{j=0}^{\infty} \tau_j(\theta) = O(1)$ by lemma D.2. By (B.15), (B.17), and (B.18), it follows for the first term in (B.13) that

$$\frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\Big)^2 \geq \frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)^2 \tag{B.19}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)\Big(\Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j\Big)\right] \tag{B.20}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}\Big)\right] \tag{B.21}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\Big)\right] \tag{B.22}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}\Big)\Big(\Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j\Big)\right] \tag{B.23}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\Big)\right] \tag{B.24}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\Big)\right]. \tag{B.25}$$

From (B.12), it immediately follows that (B.21) to (B.25) are $O_p(1)$, as $d - d_0 + 1 > 0$ and $d - d_0 > -1$ for all $\theta \in \Theta_2(\kappa_2, \kappa_3)$. In addition, as $\sum_{j=1}^{\infty}\alpha_j = O((1+\log t)^2 t^{\max(-d,-\zeta)})$ and as $\sum_{j=0}^{\infty}\tau_j(\theta)$ is bounded away from zero by assumption 3, it follows that (B.19) asymptotically dominates (B.20), so that the rate of convergence of (B.13) will depend solely on (B.19). The asymptotic probability limit of the first term (B.19) is derived analogously to Nielsen (2015, pp. 163f) by defining $w_t = \sum_{i=0}^{N-1}\pi_i(d-d_0)\eta_{t-i}\sum_{j=0}^{\infty}\tau_j(\theta)$ and $u_t = \sum_{i=N}^{t-1}\pi_i(d-d_0)\eta_{t-i}\sum_{j=0}^{\infty}\tau_j(\theta)$ for some $N \geq 1$ to be determined. Then $\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta) = w_t + u_t$, and it holds for (B.19)

$$\frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)^2 \geq \frac{1}{n}\sum_{t=N+1}^{n}\left(w_t^2 + 2w_tu_t\right). \tag{B.26}$$

As shown by Nielsen (2015, p. 164), for some κ satisfying $\max(\kappa_2,\kappa_3) \leq \kappa < 1/2$, setting $N = n^\alpha$ with $0 < \alpha < \min\left(\frac{1/2-\kappa}{1/2+\kappa}, \frac{1/2}{1/2+2\kappa}\right)$, it holds by Nielsen (2015, eqn. B.4 in lemma B.2) that $n^{-1}\sum_{t=n^\alpha+1}^{n} w_tu_t \xrightarrow{p} 0$ uniformly in $\theta \in \Theta_2(\kappa,\kappa) \supseteq \Theta_2(\kappa_2,\kappa_3)$. As also shown by Nielsen (2015, p. 164), the other term in (B.26) satisfies

$$\sup_{\theta\in\Theta_2(\kappa,\kappa)}\left|\frac{1}{n}\sum_{t=n^\alpha+1}^{n} w_t^2 - \sigma_{\eta,0}^2\Big(\sum_{j=0}^{\infty}\tau_j(\theta)\Big)^2\sum_{j=0}^{n^\alpha-1}\pi_j^2(d-d_0)\right| \xrightarrow{p} 0, \tag{B.27}$$

as $n \to \infty$, and by Nielsen (2015, lemma A.3) the latter sum is bounded from below by $\sum_{j=0}^{n^\alpha-1}\pi_j^2(d-$

$d_0) \geq 1 + K\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}$ for some $K > 0$. The limit of the fraction $\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}$ is discussed by Nielsen (2015, p. 165): It increases in n from zero (for $n = 2$) to $1/(2\kappa_3)$ as $n \to \infty$, and decreases in κ_3 from $\alpha \log(n-1)$ for $\kappa_3 = 0$ to zero for $\kappa_3 \to 1/2$. Consequently $\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3} \to \infty$ as $(n, \kappa_3) \to (\infty, 0)$. This, together with (B.19), (B.26), and (B.27) yields that the lower bound of $\frac{1}{n}\sum_{t=1}^{n}(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j})^2$ diverges in probability for $\theta \in \Theta_2(\kappa,\kappa)$ as $(n,\kappa) \to (\infty, 0)$. By (B.13), (B.14), and (B.15) the result of Nielsen (2015, eqn. 25) for ARFIMA models carries over to the fractional UC model: For any $K > 0$, $\delta > 0$, there exist $\bar{\kappa}_3 > 0$ and $T_2 \geq 1$ such that

$$\Pr\left(\inf_{d\in D_2(\kappa_2,\bar{\kappa}_3),\nu\in\Sigma_\nu,\varphi\in\Phi} Q(y,\theta) > K\right) \geq 1-\delta, \quad \text{for all } T \geq T_2, \tag{B.28}$$

and (B.28) holds for any $\kappa_2 \in (0, 1/2)$.

Convergence on $\Theta_1(\kappa_1)$ Finally, consider the non-stationary subset $\Theta_1(\kappa_1) = D_1(\kappa_1) \times \Sigma_\nu \times \Phi$. Starting again with (B.13) above, the second term in (B.13), by the same argument with respect to absolute summability of the coefficients as for (B.14), is now

$$\frac{1}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right)\left(\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\right) = O_p\left(1+\log(n)n^{d_0-d-1}\right), \tag{B.29}$$

for all $\theta \in \Theta_1(\kappa_1)$ by (B.12) with $d_1 = d - d_0$, $d_2 = d$, and thus is $O_p(1)$ for $d - d_0 > -1$ and $O_p(\log(n)n^{d_0-d-1})$ otherwise. As will be shown, the first term in (B.13) will asymptotically diverge at a faster rate compared to the second term above. To see this, note that the decomposition of the first term in (B.13) into $\Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j}$ and $r_{\eta,t}(\theta)$ in (B.15) and (B.16) above also applies in $\Theta_1(\kappa_1)$. Consequently, the Beveridge-Nelson decompositions in (B.17) and (B.18) also hold for $\theta \in \Theta_1(\kappa_1)$. Again, the decomposition in (B.19) to (B.25) applies, however the terms in (B.21) to (B.25) will not necessarily be $O_p(1)$, since $d - d_0$ is no longer bounded from above by -1 or by -2. However, as will become clear, the first term (B.19) asymptotically dominates all other terms in (B.20) to (B.25) and thus it will be sufficient to consider only this term.

To arrive at the desired result, consider $n^{2(d-d_0)}\sum_{t=1}^{n}(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta))^2$, a scaled version of (B.19). It follows from the Cauchy-Schwarz inequality that

$$n^{2(d-d_0)}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\right)^2 \geq \left(n^{d-d_0-1/2}\sum_{t=1}^{n}\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\right)^2, \tag{B.30}$$

where the scaling by $n^{d-d_0-1/2}$ is required for a functional central limit theorem later to hold.

The remaining proof for $\theta \in \Theta_1(\kappa_1)$ follows Nielsen (2015, pp. 168f) and shows his results for the CSS estimator for ARFIMA processes to carry over to the fractional UC model. As also shown there, from Hosoya (2005, thm. 2) a functional central limit theorem for

$$r_n(\theta) = n^{d-d_0-1/2}\sum_{t=1}^{n}\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta) = n^{d-d_0-1/2}\Delta_+^{d-d_0-1}\eta_n\sum_{j=0}^{\infty}\tau_j(\theta) \tag{B.31}$$

follows if assumptions A(i) to A(iv) of Hosoya (2005) hold. Since $0 < \sum_{j=0}^{\infty} |\tau_j(\theta)| < \infty$ and $\mathrm{E}(\eta_j|\mathcal{F}_t) = 0$ for all $j > t$, as well as $\mathrm{E}(\eta_j\eta_k|\mathcal{F}_t) - \mathrm{E}(\eta_j\eta_k) = 0$ for $j,k > t$ by assumption 1, it follows that assumptions A(i) and A(ii) of Hosoya (2005) are satisfied. By Hosoya (2005, lemma 3), assumption A(iii) of Hosoya (2005) is satisfied if η_t is a fourth-order stationary process with a bounded fourth-order cumulant spectral density, which is satisfied by assumption 1. Finally, by Hosoya (2005, thm. 3) the respective assumption A(iv) is satisfied for the fourth-order stationary process η_t if $2 > (2(d_0 - d + 1) - 1)^{-1}$ holds, which is equivalent to $d_0 - d > -1/4$ and is satisfied for all $\theta \in \Theta_1(\kappa_1)$. By Hosoya (2005, thm. 2), as $n \to \infty$

$$n^{d-d_0-1/2}\Delta_+^{d-d_0-1}\eta_{\lfloor nr\rfloor}\sum_{j=0}^{\infty}\tau_j(\theta) \Rightarrow W_{d_0-d}(r) \quad \text{in } \mathcal{D}[0,1], \tag{B.32}$$

for $r \in [0,1]$ and fixed $d \in D_1(\kappa_1)$, where $\lfloor nr\rfloor$ is the greatest integer smaller or equal to nr, $W_{d_0-d}(r) = \Gamma(d_0-d+1)^{-1}\int_0^r (r-s)^{d_0-d}\mathrm{d}W(s)$ is fractional Brownian motion of type II, and W denotes Brownian motion generated by $\eta_t \sum_{j=0}^{\infty}\tau_j(\theta)$. (B.32) is equivalent to Nielsen (2015, eqn. 30) for the univariate case. From (B.32) it follows that $r_n(\theta) \xrightarrow{d} r(\theta) = W_{d_0-d}(1)$ for fixed $d \in D_1(\kappa_1)$. Pointwise convergence $r_n(\theta)$ can be generalized to uniform convergence in $D_1(\kappa_1)$ if $r_n(\theta)$ is tight (stochastically equicontinuous) as a function of θ on $\theta \in \Theta_1(\kappa_1)$. Since the parameters φ, ν only enter $r_n(\theta)$ through $\sum_{j=0}^{\infty}\tau_j(\theta)$, it is sufficient for tightness of $r_n(\theta)$ in θ that $n^{d-d_0-1/2}\Delta_+^{d-d_0-1}\eta_n$ is tight in $(d-d_0)$. As in Nielsen (2015, pp. 169f), tightness in $(d-d_0)$ can be shown using the moment condition in Billingsley (1968, thm. 12.3) which requires to show that $r_n(\theta)$ is tight for a fixed $d-d_0$ and that $|n^{d_1-1/2}\Delta_+^{d_1-1}\eta_n - n^{d_2-1/2}\Delta_+^{d_2-1}\eta_n| \le K|d_1-d_2|$ for some constant $K > 0$ that does not depend on n, d_1, or d_2, see Nielsen (2015, pp. 169f). As noted there, the first condition is implied by pointwise convergence in probability and distribution, while the second condition holds by Nielsen (2015, lemma B.1). Consequently, $r_n(\theta) \Rightarrow r(\theta)$ in $d \in D_1(\kappa_1)$, and thus $\inf_{\theta\in\Theta_1(\kappa_1)} r_n(\theta)^2 \xrightarrow{d} \inf_{\theta\in\Theta_1(\kappa_1)} r(\theta)^2$.

Coming back to the first term of the objective function (B.13), for which a lower bound is given by the expressions (B.19) to (B.25), note that by (B.30) the first term (B.19) is bounded from below (when scaled appropriately) by

$$\inf_{\theta\in\Theta_1(\kappa_1)} \frac{1}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\right)^2 \ge n^{2(d_0-d-1/2)}\inf_{\theta\in\Theta_1(\kappa_1)} r_n(\theta)^2. \tag{B.33}$$

The probability limits of (B.21) to (B.25) can be derived by (B.12) for $d_1 = d-d_0$ and $d_2 = d-d_0+1$, and equal $O_p(1 + n^{-a}\log n)$, where $a = \min(1+d-d_0, 2+2(d-d_0))$. Thus, $a = 1+d-d_0$ if $d-d_0 > -1$, and $a = 2+2(d-d_0)$ if $d-d_0 \le -1$. In the former case, $a > 0$, so that (B.21) to (B.25) are $O_p(1)$. In the latter case, they are $O_p\left(n^{2(d_0-d-1)}\log n\right)$ and thus diverge at a slower rate than (B.19). For (B.20), note that $\sum_{j=1}^{\infty}\alpha_j = O((1+\log t)^2 t^{\max(-d,-\zeta)})$, while $\sum_{j=0}^{\infty}\tau_j(\theta)$ is bounded away from zero by assumption 3. Consequently, (B.20) will also diverge at a slower rate than (B.19). Finally, as already shown in (B.29), the second term in (B.13) is $O_p\left(1 + \log(n)n^{d_0-d-1}\right)$ and thus is also dominated by (B.19). It follows that the rate of divergence of the objective function is determined by the first term in (B.13) and is given by the divergence rate of (B.19). This, together with (B.33), yields

$$\inf_{\theta\in\Theta_1(\kappa_1)} Q(y,\theta) \geq n^{2(d_0-d-1/2)} \inf_{\theta\in\Theta_1(\kappa_1)} r_n(\theta)^2 \geq n^{2\kappa_1} \inf_{\theta\in\Theta_1(\kappa_1)} r_n(\theta)^2 \tag{B.34}$$

as $n \to \infty$. Thus, one obtains the result of Nielsen (2015, eqn. 34) that for any $K > 0$ and all $\kappa_1 > 0$

$$\Pr\left(\inf_{d\in D_1(\kappa_1),\nu\in\Sigma_\nu,\varphi\in\Phi} \frac{1}{n} Q(y,\theta) > K \right) \to 1, \quad \text{as } T \to \infty. \tag{B.35}$$

Together, (B.28) and (B.35) prove (B.3). $\square$

C Proof of theorem 4.2

Proof of theorem 4.2. Since $\hat{\theta}$ is consistent, see theorem 4.1, the asymptotic distribution theory can be derived based on the Taylor series expansion of the score function as usual

$$0 = \sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\hat{\theta}} = \sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} + \sqrt{n}\frac{\partial^2 Q(y,\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\bar{\theta}} \left(\hat{\theta} - \theta_0\right), \tag{C.1}$$

where for the entries of $\bar{\theta}$ it holds that $|\bar{\theta}_{(i)} - \theta_{0_{(i)}}| \leq |\hat{\theta}_{(i)} - \theta_{0_{(i)}}|$ for all $i = 1, ..., q+2$. The normalized score at θ_0 is

$$\sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} = \frac{2}{\sqrt{n}}\sum_{t=1}^{n} v_t(\theta_0)\frac{\partial v_t(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}, \tag{C.2}$$

with $v_t(\theta)$ denoting the prediction error as defined in (14) and (15), and its partial derivative as given in (B.11). Denote the normalized, untruncated score

$$\sqrt{n}\frac{\partial \tilde{Q}(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} = \frac{2}{\sqrt{n}}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)\frac{\partial \tilde{v}_t(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}, \tag{C.3}$$

with $\tilde{v}_t(\theta)$ as defined in (B.2). As shown in lemma D.6, the difference between truncated and untruncated score is asymptotically negligible. Therefore it is sufficient to consider the distribution of the latter. By assumption 5, the untruncated prediction error $\tilde{v}_t(\theta_0)$ is a stationary MDS when adapted to $\mathcal{F}_t^{\tilde{\xi}} = \sigma(\tilde{\xi}_s, s \leq t)$. Thus, for (C.3) a central limit theorem can be shown to apply following Nielsen (2015, p. 175): By the Cramér-Wold device it is sufficient to show that for any $q+2$-dimensional vector μ, $\mu'\sqrt{n}\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}|_{\theta=\theta_0} = \sqrt{n}\sum_{i=1}^{q+2}\mu_{(i)}\left(\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}|_{\theta=\theta_0}\right)_{(i)} = \frac{2}{\sqrt{n}}\sum_{i=1}^{q+2}\mu_{(i)}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{1,t} + \tilde{h}_{2,t})_{(i)} \xrightarrow{d} \mathrm{N}(0, 4\sigma_{v,0}^2\mu'\Omega_0\mu)$ as $n \to \infty$, with $\tilde{h}_{1,t} = \sum_{j=1}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$, as well as $\tilde{h}_{2,t} = \sum_{j=0}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$. As $\tilde{h}_{1,t}$ and $\tilde{h}_{2,t}$ are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, $\nu_t = \sum_{i=1}^{q+2}\mu_{(i)}\tilde{v}_t(\theta_0)(\tilde{h}_{1,t} + \tilde{h}_{2,t})_{(i)}$ together with $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS. Thus, by the law of large numbers for stationary and ergodic processes, it holds that

$$\frac{1}{n}\sum_{t=1}^{n}\mathrm{E}\left(\nu_t^2|\mathcal{F}_{t-1}^{\tilde{\xi}}\right)=\frac{1}{n}\sum_{t=1}^{n}\sum_{i,j=1}^{q+2}\mu_{(i)}\mu_{(j)}\sigma_{v,0}^2(\tilde{h}_{1,t}+\tilde{h}_{2,t})_{(i)}(\tilde{h}_{1,t}+\tilde{h}_{2,t})_{(j)}$$
$$=\sum_{i,j=1}^{q+2}\mu_{(i)}\mu_{(j)}\sigma_{v,0}^2\frac{1}{n}\sum_{t=1}^{n}(\tilde{h}_{1,t}+\tilde{h}_{2,t})_{(i)}(\tilde{h}_{1,t}+\tilde{h}_{2,t})_{(j)}\xrightarrow{p}\sigma_{v,0}^2\sum_{i,j=1}^{q+2}\mu_{(i)}\mu_{(j)}\Omega_{0_{(i,j)}},$$

with $\sigma_{v,0}^2=\mathrm{E}(\tilde{v}_t^2(\theta_0)|\mathcal{F}_{t-1}^{\tilde{\xi}})=\mathrm{E}(\tilde{v}_t^2(\theta_0))$, and $\Omega_{0_{(i,j)}}=\mathrm{E}\left[\frac{\partial\tilde{v}_t(\theta)}{\partial\theta_{(i)}}\big|_{\theta=\theta_0}\frac{\partial\tilde{v}_t(\theta)}{\partial\theta_{(j)}}\big|_{\theta=\theta_0}\right]$. Finally, the Lindeberg criterion is satisfied as $\tilde{v}_t(\theta_0)$ is stationary. It follows directly that $\sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta}\big|_{\theta=\theta_0}=\sqrt{n}\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}\big|_{\theta=\theta_0}+o_p(1)\xrightarrow{d}\mathrm{N}(0,4\sigma_{v,0}^2\Omega_0)$.

