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

Tobias Hartl*[a,b]

[a]University of Regensburg, 93053 Regensburg, Germany
[b]Institute for Employment Research, 90478 Nuremberg, Germany

November 2022

Abstract. This paper provides a data-driven solution to the specification of the long-run dynamics in trend-cycle decompositions. A novel state space model of form $y_t = x_t + c_t$ is introduced, allowing the unobserved trend $x_t \sim I(d)$ to be fractionally integrated of order d, whereas c_t represents an unobserved stationary cyclical component. The new model encompasses the two major specifications in the literature that either assume $x_t \sim I(1)$ in spirit of the Beveridge-Nelson decomposition, or $x_t \sim I(2)$ as for the Hodrick-Prescott filter. As d can take any value on the positive real line, the new model allows for intermediate solutions between integer-integrated specifications and thus for richer long-run dynamics. Trend and cycle can be estimated via the Kalman filter, for which a closed-form solution is provided. The integration order d is treated as unknown and is estimated jointly with the other model parameters via the conditional sum-of-squares estimator. The paper derives the asymptotic theory for parameter estimation under relatively mild assumptions, showing the conditional sum-of-squares estimator to be consistent and asymptotically normally distributed. While the proofs are carried out for a prototypical model, the asymptotic theory carries over to generalizations allowing for deterministic terms and correlated innovations, but also to (quasi-) maximum likelihood estimation. An application to annual CO2 emission reveals a smooth trend component starting to exhibit an inverted U-shape, together with cyclical CO2 emissions that are closely coupled to the business cycle.

Keywords. Unobserved components, trend-cycle decomposition, state space models, Kalman filter, long memory, conditional sum-of-squares estimator

JEL-Classification. C22, C51, Q54

*Corresponding author: Department of Economics and Econometrics, University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany, email: tobias1.hartl@ur.de
The author thanks the participants of the International Conference on Computational and Financial Econometrics 2021, the Annual Meeting of the German Statistical Society 2021, the Annual Meeting of the German Economic Society 2021, the European Meeting of the Econometric Society 2021, the Annual Symposium of the Society for Nonlinear Dynamics and Econometrics 2022, the department seminar at the University of Regensburg, the Vienna Workshop on High Dimensional Time Series in Macroeconomics and Finance 2022, and the Annual Meeting of the German Statistical Society 2022. Support through the project 356439312 financed by the German Research Foundation (DFG) is gratefully acknowledged.

1 Introduction

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

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

This paper aims to bridge these gaps by introducing a novel UC model that specifies the stochastic trend component x_t as a fractionally integrated process of order $d \in \mathbb{R}_+$, denoted as $x_t \sim I(d)$. It allows for random walk trend components (as suggested among others by Beveridge and Nelson; 1981; Harvey; 1985; Morley et al.; 2003) for $d = 1$, but also includes quadratic stochastic trend specifications (e.g. those of Clark; 1987; Hodrick and Prescott; 1997; Oh et al.; 2008) for $d = 2$. As the integration order d can take any value on the positive real line and enters the model as an unknown parameter to be estimated, the model links integer-integrated specifications seamlessly. By including non-integer d, it allows for even more general patterns of persistence between the integer cases, thus generating long memory. While the UC literature has mostly considered the aforementioned integer-integrated specifications, some generalizations to non-integer integration orders exist: For asymptotically stationary processes (i.e. $d < 1/2$) Chan and Palma (1998, 2006), Palma (2007) and Grassi and de Magistris (2014) consider approximations to long-memory processes in state space form either by truncating the autoregressive or the moving average representation of the fractional filter. Their models have

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

been found valuable for realized volatility modelling (see Ray and Tsay; 2000; Harvey; 2007; Chen and Hurvich; 2006; Varneskov and Perron; 2018) but exclude non-stationary stochastic trends that are indispensable for general UC models. Recently, Hartl and Jucknewitz (2022) studied ARMA approximations to fractionally integrated processes in state space form, also including the non-stationary domain. However, their inference is limited to Monte Carlo studies. Besides the fractional trend, the fractional UC model includes a cyclical component that is allowed to be any short memory process governed by a parameter vector to be estimated. Thus, it encompasses the usual ARMA specification in the UC literature, but also allows for a wider class of processes such as e.g. the exponential model of Bloomfield (1973). Innovations are allowed to deviate from the usual Gaussian white noise assumption in the UC literature.

The second methodological contribution of the paper is to derive the full estimation theory both for the unobserved components and the model parameters. In line with the UC literature, unobserved components are estimated by minimizing the objective function of the Kalman filter. While the literature typically relies on iterative estimates for trend and cycle via the Kalman recursions, I derive an analytical solution to the optimization problem.[2] As iterative and analytical solution to the Kalman filter only differ in their way of computation, both approaches yield the minimum variance linear unbiased estimators for trend and cycle (Durbin and Koopman; 2012, lemma 2). However, the analytical solution is computationally less costly for the fractional UC model. As an additional advantage, it yields a closed-form solution to the objective function of the conditional sum-of-squares (CSS) estimator that is used for the estimation of the integration order d as well as for the other model parameters. Under the comparably weak assumption that long- and short-run shocks are stationary martingale difference sequences, parameter estimates are shown to be consistent. Under the stronger assumption of white noise shocks they can also be shown to be asymptotically normally distributed. The latter assumption is only required when the integration order d is unknown, i.e. it can be avoided in the integer-integrated UC literature where d is assumed to be known. The proofs are complicated by non-ergodicity of the prediction errors and non-uniform convergence of the objective function. The latter is caused by a prediction error that is stationary if the estimate for d is close to the true value, whereas it becomes non-stationary if the estimate is too far off. While all proofs are carried out for the computationally superior conditional sum-of-squares (CSS) estimator, they are shown to carry over seamlessly to the (quasi-) maximum likelihood (QML) estimator that is typically utilized in the UC literature. Furthermore, estimation results are shown to also hold for models with deterministic terms and correlated trend and cycle innovations (e.g. those of Balke and Wohar; 2002; Morley et al.; 2003). The asymptotic results are supported by a Monte Carlo study that finds both CSS and QML estimates to behave consistent. In addition, parameter estimates for the integration order outperform the exact local Whittle estimator of Shimotsu (2010) that is deteriorated by the cyclical fluctuations.

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

An application to CO2 emissions illustrates the benefits from the fractional UC model: Log US annual CO2 emissions are found to be integrated of order around 1.75, which is clearly at odds with integer-integrated models. The resulting trend-cycle decomposition finds evidence for the trend component exhibiting an inverted U-shape, thus supporting the existence of an environmental Kuznets curve as well as the often hypothesized decoupling of economic activity and CO2 emissions in terms of the trend. In contrast, as a glimpse on figure 3 reveals, cyclical CO2 emissions appear to remain coupled to the business cycle, as they exhibit rich pro-cyclical dynamics. Integer-integrated benchmarks fail to grasp these stylized facts.

The remaining paper is organized as follows: Section 2 introduces the fractional UC model, relates it to the literature, and discusses the underlying assumptions. Section 3 considers the estimation of trend and cycle, while section 4 details parameter estimation via the CSS estimator. Generalizations of the fractional UC model are discussed in section 5. Section 6 investigates the finite sample properties of the proposed methods in a Monte Carlo study, while section 7 applies the fractional UC model to CO2 emissions. Section 8 concludes. The proofs for consistency and asymptotic normality are contained in the appendix.

2 Model

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

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

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

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

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

The fractional difference operator Δ_+^d solely depends on the parameter d and controls the memory of x_t. It exhibits a polynomial expansion in the lag operator L of order infinite

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

where the weights $\pi_j(d)$ are determined recursively. The motivation behind (2) and (3) is that the higher d is, the higher the effect of a past shock η_{t-j} on x_t, and the more differencing is required to eliminate the persistent impact of the past shock via (2). For this reason $x_t \sim I(d)$ is said to have long memory whenever $d > 0$ (see Hassler; 2019, for more details). The +-subscript in (2) denotes the truncation of an operator at $t \leq 0$, $\Delta_+^d x_t = \sum_{j=0}^{t-1} \pi_j(d) x_{t-j}$, which reflects the type II definition of fractionally integrated processes (Marinucci and Robinson; 1999).

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

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

The parametric form of $a(L, \varphi)$ is assumed to be known. c_t can for example be an ARMA process as usual in the UC literature, but the specification in general captures a wider class of processes, e.g. the exponential model of Bloomfield (1973).

In what follows, 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 σ-algebra $\mathcal{F}_t = \sigma(\tilde{\xi}_s, s \leq t)$ generated by the $\tilde{\xi}_s = \eta_s + \Delta^{d_0} c_s$ with $s \leq t$, 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.*

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

Assumption 1 allows for conditionally homoskedastic martingale difference sequences (MDS)

η_t and ϵ_t. This is somewhat more general than the UC literature, which typically assumes Gaussian white noise disturbances (e.g. in Harvey; 1985; Morley et al.; 2003). It also encompasses independent (but not necessarily normally distributed) disturbances as considered by Harvey et al. (1994) and Ruiz (1994) for stochastic variance models. The generalization is of great practical importance given the applications of UC models in macroeconomics and finance. However, it comes at the cost that the standard aggregation theory for moving average processes of Granger and Morris (1976), as used in the UC literature to relate UC models to ARIMA models and apply the asymptotic theory thereof (e.g. in Morley et al.; 2003, p. 236), is not applicable. Instead, the full asymptotic theory for parameter estimation has to be derived in section 4, and is inferred from the well established central limit theory for MDS.

The choice of the filtration $\mathcal{F}_t$ is to ensure that the residual in the objective function for parameter estimation asymptotically converges to a MDS: As usual in the UC literature, the objective function for parameter estimation is set up based on the one-step-ahead prediction error of y_{t+1} given the information y_s, $s \leq t$. As taking fractional differences only involves past y_s, $s \leq t$, this information set is the same as the one generated by $\xi_s = \Delta_+^{d_0} y_s = \eta_s + \Delta_+^{d_0} c_s$. As will be shown, the difference between the non-ergodic ξ_s and the ergodic $\tilde{\xi}_s = \eta_s + \Delta^{d_0} c_s$ is asymptotically negligible in probability. Thus, $\mathcal{F}_t$ asymptotically equals the σ-algebra generated by the observable y_s, $s \leq t$. It follows directly that the one-step-ahead prediction error of y_{t+1} given y_s, $s \leq t$ (and thus also given ξ_s, $s \leq t$) asymptotically becomes a MDS when adapted to the filtration $\mathcal{F}_t$. Note that this assumption is not only weaker compared to defining η_t and ϵ_t as MDS in terms of their own natural (and thus finer) filtrations. It also takes into account that only y_s, $s \leq t$ is observable at t, and at the same time ensures the prediction error to asymptotically become a MDS. Therefore, assumption 1 is also more general to Dunsmuir (1979, ch. 3), who in principle considers estimation under a single MDS disturbance but for his "signal observed with noise" model requires signal and noise disturbances to be "at least white noise".

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

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

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

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

(iii) $a(z, \varphi) = \sum_{j=0}^{\infty} a_j(\varphi) z^j$ is continuously differentiable in φ, and the 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 $\sum_{j=0}^{\infty} |b_j(\varphi)|$ is bounded and bounded away from zero. By the Lipschitz condition it holds that

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

The rate for $a_j(\varphi)$ follows directly from assumption 3(ii), while the one for $b_j(\varphi)$ follows by Zygmund (2002, pp. 46 and 71). The convergence rate for the partial derivative $\dot{a}_j(\varphi)$ is an immediate consequence of compactness of Φ and continuity of $\partial a_j(\varphi)/\partial\varphi'$. Assumption 3 imposes some smoothness on the linear coefficients in $a(L,\varphi)$, and thus also on $b(L,\varphi)$. It is satisfied by any stationary and invertible ARMA process. For ARFIMA models, where aggregation of long- and short-run shocks is not a problem, the asymptotic estimation theory is well established under assumptions similar to 1, 2, and 3, see Hualde and Robinson (2011) and Nielsen (2015).

3 Filtering and smoothing

The system introduced in (1), (2), and (4) forms a state space model, with (1) being the measurement equation and (2), (4) being the state equations for trend and cycle.[3] This opens the way for the Kalman filter, a powerful set of algorithms for filtering, predicting, and smoothing the latent components x_t and c_t, but also for parameter estimation. In this section, I derive an analytical solution to the optimization problem of the Kalman filter and smoother. As will become clear at the end of this section, it has two decisive advantages over the usual recursive algorithm for filtering and smoothing: 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 $b_+(L,\varphi)y_t = b(L,\varphi)y_t\mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} b_j(\varphi)y_{t-j}$, where $\mathbb{1}(t \geq 1)$ is the indicator function that takes the value one for positive subscripts of y_{t-j}, else zero. For the optimization problem of the Kalman filter, it will be helpful to define the $t \times t$ difference matrix $S_{d,t}$ and the $t \times t$ coefficient matrix $B_{\varphi,t}$

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

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 difference matrix of Burman and Shumway (2009), and it holds that $S_dS_{-d} = I$ and $S_0 = I$. In the following, I show the closed-form solutions for the updating step of the Kalman filter to be given by

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

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

[3] Section 5 outlines the state space representation and illustrates the dimensions of the system matrices. For further details on state space models and the Kalman filter see Harvey (1990, ch. 3)

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

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

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

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

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

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

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

Here, the first residual in (10) stems from plugging (4) into the measurement equation and solving for ϵ_j, while the second is from solving (2) for η_j. Analogously, the first term in (11) follows from inserting (2) into (1) and solving for η_j, while the second follows from solving (4) for ϵ_j. Constant terms are omitted. As x_t and c_t are estimated based on all observations until period t, it holds that $\hat{x}_{t:1}(y_{t:1}, \psi) = y_{t:1} - \hat{c}_{t:1}(y_{t:1}, \psi)$. Under Gaussian η_t and ϵ_t, the optimization problems in (10) and (11) yield the conditional expectations $\hat{x}_{t:1}(y_{t:1}, \psi) = \mathrm{E}_\psi(x_{t:1}|y_{t:1})$ and $\hat{c}_{t:1}(y_{t:1}, \psi) = \mathrm{E}_\psi(c_{t:1}|y_{t:1})$, see Durbin and Koopman (2012, lemma 1), where the expected value operator $\mathrm{E}_\psi(z_t)$ of an arbitrary random variable z_t denotes that expectation is taken with respect to the distribution of z_t given ψ. If η_t, ϵ_t are not normally distributed, the optimization problems (10) and (11) remain valid. The filtered $\hat{x}_{t:1}(y_{t:1}, \psi)$, $\hat{c}_{t:1}(y_{t:1}, \psi)$ are the projection of $x_{t:1}$ and $c_{t:1}$ on the span of $y_{t:1}$, and are the minimum variance linear unbiased estimators for $x_{t:1}$ and $c_{t:1}$ given the observable information $y_1, ..., y_t$ (Durbin and Koopman; 2012, lemma 2). For $t = n$, $d = 2$, $b(L, \varphi) = 1$, $\lambda = \sigma_\epsilon^2/\sigma_\eta^2$, (10) becomes the HP filter with λ being the tuning parameter. Thus, the HP filter constitutes a special case of the fractional UC model.

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

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

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

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

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

Note that filtered, predicted and smoothed x_t and c_t can be calculated either via the analytical solution above, or recursively by executing the Kalman recursions (see e.g. Harvey; 1990, ch. 3). Both approaches yield identical results and only differ in their way of computation. However, the analytical solution bears two decisive advantages over the traditional recursions: (i) It is computationally superior for fractional trends as in (2). As the state vector of the fractional trend in (2) is of dimension $n-1$, the dimension of the state vector for both trend and cycle is of dimension $m \geq n-1$. Thus, each recursion of the Kalman filter involves multiple multiplications of $(m \times m)$-dimensional covariance and system matrices, and each multiplication needs $2m^3 - m^2$ flops. The analytical solution also requires the costly computation of an $(n \times n)$ inverse, however the underlying matrix is symmetric, positive definite, and thus the Cholesky decomposition can be used to reduce the complexity to $2/3n^3 + 2n^2$ flops per iteration. As $m \geq n - 1$, the analytical solution speeds up the computation considerably. This allows to run the otherwise computationally infeasible Monte Carlo studies in section 6. (ii) The solution allows to derive an objective function for parameter estimation that does not depend on the Kalman recursions and thus is easier to analyze. As usual, the objective function for parameter estimation is set up based on the one-step-ahead prediction error, that is obtained by plugging (8) and (9) into the measurement equation (1). As (8) and (9) only depend on the observable $y_1, ..., y_t$ as well as on the model parameters, the objective function will not depend on a recursive solution for trend and cycle estimates. This greatly simplifies the asymptotic theory for parameter estimation in section 4, as all coefficients and their convergence rates are known.

4 Parameter estimation

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

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

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

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

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

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

As the objective function is proportional to the exponent in the (quasi-) likelihood function, optimization of (16) is similar to (quasi-) maximum likelihood estimation that is typically applied in the state space literature, see e.g. Durbin and Koopman (2012, ch. 7). While the latter allows for a time-varying variance of the prediction error, (16) implicitly assumes a constant prediction error variance. However, as subsection 5.3 discusses in greater detail, the filtered prediction error variance of the fractional UC model converges to its steady state solution at an exponential rate. Thus, (16) and (quasi-) maximum likelihood estimation are asymptotically equivalent. Differences only occur due to a different weighting of prediction errors at the very beginning of the sample. However, (16) is computationally much simpler, as it avoids the Kalman recursions for the prediction error variance. Furthermore, (16) equals the parameter estimates based on the steady-state Kalman filter after some burn-in period, see Harvey (1990, ch. 4.2.2).

While the asymptotic theory for CSS estimation is well established for autoregressive fractionally integrated moving average (ARFIMA) models, see Hualde and Robinson (2011) and Nielsen (2015), only little is known about the asymptotic theory for unobserved components models of such generality. For the sub-class of $I(1)$ UC models with Gaussian white noise shocks η_t and ϵ_t, the asymptotic theory can be inferred from the ARIMA literature (Harvey and Peters; 1990;

Morley et al.; 2003), as for $d = 1$, $\Delta_+^d y_t = \eta_t + \Delta_+^d c_t$ becomes an ARIMA model by the aggregation theory for moving average processes of Granger and Morris (1976). Unfortunately, the latter does no longer hold for MDS η_t and ϵ_t and/or non-integer $d \in \mathbb{R}_+$. Hence, the asymptotic theory for parameter estimation has to be derived from scratch for fractional UC models. While the proofs in this section are carried out for the (simpler) CSS estimator, they are shown to carry over to the traditional (quasi-) maximum likelihood estimator in subsection 5.3. By the encompassing nature of the fractional UC model, the results derived below also hold for CSS and (quasi-) maximum likelihood estimation of all sub-classes of UC models such as e.g. integer-integrated models with MDS shocks.

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

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

Second, even within the asymptotically stationary region of the parameter space, the forecast errors are non-ergodic, as can be seen from (14) and (15): The truncated fractional differencing polynomial Δ_+^d includes more lags as t increases, and thus $\xi_t(d) = \Delta_+^{d-d_0}\eta_t + \Delta_+^d c_t$ is non-ergodic. In addition, $\tau_j(\theta, t)$ in (15) depends on t. Consequently, even for $d_0 - d < 1/2$, a law of large numbers for stationary and ergodic series does not apply directly to $v_t(\theta)$. I tackle this problem by showing the difference between the prediction error in (14), and the untruncated and ergodic $\tilde{v}_t(\theta) = \sum_{j=0}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d)$, 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, while the coefficients $\tau_j(\theta)$ stem from the ∞-vector $(\tau_1(\theta), \tau_2(\theta) \cdots) = \nu(b_1(\varphi) - \pi_1(d), b_2(\varphi) - \pi_2(d), \cdots)(B'_{\varphi,\infty} B_{\varphi,\infty} + \nu S'_{d,\infty} S_{d,\infty})^{-1} S'_{d,\infty}$, and $\tau_0(\theta) = 1$. As $\tilde{v}_t(\theta)$ is stationary and ergodic within the stationary region of the parameter space, it follows that a weak law of large numbers applies to the objective function. The final part of the proof is to strengthen pointwise convergence in probability to weak convergence, which yields the desired result of theorem 4.2.

