Fractional unobserved components and factor models: econometric theory and applications

Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft

eingereicht an der

Fakultät für Wirtschaftswissenschaften der Universität Regensburg

vorgelegt von Tobias Hartl

Berichterstatter:

Prof. Dr. Rolf Tschernig, Universität Regensburg
Prof. Dr. Uwe Hassler, Goethe-Universität Frankfurt

Tag der Disputation: 04. Juli 2023

Acknowledgments

First of all, I would like to thank Prof. Rolf Tschernig for giving me the opportunity to work on my dissertation under his supervision. Over the past five years, I have had the opportunity to learn, research, teach, and grow as a scientist in an inspiring, supportive, and friendly environment. Through numerous discussions, collaborative research, and the much advice I have received, I have benefited from my work at the Chair of Econometrics as a researcher, but also as a person. In addition, I would like to thank my co-supervisor, Prof. Uwe Hassler, for helping me develop a proof strategy for the asymptotic theory of fractional unobserved components models, as well as for giving me a lot of advice on how to get along in academia.

Second, I would like to thank Roland Jucknewitz for sparking my enthusiasm for long memory, as well as Prof. Enzo Weber for the many fruitful discussions about unobserved components models and their applications.

Third, I would like to thank the many colleagues at the University of Regensburg and the Institute for Employment Research Nuremberg. Not only was I able to learn about new facets of economics and econometrics, but also benefited greatly from the support, motivation and friendship I was able to experience. I appreciate having worked with Dominik Ammon, Timon Hellwagner, Nicole Lechermann, and Christoph Rust.

Fourth, I would like to thank the German Research Foundation (DFG) for funding our project "Fractional Unobserved Components and Factor Models for Macroeconomic Analysis and Forecasting", which led to this dissertation.

Finally, I would like to thank my family, especially Alina. Without their unlimited support and understanding, none of this would have been possible.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis develops generalizations of unobserved components and factor models to account for long memory. Long memory describes a strongly persistent and often non-stationary behavior, as found for numerous time series in macroeconomics, finance, climate research, and beyond (for an introduction, see Hassler; 2019). Traditional unobserved components and factor models are limited to integer-integrated processes, and often make strong assumptions about the integration orders of the data under study. Generalizing these models to account for fractional integration seamlessly links integer-integrated specifications, allows for intermediate solutions between integer integration orders, and provides a data-driven solution to the specification of the long-run dynamics.

The introduction first outlines the general frameworks of unobserved components and factor models, and introduces the concept of long memory, along with a brief motivation for incorporating long memory into unobserved components and factor models. A second subsection summarizes the three essays contained in this dissertation and details the contribution to the literature.

1.1 Methodological framework

To introduce unobserved components models, consider the observable, univariate time series y_t that is measured regularly at $t = 1, ..., n$. Unobserved components models build on the framework as introduced by Harvey (1985) and Clark (1987), and assume that y_t is generated as the sum of two latent components, a trend τ_t and a cycle c_t, such that

$$y_t = \tau_t + c_t, \qquad t = 1, ..., n.$$

Both τ_t and c_t are unobserved, and are distinguished by their spectra: The cyclical component c_t is assumed to be a stationary process with mean zero and is to capture the short-run dynamics of y_t. Typically, it is represented as a stationary autoregressive process of order p (see, for instance, Clark; 1987; Morley et al.; 2003; Oh et al.; 2008)

$$a(L)c_t = \epsilon_t, \tag{1.1}$$

where $a(L) = 1 - a_1L - ... - a_pL^p$, L is the lag operator, and ϵ_t is assumed to be at least white noise, often Gaussian. However, some alternatives to (1.1) have been considered in the literature, such as ARMA models and mixtures of sine and cosine waves, see Harvey (1989, ch. 2).

For the trend component, the literature typically assumes that τ_t is generated by a linear deterministic trend $\mu_t = \mu_0 + \mu_1 t$, plus a non-stationary stochastic trend x_t that is integrated of order d, such that

$$\tau_t = \mu_t + x_t, \qquad \Delta^d x_t = \eta_t, \tag{1.2}$$

where $\Delta^d = (1-L)^d$, $d \in \mathbb{N}$ is an integer that defines the integration order of x_t and is assumed to be known, and η_t is again at least white noise, often Gaussian.

For log US real GDP, which is arguably the most important application of unobserved components models, the field is divided into two main groups with respect to the specification of d: On the one hand, models in the spirit of Beveridge and Nelson (1981) and Harvey (1985) assume $d = 1$, i.e. the stochastic long-run dynamics are modeled by a random walk. On the other hand, models based on Clark (1987) and Hodrick and Prescott (1997) assume $d = 2$, i.e. x_t becomes a quadratic stochastic trend. Although the latter specification is clearly at odds with the long-run properties of log US GDP, which tend to support setting $d = 1$, choosing $d = 2$ forces the estimated trend component to be very smooth, leaving rich dynamics to be captured by the cycle. Conversely, setting $d = 1$ yields an estimated trend component that is erratic, along with a noisy cyclical component, once η_t and ϵ_t are allowed to be correlated, see Morley et al. (2003).

A major limitation of unobserved components models is the dichotomy between specifying the stochastic trend as a random walk and specifying it as a quadratic stochastic trend. Both are rather extreme cases of long-run dynamics, where past shocks enter the trend either as an unweighted sum, or with quadratically increasing weights. While other parameters are defined as continuous and are subject to

estimation, the unobserved components literature treats d as discrete and known, thus allowing for no intermediate solutions between specifying the trend as a random walk and specifying it as a quadratic stochastic trend. As the coexistence of $I(1)$ and $I(2)$ trend specifications for log US real GDP illustrates, there is often no consensus on the appropriate choice of d, and estimates of trend and cycle differ massively between the two specifications.

This thesis addresses this limitation by generalizing unobserved components models to account for fractional integration. The concept of fractional integration allows for integration orders $d \in \mathbb{R}$, i.e. also for intermediate solutions between the integer integration orders. For $d \in \mathbb{R}$, the differencing operator in (1.2) exhibits a polynomial expansion in L of order infinite

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

where the weights $\pi_j(d)$ are determined recursively. The concept of fractional integration encompasses integer integration orders as special cases, seamlessly links the random walk trend specification with the quadratic stochastic trend specification, and adds flexibility to the weighting of past long-run innovations. Since d is defined on the real line, it can be treated as an unknown parameter and can be estimated jointly with the other model parameters. Thus, the fractional generalization of unobserved components models provides a data-driven solution to the specification of the long-run dynamics.

A second methodological focus of this thesis is on factor models, in which latent variables also play a key role. In contrast to univariate unobserved components models, factor models are typically used to model high-dimensional data, both in the time domain and in the cross section. For an N-dimensional vector of observable variables $Y_t = (y_{1,t}, ..., y_{N,t})'$[1], which are measured regularly at $t = 1, ..., T$, the basic structure of a factor model is given by

$$Y_t = f(\chi_t) + u_t, \qquad t = 1, ..., T,$$

where u_t is an N-vector holding the idiosyncratic component, and $f(\chi_t)$ is the common component that is driven by the r common factors $\chi_t = (\chi_{1,t}, ..., \chi_{r,t})'$. While traditional factor models treat all observable variables as $I(0)$ (see Bai and Ng; 2008,

[1] In chapter 4 of this thesis, the observable variables are denoted as y_t. However, in the introduction I use a capital Y_t to avoid confusion with the notation of the unobserved components model.

for an overview), there are several generalizations to $I(1)$ processes, see Bai (2004), Bai and Ng (2004), Eickmeier (2009), Chang et al. (2009), Banerjee et al. (2014), and Barigozzi et al. (2021) among others. However, such models again rule out intermediate solutions between the rather extreme cases of short memory and perfect memory. For fractionally integrated data, the literature has so far mostly considered semiparametric factor models, where the factors are estimated via principal components (see Morana; 2004; Luciani and Veredas; 2015; Cheung; 2022; Ergemen; 2023). Recently, Hartl and Jucknewitz (2021) developed a parametric factor model that allows for stationary and non-stationary factors, fractional integration, and heterogeneous integration orders among the factors and data. Building on their work, this thesis investigates several fractionally integrated factor model formulations and examines their forecast performance for the high-dimensional macroeconomic data set of McCracken and Ng (2016).

1.2 Overview and contribution

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

The first paper addresses the aforementioned limitation of unobserved components models that arises by the dichotomy between specifying the stochastic trend as a random walk and specifying it as a quadratic stochastic trend. I introduce a novel unobserved components model that generalizes the stochastic trend component x_t in (1.2) to a fractionally integrated process of order $d \in \mathbb{R}_+$, denoted as $x_t \sim I(d)$. The model encompasses the two main specifications in the literature, which assume either that the trend component is $I(1)$ (e.g. Beveridge and Nelson; 1981; Harvey; 1985; Morley et al.; 2003), or that it is $I(2)$ (e.g. Clark; 1987; Hodrick and Prescott; 1997; Oh et al.; 2008). Since d can take any value on the positive real line, the model allows for intermediate solutions between the integer-integrated specifications and thus for even more general patterns of persistence. In addition to the fractional trend, the model allows for flexible parameterizations of the cyclical component, including stationary ARMA specifications.