Next, consider the second derivatives in (C.1). By Johansen and Nielsen (2010, lemma A.3), the Hessian matrix in (C.1) can be evaluated at the true parameters θ_0 if $\hat{\theta}$ is consistent and if the second derivatives are tight (stochastically equicontinuous). As also discussed by Nielsen (2015) for the CSS estimator of ARFIMA models, tightness holds for the second derivatives if its derivatives are uniformly dominated in $d\in D_3$ as defined in the proof of theorem 4.1, $\nu\in\Sigma_\nu$ as defined in section 4, and $\varphi\in N_\delta(\varphi_0)$ as defined in assumptions 2 and 4, by a random variable $B_n=O_p(1)$, see Newey (1991, cor. 2.2). This holds by lemma D.7. Therefore, the second derivative in (C.1) can be evaluated at the true value θ_0

$$\frac{\partial^2 Q(y,\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\bigg|_{\theta=\theta_0}=\frac{2}{n}\sum_{t=1}^{n}\frac{\partial v_t(\theta)}{\partial\theta_{(k)}}\bigg|_{\theta=\theta_0}\frac{\partial v_t(\theta)}{\partial\theta_{(l)}}\bigg|_{\theta=\theta_0}+\frac{2}{n}\sum_{t=1}^{n}v_t(\theta_0)\frac{\partial^2 v_t(\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\bigg|_{\theta=\theta_0},\tag{C.4}$$

$k,l=1,2,...,q+2$. By lemma D.8, as $t\to\infty$,

$$\mathrm{E}\left[\left(\frac{\partial\tilde{v}_t(\theta)}{\partial\theta}-\frac{\partial v_t(\theta)}{\partial\theta}\right)\bigg|_{\theta=\theta_0}\left(\frac{\partial\tilde{v}_t(\theta)}{\partial\theta'}-\frac{\partial v_t(\theta)}{\partial\theta'}\right)\bigg|_{\theta=\theta_0}\right]\xrightarrow{p}0.$$

From the law of large numbers for stationary and ergodic processes, it then holds for the first term in (C.4) that $\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\tilde{v}_t(\theta)}{\partial\theta}\frac{\partial\tilde{v}_t(\theta)}{\partial\theta'}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial v_t(\theta)}{\partial\theta}\frac{\partial v_t(\theta)}{\partial\theta'}+o_p(1)$. In addition, by lemma D.9 the second term in (C.4) is $\frac{2}{n}\sum_{t=1}^{n}v_t(\theta_0)\frac{\partial^2 v_t(\theta)}{\partial\theta\partial\theta'}\big|_{\theta=\theta_0}=\frac{2}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\frac{\partial^2\tilde{v}_t(\theta)}{\partial\theta\partial\theta'}\big|_{\theta=\theta_0}+o_p(1)$. As $(\tilde{v}_t(\theta_0),\mathcal{F}_t^{\tilde{\xi}})$ is a stationary MDS, while the second partial derivatives are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, it holds that $\frac{2}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\frac{\partial^2\tilde{v}_t(\theta)}{\partial\theta\partial\theta'}\big|_{\theta=\theta_0}=o_p(1)$. Taken together, this implies for (C.4) that

$$\frac{\partial^2 Q(y,\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\bigg|_{\theta=\theta_0}=\frac{2}{n}\sum_{t=1}^{n}\frac{\partial\tilde{v}_t(\theta)}{\partial\theta_{(k)}}\bigg|_{\theta=\theta_0}\frac{\partial\tilde{v}_t(\theta)}{\partial\theta_{(l)}}\bigg|_{\theta=\theta_0}+o_p(1).\tag{C.5}$$

Finally, from the law of large numbers, it follows that $\frac{\partial^2 Q(y,\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\big|_{\theta=\theta_0}\xrightarrow{p}2\Omega_{0_{(k,l)}}$. Thus, solving (C.1) for $\sqrt{n}(\hat{\theta}-\theta_0)$ yields the desired result

$$\sqrt{n}(\hat{\theta}-\theta_0)=-\left[\frac{\partial^2 Q(y,\theta)}{\partial\theta\partial\theta'}\right]_{\theta=\bar{\theta}}^{-1}\sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta'}\bigg|_{\theta=\theta_0}\xrightarrow{d}\mathrm{N}(0,\sigma_{v,0}^2\Omega_0^{-1}).$$

$\square$

D Additional lemmas

In what follows, let $z_{(j)}$ denote the j-th entry for some vector z, and let $Z_{(i,j)}$ denote the (i, j)-th entry (i.e. the entry in row i and column j) for some matrix Z.

Lemma D.1 (Convergence rates of $\pi_j(d)$, $b_j(\varphi)$, and related vector and matrix entries)**.** *It holds that*

$$\pi_j(d) = O(j^{-d-1}), \tag{D.1}$$

$$b_j(\varphi) = O(j^{-\zeta-1}), \tag{D.2}$$

$$(B'_{\varphi,t}B_{\varphi,t})_{(i,j)} = \begin{cases} O(|i-j|^{-\zeta-1}) & \text{for } i \neq j, \\ O(1) & \text{for } i = j, \end{cases} \tag{D.3}$$

$$(S'_{d,t}S_{d,t})_{(i,j)} = \begin{cases} O(|i-j|^{-d-1}) & \text{for } i \neq j, \\ O(1) & \text{for } i = j, \end{cases} \tag{D.4}$$

$$(B'_{\varphi,t}B_{\varphi,t})^{-1}_{(i,j)} = \begin{cases} O(|i-j|^{-\zeta-1}) & \text{for } i \neq j, \\ O(1) & \text{for } i = j, \end{cases} \tag{D.5}$$

$$(B_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t})^{-1}_{(i,j)} = \begin{cases} O(|i-j|^{\max(-d,-\zeta)-1}) & \text{for } i \neq j, \\ O(1) & \text{for } i = j, \end{cases} \tag{D.6}$$

$$(B'_{\varphi,t}\beta_t)_{(j)} = O((t-j+1)^{-\zeta-1}), \tag{D.7}$$

$$(S'_{d,t}s_t)_{(j)} = O((t-j+1)^{-d-1}), \tag{D.8}$$

with $\pi_j(d)$ as defined in (3)*, $b_j(\varphi)$ as defined below assumption 3, $B_{\varphi,t}$ and $S_{d,t}$ as defined in* (5)*, and $\beta'_t = (b_t(\varphi) \cdots b_1(\varphi))$, $s'_t = (\pi_t(d) \cdots \pi_1(d))$.*

Proof of Lemma D.1. (D.1) follows by Johansen and Nielsen (2010, lemma B.3) while (D.2) follows by assumption 3. (D.3) follows from (D.2) by $(B'_{\varphi,t}B_{\varphi,t})_{(i,j)} = \sum_{k=0}^{\min(i,j)-1} b_k(\varphi)b_{k+|i-j|}(\varphi) = O(|i-j|^{-\zeta-1})\sum_{k=0}^{\min(i,j)-1} b_k(\varphi) = O(|i-j|^{-\zeta-1})$ for $i \neq j$, and $(B'_{\varphi,t}B_{\varphi,t})_{(i,i)} = \sum_{k=0}^{i-1} b_k^2(\varphi) = O(1)$. The proof for (D.4) is analogous and follows from (D.1), as $(S'_{d,t}S_{d,t})_{(i,j)} = \sum_{k=0}^{\min(i,j)-1} \pi_k(d)\pi_{k+|i-j|}(d) = O(|i-j|^{-d-1})$ for $i \neq j$, $(S'_{d,t}S_{d,t})_{(i,i)} = O(1)$.

To derive the convergence rates for the entries of $(B'_{\varphi,t}B_{\varphi,t})^{-1}$ and $(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t})^{-1}$ in (D.5) and (D.6), note that as $t \to \infty$, $B'_{\varphi,t}B_{\varphi,t}$ and $B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t}$ converge to the Toeplitz matrices[10] $T_t(f_1)$ and $T_t(f_2)$ with symbols $f_1(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_1(j)e^{i\lambda j}$, $\gamma_1(j) = \sum_{k=0}^{\infty} b_k(\varphi)b_{k+j}(\varphi)$, $f_2(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_2(j)e^{i\lambda j}$, $\gamma_2(j) = \sum_{k=0}^{\infty}[b_k(\varphi)b_{k+j}(\varphi) + \nu\pi_k(d)\pi_{k+j}(d)]$, where $\gamma_1(j) = O(j^{-\zeta-1})$ and $\gamma_2(j) = O(j^{\max(-d,-\zeta)-1})$ as $j \to \infty$. Consequently, $(B'_{\varphi,t}B_{\varphi,t})^{-1}$ and $(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t})^{-1}$ converge to the Toeplitz matrices $T_t(1/f_1)$ and $T_t(1/f_2)$ that exist by assumption 3. Denote the respective spectral densities as $1/f_1(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_3(j)e^{i\lambda j}$ and $1/f_4(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_4(j)e^{i\lambda j}$.

[10] Gray (2006) provides a good overview about the asymptotic behavior of Toeplitz matrices.

Then the convergence rate of $\gamma_3(j)$ can be obtained from the partial derivative $(\partial/\partial\lambda)[1/f_1(\lambda)] = (2\pi)^{-1}\sum_{j=0}^{\infty} ij\gamma_3(j)e^{i\lambda j} = -f_1(\lambda)^{-2}(2\pi)^{-1}\sum_{j=0}^{\infty} ij\gamma_1(j)e^{i\lambda j}$, where $j\gamma_1(j) = O(j^{-\zeta})$, so that $j\gamma_3(j) = O(j^{-\zeta})$ as $f_1(\lambda)$ is bounded away from zero by assumption 3. It follows that $\gamma_3(j) = O(j^{-\zeta-1})$. Similarly, it can be shown that $\gamma_4(j) = O(j^{\max(-d,-\zeta)-1})$. As the j-th descending diagonals of $(B'_{\varphi,t}B_{\varphi,t})^{-1}$ and $(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t})^{-1}$ converge to $\gamma_3(j)$ and $\gamma_4(j)$ as $t \to \infty$, one has (D.5) and (D.6).
(D.7) follows immediately from (D.2), since $(B'_{\varphi,t}\beta_t)_{(j)} = \sum_{k=0}^{j-1} b_k(\varphi)b_{t-j+k+1}(\varphi) = O((t-j+1)^{-\zeta-1})\sum_{k=0}^{j-1} b_k(\varphi) = O((t-j+1)^{-\zeta-1})$, while (D.8) follows immediately from (D.1) by $(S'_{d,t}s_{t+1})_{(j)} = \sum_{k=0}^{j-1}\pi_k(d)\pi_{t-j+k+1}(d) = O((t-j+1)^{-d-1})\sum_{k=0}^{j-1}\pi_k(d) = O((t-j+1)^{-d-1})$. $\square$

Lemma D.2 (Convergence rates of $\tau_j(\theta,t)$)**.** *For the coefficients $\tau_j(\theta,t)$ as defined in* (15) *and below, it holds that*

$$\tau_j(\theta,t) = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right). \tag{D.9}$$

Proof of Lemma D.2. To prove (D.9), consider $\tau_j(\theta,t)$ as defined in (15) and below

$$\tau_j(\theta,t) = \nu\sum_{k=1}^{t}\left[\begin{pmatrix} b_1(\varphi)-\pi_1(d) & \cdots & b_t(\varphi)-\pi_t(d)\end{pmatrix}(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}\right]_{(k)} S_{d,t_{(j,k)}}. \tag{D.10}$$

The left term in (D.10) is

$$\begin{aligned}
&\left[\begin{pmatrix} b_1(\varphi)-\pi_1(d) & \cdots & b_t(\varphi)-\pi_t(d)\end{pmatrix}(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}\right]_{(k)}\\
&= (b_k(\varphi)-\pi_k(d))\,(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}_{(k,k)}\\
&+\sum_{i=1}^{k-1}(b_i(\varphi)-\pi_i(d))\,(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}_{(i,k)}\\
&+\sum_{i=k+1}^{t}(b_i(\varphi)-\pi_i(d))\,(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}_{(i,k)}.
\end{aligned} \tag{D.11}$$

Note that $\pi_k(d) = O(k^{-d-1})$, $b_k(\varphi) = O(k^{-\zeta-1})$, $(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}_{(k,k)} = O(1)$, and $(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}_{(i,k)} = O(|i-k|^{\max(-d,-\zeta)-1})$ for $i \neq k$ by (D.1), (D.2), and (D.6). Thus, the first term in (D.11) is $O\left(k^{\max(-d,-\zeta)-1}\right)$, while the second term is $\sum_{i=1}^{k-1} O\left(i^{\max(-d,-\zeta)-1}(k-i)^{\max(-d,-\zeta)-1}\right) = O\left((1+\log k)k^{\max(-d,-\zeta)-1}\right)$, where the last equality follows from Johansen and Nielsen (2010, lemma B.4), who show that $\sum_{i=1}^{k-1} i^{\max(-d,-\zeta)-1}(k-i)^{\max(-d,-\zeta)-1} = O((1+\log k)k^{\max(-d,-\zeta)-1})$. Analogously, it holds for the third term in (D.11) that $\sum_{i=k+1}^{t} O\left(i^{\max(-d,-\zeta)-1}(i-k)^{\max(-d,-\zeta)-1}\right) = O((k+1)^{\max(-d,-\zeta)-1}\sum_{i=k+1}^{t}(i-k)^{\max(-d,-\zeta)-1}) = O((k+1)^{\max(-d,-\zeta)-1})$. Therefore

$$\begin{aligned}
&\left[\begin{pmatrix} b_1(\varphi)-\pi_1(d) & \cdots & b_t(\varphi)-\pi_t(d)\end{pmatrix}(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}\right]_{(k)}\\
&= O\left((1+\log k)k^{\max(-d,-\zeta)-1}\right).
\end{aligned} \tag{D.12}$$

By plugging (D.12) into (D.10) and using (5) together with (D.1), one obtains

$$\begin{aligned}
&\Big[\Big(b_1(\varphi)-\pi_1(d) \quad \cdots \quad b_t(\varphi)-\pi_t(d)\Big)\,(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}S'_{d,t}\Big]_{(j)}\\
&=\sum_{k=j}^{t}\Big[\Big(b_1(\varphi)-\pi_1(d) \quad \cdots \quad b_t(\varphi)-\pi_t(d)\Big)\,(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}\Big]_{(k)}\pi_{k-j}(d)\\
&=O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\Big)+O\Big(\sum_{k=j+1}^{t}(1+\log k)k^{\max(-d,-\zeta)-1}(k-j)^{-d-1}\Big)\\
&=O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\Big)+O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\sum_{k=1}^{t-j}k^{-d-1}\Big)\\
&=O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right), \qquad (D.13)
\end{aligned}$$

since $\sum_{k=1}^{t-j}k^{-d-1}=O(1)$ for all $d>0$. This proves (D.9).

□

Lemma D.3 (Convergence of $\tau_j(\theta,t)$ as $t\to\infty$)**.** *For the coefficients $\tau_j(\theta,t)$ as defined in* (15) *and below, it holds that*

$$\tau_j(\theta,t)=\tau_j(\theta,t+1)+r_{\tau,j,t+1}(\theta), \qquad (D.14)$$

where $r_{\tau,j,t+1}(\theta)=O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1}\big)$.