With a consistent parameter estimator at hand, I next turn to the derivation of the asymptotic distribution of the CSS estimator. For this purpose, the underlying assumptions need to be strengthened.

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

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

Assumption 5. *η_t and ϵ_t are white noise, i.e. $\eta_t \sim$ i.i.d.$(0, \sigma_\eta^2)$, and $\epsilon_t \sim$ i.i.d.$(0, \sigma_\epsilon^2)$.*

Assumption 5 strengthens the distributional assumptions on long- and short-run shocks. It is required, since the asymptotic distribution of the CSS estimator is inferred from a central limit theorem for the score function of the untruncated residual $\tilde{v}_t(\theta)$. Hence, the difference between truncated and untruncated score function has to be asymptotically negligible in probability, as shown in lemma D.6. This requires

$$n^{-1/2} \sum_{t=1}^{n} \eta_t \left(\frac{\partial \Delta^{d-d_0} \eta_t}{\partial d} - \frac{\partial \Delta_+^{d-d_0} \eta_t}{\partial d} \right) \Bigg|_{\theta=\theta_0} = o_p(1), \tag{17}$$

where $\left(\frac{\partial \Delta^{d-d_0} \eta_t}{\partial d} - \frac{\partial \Delta_+^{d-d_0} \eta_t}{\partial d} \right) \Big|_{\theta=\theta_0} = \sum_{j=1}^{\infty} j^{-1} \eta_{t-j} - \sum_{j=1}^{t-1} j^{-1} \eta_{t-j} = \sum_{j=t}^{\infty} j^{-1} \eta_{t-j}$, see Robinson (2006, pp. 135f). Under assumption 5, the latter term is $o_p(1)$, and thus (17) holds. While Robinson (2006, p. 134f) also imposed the white noise assumption to derive the asymptotic distribution of the CSS estimator for ARFIMA models, Hualde and Robinson (2011), Nielsen (2015) and Hualde and Nielsen (2020) were recently able to relax the assumption to MDS disturbances (in terms of their own natural filtration). For the results to carry over, this would require to assume both η_t and ϵ_t to be MDS in terms of their own natural (and thus finer) filtration instead of $\mathcal{F}_t$. Doing so, the term in parentheses in (17) would become measurable, and (17) could be proven analogously to Hualde and Robinson (2011, p. 3170). However, defining η_t and ϵ_t as MDS in terms of their own natural filtration is at odds with only $y_1, ..., y_t$ being observable at period t, and also at odds with the construction of the Kalman filter (that builds on the projections of x_t, c_t onto the span of $y_1, ..., y_t$). Hence, assumption 5 seems to be more appropriate for UC models. For integer-integrated UC models (e.g. the widely applied models of Harvey; 1985; Morley et al.; 2003), assumption 5 is not required, as it only matters for the estimation of d_0. Therefore, it can be omitted when assuming d_0 to be known.

Theorem 4.2. *For the model in* (1), (2), *and* (4), *under assumptions 1 to 5, the estimator* $\hat{\theta}$ *as defined via* (16) *is asymptotically normally distributed, i.e.* $\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} \mathrm{N}(0, \Omega_0^{-1})$ *as* $n \to \infty$, *with variance-covariance matrix* $\Omega_{0_{(i,j)}} = \mathrm{Var}(v_t(\theta_0))\,\mathrm{E}\left(\frac{\partial \tilde{v}_t(\theta)}{\partial \theta_{(i)}}\big|_{\theta=\theta_0} \frac{\partial \tilde{v}_t(\theta)}{\partial \theta_{(j)}}\big|_{\theta=\theta_0}\right)$, $i, j = 1, ..., q+2$.

The proof of theorem 4.2 is contained in Appendix C. As usual, the asymptotic distribution of the CSS estimator is inferred from a Taylor expansion of the score function around θ_0. Analogously to Robinson (2006) and Hualde and Robinson (2011), the normalized score at θ_0 is first shown to be asymptotically equivalent to the score function of the untruncated, stationary and ergodic residual $\sqrt{n}(\partial \tilde{Q}(y,\theta)/\partial\theta)\big|_{\theta=\theta_0} = (2/\sqrt{n}) \sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\partial \tilde{v}_t(\theta)/\partial\theta)\big|_{\theta=\theta_0}$. Next, a UWLLN is shown to hold for the Hessian matrix, so that it can be evaluated at θ_0 in the Taylor expansion, and the difference between the truncated and untruncated Hessian matrix is shown to be asymptotically negligible in probability. Therefore, both score and Hessian matrix in the Taylor expansion can be replaced by their untruncated counterparts. While a weak law of large numbers applies to the untruncated Hessian matrix, a central limit theorem for martingale difference sequences applies to the untruncated score and yields the asymptotic distribution. Finally, while theorem 4.2 does not provide an analytical expression for the covariance matrix of the CSS estimator, it shows the usual estimators via the numerical Hessian or Fisher information matrix to be valid for the estimation of $\mathit{\Omega}_0^{-1}$.

5 Generalizations

One key advantage of fractional unobserved components models is their state space representation: It makes the Kalman filter and smoother applicable, enables (quasi-) maximum likelihood estimation of the model parameters, allows to diffusely initialize the filter and to seamlessly add additional structural components to the model. In addition, several useful methods and generalizations that are beyond the scope of this paper become available, such as frequency-domain optimization, additional observable explanatory variables, time-varying and non-linear models, and mixed frequency among others, see Harvey (1990) for an overview. In this section, I outline some generalizations of the fractional UC model that are of immediate applied relevance: Subsection 5.1 introduces deterministic components to the model, while subsection 5.2 allows for correlated trend and cycle innovations. Subsection 5.3 generalizes parameter estimation to the (quasi-) maximum likelihood estimator. For all three modifications, the asymptotic results in section 4 are shown to carry over. However, before turning to the three generalizations, I first introduce the state space representation of the fractional UC model.

The basic state space representation is of the form

$$y_t = Z\alpha_t + u_t, \tag{18}$$

$$\alpha_t = T\alpha_{t-1} + R\zeta_t, \tag{19}$$

where the states may be partitioned into $\alpha_t = (\alpha_t^{(x)\prime}, \alpha_t^{(c)\prime}, \alpha_t^{(r)\prime})'$, with $(n-1)$-vectors for trend

$\alpha_t^{(x)} = (x_t, x_{t-1}, ..., x_{t-n+2})'$, and cycle $\alpha_t^{(c)} = (c_t, c_{t-1}, ..., c_{t-n+2})'$. The observation matrix is $Z = (Z^{(x)}, Z^{(c)}, Z^{(r)})$, where $Z^{(x)} = (1, 0, ..., 0)$, $Z^{(c)} = (1, 0, ..., 0)$ are $(n-1)$-dimensional row vectors picking the first entry of $\alpha_t^{(x)}$ and $\alpha_t^{(c)}$. For the transition equation (19), one has $T = \text{diag}(T^{(x)}, T^{(c)}, T^{(r)})$, $R = \text{diag}(R^{(x)}, R^{(c)}, R^{(r)})$,

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

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

5.1 Deterministic components

In practice, deterministic components such as an intercept and a linear time trend often need to be taken into account. As will become clear, such terms can be added straightforwardly to the state space framework and their estimation can be carried out efficiently by a combination of the Kalman filter, the GLS estimator, and the CSS estimator. Parameter estimates will be consistent, given that the deterministic terms diverge at a rate similar to the stochastic trend.

Deterministic components can be taken into account either by detrending the data prior to estimating the fractional UC model, or by adding the components to the state space model. However, prior detrending biases the estimates both for deterministic and stochastic trend whenever the data is non-stationary, and thus should be avoided (Harvey; 1990, ch. 6.1.3). An alternative is to include the deterministic terms into the state vector and to explicitly model their dynamics via the state equation (19). However, state space models with deterministic components in the state vector are not stabilisable, thus the Kalman filter does not converge to its steady state solution and the CSS estimator is not applicable, see Harvey (1990, ch. 4.2.5). Following the suggestion there, I place the deterministic terms directly in the measurement equation (18). This allows the estimation of deterministic terms by the computationally efficient GLS estimator and does not hinder the steady state convergence of the Kalman filter. The remaining parameters θ_0 can be estimated via CSS as described in section 4, with the asymptotic theory being unaffected.

To model the deterministic terms, I set $u_t = \mu' w_t$ in the measurement equation (18), where w_t is a non-stochastic k-vector, and μ is a k-vector of unknown parameters to be estimated. The modified measurement equation is then $y_t = \mu' w_t + Z\alpha_t$. Letting $W = (w_1, ..., w_n)'$ denote the $n \times k$ matrix collecting all w_t, and $V = \text{Var}(x_{1:n} + c_{1:n})$ denote the variance-covariance matrix of $x_{1:n} + c_{1:n}$, a GLS estimator for μ is given by $\tilde{\mu} = (W'V^{-1}W)^{-1}W'V^{-1}y_{1:n}$, see Harvey (1990, ch. 3.4.2). As also shown there, it is not necessary to compute V^{-1}. To see this, assume for the moment that $y_t - \mu' w_t$ was observable. The Kalman filter, when applied to $y_t - \mu' w_t$, yields the filtered values for trend and cycle in (6) to (9), together with the prediction errors as denoted by $v_t^*(\theta)$ in the following for the modified model. These prediction errors equal 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 variance-covariance matrix of $x_{1:n} + c_{1:n}$ conditional on ψ. As the Kalman filter is linear, it can be applied to the observable y_t and w_t separately, yielding $F(\theta)y_{1:n} = y^*(\theta)$ and $F(\theta)W = W^*(\theta)$ as prediction errors. The GLS estimator $\tilde{\mu}$ follows then from regressing $y^*(\theta) = (y_1^*(\theta), ..., y_n^*(\theta))'$ on $W^*(\theta) = (w_1^*(\theta), ..., w_n^*(\theta))'$, see Harvey (1990, ch. 3.4.2). The concentrated CSS estimator $\tilde{\theta} = (\tilde{d}, \tilde{\nu}, \tilde{\varphi}')'$ follows from minimizing the modified sum of squared prediction errors

$$\tilde{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{t=1}^{n} v_t^*(\theta)^2, \tag{20}$$

and $v_t^*(\theta) = y_t^*(\theta) - \tilde{\mu}' w_t^*(\theta)$ is the GLS residual. Asymptotic standard errors can be obtained from the Fisher information matrix (Harvey; 1990, ch. 4.5.3 and ch. 7.3).

To derive the asymptotic properties of both the GLS estimator $\tilde{\mu}$ and the concentrated CSS estimator (20), let the j-th term in w_t be $w_{j,t} = O(t^{\beta_j})$, $t \geq 1$, $\beta_j \in \mathbb{R}$, such that $w_{j,t}$ is a polynomial trend. I will only consider $-1 < \beta_j \leq d_0$ for all j, as the lower bound is required for $\Delta_+^{d_0} t^{\beta_j} = O(t^{\beta_j - d_0})$ to hold, see Robinson (2005), while the upper bound ensures that the fractional stochastic trend is not drowned by the deterministic terms. This guarantees the results on consistency and asymptotic normality of the CSS estimator in theorems 4.1 and 4.2 to remain valid. However, at least for CSS estimation of ARFIMA models, Hualde and Nielsen (2020) recently derived the asymptotic theory also allowing for deterministic trends of higher power, $\beta_j > d_0$. As the focus of this paper is not on the deterministic components, showing their results to carry over to fractional unobserved components models is left open for future research.

Within $-1 < \beta_j \leq d_0$, note that the arguments for consistency of the CSS estimator of θ_0 remain unchanged: $y^*(\theta) = F(\theta)y_{1:n}$ is $I(d_0 - d)$ and precisely equals the initial prediction error (14) in section 3 if y_t contains no deterministic terms, as $F(\theta)y_{1:n}$ is the residual from applying the Kalman filter as defined in section 3 to $y_{1:n}$ given the parameters θ. If deterministic terms are present in y_t, then $y^*(\theta) = F(\theta)y_{1:n}$ equals the prediction error (14) shifted either by a constant, or by an $o(1)$ term (depending on how close β_j is to d_0, as will become clear). Therefore, also the prediction error $v_t^*(\theta) = [y^*(\theta) - (W^{*'}(\theta)W^*(\theta))^{-1}W^{*'}(\theta)y^*(\theta)]_{(t)}$ is $I(d_0 - d)$. Thus, both $y_t^*(\theta)$ and $v_t^*(\theta)$ are asymptotically stationary for $d_0 - d < 1/2$, else non-stationary. By the same proof as for (B.1), the objective function (20) can be shown to converge in probability whenever $d_0 - d >$

$-1/2$, and to diverge in the opposite case. Therefore, the probability of the CSS estimator to converge within the non-stationary region of the parameter space is asymptotically zero. Hence, it is sufficient to consider the region of the parameter space where $v_t^*(\theta)$ is asymptotically stationary. Within this region, the same proof as for theorem 4.1 applies, showing that a UWLLN holds for the objective function. Thus, $\tilde{\theta}$ is consistent. This result is somewhat obvious, as the assumption on β_j ensures the filtered $y_t^*(\theta)$ to contain at most deterministic terms of order $O(1)$.

Consistency for the GLS estimator $\tilde{\mu}$ is not so trivial. Define $u^*(\theta) = (u_1^*, ..., u_n^*)' = F(\theta)(x_{1:n} + c_{1:n})$ as the residual from applying the Kalman filter to the true $x_{1:n}$ and $c_{1:n}$. $u_t^*(\theta)$ would equal the prediction error $v_t^*(\theta)$ if no deterministic terms were present. The GLS estimates $\tilde{\mu}$ are thus

$$\begin{aligned}\tilde{\mu} &= (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})F(\tilde{\theta})y_{n:1} = (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})F(\tilde{\theta})\left[W\mu_0 + x_{1:n} + c_{1:n}\right] \\ &= \mu_0 + (W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta}), \end{aligned} \tag{21}$$

where μ_0 denotes the true coefficients to be estimated. $\tilde{\mu}$ is consistent if and only if the latter term in (21) is $o_p(1)$, i.e. the bias converges to zero as $n \to \infty$. For the purpose of illustration, I now focus on a single deterministic term only, such that $W^*(\tilde{\theta}) = (w_1^*(\tilde{\theta}), ..., w_n^*(\tilde{\theta}))'$. However the results directly generalize to multiple deterministic terms. 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 (21) can be written as

$$(W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta}) = \left(\frac{\sum_{t=1}^n {w_t^*}^2(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}\right)^{-1} \frac{\sum_{t=1}^n w_t^*(\tilde{\theta})u_t^*(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}, \tag{22}$$

where $\frac{\sum_{t=1}^n {w_t^*}^2(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}$ is bounded from above and below as $n \to \infty$. In contrast, by Hualde and Nielsen (2020, eqn. (S.88)), $\frac{\sum_{t=1}^n w_t^* u_t^*(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}} = o_p(1)$ if and only if $d_0 - 1/2 < \beta$. Thus, the GLS estimator for the deterministic terms is only consistent when deterministic and stochastic trends diverge at similar rates. In addition, as also can be seen from (22), the power of the deterministic term affects the rate of convergence of the GLS estimator: As $\frac{\sum_{t=1}^n w_t^*(\tilde{\theta})u_t^*(\tilde{\theta})}{n^{1/2+\beta-\tilde{d}}}$ converges in distribution when $n \to \infty$, see Hualde and Nielsen (2020, proof of cor. 1), it follows that the GLS estimator converges at rate $n^{1/2+(\beta-d_0)}$ as $n \to \infty$, and thus the rate is slower than the standard $\sqrt{n}$-convergence whenever the deterministic terms are dominated by the stochastic trend.

In summary, any trend of order $d_0 - 1/2 < \beta_j \leq d_0$ can be estimated consistently, and the rate of convergence of the GLS estimator will be faster the closer β_j is to d_0. This is in line with the well-established findings in the literature, that an intercept (i.e. $\beta_j = 0$) cannot be estimated consistently for time series with unit roots ($d_0 = 1$), whereas a linear trend ($\beta_j = 1$) can be estimated consistently. In addition, the convergence rate matches the findings of Robinson (2005) for semiparametric long memory models with deterministic components, of Hualde and

Nielsen (2020) for parametric ARFIMA models with deterministic components, and the general literature on the estimation of the sample mean for fractionally integrated processes, see e.g. Hassler (2019, ch. 7).

5.2 Correlated trend and cycle innovations

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

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

The new optimization problem then becomes

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

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

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

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

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

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

$$\begin{aligned}\hat{c}_{t:1}(y_{t:1}, \tilde{\psi}) = &\left[\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}'B_{\varphi,t} + B_{\varphi,t}'S_{d,t}) + \sigma_\epsilon^2 S_{d,t}'S_{d,t}\right]^{-1} \\ &\times\left(\sigma_\epsilon^2 S_{d,t}'S_{d,t} + \sigma_{\eta\epsilon}B_{\varphi,t}'S_{d,t}\right)y_{t:1}.\end{aligned} \tag{24}$$

Obviously, (23) and (24) 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 (23) and (24) 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 $\tilde{\theta} = (d, \nu, \nu_2, \varphi')'$. This is necessary for the CSS estimator to be identified, however the (quasi-) maximum likelihood estimator derived in subsection 5.3 can be used to estimate $\tilde{\psi}$ 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

$$\begin{aligned} v_{t+1}(\tilde{\psi}) &= \Delta_+^d y_{t+1} + (b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d)) \\ &\times \left[\sigma_\eta^2 B_{\varphi,t}' B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}' B_{\varphi,t} + B_{\varphi,t}' S_{d,t}) + \sigma_\epsilon^2 S_{d,t}' S_{d,t}\right]^{-1} \left(\sigma_\epsilon^2 S_{d,t}' + \sigma_{\eta\epsilon} B_{\varphi,t}'\right) S_{d,t} y_{t:1}. \end{aligned} \tag{25}$$

Based on (25), a conditional sum-of-squares estimator for $\tilde{\psi}_0$ can be set up. Note that y_{t+1} enters (25) in fractional differences, and note also that all terms in (25) exhibit the same convergence rates as for the case with uncorrelated errors. Thus, consistency and asymptotic normality of the CSS estimator with correlated innovations can be shown by carrying out the same proofs as for the untruncated model summarized in section 4. Finally, as noted by Morley et al. (2003), for the integer-integrated case $d_0 = 1$, the model is not identified if c_t follows an AR(p) with $p < 2$, as the autocovariance function of Δy_t dies out after lag one. For non-integer integration orders, identification is not a problem, as the autocovariance function of $\Delta_+^d y_t$ dies out only at lag t.

5.3 Maximum likelihood estimation

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

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

see Harvey (1990, ch. 3.4). Now, when the Kalman filter converges to its steady state solution at an exponential rate, the ML estimator asymptotically becomes independent of the initialization of the Kalman filter, see Harvey (1990, ch. 3.4.2), and $\sigma_{v_t}^2$ converges to a constant. Thus, neither the initialization of the Kalman filter, nor the time-dependence of $\sigma^2_{v_t}$ asymptotically matters, and therefore the CSS estimator in (16) exhibits the same asymptotic distribution as the (quasi-) maximum likelihood estimator, see Harvey (1990, p. 129).

