Robust trend estimation for strongly persistent data with unobserved memory

Tobias Hartl*[a]

[a]Department of Quantitative Economics, Maastricht University

March 2025

Abstract. Economic analysis is often based on pre-filtered, de-trended, or seasonally adjusted data. Underlying filtering methods make strong assumptions about the memory of the series to be filtered, and inference about the memory is limited particularly when persistent cyclical variation overshadows the trend. This paper introduces a data-driven method for filtering persistent series that requires no prior assumptions about the memory, thus is robust to the actual memory of the data. It makes three primary contributions: first, it generalizes unobserved components (UC) models to fractionally integrated trends, making prior assumptions about the trend memory redundant while retaining the advantages of the state space structure of UC models; second, it establishes the asymptotic estimation theory for fractional UC models under mild assumptions; and third, it presents a computationally efficient estimator for the trend by deriving the closed-form solution to the Kalman filter optimization problem.

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

JEL-Classification. C32, C51, Q54

*Permanent address: Department of Quantitative Economics, School of Business and Economics, Maastricht University, 6200 MD Maastricht, The Netherlands, email: tobias.hartl@maastrichtuniversity.nl

1 Introduction

What is the extent of global warming? As a glimpse at figure 3 reveals, temperature is characterized by rich, persistent cyclical dynamics that overshadow the permanent temperature rise, i.e. the trend. Estimating the trend of such series is difficult, because both cycle and trend generate spectral mass at the low frequencies, making them hard to distinguish. As a solution, the literature has proposed a variety of unobserved components (UC) models that impose structure by formulating models for trend and cyclical dynamics. While cycles are typically well approximated by trigonometric functions or ARMA models, the choice of the trend is far from trivial: state-of-the-art UC models specify the trend either as a random walk in the spirit of Beveridge and Nelson (1981) and Harvey (1985), or as a quadratic stochastic trend as suggested by Clark (1987) and Hodrick and Prescott (1997). Because of the overlapping spectral densities of trend and cycle, statistical inference about the memory of the trend is limited, and chances of model specification errors are high. This is crucial, as economic analysis is often based on pre-filtered series.

This paper proposes a flexible, data-driven method for decomposing time series into trend and cycle that is agnostic about and robust to the memory of the trend: the underlying state space model retains the structure of an UC model, but allows the trend to be fractionally integrated of order $d \in \mathbb{R}_+$, where d is to be estimated. It encompasses traditional UC models for integer d and seamlessly links them by allowing for intermediate solutions. It provides statistical inference about d robust against persistent cyclical dynamics, bears all the advantages from the state space structure, and requires no bandwidth choice. Besides the trend, the model includes a cyclical component that allows for ARMA specifications typical in the UC literature, but also for a broader class of processes.

As a second contribution, this paper derives the asymptotic estimation theory for fractional UC models under relatively mild assumptions: Allowing for fractional trends with unknown memory d complicates parameter estimation, since underestimating the memory can cause non-stationary residuals. As a consequence, the objective function does not uniformly converge on the full parameter space, but can be shown to diverge within the region where the residuals are non-stationary. This is addressed by showing that the probability of the estimator to converge within the non-stationary region of the parameter space is asymptotically zero, so that the relevant parameter space reduces to the region where the residuals are asymptotically stationary, and where the estimators for d and the other model parameters can be shown to be consistent. Under the somewhat stronger assumption that the residuals are martingale difference sequences, asymptotic normality can be established.

The third contribution of the paper is to propose a computationally feasible estimator for the trend in fractional UC models by deriving the closed-form solution to the optimization problem of the Kalman filter. So far, the literature on state space models has only considered approximations to fractionally integrated processes: For asymptotically stationary processes (i.e. $d < 1/2$) Chan and Palma (1998, 2006), Palma (2007) and Grassi and de Magistris (2014) approximate fractionally integrated processes 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, also including the non-stationary domain. At contrast, by deriving a closed-form solution to the Kalman filter optimization problem, this paper avoids the computationally infeasible Kalman recursions, and thus is the first to propose a computationally feasible, exact state space model encompassing fractionally integrated trends. For $d \in \mathbb{N}$, a closed-form solution to the Kalman filter has been derived by Burman and Shumway (2009), which is generalized to $d \in \mathbb{R}_+$ in this paper.

All theoretical results are established for a prototypical model consisting only of a fractional trend and a cyclical component. Later, they are shown to extend seamlessly to models with deterministic terms and correlated innovations.

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 applies the fractional UC model to realized volatility and temperature anomalies, whereas section 7 concludes. The proofs for consistency and asymptotic normality, as well as Monte Carlo inference, are contained in the online appendix. The code for this paper, a computationally efficient R package containing all necessary functions for fractional UC models, and further simulation results, are available at `https://github.com/tobiashartl/fracUCM`.

2 Model

Building on the unobserved components (UC) literature, I consider a prototypical model that decomposes an observable time series $\{y_t\}_{t=1}^n$ into unobserved trend x_t and cycle c_t

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

Cycle c_t and trend x_t are distinguished by their spectra: The cycle (or short-run component) c_t is a mean zero stationary process capturing the transitory features of y_t, whereas for any t, the autocovariance function of the trend (or long-run component) x_t decays at a rate slower than exponential (Marinucci and Robinson; 1999), reflecting the persistent, possibly non-stationary features of y_t.

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

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

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

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

The motivation behind (2) and (3) is that the higher d, the more persistent the effect of a past shock η_{t-j} on x_t. For this reason $x_t \sim I(d)$ is said to have long memory whenever $d > 0$ (see Hassler; 2019). 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 the trend specifications in the literature: For $d = 1$, it nests the random walk 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 HP filter of Hodrick and Prescott (1997) 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, 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). By restricting c_t to short memory processes, (4) excludes the fractional sinusoidal waveform process, a novel model for persistent cyclical variation as suggested by Proietti and Maddanu (2024).

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 $\varPsi = D \times \varSigma_\eta \times \varSigma_\epsilon \times \varPhi$ denote the parameter space of $\psi \in \varPsi$, where $D = \{d \in \mathbb{R} | 0 < d_{min} \leq d \leq d_{max} < \infty\}$, $\varSigma_\eta = \{\sigma_\eta^2 \in \mathbb{R} | 0 < \sigma_{\eta,min}^2 \leq \sigma_\eta^2 \leq \sigma_{\eta,max}^2 < \infty\}$, $\varSigma_\epsilon = \{\sigma_\epsilon^2 \in \mathbb{R} | 0 < \sigma_{\epsilon,min}^2 \leq \sigma_\epsilon^2 \leq \sigma_{\epsilon,max}^2 < \infty\}$, and $\varPhi \subseteq \mathbb{R}^q$ 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 $\psi_0 \in \varPsi$.*

Assumption 1 allows for conditionally homoscedastic martingale difference sequences (MDS) η_t and ϵ_t, which is somewhat more general than the UC literature. Independence of the shocks is only assumed in order to make the derivation of the asymptotic theory in section 4 more readable. The assumption 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 $d \in D$. While $d > 0$ is assumed for x_t to be a long-run component, all results can be generalized to $d \in \mathbb{R}$ as long as trend and cycle can be distinguished by the autocovariance function of y_t.

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

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

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 $\dot{a}_j(\varphi)$ follows from 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 estimation theory is well established under assumptions similar to 1, 2, and 3, see Hualde and Robinson (2011) and Nielsen (2015). While generalizations of the model are considered in section 5, I will keep the model as simple as possible in the following for the sake of clarity.

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. This opens the way to the Kalman recursions, a powerful set of algorithms for filtering and smoothing the latent components x_t and c_t. In this section, I derive an analytical solution to the optimization problem of the Kalman filter. As will become clear, the analytical solution has two decisive advantages over the usual recursive algorithm: for fractional trends 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 $b_+(L, \varphi)c_t = b(L, \varphi)c_t \mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} b_j(\varphi)c_{t-j}$. For any z_t, define the t-vector $z_{t:1} = (z_t, ..., z_1)'$, and define the $t \times t$ 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 $S_{d,t}^{-1} = S_{-d,t}$, $S_{d,t}S_{-d,t} = I$, 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 transition 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). A closed-form solution for the Kalman smoother for x_t and c_t follows from (6), (7) by setting $t = n$.

For (6) and (7), note that the Kalman filter is the minimum variance linear unbiased estimator that is $\hat{x}_{t:1}(y_{t:1},\psi) = \mathrm{Cov}_\psi(x_{t:1}, y_{t:1})\,\mathrm{Var}_\psi(y_{t:1})^{-1}y_{t:1}$ for the trend (see Durbin and Koopman; 2012, lemma 2). The ψ-subscript, e.g. on $\mathrm{E}_\psi(z_t)$, denotes that expectation is taken with respect to the distribution of z_t given ψ. Under assumption 1, $\mathrm{Cov}_\psi(x_{t:1}, y_{t:1}) = \sigma_\eta^2(S_{d,t}'S_{d,t})^{-1}$, and $\mathrm{Var}_\psi(y_{t:1}) = \sigma_\eta^2(S_{d,t}'S_{d,t})^{-1} + \sigma_\epsilon^2(B_{\varphi,t}'B_{\varphi,t})^{-1}$, so that

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

Since $\hat{x}_{t:1}(y_{t:1},\psi) = y_{t:1} - \hat{c}_{t:1}(y_{t:1},\psi)$, $\hat{c}_{t:1}(y_{t:1},\psi)$ follows immediately. This proves the first equality of (6) and (7). For the second, 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, $0 < \nu < \infty$. Defining $\theta = (d, \nu, \varphi')'$, one has $\hat{x}_{t:1}(y_{t:1},\psi) = \hat{x}_{t:1}(y_{t:1},\theta)$ and $\hat{c}_{t:1}(y_{t:1},\psi) = \hat{c}_{t:1}(y_{t:1},\theta)$. For $t = n$, $d = 2$, $b(L,\varphi) = 1$, (6) and (7) equal the HP filter with tuning parameter ν, whereas for normally distributed errors, they are the conditional expected value of trend and cycle (conditional on $y_{t:1}$).

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. They follow immediately from (6) and (7) by setting $t = n$, and equal the Kalman smoother.

Note that the filtered, predicted and smoothed x_t and c_t can be computed either via the analytical solution or by the Kalman recursions (see Harvey; 1989, ch. 3, for the latter). However, the former is faster to compute 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 $m \geq n-1$. Each recursion of the Kalman filter involves multiple multiplications of $(m \times m)$-dimensional covariance and system matrices, each requiring $2m^3 - m^2$ flops (Hunger; 2007). The analytical solution requires the expensive computation of an $(n \times n)$ Cholesky inverse, taking $n^3 + n^2 + n$ flops per iteration (Hunger; 2007). Since $m \geq n-1$, the analytical solution speeds up the computation considerably: For $n = 250$, the analytical solution $\hat{x}_{250:1}(y_{t:1}, \varphi)$ takes about 0.1 seconds, and outperforms the Kalman recursions by a factor of 300, whereas for $n = 500$ it takes about 0.8 seconds and improves performance by a factor of about 600. This is severe, as the application in subsection 6.2, with more than 2,000 observations, takes about six hours to compute. Moreover, the closed-form solution allows to derive an objective function for parameter estimation that does not depend on the Kalman recursions and is thus easier to analyze. This simplifies the derivation of the asymptotic estimation theory in section 4.

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 prediction 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)

$$\begin{aligned} v_{t+1}(\theta) =& \nu \left(b_1(\varphi) - \pi_1(d) \ \cdots \ b_t(\varphi) - \pi_t(d)\right) (B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t})^{-1} S_{d,t}' S_{d,t} y_{t:1} \\ &+ \Delta_+^d y_{t+1} = \sum_{j=0}^{t} \tau_j(\theta, t) \xi_{t+1-j}(d), \end{aligned} \tag{10}$$

where $\xi_{t+1}(d) = \Delta_+^d y_{t+1} = \Delta_+^{d-d_0}\eta_{t+1} + \Delta_+^d c_{t+1}$ and $\xi_{t:1}(d) = (\xi_t(d)\cdots\xi_1(d))' = S_{d,t}y_{t:1}$ denote the differenced observable variable, $\tau_0(\theta,t) = 1$, and $(\tau_1(\theta,t)\cdots\tau_t(\theta,t)) = \nu(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}S'_{d,t}$. Minimizing the sum of squared prediction errors yields the conditional sum-of-squares (CSS) estimator for θ_0

$$\hat{\theta} = \arg\min_{\theta\in\Theta} Q(y,\theta), \qquad Q(y,\theta) = \frac{1}{n}\sum_{t=1}^{n} v_t^2(\theta). \tag{11}$$

$Q(y,\theta)$ is proportional to the exponent in the quasi-likelihood function. Therefore, as discussed in subsection 5.3, (11) and quasi-maximum likelihood (QML) estimation are asymptotically equivalent. However, (11) is computationally much simpler, because it avoids the Kalman recursions for the prediction error variance.

While the asymptotic theory for CSS estimation is well established for ARFIMA models, see Hualde and Robinson (2011) and Nielsen (2015), little is known about the asymptotic theory for UC models of such generality. For the sub-class of $I(1)$ UC models with Gaussian white noise innovations, 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. While the proofs in this section are given for the (simpler) CSS estimator, they carry over seamlessly to QML estimation, see subsection 5.3.

Theorem 4.1. *For the model in* (1), (2), *and* (4), *and under assumptions 1 to 3, the estimator* $\hat{\theta}$ *as defined via* (11) *is consistent,* $\hat{\theta} \xrightarrow{p} \theta_0$ *as* $n \to \infty$.

The proof is contained in Appendix A. 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 (10): First, as can be seen from (10), 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 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 asymptotically reduces to the region where $d_0 - d < 1/2$.

Second, even within the asymptotically stationary region of the parameter space, the prediction errors are non-ergodic, as can be seen from (10). Consequently, a weak law of large numbers (WLLN) for stationary and ergodic series does not apply directly to $v_t(\theta)$. I tackle this problem by showing the difference between the prediction error (10) and the untruncated and ergodic $\tilde{v}_t(\theta) = \sum_{j=0}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d)$ to be asymptotically negligible in probability, where $\tilde{\xi}_t(d) = \Delta^{d-d_0}\eta_t + \Delta^d c_t$ is the untruncated residual, 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 WLLN applies to the objective function. The final part is to strengthen pointwise convergence in probability to weak convergence.

To derive the asymptotic distribution, 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). It ensures absolute summability of the partial derivatives, which is used for uniform convergence of the Hessian matrix. The convergence rates of the partial derivatives are a direct consequence of compactness of

$N_\delta(\varphi_0)$ and continuity of the derivatives. Assumption 4 still includes the class of stationary ARMA processes, and even allows for a slower 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* $\tilde{\xi}_s = \tilde{\xi}_s(d_0)$*.*

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 $t \to \infty$, whereas $\Delta_+^{d_0} y_t = \xi_t(d_0) = \tilde{\xi}_t + o_p(1)$ 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 equals the filtration generated by the differenced, observable variables. For assumption 5 to be satisfied, the 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 corresponds 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, and 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,* $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathrm{N}(0, \sigma_{v,0}^2 \Omega_0^{-1})$ *as* $n \to \infty$*, with* $\sigma_{v,0}^2 = \lim_{t\to\infty} \mathrm{Var}(v_t(\theta_0)) = \mathrm{Var}(\tilde{v}_t(\theta_0))$*, 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 B. 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. While a WLLN 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 be 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 recursions applicable, enables quasi-maximum likelihood estimation, allows to diffusely initialize the filter (see Durbin and Koopman (2012, ch. 5)), and to seamlessly add additional structural components to the model. In addition, several useful methods and generalizations beyond the scope of this paper become available, 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 terms, while subsection 5.2 allows for correlated innovations. Subsection 5.3 considers quasi-maximum likelihood estimation. For all three modifications, the results of section 4 remain valid.