Proof of Lemma D.3. To prove (D.14), I study the impact of an increase from t to $t+1$ on $\tau_j(\theta,t+1)=\nu[(b_1(\varphi)-\pi_1(d)\cdots b_{t+1}(\varphi)-\pi_{t+1}(d))(B'_{\varphi,t+1}B_{\varphi,t+1}+\nu S'_{d,t+1}S_{d,t+1})^{-1}S'_{d,t+1}]_{(j)}$. Denote

$$B_{\varphi,t+1}=\begin{bmatrix}B_{\varphi,t} & \beta_t\\ 0_{1\times t} & 1\end{bmatrix}, \qquad S_{d,t+1}=\begin{bmatrix}S_{d,t} & s_t\\ 0_{1\times t} & 1\end{bmatrix}, \qquad (D.15)$$

with $\beta_t=(b_t(\varphi)\cdots b_1(\varphi))'$ and $s_t=(\pi_t(d)\cdots\pi_1(d))'$. Let $\Xi_{t+1}(\theta)=(B'_{\varphi,t+1}B_{\varphi,t+1}+\nu S'_{d,t+1}S_{d,t+1})^{-1}$. Then, by the Sherman-Morrison formula

$$\Xi_{t+1}(\theta)=\begin{bmatrix}\Xi_t(\theta)+R_1 & R_2\\ R'_2 & R_3\end{bmatrix}, \qquad (D.16)$$

with the block entries

$$\begin{aligned}
R_3&=[(1+\beta'_t\beta_t+\nu+\nu s'_ts_t)-(\beta'_tB_{\varphi,t}+\nu s'_tS_{d,t})\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t)]^{-1},\\
R_2&=-R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t),\\
R_1&=R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t)(\beta'_tB_{\varphi,t}+\nu s'_tS_{d,t})\Xi_t(\theta).
\end{aligned}$$

Clearly $R_3 = O(1)$, since by (D.6), (D.7) and (D.8)

$$[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)]_{(j)} = O\Big(\sum_{i=1}^{j-1}(t+1-i)^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\Big)$$
$$+ O((t+1-j)^{\max(-d,-\zeta)-1}) + O\Big(\sum_{i=1}^{t-j}(t+1-i-j)^{\max(-d,-\zeta)-1} i^{\max(-d,-\zeta)-1}\Big)$$
$$= O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right), \tag{D.17}$$

and again by (D.7) and (D.8)

$$(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)(B_{\varphi,t}'\beta_t + \nu S_{d,t}' s_t)$$
$$= O\Big(\sum_{j=1}^{t}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}(t+1-j)^{\max(-d,-\zeta)-1}\Big),$$

which is $O(1)$. This, together with $1+\beta_t'\beta_t+\nu+\nu s_t' s_t = \sum_{j=0}^t b_j^2(\varphi) + \nu\sum_{j=0}^t \pi_j^2(d) = O(1)$, yields $R_3^{-1} = O(1)$. Furthermore, R_3^{-1} is bounded away from zero, as $\Xi_t(\theta)^{-1}$ is regular by assumption 3. For R_2, by (D.17) it follows that $R_{2_{(j)}} = O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$. Finally, for R_1, by (D.17) it follows that $R_{1_{(i,j)}} = O\big((1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big)$.

Next, consider the vector

$$(b_1(\varphi)-\pi_1(d)\cdots b_{t+1}(\varphi)-\pi_{t+1}(d))(B_{\varphi,t+1}'B_{\varphi,t+1}+\nu S_{d,t+1}'S_{d,t+1})^{-1}$$
$$= \Big((b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))[\Xi_t(\theta)+R_1] + (b_{t+1}(\varphi)-\pi_{t+1}(d))R_2' \qquad R_4\Big),$$

where $R_4 = (b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2 + (b_{t+1}(\varphi)-\pi_{t+1}(d))R_3$. By (D.1) and (D.2), it holds for the terms in R_4 that $[b_{t+1}(\varphi)-\pi_{t+1}(d)]R_3 = O((t+1)^{\max(-d,-\zeta)-1})$, and $(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2 = O\big(\sum_{j=1}^t j^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big) = O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Thus $R_4 = O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Analogously, for the other terms in the above vector, one has $[(b_{t+1}(\varphi)-\pi_{t+1}(d))R_2']_{(j)} = O\big((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big)$, and $[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1]_{(j)} = O\big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\sum_{i=1}^t(1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1} i^{\max(-d,-\zeta)-1}\big) = O\big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}(1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Therefore, for $j=1,...,t$, the whole term $\tau_j(\theta,t+1)$ is

$$\tau_j(\theta,t+1) = \nu\left((b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S_{d,t}' + R_5'\right)_{(j)} = \tau_j(\theta,t) + \nu R_{5_{(j)}}, \tag{D.18}$$

where $R_5' = [b_{t+1}(\varphi)-\pi_{t+1}(d)]R_2'S_{d,t}' + R_4 s_t' + (b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1 S_{d,t}'$. For R_5

$$[R_2'S_{d,t}']_{(j)} = \sum_{i=j}^{t} R_{2_{(i)}}\pi_{i-j}(d) = R_{2_{(j)}} + \sum_{i=1}^{t-j} R_{2_{(i+j)}}\pi_i(d)$$
$$= O\Big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\Big)$$

$$
\begin{aligned}
&+ O\Big((1+\log(t+1-j))\sum_{i=1}^{t-j}(t+1-i-j)^{\max(-d,-\zeta)-1} i^{-d-1}\Big) \\
&= O\left((1+\log(t+1-j))^2 (t+1-j)^{\max(-d,-\zeta)-1}\right),
\end{aligned}
$$

so that $[(b_{t+1}(\varphi)-\pi_{t+1}(d))R_2' S_{d,t}']_{(j)} = O\big((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1}\big)$, while $[R_4 s_t']_{(j)} = O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}(t+1-j)^{-d-1}\big)$. Furthermore

$$
\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1 S_{d,t}']_{(j)} = \sum_{i=j}^{t}[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1]_{(i)}\pi_{i-j}(d) \\
=&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1]_{(j)} + \sum_{i=1}^{t-j}[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1]_{(i+j)}\pi_i(d) \\
=&O\big((1+\log(t+1))^2(t+1)^{-\min(d,\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{-\min(d,\zeta)-1}\big).
\end{aligned}
$$

Hence, $R_{5_{(j)}} = O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1}\big)$. This completes the proof of (D.14).

□

Lemma D.4 (Convergence rates for partial derivatives of $\tau_j(\theta,t)$)**.** *For the partial derivatives of the coefficients $\tau_j(\theta,t)$, as defined in* (15) *and below, it holds that*

$$
\frac{\partial \tau_j(\theta,t)}{\partial d} = O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \tag{D.19}
$$
$$
\frac{\partial \tau_j(\theta,t)}{\partial \nu} = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \tag{D.20}
$$
$$
\frac{\partial \tau_j(\theta,t)}{\partial \varphi_{(l)}} = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \tag{D.21}
$$

where $\varphi_{(l)}$ denotes the l-th entry of φ, $l=1,...,q$.

Proof of Lemma D.4. Denote $\dot\pi_j(d) = \partial\pi_j(d)/\partial d = O((1+\log j)j^{-d-1})$, see Johansen and Nielsen (2010, lemma B.3), and $\dot b_j(\varphi_{(l)}) = \partial b_j(\varphi)/\partial\varphi_{(l)} = O(j^{-\zeta-1})$ by assumption 3. Furthermore, denote the partial derivatives of $S_{d,t}$ and $B_{\varphi,t}$ as

$$
\dot S_{d,t} = \frac{\partial S_{d,t}}{\partial d} = \begin{bmatrix} 0 & \dot\pi_1(d) & \cdots & \dot\pi_{t-1}(d) \\ 0 & 0 & \cdots & \dot\pi_{t-2}(d) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \qquad \dot B_{\varphi_{(l)},t} = \frac{\partial B_{\varphi,t}}{\partial\varphi_{(l)}} = \begin{bmatrix} 0 & \dot b_1(\varphi_{(l)}) & \cdots & \dot b_{t-1}(\varphi_{(l)}) \\ 0 & 0 & \cdots & \dot b_{t-2}(\varphi_{(l)}) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},
$$

and note that $[\dot S_{d,t}' S_{d,t}]_{(1,j)} = 0$ for all $j=1,...,t$, while for $1 < i \le t$ it holds that

$$
[\dot S_{d,t}' S_{d,t}]_{(i,j)} = \begin{cases} \sum_{k=1}^{i-1}\dot\pi_k(d)\pi_{k+j-i}(d) = O((1+j-i)^{-d-1}) & \text{if } i \le j, \\ \sum_{k=0}^{j-1}\pi_k(d)\dot\pi_{k+i-j}(d) = O((1+\log(i-j))(i-j)^{-d-1}) & \text{if } i > j. \end{cases} \tag{D.22}
$$

Similarly, $[\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}]_{(1,j)} = 0$ for all $j = 1, ..., t$, while for $1 < i \leq t$ one has

$$[\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}]_{(i,j)} = \begin{cases} \sum_{k=1}^{i-1} \dot{b}_k(\varphi_{(l)})b_{k+j-i}(\varphi) = O((1+j-i)^{-\zeta-1}) & \text{if } i \leq j, \\ \sum_{k=0}^{j-1} b_k(\varphi)\dot{b}_{k+i-j}(\varphi_{(l)}) = O((i-j)^{-\zeta-1}) & \text{if } i > j. \end{cases} \tag{D.23}$$

In addition, denote $\Xi_t(\theta) = (B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t})^{-1}$ to simplify the notation. Starting with the partial derivatives $\partial\tau_j(\theta,t)/\partial d$, one has

$$\begin{aligned} \frac{\partial\tau_j(\theta,t)}{\partial d} &= -\nu^2\big[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d)) \times \Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)S'_{d,t}\big]_{(j)} \\ &\quad + \nu[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)\dot{S}'_{d,t}]_{(j)} - \nu[(\dot{\pi}_1(d) \cdots \dot{\pi}_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)}. \end{aligned} \tag{D.24}$$

For the first term, note that by (D.22) $[\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t}]_{(i,j)} = [\dot{S}'_{d,t}S_{d,t}]_{(i,j)} + [\dot{S}'_{d,t}S_{d,t}]_{(j,i)} = O((1 + \log|i-j|)|i-j|^{-d-1})$ for $i \neq j$, and $[\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t}]_{(i,i)} = O(1)$. Together with (D.12) it follows for the first terms in (D.24) that

$$\begin{aligned} &[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t})]_{(j)} \\ &\quad = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right) + O\Big(\sum_{i=1}^{j-1}(1+\log i)i^{\max(-d,-\zeta)-1}(1+\log(j-i))(j-i)^{-d-1}\Big) \\ &\qquad + O\Big(\sum_{i=j+1}^{t}(1+\log i)i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\Big) \\ &\quad = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \end{aligned} \tag{D.25}$$

where for the last equality, note that the second term satisfies $\sum_{i=1}^{j-1} i^{\max(-d,-\zeta)-1}(j-i)^{-d-1} = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right)$, see Johansen and Nielsen (2010, lemma B.4), and that it dominates the first and third term above. Taking into account the next product term for the first term in (D.24), by (D.6) and (D.25)

$$\begin{aligned} &[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)]_{(j)} \\ &\quad = O\Big((1+\log j)^3 j^{\max(-d,-\zeta)-1}\Big) + O\Big(\sum_{i=1}^{j-1}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\Big) \\ &\qquad + O\Big(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{\max(-d,-\zeta)-1}\Big) \\ &\quad = O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \end{aligned} \tag{D.26}$$

where the proof is the same as for (D.25) besides the additional log-factor. Adding the last term, it follows by (D.1) and (D.26) that

$$[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)S'_{d,t}]_{(j)}$$

$$
\begin{aligned}
&= \sum_{i=j}^{t} [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t} S_{d,t} + S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)]_{(i)}\pi_{i-j}(d) \\
&= O\Big((1+\log j)^4 j^{\max(-d,-\zeta)-1}\Big) + O\Big(\sum_{i=j+1}^{t} (1+\log i)^4 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\Big) \\
&= O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \qquad \text{(D.27)}
\end{aligned}
$$

where the second equality uses $\pi_0(d) = 1$ to obtain the first term, while the last equality uses $\sum_{i=1}^{t-j} i^{-d-1} = O(1)$, which holds for all $d > 0$. Consequently, the first term in (D.24) is bounded by $O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right)$. Turning to the second term in (D.24), by (D.12)

$$
\begin{aligned}
&[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)\dot{S}'_{d,t}]_{(j)} \\
&= \sum_{i=j+1}^{t} [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)]_{(i)}\dot{\pi}_{i-j}(d) \\
&= O\Big(\sum_{i=j+1}^{t} (1+\log i) i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\Big) \\
&= O\left((1+\log j) j^{\max(-d,-\zeta)-1}\right), \qquad \text{(D.28)}
\end{aligned}
$$

where the last equality follows from $\sum_{i=1}^{t-j}(1+\log i)i^{-d-1} = O(1)$ for all $d > 0$. By an analogous proof, the third term in (D.24) is

$$
\begin{aligned}
[(\dot{\pi}_1(d) \cdots \dot{\pi}_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)} &= \sum_{i=j}^{t} [(\dot{\pi}_1(d) \cdots \dot{\pi}_t(d))\Xi_t(\theta)]_{(i)}\pi_{i-j}(d) \\
&= O\Big((1+\log j)^2 j^{\max(-d,-\zeta)-1}\Big) + O\Big(\sum_{i=j+1}^{t} (1+\log i)^2 i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\Big) \\
&= O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right). \qquad \text{(D.29)}
\end{aligned}
$$

Together, (D.27), (D.28), and (D.29) yield (D.19).

To prove (D.20), consider the partial derivatives $\partial\tau_j(\theta,t)/\partial\nu$, for which

$$
\begin{aligned}
\frac{\partial\tau_j(\theta,t)}{\partial\nu} &= [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)} \qquad \text{(D.30)} \\
&\quad - \nu[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)S'_{d,t}]_{(j)}. \qquad \text{(D.31)}
\end{aligned}
$$

By (D.13) the first term (D.30) is $O\big((1+\log j)j^{\max(-d,-\zeta)-1}\big)$, while by (D.4) and (D.12), it holds for the second term (D.31) that

$$
\begin{aligned}
&[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}]_{(j)} = O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\Big) \\
&\quad + O\Big(\sum_{i=1}^{j-1}(1+\log i)i^{\max(-d,-\zeta)-1}(j-i)^{-d-1}\Big) + O\Big(\sum_{i=j+1}^{t}(1+\log i)i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\Big)
\end{aligned}
$$

$$=O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right), \tag{D.32}$$

and the proof is analogous to (D.25) besides one log-factor. Furthermore, by (D.6) and (D.32)

$$\begin{aligned}
[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)]_{(j)} &= O\Big((1+\log j)^2 j^{\max(-d,-\zeta)-1}\Big)\\
&+O\Big(\sum_{i=1}^{j-1}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\Big)\\
&+O\Big(\sum_{i=j+1}^{t}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(i-j)^{\max(-d,-\zeta)-1}\Big)\\
&=O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right),
\end{aligned} \tag{D.33}$$

where again the proof is analogous to (D.26) besides one log-factor. From (D.1) and (D.33) it then follows for (D.31) that

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)S'_{d,t}]_{(j)}\\
&=O\Big((1+\log j)^3 j^{\max(-d,-\zeta)-1}\Big)+O\Big(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\Big)\\
&=O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right),
\end{aligned} \tag{D.34}$$

and the proof can be carried out the same way as (D.27). Thus, (D.20) holds.