For the Kalman filter to converge to its steady state solution at an exponential rate, it is sufficient that the state space model is detectable and stabilizable (Harvey; 1990, p. 116). Detectability holds if there exists a matrix D such that all eigenvalues of $T - DZ$ are bounded below unity in absolute value, while stabilizability holds if there exists a matrix $\tilde{D}$ such that all eigenvalues of $T + R\tilde{D}$ are bounded below unity in absolute value (see Harvey; 1990, p. 116). The former condition is easy to show, as setting $D = T^{(\cdot,1)}$ to be the first column of T yields a reduced rank of $T - DZ$, so that all eigenvalues are zero. Similarly, setting the first row of $\tilde{D}$ equal to the negative first row of T, $\tilde{D}^{(1,\cdot)} = -T^{(1,\cdot)}$, as well as the second row of $\tilde{D}$ equal to the negative n-th row of T, $\tilde{D}^{(2,\cdot)} = -T^{(n,\cdot)}$, yields a reduced rank of $T + R\tilde{D}$, so that all eigenvalues are zero. It follows that the state space model is also stabilizable. Thus, as $n \to \infty$, the CSS estimator and the QML estimator asymptotically become identical, which was also pointed out by Harvey (1990, p. 187) for integer-integrated models. Consequently, the results in section 4 also hold for the QML estimator. Finally, while computational efficiency clearly favors the CSS estimator that avoids the Kalman recursions for the conditional variance of the state vector, the QML estimator may be favorable in finite samples where initialization of the Kalman filter plays a non-negligible role. In particular, a combination of the QML estimator for an initial burn-in period together with the CSS estimator once the filtered prediction error variance has sufficiently converged seems promising: It combines the possibility of diffuse initialization and thus assigns a lower weight to initial prediction errors, but switches to the computationally efficient CSS estimator once the benefits from the QML estimator have vanished. The performance of this estimator, that is typically called the steady-state filter (Harvey; 1990, p. 185f), is also examined in a Monte Carlo study in section 6 and compared to the CSS estimator.

6 Simulations

By the means of a Monte Carlo study, this section examines the finite sample estimation properties for the latent components and parameters of the fractional UC model as introduced in section 2. By considering the CSS estimator of section 4 as well as the QML estimator of subsection 5.3, the study reveals the loss in estimation accuracy of the computationally simpler CSS estimator by treating the filtered prediction error variance to be constant. Therefore, the study assigns a price tag on the computational efficiency gains and provides a guided choice for empirical researchers on when to use the CSS estimator. Furthermore, parameter estimates for the integration order are compared to the exact local Whittle estimator of Shimotsu (2010) for various choices of tuning parameters as a prominent benchmark. To see whether allowing for fractional trends matters, I also present results for the integer-integrated UC models of Harvey (1985) and Morley et al. (2003). Doing so, I examine whether fractional trends are well approximated by integer-integrated models, or whether the estimates for x_t and c_t are significantly deteriorated. Furthermore, I investigate whether misspecifying d to be unity biases the parameter estimates.

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

In contrast to the CSS estimator, the QML estimator makes use of the Kalman iterations for the prediction error variance, thereby allowing it to be time-dependent: It first initializes trend and cycle component with initial variances σ^2_η and σ^2_ϵ. Next, in a burn-in period, the QML estimator takes into account the exponential convergence of the prediction error variance by allowing it to converge to its steady-state value. As soon as the prediction error variance has converged sufficiently, i.e. it satisfies the convergence criterion $\left|\frac{\mathrm{Var}_\psi(v_{t+1}(\psi)|y_1,\dots,y_t) - \mathrm{Var}_\psi(v_t(\psi)|y_1,\dots,y_{t-1})}{\mathrm{Var}_\psi(v_t(\psi)|y_1,\dots,y_{t-1})}\right| < 0.01$, the optimization switches to the steady state Kalman filter that assumes the prediction error variance to be fixed from that point on. This avoids further iterations of the Kalman filter for the prediction error variance, speeds up the computation, and has a negligible impact on the estimation precision. The exact local Whittle estimator of Shimotsu (2010) is introduced as a benchmark for $m = \lfloor n^j \rfloor$ Fourier frequencies, $j \in \{.50, .55, .60, .65, .70\}$.

Estimates for θ_0 are compared by the root mean squared error (RMSE), as well as by the median bias. To judge how well trend and cycle are estimated, the coefficients of determination R^2_x and R^2_c from regressing x_t and c_t on their respective estimates from the Kalman smoother are reported, both for the underlying CSS and maximum likelihood estimates.

6.1 Fractional UC model with uncorrelated innovations

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

$$y_t = x_t + c_t, \qquad \Delta^d_+ x_t = \eta_t, \qquad c_t - b_1 c_{t-1} - b_2 c_{t-2} = \epsilon_t, \tag{26}$$

where $\eta_t \sim \mathrm{NID}(0, 1)$, $\epsilon_t \sim \mathrm{NID}(0, \nu)$ are uncorrelated. The cyclical coefficients are set to $b_{1,0} = 1.6$, $b_{2,0} = -0.8$ to reflect strong cyclical patterns. Starting values for the numerical optimization are set to $\theta_{start} = (d_{start}, \nu_{start}, b_{1_{start}}, b_{2_{start}})' = (1, 1, 0.5, -0.5)'$, both for the CSS and the QML estimator. Note that for the QML estimator this implies assuming $\sigma^2_{\eta,0} = 1$ to be known, as only ν_0 is estimated. Although this assumption is usually violated, it allows for a fairer comparison between CSS and QML estimator, which is the core of this first simulation study. The $I(1)$ UC model is initialized analogously using $(\nu_{start}, b1_{start}, b2_{start})' = (1, 0.5, -0.5)'$.

Table A.1 shows RMSE and median bias for the estimated integration orders via CSS, QML and the exact local Whittle estimator. It illustrates convergence of the RMSE as n increases, which is in line with the theoretical results on consistency. As can be expected by the parametric nature, the fractional UC models yield a smaller RMSE compared to the nonparametric Whittle estimators. The differences are particularly striking for higher ν_0, where the signal of the fractional trend is drowned by strong cyclical variation, as well as for high n. In direct comparison, the QML estimator slightly outperforms the CSS estimator for the estimation of the integration order, however differences are rather small. Both CSS and QML estimator seem to exhibit little to no bias for d_0, while the cyclical dynamics induce a strong bias on the exact local Whittle estimator.

Tables A.2 and A.3 contain RMSE and median bias for ν_0 and the autoregressive parameters estimated via CSS and QML estimator, both for the fractional UC model and an $I(1)$ UC benchmark setting $d = 1$. While there was hardly a difference between CSS and QML in terms of the integration order estimate, for ν_0 both bias and variance are significantly smaller for the QML estimator. For $b_{1,0}$ and $b_{2,0}$, again little to no differences between CSS and QML are visible. The direct comparison with the $I(1)$ benchmark reveals that there is little to no difference for the estimation of $b_{1,0}$ and $b_{2,0}$, whereas ν_0 is typically estimated with a higher precision via the fractional UC model whenever $d_0 \neq 1$.

Table A.4 compares the estimates for x_t and c_t for the fractional UC model and the $I(1)$ UC benchmark (that sets $d = 1$). As before, it considers parameter estimates via the CSS and the QML estimator. As can be seen, differences between the coefficients of determination are negligible for CSS and QML. When comparing the $I(1)$ benchmark with the fractional UC model, one finds that for $d = 1$ there is no efficiency loss of the fractional model visible. For non-integer d, the coefficient of determination for x_t should not be compared to the $I(1)$ benchmark, as a high R_x^2 may also result from a spurious regression, and thus only R_c^2 is considered. There, a clear improvement from the fractional model is visible, particularly when ν is small. However, R_c^2 is still relatively high for the $I(1)$ benchmarks, so that at least for the setup under study, integer-integrated UC models are able to approximate the fractionally integrated trend.

6.2 Fractional UC model with correlated innovations

Next, (26) is extended to account for correlated innovations

$$y_t = x_t + c_t, \qquad \Delta_+^d x_t = \eta_t, \qquad c_t - b_1 c_{t-1} - b_2 c_{t-2} = \epsilon_t, \qquad \begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim \text{NID}(0, Q). \qquad (27)$$

As before, the cyclical coefficients are set to $b_{1,0} = 1.6$, $b_{2,0} = -0.8$. Q is parametrized as $Q_{1,1} = 1$, $Q_{2,2} = \nu_0 \in \{1, 5, 10\}$, thus accounting for medium to strong cyclical fluctuations. $Q_{1,2} = Q_{2,1} = -0.2\sqrt{\nu_0}$ sets the correlation between short- and long-run innovations to be slightly negative and

Note: Simulations are currently running. TBD: Add tables and interpretation constant across the different parametrizations of the model. Starting values for the numerical optimization are set to $\theta_{start} = (d_{start}, \nu_{start}, \nu_{2,start}, b_{1_{start}}, b_{2_{start}})' = (1, 1, 0, 0.5, -0.5)'$. The $I(1)$ UC model is initialized analogously using $(\nu_{start}, \nu_{2,start}, b_{1_{start}}, b_{2_{start}})' = (1, 0, 0.5, -0.5)'$. As the QML estimator becomes computationally infeasible in this setup, only the CSS estimates for the fractional UC model are shown.

7 Application

In this section, I apply the fractional UC model to log US annual CO2 emissions. Besides estimates for the memory parameter d, which may be of interest in their own right, I address the following research questions: (i) What is the trending behavior of US CO2 emissions? Does the estimate for x_t resemble the shape of the often hypothesized environmental Kuznets curve, i.e. an inverted U-shaped relation between economic development and CO2 emissions (see e.g. Harbaugh et al.; 2002), and if so, what is the current position of the US economy on this curve? (ii) What is the cyclical component of US CO2 emissions? Does it align with the business cycle, as results of Doda (2014) suggest? (iii) Is there evidence for a decoupling of economic activity and CO2 emissions (see Haberl et al.; 2020)? Does decoupling involve the cycle, the trend, or both? (iv) Is there evidence for correlation among long- and short-run shocks? If so, is it positive or negative, and can we assign an interpretation to the correlation structure? (v) Are there any additional insights that follow from a fractional model in comparison to integer-integrated UC models?

Data on US annual CO2 emissions stem from the Global Carbon Project and were collected by Ritchie et al. (2020). The underlying time series spans from 1800 to 2020, consists of 221 observations, is measured in million tons, was log-transformed to account for the exponential increase, and is sketched in figure 1.

From figure 1, it becomes apparent that log CO2 emissions evolve along a rather linear time trend, that needs to be taken into account in what follows. Furthermore, prior estimates via the exact local Whittle estimator including a linear time trend find an integration order between 1.29 and 1.44, depending on the choice of the bandwidth. Note however that the Monte Carlo study of section 6 found the exact local Whittle estimator to be heavily downward-biased for similar sample sizes whenever cyclical dynamics were present.

In what follows, c_t is specified as an autoregressive process of order p, which is in line with the UC literature. The resulting fractional UC model is thus given by

$$y_t = \mu_0 + \mu_1 t + x_t + c_t, \qquad \Delta_+^d x_t = \eta_t, \qquad \sum_{j=0}^{p} b_j c_{t-j} = \epsilon_t, \tag{28}$$

where μ_0 and μ_1 account for a constant and a linear trend. Furthermore, $\text{Var}(\eta_t, \epsilon_t)' = Q$. To estimate the fractional UC model, I draw 100 combinations of starting values from uniform distributions with appropriate support ($d \in [1; 2]$, $\nu \in [1, 20]$, and ν_2 is set to force the correlation

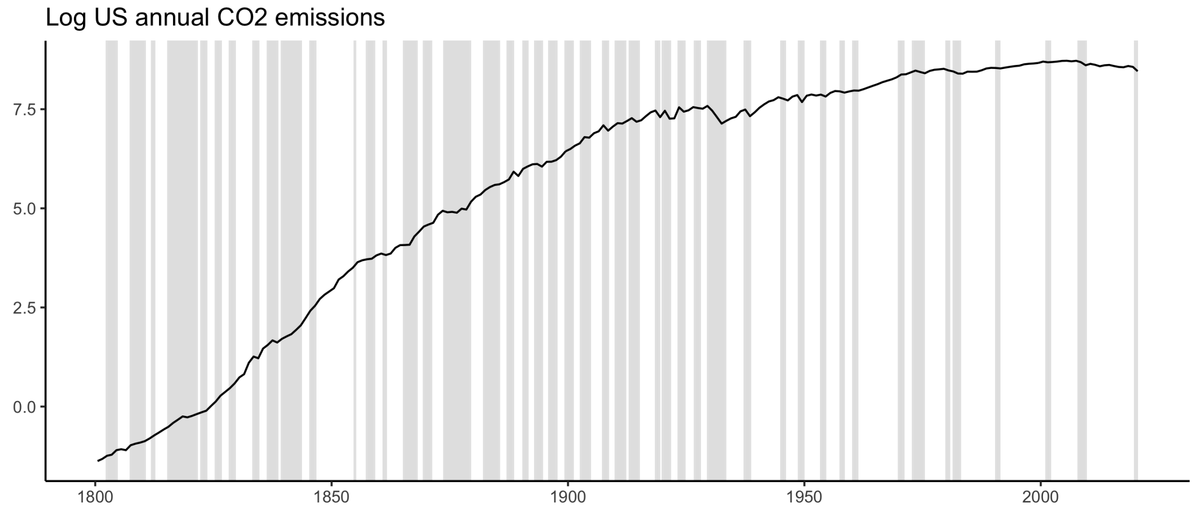

Figure 1: Log US annual CO2 emissions from 1800 to 2020. Shaded areas correspond to US recession periods. Data stem from the Global Carbon Project and were collected by Ritchie et al. (2020).

to be $\in [-0.5, 0.5]$). Autoregressive parameters are drawn randomly from the set of coefficients that ensure the cyclical AR polynomial to be stable. For each of the 100 initial values, parameter estimates are then obtained from the CSS estimator as introduced in section 4. From the 100 resulting estimates, the one corresponding to the smallest value of the objective function is chosen as the final estimate.

Table A.5 displays the estimation results for $p \in \{0, 1, 2, 3, 4, 5\}$, together with the corresponding value of the objective function, both for uncorrelated and correlated innovations. As it shows, for $p > 3$ both estimates for the correlated and the uncorrelated model are relatively stable. The integration order is found to be around 1.75, indicating that trend CO2 emissions are strongly persistent, non-mean-reverting, and clearly closer to a quadratic trend specification than to a random walk trend. However, trend CO2 growth (that is, the first difference of the estimated trend) is (conditionally) mean-reverting, as its integration order is below unity. This indicates a converging impact of a long-run shock on trend CO2 growth as $t \to \infty$, which would not be the case if $d \geq 2$. The variance ratio ν is estimated to be small, which is typical for smooth, persistent trends. Furthermore, long- and short-run innovations are found to be positively correlated. In what follows, I only consider the results from the correlated fractional UC model for $p = 5$, as it encompasses the other specifications presented in table A.5.

Figure 2 plots the smoothed trend estimate $\hat{x}_t(y_{n:1}, \hat{\theta}) + \hat{\mu}_0 + \hat{\mu}_1 t$ together with the series for log US annual CO2 emissions. The left plot shows the series in logs, while the right plot displays US annual per capita CO2 emissions in tons CO2. The fractional UC model estimates a smooth trend which is due to the relatively high $\hat{\nu}$, as well as the high integration order $\hat{d}$. As becomes apparent from the right-hand plot, the 1979 energy crisis as well as the Great Recession

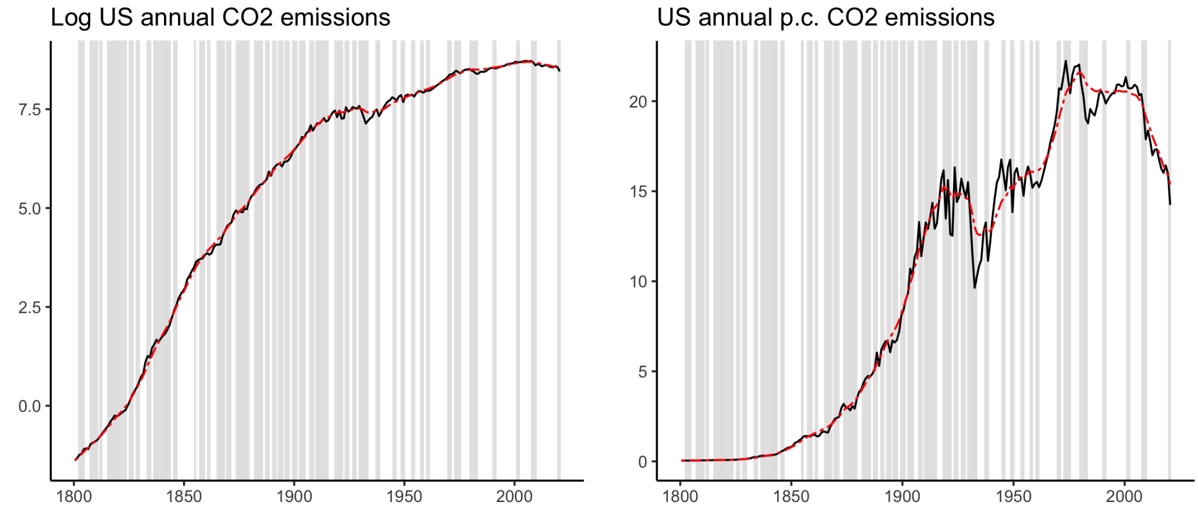

Figure 2: Trend CO2 emissions: The left plot sketches log US annual CO2 emissions (black) together with the estimated trend $\hat{x}_t(y_{n:1}, \hat{\theta}) + \hat{\mu}_0 + \hat{\mu}_1 t$ (red). The right plot shows US annual CO2 emissions per capita (measured in tons), together with the respective transformation of the trend estimate. Shaded areas correspond to US recession periods.

mark two turning points in per capita CO2 emissions: Since the former, per capita emissions are decreasing, while annual emissions for the economy as a whole are declining since the Great Recession. The turning points, together with concave trend dynamics in figure 2, support the environmental Kuznets curve hypothesis.

Figure 3 displays the smoothed estimates for the cycle $\hat{c}_t(y_{n:1}, \hat{\theta})$. In line with the high estimate $\hat{\nu}$, the fractional UC model yields rich cyclical dynamics that display a persistent behavior. Clearly, $\hat{c}_t(y_{n:1}, \hat{\theta})$ shows pro-cyclical behavior, where sharp drops occur mostly during recession periods while graduate increases in cyclical CO2 emissions happen during periods of economic recovery and prosperity. In particular the massive downturn during the Great Depression is remarkable. However, from the second half of the 20th century on, the magnitude of pro-cyclical variation has somewhat decreased. While decoupling of economic activity and CO2 emissions may hold for the long-run behavior as figure 2 suggests, figure 3 indicates that cyclical CO2 emissions and the business cycle remain coupled. Finally, table A.5 shows medium positive correlation between long- and short-run innovations. One possible explanation, that is also supported by figures 2 and 3 is that recession periods do not only yield a decline of cyclical economic activity and thus also of cyclical CO2 emissions. Instead, they may also have a permanent impact on the economy, e.g. by replacing outdated technologies with newer ones, by a permanent reduction of the workforce, or by a transformation of energy production. The positive relation between permanent and transitory shocks calls for further investigation, and I leave this open to future research.