Before going into the details, I outline the basic state space representation

$$y_t = Z\alpha_t + u_t, \tag{12}$$

$$\alpha_t = T\alpha_{t-1} + R\zeta_t. \tag{13}$$

The states may be partitioned into $\alpha_t = (\alpha_t^{(x)\prime}, \alpha_t^{(c)\prime}, \alpha_t^{(r)\prime})'$, where $\alpha_t^{(r)}$ accounts for additional, not further specified components as explained at the end of the paragraph. Beyond $\alpha_t^{(r)}$, α_t holds the $(n-1)$-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 dimension $(n-1)$ results from x_n, and c_n depending on $(n-1)$ lags, so that a minimal, time-invariant state space model requires state vectors of that dimension. 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 (13), one has $T = \mathrm{diag}(T^{(x)}, T^{(c)}, T^{(r)})$, $R = \mathrm{diag}(R^{(x)}, R^{(c)}, R^{(r)})$,

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

and $R^{(x)} = (1, 0, ..., 0)'$, $R^{(c)} = (1, 0, ..., 0)'$ are $(n-1)$-vectors picking the respective entries of $\zeta_t = (\eta_t, \epsilon_t, \zeta_t^{(r)'})'$. Consistent with the type-II definition, α_t is initialized deterministically with $x_j = c_j = 0$ for all $j \leq 0$, so that in $\alpha_t^{(x)} = T^{(x)}\alpha_{t-1}^{(x)}$ only $x_{t-1}, ..., x_1$ enter, whereas all other x_j are zero. 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 and seasonal components among others. 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

Deterministic components are present in the vast majority of UC models and can be added to the fractional UC model following Harvey (1989, ch. 3.4.2): they are placed directly into the measurement equation (12), and are estimated via generalized least squares (GLS). All remaining parameters θ_0 can be estimated as described in section 4, with the asymptotic theory being unaffected. In what follows, I summarize the estimation of deterministic components in the fractional UC model, where further details are given in Harvey (1989, ch. 3.4.2). The end of the subsection discusses consistency of the GLS estimator, where it is shown that deterministic trends of order similar to x_t can be estimated consistently, whereas trends of order below $d_0 - 1/2$ are differenced away whenever x_t is nonstationary.

Let $u_t = \mu' w_t$ in (12), where w_t is an observable deterministic k-vector, and μ is a k-vector of unknown parameters to be estimated. The new measurement equation is $y_t = \mu' w_t + Z\alpha_t$. Let $W = (w_1, ..., w_n)'$, $\tilde{u}_t = Z\alpha_t$, $\tilde{u} = (\tilde{u}_1, ..., \tilde{u}_n)'$ and $V = \mathrm{Var}(x_{1:n} + c_{1:n}) = \mathrm{Var}(\tilde{u})$. (12) then can be written as the regression model

$$y_{1:n} = W\mu + \tilde{u}, \qquad \mathrm{Var}(\tilde{u}) = V,$$

with the GLS estimator for μ being $\tilde{\mu} = (W'V^{-1}W)^{-1}W'V^{-1}y_{1:n}$, see Harvey (1989, ch. 3.4.2). As also shown there, it is not necessary to compute V^{-1}: 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 $\tilde{\theta} = \arg\min_\theta \frac{1}{n}\sum_{t=1}^n \tilde{v}_t^*(\theta)^2$, and $\tilde{v}_t^*(\theta) = y_t^*(\theta) - \tilde{\mu}' w_t^*(\theta)$ is the GLS residual.

To derive the asymptotic properties of both GLS estimator $\tilde{\mu}$ and concentrated CSS estimator, 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.

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 equals the prediction error (10) in section 3 shifted either by a constant, or by an $o(1)$ term. Therefore $\tilde{v}_t^*(\theta) = [y^*(\theta) - W^*(\theta)(W^{*'}(\theta)W^*(\theta))^{-1}W^{*'}(\theta)y^*(\theta)]_{(t)}$ is $I(d_0 - d)$. Both $y_t^*(\theta)$ and $\tilde{v}_t^*(\theta)$ are asymptotically stationary for $d_0 - d < 1/2$, otherwise non-stationary. By the same proof as for (A.1) in appendix A, the objective function of the concentrated CSS estimator can be shown to converge in probability whenever $d_0 - d > -1/2$, and to diverge otherwise. Thus, it is sufficient to consider the region of the parameter space where $\tilde{v}_t^*(\theta)$ is asymptotically stationary. There, the same proof as for theorem 4.1 applies, and $\tilde{\theta}$ is consistent.

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 $c_{1:n}$. The GLS estimates $\tilde{\mu}$ are

$$\tilde{\mu} = (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})F(\tilde{\theta})y_{n:1} = \mu_0 + (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta}), \tag{14}$$

where μ_0 denotes the true coefficients to be estimated. $\tilde{\mu}$ is consistent if and only if the latter term in (14) is $o_p(1)$. 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 (14) can be written as

$$(W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta}) = \left(\frac{\sum_{t=1}^n w_t^{*^2}(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}\right)^{-1} \frac{\sum_{t=1}^n w_t^*(\tilde{\theta})u_t^*(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}, \tag{15}$$

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 $d_0 - 1/2 < \beta$. As also can be seen from (15), 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 $n^{1/2+(\beta-d_0)}$.

In summary, trends of order $d_0 - 1/2 < \beta_j \leq d_0$ are estimated consistently, while deterministic terms of order below $d_0 - 1/2$ are differenced away asymptotically, such that CSS and QML estimator retain their asymptotic properties. This aligns to the well-established finding that for series with unit roots ($d_0 = 1$), a linear trend ($\beta_j = 1$) can be estimated consistently, whereas an intercept (i.e. $\beta_j = 0$) cannot. Moreover, the convergence rate matches the findings of Robinson (2005) for semiparametric long memory models, of Hualde and Nielsen (2020) for parametric ARFIMA models, and the literature on the estimation of the sample mean for fractionally integrated processes (Hassler; 2019, ch. 7).

5.2 Correlated trend and cycle innovations

As shown by Morley et al. (2003), for integer-integrated UC 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 correlated innovations

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

Let $\tilde{\psi} = (d, \sigma_\eta^2, \sigma_{\eta\epsilon}, \sigma_\epsilon^2, \varphi')'$ denote the new parameter vector that now also includes the covariance $\sigma_{\eta\epsilon}$. Similar to section 3, to derive the Kalman filter estimate for the trend, consider the minimum variance linear unbiased estimator $\hat{x}_{t:1}(y_{t:1}, \tilde{\psi}) = \operatorname{Cov}_{\tilde{\psi}}(x_{t:1}, y_{t:1}) \operatorname{Var}_{\tilde{\psi}}(y_{t:1})^{-1} y_{t:1}$, where now $\operatorname{Cov}_{\tilde{\psi}}(x_{t:1}, y_{t:1}) = \sigma_\eta^2 (S_{d,t}' S_{d,t})^{-1} + \sigma_{\eta\epsilon}(S_{d,t}' B_{\varphi,t})^{-1}$, and $\operatorname{Var}_{\tilde{\psi}}(y_{t:1}) = \sigma_\eta^2 (S_{d,t}' S_{d,t})^{-1} +$

$\sigma_{\eta\epsilon}\left[(S'_{d,t}B_{\varphi,t})^{-1} + (B'_{\varphi,t}S_{d,t})^{-1}\right] + \sigma_\epsilon^2(B'_{\varphi,t}B_{\varphi,t})^{-1}$. Plugging in and rearranging yields

$$\begin{aligned}\hat{x}_{t:1}(y_{t:1},\tilde{\psi}) &= \left[\sigma_\eta^2 B'_{\varphi,t}B_{\varphi,t} + \sigma_{\eta\epsilon}(S'_{d,t}B_{\varphi,t} + B'_{\varphi,t}S_{d,t}) + \sigma_\epsilon^2 S'_{d,t}S_{d,t}\right]^{-1} \\ &\quad \times \left(\sigma_\eta^2 B'_{\varphi,t}B_{\varphi,t} + \sigma_{\eta\epsilon}S'_{d,t}B_{\varphi,t}\right) y_{t:1},\end{aligned} \tag{16}$$

and by using $y_{t:1} = \hat{x}_{t:1}(y_{t:1},\tilde{\psi}) + \hat{c}_{t:1}(y_{t:1},\tilde{\psi})$

$$\begin{aligned}\hat{c}_{t:1}(y_{t:1},\tilde{\psi}) &= \left[\sigma_\eta^2 B'_{\varphi,t}B_{\varphi,t} + \sigma_{\eta\epsilon}(S'_{d,t}B_{\varphi,t} + B'_{\varphi,t}S_{d,t}) + \sigma_\epsilon^2 S'_{d,t}S_{d,t}\right]^{-1} \\ &\quad \times \left(\sigma_\epsilon^2 S'_{d,t}S_{d,t} + \sigma_{\eta\epsilon}B'_{\varphi,t}S_{d,t}\right) y_{t:1}.\end{aligned} \tag{17}$$

Obviously, (16) and (17) 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 (16) and (17) 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 QML estimator as derived in subsection 5.3 can be used to estimate $\tilde{\psi}_0 = (d_0, \sigma^2_{\eta,0}, \sigma_{\eta\epsilon,0}, \sigma^2_{\epsilon,0}, \varphi'_0)$ 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

$$\begin{aligned}&v_{t+1}(\tilde{\psi}) = \Delta_+^d y_{t+1} + (b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d)) \\ &\times \left[\sigma_\eta^2 B'_{\varphi,t}B_{\varphi,t} + \sigma_{\eta\epsilon}(S'_{d,t}B_{\varphi,t} + B'_{\varphi,t}S_{d,t}) + \sigma_\epsilon^2 S'_{d,t}S_{d,t}\right]^{-1} \left(\sigma_\epsilon^2 S'_{d,t} + \sigma_{\eta\epsilon}B'_{\varphi,t}\right) S_{d,t}y_{t:1}.\end{aligned} \tag{18}$$

Based on (18), a CSS estimator for $\bar{\theta}_0 = (d_0,\nu_0,\nu_{2,0},\varphi'_0)$ can be set up. Note that y_{t+1} enters (18) in fractional differences, and that all terms in (18) have the same convergence rates as for the case with uncorrelated errors. Thus, the asymptotic theory for CSS estimation with correlated innovations follows by the same proofs as in section 4. Finally, as noted by Morley et al. (2003), for $d_0 = 1$, 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, leaving only two autocovariance equations for the estimation of the three parameters in Q. This is no problem for integration orders $d_0 \in \mathbb{R}\setminus\{1\}$: by (3), the autocovariance function of $\Delta_+^{d_0} y_n$ dies out at lag n, which leaves sufficient equations to identify Q even if c_t is white noise.

5.3 Maximum likelihood estimation

For the QML estimator, denote $\psi = (d, \sigma_\eta^2, \sigma_\epsilon^2, \varphi)'$ the vector holding the model parameters of the fractional UC model, and let $\mathrm{Var}_\psi(v_t(\psi)) = \sigma_{v_t}^2$ denote the (hypothetical) variance of $v_t(\psi)$. A recursive solution for $\sigma_{v_t}^2$ can either be obtained from the Kalman filter, see Durbin and Koopman (2012, ch. 4.3), or from the law of total variance $\sigma_{v_{t+1}}^2 = \mathrm{Var}_\psi(y_{t+1}) - \mathrm{Cov}_\psi(x_{t+1}, y_{t:1}) \mathrm{Var}_\psi(y_{t:1})^{-1} \mathrm{Cov}_\psi(y_{t:1}, x_{t+1}) - \mathrm{Cov}_\psi(c_{t+1}, y_{t:1}) \mathrm{Var}_\psi(y_{t:1})^{-1} \mathrm{Cov}_\psi(y_{t:1}, c_{t+1})$. The (scaled) quasi-log likelihood is then

$$n^{-1} \log L(\psi) = -\frac{1}{2n} \sum_{t=1}^{n} \log \sigma_{v_t}^2 - \frac{1}{2n} \sum_{t=1}^{n} \frac{v_t^2(\psi)}{\sigma_{v_t}^2}. \tag{19}$$

As will be shown, (19) asymptotically becomes independent of the initialization of $\sigma_{v_t}^2$, which is referred to as steady state convergence of the Kalman filter and is sufficient for CSS and QML to have the same asymptotic distribution (Harvey; 1989, p. 129). For univariate, integer-integrated UC models, equivalence of CSS and QML is well established in the literature, see Harvey (1989, ch. 4.2.2), and is typically shown by verifying that the model is detectable and stabilizable (Harvey; 1989, ch. 3.3). These two conditions are sufficient for an invertible reduced form with a stationary and ergodic prediction error, so that steady state convergence of the prediction error variance follows by the law of large numbers. However, as (i) the fractional UC model makes no prior assumption about the integration order d_0 (as required for stabilizability), and (ii) the dimension of the state space model (12) and (13) increases with n, making the prediction error (10) non-ergodic, I establish equivalence of CSS and QML for the fractional UC model under milder conditions.

First, by (A.1) in appendix A, the sum of squared prediction errors in (19) diverges in probability whenever $d < d_0 - 1/2$. Since $\frac{1}{2n} \sum_{t=1}^{n} \log \sigma_{v_t}^2$ must be bounded for (19) to be finite, it holds that the probability, that the QML estimator converges within the region where $d < d_0 - 1/2 + \kappa_3$ holds, is zero asymptotically for some $0 < \kappa_3 < 1/2$, see (A.3) in appendix A. As in the proof of theorem 4.1, the relevant parameter space asymptotically reduces to the region where $d - d_0 \geq -1/2 + \kappa_3$ holds, and where $v_t(\psi)$ is asymptotically stationary. Within this region, by (A.4) in appendix A, the sum of squared prediction errors uniformly converges in probability to $\mathrm{E}[\tilde{v}_t^2(\psi)] = \sigma^2_{\tilde{v}}$, the steady state solution of $\sigma^2_{v_t}$, with $\tilde{v}_t(\psi)$ being the untruncated error as given in (A.2) in appendix A. Second, by a Taylor approximation for the second term in (19)

$$\frac{1}{2n}\sum_{t=1}^{n}\frac{v_t^2(\psi)}{\sigma^2_{v_t}} = \frac{1}{2n}\sum_{t=1}^{n}\frac{v_t^2(\psi)}{\sigma^2_{\tilde{v}}} + \frac{1}{2n\sigma^4_{\tilde{v}}}\sum_{t=1}^{n} O\left[v_t^2(\psi)(\sigma^2_{v_t} - \sigma^2_{\tilde{v}})\right] = \frac{1}{2n}\sum_{t=1}^{n}\frac{v_t^2(\psi)}{\sigma^2_{\tilde{v}}} + o_p(1),$$

where the last equality follows from $n^{-1}\sum_{t=1}^{n} v_t^2(\psi)(\sigma^2_{v_t} - \sigma^2_{\tilde{v}}) = n^{-1}\sum_{t=1}^{n}[v_t^2(\psi) - \sigma^2_{\tilde{v}}](\sigma^2_{v_t} - \sigma^2_{\tilde{v}}) + n^{-1}\sigma^2_{\tilde{v}}\sum_{t=1}^{n}(\sigma^2_{v_t} - \sigma^2_{\tilde{v}}) = n^{-1}\sum_{t=1}^{n}[v_t^2(\psi) - \sigma^2_{\tilde{v}}](\sigma^2_{v_t} - \sigma^2_{\tilde{v}}) + o(1)$, and $n^{-1}\sum_{t=1}^{n}[v_t^2(\psi) - \sigma^2_{\tilde{v}}]|\sigma^2_{v_t} - \sigma^2_{\tilde{v}}| \leq cn^{-1}\sum_{t=1}^{n}[v_t^2(\psi) - \sigma^2_{\tilde{v}}] = o_p(1)$ for some $0 < c < \infty$, see (A.4) in appendix A. Analogously, for the first term in (19), it holds that $n^{-1}\sum_{t=1}^{n}\log\sigma^2_{v_t} = \log\sigma^2_{\tilde{v}} + n^{-1}\sigma^{-2}_{\tilde{v}}\sum_{t=1}^{n} O\left(\sigma^2_{v_t} - \sigma^2_{\tilde{v}}\right) = \log\sigma^2_{\tilde{v}} + o(1)$. It follows for the quasi-log likelihood in (19)

$$n^{-1}\log L(\psi) = -\frac{1}{2}\log\sigma^2_{\tilde{v}} - \frac{1}{2n\sigma^2_{\tilde{v}}}\sum_{t=1}^{n} v_t^2(\psi) + o_p(1),$$

which is a scaled translation of the CSS objective function asymptotically. Consequently, the asymptotic theory in section 4 also holds for the QML estimator.

6 Application

Two applications of the fractional UC model are presented: Subsection 6.1 examines the memory of realized volatility, whereas subsection 6.2 decomposes sea surface temperature into trend and cycle to quantify the permanent effect of climate change on temperature.