Estimates for trend and cycle are obtained from the Kalman filter and smoother, as is common in the unobserved components literature. However, the fractional generalization of the trend comes at the cost of making the state vector high-dimensional, in contrast to traditional, integer-integrated models. Consequently, running the

Kalman recursions for the fractional unobserved components model is computationally expensive. To speed up the computation, I derive an analytical solution to the optimization problem of the Kalman filter, which allows the filtered and smoothed trend and cycle to be computed directly, thus bypassing the Kalman recursions. This has the additional advantage that I obtain an analytical expression for the prediction error that depends only on the model parameters and the observable data, and not on recursive solutions for the filtered trend and cycle. Since the prediction error enters the objective function for parameter estimation, this simplifies the derivation of the asymptotic estimation theory.

Parameters can be estimated using either the conditional sum-of-squares estimator or the quasi-maximum likelihood estimator. Both estimators minimize the sum of squared prediction errors, and are asymptotically equivalent. A key part of this paper is to derive the asymptotic theory for parameter estimation, where I show that both estimators are consistent and asymptotically normally distributed. The derivation of the asymptotic theory is complicated by the non-ergodicity of the prediction errors and the non-uniform convergence of the objective function. The finite sample properties of the estimators are evaluated in a Monte Carlo study, which supports the results on consistency for both estimators.

The fractional unobserved components model is then applied to model log annual US carbon emissions. Estimation results indicate an integration order of 1.75, suggesting that integer-integrated models are misspecified for log annual US carbon emissions. The estimated trend is smooth and has two major turning points, the 1979 energy crisis and the Great Recession. Since the former, per capita emissions decline, while total annual US emissions decline since the Great Recession. This supports the environmental Kuznets curve hypothesis, i.e. an inverted U-shaped relationship between economic development and carbon emissions. The estimated cyclical component shows a persistent behavior and appears to be closely coupled to the business cycle.

1.2.2 Solving the unobserved components puzzle: a fractional approach to measuring the business cycle

The second paper is joint work with Rolf Tschernig (University of Regensburg) and Enzo Weber (University of Regensburg and Institute for Employment Research (IAB) Nuremberg). It addresses the puzzling estimates for the business cycle obtained from traditional unobserved components models with integer-integrated trend components: Once correlation between trend- and cycle innovations is allowed for,

the results from unobserved components models with an $I(1)$ stochastic trend are virtually identical to the Beveridge-Nelson decomposition of log US real GDP. As shown by Morley et al. (2003), the estimated correlation between long- and short-run innovations is close to -1, the estimated trend is volatile, and the estimated cycle is noisy, lacking a clear cyclical pattern and missing the NBER chronology. At the same time, unobserved components models that either force the correlation between long- and short-run innovations to be zero (e.g. the model of Harvey and Jäger; 1993), or restrict the variance-ratio between long- and short-run innovations to be small (e.g. the model of Kamber et al.; 2018), yield economically plausible decompositions into trend and cycle, but exhibit a smaller likelihood.

We provide evidence that the puzzling results are an artifact generated by the presence of a smooth fractionally integrated trend in log US real GDP with an integration order greater than one but less than two. The correlated unobserved components model of Morley et al. (2003) then misspecifies the integration order, which upward-biases the variance estimate for the long-run innovations, leading to an estimate for the trend component that is erratic. By allowing for correlation between long- and short-run innovations, the cyclical component can adjust for the erratic behavior of the trend, which also makes the cycle noisy. Conversely, unobserved components models with a stochastic trend that is $I(2)$ (e.g. the models of Hodrick and Prescott; 1997; Oh et al.; 2008) will estimate a trend that is too smooth, and attribute too much variation in GDP to the cycle.

We revisit the problem of decomposing log GDP into trend and cycle using the fractional unobserved components model as introduced in chapter 2 of this thesis. The model allows for intermediate solutions between an $I(1)$ and an $I(2)$ specification of the trend component. In addition, it allows the integration order to be estimated jointly with the other model parameters, thus providing inference on the appropriate specification of the trend in unobserved components models for log GDP. We consider several specifications for the cyclical component, including ARMA models as well as autoregressive models that replace the lag operator with the fractional lag operator as suggested by Granger (1986).

For the preferred specification in terms of the Bayesian information criterion, we reject the hypotheses that log GDP is $I(1)$ and $I(2)$, respectively, and estimate an integration order of about 1.3. The business cycle estimate from the fractional unobserved components model rises gradually in periods of economic upswing, falls sharply during the NBER recession periods, and exhibits the same turning points as the theory-based output gap measure of the US Congressional Budget Office. In

addition to the latter, the fractional model reveals an overheating economy in the run-up to the Great Recession, as also found by Barigozzi and Luciani (2021).

1.2.3 Macroeconomic forecasting with fractional factor models

The third paper combines high-dimensional factor models with fractional integration methods. Factor models are popular in fields where rich data sets with strong cross sectional dependencies are evident, as they efficiently bundle common dynamics into a typically small number of common factors. In contrast to the cross section, strong dependencies in the time domain have received comparably little attention: The bulk of the literature on factor models assumes stationary factors and data, and typically pre-processes the latter by taking first or second differences to ensure stationarity. While there are some generalizations to non-stationary factors and data, these are typically semiparametric models (see the discussion in subsection 1.1).

The core of the paper is to infer whether the combination of parametric factor models and fractional integration methods improves the forecast performance for macroeconomic data. I consider three different factor models that allow for possibly non-stationary, fractionally integrated data and factors with possibly heterogeneous integration orders. The first model generalizes the non-stationary dynamic factor model of Barigozzi et al. (2021) with $I(1)$ factors to fractionally integrated factors. As a second model, I examine the fractional components model of Hartl and Jucknewitz (2021), which distinguishes between long memory factors that are purely fractionally integerated noise, and short memory factors that are modeled as autoregressive processes. The third model is a dynamic factor model that is set up for the observable variables in fractional differences. For all models, I derive the state space representation, so that factors, loadings, and integration orders can be jointly estimated by a combination of the Kalman recursions and maximum likelihood. To keep the dimension of the state vector manageable, I use approximations to the fractional differencing operator as proposed by Hartl and Jucknewitz (2022) among others.

In a pseudo out-of-sample forecast experiment for the US macroeconomic data set of McCracken and Ng (2016), I study the forecast performance of the different fractional factor models. I consider a total of 112 macroeconomic variables in monthly frequency and perform forecasts from one to twelve steps ahead. In order to relate the forecast performance to the existing literature, I also include several benchmark models as competitors, such as the approximate dynamic factor model of Stock and Watson (2002), and the factor-augmented error-correction model of Banerjee and Marcellino (2009). For several variables, I find that fractional factor

models can significantly improve the forecasts in terms of the mean squared prediction error (MPSE). In particular the fractional components model of Hartl and Jucknewitz (2021) often yields a smaller MSPE than its competitors, with reductions of more than 50% possible.

Chapter 2

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

2.1 Introduction

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

The literature on trend-cycle decompositions has been shaped by the seminal works of Beveridge and Nelson (1981), Harvey (1985), Clark (1987), and Hodrick and Prescott (1997). Since then, a variety of unobserved components (UC) models have been proposed, and often the integration order of the trend was subject to

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

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

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

While the UC literature has mostly considered integer-integrated specifications, there are some generalizations to non-integer integration orders: 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 by truncating either the autoregressive or the moving average representation of the fractional differencing polynomial. Their models have been found valuable for realized volatility modeling (see Ray and Tsay; 2000; 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.

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

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

[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 carbon emissions illustrates the benefits from the fractional UC model: Log annual US carbon emissions are estimated to be integrated of order around 1.75, which is clearly at odds with integer-integrated models. The resulting trend-cycle decomposition finds evidence that the trend component exhibits an inverted U-shape, supporting the existence of an environmental Kuznets curve as well as the often hypothesized decoupling of economic activity and carbon emissions in terms of the trend. In contrast, as a glimpse on figure 2.3 reveals, cyclical emissions appear to remain coupled to the business cycle, as they exhibit rich pro-cyclical dynamics. Integer-integrated benchmarks fail to capture these stylized facts.

The rest of the paper is organized as follows: Section 2.2 introduces the fractional UC model and discusses the underlying assumptions. Section 2.3 discusses trend and cycle estimation, while section 2.4 details parameter estimation using the CSS estimator. Generalizations of the fractional UC model are discussed in section 2.5. Section 2.6 examines the finite sample properties of the proposed methods in a Monte Carlo study, while section 2.7 applies the fractional UC model to carbon emissions. Section 2.8 concludes. The proofs for consistency and asymptotic normality are contained in the appendix.