Turning to (D.21), consider the partial derivatives $\partial\tau_j(\theta,t)/\partial\varphi_{(l)}$, where

$$\frac{\partial\tau_j(\theta,t)}{\partial\varphi_{(l)}} = \nu[(\dot b_1(\varphi_{(l)})\cdots \dot b_t(\varphi_{(l)}))\Xi_t(\theta)S'_{d,t}]_{(j)} \tag{D.35}$$

$$-\nu[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot B'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot B_{\varphi_{(l)},t})\Xi_t(\theta)S'_{d,t}]_{(j)}. \tag{D.36}$$

By assumption 3, the partial derivatives are of order $\dot b_j(\varphi_{(l)}) = \partial b_j(\varphi)/\partial\varphi_{(l)} = O(j^{-\zeta-1})$, so that for the first term (D.35), analogously to (D.12)

$$[(\dot b_1(\varphi_{(l)})\cdots \dot b_t(\varphi_{(l)}))\Xi_t(\theta)]_{(j)} = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right),$$

and, analogously to (D.13)

$$[(\dot b_1(\varphi_{(l)})\cdots \dot b_t(\varphi_{(l)}))\Xi_t(\theta)S_{d,t}]_{(j)} = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right), \tag{D.37}$$

so that (D.37) determines the rate of (D.35). Next, consider (D.36), for which one has by (D.12) and (D.23)

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot B'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot B_{\varphi_{(l)},t})]_{(j)}\\
&=O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\Big)+O\Big(\sum_{i=1}^{j-1}(1+\log i)i^{\max(-d,-\zeta)-1}(j-i)^{-\zeta-1}\Big)
\end{aligned}$$

$$+ O\Big(\sum_{i=j+1}^{t} (1+\log i) i^{\max(-d,-\zeta)-1} (i-j)^{-\zeta-1} \Big) = O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1} \right), \tag{D.38}$$

where the proof is identical to (D.25). By the same proof as for (D.26), by (D.6) and (D.38)

$$\begin{aligned}
&[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t} B_{\varphi,t} + B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta)]_{(j)} \\
&\quad = O\Big((1+\log j)^2 j^{\max(-d,-\zeta)-1} \Big) \\
&\qquad + O\Big(\sum_{i=1}^{j-1} (1+\log i)^2 i^{\max(-d,-\zeta)-1} (j-i)^{\max(-d,-\zeta)-1} \Big) \\
&\qquad + O\Big(\sum_{i=j+1}^{t} (1+\log i)^2 i^{\max(-d,-\zeta)-1} (i-j)^{\max(-d,-\zeta)-1} \Big) \\
&\qquad = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1} \right).
\end{aligned} \tag{D.39}$$

Finally, again by using the same proof as for (D.27), by (D.1) and (D.38)

$$\begin{aligned}
&[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t} B_{\varphi,t} + B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta) S'_{d,t}]_{(j)} \\
&= O\Big((1+\log j)^3 j^{\max(-d,-\zeta)-1} \Big) + O\Big(\sum_{i=j+1}^{t} (1+\log i)^3 i^{\max(-d,-\zeta)-1} (i-j)^{-d-1} \Big) \\
&= O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1} \right).
\end{aligned} \tag{D.40}$$

Together, (D.37) and (D.40) yield (D.21). □

Lemma D.5 (Convergence of the partial derivatives of $\tau_j(\theta,t)$ to $\tau_j(\theta)$)**.** *For the partial derivatives of $\tau_j(\theta,t)$, it holds that*

$$\frac{\partial \tau_j(\theta,t)}{\partial \theta}\Big|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = \sum_{k=t+1}^{\infty} \frac{\partial r_{\tau,j,k}(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = O\left((1+\log t)^5 t^{\max(-d_0-\zeta)-1} \right), \tag{D.41}$$

with $r_{\tau,j,k}(\theta)$ as given in lemma D.3.

Proof of lemma D.5. From (D.18) and below $r_{\tau,j,t+1}(\theta) = -\nu R_{5_{(j)}}$, where

$$\begin{aligned}
R_{5_{(j)}} =& [(b_{t+1}(\varphi) - \pi_{t+1}(d)) \left(R'_2 S'_{d,t} + R_3 s'_t \right)]_{(j)} \\
&+ [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d)) \left(R_2 s'_t + R_1 S'_{d,t} \right)]_{(j)},
\end{aligned}$$

and with $B_{\varphi,t}$ and $S_{d,t}$ as defined in (5), $\beta'_t = (b_t(\varphi) \cdots b_1(\varphi))$, $s'_t = (\pi_t(d) \cdots \pi_1(d))$ as given in lemma D.1, and R_1, R_2, R_3 as stated below (D.16). The partial derivative of $R_{5_{(j)}}$ w.r.t. the l-th entry $\theta_{(l)}$ is thus given by

$$\frac{\partial R_{5_{(j)}}}{\partial \theta_{(l)}} = \left[\frac{\partial (b_{t+1}(\varphi) - \pi_{t+1}(d))}{\partial \theta_{(l)}} \left(R'_2 S'_{d,t} + R_3 s'_t \right) \right]_{(j)} \tag{D.42}$$

$$+\left[\left(\frac{\partial(b_1(\varphi)-\pi_1(d))}{\partial\theta_{(l)}}\cdots\frac{\partial(b_t(\varphi)-\pi_t(d))}{\partial\theta_{(l)}}\right)\left(R_2 s_t'+R_1 S_{d,t}'\right)\right]_{(j)} \tag{D.43}$$

$$+\left[(b_{t+1}(\varphi)-\pi_{t+1}(d))\left(R_2'\frac{\partial S_{d,t}'}{\partial\theta_{(l)}}+R_3\frac{\partial s_t'}{\partial\theta_{(l)}}\right)\right]_{(j)} \tag{D.44}$$

$$+\left[((b_1(\varphi)-\pi_1(d))\cdots(b_t(\varphi)-\pi_t(d)))\left(R_2\frac{\partial s_t'}{\partial\theta_{(l)}}+R_1\frac{\partial S_{d,t}'}{\partial\theta_{(l)}}\right)\right]_{(j)} \tag{D.45}$$

$$+\left[(b_{t+1}(\varphi)-\pi_{t+1}(d))\left(\frac{\partial R_2'}{\partial\theta_{(l)}}S_{d,t}'+\frac{\partial R_3}{\partial\theta_{(l)}}s_t'\right)\right]_{(j)} \tag{D.46}$$

$$+\left[((b_1(\varphi)-\pi_1(d))\cdots(b_t(\varphi)-\pi_t(d)))\left(\frac{\partial R_2}{\partial\theta_{(l)}}s_t'+\frac{\partial R_1}{\partial\theta_{(l)}}S_{d,t}'\right)\right]_{(j)}. \tag{D.47}$$

As noted in the proof of lemma D.4, the partial derivative of $\pi_j(d)$ only adds a log-factor to the convergence rate of $\pi_j(d)$, i.e. $\partial\pi_j(d)/\partial d = O((1+\log j)j^{-d-1})$, see Johansen and Nielsen (2010, lemma B.3), while $\partial b_j(\varphi)/\partial\varphi_{(l)} = O(j^{-\zeta-1})$ by assumption 3. Thus, the convergence rates of (D.42) and (D.43) can be derived analogously to the proof of lemma D.3. This yields that (D.42) is $O((1+\log(t+1))(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1})$, while (D.43) is $O\big((1+\log(t+1))^3(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1}\big)$, and the additional $(1+\log(t+1))$ term stems from $\partial\pi_j(d)/\partial d$. Analogously, the partial derivatives of s_t and $S_{d,t}$ only add a log-factor to the convergence rates as derived in the proof of lemma D.3. Thus, it holds that (D.44) is $O((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1})$, while (D.45) is $O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1}\big)$, and the additional $(1+\log(t+1-j))$ term stems from $\partial s_t'/\partial d$ and $\partial S_{d,t}'/\partial d$. For the last two terms (D.46) and (D.47), note that $R_3 = O(1)$ as shown in (D.17) and below. Since $\beta_t'(\partial\beta_t/\partial\theta_{(l)})$, $s_t'(\partial s_t/\partial\theta_{(l)})$, $s_t's_t$, $(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)\partial(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})'/\partial\theta_{(l)}$, and $(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})(\partial\Xi_t(\theta)/\partial\theta_{(l)})(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})'$ are $O(1)$, it follows that

$$\frac{\partial R_3}{\partial\theta_{(l)}} = -(R_3)^2\frac{\partial}{\partial\theta_{(l)}}\left[(1+\beta_t'\beta_t+\nu+\nu s_t's_t)-(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)(B_{\varphi,t}'\beta_t+\nu S_{d,t}'s_t)\right] = O(1).$$

For the partial derivatives of $R_{2_{(j)}}$, consider

$$\frac{\partial R_{2_{(j)}}}{\partial\theta_{(l)}} = -\frac{\partial R_3}{\partial\theta_{(l)}}\left[(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)\right]_{(j)} - R_3\left[(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\frac{\partial\Xi_t(\theta)}{\partial\theta_{(l)}}\right]_{(j)} \tag{D.48}$$

$$-R_3\Big[(\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}}+\frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t}+\frac{\partial\nu}{\partial\theta_{(l)}}s_t'S_{d,t}+\nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t}+\nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}})\Xi_t(\theta)\Big]_{(j)}, \tag{D.49}$$

where the first term in (D.48) is $O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$ by (D.17) and by $\partial R_3/\partial\theta_{(l)} = O(1)$. For the second term in (D.48), one has $[(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)]_{(j)} = O((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1})$ from (D.17). Together with $\partial\Xi_t(\theta)/\partial\theta_{(l)} = -\Xi_t(\theta)[(\partial/\partial\theta_{(l)})(B_{\varphi,t}'B_{\varphi,t}+\nu S_{d,t}'S_{d,t})]\Xi_t(\theta)$, (D.22) and (D.23), it follows that

$$\left\{(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)\left[\frac{\partial}{\partial\theta_{(l)}}\left(B_{\varphi,t}'B_{\varphi,t}+\nu S_{d,t}'S_{d,t}\right)\right]\right\}_{(j)}$$

$$
\begin{aligned}
=&O\Big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\Big)\\
+&O\Big(\sum_{k=1}^{j-1}(1+\log(t+1-k))(t+1-k)^{\max(-d,-\zeta)-1}\times(1+\log(j-k))(j-k)^{\max(-d,-\zeta)-1}\Big)\\
+&O\Big(\sum_{k=1}^{t-j}(1+\log(t+1-j-k))(t+1-j-k)^{\max(-d,-\zeta)-1}\times(1+\log k)k^{\max(-d,-\zeta)-1}\Big)\\
=&O\left((1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1}\right).
\end{aligned}
$$

Finally, using (D.6), one obtains

$$
\begin{aligned}
&\left\{(\beta_t' B_{\varphi,t}+\nu s_t' S_{d,t})\Xi_t(\theta)\left[\frac{\partial}{\partial\theta_{(l)}}\left(B_{\varphi,t}' B_{\varphi,t}+\nu S_{d,t}' S_{d,t}\right)\right]\Xi_t(\theta)\right\}_{(j)}\\
&=O\left((1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1}\right),
\end{aligned}
\tag{D.50}
$$

which yields the binding rate of convergence for the second term in (D.48). For (D.49)

$$
\begin{aligned}
&\left(\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}}+\frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t}+\frac{\partial\nu}{\partial\theta_{(l)}}s_t' S_{d,t}+\nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t}+\nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\right)_{(j)}\\
&=O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right),
\end{aligned}
$$

by lemma D.1. Hence, using (D.6) yields an upper bound for (D.49)

$$
\begin{aligned}
&\left[\left(\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}}+\frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t}+\frac{\partial\nu}{\partial\theta_{(l)}}s_t' S_{d,t}+\nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t}+\nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\right)\Xi_t(\theta)\right]_{(j)}\\
&=O\left((1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1}\right).
\end{aligned}
\tag{D.51}
$$

Together, the rates of convergence of (D.48) and (D.49) yield

$$
\frac{\partial R_{2_{(j)}}}{\partial\theta_{(l)}}=O\left((1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1}\right). \tag{D.52}
$$

For the partial derivatives of R_1, note that

$$
\frac{\partial R_{1_{(i,j)}}}{\partial\theta_{(l)}}=-\frac{\partial R_{2_{(i)}}}{\partial\theta_{(l)}}\left[(\beta_t' B_{\varphi,t}+\nu s_t' S_{d,t})\Xi_t(\theta)\right]_{(j)}-R_{2_{(i)}}\left[(\beta_t' B_{\varphi,t}+\nu s_t' S_{d,t})\frac{\partial\Xi_t(\theta)}{\partial\theta_{(l)}}\right]_{(j)} \tag{D.53}
$$

$$
-R_{2_{(i)}}\Big[\Big(\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}}+\frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t}+\frac{\partial\nu}{\partial\theta_{(l)}}s_t' S_{d,t}+\nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t}+\nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\Big)\Xi_t(\theta)\Big]_{(j)}. \tag{D.54}
$$

From (D.17) and (D.52), the first term in (D.53) is $O((1+\log(t+1-i))^4(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1})$. Similarly, using (D.50) and the convergence rate of $R_{2_{(i)}}$ as derived in the proof of lemma D.3, the second term in (D.53) is $O((1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1})$. By (D.51), it follows that (D.54) is

$O((1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1})$. Thus

$$\begin{aligned}\frac{\partial R_{1_{(i,j)}}}{\partial\theta_{(l)}} = O\Big((1+\log(t+1-i))^4(t+1-i)^{\max(-d,-\zeta)-1} \\ \times (1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1}\Big).\end{aligned} \tag{D.55}$$

With (D.52) at hand, it follows directly for (D.46) that

$$\left(\frac{\partial R_2'}{\partial\theta_{(l)}}S_{d,t}' + \frac{\partial R_3}{\partial\theta_{(l)}}s_t'\right)_{(j)} = O\left((1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\right).$$

By (D.1) and (D.2), it follows that (D.46) is $O\big((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\big)$. For (D.47), it follows from (D.52) and (D.55) that $\left(\frac{\partial R_2}{\partial\theta_{(l)}}s_t' + \frac{\partial R_1}{\partial\theta_{(l)}}S_{d,t}'\right)_{(i,j)} = O\big((1+\log(t+1-i))^4(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\big)$. Again using (D.1) and (D.2), it thus follows that (D.47) is $O\big((1+\log(t+1))^5(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\big)$. Together, this implies for (D.41) that

$$\begin{aligned}\frac{\partial r_{\tau,j,t+1}(\theta)}{\partial\theta_{(l)}} = O\big((1+\log(t+1))^5(t+1)^{\max(-d,-\zeta)-1} \\ \times (1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\big),\end{aligned}$$

and thus $\frac{\partial}{\partial\theta}\sum_{k=t+1}^{\infty} r_{\tau,j,k}(\theta)\big|_{\theta=\theta_0} = O\left((1+\log t)^5 t^{\max(-d_0-\zeta)-1}\right)$. □

Lemma D.6. *For the truncated score function as given in* (C.2)*, and the untruncated score function as given in* (C.3)*, it holds for all $\theta\in\Theta_3(\kappa_3)$ that*

$$\sqrt{n}\left[\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} - \frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\right] = o_p(1). \tag{D.56}$$

Proof of lemma D.6. Define $h_{1,t} = \sum_{j=1}^{t-1}\frac{\partial\tau_j(\theta,t)}{\partial\theta}\Big|_{\theta=\theta_0}\xi_{t-j}(d_0)$, $\tilde{h}_{1,t} = \sum_{j=1}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$, as well as $h_{2,t} = \sum_{j=0}^{t-1}\tau_j(\theta_0,t)\frac{\partial\xi_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$, and $\tilde{h}_{2,t} = \sum_{j=0}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$. Then plugging (C.2), (C.3) into (D.56) and using (B.11) yields

$$\begin{aligned}&\sqrt{n}\left[\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} - \frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\right] \\ &= \frac{2}{\sqrt{n}}\left[\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{1,t}-h_{1,t}) + \sum_{t=1}^{n}h_{1,t}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)\right] \\ &+ \frac{2}{\sqrt{n}}\left[\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{2,t}-h_{2,t}) + \sum_{t=1}^{n}h_{2,t}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)\right],\end{aligned} \tag{D.57}$$

so that it remains to be shown that all four terms in (D.57) are $o_p(1)$.

For the proofs it will be very useful to note that $\tilde{v}_t(\theta_0)$ adapted to the filtration $\mathcal{F}_t^{\tilde{\xi}} = \sigma(\tilde{\xi}_s, s\leq t)$ is a stationary martingale difference sequence (MDS), as explained in the proof of theorem 4.2. Note in addition that all $\tilde{h}_{1,t}$, $\tilde{h}_{2,t}$ are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, as $\tau_0 = \pi_0 = 1$ are invariant w.r.t. θ.