Both subsections consider the fractional UC model of form

$$y_t = x_t + c_t, \qquad \Delta^d_+ x_t = \mu + \eta_t, \qquad \sum_{j=0}^{p} b_j c_{t-j} = \epsilon_t, \tag{20}$$

where $b_0 = 1$, and thus c_t is a stationary AR(p), 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. Moreover, $\mathrm{Var}(\eta_t, \epsilon_t)' = Q$ may be non-diagonal, see subsection 5.2. Seasonal dummies were found insignificant for both applications and thus were excluded.

Parameters of the fractional UC model are estimated via QML as described in subsection 5.3. While the Monte Carlo studies in appendix D show a similar performance of CSS and QML for most parameters, estimates for Q_0 from QML are found more accurate. Starting values for numerical optimization are obtained by drawing 100 combinations from uniform distributions with appropriate support: d is drawn from $[0; 2]$ for realized volatility and from $[1/2; 2]$ for temperature, as estimates from the Whittle estimator of Shimotsu (2010) for $m = \lfloor n^\alpha \rfloor$ Fourier frequencies, $\alpha = 0.33, 0.34, ..., 0.90$, were all within $[0.36; 0.67]$ and $[0.60; 1.11]$ for realized volatility and temperature respectively. Q is drawn from $\log \sigma_\epsilon^2 \in [-5; 15]$, $\log \sigma_\eta^2 \in [-10; 10]$ for log realized volatility (RV), and from $\log \sigma_\epsilon^2 \in [-10; 0]$, $\log \sigma_\eta^2 \in [-25; -5]$ for temperature. The upper boundary follows from $\log \widehat{\mathrm{Var}}(y_t) = 9.09$ for log RV and $\log \widehat{\mathrm{Var}}(y_t) = -2.57$ for temperature, whereas the lower boundary gives room for strongly persistent cycles and trends with high memory. Autoregressive parameters are drawn randomly from the set of coefficients that ensure stability. Increasing the grid size to 1000 did not alter the results. To reduce computational burdens and speed up the grid search, I use ARMA(3, 3) approximations for the fractional differencing operator as suggested by Hartl and Jucknewitz (2022) . The estimate that maximizes the likelihood of the (approximate) fractional UC model is then taken as starting value for the numerical likelihood maximization of the (exact) fractional UC model.

6.1 Decomposing realized financial volatility

Realized volatility (RV), defined as the sum of squared return realizations over a finite time interval, is characterized by strong persistence at the stationary to nonstationary boundary, and is among the most prominent use cases of long memory models, see Maasoumi and McAleer (2008) for an overview. Andersen et al. (2001) estimate the integration orders of daily RV for the 30 DIJA stocks to be in $[0.29, 0.46]$, whereas the integration order estimates of Christensen and Nielsen (2006) and Nielsen (2007) for weekly RV for the

S&P500 are in $[0.42, 0.48]$. Applying the fractional UC model, this subsection examines whether RV is indeed asymptotically stationary, or whether a possibly nonstationary trend in RV is overshadowed by a persistent but stationary additive cycle. Beyond memory, out-of-sample forecast properties are assessed.

I construct a monthly RV measure by sampling daily S&P500 returns from January 1, 1960 to October 31, 2024 and calculating

$$\sigma^2_{RV,t} = \frac{1}{K_t} \sum_{k=1}^{K_t} r^2_{t,k}, \qquad t = 1, ..., n, \tag{21}$$

where $r_{t,k}$ are daily returns in month t, sampled at $k = 1, ..., K_t$ trading days. Considering volatility at the monthly (rather than the daily) frequency is more in line with x_t capturing long-term market volatility, yields a sample size typical for macroeconomic and financial data where UC models are widely applied, and - as will become clear - by the self-similar nature of fractionally integrated processes yields an integration order estimate similar to the results for daily volatility in the literature. In line with the type II definition of long memory, I shift the log-transformed volatility series $y_t = \log \sigma^2_{RV,t} - \log \sigma^2_{RV,1}$, so that $y_1 = 0$.

Since a focus lies on the out-of-sample properties, (20) is specified according to the BIC, which suggests $\mu = 0$, Q diagonal, and $p = 1$. The QML estimator for the fractional UC model estimates $\hat{d} = 0.45$, together with a 95% confidence interval $[0.31; 0.58]$. This supports the results of Christensen and Nielsen (2006) and Nielsen (2007) that log RV is asymptotically stationary with memory slightly below the stationary to nonstationary boundary. Consequently, x_t should not be interpreted as a trend. The cyclical coefficient $\hat{b}_1 = -0.93$ (standard error (se): 0.06) indicates strong persistence of cyclical volatility. By $\hat{\sigma}^2_\eta = 3927.41$ (se: 504.30) and $\hat{\sigma}^2_\epsilon = 284.74$ (se: 351.66), the fractional UC model attributes more variation to long-run innovations as compared to cyclical innovations.

Figure 1 plots the estimated long-run component of log RV, whereas figure 2 shows the cyclical component. The fractional UC model attributes comparably much variation to the

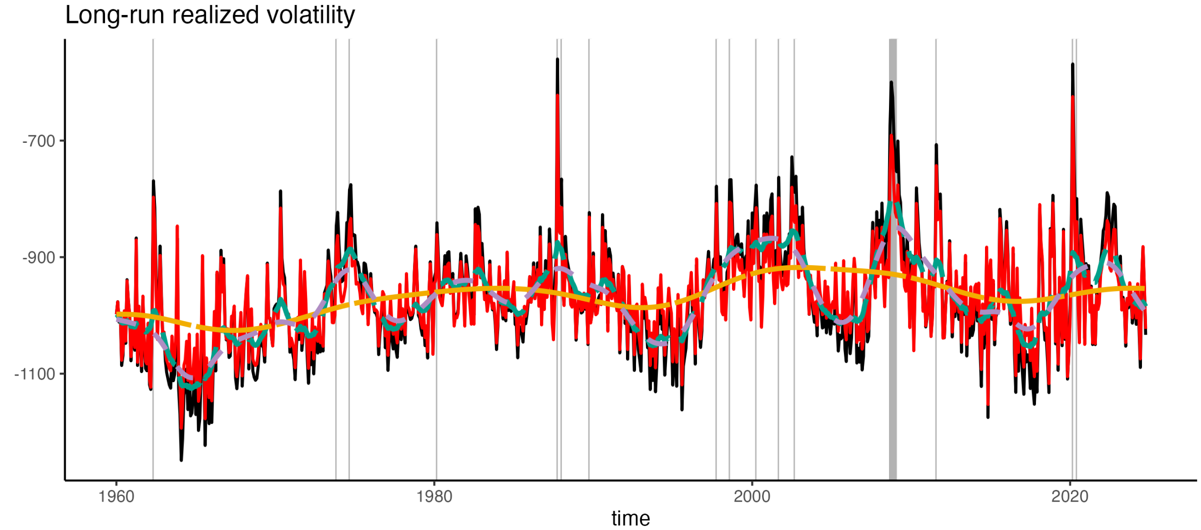

Figure 1: Monthly log RV $\log \sigma^2_{RV,t}$(black) together with the estimated long-run component $\hat{x}_t(y_{n:1}, \hat{\psi}_{QML}) + \log \sigma^2_{RV,1}$ (red, solid) from the fractional UC model, the $I(1)$ UC model (green), the $I(2)$ UC model (yellow), and the HP filter with $\lambda = 14,400$ (purple). Shaded areas correspond to stock market crash periods, defined as months where the S&P500 lost more than 5% on one single trading day, or more than 10% over the entire month.

estimate of x_t, which makes the fractional component rough. This is at contrast to integer-integrated UC models that, by imposing $d = 1$ or even $d = 2$, over-smooth the long-run component. Figure 2 shows a smooth cyclical estimate for the fractional UC model, much at contrast to erratic estimates from the integer-integrated benchmarks. It suggests steep increases in cyclical market volatility during crash periods are followed by a slow recovery.

Even though the focus of UC models is structural analysis and decomposition rather than forecasting, evaluating the out-of-sample properties is crucial to avoid overfitting and end-of-sample distortions. I consider an out-of-sample forecast experiment where the first half of the data (372 observations until December 1990) is treated as training set and is used to specify (20) by minimizing the BIC, analogously to the full sample. This yields $p = 1$, $\mu = 0$, and Q diagonal, as before. Next, applying a recursive window scheme, one observation is added at a time, parameter estimates are updated (without updating the

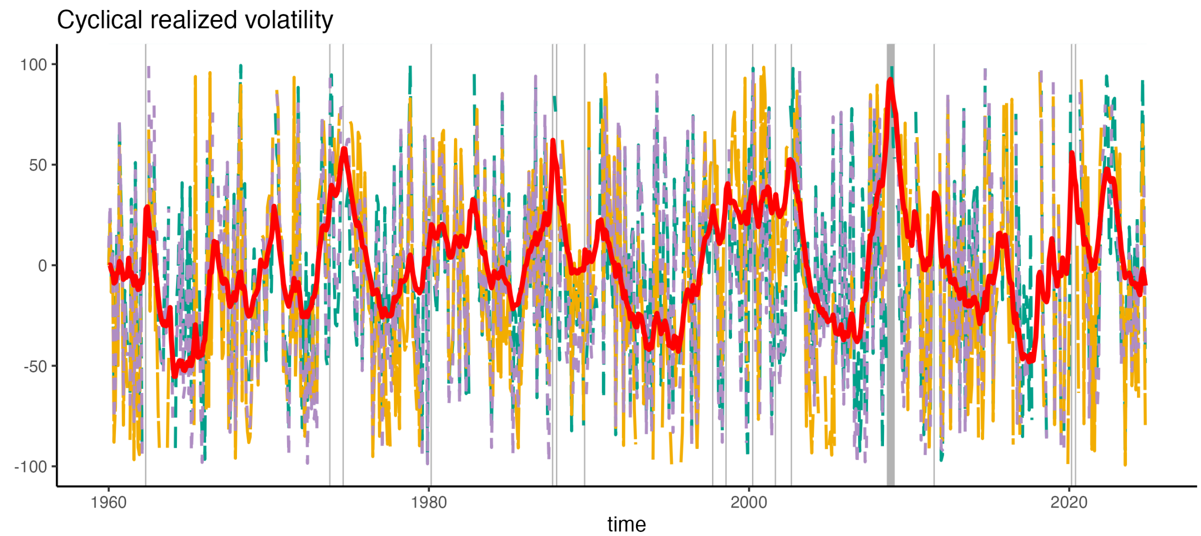

Figure 2: Estimated cyclical log volatility $\hat{c}_t(y_{n:1}, \hat{\psi}_{QML})$ from the fractional UC model (red), the $I(1)$ UC model (green), the $I(2)$ UC model (yellow), and the HP filter with $\lambda = 14,400$ (purple). Shaded areas correspond to crash periods, see figure 1.

model specification), and log RV forecasts for horizons $h \in \{1, 3, 6, 12, 24, 48, 60, 90, 120\}$ (months) are calculated. As benchmarks, I consider the random walk forecast (RW), and autoregressive integrated (ARI), autoregressive integrated moving average (ARIMA), and autoregressive fractionally integrated moving average (ARFIMA) models. For all benchmarks, the lag orders of AR and MA polynomials as well as the integration orders for ARI and ARIMA are chosen either via the AIC or the BIC.

Table 1 compares the forecast performance for the different models. The fractional UC model shows a clear dominance over ARI(MA) and RW models, especially for $h > 12$. ARFIMA benchmarks perform similar to the fractional UC model, indicating that gains in prediction power are mostly from the non-integer d. As most of the variation is captured by the long-run component x_t, the similarity with the ARFIMA is reasonable.

6.2 Tracking climate change

The second application considers monthly global sea surface temperature anomalies. Trends and cycles of climate time series have recently attracted attention in the econometric litera-

h	fUC	ARIMA (1,1,1)	ARIMA (2,1,3)	ARI (2,1)	ARI (4,1)	ARFIMA (0,d,0)	ARIMA (4,d,3)	RW
1	1	1.01	1.02**	1.04**	1.02	1.00	1.01	1.11**
3	1	1.01	1.03*	1.11**	1.05*	1.00	1.01	1.18**
6	1	1.03	1.03	1.11*	1.06	1.01	1.01	1.20**
12	1	1.05	1.04	1.16*	1.11	1.01	1.03	1.23**
24	1	1.10	1.07	1.19**	1.15*	1.02	1.04	1.26**
48	1	1.19**	1.13	1.25**	1.23**	1.03	1.05	1.30**
60	1	1.24**	1.14	1.34**	1.30**	1.02	1.04	1.39**
90	1	1.21**	1.12	1.31**	1.27**	1.02	1.02	1.36**
120	1	1.07	1.02	1.19	1.15	1.02	1.02	1.26

Table 1: (Normalized) L_2 loss for the h months ahead forecasts of log RV for the fUC model (fUC) with $p = 1$ (normalized to unity). Benchmarks are the ARIMA(a, $\tilde{d}$, b), the ARI(a, $\tilde{d}$), the ARFIMA(a, d, b), and the random walk (RW). * and ** indicate significance of the Diebold-Mariano test at the 5% and 1% level.

ture, see Chang et al. (2020), Gadea Rivas and Gonzalo (2020), and Proietti and Maddanu (2024), however fractional trends have not played a role so far. 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 can be matched with well-understood cyclical climate phenomena, such as El Niño. Data stem from the National Centers for Environmental Information and are based on Huang et al. (2017). The monthly series spans from January 1850 to July 2023, and measures the deviation from the 1901–2000 average in degrees Celsius.

Table 2 contains the estimation results for the fractional UC model and the two integer-integrated benchmarks. All models allow for $p = 4$ autoregressive lags in (20), which is supported by the AIC and BIC for the fractional UC model, and the AIC for the benchmarks (for which the BIC suggests $p = 2$). The QML estimator for the fractional UC model yields $\hat{d}_{QML} = 1.753$, together with a 95% confidence interval $[1.634; 1.872]$, supporting a specification of the 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-

	$I(d)$		$I(1)$		$I(2)$	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
d_0	1.753	0.061				
$\sigma^2_{\eta,0}$	1.351E-08	1.527E-08	1.032E-04	4.499E-05	6.179E-10	7.081E-10
$\sigma_{\eta\epsilon,0}$	-2.202E-06	2.620E-06	-5.465E-04	1.402E-04	-1.094E-06	6.279E-07
$\sigma^2_{\epsilon,0}$	1.981E-03	6.171E-05	2.901E-03	2.313E-04	1.955E-03	6.103E-05
$b_{1,0}$	-1.024	0.022	-0.997	0.020	-1.033	0.019
$b_{2,0}$	0.101	0.031	0.094	0.024	0.137	0.014
$b_{3,0}$	0.064	0.031	0.027	0.007	0.018	0.000
$b_{4,0}$	-0.063	0.022	-0.033	0.012	-0.040	0.006
ν_0	1.466E+05		28.115		3.163E+06	
$\nu_{2,0}$	-162.993		-5.296		-1.771E+03	
ρ_0	-0.426		-0.999		-0.996	
$\log L(\psi)$	5436.6		5428.1		5430.4	

Table 2: QML estimation results from the fractional UC model, the $I(1)$ UC model, and the $I(2)$ UC model. Estimates for ν_0, $\nu_{2,0}$, ρ_0 are calculated based on the estimates of Q_0. Standard errors are from the numerical Hessian.

run innovations $\hat{\nu}_{QML} = \hat{\sigma}^2_{\epsilon QML}/\hat{\sigma}^2_{\eta QML} = 146621$ falls into a similar range as the tuning parameter of the HP filter for monthly data, 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 test statistic of the likelihood ratio test for the respective hypothesis is 31.4. Estimates for the AR coefficients suggest a persistent cyclical pattern, with the greatest eigenvalue of the AR polynomial being 0.92.

AR coefficient estimates are similar for the three models, whereas estimates for Q_0 are strongly biased for the benchmarks: If d is assumed lower than d_0, 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 d is assumed higher than d_0, the estimate for $\sigma^2_{\eta,0}$ will be downward-biased. Moreover, both benchmarks converge towards the corner solution of perfectly correlated innovations.