2.2 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

$$y_t = x_t + c_t. \tag{2.1}$$

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

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

$$\Delta_+^d x_t = \eta_t. \tag{2.2}$$

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

$$\Delta^d = (1-L)^d = \sum_{j=0}^{\infty} \pi_j(d) L^j, \qquad \pi_j(d) = \begin{cases} \frac{j-d-1}{j}\pi_{j-1}(d) & j = 1, 2, ..., \\ 1 & j = 0, \end{cases} \tag{2.3}$$

where the weights $\pi_j(d)$ are determined recursively. The motivation behind (2.2) and (2.3) is that the higher d, the greater the effect of a past shock η_{t-j} on x_t, and the more differencing is required to eliminate the persistent impact of the past shock via (2.2). 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 (2.2) denotes the truncation of an operator at $t \leq 0$, $\Delta_+^d x_t = \Delta^d x_t \mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} \pi_j(d) x_{t-j}$, where $\mathbb{1}(t \geq 1)$ is the indicator function that takes the value one for positive subscripts of x_{t-j}, otherwise zero. The truncated fractional difference operator reflects the type II definition of fractionally integrated processes (Marinucci and Robinson; 1999) and is required to treat the asymptotically stationary case alongside the non-stationary case.

Equation (2.2) encompasses several trend specifications in the literature: For $d = 1$, it nests the random walk trend model as considered by Harvey (1985), Balke and Wohar (2002), and Morley et al. (2003) among others. For $d = 2$, one has the double-drift model of Clark (1987) and Oh et al. (2008), but also the filter of Hodrick and Prescott (1997, HP filter in what follows) as will become clear. For $d \in \mathbb{N}$, the model of Burman and Shumway (2009) is obtained. Allowing for $d \in \mathbb{R}_+$ seamlessly links these integer-integrated models and allows for far more general dynamics of the trend: For $0 < d < 1/2$, it covers stationary and strongly persistent processes as considered by Ray and Tsay (2000), Chen and Hurvich (2006), and Varneskov and Perron (2018) for realized volatility modeling. For $1/2 < d < 1$, it allows for non-stationary but mean-reverting processes, while $d \geq 1$ yields non-stationary non-mean-reverting processes that are indispensable for trend-cycle decompositions of macroeconomic variables among others. Since d enters the model as an unknown parameter to be estimated, the model allows for a data-driven choice of d and provides statistical inference on the appropriate specification of UC models.

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

$$c_t = a(L, \varphi)\epsilon_t = \sum_{j=0}^{\infty} a_j(\varphi)\epsilon_{t-j}. \tag{2.4}$$

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

In what follows, the model (2.1), (2.2), and (2.4) is analyzed under the following assumptions:

Assumption 2.1 (Errors)**.** *The errors ϵ_t, η_t are stationary and ergodic with finite moments up to order four and absolutely summable autocovariance function. For the joint σ-algebra $\mathcal{F}_t = \sigma((\eta_s, \epsilon_s), s \leq t)$, it holds that $\mathrm{E}(\epsilon_t|\mathcal{F}_{t-1}) = 0$, $\mathrm{E}(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\epsilon^2$, and $\mathrm{E}(\eta_t|\mathcal{F}_{t-1}) = 0$, $\mathrm{E}(\eta_t^2|\mathcal{F}_{t-1}) = \sigma_\eta^2$. Furthermore, conditional on $\mathcal{F}_{t-1}$, the third and fourth moments of ϵ_t, η_t are finite and equal their unconditional moments. Finally, ϵ_t and η_t are independent.*

Assumption 2.2 (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 2.1 allows for conditionally homoscedastic martingale difference sequences (MDS) η_t and ϵ_t. This is somewhat more general than the UC literature, which typically assumes Gaussian white noise disturbances (e.g. in Morley et al.; 2003). The generalization is of great practical importance given the applications of UC models in macroeconomics and finance.

Assumption 2.2 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 2.3 (Stability of $a(L, \varphi)$)**.** *For all $\varphi \in \Phi$ and all z in the complex unit disc $\{z \in \mathbb{C} : |z| \leq 1\}$ it holds that*

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

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

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

Under assumption 2.3, $a(L,\varphi)^{-1} = b(L,\varphi) = \sum_{j=0}^{\infty} b_j(\varphi)L^j$ exists, is well defined, and the sum $\sum_{j=0}^{\infty} |b_j(\varphi)|$ is bounded and bounded away from zero. By the Lipschitz condition it holds that

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

The rate for $a_j(\varphi)$ follows directly from assumption 2.3(ii), while that for $b_j(\varphi)$ follows from Zygmund (2002, pp. 46 and 71). The convergence rate for the partial derivative $\dot{a}_j(\varphi)$ is a direct consequence of compactness of Φ and continuity of $\partial a_j(\varphi)/\partial\varphi'$. Assumption 2.3 imposes some smoothness on the linear coefficients in $a(L,\varphi)$, and thus also on $b(L,\varphi)$. It is satisfied by any stationary and invertible ARMA process. For ARFIMA models, the asymptotic estimation theory is well established under assumptions similar to 2.1, 2.2, and 2.3, see Hualde and Robinson (2011) and Nielsen (2015).

2.3 Filtering and smoothing

The system introduced in (2.1), (2.2), and (2.4) forms a state space model, where (2.1) is the measurement equation and (2.2), (2.4) are the state equations for trend and cycle.[3] This opens the way to the Kalman filter, a powerful set of algorithms for filtering, predicting, and smoothing the latent components x_t and c_t, but also for parameter estimation. In this section, I derive an analytical solution to the optimization problem of the Kalman filter and smoother. As will become clear at the end of this section, the analytical solution has two decisive advantages over the usual recursive algorithm 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 $b_+(L,\varphi)c_t = b(L,\varphi)c_t\mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} b_j(\varphi)c_{t-j}$. Furthermore, collect $x_{t:1} = (x_t, ..., x_1)'$ and $c_{t:1} = (c_t, ..., c_1)'$, and define the $t \times t$ differencing matrix $S_{d,t}$

[3] Section 2.5 outlines the state space representation and illustrates the dimensions of the system matrices. For further details on state space models and the Kalman filter, see Harvey (1989, ch. 3).

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

$$\begin{aligned} S_{d,t} &= \begin{bmatrix} \pi_0(d) & \pi_1(d) & \cdots & \pi_{t-1}(d) \\ 0 & \pi_0(d) & \cdots & \pi_{t-2}(d) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_0(d) \end{bmatrix}, \\ B_{\varphi,t} &= \begin{bmatrix} b_0(\varphi) & b_1(\varphi) & \cdots & b_{t-1}(\varphi) \\ 0 & b_0(\varphi) & \cdots & b_{t-2}(\varphi) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_0(\varphi) \end{bmatrix}, \end{aligned} \tag{2.5}$$

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

$$\hat{x}_{t:1}(y_{t:1}, \psi) = \left(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t}\right)^{-1} B'_{\varphi,t}B_{\varphi,t}y_{t:1} = \hat{x}_{t:1}(y_{t:1}, \theta), \tag{2.6}$$

$$\hat{c}_{t:1}(y_{t:1}, \psi) = \nu\left(B'_{\varphi,t}B_{\varphi,t} + \nu S'_{d,t}S_{d,t}\right)^{-1} S'_{d,t}S_{d,t}y_{t:1} = \hat{c}_{t:1}(y_{t:1}, \theta), \tag{2.7}$$

where the fraction $\nu = \sigma_\epsilon^2/\sigma_\eta^2$ controls for the variance ratio of the innovations, $\hat{x}_{t:1}(y_{t:1}, \psi) = (\hat{x}_t(y_{t:1}, \psi), ..., \hat{x}_1(y_{t:1}, \psi))'$, $\hat{c}_{t:1}(y_{t:1}, \psi) = (\hat{c}_t(y_{t:1}, \psi), ..., \hat{c}_1(y_{t:1}, \psi))'$ collect the filtered trend and cycle, and $\theta = (d, \nu, \varphi')'$. (2.6) and (2.7) 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 (2.6) and (2.7) into the state equations (2.2) and (2.4)

$$\hat{x}_{t+1}(y_{t:1}, \theta) = -\begin{pmatrix} \pi_1(d) & \cdots & \pi_t(d) \end{pmatrix} \hat{x}_{t:1}(y_{t:1}, \theta), \tag{2.8}$$

$$\hat{c}_{t+1}(y_{t:1}, \theta) = -\begin{pmatrix} b_1(\varphi) & \cdots & b_t(\varphi) \end{pmatrix} \hat{c}_{t:1}(y_{t:1}, \theta). \tag{2.9}$$

Together, the updating equations (2.6), (2.7) and the prediction equations (2.8), (2.9) 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 (2.6), (2.7) by setting $t = n$. They are identical to those obtained by the Kalman smoother.