Finally, I investigate to what extent the fractional UC model reveals new information about the trending and cyclical behavior of CO2 emissions by comparing the above results to integer-

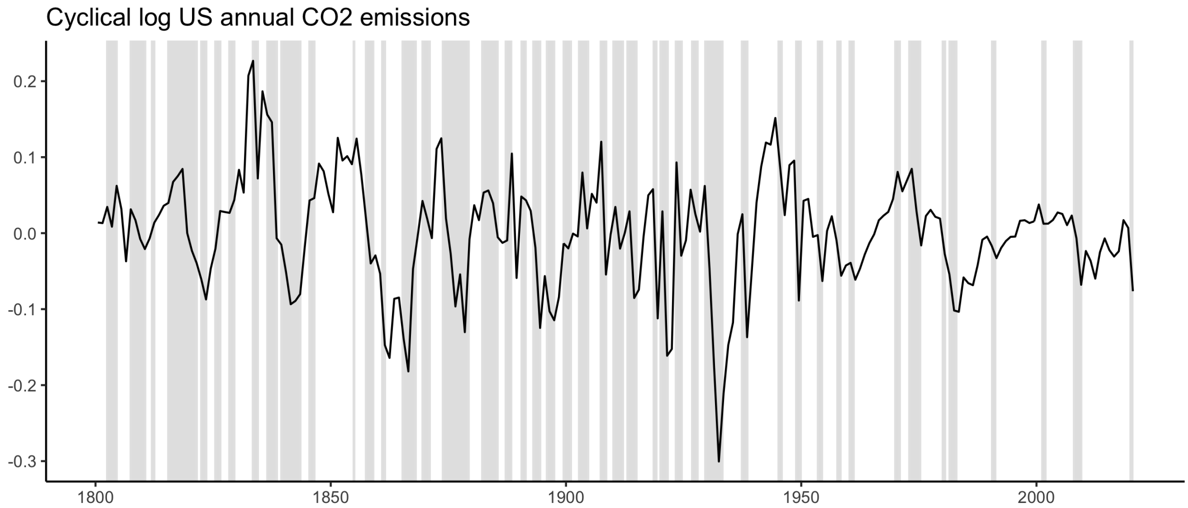

Figure 3: Estimated cyclical log US annual CO2 emissions $\hat{c}_t(y_{n:1}, \hat{\theta})$. Shaded areas correspond to US recession periods.

integrated benchmark models. I consider the model of Harvey (1985), that assumes x_t to be a random walk and c_t to be an autoregressive process, the correlated UC model of Morley et al. (2003), that generalizes the model of Harvey (1985) to account for correlated innovations, and the HP filter of Hodrick and Prescott (1997), that assumes x_t to be $I(2)$. The former two models are obtained by setting $d = 1$ in (28), while the HP filter follows from setting $t = n$, $d = 2$, $b(L, \varphi) = 1$, and $\lambda = \sigma_\epsilon^2/\sigma_\eta^2$ in (10), as also discussed in section 3.

Estimates for the $I(1)$ UC model are obtained analogously to the fractional UC model: I draw 100 combinations of starting values from uniform distributions, where $\sigma_\eta^2, \sigma_\epsilon^2 \in [0.0001, 0.01]$ was found to be appropriate. All other parameters are initialized as before. Each of the 100 combinations enters as starting values for the QML estimator, and the QML estimates yielding the highest log likelihood are chosen as final estimates. Note that while the fractional UC model was estimated by the CSS estimator, I use the QML estimator for the benchmarks to be in line with the empirical literature. Furthermore, the $I(1)$ specification keeps the state vector small, and thus computational simplicity is not a relevant argument to the $I(1)$ model.

Table A.6 contains the parameter estimates both for the uncorrelated and the correlated $I(1)$ UC model. For the latter, the correlation coefficients converge to -1, and thus the covariance matrix of long- and short-run shocks is nearly singular. As can be seen from the estimated coefficients of the cyclical component, the $I(1)$ trend clearly cannot grasp the long-run dynamics of log US annual CO2 emissions. Instead, the model attributes additional long-run dynamics to the cycle, forcing it to exhibit near-unit-root behavior. Therefore, the estimated cyclical components of all parametrizations of table A.6 evolve non-mean-reverting. This can also be seen from figure 4, that sketches the estimated cycle for $p = 5$.

While the $I(1)$ specification is clearly at odds with the memory parameter estimates in table

A.5, the $I(2)$ trend assumption of the HP filter can be expected to better match the long-run dynamics of log US annual CO2 emissions. Instead of estimating a parametric model, the HP filter requires to set a tuning parameter λ that penalizes the cyclical dynamics. As shown before, it can be interpreted as the variance ratio of short- and long-run innovations under the restrictions of the HP filter. Thus, the higher the λ, the more variation is attributed to the cyclical component. Following Ravn and Uhlig (2002), I set $\lambda = 6.25$, which is typically chosen by the empirical literature for annual data, and was also set by Doda (2014) for decomposing log CO2 emissions into trend and cycle via the HP filter. Figure 5 presents the estimated cycle from the HP filter together with the estimate from the fractional UC model with correlated innovations. Clearly, the HP filter attributes fewer variation to c_t compared to the fractional UC model. It lacks the persistent patterns of peaks and troughs, behaves comparably noisy, and misses the cyclical patterns at the current frontier of the data. Thus, in comparison to both, $I(1)$ and $I(2)$ trend specifications, the fractional UC model offers additional insights on permanent and cyclical dynamics of US annual CO2 emissions.

8 Conclusion

This paper has introduced a novel unobserved components model, where the trend component is specified as a type II fractionally integrated process of unknown memory d. It encompasses the bulk of unobserved components models in the literature, allows for richer long-run dynamics beyond the typical integer-integrated specifications, and for a data-dependent specification of the trend by treating d as an unknown parameter to be estimated. Estimates for trend and cycle were derived as the analytical solution to the optimization problem of the Kalman filter. The conditional sum-of-squares estimator was introduced to estimate the model parameters, and was shown to be consistent and asymptotically normally distributed. In an application to US annual CO2 emissions, the fractional unobserved components model estimated a smooth trend component that started to exhibit an inverted U-shape, together with a cycle that was closely linked to the business cycle.

To applied researchers, the fractional unobserved components model offers a robust, flexible, and data-driven way for signal extraction of data of unknown persistence. It neither requires prior assumptions about the integration order of a process, nor the choice of some tuning parameter. Therefore, it provides a solution to model specification in the unobserved components literature, calling for further applications beyond CO2 emissions.

A Figures and Tables

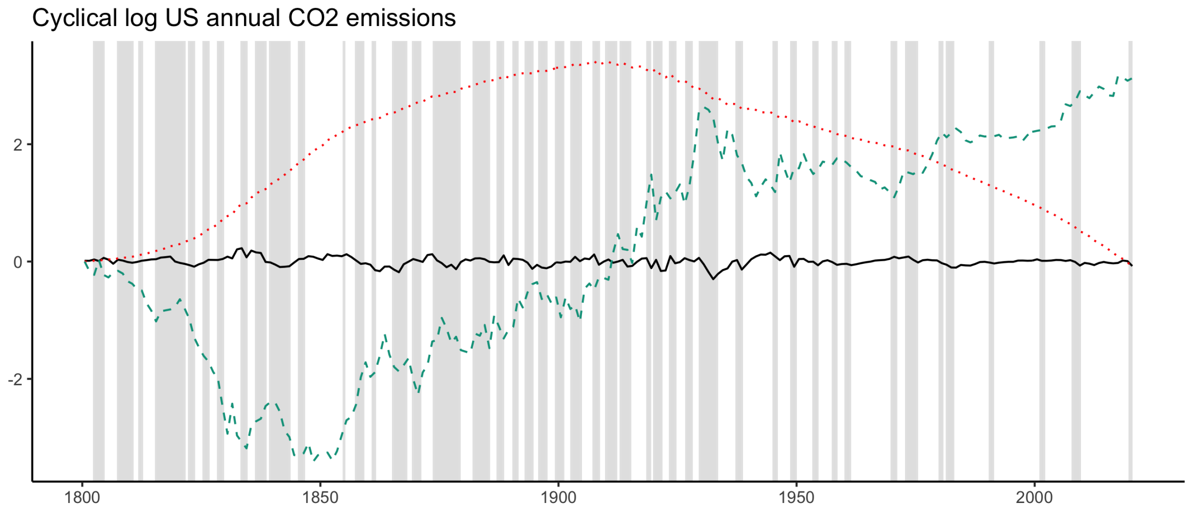

Figure 4: Estimated cycle of the $I(1)$ UC model with uncorrelated shocks of Harvey (1985) (red, solid), of the correlated $I(1)$ UC model of Morley et al. (2003) (blue, dashed), and of the fractional UC model with correlated innovations (black, solid). Each model considers $p = 5$ lags of the cyclical AR polynomial. Shaded areas correspond to US recession periods.

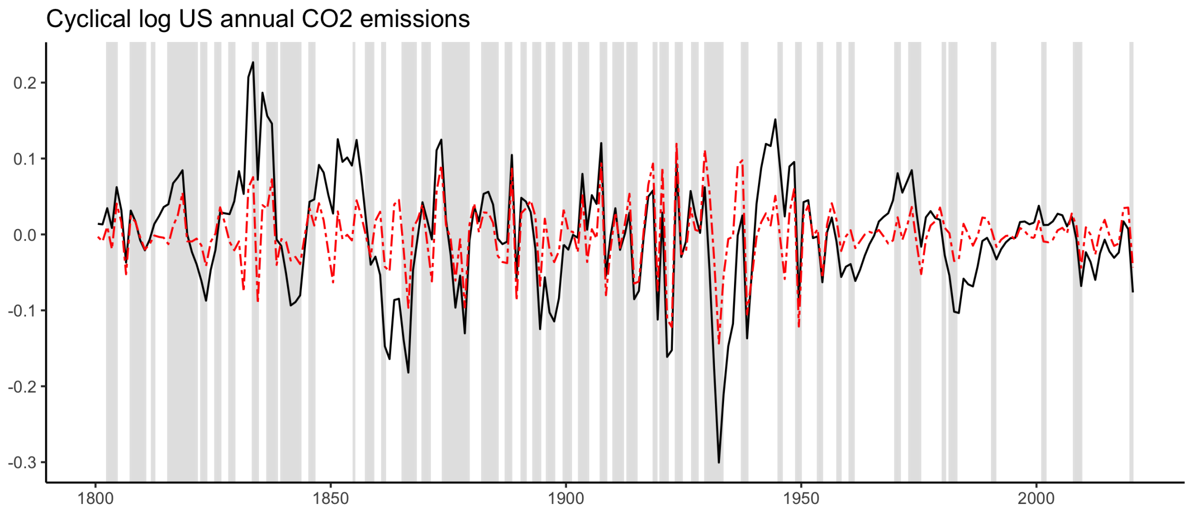

Figure 5: Estimated cyclical component of the HP filter with $\lambda = 6.25$ as suggested by Ravn and Uhlig (2002) (red, dashed), and of the fractional UC model with correlated innovations $\hat{c}_t(y_{n:1}, \hat{\theta})$ (black, solid). Shaded areas correspond to US recession periods.

			RMSE							bias						
n	ν_0	d_0	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$	$\hat{d}_{CSS}$	$\hat{d}_{QML}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$
100	1	.75	.289	.228	.638	.579	.410	.228	.574	-.026	-.006	-.675	-.569	-.363	.032	.509
		1.00	.259	.228	.681	.614	.460	.222	.397	-.013	-.012	-.652	-.577	-.408	-.108	.316
		1.25	.260	.494	.651	.591	.464	.258	.258	-.004	-.005	-.602	-.546	-.415	-.186	.149
	5	.75	.379	.289	.714	.673	.507	.268	.743	-.062	-.037	-.750	-.727	-.477	.047	.702
		1.00	.328	.263	.871	.810	.638	.289	.526	-.013	-.031	-.896	-.806	-.602	-.162	.464
		1.25	.264	.205	.903	.842	.694	.382	.338	-.014	-.031	-.872	-.805	-.659	-.330	.233
	10	.75	.401	.380	.726	.690	.527	.276	.773	-.081	-.073	-.750	-.750	-.501	.043	.729
		1.00	.370	.322	.919	.866	.692	.309	.549	-.035	-.049	-.976	-.880	-.664	-.180	.490
		1.25	.295	.308	.995	.934	.779	.426	.354	-.033	-.036	-.978	-.907	-.743	-.375	.251
200	1	.75	.166	.148	.618	.642	.568	.389	.139	-.021	-.015	-.622	-.653	-.550	-.363	.030
		1.00	.118	.122	.598	.637	.563	.415	.153	-.014	-.010	-.556	-.610	-.534	-.388	-.099
		1.25	.128	.138	.530	.584	.526	.407	.200	-.005	-.008	-.486	-.552	-.498	-.378	-.166
	5	.75	.285	.250	.722	.732	.697	.521	.164	-.036	-.024	-.750	-.750	-.737	-.504	.037
		1.00	.205	.188	.821	.852	.784	.615	.221	-.012	-.020	-.817	-.850	-.769	-.597	-.168
		1.25	.181	.192	.786	.835	.773	.640	.335	-.003	-.016	-.749	-.812	-.751	-.618	-.315
	10	.75	.336	.247	.736	.743	.719	.553	.169	-.041	-.045	-.750	-.750	-.750	-.542	.036
		1.00	.242	.203	.890	.914	.857	.683	.241	-.019	-.031	-.914	-.941	-.855	-.665	-.187
		1.25	.192	.193	.890	.934	.870	.729	.384	-.016	-.019	-.865	-.911	-.849	-.710	-.362
300	1	.75	.128	.110	.508	.607	.603	.494	.216	-.007	-.006	-.482	-.596	-.590	-.474	-.197
		1.00	.090	.100	.448	.577	.581	.487	.272	-.004	-.000	-.405	-.547	-.560	-.470	-.256
		1.25	.132	.206	.369	.515	.534	.457	.290	.001	.001	-.318	-.486	-.511	-.439	-.273
	5	.75	.232	.157	.671	.724	.723	.650	.305	-.009	-.012	-.696	-.750	-.750	-.648	-.286
		1.00	.160	.124	.682	.795	.796	.701	.431	.000	-.006	-.655	-.777	-.781	-.686	-.421
		1.25	.155	.119	.611	.754	.769	.691	.491	-.002	-.004	-.577	-.726	-.743	-.673	-.477
	10	.75	.276	.225	.707	.739	.739	.687	.326	-.013	-.023	-.750	-.750	-.750	-.701	-.310
		1.00	.185	.186	.771	.869	.870	.778	.484	-.003	-.005	-.756	-.870	-.867	-.768	-.475
		1.25	.161	.154	.715	.850	.862	.782	.568	-.010	-.007	-.687	-.823	-.839	-.764	-.555

Table A.1: Root mean squared prediction errors (RMSE) and median bias for the integration order estimates of the fractional UC model with uncorrelated innovations in subsection 6.1. The different columns indicate integration order estimates via the CSS estimator ($\hat{d}_{CSS}$), the QML estimator $\hat{d}_{QML}$, and the exact local Whittle estimator of Shimotsu (2010) with tuning parameter α ($\hat{d}^{EW}_{\alpha}$).

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}_{1_{QML}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}_{2_{QML}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
100	1	.75	13.345	.426	35.064	.414	.121	.107	.178	.142	.115	.122	.176	.137
		1.00	22.242	.392	19.960	.313	.142	.114	.121	.104	.125	.129	.130	.111
		1.25	26.250	.427	28.616	.489	.152	.180	.137	.159	.173	.221	.195	.226
	5	.75	37.191	1.241	62.781	3.683	.089	.096	.112	.108	.089	.095	.108	.114
		1.00	43.868	1.804	37.452	1.015	.085	.101	.089	.100	.083	.085	.092	.109
		1.25	42.022	1.007	21.794	2.631	.081	.091	.082	.123	.085	.086	.091	.153
	10	.75	48.294	2.900	69.021	2.793	.103	.106	.095	.147	.086	.101	.092	.142
		1.00	51.038	2.413	46.118	3.031	.079	.110	.080	.126	.079	.107	.079	.135
		1.25	48.087	3.259	25.512	4.288	.081	.131	.076	.173	.083	.131	.081	.180
200	1	.75	3.334	.289	26.616	.323	.075	.090	.134	.095	.084	.107	.135	.092
		1.00	11.170	.255	8.116	.225	.076	.090	.070	.070	.083	.113	.074	.079
		1.25	15.620	.239	16.848	.434	.090	.096	.069	.116	.094	.122	.094	.168
	5	.75	25.979	1.367	64.184	.836	.059	.072	.082	.072	.059	.089	.083	.079
		1.00	33.814	1.637	23.688	.877	.057	.065	.059	.059	.060	.080	.062	.070
		1.25	34.675	1.708	8.199	1.490	.061	.068	.059	.117	.064	.083	.063	.151
	10	.75	41.458	1.987	72.690	4.786	.059	.058	.071	.106	.065	.053	.077	.107
		1.00	45.029	1.625	34.472	4.817	.063	.060	.057	.151	.073	.054	.064	.157
		1.25	43.429	1.711	12.641	4.030	.052	.070	.060	.123	.058	.055	.066	.154
300	1	.75	.913	.239	2.193	.283	.058	.063	.107	.082	.065	.077	.108	.081
		1.00	2.850	.244	4.777	.190	.058	.058	.060	.058	.065	.073	.065	.065
		1.25	7.839	.255	14.689	.375	.090	.095	.081	.147	.103	.125	.098	.214
	5	.75	19.731	.995	62.337	.725	.048	.049	.074	.056	.051	.049	.077	.062
		1.00	27.762	.667	17.423	.687	.050	.044	.047	.058	.056	.044	.052	.071
		1.25	29.982	.733	4.445	.826	.044	.049	.056	.066	.049	.052	.060	.086
	10	.75	37.220	1.568	73.606	1.830	.048	.071	.058	.177	.052	.076	.063	.212
		1.00	39.479	1.885	26.651	2.071	.047	.065	.046	.174	.053	.058	.052	.209
		1.25	37.784	1.707	8.908	4.767	.049	.078	.054	.196	.053	.072	.062	.238

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

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}_{1_{QML}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}_{2_{QML}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
100	1	.75	.064	-.008	-.230	.164	-.012	-.013	-.015	-.061	-.004	.005	-.010	.045
		1.00	-.004	.039	-.078	-.005	-.009	-.017	-.007	-.014	-.003	.007	-.008	.007
		1.25	-.007	.048	-.238	.080	-.009	-.017	.028	.015	-.002	.017	-.027	.027
	5	.75	1.451	-.233	29.748	.296	-.009	-.008	-.043	-.032	-.001	.000	.032	.023
		1.00	.676	-.195	.131	.052	-.004	-.009	-.008	-.015	-.002	.001	.000	.012
		1.25	-.056	-.121	-3.075	.148	-.002	-.011	.031	.009	-.008	.007	-.036	.016
	10	.75	4.023	-.604	89.903	.307	-.005	-.004	-.030	-.025	-.005	-.000	.024	.020
		1.00	3.226	-.413	2.023	.073	-.003	-.006	-.007	-.015	-.006	.002	-.000	.013
		1.25	-.344	-.262	-6.835	.311	.002	-.011	.025	.003	-.011	.006	-.032	.016
200	1	.75	.027	.028	-.215	.172	-.008	-.008	-.014	-.051	.004	.008	.001	.043
		1.00	.026	.096	-.006	.008	-.004	-.013	-.003	-.009	.002	.011	.000	.009
		1.25	.009	.020	-.370	.060	-.003	-.008	.044	.025	.002	.010	-.031	.012
	5	.75	.851	-.132	43.591	.274	-.006	-.004	-.048	-.024	.004	.003	.045	.019
		1.00	.309	-.098	.198	.002	-.003	-.005	-.005	-.008	.003	.005	.005	.008
		1.25	.222	-.085	-3.088	.080	-.001	-.004	.038	.013	-.002	.004	-.035	.010
	10	.75	2.758	-.270	89.909	.293	-.004	-.002	-.033	-.018	.001	.001	.030	.014
		1.00	1.159	-.148	.898	.059	-.001	-.003	-.006	-.010	.001	.004	.005	.009
		1.25	.668	-.133	-6.983	.254	.001	-.005	.030	.007	-.004	.004	-.029	.014
300	1	.75	-.002	.039	-.234	.162	-.001	-.005	-.007	-.043	-.000	.003	-.006	.037
		1.00	.002	.121	-.033	.002	-.001	-.012	-.001	-.004	-.002	.010	-.005	.003
		1.25	-.007	.011	-.390	.069	-.001	-.005	.045	.024	-.003	.003	-.034	.007
	5	.75	.271	-.051	42.045	.260	-.001	-.001	-.046	-.020	-.000	.001	.046	.015
		1.00	.022	-.071	-.162	-.002	-.000	-.001	-.001	-.004	-.001	.002	.001	.004
		1.25	.108	-.064	-3.173	.058	.000	-.003	.041	.015	-.003	.003	-.039	.004
	10	.75	1.420	-.140	89.910	.237	.001	-.001	-.028	-.016	-.001	.000	.029	.012
		1.00	.057	-.088	-.006	-.006	.002	-.001	-.001	-.005	-.001	.002	.002	.006
		1.25	-.294	-.069	-7.135	.199	.004	-.002	.034	.008	-.005	.003	-.034	.011