Figure 3 plots the smoothed trend estimate $\hat{x}_t(y_{n:1}, \hat{\psi}_{QML})$ together with those from

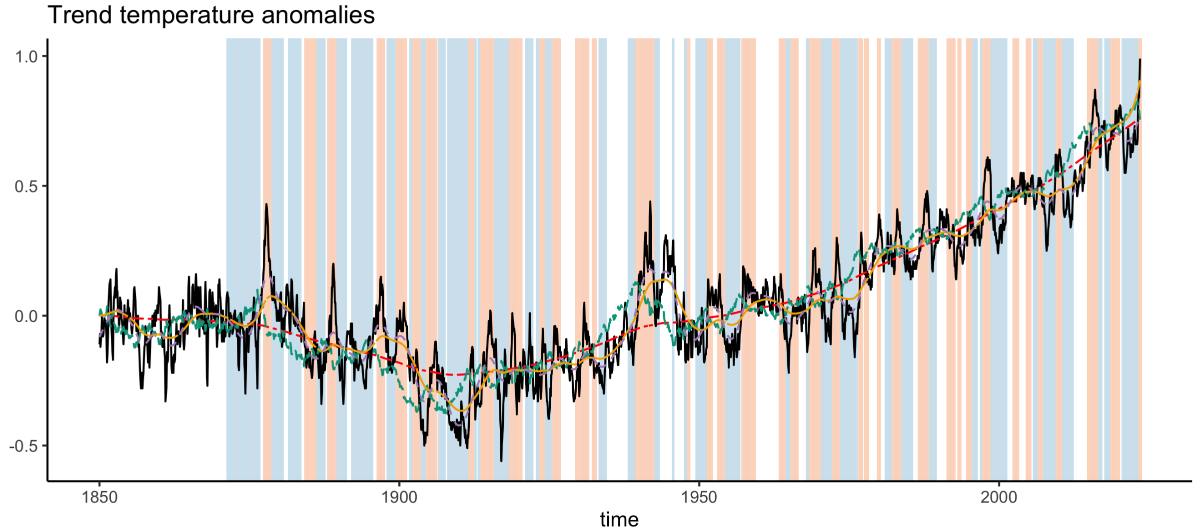

Figure 3: Temperature anomalies (black) together with the estimated trend $\hat{x}_t(y_{n:1}, \hat{\psi}_{QML})$ from the fractional UC model (red), the $I(1)$ UC model (green), the $I(2)$ UC model (yellow), and the HP filter with $\lambda = 14,400$ (purple). Shaded areas correspond to warm (red) and cold (blue) periods based on a threshold of $\pm 1/2$ degrees for the Oceanic Niño Index (ONI).[1]

integer-integrated benchmarks against y_t. In line with the high $\hat{\nu}_{QML}$, it reveals a comparably smooth trend estimate for the fractional UC model. At least since the mid 20th century trend temperature anomalies are strictly increasing, leading to a +0.76 degrees Celsius increase for the last observational period. Contrary to the fractional model, the benchmarks attribute significant short-run variation to the trend component: the $I(1)$ UC model yields a much more erratic trend that behaves countercyclical, whereas HP filter and the $I(2)$ benchmark attribute more of the overall variation to the trend component. Obviously, the additional short-run dynamics in the benchmark models are generated by the (almost) perfect correlation that ties trend and cycle component together.

[1]From 1950 on, the ONI is reported by the National Weather Service (`https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php`). Before 1950, I use the extended multivariate ENSO Index (MEI.ext) of Wolter and Timlin (2011) (`https://psl.noaa.gov/enso/mei.ext/`). Since the MEI.ext is a bi-monthly rolling average, a month is considered a cold (warm) month once bi-monthly rolling averages of current and next month both cross the ±0.5 degrees threshold.

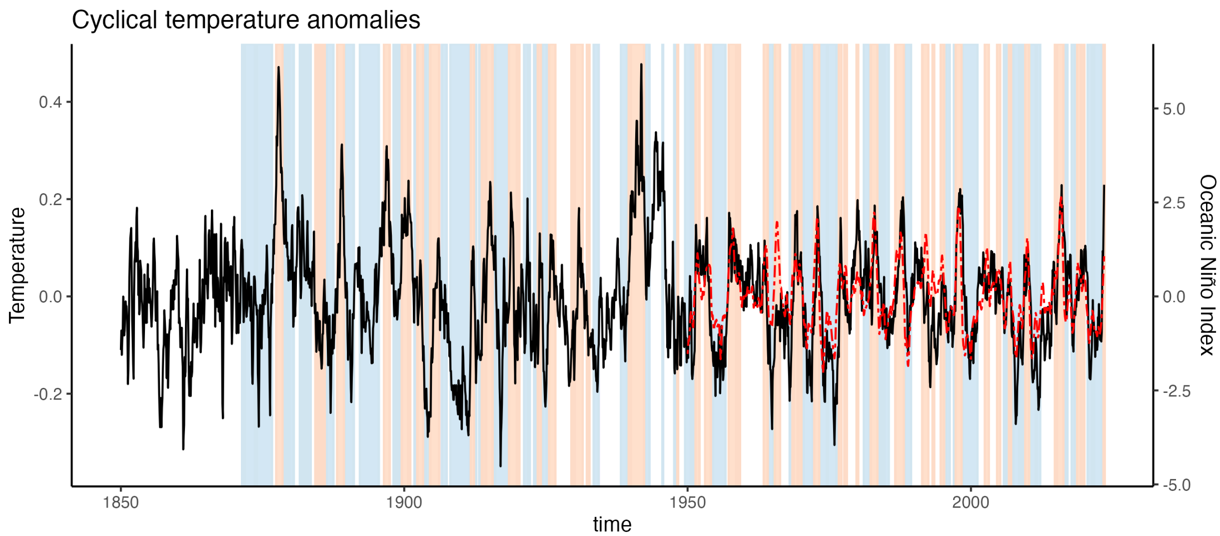

Figure 4: Estimated cyclical sea surface temperature anomalies $\hat{c}_t(y_{n:1}, \hat{\psi}_{QML})$ from the fractional UC model (black), together with the Oceanic Niño Index (red, dashed). Shaded areas correspond to warm (red) and cold (blue) periods according to the ONI, see figure 3.

Figure 4 shows the smoothed cyclical component $\hat{c}_t(y_{n:1}, \hat{\psi}_{QML})$ for the fractional UC model. As already noted, estimates for the AR parameters as well as for the variance 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: peaks during El Niño are followed by troughs during La Niña, the overall dynamics match the ONI series closely, and the correlation between $\hat{c}_t(y_{n:1}, \hat{\psi}_{QML})$ and the ONI is 0.67.

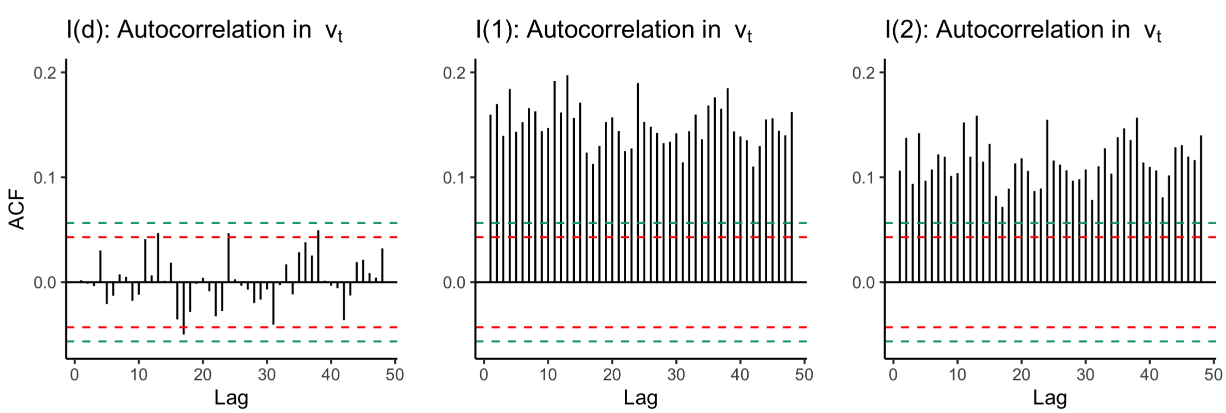

Figure 5: Estimated autocorrelation function of the prediction errors, together with 5% (red) and 1% (blue) confidence bands.

Finally, figure 5 plots the estimated autocorrelation function for the one-step ahead forecast errors of the fractional UC model and the two benchmarks. Misspecifying the integration order to be 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.

7 Conclusion

This paper introduces a novel unobserved components model allowing the trend to be fractionally integrated. The model encompasses the bulk of UC models in the literature, allows for richer long-run dynamics beyond integer-integrated specifications, and for a data-dependent specification of the trend. To applied researchers, the model offers a robust, flexible, and data-driven method for signal extraction. It neither requires prior assumptions about the integration order, nor the choice of any tuning parameter. To econometric theorists, the new asymptotic results provide a starting point for the assessment of more sophisticated models: while the theory can be expected to carry over seamlessly to (identified) univariate models with ARFIMA trends and multivariate models when cointegration is ruled out, multivariate fractionally cointegrated models as considered by Hartl and Jucknewitz (2023) pose an interesting and empirically relevant challenge to future research.

Acknowledgments

This research has partly been conducted at Princeton University, the University of Regensburg, and the Institute for Employment Research Nuremberg. Support through project 356439312 financed by the German Research Foundation is gratefully acknowledged.

References

Andersen, T., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics* **61**(1): 43–76.

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.

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.

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

Christensen, B. J. and Nielsen, M. Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting, *Journal of Econometrics* **133**(1): 343–371.

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.

Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter, *The Review of Economics and Statistics* **100**(5): 831–843.

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

Hartl, T. and Jucknewitz, R. (2023). Multivariate fractional components analysis, *Journal of Financial Econometrics* **21**(3): 880–914.

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.

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 Nielsen, M. Ø. (2022). Truncated sum-of-squares estimation of fractional time series models with generalized power law trend, *Electronic Journal of Statistics* **16**: 2884–2946.

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.

Johansen, S. and Nielsen, M. Ø. (2012). A necessary moment condition for the fractional functional central limit theorem, *Econometric Theory* **28**(3): 671–679.

Johansen, S. and Nielsen, M. Ø. (2019). Nonstationary cointegration in the fractionally cointegrated VAR model, *Journal of Time Series Analysis* **40**(4): 519–543.

Maasoumi, E. and McAleer, M. (2008). Realized volatility and long memory: An overview, *Econometric Reviews* **27**(1-3): 1–9.

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. (2007). Local Whittle analysis of stationary fractional cointegration and the implied-realized volatility relation, *Journal of Business & Economic Statistics* **25**(4): 427–446.

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.

Perron, P. and Wada, T. (2009). Let's take a break: Trends and cycles in US real GDP, *Journal of Monetary Economics* **56**(6): 749–765.

Proietti, T. and Maddanu, F. (2024). Modelling cycles in climate series: The fractional sinusoidal waveform process, *Journal of Econometrics* **239**(1): 105299.

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. (2010). Exact local Whittle estimation of fractional integration with unknown mean and time trend, *Econometric Theory* **26**(2): 501–540.

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.

A Online Appendix: Proof of theorem 4.1

Proof of theorem 4.1. Theorem 4.1 holds if the objective function (11) 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 (10), 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

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

meaning that the probability limit only exists for $d - d_0 > -1/2$, while it does not exist in the opposite case, in which the function increases without bound. $\tilde{v}_t(\theta)$ denotes the untruncated forecast error

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

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 ∞-dimensional 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 (A.2) by letting the dimension of the

t-dimensional coefficient vector go to infinity. Hence, while the truncated forecast errors in (10) are non-ergodic, the untruncated errors (A.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 (A.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 (A.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

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

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 $\kappa_3 \in (0, 1/2)$

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

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 (A.3) and (A.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 (10) 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 (10) 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 (10) and (A.2) to be

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

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 (A.5) that the WLLN for stationary and ergodic processes carries over from $\tilde{v}_{t+1}(\theta)$ to $v_{t+1}(\theta)$

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

(A.6) can be generalized to uniform convergence by showing that a WLLN also holds for the supremum of the absolute gradient, which yields (A.4). From (A.3) and (A.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 (A.4) and (A.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 (A.5), I first show that

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

where $\phi_{\eta,j}(\theta,t)$ is $O((1+\log j)^3 j^{\max(-d+d_0,-\zeta)-1})$ for $j > t$, and $O((1+\log(t+1))^2(t+1)^{\max(-d+d_0,-\zeta)-1})$ for $j \leq t$, whereas $\phi_{\epsilon,j}(\theta,t)$ is $O((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1})$ for $j \leq t$, 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 (A.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

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

where the coefficients are given by $\phi_{1,\eta,j}(\theta,t) = \sum_{k=0}^{t}\tau_k(\theta,t)\pi_{j-k}(d-d_0)$, 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, (C.1), lemma C.2, and $j > t$, the coefficients in (A.8) are $\phi_{1,\eta,t} = O((1+\log j)^2 j^{\max(-d+d_0,-\zeta)-1})$, and $\phi_{1,\epsilon,t} = O((1+\log j)^3 j^{\max(-d,-\zeta)-1})$.

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

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

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

For the third term in (A.7), by lemma C.3

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

By lemma C.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 C.1 together with Johansen and Nielsen (2010, lemma B.4). Hence, $\phi_{3,\eta,j}(\theta,t) = -\sum_{k=0}^{\min(j,t)}\left(\sum_{m=t+1}^{\infty}r_{\tau,k,m}(\theta)\right)\pi_{j-k}(d-d_0)$ is $O\left((1+\log(t+1))^2(t+1)^{\max(-d+d_0,-\zeta)-1}\right)$ whenever $j \leq t$, whereas for $j > t$ it is $O\left((1+\log j)^3 j^{\max(-d+d_0,-\zeta)-1}\right)$.

Similarly, for $j \leq t$, the $\phi_{3,\epsilon,j}(\theta,t) = \sum_{k=0}^{\min(j,t)}\left(\sum_{m=t+1}^{\infty}r_{\tau,k,m}(\theta)\right)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d)$ are $O\left((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\right)$, and for $j > t$ they are $O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right)$.

Together, (A.8), (A.9), (A.10) and the rates established below prove (A.7).

(A.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

$$\begin{aligned}
\mathrm{E}[(\tilde{v}_{t+1}(\theta)-v_{t+1}(\theta))^2] &= \sum_{j=1}^{\infty}\left[\phi_{\eta,j}^2(\theta,t)\,\mathrm{E}(\eta_{t+1-j}^2)+\phi_{\epsilon,j}^2(\theta,t)\,\mathrm{E}(\epsilon_{t+1-j}^2)\right]\\
&= \sum_{j=1}^{t} O\left((1+\log(t+1))^4(t+1)^{2\max(-d+d_0,-\zeta)-2}\right) \qquad \text{(A.11)}\\
&\quad + \sum_{j=t+1}^{\infty} O\left((1+\log(t+1))^8(t+1)^{2\max(-d+d_0,-\zeta)-2}\right) = o(1),
\end{aligned}$$

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 $\theta \in \Theta_3(\kappa_3)$. (A.5) follows directly. From the law of large numbers for stationary and ergodic processes, (A.6) follows immediately.