To prove (2.6) and (2.7), I first consider the objective function of the Kalman filter, which follows from maximizing the quasi-log likelihood of (2.1), (2.2), and (2.4) 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

$$\hat{x}_{t:1}(y_{t:1}, \psi) = \arg\min_{x_{t:1}} \frac{1}{t} \sum_{j=1}^{t} \left\{ \frac{1}{\sigma_\epsilon^2} \left[b_+(L, \varphi)(y_j - x_j) \right]^2 + \frac{1}{\sigma_\eta^2} \left(\Delta_+^d x_j \right)^2 \right\}, \tag{2.10}$$

$$\hat{c}_{t:1}(y_{t:1}, \psi) = \arg\min_{c_{t:1}} \frac{1}{t} \sum_{j=1}^{t} \left\{ \frac{1}{\sigma_\eta^2} \left[\Delta_+^d (y_j - c_j) \right]^2 + \frac{1}{\sigma_\epsilon^2} \left(b_+(L, \varphi) c_j \right)^2 \right\}. \tag{2.11}$$

Here, the first residual in (2.10) stems from plugging (2.4) into the measurement equation and solving for ϵ_j, while the second is from (2.2). Analogously, the first term in (2.11) follows from inserting (2.2) into (2.1) and solving for η_j, while the second follows from solving (2.4) for ϵ_j. Constant terms are omitted. As x_t and c_t are estimated based on all observations until period t, it holds that $\hat{x}_{t:1}(y_{t:1}, \psi) = y_{t:1} - \hat{c}_{t:1}(y_{t:1}, \psi)$. If η_t and ϵ_t are assumed to be Gaussian, the optimization problems in (2.10) and (2.11) 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 (2.10) and (2.11) remain valid. The filtered $\hat{x}_{t:1}(y_{t:1}, \psi)$, $\hat{c}_{t:1}(y_{t:1}, \psi)$ are the projections of $x_{t:1}$ and $c_{t:1}$ on the span of $y_{t:1}$, and are the minimum variance linear unbiased estimators for $x_{t:1}$ and $c_{t:1}$ given the observable information $y_1, ..., y_t$ (Durbin and Koopman; 2012, lemma 2). For $t = n$, $d = 2$, $b(L, \varphi) = 1$, $\nu = \sigma_\epsilon^2 / \sigma_\eta^2$, (2.10) becomes the HP filter with ν being the tuning parameter. Thus, the HP filter constitutes a special case of the fractional UC model.

From (2.5), a matrix representation of (2.10) and (2.11) follows

$$\hat{x}_{t:1}(y_{t:1}, \psi) = \arg\min_{x_{t:1}} \frac{1}{t} \left\{ \frac{1}{\sigma_\epsilon^2} \left\| B_{\varphi,t}(y_{t:1} - x_{t:1}) \right\|^2 + \frac{1}{\sigma_\eta^2} x_{t:1}' S_{d,t}' S_{d,t} x_{t:1} \right\}, \tag{2.12}$$

$$\hat{c}_{t:1}(y_{t:1}, \psi) = \arg\min_{c_{t:1}} \frac{1}{t} \left\{ \frac{1}{\sigma_\eta^2} \left\| S_{d,t}(y_{t:1} - c_{t:1}) \right\|^2 + \frac{1}{\sigma_\epsilon^2} c_{t:1}' B_{\varphi,t}' B_{\varphi,t} c_{t:1} \right\}, \tag{2.13}$$

where $\|\cdot\|$ denotes the Euclidean norm. Calculating the derivative of (2.12) and (2.13) and solving for x_t and c_t yields (2.6) and (2.7). Note that (2.6) and (2.7) do not depend on the exact magnitudes of σ_η^2 and σ_ϵ^2, but only on their ratio ν, $0 < \nu < \infty$. Thus, for any positive constant $K > 0$, the parameter vector $\psi^* = (d, K\sigma_\eta^2, K\sigma_\epsilon^2, \varphi')'$ yields the same estimates $\hat{x}_{t:1}(y_{t:1}, \psi^*)$, $\hat{c}_{t:1}(y_{t:1}, \psi^*)$ as (2.6) and (2.7). By defining the parameter vector $\theta = (d, \nu, \varphi')'$, one has $\hat{x}_{t:1}(y_{t:1}, \psi) = \hat{x}_{t:1}(y_{t:1}, \theta)$ and $\hat{c}_{t:1}(y_{t:1}, \psi) = \hat{c}_{t:1}(y_{t:1}, \theta)$. This will be helpful for parameter estimation in section 2.4, since the conditional sum-of-squares estimator is not identified for ψ. Also, using

θ reduces the dimension of the parameter vector, which speeds up the optimization. However, ψ can also be estimated directly by maximum likelihood as will be shown in subsection 2.5.3.

From the filtered latent components in (2.6) and (2.7), the one-step ahead predictions for x_{t+1} and c_{t+1} follow immediately by plugging (2.6) and (2.7) into the state equations (2.2) and (2.4). This yields (2.8) and (2.9). While (2.6), (2.7), (2.8), and (2.9) are required for parameter estimation, as discussed in the next section, estimates for x_t and c_t typically reported are the projections of x_t and c_t on the span of $y_1, ..., y_n$, i.e. on the full sample information. They follow immediately from (2.6) and (2.7) by setting $t = n$, and are identical to the Kalman smoother.

Note that the filtered, predicted and smoothed x_t and c_t can be computed either via the analytical solution above or recursively by executing the Kalman recursions (see Harvey; 1989, ch. 3, for the latter). Both approaches yield identical results and only differ in the way they are computed. However, the analytical solution has two decisive advantages over the traditional recursions: (i) It is computationally superior for fractional trends. As the state vector of the fractional trend in (2.2) is of dimension $n - 1$, the dimension of the state vector for both trend and cycle is of dimension $m \geq n - 1$. Thus, each recursion of the Kalman filter involves multiple multiplications of $(m \times m)$-dimensional covariance and system matrices, and each multiplication requires $2m^3 - m^2$ flops (Hunger; 2007). The analytical solution also requires the expensive computation of an $(n \times n)$ inverse, however the underlying matrix is symmetric, positive definite, and thus the Cholesky decomposition can be used to reduce the complexity to $n^3 + n^2 + n$ flops per iteration (Hunger; 2007). Since $m \geq n - 1$, the analytical solution speeds up the computation considerably. This allows to run the Monte Carlo studies in section 2.6, which would otherwise be computationally infeasible. (ii) The solution allows to derive an objective function for parameter estimation that does not depend on the Kalman recursions and is thus easier to analyze. As usual, the objective function for parameter estimation is set up based on the one-step ahead prediction error, that is obtained by plugging (2.8) and (2.9) into the measurement equation (2.1). Since (2.8) and (2.9) depend only on the observable $y_1, ..., y_t$ as well as on the model parameters, the objective function does not depend on a recursive solution for the filtered trend and cycle. This greatly simplifies the asymptotic theory for parameter estimation in section 2.4, since the convergence rates of all coefficients are either known, or can be derived immediately.

2.4 Parameter estimation

To estimate $\theta_0 = (d_0, \nu_0, \varphi_0')'$, denote $\Theta = D \times \Sigma_\nu \times \Phi$ the respective parameter space, where $\Sigma_\nu = \{\nu \in \mathbb{R} | 0 < \nu_{min} \leq \nu \leq \nu_{max} < \infty\}$, and D, Φ as defined in assumption 2.2. By assumption 2.2, Θ is convex and compact. As usual in the state space literature, I set up the objective function for parameter estimation based on the one-step ahead forecast error for y_{t+1}, denoted as $v_{t+1}(\theta) = y_{t+1} - \hat{x}_{t+1}(y_{t:1}, \theta) - \hat{c}_{t+1}(y_{t:1}, \theta)$. By plugging in (2.8) and (2.9), $v_{t+1}(\theta)$ can be represented as

$$\begin{aligned} v_{t+1}(\theta) = \Delta_+^d y_{t+1} + \nu \left(b_1(\varphi) - \pi_1(d) \ \cdots \ b_t(\varphi) - \pi_t(d)\right) \\ \times (B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t})^{-1} S_{d,t}' S_{d,t} y_{t:1}. \end{aligned} \tag{2.14}$$