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

n	ν_0	d_0	Trend R^2_{CSS}	Trend R^2_{QML}	Trend $R^{I(1)^2}_{CSS}$	Trend $R^{I(1)^2}_{QML}$	Cycle R^2_{CSS}	Cycle R^2_{QML}	Cycle $R^{I(1)^2}_{CSS}$	Cycle $R^{I(1)^2}_{QML}$
100	1	.75	.476	.518	.474	.521	.831	.845	.809	.841
		1.00	.738	.774	.756	.783	.762	.782	.772	.788
		1.25	.901	.918	.867	.870	.688	.682	.608	.544
	5	.75	.267	.294	.311	.325	.938	.945	.935	.943
		1.00	.574	.604	.606	.629	.897	.907	.903	.916
		1.25	.822	.838	.818	.809	.853	.865	.847	.787
	10	.75	.205	.217	.277	.276	.960	.964	.961	.965
		1.00	.488	.507	.543	.558	.930	.936	.935	.945
		1.25	.768	.774	.781	.765	.892	.896	.895	.841
200	1	.75	.610	.633	.588	.626	.846	.854	.827	.848
		1.00	.867	.875	.870	.876	.792	.797	.796	.800
		1.25	.967	.969	.940	.928	.729	.737	.671	.572
	5	.75	.363	.399	.383	.404	.943	.944	.940	.941
		1.00	.734	.750	.740	.760	.908	.910	.909	.915
		1.25	.927	.928	.924	.901	.868	.866	.865	.783
	10	.75	.278	.303	.330	.323	.964	.965	.965	.962
		1.00	.653	.674	.667	.688	.935	.936	.936	.941
		1.25	.898	.897	.894	.870	.903	.901	.899	.834
300	1	.75	.681	.693	.660	.683	.854	.859	.834	.848
		1.00	.908	.913	.911	.912	.799	.803	.803	.804
		1.25	.982	.983	.959	.952	.738	.739	.678	.551
	5	.75	.461	.483	.459	.474	.944	.947	.941	.940
		1.00	.809	.821	.812	.824	.910	.914	.912	.914
		1.25	.958	.960	.958	.945	.873	.875	.872	.815
	10	.75	.367	.381	.396	.379	.964	.965	.965	.961
		1.00	.751	.757	.755	.754	.937	.937	.938	.936
		1.25	.939	.938	.937	.913	.903	.904	.902	.819

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

$Q(y,\hat{\theta})$	$\hat{d}$	$\hat{\nu}$	$\hat{\nu}_2$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\widehat{\mathrm{Corr}(\eta_t, \epsilon_t)}$
1.1349	1.3999	0.7903							
1.1272	1.4674	1.7286		-0.2911					
1.1141	1.3898	0.3410		0.3415	0.3560				
1.1082	1.5420	3.7698		-0.5436	-0.0595	-0.1822			
1.0923	1.8722	48.4906		-0.7278	-0.0419	-0.1618	0.1378		
1.0880	1.8413	43.8545		-0.7118	-0.0493	-0.1728	0.1190	0.0678	
1.1258	1.6157	4.6707	-2.1612						-1.0000
1.1146	1.6049	7.0201	-2.6495	-0.2703					-1.0000
1.1144	1.6096	7.6008	-2.7569	-0.3096	-0.0205				-1.0000
1.1050	1.5483	2.7716	1.0818	-0.5675	-0.0279	-0.2077			0.6498
1.0941	1.7443	16.7184	2.7274	-0.7281	-0.0766	-0.1816	0.1517		0.6670
1.0894	1.7313	14.5846	1.9590	-0.7123	-0.0612	-0.1890	0.0900	0.0892	0.5130

Table A.5: Estimation results for the fractional UC model of log US CO2 emission via the CSS estimator for uncorrelated and correlated innovations. Correlations are estimated using $\nu = \sigma_\epsilon^2/\sigma_\eta^2$, $\nu_2 = \sigma_{\eta\epsilon}/\sigma_\eta^2$, and thus $\widehat{\mathrm{Corr}(\eta_t, \epsilon_t)} = \hat{\nu}_2/\sqrt{\hat{\nu}}$

$\log L(\psi)$	$Q(y,\hat{\psi})$	$\hat{\nu}$	$\hat{\nu}_2$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\widehat{\mathrm{Corr}(\eta_t, \epsilon_t)}$
-245.6263	1.3420	0.0001							
-245.6292	1.3419	0.0217		-0.9997					
-264.3214	1.1333	0.0039		-1.9998	1.0000				
-264.3018	1.1335	0.0084		-1.6303	0.2620	0.3686			
-267.4912	1.1003	0.0524		-0.6677	-0.7618	-0.4631	0.8932		
-266.3117	1.1149	0.2964		-0.8260	-0.2469	-0.4881	0.0683	0.4934	
-247.3099	1.3287	0.6451	-0.7296						-0.9083
-263.6369	1.1407	0.9681	-0.9797	-0.9935					-0.9957
-266.1118	1.1251	0.8804	-0.9369	-1.1702	0.1900				-0.9985
-266.1579	1.1225	0.9419	-0.9697	-0.8581	-0.1582	0.0300			-0.9992
-267.9061	1.1029	1.0745	-1.0305	-0.9389	-0.1141	-0.0076	0.0741		-0.9941
-267.7790	1.1031	0.8773	-0.9275	-1.0156	0.0342	0.1270	-0.1802	0.0507	-0.9903

Table A.6: Estimation results for the $I(1)$ UC model of log US CO2 emission via the QML estimator for uncorrelated and correlated innovations. While optimization is conducted over σ_η^2, $\sigma_{\eta\epsilon}$, σ_ϵ^2, the transformed $\nu = \sigma_\epsilon^2/\sigma_\eta^2$, $\nu_2 = \sigma_{\eta\epsilon}/\sigma_\eta^2$ are reported.

B Proof of theorem 4.1

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

However, as can be seen from (15), uniform convergence of the objective function fails around the point $d = d_0 - 1/2$: As 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 are asymptotically stationary whenever $d > d_0 - 1/2$, and non-stationary otherwise. Subsequently, I will show the pointwise probability limit of $Q(y, \theta)$ to be given by

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

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

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

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

To deal with non-uniform convergence in (B.1), I adopt the strategy of proof of Nielsen (2015), who tackles non-uniform convergence of the CSS estimator for 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 $0 < \kappa_1 < \kappa_2 < \kappa_3 < 1/2$ that will be determined later. Note that $\cup_{i=1}^3 D_i = D$. Within D_1 and D_3 convergence is uniform, while it is non-uniform within the overlapping D_2 that covers both stationary and non-stationary forecast errors. Denote the partitioned parameter spaces for θ as $\Theta_j = D_j \times \Sigma_\nu \times \Phi$, $j = 1, 2, 3$. Non-uniform convergence of (B.1) is then tackled similar to Nielsen (2015) by showing that for a given constant $K > 0$ there always exists a fixed $\bar{\kappa} > 0$ such that

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

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

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

which holds if both the objective function and the supremum of its absolute gradient satisfy a WLLN (Wooldridge; 1994, thm. 4.2). While the results in (B.3) and (B.4) are well established for the CSS estimator in the ARFIMA literature, see Hualde and Robinson (2011) and Nielsen (2015), showing them to carry over to the fractional UC model requires some additional effort. In particular, despite being asymptotically stationary for $\theta \in \Theta_3(\kappa_3)$, the forecast errors in (14) are not ergodic for two reasons: First, as 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 non-ergodic. Second, also the $\tau_j(\theta, t)$ in (15) depend on t. Therefore, even within $\Theta_3(\kappa_3)$ a WLLN for stationary and ergodic processes does not immediately apply. I tackle these problems by showing the expected difference between (15) and (B.2) to be

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

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

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

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

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

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

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

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

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

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

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

where $\phi_{2,\eta,j}(\theta,t) = \sum_{k=0}^{j-t-1} \pi_k(d-d_0)\tau_{j-k}(\theta) = O((1+\log j)^2 j^{\max(-d+d_0,-\zeta)-1})$ 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 D.1 and lemma D.2.

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

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

By lemma D.3, it holds that $\sum_{m=t+1}^{\infty} r_{\tau,k,m}(\theta) = O((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1})$, while $\pi_j(d-d_0) = O(j^{-d+d_0-1})$ and $\sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d) = O((1+\log(j-k))(j-k)^{\max(-d,-\zeta)-1})$, see lemma D.1 together with Johansen and Nielsen (2010, lemma B.4). Consequently, $\phi_{3,\eta,j}(\theta,t) = \sum_{k=0}^{\min(j,t)}\left(\sum_{m=t+1}^{\infty} r_{\tau,k,m}(\theta)\right)\pi_{j-k}(d-d_0)$ is $O\left((1+\log(t+1))^2(t+1)^{\max(-d+d_0,-\zeta)-1}\right)$ for $j \le t$, and $O\left((1+\log j)^3 j^{\max(-d+d_0,-\zeta)-1}\right)$ for $j > t$, since $d - d_0 > -1/2$ for all $\theta \in \Theta_3(\kappa_3)$. Analogously, the other coefficient $\phi_{3,\epsilon,j}(\theta,t) = \sum_{k=0}^{\min(j,t)}\left(\sum_{m=t+1}^{\infty} r_{\tau,k,m}(\theta)\right)\sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d)$ is $O\left((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\right)$ for $j \le t$, and $O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right)$ for $j > t$. Together, (B.8), (B.9), (B.10) and the rates established below prove (B.7).

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

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

For the first sum, note that

$$\sum_{j,k=1}^{\infty} \phi_{\eta,j}(\theta,t)\phi_{\eta,k}(\theta,t)\,\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k})$$

$$\leq K(1+\log(t+1))^4(t+1)^{2\max(-d+d_0,-\zeta)-2}\sum_{j=1}^{t}\sum_{k=1}^{t}\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k})$$

$$+ K(1+\log(t+1))^2(t+1)^{\max(-d+d_0,-\zeta)-1}\sum_{j=1}^{t}\sum_{k=t+1}^{\infty}(1+\log k)^3k^{\max(-d+d_0,-\zeta)-1}\,\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k})$$

$$+ K\sum_{j=t+1}^{\infty}\sum_{k=t+1}^{\infty}(1+\log j)^3j^{\max(-d+d_0,-\zeta)-1}(1+\log k)^3k^{\max(-d+d_0,-\zeta)-1}\,\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k}).$$

By absolute summability of the autocovariance function it holds that $\sum_{k=1}^{t}\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k}) = O(1)$. Consequently, it holds for the first term that $\sum_{j=1}^{t}\sum_{k=1}^{t}\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k}) = O(t)$. This, together with $d - d_0 > -1/2$ for all $\theta \in \Theta_3(\kappa_3)$, yields that the first term is $o(1)$. For the second term it holds that $\sum_{j=1}^{t}\sum_{k=t+1}^{\infty}(1+\log k)^3k^{\max(-d+d_0,-\zeta)-1}\,\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k}) = \sum_{k=t+1}^{\infty}(1+\log k)^3k^{\max(-d+d_0,-\zeta)-1}O(1) = O((1+\log(t+1))^3(t+1)^{\max(-d+d_0,-\zeta)})$. Again, by $d-d_0 > -1/2$ for all $\theta \in \Theta_3(\kappa_3)$, the second term is $o(1)$. Finally, for the third sum it holds by absolute summability of the autocovariance function that $\sum_{j=t+1}^{\infty}\sum_{k=t+1}^{\infty}(1+\log j)^3j^{\max(-d+d_0,-\zeta)-1}(1+\log k)^3k^{\max(-d+d_0,-\zeta)-1}\,\mathrm{E}(\eta_{t+1-j}\eta_{t+1-k}) \leq K(1+\log(t+1))^3(t+1)^{\max(-d+d_0,-\zeta)-1}\sum_{j=t+1}^{\infty}(1+\log j)^3j^{\max(-d+d_0,-\zeta)-1} \leq K(1+\log(t+1))^6(t+1)^{2\max(-d+d_0,-\zeta)-1} = o(1)$. Similarly, it can be shown that $\sum_{j,k=1}^{\infty}\phi_{\epsilon,j}(\theta,t)\phi_{\epsilon,k}(\theta,t)\,\mathrm{E}(\epsilon_{t+1-j}\epsilon_{t+1-k}) = o(1)$. Since $\zeta > 0$ and $d-d_0+1/2 > \kappa_3 > 0$ for all $\theta \in \Theta_3(\kappa_3)$, (B.5) follows directly. From the law of large numbers for stationary and ergodic processes, (B.6) follows immediately.

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

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

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and by Nielsen (2015, lemma A.3) the latter sum is bounded from below by $\sum_{j=0}^{n^{\alpha}-1}\pi_j^2(d-d_0)\geq 1+K\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}$ for some $K>0$. The limit of the fraction $\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}$ is discussed by Nielsen (2015, p. 165): It increases in n from zero (for $n=2$) to $1/(2\kappa_3)$ as $n\to\infty$, and decreases in κ_3 from $\alpha\log(n-1)$ for $\kappa_3=0$ to zero for $\kappa_3\to 1/2$. Consequently $\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}\to\infty$ as $(n,\kappa_3)\to(\infty,0)$. This, together with (B.19), (B.26), and (B.27) yields that the lower bound of $\frac{1}{n}\sum_{t=1}^{n}(\Delta_{+}^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j})^2$ diverges in probability for $\theta\in\Theta_2(\kappa,\kappa)$ as $(n,\kappa)\to(\infty,0)$. By (B.13), (B.14), and (B.15) the result of Nielsen (2015, eqn. 25) for ARFIMA models carries over to the fractional UC model: For any $K>0$, $\delta>0$, there exist $\bar{\kappa}_3>0$ and $T_2\geq 1$ such that

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C Proof of theorem 4.2

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

□

D Additional lemmas

Lemma D.1 (Convergence rates of $\pi_j(d)$, $b_j(\varphi)$, and related vector and matrix entries)**.** *For some vector z, let $z_{(j)}$ denote its j-th entry, and for some matrix Z, let $Z_{(i,j)}$ denote the (i,j)-th entry (i.e. the entry in row i and column j). Then*

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

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

$$(B'_{\varphi,t}B_{\varphi,t})_{(i,j)} = O(|i-j|^{-\zeta-1}) \quad \textit{for } i \neq j \quad \textit{and} \quad (B'_{\varphi,t}B_{\varphi,t})_{(i,i)} = O(1), \tag{D.3}$$

$$(S'_{d,t}S_{d,t})_{(i,j)} = O(|i-j|^{-d-1}) \quad \textit{for } i \neq j \quad \textit{and} \quad (S'_{d,t}S_{d,t})_{(i,i)} = O(1), \tag{D.4}$$

$$(B'_{\varphi,t}B_{\varphi,t})^{-1}_{(i,j)} = O(|i-j|^{-\zeta-1}) \quad \textit{for } i \neq j \quad \textit{and} \quad (B'_{\varphi,t}B_{\varphi,t})^{-1}_{(i,i)} = O(1), \tag{D.5}$$

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

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

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

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

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

To derive the convergence rates for the entries of $(B'_{\varphi,t}B_{\varphi,t})^{-1}$ and $(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}$ in (D.5) and (D.6), note that as $t \to \infty$, $B'_{\varphi,t}B_{\varphi,t}$ and $B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t}$ converge to the Toeplitz matrices[4] $T_t(f_1)$ and $T_t(f_2)$ with symbols $f_1(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_1(j)e^{i\lambda j}$, $\gamma_1(j) = \sum_{k=0}^{\infty} b_k(\varphi)b_{k+j}(\varphi)$, and $f_2(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_2(j)e^{i\lambda j}$, $\gamma_2(j) = \sum_{k=0}^{\infty}[b_k(\varphi)b_{k+j}(\varphi) + \nu\pi_k(d)\pi_{k+j}(d)]$, where $\gamma_1(j) = O(j^{-\zeta-1})$ and $\gamma_2(j) = O(j^{\max(-d,-\zeta)-1})$ as $j \to \infty$. Consequently, $(B'_{\varphi,t}B_{\varphi,t})^{-1}$ and $(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t})^{-1}$ converge to the Toeplitz matrices $T_t(1/f_1)$ and $T_t(1/f_2)$ that exist by assumption 3. Denote the respective spectral densities as $1/f_1(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_3(j)e^{i\lambda j}$ and $1/f_4(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_4(j)e^{i\lambda j}$. Then the convergence rate of $\gamma_3(j)$ can be obtained from the partial derivative $(\partial/\partial\lambda)[1/f_1(\lambda)] = (2\pi)^{-1}\sum_{j=0}^{\infty} ij\gamma_3(j)e^{i\lambda j} = -f_1(\lambda)^{-2}(2\pi)^{-1}\sum_{j=0}^{\infty} ij\gamma_1(j)e^{i\lambda j}$, where $j\gamma_1(j) = O(j^{-\zeta})$, so that $j\gamma_3(j) = O(j^{-\zeta})$ as $f_1(\lambda)$ is bounded away from zero by assumption 3. It follows that $\gamma_3(j) = O(j^{-\zeta-1})$. Similarly, it can be shown that $\gamma_4(j) = O(j^{\max(-d,-\zeta)-1})$. As the j-th descending diagonals of $(B'_{\varphi,t}B_{\varphi,t})^{-1}$ and $(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t})^{-1}$ converge to $\gamma_3(j)$ and $\gamma_4(j)$ as $t \to \infty$, one has (D.5) and (D.6).