(A.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 (A.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

$$\begin{aligned}
\frac{\partial Q(y,\theta)}{\partial \theta_{(l)}} &= \frac{2}{n}\sum_{t=1}^{n} v_t(\theta)\frac{\partial v_t(\theta)}{\partial \theta_{(l)}},\\
\frac{\partial v_t(\theta)}{\partial \theta_{(l)}} &= \sum_{j=1}^{t-1}\frac{\partial \tau_j(\theta,t)}{\partial \theta_{(l)}}\xi_{t-j}(d) + \sum_{j=0}^{t-1}\tau_j(\theta,t)\frac{\partial \xi_{t-j}(d)}{\partial \theta_{(l)}},
\end{aligned} \qquad \text{(A.12)}$$

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

$$\sup_{(d_1,d_2,\nu,\varphi)\in\tilde{\Theta}} \left| \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\partial^k \Delta_+^{d_1}}{\partial d_1^k} \sum_{m=0}^{\infty} \tilde{\tau}_{i,m}(\theta) z_{i,t-m}(\theta) \right] \left[\frac{\partial^l \Delta_+^{d_2}}{\partial d_2^l} \sum_{m=0}^{\infty} \tilde{\tau}_{j,m}(\theta) z_{j,t-m}(\theta) \right] \right| = \begin{cases} O_p(1) & \text{for } a > 0, \\ O_p((\log n)^{1+k+l} n^{-a}) & \text{for } a \leq 0, \end{cases} \tag{A.13}$$

$i,j = 1,2$, $k,l = 1,2,...$, as shown by Nielsen (2015, lemma B.3). Now, note that by lemmas C.2 and C.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 $\sum_{j=0}^{t-1} \tau_j(\theta,t)L^j a(L,\varphi_0)$ 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 (A.12) satisfy the conditions for (A.13) with $a > 0$. By (A.13), 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, (A.6) holds uniformly in $\theta \in \Theta_3(\kappa_3)$. As this holds for any κ_3, this proves (A.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 (11), together with (10), it holds that

$$\begin{aligned} Q(y,\theta) =& \frac{1}{n} \sum_{t=1}^{n} \left[\sum_{j=0}^{t-1} \tau_j(\theta,t)\xi_{t-j}(d) \right]^2 \geq \frac{1}{n} \sum_{t=1}^{n} \left(\Delta_+^{d-d_0} \sum_{j=0}^{t-1} \tau_j(\theta,t)\eta_{t-j} \right)^2 \\ &+ \frac{2}{n} \sum_{t=1}^{n} \left(\Delta_+^{d-d_0} \sum_{j=0}^{t-1} \tau_j(\theta,t)\eta_{t-j} \right) \left(\Delta_+^{d} \sum_{j=0}^{t-1} \tau_j(\theta,t)c_{t-j} \right), \end{aligned} \tag{A.14}$$

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 (A.14), by lemma C.2 $\sum_{j=0}^{t-1}|\tau_j(\theta,t)| < \infty$, and by assumption 3 and lemma C.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 (A.13)

$$\sup_{\theta\in\Theta_2(\kappa_2,\kappa_3)}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right]\left[\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\right]\right| = O_p(1). \tag{A.15}$$

Next, consider the first term in (A.14), for which one has by lemma C.3

$$\begin{aligned}\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j} &= \Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta)\eta_{t-j} + \Delta_+^{d-d_0}\sum_{j=1}^{t-1}\left(\sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta)\right)\eta_{t-j}\\ &= \Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j} + r_{\eta,t}(\theta),\end{aligned} \tag{A.16}$$

where

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

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 C.2 and C.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 $r_{\eta,t}(\theta)$

$$r_{\eta,t}(\theta) = \Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j + \Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}, \qquad \alpha_j^* = -\sum_{i=j+1}^{\infty}\alpha_i, \tag{A.18}$$

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 (A.16)

$$\Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j} = \Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta) + \Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}, \tag{A.19}$$

where $\tau_j^*(\theta) = -\sum_{i=j+1}^{\infty}\tau_i(\theta)$, and $\sum_{j=0}^{\infty}\tau_j(\theta) = O(1)$ by lemma C.2. By (A.16), (A.18), and (A.19), it follows for the first term in (A.14) that

$$\frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\Big)^2 \geq \frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)^2 \tag{A.20}$$

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

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

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

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

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\left(\Delta_{+}^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}\right)\left(\Delta_{+}^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\right)\right] \tag{A.25}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\left(\Delta_{+}^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j\right)\left(\Delta_{+}^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\right)\right]. \tag{A.26}$$

From (A.13), it immediately follows that (A.22) to (A.26) 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 (A.20) asymptotically dominates (A.21), so that the rate of convergence of (A.14) will depend solely on (A.20). The asymptotic probability limit of the first term (A.20) 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 (A.20)

$$\frac{1}{n}\sum_{t=1}^{n}\left(\Delta_{+}^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\right)^2\geq\frac{1}{n}\sum_{t=N+1}^{n}\left(w_t^2+2w_tu_t\right). \tag{A.27}$$

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 (A.27) satisfies

$$\sup_{\theta\in\Theta_2(\kappa,\kappa)}\left|\frac{1}{n}\sum_{t=n^{\alpha}+1}^{n}w_t^2-\sigma_{\eta,0}^2\left(\sum_{j=0}^{\infty}\tau_j(\theta)\right)^2\sum_{j=0}^{n^{\alpha}-1}\pi_j^2(d-d_0)\right|\xrightarrow{p}0, \tag{A.28}$$

as $n\to\infty$, and by Nielsen (2015, lemma A.3) the latter sum is bounded from below $\sum_{j=0}^{n^{\alpha}-1}\pi_j^2(d-d_0)\geq 1+K\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}$ 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 (A.20), (A.27), and (A.28) 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 (A.14), (A.15), and (A.16) 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

$$\Pr\left(\inf_{d\in D_2(\kappa_2,\bar{\kappa}_3),\nu\in\Sigma_\nu,\varphi\in\Phi} Q(y,\theta) > K\right) \geq 1-\delta, \quad \text{for all } T \geq T_2, \tag{A.29}$$

and (A.29) 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 (A.14) above, the second term in (A.14), by the same argument with respect to absolute summability of the coefficients as for (A.15), is now

$$\frac{1}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right)\left(\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\right) = O_p\left(1+\log(n)n^{d_0-d-1}\right), \tag{A.30}$$

for all $\theta \in \Theta_1(\kappa_1)$ by (A.13) 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 (A.14) will asymptotically diverge at a faster rate than the second term above. To see this, note that the decomposition of the first term in (A.14) into $\Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j}$ and $r_{\eta,t}(\theta)$ in (A.16) and (A.17) above also applies in $\Theta_1(\kappa_1)$. Consequently, the Beveridge-Nelson decompositions in (A.18) and (A.19) also hold for $\theta \in \Theta_1(\kappa_1)$. Again, the decomposition in (A.20) to (A.26) applies, however the terms in (A.22) to (A.26) 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 (A.20) asymptotically dominates all other terms in (A.21) to (A.26) 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 (A.20). It follows from the Cauchy-Schwarz inequality that

$$n^{2(d-d_0)} \sum_{t=1}^{n} \left(\Delta_+^{d-d_0} \eta_t \sum_{j=0}^{\infty} \tau_j(\theta) \right)^2 \geq \left(n^{d-d_0-1/2} \sum_{t=1}^{n} \Delta_+^{d-d_0} \eta_t \sum_{j=0}^{\infty} \tau_j(\theta) \right)^2, \tag{A.31}$$

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 $\theta \in \Theta_1(\kappa_1)$ is to show that (A.31) weakly converges and is bounded away from zero. As noted by Hualde and Nielsen (2022), the natural way of showing this would be to prove weak convergence of the right-hand side of (A.31). However, note that $\Delta_+^{d-d_0}\eta_t \sim I(d_0 - d)$, and $d_0 - d \geq 1/2 + \kappa_1$ for an arbitrarily small $\kappa_1 > 0$, so that for a functional CLT to hold, one requires $q \geq \max(2, 1/(d_0 - d - 1/2))$ finite moments of η_t (see Johansen and Nielsen; 2012). This condition becomes very strong once $d_0 - d$ is arbitrarily close to 1/2. A much weaker condition is proposed by Johansen and Nielsen (2019, lemma 2): There, it is shown that

$$n^{2(d-d_0)} \sum_{t=1}^{n} \left(\Delta_+^{d-d_0} \eta_t \sum_{j=0}^{\infty} \tau_j(\theta) \right)^2 \geq \left(\frac{\pi^2}{4} + O(n^{-1}) \right) n^{2(d-d_0-1)} \quad \sum_{t=1}^{n} \left(\Delta_+^{d-d_0-1} \eta_t \sum_{j=0}^{\infty} \tau_j(\theta) \right)^2,$$

where the term on the right-hand side has memory $1 + d_0 - d \geq 3/2 + \kappa_1$. Consequently, no strong moment condition is required. Note that Johansen and Nielsen (2019) assume the existence of eight moments, but this is not required for establishing the above limit. Hence, assumption 1 does not need to be strengthened. By Johansen and Nielsen (2019, lemma 3), the right-hand side of the above inequality is bounded away from zero in probability, and (A.20) is $O_p(n^{2(d_0-d-1/2)})$. Since $d_0 - d \geq 1/2 + \kappa_1$ for all $d \in D_1(\kappa_1)$, it holds that for any $K > 0$ and all $\kappa_1 > 0$

$$\Pr\left(\inf_{d \in D_1(\kappa_1), \nu \in \Sigma_\nu, \varphi \in \Phi} \frac{1}{n} Q(y, \theta) > K \right) \to 1, \quad \text{as } T \to \infty. \tag{A.32}$$

Together, (A.29) and (A.32) prove (A.3). $\square$

B Online Appendix: 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

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

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

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

with $v_t(\theta)$ denoting the prediction error as defined in (10), and its partial derivative as given in (A.12). Denote the normalized, untruncated score

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

with $\tilde{v}_t(\theta)$ as defined in (A.2). As shown in lemma C.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 (B.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, $\mu'\sqrt{n}\frac{\partial \tilde{Q}(y,\theta)}{\partial \theta}\big|_{\theta=\theta_0} = \sqrt{n}\sum_{i=1}^{q+2}\mu_{(i)}\left(\frac{\partial \tilde{Q}(y,\theta)}{\partial \theta}\big|_{\theta=\theta_0}\right)_{(i)} = \frac{2}{\sqrt{n}}\sum_{i=1}^{q+2}\mu_{(i)}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{1,t}+\tilde{h}_{2,t})_{(i)} \xrightarrow{d} \mathrm{N}(0, 4\sigma_{v,0}^2\mu'\Omega_0\mu)$ as $n \to \infty$, with expressions $\tilde{h}_{1,t} = \sum_{j=1}^{\infty}\frac{\partial \tau_j(\theta)}{\partial \theta}\Big|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$, and $\tilde{h}_{2,t} = \sum_{j=0}^{\infty}\tau_j(\theta_0)\frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\Big|_{\theta=\theta_0}$. As the $\tilde{h}_{1,t}$ 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 (B.1). By Johansen and Nielsen (2010, lemma A.3), the Hessian matrix in (B.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 C.7. Therefore, the second derivative in (B.1) can be evaluated at the true value θ_0

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

$k, l = 1, 2, ..., q+2$. By lemma C.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 (B.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 C.9 the second term in (B.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 (B.4) that

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

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 (B.1) for $\sqrt{n}(\hat{\theta}-\theta_0)$ yields the desired result

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

□

C Online Appendix: Auxiliary 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 C.1 (Convergence rates of $\pi_j(d)$, $b_j(\varphi)$, and related vector and matrix entries)**.** *It holds that*

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

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

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

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

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

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

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

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

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 C.1. (C.1) follows by Johansen and Nielsen (2010, lemma B.3) while (C.2) follows by assumption 3. From (C.2), the (i,j)-th entries of $B_{\varphi,t}' B_{\varphi,t}$ are $(B_{\varphi,t}' B_{\varphi,t})_{(i,j)} = \sum_{k=0}^{\min(i,j)-1} b_k(\varphi) b_{k+|i-j|}(\varphi) = O(|i-j|^{-\zeta-1}) \sum_{k=0}^{\min(i,j)-1} b_k(\varphi) = O(|i-j|^{-\zeta-1})$ 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 (C.4) is analogous, 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 $i \neq j$, $(S_{d,t}' S_{d,t})_{(i,i)} = O(1)$.

For $(B_{\varphi,t}' B_{\varphi,t})^{-1}$ and $(B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t})^{-1}$, 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 two Toeplitz matrices $T_t(f_1)$ and $T_t(f_2)$, with symbols given by $f_1(\lambda) = (2\pi)^{-1} \sum_{j=0}^{\infty} \gamma_1(j) e^{i\lambda j}$, $\gamma_1(j) = \sum_{k=0}^{\infty} b_k(\varphi) b_{k+j}(\varphi)$, $f_2(\lambda) = (2\pi)^{-1} \sum_{j=0}^{\infty} \gamma_2(j) e^{i\lambda j}$, and $\gamma_2(j) = \sum_{k=0}^{\infty} [b_k(\varphi) b_{k+j}(\varphi) + \nu \pi_k(d) \pi_{k+j}(d)]$, 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 $1/f_4(\lambda) = (2\pi)^{-1} \sum_{j=0}^{\infty} \gamma_4(j) e^{i\lambda j}$. 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 (C.5) and (C.6).

(C.7) follows immediately from (C.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 (C.8) follows from (C.1) by noting that $(S_{d,t}' s_{t+1})_{(j)} = \sum_{k=0}^{j-1} \pi_k(d) \pi_{t-j+k+1}(d) = O((t-j+1)^{-d-1}) \sum_{k=0}^{j-1} \pi_k(d) = O((t-j+1)^{-d-1})$. □

Lemma C.2 (Convergence rates of $\tau_j(\theta,t)$)**.** *For the coefficients* $\tau_j(\theta,t)$ *as defined in* (10) *and below, it holds that*

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

Proof of Lemma C.2. To prove (C.9), consider $\tau_j(\theta,t)$ as defined in (10) and below

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

The left term in (C.10) is

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

For the coefficients, note $\pi_k(d) = O(k^{-d-1})$, $b_k(\varphi) = O(k^{-\zeta-1})$, $(B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1}_{(k,k)} = O(1)$, 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 (C.1), (C.2), and (C.6). It follows that the first term in (C.11) is $O\left(k^{\max(-d,-\zeta)-1}\right)$, while the second term is $\sum_{i=1}^{k-1} O\left(i^{\max(-d,-\zeta)-1} (k-i)^{\max(-d,-\zeta)-1}\right) = O\left((1+\log k) k^{\max(-d,-\zeta)-1}\right)$. This follows from Johansen and Nielsen (2010, lemma B.4), who show $\sum_{i=1}^{k-1} i^{\max(-d,-\zeta)-1} (k-i)^{\max(-d,-\zeta)-1} = O((1+\log k) k^{\max(-d,-\zeta)-1})$. Analogously, $\sum_{i=k+1}^{t} O\left(i^{\max(-d,-\zeta)-1} (i-k)^{\max(-d,-\zeta)-1}\right) = O((k+1)^{\max(-d,-\zeta)-1} \sum_{i=k+1}^{t} (i-k)^{\max(-d,-\zeta)-1}) = O((k+1)^{\max(-d,-\zeta)-1})$. Therefore

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

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

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

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

$\square$

Lemma C.3 (Convergence of $\tau_j(\theta, t)$ as $t \to \infty$). *For the coefficients $\tau_j(\theta, t)$ as defined in* (10) *and below, it holds that*

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

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

Proof of Lemma C.3. To prove (C.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

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

with t-vectors $\beta_t = (b_t(\varphi) \cdots b_1(\varphi))'$ 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

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

with the block entries

$$R_3 = [(1 + \beta_t'\beta_t + \nu + \nu s_t's_t) - (\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)(B'_{\varphi,t}\beta_t + \nu S'_{d,t}s_t)]^{-1},$$

$$R_2 = -R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t + \nu S'_{d,t}s_t),$$

$$R_1 = R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t + \nu S'_{d,t}s_t)(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta).$$

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

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

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

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

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. By (C.17), $R_{2_{(j)}} = O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$. Finally, for R_1, by (C.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

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

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 (C.1) and (C.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

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

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

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

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

$$[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))R_1 S_{d,t}']_{(j)} = \sum_{i=j}^{t}[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))R_1]_{(i)}\pi_{i-j}(d)$$

$$= O\big((1+\log(t+1))^2(t+1)^{-\min(d,\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{-\min(d,\zeta)-1}\big).$$

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 (C.14).