$v_{t+1}(\theta)$ 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 (2.14) that fully depends on θ. Let $\xi_{t+1}(d) = \Delta_+^d y_{t+1} = \Delta_+^{d-d_0} \eta_{t+1} + \Delta_+^d c_{t+1}$ and $\xi_{t:1}(d) = (\xi_t(d) \cdots \xi_1(d))' = S_{d,t} y_{t:1}$ denote the fractionally differenced y_{t+1} and $y_{t:1}$ respectively. Then, (2.14) can be written as

$$v_{t+1}(\theta) = \xi_{t+1}(d) + \sum_{j=1}^{t} \tau_j(\theta, t) \xi_{t+1-j}(d) = \sum_{j=0}^{t} \tau_j(\theta, t) \xi_{t+1-j}(d), \tag{2.15}$$

where $\tau_0(\theta, t) = 1$, and $(\tau_1(\theta, t) \cdots \tau_t(\theta, t)) = \nu(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))$ $(B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t})^{-1} S_{d,t}'$ collects the t coefficients belonging to $\xi_t(d), ..., \xi_1(d)$ in (2.15). The conditional sum-of-squares (CSS) estimator for θ_0 follows from minimizing the sum of squared forecast errors

$$\hat{\theta} = \arg\min_{\theta \in \Theta} Q(y, \theta), \qquad Q(y, \theta) = \frac{1}{n} \sum_{t=1}^{n} v_t^2(\theta). \tag{2.16}$$

Since the objective function is proportional to the exponent in the quasi-likelihood function, (2.16) is similar to the quasi-maximum likelihood estimator that is typically used in the state space literature, see e.g. Durbin and Koopman (2012, ch. 7). While the latter allows for a time-varying variance of the prediction error, (2.16) implicitly assumes a constant variance of the prediction error. However, as subsection 2.5.3 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, (2.16) and quasi-maximum likelihood estimation are asymptotically equivalent. Differences arise only due to a different weighting of prediction errors at the very beginning of

the sample. However, (2.16) is computationally much simpler, because it avoids the Kalman recursions for the prediction error variance. Furthermore, parameter estimation via the steady-state Kalman filter is identical to (2.16) after some burn-in period, see Harvey (1989, ch. 4.2.2).

While the asymptotic theory for CSS estimation is well established for autoregressive fractionally integrated moving average (ARFIMA) models, see Hualde and Robinson (2011) and Nielsen (2015), only little is known about the asymptotic theory for unobserved components models of such generality. For the sub-class of $I(1)$ UC models with Gaussian white noise shocks η_t and ϵ_t, the asymptotic theory can be inferred from the ARIMA literature (Harvey and Peters; 1990; Morley et al.; 2003). Unfortunately, no such results are available for UC models with fractional trends, so the asymptotic theory for parameter estimation of fractional UC models must be derived from scratch. While the proofs in this section are given for the (simpler) CSS estimator, it is shown in subsection 2.5.3 that they also apply to the traditional quasi-maximum likelihood estimator. Due to the encompassing nature of the fractional UC model, the results below also hold for CSS and quasi-maximum likelihood estimation of all sub-classes of UC models such as e.g. integer-integrated models with MDS shocks.

Theorem 2.4.1. *For the model in* (2.1)*,* (2.2)*, and* (2.4)*, and under assumptions 2.1 to 2.3, the estimator* $\hat{\theta}$ *as defined via* (2.16) *is consistent, i.e.* $\hat{\theta} \xrightarrow{p} \theta_0$ *as* $n \to \infty$*.*

The proof is contained in Appendix 2.A.2. While consistency ultimately follows from a uniform weak law of large numbers (UWLLN), showing that the UWLLN holds is complicated by the non-uniform convergence of the objective function within Θ, as well as by the non-ergodicity of the prediction errors in (2.14): First, as can be seen from (2.14), the prediction errors are $I(d_0 - d)$, and thus are asymptotically stationary for $d_0 - d < 1/2$, and otherwise non-stationary. In the former case, a UWLLN can be shown to hold for the objective function, while in the latter case a functional central limit theorem holds under some additional assumptions. Consequently, uniform convergence of the objective function fails around the point $d = d_0 - 1/2$. Following the idea of Nielsen (2015), I partition the parameter space D into three compact subsets, one where $v_t(\theta)$ is asymptotically non-stationary, one for stationary $v_t(\theta)$, and an overlapping subset. Next, whenever θ is not contained in the stationary region of the parameter space, I show that the objective function approaches infinity with probability converging to 1 as $n \to \infty$. Thus, the relevant region of the parameter space reduces asymptotically to the region where $d_0 - d < 1/2$ holds, and where uniform convergence of the objective function is not hindered.

Second, even within the asymptotically stationary region of the parameter space, the forecast errors are non-ergodic, as can be seen from (2.14) and (2.15): 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 (2.15) depends on t. Consequently, even for $d_0 - d < 1/2$, a law of large numbers for stationary and ergodic series does not apply directly to $v_t(\theta)$. I tackle this problem by showing that the difference between the prediction error in (2.14), and the untruncated and ergodic $\tilde{v}_t(\theta) = \sum_{j=0}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d)$, is asymptotically negligible in probability, where $\tilde{\xi}_t(d) = \Delta^{d-d_0}\eta_t + \Delta^d c_t$ is the untruncated residual, while the coefficients $\tau_j(\theta)$ stem from the ∞-vector $(\tau_1(\theta), \tau_2(\theta) \cdots) = \nu(b_1(\varphi) - \pi_1(d), b_2(\varphi) - \pi_2(d), \cdots)(B'_{\varphi,\infty}B_{\varphi,\infty} + \nu S'_{d,\infty}S_{d,\infty})^{-1}S'_{d,\infty}$, and $\tau_0(\theta) = 1$. Since $\tilde{v}_t(\theta)$ is stationary and ergodic within the stationary region of the parameter space, it follows that a weak law of large numbers applies to the objective function. The final part of the proof is to strengthen pointwise convergence in probability to weak convergence, which yields the desired result of theorem 2.4.1.

With a consistent parameter estimator at hand, I next derive the asymptotic distribution of the CSS estimator. For this purpose, assumption 2.3 needs to be strengthened.

Assumption 2.4. *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 2.4 is similar to assumption E of Nielsen (2015), and strengthens the smoothness conditions of the linear coefficients in $a(L, \varphi)$. It ensures absolute summability of the partial derivatives, which is used to prove uniform convergence of the Hessian matrix and thus to evaluate the Hessian matrix at θ_0 in the Taylor expansion of the score. The convergence rates of the (second and third) partial derivatives are a direct consequence of compactness of $N_\delta(\varphi_0)$ together with continuity of the partial derivatives. Assumption 2.4 still includes the class of stationary ARMA processes, and even allows for a slower rate of decay of the autocovariance function.

Assumption 2.5. *The true prediction error of the untruncated process $\tilde{v}_t(\theta_0)$ is a MDS when adapted to the filtration $\mathcal{F}_t^{\tilde{\xi}} = \sigma(\tilde{\xi}_s, s \leq t)$.*

So far, the Kalman filter, when applied to obtain the one-step ahead forecast

for $\tilde{\xi}_t = \eta_t + \Delta^{d_0} c_t$, yielded the projection of $\tilde{\xi}_t$ onto the span of $\tilde{\xi}_s$, $s < t$. Thus, the Kalman filter was the best linear predictor given $\mathcal{F}_{t-1}^{\tilde{\xi}}$ (in the least squares sense). Assumption 2.5 forces the prediction error to be a MDS when adapted to $\mathcal{F}_t^{\tilde{\xi}}$, which makes the Kalman filter the best predictor for $\tilde{\xi}_t$ given $\mathcal{F}_{t-1}^{\tilde{\xi}}$. Since $\tilde{v}_t(\theta_0)$ plays the role of the (asymptotic) residual for fractional UC models, assumption 2.5 fits well to the usual assumption of MDS residuals for CSS estimation, see e.g. Hualde and Robinson (2011), Nielsen (2015), and Hualde and Nielsen (2020). In the UC literature, Dunsmuir (1979, ass. C2.3) imposes the same assumption for his stationary signal plus noise model, but also discusses the possibility of relaxing the assumption (see Dunsmuir; 1979, pp. 502f).

Theorem 2.4.2. *For the model in* (2.1), (2.2), *and* (2.4), *under assumptions 2.1 to 2.5, the estimator* $\hat{\theta}$ *as defined via* (2.16) *is asymptotically normally distributed, i.e.* $\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} \mathrm{N}(0, \sigma_{v,0}^2 \Omega_0^{-1})$ *as* $n \to \infty$, *with* $\sigma_{v,0}^2 = \lim_{t\to\infty} \mathrm{Var}(v_t(\theta_0)) = \mathrm{Var}(\tilde{v}_t(\theta_0))$, *and* Ω_0 *has the* (i,j)*-th entry* $\Omega_{0_{(i,j)}} = \mathrm{E}\left(\frac{\partial \tilde{v}_t(\theta)}{\partial \theta_{(i)}}\big|_{\theta=\theta_0} \frac{\partial \tilde{v}_t(\theta)}{\partial \theta_{(j)}}\big|_{\theta=\theta_0}\right)$, $i, j = 1, ..., q+2$.