Starting with the first term of (D.57), by plugging in $h_{1,t}$ and $\tilde{h}_{1,t}$

$$\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{1,t}-h_{1,t})$$

$$= \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=1}^{t-1}\frac{\partial\tau_j(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\left(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0)\right) \tag{D.58}$$

$$+ \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=1}^{t-1}\left(\frac{\partial\tau_j(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}-\frac{\partial\tau_j(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\right)\tilde{\xi}_{t-j}(d_0) \tag{D.59}$$

$$+ \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=t}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0). \tag{D.60}$$

As $\sum_{j=t}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$ is $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, $\tilde{v}_t(\theta_0)\sum_{j=t}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$ is also a MDS. Since $\frac{\partial\tau_j(\theta)}{\partial\theta}|_{\theta=\theta_0} = O((1+\log j)^4 j^{\max(-d_0,-\zeta)-1})$, see lemma D.4, it follows that (D.60) is $o_p(1)$. In (D.59), $\tilde{v}_t(\theta_0)\sum_{j=1}^{t-1}\left(\frac{\partial\tau_j(\theta)}{\partial\theta}|_{\theta=\theta_0}-\frac{\partial\tau_j(\theta,t)}{\partial\theta}|_{\theta=\theta_0}\right)\tilde{\xi}_{t-j}(d_0)$ adapted to $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS, while the sum $\sum_{j=1}^{t-1}\left(\frac{\partial\tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}-\frac{\partial\tau_j(\theta,t)}{\partial\theta}\Big|_{\theta=\theta_0}\right)\tilde{\xi}_{t-j}(d_0) = O_p((1+\log t)^5 t^{\max(-d_0,-\zeta)})$ by lemma D.5. Hence (D.59) is $o_p(1)$. For (D.58), note that by assumption 1

$$\mathrm{E}\left\{\left[\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=1}^{t-1}\frac{\partial\tau_j(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\left(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0)\right)\right]^2\right\}$$

$$\begin{aligned}
&= \mathrm{E}\Bigg[\sum_{s,t=1}^{n}\left(\sum_{j=0}^{\infty}\eta^2_{\min(s,t)-j}\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0)\right)\\
&\qquad\times\sum_{j=0}^{\infty}\epsilon^2_{-j}\Big(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\Big)\\
&\qquad\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\Bigg]
\end{aligned} \tag{D.61}$$

$$\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\Bigg(\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\Bigg)\\
&\qquad\times\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\Bigg]
\end{aligned} \tag{D.62}$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\Bigg[\Bigg(\sum_{j=t}^{\infty}\epsilon_{t-j}^2\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\qquad\times\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\\
&\qquad\times\Bigg(\sum_{j=s}^{\infty}\epsilon_{s-j}^2\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\qquad\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\Bigg].
\end{aligned}
\tag{D.63}
$$

For (D.61), I use $\sum_{j=0}^{\infty}\eta_{\min(s,t)-j}^2\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0) = O_p(|t-s|^{\max(-d_0,-\zeta)-1})$ for $t \neq s$, else $O_p(1)$, see lemma D.2, and $\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\big|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0) = O\big((1+\log(t+j))^6(t+j)^{\max(-d_0,-\zeta)-1}\big)$, see (D.1) together with lemma D.4. This yields the upper bound for (D.61)

$$
\begin{aligned}
K\sum_{t=1}^{n}\Big(&\sum_{s=1,\ s<t}(t-s)^{\max(-d_0,-\zeta)-1}(1+\log t)^6t^{\max(-d_0,-\zeta)-1}+(1+\log t)^{12}t^{2\max(-d_0,-\zeta)-1}\\
&+\sum_{s=t+1}^{n}(s-t)^{\max(-d_0,-\zeta)-1}(1+\log t)^6t^{\max(-d_0,-\zeta)-1}\Big)\\
\le& K\sum_{t=1}^{n}(1+\log t)^6t^{\max(-d_0,-\zeta)-1}=O(1).
\end{aligned}
$$

Similarly, for the second term (D.62), by (D.1) and lemma D.2 it holds that

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon_{\min(s,t)-j}^2\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\right]\\
&\le K\sum_{j=1}^{\min(s,t)-1}(1+\log j)^3j^{-\min(d_0,\zeta)-1}(1+\log(j+|t-s|))^3(j+|t-s|)^{-\min(d_0,\zeta)-1}.
\end{aligned}
$$

Furthermore, by lemma D.4

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\right.\\
&\left.\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\right]\\
&\le K\sum_{j=1}^{\infty}(1+\log(t+j))^6(t+j)^{\max(-d_0,-\zeta)-1}(1+\log(s+j))^6(s+j)^{\max(-d_0,-\zeta)-1},
\end{aligned}
$$

so that by the same proof as for (D.61), it holds that (D.62) is also $O(1)$.

By (D.1) and lemmas D.2 and D.4, the third term (D.63) is bounded from above by

$$\begin{aligned}
&\sum_{s,t=1}^{n} \mathrm{E}\Bigg[\Bigg(\sum_{j=t}^{\infty}\epsilon_{t-j}^2\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\qquad\times\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\\
&\quad\times\Bigg(\sum_{j=s}^{\infty}\epsilon_{s-j}^2\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\qquad\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\Bigg]\\
&\le K\sum_{s,t=1}^{n}(1+\log t)^9 t^{2\max(-d_0,-\zeta)-1}(1+\log s)^9 s^{2\max(-d_0,-\zeta)-1}=O(1).
\end{aligned}$$

As all three terms (D.61) to (D.63) are $O(1)$, it follows directly by the scaling that (D.58) is $o_p(1)$. Now, since (D.58) to (D.60) are $o_p(1)$, the first term in (D.57) is also $o_p(1)$.

Next, consider the third term in (D.57). I plug in $h_{2,t}$ and $\tilde{h}_{2,t}$ which gives

$$\begin{aligned}
&\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{2,t}-h_{2,t})\\
&=\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left(\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}-\frac{\partial\xi_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}\right) &\text{(D.64)}\\
&\quad+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0} &\text{(D.65)}\\
&\quad+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}. &\text{(D.66)}
\end{aligned}$$

For (D.66), note that $(\tilde{v}_t(\theta_0),\mathcal{F}_t^{\tilde{\xi}})$ is a stationary MDS, and the sum $\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\big|_{\theta=\theta_0}$ is $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. Since $\partial\tilde{\xi}_{t-i}(d)/\partial\theta$ is $O_p(1)$ for all $d>d_0-1/2$, it follows by lemma D.2 that $\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\big|_{\theta=\theta_0}=O_p((1+\log t)t^{\max(-d_0,-\zeta)})$, and thus (D.66) is $o_p(1)$.

For (D.65), note that $\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\big|_{\theta=\theta_0}$ together with $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS. Furthermore, by lemma D.3, it holds that $\tau_j(\theta_0)-\tau_j(\theta_0,t)=O((1+\log t)^2 t^{\max(-d_0,-\zeta)-1})$. Since the partial derivatives of $\tilde{\xi}_t(d)$ are bounded in probability, $\sum_{j=0}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\big|_{\theta=\theta_0}=O_p((1+\log t)^2 t^{\max(-d_0,-\zeta)})$. Therefore, (D.65) is $o_p(1)$.

For (D.64), I use $\frac{\partial\pi_j(d-d_0)}{\partial d}\big|_{d=d_0}=-j^{-1}$ as shown by Robinson (2006, pp. 135-136) and Hualde and Robinson (2011, p. 3170). Thus, the partial derivative in (D.64) w.r.t. d is

$$\left.\frac{\partial\tilde{\xi}_t(\theta)}{\partial d}\right|_{\theta=\theta_0} - \left.\frac{\partial\xi_t(\theta)}{\partial d}\right|_{\theta=\theta_0} = -\sum_{j=t}^{\infty} j^{-1}\eta_{t-j} + \sum_{j=0}^{\infty}\epsilon_{-j}\sum_{k=0}^{j}\left.\frac{\partial\pi_{t+j-k}(d)}{\partial d}\right|_{\theta=\theta_0} a_k(\varphi_0). \tag{D.67}$$

As the partial derivatives w.r.t. all other entries in θ are zero, by assumption 1 it is sufficient to consider

$$\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left(\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial d}\right|_{\theta=\theta_0} - \left.\frac{\partial\xi_{t-j}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]^2\right\}\\
&= \sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\eta^2_{\min(s,t)-j}\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0)\right]\\
&\qquad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\frac{\tau_k(\theta_0,t)}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\frac{\tau_k(\theta_0,s)}{s+j-k}\right)\right.\\
&\qquad\quad+\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\left.\frac{\partial\pi_{j+t-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\left.\frac{\partial\pi_{j+s-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]
\end{aligned}\tag{D.68}$$

$$\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\right]\\
&\qquad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\frac{\tau_k(\theta_0,t)}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\frac{\tau_k(\theta_0,s)}{s+j-k}\right)\right.\\
&\qquad\quad+\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\left.\frac{\partial\pi_{j+t-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\left.\frac{\partial\pi_{j+s-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]
\end{aligned}\tag{D.69}$$

$$\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\left\{\left[\sum_{j=t}^{\infty}\eta^2_{t-j}\tau_j(\theta_0)\sum_{k=0}^{t-1}\frac{-\tau_k(\theta_0,t)}{j-k}+\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\right.\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]\\
&\qquad\left[\sum_{j=s}^{\infty}\eta^2_{s-j}\tau_j(\theta_0)\sum_{k=0}^{s-1}\frac{-\tau_k(\theta_0,s)}{j-k}+\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\qquad\qquad\left.\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]\right\}.
\end{aligned}\tag{D.70}$$

For (D.68), note the first expectation is $\sigma^2_{\eta,0}\sum_{j=0}^{\min(s,t)-1}\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0) = O(|t-s|^{\max(-d_0,-\zeta)-1})$ for all $t \neq s$, and $O(1)$ for $t = s$, see lemma D.2. For the other terms in (D.68), it holds that $\mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\frac{1}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\frac{1}{s+j-k}\right)\right] \le K\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}$, together with

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\frac{\partial\pi_{j+t-k-l}(d)}{\partial d}\Big|_{\theta=\theta_0}\right)\right.\\
&\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\frac{\partial\pi_{j+s-k-l}(d)}{\partial d}\Big|_{\theta=\theta_0}\right)\right]\\
&\le K\sum_{j=0}^{\infty}(1+\log(t+j))^4(t+j)^{\max(-d_0,-\zeta)-1}(1+\log(s+j))^4(s+j)^{\max(-d_0,-\zeta)-1},
\end{aligned}
$$

by lemma D.2. It follows that (D.68) is bounded from above by

$$
\begin{aligned}
&K\sum_{t=1}^{n}\left[\sum_{s=1,\ s<t}(t-s)^{\max(-d_0,-\zeta)-1}\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}\right.\\
&\qquad+\sum_{j=0}^{\infty}(1+\log(t+j))^4(t+j)^{-2}\\
&\qquad\left.+\sum_{s=t+1}^{n}(s-t)^{\max(-d_0,-\zeta)-1}\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}\right]\\
&\le K\sum_{t=1}^{n}\left[(1+\log t)t^{-1+\kappa}\right]\le Kn^{\kappa},
\end{aligned}
$$

for $0<\kappa<1/2$, since $\sum_{j=0}^{\infty}(s+j)^{-2} = O(s^{-1})$, see Chan and Palma (1998, lemma 3.2), and, as the logarithm is dominated by its powers, $\sum_{j=0}^{\infty}(1+\log(s+j))^2(s+j)^{-2} = O(s^{-1+\kappa})$ for all $0<\kappa<1/2$. For (D.69), by lemmas D.1 and D.2, the first expectation is bounded by

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\qquad\left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\right] = O(|t-s|^{\max(-d_0,-\zeta)-1}),
\end{aligned}
$$

for all $t\neq s$, and is $O(1)$ for $t=s$. Hence, by the same proof as for (D.68) the second term (D.69) is also $O(n^{\kappa})$, $0<\kappa<1/2$. For the third term (D.70) one has by lemma D.2

$$
\begin{aligned}
&\sum_{s,t=1}^{n}\mathrm{E}\left\{\left[\sum_{j=t}^{\infty}\eta^2_{t-j}\tau_j(\theta_0)\sum_{k=0}^{t-1}\frac{-\tau_k(\theta_0,t)}{j-k}+\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\right.\\
&\times\left.\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\frac{\partial\pi_{j-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right)\right]\left[\sum_{j=s}^{\infty}\eta^2_{s-j}\tau_j(\theta_0)\sum_{k=0}^{s-1}\frac{-\tau_k(\theta_0,s)}{j-k}\right.
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{j=s}^{\infty}\epsilon_{s-j}^2\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\Bigg]\Bigg\}\\
&=\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^3 j^{\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^3 j^{\max(-d_0,-\zeta)-2}\right)\right)\\
&+\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^7 j^{2\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^7 j^{2\max(-d_0,-\zeta)-2}\right)\right)\\
&+\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^3 j^{\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^7 j^{2\max(-d_0,-\zeta)-2}\right)\right)\\
&+\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^7 j^{2\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^3 j^{\max(-d_0,-\zeta)-2}\right)\right),
\end{aligned}
$$

which is $O(1)$, and thus all terms (D.68) to (D.70) are $O(n^{\kappa})$. As (D.64) is appropriately scaled, it follows that (D.64) is $o_p(1)$ and thus the third term in (D.57) is $o_p(1)$.

Next, consider the second term in (D.57) that can be decomposed into

$$
\begin{aligned}
\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{1,t}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)&=\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{1,t}\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0))\tau_j(\theta_0,t)\\
&+\frac{2}{\sqrt{n}}\sum_{t=0}^{n}h_{1,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\tilde{\xi}_{t-j}(d_0)+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{1,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0).
\end{aligned}
\tag{D.71}
$$

For the first term in (D.71), note that by assumption 1

$$
\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^{n}h_{1,t}\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0))\tau_j(\theta_0,t)\right]^2\right\}\\
&=\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\left.\frac{\partial\tau_j(\theta,\min(s,t))}{\partial\theta}\right|_{\theta=\theta_0}\left.\frac{\partial\tau_{j+|t-s|}(\theta,\max(s,t))}{\partial\theta'}\right|_{\theta=\theta_0}\eta^2_{\min(s,t)-j}\right]\\
&\qquad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\right.\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\right]
\end{aligned}
\tag{D.72}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1} \epsilon^2_{\min(s,t)-j} \left(\sum_{k=0}^{j} \left.\frac{\partial \tau_k(\theta,\min(s,t))}{\partial \theta}\right|_{\theta=\theta_0} \sum_{l=0}^{j-k} \pi_l(d_0) a_{j-k-l}(\varphi_0)\right)\right.\\
&\qquad \left.\times\left(\sum_{k=0}^{j+|t-s|} \left.\frac{\partial \tau_k(\theta,\max(s,t))}{\partial \theta'}\right|_{\theta=\theta_0} \sum_{l=0}^{j+|t-s|-k} \pi_l(d_0) a_{j+|t-s|-k-l}(\varphi_0)\right)\right]\\
&\qquad \times \mathrm{E}\left[\sum_{j=0}^{\infty} \epsilon^2_{-j} \left(\sum_{k=0}^{t-1} \tau_k(\theta_0,t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0)\right)\right.\\
&\qquad\qquad \left.\times\left(\sum_{k=0}^{s-1} \tau_k(\theta_0,s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0)\right)\right]
\end{aligned}
\tag{D.73}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\left[\left(\sum_{j=t}^{\infty} \epsilon^2_{t-j} \left(\sum_{k=0}^{t-1} \left.\frac{\partial \tau_k(\theta,t)}{\partial \theta}\right|_{\theta=\theta_0} \sum_{l=0}^{\min(j-k,t-1)} \pi_l(d_0) a_{j-k-l}(\varphi_0)\right)\right.\right.\\
&\qquad\qquad \left.\left.\times\left(\sum_{k=0}^{t-1} \tau_k(\theta_0,t) \sum_{l=0}^{j-t} a_l(\varphi_0) \pi_{j-k-l}(d_0)\right)\right)\right.\\
&\qquad \times \sum_{j=s}^{\infty} \epsilon^2_{s-j} \left(\sum_{k=0}^{s-1} \left.\frac{\partial \tau_k(\theta,s)}{\partial \theta'}\right|_{\theta=\theta_0} \sum_{l=0}^{\min(j-k,s-1)} \pi_l(d_0) a_{j-k-l}(\varphi_0)\right)\\
&\qquad\qquad \left.\times\left(\sum_{k=0}^{s-1} \tau_k(\theta_0,s) \sum_{l=0}^{j-s} a_l(\varphi_0) \pi_{j-k-l}(d_0)\right)\right)\Bigg].
\end{aligned}
\tag{D.74}
$$

For (D.72), one has for all $t \neq s$

$$
\mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1} \left.\frac{\partial \tau_j(\theta,\min(s,t))}{\partial \theta}\right|_{\theta=\theta_0} \left.\frac{\partial \tau_{j+|t-s|}(\theta,\max(s,t))}{\partial \theta'}\right|_{\theta=\theta_0} \eta^2_{\min(s,t)-j}\right] = O(|t-s|^{\max(-d_0,-\zeta)-1}),
$$

by lemma D.4, and $O(1)$ for $t = s$. Furthermore, for (D.73), the first term is bounded by

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1} \epsilon^2_{\min(s,t)-j} \left(\sum_{k=0}^{j} \left.\frac{\partial \tau_k(\theta,\min(s,t))}{\partial \theta}\right|_{\theta=\theta_0} \sum_{l=0}^{j-k} \pi_l(d_0) a_{j-k-l}(\varphi_0)\right)\right.\\
&\qquad \left.\left(\sum_{k=0}^{j+|t-s|} \left.\frac{\partial \tau_k(\theta,\max(s,t))}{\partial \theta'}\right|_{\theta=\theta_0} \sum_{l=0}^{j+|t-s|-k} \pi_l(d_0) a_{j+|t-s|-k-l}(\varphi_0)\right)\right]\\
&\qquad = O(|t-s|^{\max(-d_0,-\zeta)-1}),
\end{aligned}
$$

by lemmas D.1 and D.4 for $t \neq s$, and $O(1)$ otherwise. In addition, for both (D.72) and (D.73), by lemmas D.1 and D.2 the other remaining term is bounded by