(D.7) follows immediately from (D.2), since $(B'_{\varphi,t}\beta_t)_{(j)} = \sum_{k=0}^{j-1} b_k(\varphi)b_{t-j+k+1}(\varphi) = O((t-j+1)^{-\zeta-1})\sum_{k=0}^{j-1} b_k(\varphi) = O((t-j+1)^{-\zeta-1})$, while (D.8) follows from (D.1) by $(S'_{d,t}s_{t+1})_{(j)} = \sum_{k=0}^{j-1}\pi_k(d)\pi_{t-j+k+1}(d) = O((t-j+1)^{-d-1})\sum_{k=0}^{j-1}\pi_k(d) = O((t-j+1)^{-d-1})$. □

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

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

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

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

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

The left term in the sum of (D.10) is

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

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

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

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

$$\begin{aligned}&\left[\begin{pmatrix} b_1(\varphi)-\pi_1(d) & \cdots & b_t(\varphi)-\pi_t(d)\end{pmatrix}(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}S'_{d,t}\right]_{(j)} \\ &= \sum_{k=j}^{t}\left[\begin{pmatrix} b_1(\varphi)-\pi_1(d) & \cdots & b_t(\varphi)-\pi_t(d)\end{pmatrix}(B'_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}\right]_{(k)}\pi_{k-j}(d) \\ &= O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right)\pi_0(d)+O\left(\sum_{k=j+1}^{t}(1+\log k)k^{\max(-d,-\zeta)-1}(k-j)^{-d-1}\right) \\ &= O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right)+O\left((1+\log j)j^{\max(-d,-\zeta)-1}\sum_{k=1}^{t-j}k^{-d-1}\right)\end{aligned}$$

$$= O\left((1+\log j) j^{\max(-d,-\zeta)-1}\right), \tag{D.13}$$

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

□

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

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

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

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

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

with $\beta_t = (b_t(\varphi)\cdots b_1(\varphi))'$ and $s_t = (\pi_t(d)\cdots\pi_1(d))'$, and 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

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

with block entries $R_3 = \left[(1+\beta_t'\beta_t+\nu+\nu s_t's_t) - (\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t)\right]^{-1}$, $R_1 = R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t)(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)$, and $R_2 = -R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t)$. Clearly $R_3 = O(1)$, since by (D.6), (D.7) and (D.8)

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

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

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

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. Next, consider R_2, for which by (D.17) it follows that $R_{2_{(j)}}=O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$. Finally, for R_1, it follows by (D.17) 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

$$
\begin{aligned}
&(b_1(\varphi)-\pi_1(d)\cdots b_{t+1}(\varphi)-\pi_{t+1}(d))(B_{\varphi,t+1}'B_{\varphi,t+1}+\nu S_{d,t+1}'S_{d,t+1})^{-1}\\
&\quad=\Big((b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))[\Xi_t(\theta)+R_1]+(b_{t+1}(\varphi)-\pi_{t+1}(d))R_2' \qquad R_4\Big),
\end{aligned} \tag{D.18}
$$

where $R_4=(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2+(b_{t+1}(\varphi)-\pi_{t+1}(d))R_3$. By (D.1) and (D.2), it holds for the terms in R_4 that $[b_{t+1}(\varphi)-\pi_{t+1}(d)]R_3=O((t+1)^{\max(-d,-\zeta)-1})$, and $(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2=O\big(\sum_{j=1}^t j^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big)=O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Thus $R_4=O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Analogously, for the other terms in (D.18) one has $[(b_{t+1}(\varphi)-\pi_{t+1}(d))R_2']_{(j)}=O\big((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big)$, and $[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1]_{(j)}=O\big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\sum_{i=1}^t(1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1}i^{\max(-d,-\zeta)-1}\big)=O\big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}(1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Therefore, for $j=1,...,t$, the whole term $\tau_j(\theta,t+1)$ is

$$
\tau_j(\theta,t+1)=\nu\left((b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S_{d,t}'+R_5'\right)_{(j)}=\tau_j(\theta,t)+\nu R_{5_{(j)}}, \tag{D.19}
$$

where $R_5'=[b_{t+1}(\varphi)-\pi_{t+1}(d)]R_2'S_{d,t}'+(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2s_t'+(b_{t+1}(\varphi)-\pi_{t+1}(d))R_3s_t'+(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1S_{d,t}'$. For R_5, note that

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

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 $[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2s_t']_{(j)}=O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}(t+1-j)^{-d-1}\big)$, and $[(b_{t+1}(\varphi)-\pi_{t+1}(d))R_3(\varphi)s_t']_{(j)}=O((t+1)^{\max(-d,-\zeta)-1}(t+1-j)^{-d-1})$. Furthermore

$$
[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1S_{d,t}']_{(j)}=\sum_{i=j}^{t}[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1]_{(i)}\pi_{i-j}(d)
$$

$$
\begin{aligned}
&= [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))R_1]_{(j)} + \sum_{i=1}^{t-j} [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))R_1]_{(i+j)} \pi_i(d) \\
&= O\Big((1 + \log(t+1))^2 (t+1)^{\max(-d,-\zeta)-1} (1 + \log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1} \Big) \\
&+ O\left((1 + \log(t+1))^2 (t+1)^{\max(-d,-\zeta)-1} (1 + \log(t+1-j)) \sum_{i=1}^{t-j} (t+1-i-j)^{\max(-d,-\zeta)-1} i^{-d-1} \right) \\
&= O\left((1 + \log(t+1))^2 (t+1)^{\max(-d,-\zeta)-1} (1 + \log(t+1-j))^2 (t+1-j)^{\max(-d,-\zeta)-1} \right).
\end{aligned}
$$

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

□

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

$$
\begin{aligned}
\frac{\partial \tau_j(\theta, t)}{\partial d} &= O\left((1 + \log j)^4 j^{\max(-d,-\zeta)-1} \right), && \text{(D.20)} \\
\frac{\partial \tau_j(\theta, t)}{\partial \nu} &= O\left((1 + \log j)^3 j^{\max(-d,-\zeta)-1} \right), && \text{(D.21)} \\
\frac{\partial \tau_j(\theta, t)}{\partial \varphi_{(l)}} &= O\left((1 + \log j)^3 j^{\max(-d,-\zeta)-1} \right), && \text{(D.22)}
\end{aligned}
$$

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

Proof of Lemma D.4. In the following, 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

$$
[\dot{S}'_{d,t} S_{d,t}]_{(i,j)} = \begin{cases} \sum_{k=1}^{i-1} \dot{\pi}_k(d) \pi_{k+j-i}(d) = O((1+j-i)^{-d-1}) & \text{if } i \le j, \\ \sum_{k=0}^{j-1} \pi_k(d) \dot{\pi}_{k+i-j}(d) = O((1+\log(i-j))(i-j)^{-d-1}) & \text{if } i > j, \end{cases} \qquad \text{(D.23)}
$$

as well as

$$
[\dot{B}'_{\varphi_{(l)},t} B_{\varphi,t}]_{(i,j)} = \begin{cases} \sum_{k=1}^{i-1} \dot{b}_k(\varphi_{(l)}) b_{k+j-i}(\varphi) = O((1+j-i)^{-\zeta-1}) & \text{if } i \le j, \\ \sum_{k=0}^{j-1} b_k(\varphi) \dot{b}_{k+i-j}(\varphi_{(l)}) = O((i-j)^{-\zeta-1}) & \text{if } i > j. \end{cases} \qquad \text{(D.24)}
$$

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

$$\begin{aligned}\frac{\partial\tau_j(\theta,t)}{\partial d} &= -\nu^2[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)S'_{d,t}]_{(j)} \\ &+\nu[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)\dot{S}'_{d,t}]_{(j)} - \nu[(\dot{\pi}_1(d)\cdots\dot{\pi}_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)}.\end{aligned} \tag{D.25}$$

For the first term, note that by (D.23) $[\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t}]_{(i,j)} = [\dot{S}'_{d,t}S_{d,t}]_{(i,j)} + [\dot{S}'_{d,t}S_{d,t}]_{(j,i)} = O((1+\log|i-j|)|i-j|^{-d-1})$ for $i\neq j$, and $[\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t}]_{(i,i)} = O(1)$. Together with (D.12) it follows for the first terms in (D.25) that

$$\begin{aligned}&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t})]_{(j)} = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right) \\ &+O\left(\sum_{i=1}^{j-1}(1+\log i)i^{\max(-d,-\zeta)-1}(1+\log(j-i))(j-i)^{-d-1}\right) \\ &+O\left(\sum_{i=j+1}^{t}(1+\log i)i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\right) \\ &= O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right) + O\left((1+\log j)^2\sum_{i=1}^{j-1}i^{\max(-d,-\zeta)-1}(j-i)^{-d-1}\right) \\ &+O\left((1+\log j)j^{\max(-d,-\zeta)-1}\sum_{i=1}^{t-j}(1+\log i)i^{-d-1}\right) = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right),\end{aligned} \tag{D.26}$$

where for the last equality, note that the second term satisfies $\sum_{i=1}^{j-1} i^{\max(-d,-\zeta)-1}(j-i)^{-d-1} = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right)$, see Johansen and Nielsen (2010, lemma B.4), and that it dominates the first and third term above. Taking into account the next product term for the first term in (D.25), by (D.4) and (D.26)

$$\begin{aligned}&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)]_{(j)} \\ &\quad= O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right) + O\left(\sum_{i=1}^{j-1}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\right) \\ &\quad+O\left(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{\max(-d,-\zeta)-1}\right) \\ &\quad= O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right),\end{aligned} \tag{D.27}$$

where the proof is the same as for (D.26) besides the additional log-factor. Adding the last term, it follows by (D.1) and (D.27) that

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

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

$$= O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right) + O\left(\sum_{i=j+1}^{t}(1+\log i)^4 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\right)$$

$$= O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right) + O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\sum_{i=1}^{t-j} i^{-d-1}\right)$$

$$= O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \tag{D.28}$$

where the second equality uses $\pi_0(d) = 1$ to obtain the first term, while the last equality uses $\sum_{i=1}^{t-j} i^{-d-1} = O(1)$, which holds for all $d > 0$. Consequently, the first term in (D.25) is $O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right)$. Turning to the second term in (D.25), by (D.12)

$$[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)\dot{S}'_{d,t}]_{(j)} = \sum_{i=j+1}^{t}[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)]_{(i)}\dot{\pi}_{i-j}(d)$$

$$= O\left(\sum_{i=j+1}^{t}(1+\log i) i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\right)$$

$$= O\left((1+\log j) j^{\max(-d,-\zeta)-1}\sum_{i=1}^{t-j}(1+\log i) i^{-d-1}\right) = O\left((1+\log j) j^{\max(-d,-\zeta)-1}\right), \tag{D.29}$$

where the last equality follows from $\sum_{i=1}^{t-j}(1+\log i)i^{-d-1} = O(1)$ for all $d > 0$. By an analogous proof, the third term in (D.25) is

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

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

$$= O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right). \tag{D.30}$$

Together, (D.28), (D.29), and (D.30) yield (D.20).

To prove (D.21), consider the partial derivatives $\partial\tau_j(\theta,t)/\partial\nu$, for which

$$\frac{\partial\tau_j(\theta,t)}{\partial\nu} = [(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)} \tag{D.31}$$

$$- \nu[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)S'_{d,t}]_{(j)}. \tag{D.32}$$

By (D.13) the first term (D.31) is $O\big((1+\log j)j^{\max(-d,-\zeta)-1}\big)$, while by (D.4) and (D.12), for the second term (D.32) it holds that

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

and the proof is analogous to (D.26) besides one log-factor. Furthermore, by (D.6) and (D.33)

$$
\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)]_{(j)} = O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right)\\
&+O\left(\sum_{i=1}^{j-1}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(j-i)^{-d-1}\right)\\
&+O\left(\sum_{i=j+1}^{t}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\right) = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right),
\end{aligned}
\tag{D.34}
$$

where again the proof is analogous to (D.27) besides one log-factor. From (D.1), (D.2), and (D.34) it then follows for (D.32) that

$$
\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)S'_{d,t}]_{(j)} = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right)\\
&+O\left(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\right)\\
&= O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right),
\end{aligned}
\tag{D.35}
$$

and the proof can be carried out the same way as (D.28). Thus, (D.21) holds.

Turning to (D.22), consider the partial derivatives $\partial\tau_j(\theta,t)/\partial\varphi_{(l)}$, where

$$
\frac{\partial\tau_j(\theta,t)}{\partial\varphi_{(l)}} = \nu[(\dot{b}_1(\varphi_{(l)})\cdots\dot{b}_t(\varphi_{(l)}))\Xi_t(\theta)S'_{d,t}]_{(j)} \tag{D.36}
$$

$$
-\nu[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta)S'_{d,t}]_{(j)}. \tag{D.37}
$$

By assumption 3, the partial derivatives are of order $\dot{b}_j(\varphi_{(l)}) = \partial b_j(\varphi)/\partial\varphi_{(l)} = O(j^{-\zeta-1})$, so that for the first term (D.36), analogously to (D.12)

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

and, analogously to (D.13)

$$[(\dot{b}_1(\varphi_{(l)}) \cdots \dot{b}_t(\varphi_{(l)}))\Xi_t(\theta)S_{d,t}]_{(j)} = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right), \tag{D.38}$$

so that (D.38) determines the rate of (D.36). Next, consider (D.37), for which one has by (D.12) and (D.24)

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

where the proof is identical to (D.26). By the same proof as for (D.27), by (D.6) and (D.39)

$$\begin{aligned}
&[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t} + B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta)]_{(j)} \\
&= O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right) + O\left(\sum_{i=1}^{j-1}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\right) \\
&+ O\left(\sum_{i=j+1}^{t}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(i-j)^{\max(-d,-\zeta)-1}\right) \\
&= O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right).
\end{aligned} \tag{D.40}$$

Finally, again by using the same proof as for (D.28), by (D.1) and (D.39)

$$\begin{aligned}
&[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t} + B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta)S'_{d,t}]_{(j)} \\
&= O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right) + O\left(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\right) \\
&= O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right).
\end{aligned} \tag{D.41}$$

Together, (D.38) and (D.41) yield (D.22). □

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

$$\frac{\partial \tau_j(\theta,t)}{\partial\theta}\Big|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0} = \frac{\partial}{\partial\theta}\sum_{k=t+1}^{\infty} r_{\tau,j,k}(\theta)\Big|_{\theta=\theta_0} = O\left((1+\log t)^4 t^{\max(-d_0-\zeta)-1}\right), \tag{D.42}$$

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

Proof of lemma D.5. From (D.19) and below $r_{\tau,j,t+1}(\theta) = -\nu R_{5_{(j)}}$, where

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

and with $B_{\varphi,t}$ and $S_{d,t}$ as defined in (5), $\beta_t' = (b_t(\varphi) \cdots b_1(\varphi))$, $s_t' = (\pi_t(d) \cdots \pi_1(d))$ as given in lemma D.1, and $R_3 = \left[(1 + \beta_t'\beta_t + \nu + \nu s_t' s_t) - (\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)(B_{\varphi,t}'\beta_t + \nu S_{d,t}' s_t)\right]^{-1}$, $R_1 = R_3 \Xi_t(\theta)(B_{\varphi,t}'\beta_t + \nu S_{d,t}' s_t)(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)$, and $R_2 = -R_3 \Xi_t(\theta)(B_{\varphi,t}'\beta_t + \nu S_{d,t}' s_t)$ as stated below (D.16). The partial derivative of $R_{5_{(j)}}$ w.r.t. the l-th entry $\theta_{(l)}$ is thus given by

$$\frac{\partial R_{5_{(j)}}}{\partial \theta_{(l)}} = \left[\frac{\partial (b_{t+1}(\varphi) - \pi_{t+1}(d))}{\partial \theta_{(l)}} \left(R_2' S_{d,t}' + R_3 s_t'\right)\right]_{(j)} \tag{D.43}$$

$$+ \left[\left(\frac{\partial (b_1(\varphi) - \pi_1(d))}{\partial \theta_{(l)}} \cdots \frac{\partial (b_t(\varphi) - \pi_t(d))}{\partial \theta_{(l)}}\right) \left(R_2 s_t' + R_1 S_{d,t}'\right)\right]_{(j)} \tag{D.44}$$

$$+ \left[(b_{t+1}(\varphi) - \pi_{t+1}(d)) \left(R_2' \frac{\partial S_{d,t}'}{\partial \theta_{(l)}} + R_3 \frac{\partial s_t'}{\partial \theta_{(l)}}\right)\right]_{(j)} \tag{D.45}$$

$$+ \left[((b_1(\varphi) - \pi_1(d)) \cdots (b_t(\varphi) - \pi_t(d))) \left(R_2 \frac{\partial s_t'}{\partial \theta_{(l)}} + R_1 \frac{\partial S_{d,t}'}{\partial \theta_{(l)}}\right)\right]_{(j)} \tag{D.46}$$

$$+ \left[(b_{t+1}(\varphi) - \pi_{t+1}(d)) \left(\frac{\partial R_2'}{\partial \theta_{(l)}} S_{d,t}' + \frac{\partial R_3}{\partial \theta_{(l)}} s_t'\right)\right]_{(j)} \tag{D.47}$$

$$+ \left[((b_1(\varphi) - \pi_1(d)) \cdots (b_t(\varphi) - \pi_t(d))) \left(\frac{\partial R_2}{\partial \theta_{(l)}} s_t' + \frac{\partial R_1}{\partial \theta_{(l)}} S_{d,t}'\right)\right]_{(j)}. \tag{D.48}$$

As noted in the proof of lemma D.4, the partial derivative of $\pi_j(d)$ only adds a log-factor to the convergence rate of $\pi_j(d)$, i.e. $\partial \pi_j(d)/\partial d = O((1 + \log j) j^{-d-1})$, see Johansen and Nielsen (2010, lemma B.3), while $\partial b_j(\varphi)/\partial \varphi_{(l)} = O(j^{-\zeta-1})$ by assumption 3. Thus, the convergence rates of (D.43) and (D.44) can be derived analogously to the proof of lemma D.3. This yields that (D.43) is $O((1 + \log(t+1))(t+1)^{\max(-d,-\zeta)-1}(1 + \log(t+1-j))^2 (t+1-j)^{\max(-d,-\zeta)-1})$, while (D.44) is $O\big((1 + \log(t+1))^3 (t+1)^{\max(-d,-\zeta)-1}(1 + \log(t+1-j))^2 (t+1-j)^{\max(-d,-\zeta)-1}\big)$, and the additional $(1 + \log(t+1))$ term stems from $\partial \pi_j(d)/\partial d$. Analogously, the partial derivatives of s_t and $S_{d,t}$ only add a log-factor to the convergence rates as derived in the proof of lemma D.3. Thus, it holds that (D.45) is $O((t+1)^{\max(-d,-\zeta)-1}(1 + \log(t+1-j))^3 (t+1-j)^{\max(-d,-\zeta)-1})$, while (D.46) is $O\big((1 + \log(t+1))^2 (t+1)^{\max(-d,-\zeta)-1}(1 + \log(t+1-j))^3 (t+1-j)^{\max(-d,-\zeta)-1}\big)$, and the additional $(1 + \log(t+1-j))$ term stems from $\partial s_t'/\partial d$ and $\partial S_{d,t}'/\partial d$. For the last two terms (D.47) and (D.48), note that $R_3 = O(1)$ as shown in (D.17) and below. Since $\beta_t'(\partial \beta_t/\partial \theta_{(l)})$, $s_t'(\partial s_t/\partial \theta_{(l)})$, $s_t' s_t$, $(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)\partial(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})'/\partial \theta_{(l)}$, and $(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})(\partial \Xi_t(\theta)/\partial \theta_{(l)})(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})'$ are $O(1)$, it follows that $\partial R_3/\partial \theta_{(l)} = -(R_3)^2(\partial/\partial \theta_{(l)}) \left[(1 + \beta_t'\beta_t + \nu + \nu s_t' s_t) - (\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)(B_{\varphi,t}'\beta_t + \nu S_{d,t}' s_t)\right] = O(1)$. For the partial derivatives of $R_{2_{(j)}}$, consider

$$\frac{\partial R_{2_{(j)}}}{\partial\theta_{(l)}} = -\frac{\partial R_3}{\partial\theta_{(l)}}\left[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)\right]_{(j)} - R_3\left[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\frac{\partial\Xi_t(\theta)}{\partial\theta_{(l)}}\right]_{(j)} \tag{D.49}$$

$$- R_3\left[\left(\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}} + \frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t} + \frac{\partial\nu}{\partial\theta_{(l)}}s_t' S_{d,t} + \nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t} + \nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\right)\Xi_t(\theta)\right]_{(j)}, \tag{D.50}$$

where the first term in (D.49) is $O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$ by (D.17) and by $\partial R_3/\partial\theta_{(l)} = O(1)$. For the second term in (D.49), one has $[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)]_{(j)} = O((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1})$ from (D.17). Together with $\partial\Xi_t(\theta)/\partial\theta_{(l)} = -\Xi_t(\theta)[(\partial/\partial\theta_{(l)})(B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t})]\Xi_t(\theta)$, (D.23) and (D.24), it follows that

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

Finally, using (D.6), one obtains

$$\begin{aligned}
&\left\{(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)\left[\frac{\partial}{\partial\theta_{(l)}}\left(B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t}\right)\right]\Xi_t(\theta)\right\}_{(j)} \\
&= O\left((1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1}\right),
\end{aligned} \tag{D.51}$$

which yields the binding rate of convergence for the second term in (D.49). For (D.50)

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

by lemma D.1. Hence, using (D.6) yields an upper bound for (D.50)

$$\begin{aligned}
&\left[\left(\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}} + \frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t} + \frac{\partial\nu}{\partial\theta_{(l)}}s_t' S_{d,t} + \nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t} + \nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\right)\Xi_t(\theta)\right]_{(j)} \\
&= O\left((1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1}\right).
\end{aligned} \tag{D.52}$$

Together, the rates of convergence of (D.49) and (D.50) yield

$$\frac{\partial R_{2(j)}}{\partial \theta_{(l)}} = O\left((1+\log(t+1-j))^3 (t+1-j)^{\max(-d,-\zeta)-1}\right). \tag{D.53}$$