The proof of theorem 2.4.2 is contained in Appendix 2.A.3. As usual, the asymptotic distribution of the CSS estimator is inferred from a Taylor expansion of the score function around θ_0. Analogous to Robinson (2006) and Hualde and Robinson (2011), it is first shown that the normalized score at θ_0 is asymptotically equivalent to the score function of the untruncated, stationary and ergodic residual $\sqrt{n}(\partial \tilde{Q}(y,\theta)/\partial\theta)\big|_{\theta=\theta_0} = (2/\sqrt{n}) \sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\partial \tilde{v}_t(\theta)/\partial\theta)\big|_{\theta=\theta_0}$. Next, a UWLLN is shown to hold for the Hessian matrix, so that it can be evaluated at θ_0 in the Taylor expansion, and the difference between the truncated and untruncated Hessian matrix is shown to be asymptotically negligible in probability. Therefore, both the score and the Hessian matrix in the Taylor expansion can be replaced by their untruncated counterparts. While a weak law of large numbers applies to the untruncated Hessian matrix, a central limit theorem for martingale difference sequences applies to the score and yields the asymptotic distribution. Finally, while theorem 2.4.2 does not give an analytical expression for the covariance matrix of the CSS estimator, it shows that Ω_0^{-1} can by estimated via the numerical Hessian matrix.

2.5 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 become available that are beyond the scope of this paper, such as frequency-domain optimization, additional observable explanatory variables, time-varying and nonlinear models, and mixed-frequency models among others; see Harvey (1989) for an overview. In this section, I outline some generalizations of the fractional UC model that are of immediate applied relevance: Subsection 2.5.1 introduces deterministic components to the model, while subsection 2.5.2 allows for correlated trend and cycle innovations. Subsection 2.5.3 generalizes parameter estimation to the quasi-maximum likelihood estimator. For all three modifications, the asymptotic results of section 2.4 are shown to remain valid. However, before turning to the three generalizations, I first introduce the state space representation of the fractional UC model.

The basic state space representation has the form

$$y_t = Z\alpha_t + u_t, \tag{2.17}$$

$$\alpha_t = T\alpha_{t-1} + R\zeta_t, \tag{2.18}$$

where the states may be partitioned into $\alpha_t = (\alpha_t^{(x)\prime}, \alpha_t^{(c)\prime}, \alpha_t^{(r)\prime})'$, with $(n-1)$-vectors for trend $\alpha_t^{(x)} = (x_t, x_{t-1}, ..., x_{t-n+2})'$, and cycle $\alpha_t^{(c)} = (c_t, c_{t-1}, ..., c_{t-n+2})'$. The observation matrix is $Z = (Z^{(x)}, Z^{(c)}, Z^{(r)})$, where $Z^{(x)} = (1, 0, ..., 0)$, $Z^{(c)} = (1, 0, ..., 0)$ are $(n-1)$-dimensional row vectors picking the first entry of $\alpha_t^{(x)}$ and $\alpha_t^{(c)}$. For the transition equation (2.18), one has $T = \text{diag}(T^{(x)}, T^{(c)}, T^{(r)})$, $R = \text{diag}(R^{(x)}, R^{(c)}, R^{(r)})$,

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

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

and $R^{(x)} = (1, 0, ..., 0)'$, $R^{(c)} = (1, 0, ..., 0)'$ are $(n-1)$-vectors picking the respective entries of $\zeta_t = (\eta_t, \epsilon_t, \zeta_t^{(r)\prime})'$. 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 sea-

sonal 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 2.4, their specification in practice is left open to the applied researcher. Finally, $\text{Var}(\zeta_t) = Q$.

2.5.1 Deterministic components

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

Deterministic components can be taken into account either by detrending the data prior to estimating the fractional UC model, or by adding the components to the state space model. However, prior detrending biases the estimates for both deterministic and stochastic trends whenever the data are non-stationary, and thus should be avoided (Harvey; 1989, ch. 6.1.3). An alternative is to include the deterministic terms into the state vector and to explicitly model their dynamics via the state equation (2.18). However, state space models with deterministic components in the state vector are not stabilisable, so the Kalman filter does not converge to its steady state solution and the CSS estimator is not applicable, see Harvey (1989, ch. 4.2.5). Following the suggestion there, I place the deterministic terms directly in the measurement equation (2.17). This allows to estimate the deterministic components by the GLS estimator and does not interfere with the steady state convergence of the Kalman filter. The remaining θ_0 can be estimated via CSS as described in section 2.4, with the asymptotic theory being unaffected.

To model the deterministic terms, I set $u_t = \mu' w_t$ in the measurement equation (2.17), where w_t is a non-stochastic k-vector holding k deterministic components, and μ is a k-vector of unknown parameters to be estimated. The modified measurement equation is then $y_t = \mu' w_t + Z\alpha_t$. Letting $W = (w_1, ..., w_n)'$ denote the $n \times k$ matrix collecting all w_t, and $V = \text{Var}(x_{1:n} + c_{1:n})$ denote the variance-covariance matrix of $x_{1:n} + c_{1:n}$, the GLS estimator for μ is given by $\tilde{\mu} = (W'V^{-1}W)^{-1}W'V^{-1}y_{1:n}$, see Harvey (1989, ch. 3.4.2). As also shown there, it is not necessary to compute V^{-1}. To see this, assume for the moment that $y_t - \mu' w_t$ was observable. The Kalman filter, when applied to $y_t - \mu' w_t$, yields the filtered values for trend and

cycle in (2.6) to (2.9), together with the prediction errors as denoted by $v_t^*(\theta)$ in the following for the modified model. These prediction errors correspond to the linear filtering $F(\theta)(y_{1:n} - W\mu)$, where $F(\theta)$ from the Cholesky decomposition $V^{-1}(\psi) = F(\theta)'D^{-1}(\psi)F(\theta)$ is a p.d. lower triangular matrix with ones on the leading diagonal, $D(\psi)$ is a diagonal p.d. matrix, and $V(\psi)$ is the covariance matrix of $x_{1:n} + c_{1:n}$ conditional on ψ. Since the Kalman filter is linear, it can be applied separately to the observable y_t and w_t, yielding $F(\theta)y_{1:n} = y^*(\theta)$ and $F(\theta)W = W^*(\theta)$ as prediction errors. The GLS estimator $\tilde{\mu}$ then follows from regressing $y^*(\theta) = (y_1^*(\theta), ..., y_n^*(\theta))'$ on $W^*(\theta) = (w_1^*(\theta), ..., w_n^*(\theta))'$, see Harvey (1989, ch. 3.4.2). The concentrated CSS estimator $\tilde{\theta} = (\tilde{d}, \tilde{\nu}, \tilde{\varphi}')'$ follows from minimizing the modified sum of squared prediction errors

$$\tilde{\theta} = \arg\min_{\theta} \frac{1}{n} \sum_{t=1}^{n} v_t^*(\theta)^2, \tag{2.19}$$

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

To derive the asymptotic properties of both the GLS estimator $\tilde{\mu}$ and the concentrated CSS estimator (2.19), let the j-th term in w_t be $w_{j,t} = O(t^{\beta_j})$, $t \geq 1$, $\beta_j \in \mathbb{R}$, such that $w_{j,t}$ is a polynomial trend. I will only consider $-1 < \beta_j \leq d_0$ for all j, as the lower bound is required for $\Delta_+^{d_0} t^{\beta_j} = O(t^{\beta_j - d_0})$ to hold, see Robinson (2005), while the upper bound ensures that the fractional stochastic trend is not drowned by the deterministic terms. This guarantees that the results on consistency and asymptotic normality of the CSS estimator in theorems 2.4.1 and 2.4.2 remain valid. However, at least for CSS estimation of ARFIMA models, Hualde and Nielsen (2020) recently derived the asymptotic theory where they also allowed for deterministic trends of higher power, $\beta_j > d_0$. As the focus of this paper is not on the deterministic components, showing their results to carry over to fractional unobserved components models is left open for future research.

Note that within $-1 < \beta_j \leq d_0$, the arguments for consistency of the CSS estimator of θ_0 remain unchanged: $y^*(\theta) = F(\theta)y_{1:n}$ is $I(d_0 - d)$ and precisely equals the initial prediction error (2.14) in section 2.3 if y_t contains no deterministic terms, since $F(\theta)y_{1:n}$ is the residual from applying the Kalman filter as defined in section 2.3 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 (2.14) shifted either by a constant, or by an $o(1)$ term (depending on how close β_j is to d_0, as will become clear). Therefore, also the prediction error $v_t^*(\theta) = [y^*(\theta) - W^*(\theta)(W^{*'}(\theta)W^*(\theta))^{-1}W^{*'}(\theta)y^*(\theta)]_{(t)}$ is

$I(d_0-d)$. Thus, both $y_t^*(\theta)$ and $v_t^*(\theta)$ are asymptotically stationary for $d_0-d<1/2$, otherwise non-stationary. By the same proof as for (2.28), the objective function (2.19) can be shown to converge in probability whenever $d_0-d>-1/2$, and to diverge in the opposite case. Therefore, the probability of the CSS estimator to converge within the non-stationary region of the parameter space is asymptotically zero. Thus, it is sufficient to consider the region of the parameter space where $v_t^*(\theta)$ is asymptotically stationary. Within this region, the same proof as for theorem 2.4.1 applies, showing that a UWLLN holds for the objective function. Thus, $\tilde{\theta}$ is consistent. This result is somewhat obvious, since the assumption on β_j ensures that the filtered $y_t^*(\theta)$ contains at most deterministic terms of order $O(1)$.