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\infty} \epsilon^2_{-j} \left(\sum_{k=0}^{t-1} \tau_k(\theta_0,t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0)\right) \left(\sum_{k=0}^{s-1} \tau_k(\theta_0,s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0)\right)\right]\\
&=O\left((1+\log t)^3 t^{\max(-d_0,-\zeta)} (1+\log s)^3 s^{\max(-d_0,-\zeta)-1}\right).
\end{aligned}
$$

Consequently, (D.72) and (D.73) are $\sum_{s,t=1}^n O\big((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1}|t-s|^{\max(-d_0,-\zeta)-1}\big) = O(1)$. Finally, by lemmas D.1, D.2, and D.4, (D.74) is

$$\begin{aligned}
&\sum_{s,t=1}^n \mathrm{E}\left[\left(\sum_{j=t}^\infty \epsilon_{t-j}^2 O\left((1+\log j)^9 j^{2\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^\infty \epsilon_{s-j}^2 O\left((1+\log j)^9 j^{2\max(-d_0,-\zeta)-2}\right)\right)\right] \\
&\quad = \sum_{s,t=1}^n (1+\log t)^9 t^{2\max(-d_0,-\zeta)-1}(1+\log s)^9 s^{2\max(-d_0,-\zeta)-1} = O(1).
\end{aligned}$$

Thus, the first term in (D.71) is $o_p(1)$. For the second term in (D.71), note that by lemma D.3, $\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t)) \le K\sum_{j=1}^{t-1}\sum_{k=t+1}^\infty (1+\log k)^2(1+\log(k-j))^2 k^{\max(-d_0,-\zeta)-1}(k-j)^{\max(-d_0,-\zeta)-1} \le K\sum_{j=1}^{t-1}(1+\log t)^2 t^{\max(-d_0,-\zeta)-1}(1+\log(t-j))^2(t-j)^{\max(-d_0,-\zeta)} \le K(1+\log t)^2 t^{-1}\sum_{j=1}^{t-1} j^{\max(-d_0,-\zeta)}(t-j)^{\max(-d_0,-\zeta)}(1+\log(t-j))^2 \le K(1+\log t)^5 t^{\max(-d_0,-\zeta)-1}$, and thus $\frac{2}{\sqrt{n}}\sum_{t=1}^n h_{1,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\tilde{\xi}_{t-j}(d_0) = o_p(1)$. For the third term in (D.71)

$$\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^n h_{1,t}\sum_{j=t}^\infty \tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0)\right]^2\right\} \\
&= \sum_{s,t=1}^n \mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\eta^2_{\min(s,t)-j}\left.\frac{\partial\tau_j(\theta,\min(s,t))}{\partial\theta}\right|_{\theta=\theta_0}\left.\frac{\partial\tau_{j+|t-s|}(\theta,\max(s,t))}{\partial\theta'}\right|_{\theta=\theta_0}\right] \\
&\quad\times \mathrm{E}\left[\sum_{j=0}^\infty \eta_{-j}^2\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0) + \sum_{j=0}^\infty \epsilon_{-j}^2\left(\sum_{k=0}^j \tau_{t+k}(\theta_0)\sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right. \qquad (\text{D.75})\\
&\quad\left.\times\left(\sum_{k=0}^j \tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right] \\
&+\sum_{s,t=1}^n \mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^j\left.\frac{\partial\tau_k(\theta,\min(s,t))}{\partial\theta}\right|_{\theta=\theta_0}\sum_{l=0}^{j-k}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\right. \\
&\quad\left.\times\left(\sum_{k=0}^{j+|t-s|}\left.\frac{\partial\tau_k(\theta,\max(s,t))}{\partial\theta'}\right|_{\theta=\theta_0}\sum_{l=0}^{j+|t-s|-k}\pi_l(d_0)a_{j+|t-s|-k-l}(\varphi_0)\right)\right] \qquad (\text{D.76})\\
&\quad\times \mathrm{E}\left[\sum_{j=0}^\infty \eta_{-j}^2\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0) + \sum_{j=0}^\infty \epsilon_{-j}^2\left(\sum_{k=0}^j \tau_{t+k}(\theta_0)\sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right. \\
&\quad\left.\times\left(\sum_{k=0}^j \tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\left[\left(\left(\sum_{j=t}^{\infty}\epsilon_{t-j}^{2}\left(\sum_{k=0}^{t-1}\left.\frac{\partial\tau_k(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,t-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\right.\right.\right.\\
&\times\left.\left.\left(\sum_{k=0}^{j-t}\tau_{j+k}(\theta_0)\sum_{l=0}^{j-t-k}a_l(\varphi_0)\pi_{j-t-k-l}(d_0)\right)\right)\right.\\
&\times\left(\sum_{j=s}^{\infty}\epsilon_{s-j}^{2}\left(\sum_{k=0}^{s-1}\left.\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,s-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\right.\\
&\times\left.\left.\left(\sum_{k=0}^{j-s}\tau_{j+k}(\theta_0)\sum_{l=0}^{j-s-k}a_l(\varphi_0)\pi_{j-s-k-l}(d_0)\right)\right)\right].
\end{aligned}
\tag{D.77}
$$

For (D.75) and (D.76), it holds that

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon_{-j}^{2}\left(\sum_{k=0}^{j}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\left(\sum_{k=0}^{j}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right]\\
&\quad=O((1+\log t)^3t^{\max(-d_0,-\zeta)}(1+\log s)^3s^{\max(-d_0,-\zeta)-1}),
\end{aligned}
$$

and $\mathrm{E}\left[\sum_{j=0}^{\infty}\eta_{-j}^{2}\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)\right]=O((1+\log t)t^{-\min(d_0,\zeta)}(1+\log s)s^{-\min(d_0,\zeta)-1})$. Thus, analogously to (D.72) and (D.73), expressions (D.75) and (D.76) are $O(1)$. Also analogously to (D.74), by lemmas D.1, D.2, and D.4, (D.77) is bounded from above by

$$
\begin{aligned}
&\sum_{s,t=1}^{n}\mathrm{E}\left[\left(\sum_{j=t}^{\infty}\epsilon_{t-j}^{2}O\left((1+\log j)^6j^{\max(-d_0,-\zeta)-1}(1+\log(j-t))^3(j-t)^{\max(-d_0,-\zeta)-1}\right)\right)\right.\\
&\left.\left(\sum_{j=s}^{\infty}\epsilon_{s-j}^{2}O\left((1+\log j)^6j^{\max(-d_0,-\zeta)-1}(1+\log(j-s))^3(j-s)^{\max(-d_0,-\zeta)-1}\right)\right)\right]=O(1).
\end{aligned}
$$

Therefore, also the third term in (D.71) is $o_p(1)$. It follows that the second term in (D.57) is $o_p(1)$. Finally, consider the last term in (D.57)

$$
\begin{aligned}
\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{2,t}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)&=\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{2,t}\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0))\tau_j(\theta_0,t)\\
&+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{2,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\tilde{\xi}_{t-j}(d_0)+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{2,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0).
\end{aligned}
\tag{D.78}
$$

For the first term in (D.78), by assumption 1 it holds that

$$
\mathrm{E}\left\{\left[\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left.\frac{\partial\xi_{t-j}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0))\tau_j(\theta_0,t)\right]^2\right\}
$$

$$
\begin{aligned}
&= \sum_{s,t=1}^{n} \mathrm{E}\Bigg[\sum_{j=1}^{\min(s,t)-1} \eta^2_{\min(s,t)-j} \left(\sum_{k=1}^{j} \frac{1}{k} \tau_{j-k}(\theta_0, \min(s,t)) \right) \\
&\qquad \times \left(\sum_{k=1}^{j+|t-s|} \frac{1}{k} \tau_{j+|t-s|-k}(\theta_0, \max(s,t)) \right) \Bigg] \\
&\qquad \times \mathrm{E}\Bigg[\sum_{j=0}^{\infty} \epsilon^2_{-j} \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0) \right) \\
&\qquad \times \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0) \right) \Bigg]
\end{aligned} \tag{D.79}
$$

$$
\begin{aligned}
&+ \sum_{s,t=1}^{n} \mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1} \epsilon^2_{\min(s,t)-j} \left(\sum_{k=0}^{j} \tau_k(\theta_0, \min(s,t)) \sum_{l=0}^{j-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0) \right) \\
&\qquad \times \left(\sum_{k=0}^{j+|t-s|} \tau_k(\theta_0, \max(s,t)) \sum_{l=0}^{j+|t-s|-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j+|t-s|-k-l}(\varphi_0) \right) \Bigg] \\
&\qquad \times \mathrm{E}\Bigg[\sum_{j=0}^{\infty} \epsilon^2_{-j} \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0) \right) \\
&\qquad \times \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0) \right) \Bigg]
\end{aligned} \tag{D.80}
$$

$$
\begin{aligned}
&+ \sum_{s,t=1}^{n} \mathrm{E}\Bigg[\Bigg(\sum_{j=t}^{\infty} \epsilon^2_{t-j} \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{t-1-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0) \right) \\
&\qquad \times \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j-t} a_l(\varphi_0) \pi_{j-k-l}(d_0) \right) \Bigg) \\
&\qquad \times \Bigg(\sum_{j=s}^{\infty} \epsilon^2_{s-j} \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{s-1-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0) \right) \\
&\qquad \times \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j-s} a_l(\varphi_0) \pi_{j-k-l}(d_0) \right) \Bigg) \Bigg],
\end{aligned} \tag{D.81}
$$

while all other partial derivatives of $\xi_{t-j}(d)$ (i.e. those w.r.t. all other entries except d) are zero. By lemma D.2, the first term in (D.79) is

$$
\mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1} \eta^2_{\min(s,t)-j} \left(\sum_{k=1}^{j} \frac{1}{k} \tau_{j-k}(\theta_0, \min(s,t)) \right) \sum_{k=1}^{j+|t-s|} \frac{1}{k} \tau_{j+|t-s|-k}(\theta_0, \max(s,t)) \right] = O(|t-s|^{-1}),
$$

for $t \neq s$, and $O(1)$ otherwise. In addition, by lemmas D.1 and D.2 it holds that the first term of (D.80) is

$$
\mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1} \epsilon^2_{\min(s,t)-j} \left(\sum_{k=0}^{j} \tau_k(\theta_0, \min(s,t)) \sum_{l=0}^{j-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0) \right)
$$

$$\times \left(\sum_{k=0}^{j+|t-s|} \tau_k(\theta_0, \max(s,t)) \sum_{l=0}^{j+|t-s|-k} \left. \frac{\partial \pi_l(d)}{\partial d} \right|_{\theta=\theta_0} a_{j+|t-s|-k-l}(\varphi_0) \right) \Bigg]$$

$$= O(|t-s|^{\max(-d_0,-\zeta)-1}), \tag{D.82}$$

for $t \neq s$, and $O(1)$ otherwise. The second term in (D.79) and (D.80) is

$$\mathrm{E}\left[\sum_{j=0}^{\infty} \epsilon_{-j}^2 \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0) \right) \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0) \right) \right]$$
$$= O((1+\log t)^3 t^{\max(-d_0,-\zeta)} (1+\log s)^3 s^{\max(-d_0,-\zeta)-1})$$

Thus, analogously to (D.72), (D.73), (D.75) and (D.76), it holds that (D.79) and (D.80) are $O(1)$. Finally, (D.81) is bounded from above by

$$\sum_{s,t=1}^{n} \mathrm{E}\left[\left(\sum_{j=t}^{\infty} \epsilon_{t-j}^2 O\left((1+\log j)^4 j^{\max(-d_0,-\zeta)-1}\right) O\left((1+\log j)^3 j^{\max(-d_0,-\zeta)-1}\right) \right) \right.$$
$$\left. \times \left(\sum_{j=s}^{\infty} \epsilon_{s-j}^2 O\left((1+\log j)^4 j^{\max(-d_0,-\zeta)-1}\right) O\left((1+\log j)^3 j^{\max(-d_0,-\zeta)-1}\right) \right) \right]$$
$$= \sum_{s,t=1}^{n} O((1+\log t)^7 t^{2\max(-d_0,-\zeta)-1} (1+\log s)^7 s^{\max(-d_0,-\zeta)-1}) = O(1).$$

Hence, the first term in (D.78) is $o_p(1)$. For the second term in (D.78), by lemma D.3, $\sum_{j=1}^{t-1}(\tau_j(\theta_0) - \tau_j(\theta_0, t)) = O((1+\log t)^5 t^{\max(-d_0,-\zeta)-1})$ as already noted for the second term in (D.71), and thus $\frac{2}{\sqrt{n}} \sum_{t=1}^{n} h_{2,t} \sum_{j=1}^{t-1}(\tau_j(\theta_0) - \tau_j(\theta_0, t)) \tilde{\xi}_{t-j}(d_0) = o_p(1)$. For the third term in (D.71)

$$\mathrm{E}\left\{ \left[\sum_{t=1}^{n} h_{2,t} \sum_{j=t}^{\infty} \tau_j(\theta_0) \tilde{\xi}_{t-j}(d_0) \right]^2 \right\}$$
$$= \sum_{s,t=1}^{n} \mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1} \eta_{\min(s,t)-j}^2 \left(\sum_{k=1}^{j} \frac{1}{k} \tau_{j-k}(\theta_0, \min(s,t)) \right) \right.$$
$$\left. \times \left(\sum_{k=1}^{j+|t-s|} \frac{1}{k} \tau_{j+|t-s|-k}(\theta_0, \max(s,t)) \right) \right]$$
$$\times \mathrm{E}\left[\sum_{j=0}^{\infty} \eta_{-j}^2 \tau_{t+j}(\theta_0) \tau_{s+j}(\theta_0) \right. \tag{D.83}$$
$$+ \sum_{j=0}^{\infty} \epsilon_{-j}^2 \left(\sum_{k=0}^{j} \tau_{t+k}(\theta_0) \sum_{l=0}^{j-k} a_l(\varphi_0) \pi_{j-k-l}(d_0) \right)$$
$$\left. \times \left(\sum_{k=0}^{j} \tau_{s+k}(\theta_0) \sum_{l=0}^{j-k} a_l(\varphi_0) \pi_{j-k-l}(d_0) \right) \right]$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1} \epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j} \tau_k(\theta_0,\min(s,t)) \sum_{l=0}^{j-k} \left.\frac{\partial \pi_l(d)}{\partial d}\right|_{\theta=\theta_0} a_{j-k-l}(\varphi_0)\right)\\
&\qquad \times\left(\sum_{k=0}^{j+|t-s|} \tau_k(\theta_0,\max(s,t)) \sum_{l=0}^{j+|t-s|-k} \left.\frac{\partial \pi_l(d)}{\partial d}\right|_{\theta=\theta_0} a_{j+|t-s|-k-l}(\varphi_0)\right)\\
&\times \mathrm{E}\Bigg[\sum_{j=0}^{\infty} \eta^2_{-j}\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0) + \sum_{j=0}^{\infty} \epsilon^2_{-j}\left(\sum_{k=0}^{j} \tau_{t+k}(\theta_0)\sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad \times\left(\sum_{k=0}^{j} \tau_{s+k}(\theta_0)\sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg]
\end{aligned}
\tag{D.84}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\Bigg[\left(\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{t-1} \tau_k(\theta_0,t)\sum_{l=0}^{t-k-1} \left.\frac{\partial \pi_l(d)}{\partial d}\right|_{\theta=\theta_0} a_{j-k-l}(\varphi_0)\right)\right.\\
&\qquad \left.\times\left(\sum_{k=0}^{j-t} \tau_{t+k}(\theta_0)\sum_{l=0}^{j-t-k} a_l(\varphi_0)\pi_{j-t-k-l}(d_0)\right)\right)\\
&\quad \times\left(\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{s-1} \tau_k(\theta_0,s)\sum_{l=0}^{s-k-1} \left.\frac{\partial \pi_l(d)}{\partial d}\right|_{\theta=\theta_0} a_{j-k-l}(\varphi_0)\right)\right.\\
&\qquad \left.\times\left(\sum_{k=0}^{j-s} \tau_{s+k}(\theta_0)\sum_{l=0}^{j-s-k} a_l(\varphi_0)\pi_{j-s-k-l}(d_0)\right)\right)\Bigg].
\end{aligned}
\tag{D.85}
$$