□

Lemma C.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* (10) *and below, it holds that*

$$\frac{\partial \tau_j(\theta,t)}{\partial d} = O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \tag{C.19}$$

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

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

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

Proof of Lemma C.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

$$\dot{S}_{d,t} = \frac{\partial S_{d,t}}{\partial d} = \begin{bmatrix} 0 & \dot{\pi}_1(d) & \cdots & \dot{\pi}_{t-1}(d) \\ 0 & 0 & \cdots & \dot{\pi}_{t-2}(d) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \dot{B}_{\varphi_{(l)},t} = \frac{\partial B_{\varphi,t}}{\partial \varphi_{(l)}} = \begin{bmatrix} 0 & \dot{b}_1(\varphi_{(l)}) & \cdots & \dot{b}_{t-1}(\varphi_{(l)}) \\ 0 & 0 & \cdots & \dot{b}_{t-2}(\varphi_{(l)}) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

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

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

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

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

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

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

For the first term, note that by (C.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 (C.12) it follows for the first terms in (C.24) that

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

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 (C.24), by (C.6) and (C.25)

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

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

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

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 (C.24) is bounded by $O\left((1 + \log j)^4 j^{\max(-d,-\zeta)-1}\right)$. Turning to the second term in (C.24), by (C.12)

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

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 (C.24) is

$$
[(\dot{\pi}_1(d) \cdots \dot{\pi}_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)} = O\left((1 + \log j)^2 j^{\max(-d,-\zeta)-1}\right). \qquad (C.29)
$$

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

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

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

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

$$[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}]_{(j)} == O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right), \quad \text{(C.32)}$$

and the proof is analogous to (C.25) besides one log-factor. By (C.6) and (C.32)

$$[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)]_{(j)} = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \tag{C.33}$$

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

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

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

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

$$\frac{\partial\tau_j(\theta,t)}{\partial\varphi_{(l)}} = \nu[(\dot{b}_1(\varphi_{(l)})\cdots\dot{b}_t(\varphi_{(l)}))\Xi_t(\theta)S'_{d,t}]_{(j)} \tag{C.35}$$

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

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 (C.35), analogously to (C.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 (C.13)

$$[(\dot{b}_1(\varphi_{(l)})\cdots\dot{b}_t(\varphi_{(l)}))\Xi_t(\theta)S_{d,t}]_{(j)} = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right), \tag{C.37}$$

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

$$\begin{aligned}&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})]_{(j)}\\&\quad = O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right),\end{aligned} \tag{C.38}$$

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

$$[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t} + B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta)]_{(j)}$$
$$= O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right). \tag{C.39}$$

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

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

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

Lemma C.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*

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

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

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

$$\begin{aligned} R_{5_{(j)}} =& [(b_{t+1}(\varphi) - \pi_{t+1}(d))\left(R'_2S'_{d,t} + R_3s'_t\right)]_{(j)} \\ &+ [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\left(R_2s'_t + R_1S'_{d,t}\right)]_{(j)}, \end{aligned}$$

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 C.1, and R_1, R_2, R_3 as stated below (C.16). The partial derivative of $R_{5_{(j)}}$ w.r.t. the l-th entry $\theta_{(l)}$ is thus given by

$$\frac{\partial R_{5_{(j)}}}{\partial\theta_{(l)}} = \left[\frac{\partial(b_{t+1}(\varphi) - \pi_{t+1}(d))}{\partial\theta_{(l)}}\left(R'_2S'_{d,t} + R_3s'_t\right)\right]_{(j)} \tag{C.42}$$
$$+ \left[\left(\frac{\partial(b_1(\varphi) - \pi_1(d))}{\partial\theta_{(l)}}\cdots\frac{\partial(b_t(\varphi) - \pi_t(d))}{\partial\theta_{(l)}}\right)\left(R_2s'_t + R_1S'_{d,t}\right)\right]_{(j)} \tag{C.43}$$
$$+ \left[(b_{t+1}(\varphi) - \pi_{t+1}(d))\left(R'_2\frac{\partial S'_{d,t}}{\partial\theta_{(l)}} + R_3\frac{\partial s'_t}{\partial\theta_{(l)}}\right)\right]_{(j)} \tag{C.44}$$
$$+ \left[((b_1(\varphi) - \pi_1(d))\cdots(b_t(\varphi) - \pi_t(d)))\left(R_2\frac{\partial s'_t}{\partial\theta_{(l)}} + R_1\frac{\partial S'_{d,t}}{\partial\theta_{(l)}}\right)\right]_{(j)} \tag{C.45}$$

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

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

As noted in the proof of lemma C.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 (C.42) and (C.43) can be derived analogously to the proof of lemma C.3. This yields that (C.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 (C.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 C.3. Thus, it holds that (C.44) is $O((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1})$, while (C.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 (C.46) and (C.47), note that $R_3 = O(1)$ as shown in (C.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

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

which is $O(1)$. For the partial derivatives of $R_{2_{(j)}}$, consider

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

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

where the first term in (C.48) is $O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$ by (C.17) and by $\partial R_3/\partial\theta_{(l)} = O(1)$. For the second term in (C.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 (C.17). Together with $\partial\Xi_t(\theta)/\partial\theta_{(l)} =$

$-\Xi_t(\theta)[(\partial/\partial\theta_{(l)})(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})]\Xi_t(\theta)$, (C.22) and (C.23), it follows that

$$\begin{aligned}&\left\{(\beta'_tB_{\varphi,t}+\nu s'_tS_{d,t})\Xi_t(\theta)\left[\frac{\partial}{\partial\theta_{(l)}}\left(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t}\right)\right]\right\}_{(j)}\\&=O\left((1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1}\right).\end{aligned}$$

Finally, using (C.6), one obtains

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

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

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

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

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

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

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

For the partial derivatives of R_1, note that

$$\frac{\partial R_{1_{(i,j)}}}{\partial\theta_{(l)}}=-\frac{\partial R_{2_{(i)}}}{\partial\theta_{(l)}}\left[(\beta'_tB_{\varphi,t}+\nu s'_tS_{d,t})\Xi_t(\theta)\right]_{(j)}-R_{2_{(i)}}\left[(\beta'_tB_{\varphi,t}+\nu s'_tS_{d,t})\frac{\partial\Xi_t(\theta)}{\partial\theta_{(l)}}\right]_{(j)}\tag{C.53}$$

$$-R_{2_{(i)}}\left[\left(\beta'_t\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}}+\frac{\partial\beta'_t}{\partial\theta_{(l)}}B_{\varphi,t}+\frac{\partial\nu}{\partial\theta_{(l)}}s'_tS_{d,t}+\nu\frac{\partial s'_t}{\partial\theta_{(l)}}S_{d,t}+\nu s'_t\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\right)\Xi_t(\theta)\right]_{(j)}.\tag{C.54}$$

From (C.17) and (C.52), the first term in (C.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 (C.50) and the convergence rate of $R_{2_{(i)}}$ as derived in the proof of lemma C.3, the second term in (C.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 (C.51), it follows that (C.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

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

With (C.52) at hand, it follows directly for (C.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 (C.1) and (C.2), it follows that (C.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 (C.47), it follows from (C.52) and (C.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 (C.1) and (C.2), it follows that (C.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 (C.41) that

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

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

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

Proof of lemma C.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 (B.2), (B.3) into (C.56) and using (A.12) yields

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

so that it remains to be shown that all four terms in (C.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 (C.57), by plugging in $h_{1,t}$ and $\tilde{h}_{1,t}$

$$\begin{aligned}
&\frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{1,t} - h_{1,t}) \\
&\quad = \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=1}^{t-1} \frac{\partial \tau_j(\theta, t)}{\partial \theta}\bigg|_{\theta=\theta_0} \left(\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0)\right) && \text{(C.58)} \\
&\quad + \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=1}^{t-1} \left(\frac{\partial \tau_j(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta, t)}{\partial \theta}\bigg|_{\theta=\theta_0} \right) \tilde{\xi}_{t-j}(d_0) && \text{(C.59)} \\
&\quad + \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=t}^{\infty} \frac{\partial \tau_j(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0} \tilde{\xi}_{t-j}(d_0). && \text{(C.60)}
\end{aligned}$$

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

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

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

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

For (C.61), I use $\sum_{j=0}^{\infty}\eta^2_{\min(s,t)-j}\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0)=O_p(|t-s|^{\max(-d_0,-\zeta)-1})$ for $t\neq s$, else $O_p(1)$, see lemma C.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 (C.1) together with lemma C.4. This yields the upper bound for (C.61)

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

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

$$
\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.
$$

$$\times \left(\sum_{k=0}^{j+|t-s|} \tau_k(\theta_0) \sum_{l=0}^{j+|t-s|-k} a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0) \right) \Bigg]$$

$$\leq K \sum_{j=1}^{\min(s,t)-1} (1+\log j)^3 j^{-\min(d_0,\zeta)-1}(1+\log(j+|t-s|))^3(j+|t-s|)^{-\min(d_0,\zeta)-1}.$$

Furthermore, by lemma C.4

$$\mathrm{E}\left[\sum_{j=0}^{\infty} \epsilon_{-j}^2 \left(\sum_{k=0}^{t-1} \frac{\partial \tau_k(\theta,t)}{\partial \theta}\bigg|_{\theta=\theta_0} \sum_{l=0}^{j} a_l(\varphi_0)\pi_{j+t-k-l}(d_0) \right) \right.$$

$$\left. \times \left(\sum_{k=0}^{s-1} \frac{\partial \tau_k(\theta,s)}{\partial \theta'}\bigg|_{\theta=\theta_0} \sum_{l=0}^{j} a_l(\varphi_0)\pi_{j+s-k-l}(d_0) \right) \right]$$

$$\leq K \sum_{j=1}^{\infty} (1+\log(t+j))^6 (t+j)^{\max(-d_0,-\zeta)-1}(1+\log(s+j))^6(s+j)^{\max(-d_0,-\zeta)-1},$$

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

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

$$\sum_{s,t=1}^{n} \mathrm{E}\left[\left(\sum_{j=t}^{\infty} \epsilon_{t-j}^2 \left(\sum_{k=0}^{j} \tau_k(\theta_0) \sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0) \right) \right. \right.$$

$$\left. \times \left(\sum_{k=0}^{t-1} \frac{\partial \tau_k(\theta,t)}{\partial \theta}\bigg|_{\theta=\theta_0} \sum_{l=0}^{j-t} a_l(\varphi_0)\pi_{j-k-l}(d_0) \right) \right)$$

$$\times \left(\sum_{j=s}^{\infty} \epsilon_{s-j}^2 \left(\sum_{k=0}^{j} \tau_k(\theta_0) \sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0) \right) \right.$$

$$\left. \left. \times \left(\sum_{k=0}^{s-1} \frac{\partial \tau_k(\theta,s)}{\partial \theta'}\bigg|_{\theta=\theta_0} \sum_{l=0}^{j-s} a_l(\varphi_0)\pi_{j-k-l}(d_0) \right) \right) \right]$$

$$\leq K \sum_{s,t=1}^{n} (1+\log t)^9 t^{2\max(-d_0,-\zeta)-1}(1+\log s)^9 s^{2\max(-d_0,-\zeta)-1} = O(1).$$

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

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

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

$$= \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=0}^{t-1} \tau_j(\theta_0,t) \left(\frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\bigg|_{\theta=\theta_0} - \frac{\partial \xi_{t-j}(d)}{\partial \theta}\bigg|_{\theta=\theta_0} \right) \tag{C.64}$$

$$+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}\left(\tau_j(\theta_0)-\tau_j(\theta_0,t)\right)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0} \tag{C.65}$$

$$+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}. \tag{C.66}$$

For (C.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 C.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 (C.66) is $o_p(1)$.

For (C.65), note that $\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}\left(\tau_j(\theta_0)-\tau_j(\theta_0,t)\right)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\big|_{\theta=\theta_0}$ together with $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS. By lemma C.3, it holds that $\tau_j(\theta_0)-\tau_j(\theta_0,t)=O((1+\log t)^2t^{\max(-d_0,-\zeta)-1})$. Hence $\tilde{\xi}_t(d)$ are $\sum_{j=0}^{t-1}\left(\tau_j(\theta_0)-\tau_j(\theta_0,t)\right)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\big|_{\theta=\theta_0}=O_p((1+\log t)^2t^{\max(-d_0,-\zeta)})$, and (C.65) is $o_p(1)$.

For (C.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 (C.64) w.r.t. d is

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

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

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

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1} \epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\qquad\qquad \left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\right]\\
&\qquad \times \mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\frac{\tau_k(\theta_0,t)}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\frac{\tau_k(\theta_0,s)}{s+j-k}\right)\right. \qquad\qquad (\mathrm{C.69})\\
&\qquad\quad +\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\frac{\partial\pi_{j+t-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right)\\
&\qquad\qquad \left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\frac{\partial\pi_{j+s-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right)\right]\\
&+\sum_{s,t=1}^{n}\mathrm{E}\left\{\left[\sum_{j=t}^{\infty}\eta^2_{t-j}\tau_j(\theta_0)\sum_{k=0}^{t-1}\frac{-\tau_k(\theta_0,t)}{j-k}+\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\right.\\
&\qquad\qquad \left.\times\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\frac{\partial\pi_{j-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right)\right]\\
&\qquad\left[\sum_{j=s}^{\infty}\eta^2_{s-j}\tau_j(\theta_0)\sum_{k=0}^{s-1}\frac{-\tau_k(\theta_0,s)}{j-k}+\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\qquad\qquad \left.\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\frac{\partial\pi_{j-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right)\right]\right\}.
\end{aligned}
\tag{C.70}
$$

For (C.68), 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 C.2. For the other terms in (C.68), it holds that $\mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\frac{1}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\frac{1}{s+j-k}\right)\right] \leq K\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}$, together with

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

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

$$K\sum_{t=1}^{n}\Bigg[\sum_{s=1,\ s<t}(t-s)^{\max(-d_0,-\zeta)-1}\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}$$
$$+\sum_{s=t+1}^{n}(s-t)^{\max(-d_0,-\zeta)-1}\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}$$
$$+\sum_{j=0}^{\infty}(1+\log(t+j))^4(t+j)^{-2}\Bigg]\leq K\sum_{t=1}^{n}\left[(1+\log t)t^{-1+\kappa}\right]\leq Kn^{\kappa},$$

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 (C.69), by lemmas C.1 and C.2, the first expectation is bounded

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

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

$$\sum_{s,t=1}^{n}\mathrm{E}\Bigg\{\Bigg[\sum_{j=t}^{\infty}\eta^2_{t-j}\tau_j(\theta_0)\sum_{k=0}^{t-1}\frac{-\tau_k(\theta_0,t)}{j-k}+\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)$$
$$\times\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\Bigg]\Bigg[\sum_{j=s}^{\infty}\eta^2_{s-j}\tau_j(\theta_0)\sum_{k=0}^{s-1}\frac{-\tau_k(\theta_0,s)}{j-k}$$
$$+\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\Bigg]\Bigg\}$$
$$=\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^3j^{\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^3j^{\max(-d_0,-\zeta)-2}\right)\right)$$
$$+\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^7j^{2\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^7j^{2\max(-d_0,-\zeta)-2}\right)\right)$$
$$+\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^3j^{\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^7j^{2\max(-d_0,-\zeta)-2}\right)\right)$$
$$+\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^7j^{2\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^3j^{\max(-d_0,-\zeta)-2}\right)\right),$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Thus, the first term in (C.71) is $o_p(1)$. For the second term in (C.71), note that by lemma C.3, $\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t)) \leq K\sum_{j=1}^{t-1}\sum_{k=t+1}^{\infty}(1+\log k)^2(1+\log(k-j))^2 k^{\max(-d_0,-\zeta)-1}(k-j)^{\max(-d_0,-\zeta)-1} \leq K\sum_{j=1}^{t-1}(1+\log t)^2 t^{\max(-d_0,-\zeta)-1}(1+\log(t-j))^2(t-j)^{\max(-d_0,-\zeta)} \leq K(1+\log t)^2 t^{-1}\sum_{j=1}^{t-1} j^{\max(-d_0,-\zeta)}(t-j)^{\max(-d_0,-\zeta)}(1+\log(t-j))^2 \leq K(1+\log t)^5 t^{\max(-d_0,-\zeta)-1}$, 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 (C.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^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)\right. \qquad \text{(C.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{(C.76)}\\
&\quad\times\mathrm{E}\left[\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)\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]\\
&+\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.\\
&\quad\left.\times\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) \qquad \text{(C.77)}\\
&\quad\times\left(\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)\right.\\
&\quad\left.\left.\times\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}
$$

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

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

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 (C.72) and (C.73), expressions (C.75) and (C.76) are $O(1)$. Also analogously to (C.74), by lemmas C.1, C.2, and C.4, (C.77) is bounded from above by

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

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

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

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

$$\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left.\frac{\partial\xi_{t-j}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0))\tau_j(\theta_0,t)\right]^2\right\}\\
&=\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1}\eta_{\min(s,t)-j}^2\left(\sum_{k=1}^{j}\frac{1}{k}\tau_{j-k}(\theta_0,\min(s,t))\right)\right.\\
&\qquad\left.\times\left(\sum_{k=1}^{j+|t-s|}\frac{1}{k}\tau_{j+|t-s|-k}(\theta_0,\max(s,t))\right)\right]\\
&\qquad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\right.\\
&\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\right]
\end{aligned}\tag{C.79}$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0,\min(s,t))\sum_{l=0}^{j-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\right)\right.\\
&\qquad\left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0,\max(s,t))\sum_{l=0}^{j+|t-s|-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j+|t-s|-k-l}(\varphi_0)\right)\right] \qquad (C.80)\\
&\qquad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\right.\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\right]\\
&+\sum_{s,t=1}^{n}\mathrm{E}\left[\left(\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{t-1-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\right)\right.\right.\\
&\qquad\qquad\left.\times\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right) \qquad (C.81)\\
&\qquad\times\left(\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{s-1-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\right)\right.\\
&\qquad\qquad\left.\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right)\right],
\end{aligned}
$$

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 C.2, the first term in (C.79) is

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

for $t\neq s$, and $O(1)$ otherwise. By lemmas C.1 and C.2, the first term of (C.80) is

$$
\begin{aligned}
&\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0,\min(s,t))\sum_{l=0}^{j-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\right)\right.\\
&\qquad\left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0,\max(s,t))\sum_{l=0}^{j+|t-s|-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j+|t-s|-k-l}(\varphi_0)\right)\right]\\
&\qquad=O(|t-s|^{\max(-d_0,-\zeta)-1}), \qquad (C.82)
\end{aligned}
$$

for $t \neq s$, and $O(1)$ otherwise. The second term in (C.79) and (C.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 (C.72), (C.73), (C.75) and (C.76), it holds that (C.79) and (C.80) are $O(1)$. Finally, (C.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 (C.78) is $o_p(1)$. For the second term, by lemma C.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 (C.71). Moreover $\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{2,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\tilde{\xi}_{t-j}(d_0)=o_p(1)$. The third term in (C.71) is

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

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

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

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

Lemma C.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* (11) *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 C.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 (A.12),

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

Boundedness in probability of the third partial derivatives then follows from (A.13) 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 (A.13). As can be seen from lemma C.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 C.1, those for the first partial derivatives are contained in the proof of lemma C.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. It follows that

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

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

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

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

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

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

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

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

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

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

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 (A.13). By (A.13), $B_n = \sup_{d\in D_3,\nu\in\Sigma_\nu,\varphi\in N_\delta(\varphi_0)} \left|\frac{\partial^3 Q(y,\theta)}{\partial\theta^{(3)}}\right| = O_p(1)$. $\square$

Lemma C.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 $j \geq t$.