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

$$\begin{aligned}\tilde{\mu}&=(W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})F(\tilde{\theta})y_{n:1}\\&=(W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})F(\tilde{\theta})\left[W\mu_0+x_{1:n}+c_{1:n}\right]\\&=\mu_0+(W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta}),\end{aligned}\tag{2.20}$$

where μ_0 denotes the true coefficients to be estimated. $\tilde{\mu}$ is consistent if and only if the latter term in (2.20) is $o_p(1)$, i.e. the bias converges to zero as $n\to\infty$. For the purpose of illustration, I will focus only on a single deterministic term, such that $W^*(\tilde{\theta})=(w_1^*(\tilde{\theta}),...,w_n^*(\tilde{\theta}))'$. However, the results carry over directly to several deterministic components. First, note that by the fractional differencing via $F(\tilde{\theta})$, $w_t^*(\tilde{\theta})=O(t^{\beta-\tilde{d}})$, while $u_t^*(\tilde{\theta})\sim I(d_0-\tilde{d})$. By consistency of the concentrated CSS estimator, $u_t^*(\tilde{\theta})$ is asymptotically $I(0)$, while $w_t^*(\tilde{\theta})=O(t^{\beta-d_0})$, and thus $\sum_{t=1}^n w_t^{*^2}(\tilde{\theta})=\sum_{t=1}^n O(t^{2(\beta-d_0)})$, see Hualde and Nielsen (2020, lemma S.10). Hence, for a single deterministic component, the bias term in (2.20) can be written as

$$(W^{*'}(\tilde{\theta})W^*(\tilde{\theta}))^{-1}W^{*'}(\tilde{\theta})u^*(\tilde{\theta})=\left(\frac{\sum_{t=1}^n w_t^{*^2}(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}}\right)^{-1}\frac{\sum_{t=1}^n w_t^*(\tilde{\theta})u_t^*(\tilde{\theta})}{n^{1+2(\beta-\tilde{d})}},\tag{2.21}$$

where $n^{-1-2(\beta-\tilde{d})}\sum_{t=1}^n w_t^{*^2}(\tilde{\theta})$ is bounded from above and below as $n\to\infty$. In contrast, by Hualde and Nielsen (2020, eqn. (S.88)), $n^{-1-2(\beta-\tilde{d})}\sum_{t=1}^n w_t^*u_t^*(\tilde{\theta})=o_p(1)$ if and only if $d_0-1/2<\beta$. Thus, the GLS estimator for the deterministic terms is consistent only if the deterministic and stochastic trends diverge at similar

rates. As also can be seen from (2.21), the power of the deterministic term affects the rate of convergence of the GLS estimator: Since $n^{-1/2-(\beta-\tilde{d})}\sum_{t=1}^{n} w_t^*(\tilde{\theta})u_t^*(\tilde{\theta})$ converges in distribution when $n \to \infty$, see Hualde and Nielsen (2020, proof of cor. 1), it follows that the GLS estimator converges at the rate $n^{1/2+(\beta-d_0)}$ as $n \to \infty$, and thus the rate is slower than the standard $\sqrt{n}$-convergence whenever the deterministic terms are dominated by the stochastic trend.

In summary, any trend of order $d_0 - 1/2 < \beta_j \leq d_0$ can be estimated consistently, and the 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 finding in the literature, that an intercept (i.e. $\beta_j = 0$) cannot be estimated consistently for time series with unit roots ($d_0 = 1$), whereas a linear trend ($\beta_j = 1$) can be estimated consistently. 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).

2.5.2 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

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

The new optimization problem of the Kalman filter is then

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

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

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

where the matrix representation in the last step is derived analogously to (2.12). The solution to the optimization problem is then

$$\begin{aligned}\hat{x}_{t:1}(y_{t:1},\tilde{\psi}) = &\left[\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}'B_{\varphi,t} + B_{\varphi,t}'S_{d,t}) + \sigma_\epsilon^2 S_{d,t}'S_{d,t}\right]^{-1} \\ &\times\left(\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}S_{d,t}'B_{\varphi,t}\right)y_{t:1},\end{aligned} \tag{2.22}$$

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

$$\begin{aligned}\hat{c}_{t:1}(y_{t:1},\tilde{\psi}) = &\left[\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}'B_{\varphi,t} + B_{\varphi,t}'S_{d,t}) + \sigma_\epsilon^2 S_{d,t}'S_{d,t}\right]^{-1} \\ &\times\left(\sigma_\epsilon^2 S_{d,t}'S_{d,t} + \sigma_{\eta\epsilon}B_{\varphi,t}'S_{d,t}\right)y_{t:1}.\end{aligned} \tag{2.23}$$

Obviously, (2.22) and (2.23) equal (2.6) and (2.7) 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 (2.22) and (2.23) by σ_η^2, defining $\nu = \sigma_\epsilon^2/\sigma_\eta^2$ as well as $\nu_2 = \sigma_{\eta\epsilon}/\sigma_\eta^2$, and replacing $\tilde{\psi}$ by $\bar{\theta} = (d,\nu,\nu_2,\varphi')'$. This is necessary for the CSS estimator to be identified, however the quasi-maximum likelihood estimator derived in subsection 2.5.3 can be used to estimate $\tilde{\psi}_0 = (d_0,\sigma_{\eta,0}^2,\sigma_{\eta\epsilon,0},\sigma_{\epsilon,0}^2,\varphi_0')$, the true parameters, directly.

The objective function for the CSS estimator can be constructed analogously to section 2.4: First, the one-step ahead predictions for x_{t+1} and c_{t+1} are obtained as in (2.8) and (2.9). Next, they are subtracted from y_{t+1}, which gives the prediction error

$$\begin{aligned}v_{t+1}(\tilde{\psi}) = &\Delta_+^d y_{t+1} + (b_1(\varphi) - \pi_1(d)\cdots b_t(\varphi) - \pi_t(d)) \\ &\times\left[\sigma_\eta^2 B_{\varphi,t}'B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}'B_{\varphi,t} + B_{\varphi,t}'S_{d,t}) + \sigma_\epsilon^2 S_{d,t}'S_{d,t}\right]^{-1} \\ &\times\left(\sigma_\epsilon^2 S_{d,t}' + \sigma_{\eta\epsilon}B_{\varphi,t}'\right)S_{d,t}y_{t:1}.\end{aligned} \tag{2.24}$$

Based on (2.24), a CSS estimator for the true parameters $\bar{\theta}_0 = (d_0,\nu_0,\nu_{2,0},\varphi_0')$ can

be set up. Note that y_{t+1} enters (2.24) in fractional differences, and also note that all terms in (2.24) have the same convergence rates as for the case with uncorrelated errors. Thus, the CSS estimator with correlated innovations can be shown to be consistent and asymptotically normally distributed by carrying out the same proofs as summarized in section 2.4. Finally, as noted by Morley et al. (2003), for the integer-integrated case $d_0 = 1$, the model is not identified if c_t follows an AR(p) with $p < 2$, since the autocovariance function of Δy_t dies out after lag one. For non-integer integration orders, identification is not a problem, as the autocovariance function of $\Delta_+^d y_t$ dies out only at lag t.

2.5.3 Maximum likelihood estimation

Since the vast majority of state space models are estimated by quasi-maximum likelihood (QML), this subsection relates the CSS estimator to the QML estimator. For this purpose, denote $\psi = (d, \sigma_\eta^2, \sigma_\epsilon^2, \varphi)'$ the vector holding the model parameters of the fractional UC model. Furthermore, let $\text{Var}_\psi\left(v_t(\psi)|y_1, ..., y_{t-1}\right) = \sigma_{v_t}^2$ denote the (hypothetical) variance of $v_t(\psi)$ that is obtained when evaluating the conditional distribution of $v_t(\psi)$ at ψ. While the CSS estimator allowed to concentrate out the variance parameters $\sigma_\eta^2, \sigma_\epsilon^2$ and model only their variance ratio $\nu = \sigma_\epsilon^2/\sigma_\eta^2$, this is not possible for the QML estimator, since the levels of $\sigma_\eta^2, \sigma_\epsilon^2$ determine $\sigma_{v_t}^2$. Thus, optimization is conducted over ψ. Note further that ψ can be extended to account for correlated innovations, as described in subsection 2.5.2. A recursive solution for $\sigma_{v_t}^2$ is typically obtained from the Kalman filter, see Durbin and Koopman (2012, ch. 4.3). The quasi-log likelihood is then set up based on the conditional distribution of $v_t(\psi)$ and is given by

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

see Harvey (1989, ch. 3.4). Now, if the Kalman filter converges to its steady state solution at an exponential rate, the QML estimator is asymptotically independent of the initialization of the Kalman filter, see Harvey (1989, ch. 3.4.2), and $\sigma_{v_t}^2$ converges to a constant. Thus, neither the initialization of the Kalman filter, nor the time-dependence of $\sigma_{v_t}^2$ matters asymptotically, and therefore the CSS estimator in (2.16) has the same asymptotic distribution as the QML estimator, see Harvey (1989, p. 129).