As noted above, the first expected value in (D.83) is $O(|t-s|^{-1})$ for $s \neq t$, else $O(1)$. For the second term (D.84), note that the first expectation is $O(|t-s|^{\max(-d_0,-\zeta)-1})$ for $s \neq t$, else $O(1)$, see (D.82). Furthermore, as shown below (D.77), the second expectation in (D.83) and (D.84) is $O((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1})$, and thus (D.83) and (D.84) are $O(1)$. Finally, the last term (D.85) is $O(1)$, and the proof is identical to (D.81). Thus, also the third term in (D.78) is $o_p(1)$. This shows that (D.57) is $o_p(1)$ and completes the proof. □

Lemma D.7 (Boundedness of third partial derivatives of $Q(y,\theta)$)**.** *For $d \in D_3$ as defined in the proof of theorem 4.1, $\nu \in \Sigma_\nu$ as defined in section 4, and $\varphi \in N_\delta(\varphi_0)$ as defined in assumptions 2 and 4, the third partial derivatives of the objective function* (16) *are uniformly dominated by some random variable B_n that is $O_p(1)$,*

$$
B_n = \sup_{d\in D_3,\nu\in\Sigma_\nu,\varphi\in N_\delta(\varphi_0)} \left|\frac{\partial^3 Q(y,\theta)}{\partial \theta^{(3)}}\right| = O_p(1).
$$

Proof of lemma D.7. The third partial derivatives are

$$
\begin{aligned}
\frac{\partial^3 Q(y,\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}\partial\theta_{(m)}} &= \frac{2}{n}\sum_{t=1}^{n} \frac{\partial^2 v_t(\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\frac{\partial v_t(\theta)}{\partial\theta_{(m)}} + \frac{2}{n}\sum_{t=1}^{n} \frac{\partial v_t(\theta)}{\partial\theta_{(k)}}\frac{\partial^2 v_t(\theta)}{\partial\theta_{(l)}\partial\theta_{(m)}}\\
&\quad + \frac{2}{n}\sum_{t=1}^{n} \frac{\partial^2 v_t(\theta)}{\partial\theta_{(k)}\partial\theta_{(m)}}\frac{\partial v_t(\theta)}{\partial\theta_{(l)}} + \frac{2}{n}\sum_{t=1}^{n} v_t(\theta)\frac{\partial^3 v_t(\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}\partial\theta_{(m)}},
\end{aligned}
$$

for $k, l, m = 1, ..., q+2$, with $\partial v_t(\theta)/(\partial\theta_{(k)})$ in (B.11),

$$\frac{\partial^2 v_t(\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}} = \sum_{j=0}^{t-1}\Bigg[\frac{\partial^2\tau_j(\theta,t)}{\partial\theta_{(k)}\partial\theta_{(l)}}\xi_{t-j}(d) + \frac{\partial\tau_j(\theta,t)}{\partial\theta_{(k)}}\frac{\partial\xi_{t-j}(d)}{\partial\theta_{(l)}} + \frac{\partial\tau_j(\theta,t)}{\partial\theta_{(l)}}\frac{\partial\xi_{t-j}(d)}{\partial\theta_{(k)}} + \tau_j(\theta,t)\frac{\partial^2\xi_{t-j}(d)}{\partial\theta_{(k)}\partial\theta_{(l)}}\Bigg],$$

$$\begin{aligned}\frac{\partial^3 v_t(\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}\partial\theta_{(m)}} = \sum_{j=0}^{t-1}\Bigg[&\frac{\partial^3\tau_j(\theta,t)}{\partial\theta_{(k)}\partial\theta_{(l)}\partial\theta_{(m)}}\xi_{t-j}(d) + \frac{\partial^2\tau_j(\theta,t)}{\partial\theta_{(k)}\partial\theta_{(l)}}\frac{\partial\xi_{t-j}(d)}{\partial\theta_{(m)}} \\ &+ \frac{\partial^2\tau_j(\theta,t)}{\partial\theta_{(k)}\partial\theta_{(m)}}\frac{\partial\xi_{t-j}(d)}{\partial\theta_{(l)}} + \frac{\partial\tau_j(\theta,t)}{\partial\theta_{(k)}}\frac{\partial^2\xi_{t-j}(d)}{\partial\theta_{(l)}\partial\theta_{(m)}} \\ &+ \frac{\partial^2\tau_j(\theta,t)}{\partial\theta_{(l)}\partial\theta_{(m)}}\frac{\partial\xi_{t-j}(d)}{\partial\theta_{(k)}} + \frac{\partial\tau_j(\theta,t)}{\partial\theta_{(l)}}\frac{\partial^2\xi_{t-j}(d)}{\partial\theta_{(k)}\partial\theta_{(m)}} \\ &+ \frac{\partial\tau_j(\theta,t)}{\partial\theta_{(m)}}\frac{\partial^2\xi_{t-j}(d)}{\partial\theta_{(k)}\partial\theta_{(l)}} + \tau_j(\theta,t)\frac{\partial^3\xi_{t-j}(d)}{\partial\theta_{(k)}\partial\theta_{(l)}\partial\theta_{(m)}}\Bigg].\end{aligned}$$

Boundedness in probability of the third partial derivatives then follows from (B.12) upon verification of the absolute summability condition of the partial derivatives of $\tau_j(\theta,t)$, as the derivatives of $\xi_{t-j}(d)$ are zero for all entries of θ except for d, and as those derivatives w.r.t. d are contained in (B.12). As can be seen from lemma D.4 and its proof, the second and third partial derivatives of $\tau_j(\theta,t)$ depend on the coefficients $b_j(\varphi)$ and $\pi_j(d)$, the matrices $\Xi_t(\theta)$, $S_{d,t}$, $B_{\varphi,t}$, and their partial derivatives. While the convergence rates of the former are given in lemma D.1, those for the first partial derivatives are contained in the proof of lemma D.4. In addition, we require $\frac{\partial^2\pi_j(d)}{\partial d^2} = \ddot{\pi}_j(d) = O((1+\log j)^2 j^{-d-1})$ and $\frac{\partial^3\pi_j(d)}{\partial d^3} = \dddot{\pi}_j(d) = O((1+\log j)^3 j^{-d-1})$ (see Johansen and Nielsen; 2010, lemma B.3), $\frac{\partial^2 b_j(\varphi)}{\partial\varphi_{(k)}\partial\varphi_{(l)}} = \ddot{b}_j(\varphi_{(k,l)}) = O(j^{-\zeta-1})$ and $\frac{\partial^3 b_j(\varphi)}{\partial\varphi_{(k)}\partial\varphi_{(l)}\partial\varphi_{(m)}} = \dddot{b}_j(\varphi_{(k,l,m)}) = O(j^{-\zeta-1})$ for $k, l, m = 1, ..., q$ by assumption 4. Based on them, the convergence rates of the following matrices are obtained

$$(\ddot{S}_{d,t})_{(i,j)} = \left(\frac{\partial^2 S_{d,t}}{\partial d^2}\right)_{(i,j)} = \begin{cases}\ddot{\pi}_{j-i}(d) = O((1+\log(j-i))^2(j-i)^{-d-1}) & \text{if } i<j,\\ 0 & \text{else,}\end{cases}$$

$$(\dddot{S}_{d,t})_{(i,j)} = \left(\frac{\partial^3 S_{d,t}}{\partial d^3}\right)_{(i,j)} = \begin{cases}\dddot{\pi}_{j-i}(d) = O((1+\log(j-i))^3(j-i)^{-d-1}) & \text{if } i<j,\\ 0 & \text{else,}\end{cases}$$

$$(\ddot{S}'_{d,t}S_{d,t})_{(i,j)} = \begin{cases}\sum_{k=1}^{i-1}\ddot{\pi}_k(d)\pi_{k+j-i}(d) = O((1+j-i)^{-d-1}) & \text{if } i\le j,\\ \sum_{k=0}^{j-1}\pi_k(d)\ddot{\pi}_{k+i-j}(d) = O((1+\log(i-j))^2(i-j)^{-d-1}) & \text{else,}\end{cases}$$

$$(\ddot{S}'_{d,t}\dot{S}_{d,t})_{(i,j)} = \begin{cases}\sum_{k=1}^{i-1}\ddot{\pi}_k(d)\dot{\pi}_{k+j-i}(d) = O((1+\log(1+j-i))(1+j-i)^{-d-1}) & \text{if } i\le j,\\ \sum_{k=1}^{j-1}\dot{\pi}_k(d)\ddot{\pi}_{k+i-j}(d) = O((1+\log(i-j))^2(i-j)^{-d-1}) & \text{else,}\end{cases}$$

$$(\dddot{S}'_{d,t}S_{d,t})_{(i,j)} = \begin{cases}\sum_{k=1}^{i-1}\dddot{\pi}_k(d)\pi_{k+j-i}(d) = O((1+j-i)^{-d-1}) & \text{if } i\le j,\\ \sum_{k=0}^{j-1}\pi_k(d)\dddot{\pi}_{k+i-j}(d) = O((1+\log(i-j))^3(i-j)^{-d-1}) & \text{else,}\end{cases}$$

$$(\ddot{B}_{\varphi_{(k,l)},t})_{(i,j)} = \left(\frac{\partial^2 B_{\varphi,t}}{\partial\varphi_{(k)}\partial\varphi_{(l)}}\right)_{(i,j)} = \begin{cases}\ddot{b}_{j-i}(\varphi_{(k,l)}) = O((j-i)^{-\zeta-1}) & \text{if } i<j,\\ 0 & \text{else,}\end{cases}$$

$$(\dddot{B}_{\varphi_{(k,l,m)},t})_{(i,j)} = \left(\frac{\partial^3 B_{\varphi,t}}{\partial\varphi_{(k)}\partial\varphi_{(l)}\partial\varphi_{(m)}}\right)_{(i,j)} = \begin{cases} \dddot{b}_{j-i}(\varphi_{(k,l,m)}) = O((j-i)^{-\zeta-1}) & \text{if } i<j, \\ 0 & \text{else,} \end{cases}$$

$$(\ddot{B}'_{\varphi_{(k,l)},t}B_{\varphi,t})_{(i,j)} = \begin{cases} \sum_{m=1}^{i-1}\ddot{b}_m(\varphi_{(k,l)})b_{m+j-i}(\varphi) = O((1+j-i)^{-\zeta-1}) & \text{if } i\le j, \\ \sum_{m=0}^{j-1} b_m(\varphi)\ddot{b}_{m+i-j}(\varphi_{(k,l)}) = O((i-j)^{-\zeta-1}) & \text{else,} \end{cases}$$

$$(\ddot{B}'_{\varphi_{(k,l)},t}\dot{B}_{\varphi_{(m)},t})_{(i,j)} = \begin{cases} \sum_{h=1}^{i-1}\ddot{b}_h(\varphi_{(k,l)})\dot{b}_{h+j-i}(\varphi_{(m)}) = O((1+j-i)^{-\zeta-1}) & \text{if } i\le j, \\ \sum_{h=1}^{j-1}\dot{b}_h(\varphi_{(m)})\ddot{b}_{h+i-j}(\varphi_{(k,l)}) = O(((i-j)^{-\zeta-1}) & \text{else,} \end{cases}$$

$$(\dddot{B}'_{\varphi_{(k,l,m)},t}B_{\varphi,t})_{(i,j)} = \begin{cases} \sum_{h=1}^{i-1}\dddot{b}_h(\varphi_{(k,l,m)})b_{h+j-i}(\varphi) = O((1+j-i)^{-\zeta-1}) & \text{if } i\le j, \\ \sum_{h=0}^{j-1} b_h(\varphi)\dddot{b}_{h+i-j}(\varphi_{(k,l,m)}) = O((i-j)^{-\zeta-1}) & \text{else,} \end{cases}$$

for $k,l,m = 1,2,...,q+2$. As becomes apparent, the partial derivatives just add a log-term to the convergence rates that is always dominated by its powers and thus does not affect the convergence of the partial derivatives. It follows that the first, second and third partial derivatives of $\tau_j(\theta,t)$ are absolutely summable in j and thus satisfy the condition for (B.12). By (B.12), $B_n = \sup_{d\in D_3,\nu\in\Sigma_\nu,\varphi\in N_\delta(\varphi_0)}\left|\frac{\partial^3 Q(y,\theta)}{\partial\theta^{(3)}}\right| = O_p(1)$. □

Lemma D.8. *For the partial derivatives of $v_t(\theta)$, it holds that*

$$\left.\frac{\partial\tilde{v}_t(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial v_t(\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \sum_{j=1}^{\infty}\left[\tilde{\phi}_{\eta,j}\eta_{t-j} + \tilde{\phi}_{\epsilon,j}\epsilon_{t-j}\right]$$

where $\tilde{\phi}_{\eta,j}$ is $O((1+\log j)^2 j^{-1})$, while $\tilde{\phi}_{\epsilon,j}$ is $O((1+\log t)^5 t^{\max(-d_0,-\zeta)-1})$ for $j<t$ and $O((1+\log j)^7 j^{\max(-d_0,-\zeta)-1})$ for $j\ge t$.

Proof of lemma D.8. Consider

$$\left.\frac{\partial\tilde{v}_t(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial v_t(\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \sum_{j=1}^{t-1}\left.\frac{\partial\tau_j(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\left[\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0)\right] \tag{D.86}$$

$$+\sum_{j=1}^{t-1}\left[\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\tau_j(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\right]\tilde{\xi}_{t-j}(d_0) + \sum_{j=t}^{\infty}\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0) \tag{D.87}$$

$$+\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left[\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\xi_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0}\right] \tag{D.88}$$

$$+\sum_{j=1}^{t-1}\left[\tau_j(\theta_0) - \tau_j(\theta_0,t)\right]\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0} + \sum_{j=t}^{\infty}\tau_j(\theta_0)\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0}. \tag{D.89}$$

Since $\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0) = \sum_{k=t-j}^{\infty}\pi_k(d_0)c_{t-j-k}$, by (D.1), lemma D.4, and assumption 2, the term (D.86) is $\sum_{j=t}^{\infty}\epsilon_{t-j}\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\big|_{\theta=\theta_0}\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0) = \sum_{j=t}^{\infty}O((1+\log j)^6 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j}$. By lemma D.5, (D.1), and assumption 3, the first term in (D.87) is

$$\sum_{j=1}^{t-1}\left[\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\tau_j(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\right]\tilde{\xi}_{t-j}(d_0) = \sum_{j=1}^{t-1}\left[\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\tau_j(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\right]\eta_{t-j}$$

$$+\sum_{j=1}^{\infty}\epsilon_{t-j}\sum_{k=1}^{\min(j,t-1)}\left[\frac{\partial\tau_j(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}-\frac{\partial\tau_j(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\right]\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)$$
$$=\sum_{j=1}^{t-1}O((1+\log t)^5 t^{\max(-d_0,-\zeta)-1})(\eta_{t-j}+\epsilon_{t-j})+\sum_{j=t}^{\infty}O((1+\log j)^7 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j}.$$

For the second term in (D.87), by lemma D.4, (D.1), and assumption 3

$$\sum_{j=t}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)=\sum_{j=t}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\eta_{t-j}+\sum_{j=t}^{\infty}\epsilon_{t-j}\sum_{k=0}^{j-t}\frac{\partial\tau_{t+k}(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-t-k}a_l(\varphi_0)\pi_{j-t-k-l}(d_0)$$
$$=\sum_{j=t}^{\infty}O((1+\log j)^4 j^{\max(-d_0,-\zeta)-1})\eta_{t-j}+\sum_{j=t}^{\infty}O((1+\log j)^6 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j}.$$

Note that (D.88), (D.89) are non-zero only for the derivative w.r.t. d. For (D.88), it holds that $\frac{\partial\pi_j(d-d_0)}{\partial d}\big|_{d=d_0}=-j^{-1}$, see Robinson (2006, pp. 135-136). Thus

$$\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left[\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial d}\bigg|_{\theta=\theta_0}-\frac{\partial\xi_{t-j}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right]=-\sum_{j=t}^{\infty}\eta_{t-j}\sum_{k=0}^{t-1}\frac{\tau_k(\theta_0,t)}{j-k}$$
$$+\sum_{j=t}^{\infty}\epsilon_{t-j}\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\frac{\partial\pi_{j-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}$$
$$=\sum_{j=t}^{\infty}O((1+\log j)^2 j^{-1})\eta_{t-j}+\sum_{j=t}^{\infty}O((1+\log j)^4 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j},$$

by lemma D.2, Johansen and Nielsen (2010, lemma B.3), and assumption 3. For the first term in (D.89), by lemmas D.2, D.3, Johansen and Nielsen (2010, lemma B.3), and assumption 3

$$\sum_{j=1}^{t-1}\left[\tau_j(\theta_0)-\tau_j(\theta_0,t)\right]\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial d}\bigg|_{\theta=\theta_0}=-\sum_{j=1}^{\infty}\eta_{t-j}\sum_{k=1}^{\min(j,t-1)}\frac{\tau_k(\theta_0)-\tau_k(\theta_0,t)}{j+1-k}$$
$$+\sum_{j=0}^{\infty}\epsilon_{t-j}\sum_{k=0}^{\min(j,t-1)}(\tau_k(\theta_0)-\tau_k(\theta_0,t))\sum_{l=0}^{j-k}a_l(\varphi_0)\frac{\partial\pi_{j-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}$$
$$=\sum_{j=1}^{\infty}O((1+\log j)^2 j^{-1})\eta_{t-j}+\sum_{j=1}^{t-1}O((1+\log t)^2 t^{\max(-d_0,-\zeta)-1})\epsilon_{t-j}$$
$$+\sum_{j=t}^{\infty}O((1+\log j)^5 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j},$$

while for the second term in (D.89), by lemma D.2, Johansen and Nielsen (2010, lemma B.3), and assumption 3

$$\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial d}\bigg|_{\theta=\theta_0}=-\sum_{j=t}^{\infty}\eta_{t-j}\sum_{k=t}^{j}\frac{\tau_k(\theta_0)}{j+1-k}+\sum_{j=t}^{\infty}\epsilon_{t-j}\sum_{k=0}^{j-t}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-t-k}a_l(\varphi_0)\frac{\partial\pi_{j-t-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}$$

$$= \sum_{j=t}^{\infty} O((1+\log j)^2 j^{-1})\eta_{t-j} + \sum_{j=t}^{\infty} O((1+\log j)^4 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j}.$$

Together, the results above prove lemma D.8. $\square$

Lemma D.9. *For $v_t(\theta)$ as defined and* (15) *and $\tilde{v}_t(\theta)$ as defined in* (B.2)*, it holds that*

$$\frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \frac{\partial^2 \tilde{v}_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} - \frac{1}{n}\sum_{t=1}^{n} v_t(\theta_0) \frac{\partial^2 v_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} = o_p(1),$$

for all $i, j = 1, ..., q+2$.