Proof of lemma C.8. Consider

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

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

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

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

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 (C.1), lemma C.4, and assumption 2, the first term on the right-hand side (C.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 C.5, (C.1), and assumption 3, the first term in (C.87) is

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

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

$$=\sum_{j=1}^{t-1}O((1+\log t)^5 t^{\max(-d_0,-\zeta)-1})(\eta_{t-j}+\epsilon_{t-j}) + \sum_{j=t}^{\infty}O((1+\log j)^7 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j}.$$

For the second term in (C.87), by lemma C.4, (C.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 (C.88), (C.89) are non-zero only for the derivative w.r.t. d. For (C.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 C.2, Johansen and Nielsen (2010, lemma B.3), and assumption 3. For the first term in (C.89), by lemmas C.2, C.3, Johansen and Nielsen (2010, lemma B.3), and assumption 3

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

while for the second term in (C.89), by lemma C.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 C.8. □

Lemma C.9. *For $v_t(\theta)$ as defined in* (10) *and $\tilde{v}_t(\theta)$ as defined in* (A.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 C.9. The proof is analogous to the proof of lemma C.6 and thus is only summarized briefly. Note that there exists a constant $0<K<\infty$ such that

$$\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}=O\left((1+\log k)^K k^{\max(-d,-\zeta)-1}\right), \tag{C.90}$$
$$\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}-\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}=O\left((1+\log t)^K t^{\max(-d,-\zeta)-1}\right). \tag{C.91}$$

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

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

$$+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\left(\tilde{h}_{5,t_{(i,j)}}-h_{5,t_{(i,j)}}\right)+\frac{1}{n}\sum_{t=1}^{n}h_{5,t_{(i,j)}}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)$$
$$+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\left(\tilde{h}_{5,t_{(j,i)}}-h_{5,t_{(j,i)}}\right)+\frac{1}{n}\sum_{t=1}^{n}h_{5,t_{(j,i)}}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right),$$

and thus the different terms in (C.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 (C.92), by plugging in $\tilde{h}_{3,t_{(i,j)}}$, $h_{3,t_{(i,j)}}$

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

The latter two terms in (C.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 (C.90) and (C.91), it holds that $\sum_{k=t}^{\infty}\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}\tilde{\xi}_{t-k}(d_0)$ 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 (C.93) are also $o_p(1)$. In contrast, the first term in (C.93) is not a MDS. However, by the same proof as for (C.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, (C.93) is $o_p(1)$. For the third term in (C.92), by plugging in $\tilde{h}_{4,t_{(i,j)}}$, $h_{4,t_{(i,j)}}$

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

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 C.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 C.3. The second term can be shown to be $o_p(1)$ analogously to (C.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 (C.92), similarly to (C.93) and (C.94)

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

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

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

with $l = 3, 4, 5$. For $l = 3$, (C.96) only differs from (C.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 (C.90). For $l = 4$, (C.96) is almost identical to (C.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 (C.78) (Hualde and Robinson; 2011, lemma 4). For $l = 5$, (C.96) is again almost identical to (C.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 C.4 the derivative again only adds a log-factor to the convergence rate of $\tau_k(\theta, t)$. Thus, it follows directly from (C.71) and (C.78), together with (C.90) and Hualde and Robinson (2011, lemma 4), that (C.96) is $o_p(1)$. The two remaining terms in (C.92) are $o_p(1)$ by (C.95) and (C.96), as i, j can be interchanged. $\square$

D Online Appendix: Monte Carlo results

This section examines the finite sample properties of the proposed estimation methods via a Monte Carlo study. I first consider the prototypical fractional UC model

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

where $\eta_t \sim \text{NID}(0, 1)$, $\epsilon_t \sim \text{NID}(0, \nu)$ are uncorrelated. At the end of the section, generalizations of (D.1) are considered and compared with the baseline scenario. An overview about the estimation properties of QML and CSS estimator is given when (i) innovations are correlated; (ii) deterministic terms are present; (iii) unanticipated structural breaks in the trend occur; and (iv) unanticipated outliers occur at the end of the sample. All results and replication files are available at `https://github.com/tobiashartl/fracUCM`.

For the prototypical model, I set the cyclical coefficients in (D.1) to $b_{1,0} = 1.6$, $b_{2,0} = -0.8$ to generate persistent cyclical patterns. To allow for a better comparison of CSS and QML, $\sigma^2_{\eta,0} = 1$ is fixed and assumed known in the QML optimization, so that estimation is carried out over $\theta = (d, \nu, b_1, b_2)'$ both for CSS and QML. The study varies over $n \in \{100, 200, 300\}$, integration order $d_0 \in \{0.75, 1.00, 1.75\}$, and variance ratio

$$\nu_0 \in \left\{1, \frac{n^{-1}\sum_{t=1}^n \sum_{j=0}^{t-1} \pi_j^2(-d_0)}{\sum_{j=0}^\infty a_j^2(\varphi_0)} r^{-1}\right\}, \qquad r \in \{1, 10, 30\}, \tag{D.2}$$

with $\varphi_0 = (b_{1,0}, b_{2,0})$, and $a(L, \varphi_0) = \sum_{j=0}^\infty a_j(\varphi_0) L^j = (1 - b_{1,0}L - b_{2,0}L^2)^{-1}$. Choices for n and d_0 cover empirically relevant sample sizes and integration orders in macroeconomics and finance, and allow for a comparison with the $I(1)$ UC model when $d_0 = 1$. The choice for ν_0 is justified as follows: Trivially, setting $\nu_0 = \sigma^2_{\epsilon,0}/\sigma^2_{\eta,0} = 1$ assigns equal variation to long- and short-run innovations. By its non-stationary nature, the trend then dominates the overall variance of y_t, i.e. $\operatorname{Var}(x_t)/\operatorname{Var}(c_t) = O(t^{2d_0-1})$, which constitutes a favorable scenario for estimating d_0. At contrast, letting ν_0 depend on d_0 and φ_0 controls for the diverging variance ratio: The numerator in (D.2) is the mean variance of the trend, hence $\operatorname{Var}(c_t) = \operatorname{Var}(a(L,\varphi_0)\epsilon_t) = \nu_0 \sum_{j=0}^{\infty} a_j^2(\varphi_0) = r^{-1}n^{-1}\sum_{t=1}^{n}\sum_{j=0}^{t-1}\pi_j^2(-d_0)$ is proportional to the mean variance of x_t, and $n^{-1}\sum_{t=1}^{n}\operatorname{Var}(x_t)/\operatorname{Var}(c_t) = r$. This fixes the variance ratio of x_t and c_t (instead of $\sigma^2_{\epsilon,0}/\sigma^2_{\eta,0}$), and the lower r, the weaker the relative contribution of the trend, and the less favorable the scenario for estimating d_0 and x_t.

Each simulation consists of 1000 replications. For the QML estimator, the trend is initialized with variance zero, whereas the cycle is initialized with its long-run variance. Once the prediction error variance satisfies $\left|\frac{\operatorname{Var}_\theta(v_{t+1}(\theta)|y_1,\dots,y_t) - \operatorname{Var}_\theta(v_t(\theta)|y_1,\dots,y_{t-1})}{\operatorname{Var}_\theta(v_t(\theta)|y_1,\dots,y_{t-1})}\right| < 0.01$, the optimization switches to the steady-state Kalman filter. Both CSS and QML are initialized by first evaluating their objective function at a large grid, and choosing the grid point with the lowest value for the objective function as starting value. Benchmarks are (i) the exact local Whittle estimator of Shimotsu and Phillips (2005) with $m = \lfloor n^j \rfloor$ Fourier frequencies, $j \in \{.50, .60, .70\}$; (ii) the $I(1)$ UC model (setting $d = 1$); (iii) the approximate fractional UC model of Hartl and Jucknewitz (2022), which approximates the fractional differencing operator by an ARMA(3, 3) polynomial. Parameter estimates are compared by root mean squared error (RMSE) and bias. Moreover, the coefficients of determination R^2_x and R^2_c from regressing x_t and c_t on their estimates are reported.

Table D.1 shows RMSE and bias for the estimated integration orders. As can be expected, bias and RMSE decrease both in n and r, and are significantly smaller as compared to the nonparametric Whittle estimators. The difference is particularly striking when the signal of the trend is drowned by the cycle (i.e. r_0 small / ν_0 large), which biases the Whittle estimates towards zero, while the estimates for the fractional UC model are hardly affected. Noticeably, for the fractional UC model QML is slightly superior to CSS, and using ARMA approximations as suggested by Hartl and Jucknewitz (2022) yields estimates that are close to the exact fractional UC models in terms of RMSE and bias.

Tables D.2 and D.3 detail RMSE and bias for ν_0 and the autoregressive parameters. In addition to the fractional UC model, the tables also display the estimation results for the $I(1)$ UC benchmark. For $b_{1,0}$ and $b_{2,0}$, CSS and QML estimates for the fractional UC model behave equally well, while those from the approximate fractional UC model and the integer-integrated benchmarks come with a slightly higher RMSE, particularly for $d_0 = 0.75$. Interestingly, the estimates for ν_0 show a clear dominance of QML over CSS, as the latter comes with a much higher RMSE and strong, positive bias. QML at contrast does not appear biased. A violation $d_0 \neq 1$ in integer-integrated models has a rather small effect on the estimates for the autoregressive coefficients, but a comparably strong effect on the estimate for ν_0, both for CSS and QML, thus shifting variation from trend to cycle.

Table D.4 compares the estimates for x_t and c_t for the different models by regressing the respective Kalman smoother-based estimates on the true trend and cycle and reporting the coefficient of determination. As can be seen from table D.4, differences between the coefficients of determination are almost negligible for CSS and QML estimator of the fractional UC model, with the latter exhibiting slightly larger R^2's. 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 shows a higher R^2 than the integer-integrated benchmarks. Integer-integrated UC models provide a good approximation for $d_0 = 0.75$, but perform poorly for $d_0 = 1.75$, even when the cycle dominates (ν_0 large).

The above results are supported by a variety of further simulations all of which can be found at `https://github.com/tobiashartl/fracUCM`. In short, it is found that (i) allowing for **correlated innovations** increases the overall estimation uncertainty, however d_0 is still estimated reliably by the different estimators for the fractional UC model. At contrast, the covariance matrix of long- and short-run innovations turns out to be notoriously difficult to estimate, particularly for the CSS estimator. In simulations with non-integer d_0, integer-integrated UC models frequently converged to corner solutions where $\hat{Q}$ was almost singular. This indicates the (almost) perfect correlation between the innovations found in many applications may be an artifact generated by a misspecification of the trend.

Moreover, (ii) including a **deterministic trend** downward-biases the estimates for d_0 both via CSS and QML whenever $r = 1$. This is somewhat natural, as $r = 1$ generates an observable time series with a very smooth trend and strong cyclical patterns. The fractional UC model then attributes some of the variation of the stochastic trend to the deterministic term, so that the estimate for the stochastic trend can take up some of the variation generated by the cyclical component. This goes in line with a downward-biased estimate for d_0, however the bias vanishes for higher n and r.

As shown by Perron and Wada (2009), a deterministic trend with an unanticipated break is hard to distinguish from a random walk with drift. Hence, scenario (iii) adds an **unanticipated trend break** to the setup. This increases the RMSE relative to scenario (ii) for all parameters and simulations, however no additional bias is induced by the break. Moreover, the relative ordering of the different estimators, as well as the overall conclusions from the baseline scenario and scenario (ii) remain unaffected.

Many trend-cycle decompositions, in particular the HP filter, suffer from end-of-sample distortions (Hamilton; 2018). To assess the robustness of the fractional UC model, the last simulation adds (iv) an **unanticipated outlier** with variance $10\nu_0$ to a single y_t, $t \in \{\lfloor 0.9T \rfloor, \lfloor 0.9T \rfloor + 1, ..., T\}$. Naturally, adding an outlier increases the RMSE for all parameters, and the higher ν_0, the higher the increase in the RMSE. Moreover, the unanticipated outlier adds a slight downward-bias to the QML and CSS estimate $\hat{d}$, however the bias is much smaller as compared to the nonparametric benchmarks and vanishes as n increases. All other conclusions from the baseline scenario remain the same.