For the partial derivatives of R_1, note that

$$\frac{\partial R_{1(i,j)}}{\partial \theta_{(l)}} = -\frac{\partial R_{2(i)}}{\partial \theta_{(l)}} \left[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)\right]_{(j)} - R_{2(i)} \left[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\frac{\partial \Xi_t(\theta)}{\partial \theta_{(l)}}\right]_{(j)} \tag{D.54}$$

$$- R_{2(i)} \left[\left(\beta_t' \frac{\partial B_{\varphi,t}}{\partial \theta_{(l)}} + \frac{\partial \beta_t'}{\partial \theta_{(l)}} B_{\varphi,t} + \frac{\partial \nu}{\partial \theta_{(l)}} s_t' S_{d,t} + \nu \frac{\partial s_t'}{\partial \theta_{(l)}} S_{d,t} + \nu s_t' \frac{\partial S_{d,t}}{\partial \theta_{(l)}}\right) \Xi_t(\theta)\right]_{(j)}. \tag{D.55}$$

From (D.17) and (D.53), the first term in (D.54) is $O((1+\log(t+1-i))^3(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1})$. Similarly, using (D.51) and the convergence rate of $R_{2(i)}$ as derived in the proof of lemma D.3, the second term in (D.54) is $O((1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1})$. By (D.52), it follows that (D.55) 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

$$\frac{\partial R_{1(i,j)}}{\partial \theta_{(l)}} = O\left((1+\log(t+1-i))^3(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^3(t+1-j)^{\max(-d,-\zeta)-1}\right). \tag{D.56}$$

With (D.53) at hand, it follows directly for (D.47) that

$$\left(\frac{\partial R_2'}{\partial \theta_{(l)}} S_{d,t}' + \frac{\partial R_3}{\partial \theta_{(l)}} s_t'\right)_{(j)} = O\left((1+\log(t+1-j))^4 (t+1-j)^{\max(-d,-\zeta)-1}\right).$$

By (D.1) and (D.2), (D.47) is then $O\left((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1}\right)$. For (D.48), it follows from (D.53) and (D.56) that $\left(\frac{\partial R_2}{\partial \theta_{(l)}} s_t' + \frac{\partial R_1}{\partial \theta_{(l)}} S_{d,t}'\right)_{(i,j)} = O\big((1+\log(t+1-i))^3(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1}\big)$. Again using (D.1) and (D.2), it thus follows that (D.48) is $O\big((1+\log(t+1))^4(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1}\big)$. Together, this implies for (D.42) that

$$\frac{\partial r_{\tau,j,t+1}(\theta)}{\partial \theta_{(l)}} = O\big((1+\log(t+1))^4(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1}\big),$$

and thus $\frac{\partial}{\partial \theta} \sum_{k=t+1}^{\infty} r_{\tau,j,k}(\theta)\big|_{\theta=\theta_0} = O\left((1+\log t)^4 t^{\max(-d_0-\zeta)-1}\right)$. □

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

$$\sqrt{n}\left[\left.\frac{\partial \tilde{Q}(y,\theta)}{\partial \theta}\right|_{\theta=\theta_0} - \left.\frac{\partial Q(y,\theta)}{\partial \theta}\right|_{\theta=\theta_0}\right] = o_p(1). \tag{D.57}$$

Proof of lemma D.6. To simplify the notation, denote $h_{1,t} = \sum_{j=1}^{t-1} \frac{\partial \tau_j(\theta,t)}{\partial\theta}\Big|_{\theta=\theta_0} \xi_{t-j}(d_0)$, $\tilde{h}_{1,t} = \sum_{j=1}^{\infty} \frac{\partial \tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0} \tilde{\xi}_{t-j}(d_0)$, $h_{2,t} = \sum_{j=0}^{t-1} \tau_j(\theta_0,t) \frac{\partial \xi_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$, and $\tilde{h}_{2,t} = \sum_{j=0}^{\infty} \tau_j(\theta_0) \frac{\partial \tilde{\xi}_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$. Then plugging (C.2), (C.3) into (D.57) and using (B.11) yields

$$\begin{aligned} &\sqrt{n}\left[\frac{\partial \tilde{Q}(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} - \frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\right] \\ &= \frac{2}{\sqrt{n}}\left[\sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{1,t} - h_{1,t}) + \sum_{t=1}^{n} h_{1,t}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right)\right] \\ &+ \frac{2}{\sqrt{n}}\left[\sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{2,t} - h_{2,t}) + \sum_{t=1}^{n} h_{2,t}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right)\right], \end{aligned} \tag{D.58}$$

so that it remains to be shown that all four terms in (D.58) are $o_p(1)$.

For the proofs it will be very useful to note that $\tilde{v}_t(\theta_0)$ adapted to the filtration $\mathcal{F}_t = \sigma(\tilde{\xi}_s, s \leq t)$ is a stationary martingale difference sequence (MDS) as can be seen from (B.2) and below: There, the projection of $\tilde{\xi}_t(d_0) = \tilde{\xi}_t$ on the span of all past $\tilde{\xi}_s$, $s < t$, is subtracted from $\tilde{\xi}_t$. Thus, $\tilde{v}_t(\theta_0)$ is orthogonal to all $\tilde{\xi}_s$ for $s < t$ and is a MDS when adapted to the sequence $\mathcal{F}_t$. Note in addition that all $\tilde{h}_{1,t}$, $\tilde{h}_{2,t}$ are $\mathcal{F}_{t-1}$-measurable, as $\tau_0 = \pi_0 = 1$ are invariant w.r.t. θ.

Starting with the first term of (D.58), by plugging in $h_{1,t}$ and $\tilde{h}_{1,t}$, I decompose

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

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

$$+ \frac{2}{\sqrt{n}}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)\sum_{j=t}^{\infty} \frac{\partial \tau_j(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0). \tag{D.61}$$

Note that $(\tilde{v}_t(\theta), \mathcal{F}_t)$ is a stationary MDS, while $\sum_{j=t}^{\infty} \frac{\partial \tau_j(\theta)}{\partial\theta}\big|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$ is $\mathcal{F}_{t-1}$-measurable. Hence, $\tilde{v}_t(\theta_0)\sum_{j=t}^{\infty} \frac{\partial \tau_j(\theta)}{\partial\theta}\big|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$ 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 D.4, it follows that (D.61) is $o_p(1)$. In (D.60), $\tilde{v}_t(\theta_0)\sum_{j=1}^{t-1}\left(\frac{\partial \tau_j(\theta)}{\partial\theta}\big|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta,t)}{\partial\theta}\big|_{\theta=\theta_0}\right)\tilde{\xi}_{t-j}(d_0)$ adapted to $\mathcal{F}_t$ is a MDS, while $\sum_{j=1}^{t-1}\left(\frac{\partial \tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta,t)}{\partial\theta}\Big|_{\theta=\theta_0}\right)\tilde{\xi}_{t-j}(d_0) = O_p((1+\log t)^4 t^{\max(-d_0,-\zeta)})$ by lemma D.5. Hence (D.60) is $o_p(1)$. For (D.59), note that

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

$$
\begin{aligned}
=&\mathrm{E}\Bigg[\sum_{s,t=1}^{n}\Big(\sum_{j=0}^{\infty}\eta^2_{\min(s,t)-j}\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0)\Big)\sum_{j=0}^{\infty}\epsilon^2_{-j}\Big(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\Big)\\
&\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\Bigg] \qquad \text{(D.62)}\\
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\left(\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\right)\\
&\times\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right) \qquad \text{(D.63)}\\
&\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\Bigg]\\
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\left(\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right)\\
&\times\left(\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right)\Bigg]. \qquad \text{(D.64)}
\end{aligned}
$$

For (D.62), 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 D.2, as well as $\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\big|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)=O\big((1+\log(t+j))^6(t+j)^{\max(-d_0,-\zeta)-1}\big)$, see (D.1) together with lemma D.4. This yields the upper bound for (D.62)

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

Similarly, for the second term (D.63), by (D.1) and lemma D.2 it holds that

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

Furthermore, by lemma D.4

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

so that by the same proof as for (D.62), it holds that (D.63) is also $O(1)$.

By (D.1) and lemmas D.2 and D.4, the third term (D.64) is bounded from above by

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

As all three terms (D.62) to (D.64) are $O(1)$, it follows directly by the scaling that (D.59) is $o_p(1)$. Now, since (D.59) to (D.61) are $o_p(1)$, the first term in (D.58) is also $o_p(1)$.

Next, consider the third term in (D.58). I plug in $h_{2,t}$ and $\tilde{h}_{2,t}$ which gives

$$\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{2,t}-h_{2,t})=\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left(\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}-\frac{\partial\xi_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}\right) \tag{D.65}$$
$$+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0} \tag{D.66}$$
$$+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}. \tag{D.67}$$

(D.67) is $o_p(1)$, since $(\tilde{v}_t(\theta_0),\mathcal{F}_t)$ is a stationary MDS, and $\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}|_{\theta=\theta_0}$ is $\mathcal{F}_{t-1}$-measurable. Since $\partial\tilde{\xi}_{t-i}(d)/\partial\theta$ is $O_p(1)$ for all $d>d_0-1/2$, it follows by lemma D.2 that $\sum_{j=t}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}|_{\theta=\theta_0}=O_p((1+\log t)t^{\max(-d_0,-\zeta)})$, and thus (D.67) is $o_p(1)$.

For (D.66), note that $\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}|_{\theta=\theta_0}$ together with $\mathcal{F}_t$ is a MDS. Furthermore, by lemma D.3, $\tau_j(\theta_0)-\tau_j(\theta_0,t)=O((1+\log t)^2t^{\max(-d_0,-\zeta)-1})$. Since the partial derivatives of $\tilde{\xi}_t(d)$ are bounded in probability for all $d-d_0>-1/2$, it follows that $\sum_{j=0}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}|_{\theta=\theta_0}=O_p((1+\log t)^2t^{\max(-d_0,-\zeta)})$. Therefore, (D.66) is $o_p(1)$.

For (D.65), I use $\frac{\partial\pi_j(d-d_0)}{\partial d}|_{d=d_0}=-j^{-1}$ as shown by Robinson (2006, pp. 135-136) and Hualde and Robinson (2011, p. 3170). Thus, the partial derivative in (D.65) w.r.t. d is

$$\left.\frac{\partial\tilde{\xi}_t(\theta)}{\partial d}\right|_{\theta=\theta_0} - \left.\frac{\partial\xi_t(\theta)}{\partial d}\right|_{\theta=\theta_0} = -\sum_{j=t}^{\infty} j^{-1}\eta_{t-j} + \sum_{j=0}^{\infty}\epsilon_{-j}\sum_{k=0}^{j}\left.\frac{\partial\pi_{t+j-k}(d)}{\partial d}\right|_{\theta=\theta_0} a_k(\varphi_0). \tag{D.68}$$

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

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

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

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

$$\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\left\{\left[\sum_{j=t}^{\infty}\eta^2_{t-j}\tau_j(\theta_0)\sum_{k=0}^{t-1}\frac{-\tau_k(\theta_0,t)}{j-k} + \sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\right.\\
&\quad\times\left.\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]\left[\sum_{j=s}^{\infty}\eta^2_{s-j}\tau_j(\theta_0)\sum_{k=0}^{s-1}\frac{-\tau_k(\theta_0,s)}{j-k}\right.\\
&\quad\left.\left.+\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]\right\}.
\end{aligned} \tag{D.71}$$

For (D.69), note the first expectation is $\sigma^2_\eta\sum_{j=0}^{\min(s,t)-1}\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0) = O(|t-s|^{\max(-d_0,-\zeta)-1})$ for all $t\neq s$, and $O(1)$ for $t=s$, see lemma D.2. In addition, it holds for the other terms that $\mathrm{E}\big[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\frac{1}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\frac{1}{s+j-k}\right)\big] \leq K\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}$ and $\mathrm{E}\big[\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\frac{\partial\pi_{j+t-k-l}(d)}{\partial d}\big|_{\theta=\theta_0}\right)$ $\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\frac{\partial\pi_{j+s-k-l}(d)}{\partial d}\big|_{\theta=\theta_0}\right)\big] \leq K\sum_{j=0}^{\infty}(1+\log(t+j))^4(t+j)^{\max(-d_0,-\zeta)-1}(1+$

$\log(s+j))^4(s+j)^{\max(-d_0,-\zeta)-1}$ by lemma D.2. It follows that (D.69) is bounded from above by

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

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 (D.70), by lemmas D.1 and D.2, the first expectation is bounded by

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

Hence, by the same proof as for (D.69) the second term (D.70) is also $O(n^{\kappa})$, $0<\kappa<1/2$. For the third term (D.71) one has by lemma D.2

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

$$
+\sum_{s,t=1}^{n}\left(\sum_{j=t}^{\infty}O\left((1+\log j)^{7}j^{2\max(-d_0,-\zeta)-2}\right)\right)\left(\sum_{j=s}^{\infty}O\left((1+\log j)^{3}j^{\max(-d_0,-\zeta)-2}\right)\right)=O(1),
$$

and thus all terms (D.69) to (D.71) are $O(n^{\kappa})$. As (D.65) is appropriately scaled, it follows that (D.65) is $o_p(1)$ and thus the third term in (D.58) is $o_p(1)$.

Next, consider the second term in (D.58) that can be decomposed into

$$
\begin{aligned}
\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{1,t}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)&=\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{1,t}\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0)-\xi_{t-j}(d_0))\tau_j(\theta_0,t)\\
&+\frac{2}{\sqrt{n}}\sum_{t=0}^{n}h_{1,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\tilde{\xi}_{t-j}(d_0)+\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{1,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0).
\end{aligned}
\tag{D.72}
$$

For the first term in (D.72), note that

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

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

$$
\begin{aligned}
&\times\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\right]\\
&+\sum_{s,t=1}^{n}\mathrm{E}\left[\left(\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{t-1}\left.\frac{\partial\tau_k(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,t-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\right.\right.\\
&\times\left.\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right)\\
&\times\left.\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{s-1}\left.\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,s-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right)\right].
\end{aligned}
\tag{D.75}
$$

For (D.73), it holds that $\mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1}\frac{\partial\tau_j(\theta,\min(s,t))}{\partial\theta}\big|_{\theta=\theta_0}\frac{\partial\tau_{j+|t-s|}(\theta,\max(s,t))}{\partial\theta'}\big|_{\theta=\theta_0}\eta^2_{\min(s,t)-j}\right]=O(|t-s|^{\max(-d_0,-\zeta)-1})$ by lemma D.4. Furthermore, for (D.74), one has by lemmas D.1 and D.4 that the first term is bounded by $\mathrm{E}\big[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\big(\sum_{k=0}^{j}\frac{\partial\tau_k(\theta,\min(s,t))}{\partial\theta}\big|_{\theta=\theta_0}\sum_{l=0}^{j-k}\pi_l(d_0)a_{j-k-l}(\varphi_0)\big)$

$\big(\sum_{k=0}^{j+|t-s|} \frac{\partial \tau_k(\theta,\max(s,t))}{\partial\theta'}\big|_{\theta=\theta_0} \sum_{l=0}^{j+|t-s|-k} \pi_l(d_0)a_{j+|t-s|-k-l}(\varphi_0)\big)\big] = O(|t-s|^{\max(-d_0,-\zeta)-1})$. In addition, for both (D.73) and (D.74), by lemmas D.1 and D.2 the other remaining term is $\mathrm{E}\big[\sum_{j=0}^{\infty}\epsilon_{-j}^2\big(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\big)\big(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\big)\big] = O((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1})$. Consequently, both (D.73) and (D.74) are $\sum_{s,t=1}^{n} O\left((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1}|t-s|^{\max(-d_0,-\zeta)-1}\right) = O(1)$. Finally, by lemmas D.1, D.2, and D.4, (D.75) is

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

Thus, the first term in (D.72) is $o_p(1)$. For the second term in (D.72), note that by lemma D.3, $\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t)) \le K\sum_{j=1}^{t-1}\sum_{k=t+1}^{\infty}(1+\log k)^2(1+\log(k-j))^2 k^{\max(-d_0,-\zeta)-1}(k-j)^{\max(-d_0,-\zeta)-1} \le K\sum_{j=1}^{t-1}(1+\log t)^2 t^{\max(-d_0,-\zeta)-1}(1+\log(t-j))^2(t-j)^{\max(-d_0,-\zeta)} \le K(1+\log t)^2 t^{-1}\sum_{j=1}^{t-1} j^{\max(-d_0,-\zeta)}(t-j)^{\max(-d_0,-\zeta)}(1+\log(t-j))^2 \le K(1+\log t)^5 t^{\max(-d_0,-\zeta)-1}$, and thus $\frac{2}{\sqrt{n}}\sum_{t=1}^{n}h_{1,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0)-\tau_j(\theta_0,t))\tilde{\xi}_{t-j}(d_0) = o_p(1)$. For the third term in (D.72)

$$\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^{n}h_{1,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0)\right]^2\right\}\\
&=\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\eta_{\min(s,t)-j}^2\frac{\partial\tau_j(\theta,\min(s,t))}{\partial\theta}\bigg|_{\theta=\theta_0}\frac{\partial\tau_{j+|t-s|}(\theta,\max(s,t))}{\partial\theta'}\bigg|_{\theta=\theta_0}\right]\\
&\quad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\eta_{-j}^2\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)+\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{j}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\qquad\qquad(\mathrm{D.76})\\
&\quad\left.\times\left(\sum_{k=0}^{j}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right]\\
&+\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon_{\min(s,t)-j}^2\left(\sum_{k=0}^{j}\frac{\partial\tau_k(\theta,\min(s,t))}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-k}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\right.\\
&\quad\left.\times\left(\sum_{k=0}^{j+|t-s|}\frac{\partial\tau_k(\theta,\max(s,t))}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j+|t-s|-k}\pi_l(d_0)a_{j+|t-s|-k-l}(\varphi_0)\right)\right]\qquad\qquad(\mathrm{D.77})\\
&\quad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\eta_{-j}^2\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)+\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{j}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\quad\left.\times\left(\sum_{k=0}^{j}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\left[\left(\sum_{j=t}^{\infty} \epsilon_{t-j}^{2}\left(\sum_{k=0}^{t-1} \frac{\partial \tau_k(\theta,t)}{\partial \theta}\bigg|_{\theta=\theta_0} \sum_{l=0}^{\min(j-k,t-1)} \pi_l(d_0) a_{j-k-l}(\varphi_0)\right)\right.\right.\\
&\quad \times \left.\left(\sum_{k=0}^{j-t} \tau_{j+k}(\theta_0) \sum_{l=0}^{j-t-k} a_l(\varphi_0)\pi_{j-t-k-l}(d_0)\right)\right)\\
&\quad \times \left(\sum_{j=s}^{\infty} \epsilon_{s-j}^{2}\left(\sum_{k=0}^{s-1} \frac{\partial \tau_k(\theta,s)}{\partial \theta'}\bigg|_{\theta=\theta_0} \sum_{l=0}^{\min(j-k,s-1)} \pi_l(d_0) a_{j-k-l}(\varphi_0)\right)\right.\\
&\quad \times \left.\left.\left(\sum_{k=0}^{j-s} \tau_{j+k}(\theta_0) \sum_{l=0}^{j-s-k} a_l(\varphi_0)\pi_{j-s-k-l}(d_0)\right)\right)\right].
\end{aligned}
\tag{D.78}
$$

For (D.76) and (D.77), it holds that $\mathrm{E}[\sum_{j=0}^{\infty} \epsilon_{-j}^{2} \left(\sum_{k=0}^{j} \tau_{t+k}(\theta_0) \sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)$ $\times \left(\sum_{k=0}^{j} \tau_{s+k}(\theta_0) \sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)] = O((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1})$ and $\mathrm{E}\left[\sum_{j=0}^{\infty} \eta_{-j}^{2}\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)\right] = O((1+\log t)t^{\max(-d_0,-\zeta)}(1+\log s)s^{\max(-d_0,-\zeta)-1})$. Thus, analogously to (D.73) and (D.74), expressions (D.76) and (D.77) are $O(1)$. Also analogously to (D.75), by lemmas D.1, D.2, and D.4, (D.78) 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 (D.72) is $o_p(1)$. It follows that the second term in (D.58) is $o_p(1)$. Finally, consider the last term in (D.58)

$$
\begin{aligned}
\frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) &= \frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0))\tau_j(\theta_0,t)\\
&+ \frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0) - \tau_j(\theta_0,t))\tilde{\xi}_{t-j}(d_0) + \frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0).
\end{aligned}
\tag{D.79}
$$