Proof of lemma D.9. The proof is analogous to the proof of lemma D.6 and thus is only summarized briefly. It will be helpful to note that there exists a constant $0 < K < \infty$ such that

$$\frac{\partial^2 \tau_k(\theta, t)}{\partial\theta_{(i)}\partial\theta_{(j)}} = O\left((1+\log k)^K k^{\max(-d,-\zeta)-1}\right), \tag{D.90}$$

$$\frac{\partial^2 \tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}} - \frac{\partial^2 \tau_k(\theta, t)}{\partial\theta_{(i)}\partial\theta_{(j)}} = O\left((1+\log t)^K t^{\max(-d,-\zeta)-1}\right). \tag{D.91}$$

(D.90) can be seen directly from the proof of lemma D.4, as the second partial derivatives only add a log-factor to the convergence rates in lemma D.4. (D.91) can be shown analogously to the proof of lemma D.5, where again the second partial derivatives only add a log-factor to the convergence rates in lemma D.5. To simplify the notation, define $h_{3,t_{(i,j)}} = \sum_{k=1}^{t-1} \frac{\partial^2 \tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0} \xi_{t-k}(d_0)$, $h_{4,t_{(i,j)}} = \sum_{k=1}^{t-1} \tau_k(\theta_0, t) \frac{\partial^2 \xi_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, $h_{5,t_{(i,j)}} = \sum_{k=1}^{t-1} \frac{\partial \tau_k(\theta,t)}{\partial\theta_{(i)}}\big|_{\theta=\theta_0} \frac{\partial \xi_{t-k}(d)}{\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, as well as $\tilde{h}_{3,t_{(i,j)}} = \sum_{k=1}^{\infty} \frac{\partial^2 \tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0} \tilde{\xi}_{t-k}(d_0)$, $\tilde{h}_{4,t_{(i,j)}} = \sum_{k=1}^{\infty} \tau_k(\theta_0) \frac{\partial^2 \tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, $\tilde{h}_{5,t_{(i,j)}} = \sum_{k=1}^{\infty} \frac{\partial \tau_k(\theta)}{\partial\theta_{(i)}}\big|_{\theta=\theta_0} \frac{\partial \tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\big|_{\theta=\theta_0}$. The term of interest then can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \frac{\partial^2 \tilde{v}_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} - \frac{1}{n}\sum_{t=1}^{n} v_t(\theta_0) \frac{\partial^2 v_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \\
&= \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)\left(\tilde{h}_{3,t_{(i,j)}} - h_{3,t_{(i,j)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{3,t_{(i,j)}}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
&+ \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)\left(\tilde{h}_{4,t_{(i,j)}} - h_{4,t_{(i,j)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{4,t_{(i,j)}}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
&+ \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)\left(\tilde{h}_{5,t_{(i,j)}} - h_{5,t_{(i,j)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{5,t_{(i,j)}}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
&+ \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)\left(\tilde{h}_{5,t_{(j,i)}} - h_{5,t_{(j,i)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{5,t_{(j,i)}}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right),
\end{aligned} \tag{D.92}$$

and thus the different terms in (D.92) can be considered separately and will be shown to be $o_p(1)$. Note that $\tilde{v}_t(\theta_0)$ adapted to the filtration $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS as explained in the proof of theorem 4.2, while $\tilde{h}_{3,t_{(i,j)}}$, $\tilde{h}_{4,t_{(i,j)}}$, $\tilde{h}_{5,t_{(i,j)}}$ are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. Starting with the first term in (D.92), by plugging in

$\tilde{h}_{3,t_{(i,j)}}$, $h_{3,t_{(i,j)}}$

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{3,t_{(i,j)}}-h_{3,t_{(i,j)}}) &= \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\left(\tilde{\xi}_{t-k}(d_0)-\xi_{t-k}(d_0)\right)\\ &+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\left(\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}-\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\right)\tilde{\xi}_{t-k}(d_0)\\ &+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=t}^{\infty}\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\tilde{\xi}_{t-k}(d_0).\end{aligned}\tag{D.93}$$

The latter two terms in (D.93) are MDS when adapted to $\mathcal{F}_t^{\tilde{\xi}}$, as $(\tilde{v}_t(\theta_0),\mathcal{F}_t^{\tilde{\xi}})$ is a stationary MDS and as the other terms are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. By (D.90) and (D.91), $\sum_{k=t}^{\infty}\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}\tilde{\xi}_{t-k}(d_0)$ as well as $\sum_{k=1}^{t-1}\left(\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}-\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}\right)\tilde{\xi}_{t-k}(d_0)$ are $o_p(1)$. Hence, the latter two terms in (D.93) are also $o_p(1)$. In contrast, the first term in (D.93) is not a MDS. However, by the same proof as for (D.58) (replacing the first partial derivative of $\tau_k(\theta,t)$ by the second partial derivative and noting that this only adds a log-factor to the convergence rate) it can also be shown to be $o_p(1)$. Thus, (D.93) is $o_p(1)$. For the third term in (D.92), by plugging in $\tilde{h}_{4,t_{(i,j)}}$, $h_{4,t_{(i,j)}}$

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{4,t_{(i,j)}}-h_{4,t_{(i,j)}}) &= \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}(\tau_k(\theta_0)-\tau_k(\theta_0,t))\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\\ &+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\tau_k(\theta_0,t)\left(\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}-\frac{\partial^2\xi_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\right)\bigg|_{\theta=\theta_0}\\ &+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=t}^{\infty}\tau_k(\theta_0)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0},\end{aligned}\tag{D.94}$$

where the first and third term are MDS when adapted to $\mathcal{F}_t^{\tilde{\xi}}$, as $\tilde{v}_t(\theta_0)$ is a MDS and the remaining term is $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. The third term is $o_p(1)$, because $\sum_{k=t}^{\infty}\tau_k(\theta_0)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$ is $o_p(1)$ by lemma D.2, and by Hualde and Robinson (2011, lemma 4). The first term is $o_p(1)$ since $(\tau_k(\theta_0)-\tau_k(\theta_0,t))\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$ is $o_p(1)$ by lemma D.3. The second term can be shown to be $o_p(1)$ analogously to (D.64) by replacing the first partial derivatives of $\tilde{\xi}_t(d)$ with the second partial derivatives, as this only adds a log-factor to the convergence rate, see Hualde and Robinson (2011, lemma 4). For the fifth term in (D.92), similarly to (D.93) and (D.94)

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{5,t_{(i,j)}}-h_{5,t_{(i,j)}}) &= \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=t}^{\infty}\frac{\partial\tau_k(\theta_0)}{\partial\theta_{(i)}}\bigg|_{\theta=\theta_0}\frac{\partial\tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\\ &+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta_{(i)}}\bigg|_{\theta=\theta_0}\left(\frac{\partial\tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}-\frac{\partial\xi_{t-k}(d)}{\partial\theta_{(j)}}\right)\bigg|_{\theta=\theta_0}\\ &+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\left(\frac{\partial\tau_k(\theta)}{\partial\theta_{(i)}}-\frac{\partial\tau_k(\theta,t)}{\partial\theta_{(i)}}\right)\bigg|_{\theta=\theta_0}\frac{\partial\tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\bigg|_{\theta=\theta_0},\end{aligned}\tag{D.95}$$

where the first and third term are MDS as before. The first term is $o_p(1)$ by lemma D.4, while the third term is $o_p(1)$ by lemma D.5. The second term can be shown to be $o_p(1)$ analogously to (D.64) using (D.67), as the partial derivatives of $\tau_k(\theta, t)$ only add a log-factor to the convergence rates, see lemma D.4. Thus, (D.95) is also $o_p(1)$. The second, fourth and sixth term in (D.92) can be written as

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) &= \frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=0}^{t-1}(\tilde{\xi}_{t-k}(d_0) - \xi_{t-k}(d_0))\tau_k(\theta_0, t) \\ &+ \frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=1}^{t-1}(\tau_k(\theta_0) - \tau_k(\theta_0, t))\tilde{\xi}_{t-k}(d_0) + \frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=t}^{\infty}\tau_k(\theta_0)\tilde{\xi}_{t-k}(d_0),\end{aligned} \tag{D.96}$$

with $l = 3, 4, 5$. For $l = 3$, (D.96) only differs from (D.71) as it contains the second partial derivatives of $\tau_k(\theta, t)$ in $h_{3,t_{(i,j)}}$. However, they only add a log-factor to the convergence rates of the first partial derivatives, see (D.90). For $l = 4$, (D.96) is almost identical to (D.78), where the only difference is that the former considers the second partial derivatives of $\xi_t(d)$ via $h_{4,t_{(i,j)}}$. Again, the second partial derivatives only add a log-factor to the convergence rates in (D.78) (Hualde and Robinson; 2011, lemma 4). For $l = 5$, (D.96) is again almost identical to (D.78) but now includes the first partial derivative of $\tau_k(\theta, t)$ via $h_{5,t_{(i,j)}}$. As for the other terms, by lemma D.4 the derivative again only adds a log-factor to the convergence rate of $\tau_k(\theta, t)$. Thus, it follows directly from (D.71) and (D.78), together with (D.90) and Hualde and Robinson (2011, lemma 4), that (D.96) is $o_p(1)$. The two remaining terms in (D.92) are $o_p(1)$ by (D.95) and (D.96), as i, j can be interchanged. This completes the proof.

□

Acknowledgments

The research of this paper has partly been conducted while the author was a visiting postdoctoral researcher at Princeton University. The author thanks the participants of the International Conference on Computational and Financial Econometrics 2021, the Annual Meeting of the German Statistical Society 2021, the Annual Meeting of the German Economic Society 2021, the European Meeting of the Econometric Society 2021, the Annual Symposium of the Society for Nonlinear Dynamics and Econometrics 2022, the department seminar at the University of Regensburg, the Vienna Workshop on High Dimensional Time Series in Macroeconomics and Finance 2022, the Annual Meeting of the German Statistical Society 2022, the International Conference on Computational and Financial Econometrics 2022, and the Annual Conference of the International Association of Applied Econometrics 2023 for valuable comments. Support through the project 356439312 financed by the German Research Foundation (DFG) is gratefully acknowledged.

References

Balke, N. S. and Wohar, M. E. (2002). Low-frequency movements in stock prices: A state-space decomposition, *The Review of Economics and Statistics* **84**(4): 649–667.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle', *Journal of Monetary Economics* **7**(2): 151–174.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series, *Biometrica* **60**(2): 217–226.

Burman, P. and Shumway, R. H. (2009). Estimation of trend in state-space models: Asymptotic mean square error and rate of convergence, *The Annals of Statistics* **37**(6B): 3715–3742.

Chan, N. H. and Palma, W. (1998). State space modeling of long-memory processes, *The Annals of Statistics* **26**(2): 719–740.

Chan, N. H. and Palma, W. (2006). Estimation of long-memory time series models: A survey of different likelihood-based methods, *in* T. Fomby and D. Terrell (eds), *Econometric Analysis of Financial and Economic Time Series*, Emerald Group Publishing Limited, Bingley.

Chang, Y., Kaufmann, R. K., Kim, C. S., Miller, J. I., Park, J. Y. and Park, S. (2020). Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate, *Journal of Econometrics* **214**(1): 274–294.

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.

Chen, W. W. and Hurvich, C. M. (2006). Semiparametric estimation of fractional cointegration subspaces, *The Annals of Statistics* **34**(6): 2939–2979.

Clark, P. K. (1987). The cyclical component of U.S. economic activity, *The Quarterly Journal of Economics* **102**(4): 797–814.

Dunsmuir, W. (1979). A central limit theorem for parameter estimation in stationary vector time series and its application to models for a signal observed with noise, *The Annals of Statistics* **7**(3): 490–506.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods: Second Edition*, Oxford University Press, Oxford.

Gadea Rivas, M. D. and Gonzalo, J. (2020). Trends in distributional characteristics: Existence of global warming, *Journal of Econometrics* **214**(1): 153–174.

Grassi, S. and de Magistris, P. S. (2014). When long memory meets the Kalman filter: A comparative study, *Computational Statistics & Data Analysis* **76**: 301–319.

Gray, R. M. (2006). Toeplitz and circulant matrices: A review, *Foundations and Trends in Communications and Information Theory* **2**(3): 155–239.

Hartl, T. and Jucknewitz, R. (2022). Approximate state space modelling of unobserved fractional components, *Econometric Reviews* **41**(1): 75–98.

Harvey, A. C. (1985). Trends and cycles in macroeconomic time series, *Journal of Business & Economic Statistics* **3**(3): 216–227.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Harvey, A. C. (2007). Long memory in stochastic volatility, *in* J. Knight and S. Satchell (eds), *Forecasting Volatility in the Financial Markets*, 3 edn, Butterworth-Heinemann Finance, Oxford, pp. 351–363.

Harvey, A. C. and Peters, S. (1990). Estimation procedures for structural time series models, *Journal of Forecasting* **9**(2): 89–108.

Hassler, U. (2019). *Time Series Analysis with Long Memory in View*, Wiley Series in Probability and Statistics, Wiley, Hoboken, NJ.

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.

Hosoya, Y. (2005). Fractional invariance principle, *Journal of Time Series Analysis* **26**(3): 463–486.

Hualde, J. and Nielsen, M. Ø. (2020). Truncated sum of squares estimation of fractional time series models with deterministic trends, *Econometric Theory* **36**(4): 751–772.

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.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S. and Zhang, H.-M. (2017). Extended reconstructed sea surface temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons, *Journal of Climate* **30**: 8179–8205.

Hunger, R. (2007). Floating point operations in matrix-vector calculus, *Working paper*, Technical University of Munich.

Johansen, S. and Nielsen, M. Ø. (2010). Likelihood inference for a nonstationary fractional autoregressive model, *Journal of Econometrics* **158**(1): 51–66.

Marinucci, D. and Robinson, P. M. (1999). Alternative forms of fractional Brownian motion, *Journal of Statistical Planning and Inference* **80**(1–2): 111–122.

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.

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods*, Wiley, Hoboken, NJ.

Proietti, T. and Maddanu, F. (2022). Modelling cycles in climate series: The fractional sinusoidal waveform process, *Journal of Econometrics* (forthcoming).

Ray, B. K. and Tsay, R. S. (2000). Long-range dependence in daily stock volatilities, *Journal of Business & Economic Statistics* **18**(2): 254–262.

Robinson, P. M. (2005). Efficiency improvements in inference on stationary and nonstationary fractional time series, *Annals of Statistics* **33**(4): 1800–1842.

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. and Phillips, P. C. B. (2005). Exact local Whittle estimation of fractional integration, *The Annals of Statistics* **33**(4): 1890–1933.

Varneskov, R. T. and Perron, P. (2018). Combining long memory and level shifts in modelling and forecasting the volatility of asset returns, *Quantitative Finance* **18**(3): 371–393.

Wolter, K. and Timlin, M. S. (2011). El Niño Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext), *International Journal of Climatology* **31**(7): 1074–1087.

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.

Zygmund, A. (2002). *Trigonometric Series*, Cambridge University Press, Cambridge.