				RMSE						bias					
n	r	d_0	ν_0	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}_{ARMA}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.70}$	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}_{ARMA}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.70}$
100		.75	1	.27	.13	.42	.64	.41	.57	.01	-.02	.14	-.62	-.36	.52
		1	1	.33	.13	.13	.68	.46	.40	.03	-.03	-.02	-.65	-.42	.33
		1.75	1	.21	.12	.18	.55	.43	.16	.03	-.03	-.00	-.50	-.38	-.06
	1	.75	.65	.23	.12	.40	.60	.37	.50	-.00	-.02	.11	-.58	-.32	.45
		1	3.82	.48	.18	.23	.85	.61	.52	.13	-.06	-.01	-.83	-.58	.45
		1.75	1054	.64	.41	.33	1.58	1.36	.36	-.16	-.18	-.11	-1.57	-1.35	-.23
	10	.75	.07	.15	.11	.22	.34	.21	.18	-.00	-.01	.03	-.24	-.10	.10
		1	.38	.23	.11	.11	.54	.34	.28	-.00	-.03	-.02	-.49	-.28	.21
		1.75	105	.48	.24	.22	1.31	1.13	.34	-.07	-.07	-.09	-1.29	-1.11	-.22
	30	.75	.02	.13	.10	.12	.27	.19	.14	-.01	-.02	-.00	-.12	-.04	.04
		1	.13	.20	.11	.10	.38	.24	.18	-.00	-.02	-.01	-.30	-.15	.10
		1.75	35.13	.41	.19	.20	1.14	.98	.32	-.07	-.05	-.04	-1.12	-.96	-.21
200		.75	1	.13	.09	.18	.62	.57	.14	-.02	-.02	.02	-.60	-.56	.03
		1	1	.17	.08	.08	.60	.56	.15	.00	-.02	-.01	-.57	-.55	-.10
		1.75	1	.16	.07	.13	.39	.44	.24	.03	-.01	.04	-.34	-.41	-.22
	1	.75	.93	.14	.08	.19	.61	.56	.14	-.01	-.02	.02	-.59	-.55	.03
		1	7.59	.42	.15	.63	.87	.83	.23	.14	-.04	.38	-.86	-.82	-.18
		1.75	5872	.60	.36	.34	1.60	1.58	.96	-.22	-.15	-.20	-1.60	-1.58	-.95
	10	.75	.09	.11	.07	.10	.32	.26	.10	-.01	-.01	.00	-.26	-.22	.01
		1	.76	.14	.08	.07	.56	.52	.14	-.00	-.02	-.01	-.53	-.50	-.08
		1.75	587	.46	.17	.18	1.35	1.35	.89	-.09	-.05	-.03	-1.34	-1.34	-.88
	30	.75	.03	.09	.07	.09	.23	.17	.10	-.01	-.01	.00	-.13	-.10	.01
		1	.25	.10	.07	.06	.39	.35	.11	-.01	-.01	-.01	-.34	-.33	-.04
		1.75	196	.40	.13	.16	1.20	1.21	.83	-.06	-.03	-.08	-1.19	-1.20	-.82
300		.75	1	.10	.07	.12	.51	.60	.22	-.02	-.01	.01	-.49	-.59	-.19
		1	1	.10	.07	.06	.45	.58	.27	-.01	-.01	-.01	-.42	-.57	-.26
		1.75	1	.13	.06	.11	.24	.43	.28	.03	-.01	.05	-.18	-.41	-.27
	1	.75	1.14	.11	.07	.13	.53	.62	.22	-.01	-.01	.01	-.51	-.61	-.20
		1	11.37	.33	.13	.90	.79	.88	.49	.10	-.03	.74	-.77	-.88	-.48
		1.75	16099	.56	.31	.46	1.55	1.63	1.27	-.24	-.12	-.31	-1.54	-1.63	-1.27
	10	.75	.11	.09	.06	.06	.25	.31	.10	-.01	-.01	.00	-.19	-.29	-.05
		1	1.14	.10	.07	.06	.47	.60	.29	-.01	-.01	-.01	-.44	-.59	-.27
		1.75	1610	.44	.14	.16	1.28	1.42	1.14	-.09	-.04	-.06	-1.27	-1.41	-1.13
	30	.75	.04	.09	.06	.06	.19	.19	.08	-.01	-.01	.00	-.09	-.15	-.02
		1	.38	.07	.06	.05	.32	.44	.18	-.01	-.01	-.01	-.27	-.42	-.15
		1.75	537	.38	.11	.16	1.14	1.29	1.05	-.08	-.02	-.01	-1.12	-1.28	-1.04

Table D.1: Root mean squared errors (RMSE) and bias for the integration order estimates.

n	r	d_0	ν_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}_{ARMA}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}_{1_{QML}}$	$\hat{b}_{1_{ARMA}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}_{2_{QML}}$	$\hat{b}_{2_{ARMA}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
100		.75	1	83463	.31	.42	77406	.41	.13	.10	.12	.19	.14	.13	.09	.12	.18	.13
		1	1	70231	.30	.30	28468	.31	.13	.09	.09	.12	.09	.12	.08	.09	.12	.09
		1.75	1	575	.28	.32	24535	273	.12	.11	.25	.20	.19	.11	.09	.23	.19	.08
	1	.75	.65	80051	.27	.31	49436	.35	.14	.11	.13	.21	.17	.14	.11	.12	.21	.16
		1	3.82	164888	.75	.76	54935	.77	.09	.08	.08	.09	.08	.09	.07	.08	.09	.08
		1.75	1054	295220	153	148	213664	313	.07	.07	.07	.07	.07	.07	.07	.07	.07	.09
	10	.75	.07	39234	.10	.10	14238	.15	.40	.28	.36	.86	.84	.22	.20	.20	.38	.38
		1	.38	39021	.21	.19	13768	.20	.16	.13	.14	.15	.13	.15	.11	.13	.16	.13
		1.75	105	111571	18.52	15.09	73987	197	.07	.07	.07	.08	.11	.07	.07	.07	.08	.12
	30	.75	.02	15956	.05	.06	24704	.08	.63	.50	.66	1.35	1.44	.26	.28	.26	.54	.44
		1	.13	7165	.14	.12	14749	.15	.22	.24	.25	.27	.22	.19	.17	.17	.21	.19
		1.75	35.13	67152	6.87	6.70	50229	261	.08	.07	.08	.09	.12	.08	.07	.07	.10	.11
200		.75	1	9550	.25	.26	74011	.32	.08	.06	.06	.14	.09	.08	.06	.06	.14	.09
		1	1	32786	.25	.21	18973	.22	.07	.06	.06	.07	.06	.08	.06	.06	.07	.06
		1.75	1	21.97	.19	.23	35412	373	.07	.07	.10	.25	.24	.08	.06	.12	.21	.11
	1	.75	.93	45783	.24	.24	61783	.31	.08	.06	.06	.14	.09	.08	.06	.06	.14	.09
		1	7.59	211140	4.62	1.59	70367	2.18	.05	.05	.08	.05	.05	.05	.05	.07	.05	.05
		1.75	5872	445062	3386	600	116294	13792	.04	.05	.04	.04	.06	.05	.05	.04	.04	.08
	10	.75	.09	31647	.09	.08	28358	.18	.19	.15	.16	.40	.52	.15	.14	.14	.27	.40
		1	.76	24604	.23	.18	2201	.19	.08	.06	.06	.08	.06	.08	.06	.06	.08	.06
		1.75	587	207156	71.21	60.16	30899	241	.05	.04	.04	.05	.08	.05	.04	.05	.06	.10
	30	.75	.03	462	.05	.05	29010	.08	.36	.32	.39	1.14	1.22	.21	.21	.20	.55	.41
		1	.25	414	.15	.11	14433	.12	.13	.10	.10	.12	.09	.13	.10	.11	.14	.09
		1.75	196	107572	27.63	26.51	7811	303	.05	.04	.04	.06	.12	.05	.05	.05	.07	.13
300		.75	1	1659	.20	.19	43706	.28	.06	.05	.05	.11	.08	.06	.05	.05	.11	.07
		1	1	725	.22	.17	5239	.18	.06	.05	.05	.06	.05	.06	.05	.05	.06	.05
		1.75	1	10.34	.15	.21	4933	448	.06	.05	.06	.30	.24	.06	.04	.08	.21	.12
	1	.75	1.14	2188	.23	.21	62138	.29	.06	.05	.05	.11	.07	.06	.05	.05	.11	.07
		1	11.37	160382	2.82	1.77	106537	1.88	.04	.04	.10	.04	.04	.04	.04	.07	.04	.04
		1.75	16099	470811	13285	1806	83164	27793	.04	.08	.04	.04	.08	.04	.05	.04	.04	.06
	10	.75	.11	144	.08	.07	17543	.20	.18	.10	.11	.27	.45	.13	.10	.10	.24	.40
		1	1.14	654	.23	.19	1084	.19	.06	.05	.05	.06	.05	.06	.05	.04	.06	.05
		1.75	1610	299401	165	134	1618	703	.04	.04	.04	.04	.08	.04	.04	.04	.05	.09
	30	.75	.04	31528	.06	.05	18430	.09	.21	.20	.24	.94	1.05	.18	.18	.18	.48	.44
		1	.38	2.05	.16	.11	1.02	.11	.07	.07	.06	.07	.06	.08	.07	.06	.08	.06
		1.75	537	155494	59.27	46.85	533	254	.04	.04	.04	.05	.10	.04	.04	.04	.06	.10

Table D.2: Root mean squared errors (RMSE) for the other parameter estimates.

n	r	d_0	ν_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}_{ARMA}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}_{1_{QML}}$	$\hat{b}_{1_{ARMA}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}_{2_{QML}}$	$\hat{b}_{2_{ARMA}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
100		.75	1	8657	.00	.03	15815	.19	-.03	-.02	-.04	.08	.08	.03	.01	.03	-.06	-.06
		1	1	7826	.03	-.05	3048	.02	-.03	-.02	-.02	.02	.03	.02	.01	.01	-.02	-.02
		1.75	1	37.02	-.01	.04	1866	150	-.00	-.02	-.07	-.10	-.17	-.03	.01	.08	.04	-.03
	1.0	.75	.65	7284	.02	.02	8468	.16	-.03	-.02	-.04	.07	.09	.02	.01	.03	-.05	-.08
		1	3.82	35292	-.14	-.14	8829	.05	-.02	-.01	-.02	.02	.03	.02	.01	.01	-.02	-.02
		1.75	1054	118564	25.24	-21.74	54801	204	-.01	-.02	-.02	-.01	-.02	.00	.01	.01	.01	-.05
	10.0	.75	.07	1897	-.00	-.01	1296	.01	-.04	-.01	-.07	.32	.33	-.07	-.06	-.07	.01	-.02
		1	.38	2972	.03	-.03	1183	.01	-.03	-.03	-.02	.02	.03	.01	.01	-.00	-.01	-.02
		1.75	105	24840	7.65	-2.22	8069	111	-.00	-.02	-.02	-.04	-.07	-.00	.01	.01	.05	-.08
	30.0	.75	.02	518	.00	.00	2297	.01	-.15	-.10	-.26	.88	.91	-.07	-.06	-.08	-.17	-.06
		1	.13	495	.00	-.03	1168	-.01	-.02	-.02	-.02	.03	.01	-.04	-.04	-.06	.02	.04
		1.75	35.13	8422	2.47	-.15	5954	147	.00	-.02	-.02	-.06	-.09	-.01	.01	.02	.06	-.08
200		.75	1	372	.03	.00	8847	.18	-.01	-.01	-.01	.05	.06	.01	.01	.01	-.05	-.05
		1	1	1660	.10	-.02	650	.01	-.01	-.01	-.01	.01	.01	.01	.01	.01	-.01	-.01
		1.75	1	4.10	.02	.11	1967	244	.00	-.01	-.03	-.09	-.21	-.01	.01	.06	.04	.01
	1.0	.75	.93	2375	.03	.00	7116	.18	-.01	-.01	-.01	.05	.06	.01	.01	.01	-.04	-.05
		1	7.59	49750	.20	.03	9136	.16	-.01	-.01	-.03	.01	.01	.01	.01	.02	-.01	-.01
		1.75	5872	215127	215	-57.78	15729	1765	-.00	-.01	-.01	-.01	-.03	-.00	.01	.01	.01	-.04
	10.0	.75	.09	1068	.02	.01	1081	.06	-.00	-.02	-.02	.02	.22	-.02	.01	-.00	.07	-.13
		1	.76	1035	.10	-.01	139	.01	-.01	-.02	-.01	.01	.01	.01	.02	.01	-.01	-.01
		1.75	587	59376	33.63	-5.97	996	208	.00	-.01	-.01	-.03	-.07	.00	.01	.01	.04	-.08
	30.0	.75	.03	18.56	.00	.00	1879	.01	-.02	-.02	-.09	.66	.67	-.08	-.05	-.06	-.18	-.06
		1	.25	22.40	.05	-.01	774	.01	-.02	-.03	-.01	.02	.01	.01	.03	.01	-.02	-.01
		1.75	196	20082	17.04	-1.74	141	227	.00	-.01	-.01	-.05	-.08	.00	.01	.01	.06	-.09
300		.75	1	67.04	.03	.00	3488	.17	-.01	-.01	-.01	.04	.05	.01	.01	.01	-.03	-.04
		1	1	24.49	.12	-.01	166	.01	-.01	-.01	-.00	.01	.01	.01	.01	.00	-.00	-.01
		1.75	1	1.93	.01	.12	267	325	.00	-.01	-.02	-.08	-.21	-.01	.00	.05	.04	.01
	1.0	.75	1.14	84.58	.04	-.00	5875	.18	-.01	-.01	-.01	.04	.05	.01	.01	.01	-.04	-.04
		1	11.37	29508	.08	-.73	15376	.12	-.00	-.00	-.06	.00	.01	.00	.00	.04	-.00	-.01
		1.75	16099	237664	1441	226	-4223	5285	.00	-.01	.00	-.00	-.03	-.00	.00	.01	.01	-.04
	10.0	.75	.11	5.90	.02	.01	925	.09	-.02	-.02	-.02	.01	.24	.01	.02	.01	.03	-.20
		1	1.14	25.71	.11	-.02	44.23	.01	-.01	-.01	-.00	.01	.01	.01	.01	.00	-.00	-.01
		1.75	1610	106266	71.73	-16.15	-1588	556	.00	-.00	-.00	-.03	-.07	-.00	.00	.00	.03	-.07
	30.0	.75	.04	999	.01	.01	894	.02	.01	-.00	-.04	.44	.54	-.05	-.02	-.03	-.13	-.11
		1	.38	.16	.07	-.01	.09	.01	-.01	-.02	-.00	.01	.01	.01	.02	.00	-.00	-.01
		1.75	537	33899	35.92	-3.85	-533	231	.00	-.00	-.00	-.04	-.08	-.00	.00	.00	.05	-.09

Table D.3: Bias for the other parameter estimates.

				Trend					Cycle				
n	r	d_0	ν_0	R^2_{CSS}	R^2_{QML}	R^2_{ARMA}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$	R^2_{CSS}	R^2_{QML}	R^2_{ARMA}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$
100		.75	1	.56	.59	.59	.52	.58	.78	.80	.80	.75	.80
		1	1	.61	.64	.65	.62	.66	.88	.90	.90	.89	.90
		1.75	1	.78	.79	.80	.78	.74	.96	.96	.96	.95	.78
	1	.75	.65	.50	.53	.53	.46	.52	.83	.85	.85	.81	.84
		1	3.82	.75	.78	.78	.75	.79	.77	.79	.79	.77	.79
		1.75	1054	.99	.99	.99	.91	.80	.55	.60	.58	.25	.04
	10	.75	.07	.89	.90	.91	.86	.91	.16	.16	.14	.09	.09
		1	.38	.89	.91	.91	.88	.91	.39	.40	.39	.38	.40
		1.75	105	.94	.95	.95	.92	.70	.84	.89	.89	.80	.24
	30	.75	.02	.83	.83	.83	.80	.83	.33	.35	.33	.26	.27
		1	.13	.83	.85	.85	.82	.85	.62	.65	.65	.63	.65
		1.75	35.13	.91	.91	.92	.91	.70	.89	.92	.92	.89	.40
200		.75	1	.63	.65	.65	.58	.64	.84	.85	.85	.82	.84
		1	1	.69	.71	.70	.69	.72	.92	.93	.92	.93	.93
		1.75	1	.82	.83	.82	.82	.75	.98	.98	.98	.97	.78
	1	.75	.93	.62	.64	.64	.57	.63	.85	.86	.86	.82	.85
		1	7.59	.87	.88	.88	.87	.88	.79	.80	.80	.80	.81
		1.75	5872	1	1	1	.95	.81	.60	.63	.61	.24	.01
	10	.75	.09	.91	.92	.92	.87	.92	.20	.21	.19	.11	.12
		1	.76	.93	.94	.94	.92	.94	.59	.61	.61	.59	.61
		1.75	587	.96	.97	.97	.97	.74	.91	.94	.94	.92	.25
	30	.75	.03	.86	.87	.87	.83	.85	.45	.48	.48	.33	.39
		1	.25	.88	.89	.89	.88	.89	.76	.77	.77	.77	.78
		1.75	196	.94	.94	.94	.94	.77	.94	.96	.96	.94	.41
300		.75	1	.67	.68	.68	.62	.67	.87	.87	.87	.84	.86
		1	1	.74	.75	.74	.74	.76	.94	.94	.93	.94	.95
		1.75	1	.85	.86	.83	.85	.74	.98	.98	.96	.98	.79
	1	.75	1.14	.69	.70	.70	.65	.69	.86	.86	.86	.83	.85
		1	11.37	.91	.91	.91	.91	.92	.80	.81	.81	.80	.81
		1.75	16099	1	1	1	.97	.83	.62	.64	.62	.24	.01
	10	.75	.11	.92	.93	.93	.88	.93	.27	.29	.27	.15	.15
		1	1.14	.94	.95	.95	.94	.95	.68	.69	.69	.68	.69
		1.75	1610	.97	.98	.98	.97	.77	.94	.96	.95	.94	.25
	30	.75	.04	.87	.88	.88	.84	.87	.54	.56	.55	.41	.49
		1	.38	.90	.91	.91	.90	.91	.81	.82	.82	.82	.82
		1.75	537	.95	.95	.96	.95	.76	.96	.97	.97	.96	.42

Table D.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.