For the first term in (D.79) it holds that

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

$$
\begin{aligned}
&+\sum_{s,t=1}^{n} \mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1} \epsilon_{\min(s,t)-j}^2 \left(\sum_{k=0}^{j} \tau_k(\theta_0, \min(s,t)) \sum_{l=0}^{j-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0)\right) \\
&\quad \times \left(\sum_{k=0}^{j+|t-s|} \tau_k(\theta_0, \max(s,t)) \sum_{l=0}^{j+|t-s|-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j+|t-s|-k-l}(\varphi_0)\right)\Bigg] \qquad \text{(D.81)} \\
&\quad \times \mathrm{E}\left[\sum_{j=0}^{\infty} \epsilon_{-j}^2 \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0)\right) \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0)\right)\right] \\
&+\sum_{s,t=1}^{n} \mathrm{E}\Bigg[\left(\sum_{j=t}^{\infty} \epsilon_{t-j}^2 \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{t-1-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0)\right) \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j-t} a_l(\varphi_0) \pi_{j-k-l}(d_0)\right)\right) \\
&\quad \times \left(\sum_{j=s}^{\infty} \epsilon_{s-j}^2 \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{s-1-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0)\right) \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j-s} a_l(\varphi_0) \pi_{j-k-l}(d_0)\right)\right)\Bigg], \qquad \text{(D.82)}
\end{aligned}
$$

while all other partial derivatives of $\xi_{t-j}(d)$ (i.e. those w.r.t. all other entries except d) are zero. Due to the convergence rates for $\tau_j(\theta, t)$ as derived in lemma D.2, the first term in (D.80) is $\mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1} \eta_{\min(s,t)-j}^2 \left(\sum_{k=1}^{j} \frac{1}{k} \tau_{j-k}(\theta_0, \min(s,t))\right) \left(\sum_{k=1}^{j+|t-s|} \frac{1}{k} \tau_{j+|t-s|-k}(\theta_0, \max(s,t))\right)\right] = O(|t-s|^{-1})$. In addition, by lemmas D.1 and D.2 it holds that the first term of (D.81) is

$$
\begin{aligned}
&\mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1} \epsilon_{\min(s,t)-j}^2 \left(\sum_{k=0}^{j} \tau_k(\theta_0, \min(s,t)) \sum_{l=0}^{j-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j-k-l}(\varphi_0)\right) \\
&\times \left(\sum_{k=0}^{j+|t-s|} \tau_k(\theta_0, \max(s,t)) \sum_{l=0}^{j+|t-s|-k} \frac{\partial \pi_l(d)}{\partial d}\bigg|_{\theta=\theta_0} a_{j+|t-s|-k-l}(\varphi_0)\right)\Bigg] = O(|t-s|^{\max(-d_0,-\zeta)-1}). \qquad \text{(D.83)}
\end{aligned}
$$

The second term in (D.80) and (D.81) is $\mathrm{E}[\sum_{j=0}^{\infty} \epsilon_{-j}^2 \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0)\right)$ $\left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0)\right)] = O((1+\log t)^3 t^{\max(-d_0,-\zeta)} (1+\log s)^3 s^{\max(-d_0,-\zeta)-1})$. Thus, analogously to (D.73), (D.74), (D.76) and (D.77), it holds that (D.80) and (D.81) are $O(1)$. Finally, (D.82) is bounded from above by

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

Hence, the first term in (D.79) is $o_p(1)$. For the second term in (D.79), by lemma D.3, $\sum_{j=1}^{t-1}(\tau_j(\theta_0) - \tau_j(\theta_0, t)) = O((1+\log t)^5 t^{\max(-d_0,-\zeta)-1})$ as already noted for the second term in (D.72), and thus

$\frac{2}{\sqrt{n}} \sum_{t=1}^{n} h_{2,t} \sum_{j=1}^{t-1} (\tau_j(\theta_0) - \tau_j(\theta_0, t))\tilde{\xi}_{t-j}(d_0) = o_p(1)$. For the third term in (D.72)

$$
\begin{aligned}
&\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^2_{\min(s,t)-j} \left(\sum_{k=1}^{j} \frac{1}{k}\tau_{j-k}(\theta_0, \min(s,t))\right) \left(\sum_{k=1}^{j+|t-s|} \frac{1}{k}\tau_{j+|t-s|-k}(\theta_0, \max(s,t))\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 \text{(D.84)} \\
&\quad \times \left.\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} \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. \\
&\quad \times \left.\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 \text{(D.85)} \\
&\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 \times \left.\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} \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) \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)\right. \\
&\quad \times \left.\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) \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]. \qquad \text{(D.86)}
\end{aligned}
$$

As noted above, the first expected value in (D.84) is $O(|t-s|^{-1})$ for $s \neq t$, else $O(1)$. For the second term (D.85), note that the first expectation is $O(|t-s|^{\max(-d_0,-\zeta)-1})$ for $s \neq t$, else $O(1)$, see (D.83). Furthermore, as shown below (D.78), the second expectation in (D.84) and (D.85) is $O((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1})$, and thus (D.84) and (D.85) are $O(1)$. Finally, the last term (D.86) is $O(1)$, and the proof is identical to (D.82). Thus, also the third term in (D.79) is $o_p(1)$. This shows that (D.58) is $o_p(1)$ and completes the proof. □

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

where $\tilde{\phi}_{\eta,j}$ is $O((1+\log j)^2 j^{-1})$, while $\tilde{\phi}_{\epsilon,j}$ is $O((1+\log t)^4 t^{\max(-d_0,-\zeta)-1})$ for $j < t$ and $O((1+\log j)^6 j^{\max(-d_0,-\zeta)-1})$ for $j \ge t$.

Proof of lemma D.8. Consider

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

$$+\sum_{j=1}^{t-1}\left[\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\tau_j(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\right]\tilde{\xi}_{t-j}(d_0) + \sum_{j=t}^{\infty}\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0) \tag{D.88}$$

$$+\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left[\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\xi_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0}\right] \tag{D.89}$$

$$+\sum_{j=1}^{t-1}\left[\tau_j(\theta_0) - \tau_j(\theta_0,t)\right]\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0} + \sum_{j=t}^{\infty}\tau_j(\theta_0)\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0}. \tag{D.90}$$

Since $\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0) = \sum_{k=t-j}^{\infty} \pi_k(d_0) c_{t-j-k}$, by (D.1), lemma D.4, and assumption 2, (D.87) is $\sum_{j=t}^{\infty} \epsilon_{t-j} \sum_{k=0}^{t-1} \frac{\partial \tau_k(\theta,t)}{\partial\theta}\big|_{\theta=\theta_0} \sum_{l=0}^{j-t} a_l(\varphi_0)\pi_{j-k-l}(d_0) = \sum_{j=t}^{\infty} O((1+\log j)^6 j^{\max(-d_0,-\zeta)-1})\epsilon_{t-j}$. By lemma D.5, (D.1), and assumption 3, the first term in (D.88) is

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

For the second term in (D.88), by lemma D.4, (D.1), and assumption 3

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

Note that (D.89), (D.90) are non-zero only for the derivative w.r.t. d. For (D.89), it holds that $\frac{\partial\pi_j(d-d_0)}{\partial d}\big|_{d=d_0} = -j^{-1}$, see Robinson (2006, pp. 135-136). Thus

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

by lemma D.2, Johansen and Nielsen (2010, lemma B.3), and assumption 3. For the first term in (D.90), by lemmas D.2, D.3, Johansen and Nielsen (2010, lemma B.3), and assumption 3

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

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

while for the second term in (D.90), by lemma D.2, Johansen and Nielsen (2010, lemma B.3), and assumption 3

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

Together, the results above prove lemma D.8. □

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

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

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

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

$$\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}} = O\left((1+\log k)^K k^{\max(-d,-\zeta)-1}\right), \tag{D.91}$$
$$\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}} - \frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}} = O\left((1+\log t)^K t^{\max(-d,-\zeta)-1}\right). \tag{D.92}$$

(D.91) can be seen directly from the proof of lemma D.4, as the second partial derivatives only add a log-factor to the convergence rates in lemma D.4. (D.92) can be shown analogously to the proof of lemma D.5, where again the second partial derivatives only add a log-factor to the convergence rates in lemma D.5. To simplify the notation, denote $h_{3,t_{(i,j)}} = \sum_{k=1}^{t-1}\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}\xi_{t-k}(d_0)$, $h_{4,t_{(i,j)}} = \sum_{k=1}^{t-1}\tau_k(\theta_0,t)\frac{\partial^2\xi_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, $h_{5,t_{(i,j)}} = \sum_{k=1}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta_{(i)}}\big|_{\theta=\theta_0}\frac{\partial\xi_{t-k}(d)}{\partial\theta_{(j)}}\big|_{\theta=\theta_0}$ for the partial derivatives of the truncated prediction error, and $\tilde{h}_{3,t_{(i,j)}} = \sum_{k=1}^{\infty}\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}\tilde{\xi}_{t-k}(d_0)$, $\tilde{h}_{4,t_{(i,j)}} = \sum_{k=1}^{\infty}\tau_k(\theta_0)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, $\tilde{h}_{5,t_{(i,j)}} = \sum_{k=1}^{\infty}\frac{\partial\tau_k(\theta)}{\partial\theta_{(i)}}\big|_{\theta=\theta_0}\frac{\partial\tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\big|_{\theta=\theta_0}$ for the untruncated prediction error. The term of interest then can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\frac{\partial^2\tilde{v}_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}-\frac{1}{n}\sum_{t=1}^{n}v_t(\theta_0)\frac{\partial^2 v_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\\
&=\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\left(\tilde{h}_{3,t_{(i,j)}}-h_{3,t_{(i,j)}}\right)+\frac{1}{n}\sum_{t=1}^{n}h_{3,t_{(i,j)}}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)\\
&+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\left(\tilde{h}_{4,t_{(i,j)}}-h_{4,t_{(i,j)}}\right)+\frac{1}{n}\sum_{t=1}^{n}h_{4,t_{(i,j)}}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)\\
&+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\left(\tilde{h}_{5,t_{(i,j)}}-h_{5,t_{(i,j)}}\right)+\frac{1}{n}\sum_{t=1}^{n}h_{5,t_{(i,j)}}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right)\\
&+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\left(\tilde{h}_{5,t_{(j,i)}}-h_{5,t_{(j,i)}}\right)+\frac{1}{n}\sum_{t=1}^{n}h_{5,t_{(j,i)}}\left(\tilde{v}_t(\theta_0)-v_t(\theta_0)\right),
\end{aligned}\tag{D.93}$$

and thus the different terms in (D.93) can be considered separately and will be shown to be $o_p(1)$. Note that $\tilde{v}_t(\theta_0)$ adapted to the filtration $\mathcal{F}_t$ is a MDS, while $\tilde{h}_{3,t_{(i,j)}}$, $\tilde{h}_{4,t_{(i,j)}}$, $\tilde{h}_{5,t_{(i,j)}}$ are $\mathcal{F}_{t-1}$-measurable. Starting with the first term in (D.93), by plugging in $\tilde{h}_{3,t_{(i,j)}}$, $h_{3,t_{(i,j)}}$

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{3,t_{(i,j)}}-h_{3,t_{(i,j)}})=\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\left(\tilde{\xi}_{t-k}(d_0)-\xi_{t-k}(d_0)\right)\\
&+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\left(\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}-\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\right)\tilde{\xi}_{t-k}(d_0)\\
&+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=t}^{\infty}\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\tilde{\xi}_{t-k}(d_0).
\end{aligned}\tag{D.94}$$

The latter two terms in (D.94) are MDS when adapted to $\mathcal{F}_t$, as $\tilde{v}_t(\theta_0)$ is a stationary MDS and as the other terms are $\mathcal{F}_{t-1}$-measurable. By (D.91) and (D.92), $\sum_{k=t}^{\infty}\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}\tilde{\xi}_{t-k}(d_0)$ as well as $\sum_{k=1}^{t-1}\left(\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}-\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}\right)\tilde{\xi}_{t-k}(d_0)$ are $o_p(1)$. Hence, the latter two terms in (D.94) are also $o_p(1)$. In contrast, the first term in (D.94) is not a MDS. However, by the same proof as for (D.59) (replacing the first partial derivative of $\tau_k(\theta,t)$ by the second partial derivative and noting that this only adds a log-factor to the convergence rate) it can also be shown to be $o_p(1)$. Thus, (D.94) is $o_p(1)$. For the third term in (D.93), by plugging in $\tilde{h}_{4,t_{(i,j)}}$, $h_{4,t_{(i,j)}}$

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{4,t_{(i,j)}}-h_{4,t_{(i,j)}})=\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\left(\tau_k(\theta_0)-\tau_k(\theta_0,t)\right)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0}\\
&+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\tau_k(\theta_0,t)\left(\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}-\frac{\partial^2\xi_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\right)\bigg|_{\theta=\theta_0}\\
&+\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=t}^{\infty}\tau_k(\theta_0)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0},
\end{aligned}\tag{D.95}$$

where the first and third term are MDS, as $\tilde{v}_t(\theta_0)$ is a MDS and the remaining term is $\mathcal{F}_{t-1}$-measurable. The third term is $o_p(1)$, as $\sum_{k=t}^{\infty} \tau_k(\theta_0) \frac{\partial^2 \tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}|_{\theta=\theta_0}$ is $o_p(1)$ by lemma D.2, and by Hualde and Robinson (2011, lemma 4). The first term is $o_p(1)$ since $(\tau_k(\theta_0) - \tau_k(\theta_0, t)) \frac{\partial^2 \tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}|_{\theta=\theta_0}$ is $o_p(1)$ by lemma D.3. The second term can be shown to be $o_p(1)$ analogously to (D.65) by replacing the first partial derivatives of $\tilde{\xi}_t(d)$ with the second partial derivatives, as this only adds a log-factor to the convergence rate, see Hualde and Robinson (2011, lemma 4). For the fifth term in (D.93), similar to (D.94) and (D.95)

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{5,t_{(i,j)}} - h_{5,t_{(i,j)}}) = \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{k=t}^{\infty} \frac{\partial \tau_k(\theta_0)}{\partial\theta_{(i)}}\bigg|_{\theta=\theta_0} \frac{\partial \tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{k=1}^{t-1} \frac{\partial \tau_k(\theta, t)}{\partial\theta_{(i)}}\bigg|_{\theta=\theta_0} \left(\frac{\partial \tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}} - \frac{\partial \xi_{t-k}(d)}{\partial\theta_{(j)}} \right)\bigg|_{\theta=\theta_0} \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{k=1}^{t-1} \left(\frac{\partial \tau_k(\theta)}{\partial\theta_{(i)}} - \frac{\partial \tau_k(\theta, t)}{\partial\theta_{(i)}} \right)\bigg|_{\theta=\theta_0} \frac{\partial \tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\bigg|_{\theta=\theta_0},
\end{aligned} \tag{D.96}$$

where the first and third term are MDS as before. The first term is $o_p(1)$ by lemma D.4, while the third term is $o_p(1)$ by lemma D.5. The second term can be shown to be $o_p(1)$ analogously to (D.65) using (D.68), as the partial derivatives of $\tau_k(\theta, t)$ only add a log-factor to the convergence rates, see lemma D.4. Thus, (D.96) is also $o_p(1)$. The second, fourth and sixth term in (D.93) can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) = \frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=0}^{t-1} (\tilde{\xi}_{t-k}(d_0) - \xi_{t-k}(d_0))\tau_k(\theta_0, t) \\
&+ \frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=1}^{t-1} (\tau_k(\theta_0) - \tau_k(\theta_0, t))\tilde{\xi}_{t-k}(d_0) + \frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=t}^{\infty} \tau_k(\theta_0)\tilde{\xi}_{t-k}(d_0),
\end{aligned} \tag{D.97}$$

with $l = 3, 4, 5$. For $l = 3$, (D.97) only differs from (D.72) as it contains the second partial derivatives of $\tau_k(\theta, t)$ in $h_{3,t_{(i,j)}}$. However, they only add a log-factor to the convergence rates of the first partial derivatives, see (D.91). For $l = 4$, (D.97) is almost identical to (D.79), where the only difference is that the former considers the second partial derivatives of $\xi_t(d)$ via $h_{4,t_{(i,j)}}$. Again, the second partial derivatives only add a log-factor to the convergence rates in (D.79) (Hualde and Robinson; 2011, lemma 4). For $l = 5$, (D.97) is again almost identical to (D.79) but now includes the first partial derivative of $\tau_k(\theta, t)$ via $h_{5,t_{(i,j)}}$. As for the other terms, by lemma D.4 the derivative again only adds a log-factor to the convergence rate of $\tau_k(\theta, t)$. Thus, it follows directly from (D.72) and (D.79), together with (D.91) and Hualde and Robinson (2011, lemma 4), that (D.97) is $o_p(1)$. The two remaining terms in (D.93) are $o_p(1)$ by (D.96) and (D.97), as i, j can be interchanged. This completes the proof.

□

References

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

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

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

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

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

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

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

Chang, Y., Miller, J. I. and Park, J. Y. (2009). Extracting a common stochastic trend: Theory with some applications, *Journal of Econometrics* **150**(2): 231–247.

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

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

Doda, B. (2014). Evidence on business cycles and co2 emissions, *Journal of Macroeconomics* **40**: 214–227.

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.

Granger, C. W. J. and Morris, M. J. (1976). Time series modelling and interpretation, *Journal of the Royal Statistical Society. Series A (General)* **139**(2): 246–257.

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

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

Haberl, H., Wiedenhofer, D., Virág, D., Kalt, G., Plank, B., Brockway, P., Fishman, T., Hausknost, D., Krausmann, F., Leon-Gruchalski, B., Mayer, A., Pichler, M., Schaffartzik, A., Sousa, T., Streeck, J. and Creutzig, F. (2020). A systematic review of the evidence on decoupling of GDP, resource use and GHG emissions, part II: synthesizing the insights, *Environmental Research Letters* **15**: 065003.

Harbaugh, W. T., Levinson, A. and Wilson, D. M. (2002). Reexamining the empirical evidence for an environmental Kuznets curve, *The Review of Economics and Statistics* **84**(3): 541–551.

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

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

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

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

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.

Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models, *The Review of Economic Studies* **61**(2): 247–264.

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

Hodrick, R. J. and Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation, *Journal of Money, Credit and Banking* **29**(1): 1–16.

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

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

Hualde, J. and Robinson, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models, *The Annals of Statistics* **39**(6): 3152–3181.

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

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

Morley, J. C., Nelson, C. R. and Zivot, E. (2003). Why are the Beveridge-Nelson and unobserved-components decompositions of GDP so different?, *The Review of Economics and Statistics* **85**(2): 235–243.

Newey, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity, *Econometrica* **59**(4): 1161–1167.

Nielsen, M. Ø. (2015). Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time-series models, *Journal of Time Series Analysis* **36**(2): 154–188.

Oh, K. H., Zivot, E. and Creal, D. (2008). The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics, *Journal of Econometrics* **146**(2): 207–219.

Oh, K. W. and Zivot, E. (2006). The Clark model with correlated components, *Technical report*, University of Washington.

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

Ravn, M. O. and Uhlig, H. (2002). On adjusting the Hodrick-Prescott filter for the frequency of observations, *The Review of Economics and Statistics* **84**(2): 371–376.

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

Ritchie, H., Roser, M. and Rosado, P. (2020). CO2 and greenhouse gas emissions, *Our World in Data* . https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions.

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.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics* **63**(1): 289–306.

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

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.

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, Cambrdige.