For the Kalman filter to converge to its steady state solution at an exponential rate, it is sufficient that the state space model is detectable and stabilizable (Har-

vey; 1989, ch. 3.3.3). Detectability is implied by observability, while stabilizability is implied by controllability (Harvey; 1989, ch. 3.3.1). The state space model as introduced at the beginning of this section is controllable if $\text{Rank}(G, TG, ..., T^{m-1}G) = m$, where m is the dimension of α_t, and $G = RS'$ where S is the upper-triangular matrix from the Cholesky decomposition of the covariance matrix $Q = S'S$ (Harvey; 1989, ch. 3.3.1). The rank condition can be verified by simple algebra, and depends crucially on Q having full rank. Controllability means that given a realization of α_t at some period t, the innovations ζ_{t+j}, $j = 1, ..., m$, can be chosen such that an arbitrarily prescribed value α^*_{t+m} is obtained. Since in each period a new innovation enters (2.18) for both x_t and c_t, their states in α_{t+m} can be controlled by controlling ζ_{t+j}. Thus, the state space model is controllable. Similarly, the state space model is observable if $\text{Rank}(Z', T'Z', ..., (T')^{m-1}Z') = m$ (Harvey; 1989, ch. 3.3.1), which again can be verified algebraically. The idea of observability is that α_t can be uniquely determined if $y_t, ..., y_{t+m-1}$, as well as $\zeta_t, ..., \zeta_{t+m-1}$ are known. This is easy to see: Suppose y_{t+j} is known for some $j > 0$. Then $\Delta^d_+ y_{t+j} = \eta_{t+j} + \Delta^d_+ c_{t+j}$ can be calculated. With η_{t+j} at hand, we can directly calculate c_{t+j}, and thus also x_{t+j}. It follows that the system is observable. Thus, as $n \to \infty$, the CSS estimator and the QML estimator become identical, which was also pointed out by Harvey (1989, p. 187) for integer-integrated models. Consequently, the results in section 2.4 also hold for the QML estimator.

Finally, while computational efficiency clearly favors the CSS estimator, which avoids the Kalman recursions for the conditional variance of the state vector, the QML estimator may be advantageous in finite samples where the initialization of the Kalman filter plays a non-negligible role. In particular, a combination of the QML estimator, for an initial burn-in period, and the CSS estimator, once the filtered prediction error variance has sufficiently converged, seems promising: It combines the possibility of diffuse initialization and thus assigns a lower weight to initial prediction errors, but switches to the computationally efficient CSS estimator once the benefits of the QML estimator have vanished. The performance of this estimator, typically called the steady-state filter (Harvey; 1989, p. 185f), is also examined in a Monte Carlo study in section 2.6 and compared to the CSS estimator.

2.6 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 2.2. By considering both the CSS estimator of section

2.4 and the QML estimator of subsection 2.5.3, the study demonstrates the loss of estimation accuracy of the computationally simpler CSS estimator by treating the filtered prediction error variance to be constant. Thus, the study puts a price tag on the computational efficiency gains and provides empirical researchers with guidance on when to use the CSS estimator. Furthermore, the parameter estimates for the integration order are compared to the exact local Whittle estimator of Shimotsu and Phillips (2005) for various choices of tuning parameters as a prominent benchmark. To see whether allowing for fractional trends matters, I also present results for the integer-integrated UC models in the spirit of Harvey (1985) and Morley et al. (2003). Doing so, I examine whether fractional trends are well approximated by integer-integrated models, or whether the estimates for x_t and c_t are significantly biased. Furthermore, I investigate whether misspecifying d as one biases the parameter estimates.

Two different data-generating mechanisms are considered: Subsection 2.6.1 simulates data based on the fractionally integrated UC model with uncorrelated trend and cycle innovations as introduced in section 2.2, while subsection 2.6.2 in addition allows for correlated innovations as discussed in subsection 2.5.2. Both studies vary over the sample size $n \in \{100, 200, 300\}$, the integration order $d_0 \in \{0.75, 1.00, 1.25\}$, and the variance ratio of trend and cycle innovations $\nu_0 = \frac{\sigma^2_{\epsilon,0}}{\sigma^2_{\eta,0}} \in \{1, 5, 10\}$. Thus, they capture small to medium sized samples as typical in empirical applications of UC models, allow for non-stationary mean-reverting trends as well as for non-mean-reverting trends, and reflect situations where short- and long-run shocks are of equal magnitude as well as situations where the long-run shocks are drowned by the short-run dynamics. Each simulation consists of $R = 1000$ replications.

Unlike the CSS estimator, the QML estimator uses the Kalman iterations for the variance of the prediction error, thereby allowing it to be time-dependent: In the Kalman filter, trend and cycle are first initialized with variances σ^2_η and σ^2_ϵ. Then, in a burn-in period, the QML estimator takes into account the exponential convergence of the prediction error variance by allowing it to converge to its steady-state value. Once the prediction error variance has converged sufficiently, i.e. it satisfies $\left|\frac{\mathrm{Var}_\psi(v_{t+1}(\psi)|y_1,\ldots,y_t) - \mathrm{Var}_\psi(v_t(\psi)|y_1,\ldots,y_{t-1})}{\mathrm{Var}_\psi(v_t(\psi)|y_1,\ldots,y_{t-1})}\right| < 0.01$, the optimization switches to the steady state Kalman filter, which assumes the prediction error variance to be 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 accuracy. The exact local Whittle estimator of Shimotsu and Phillips (2005) 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 assess how well trend and cycle are estimated, the coefficients of determination R_x^2 and R_c^2 from regressing x_t and c_t on their respective estimates from the Kalman smoother are reported for both the CSS and QML estimates.

2.6.1 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

$$y_t = x_t + c_t, \qquad \Delta_+^d x_t = \eta_t, \qquad c_t - b_1 c_{t-1} - b_2 c_{t-2} = \epsilon_t, \tag{2.25}$$

where $\eta_t \sim \text{NID}(0,1)$, $\epsilon_t \sim \text{NID}(0,\nu)$ 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 $\theta_{start} = (d_{start}, \nu_{start}, b_{1_{start}}, b_{2_{start}})' = (1, 1, 0.5, -0.5)'$, for both the CSS and the QML estimator. Note that for the QML estimator this implies assuming that $\sigma_{\eta,0}^2 = 1$ is known, since only ν_0 is estimated. Although this assumption is usually violated, it allows for a fairer comparison between the CSS and the QML estimator, which is the focus of this first simulation study. The $I(1)$ UC model is initialized analogously using $(\nu_{start}, b_{1_{start}}, b_{2_{start}})' = (1, 0.5, -0.5)'$.

Table 2.1 shows the RMSE and the median bias for the estimated integration orders for the CSS estimator, the QML estimator, and the exact local Whittle estimator. As can be seen, the RMSE decreases as n increases, which is in line with the theoretical results on consistency. As can be expected from the parametric nature, the fractional UC models yield a smaller RMSE as compared to the nonparametric Whittle estimator. The differences are particularly striking for higher ν_0, where the signal of the fractional trend is drowned by a strong cyclical variation, and for high n. In a direct comparison, the QML estimator slightly outperforms the CSS estimator for the estimation of the integration order, but the differences are rather small. Both the CSS and the QML estimator appear to have little or no bias for d_0, while the cyclical dynamics induce a strong bias on the exact local Whittle estimates.

Tables 2.2 and 2.3 contain the RMSE and the median bias for ν_0 and the autoregressive parameters, for both the CSS and the QML estimator. In addition to the fractional UC model, the table also displays the estimation results for an $I(1)$ UC benchmark that sets $d = 1$. While there is little difference between the CSS and the QML estimator in terms of the integration order estimate, for ν_0 both the bias and the RMSE are significantly smaller for the QML estimator. For $b_{1,0}$ and $b_{2,0}$, the CSS

estimator and the QML estimator perform similarly. The direct comparison with the $I(1)$ benchmark reveals that there is little to no difference for the estimation of $b_{1,0}$ and $b_{2,0}$, while ν_0 is typically estimated with a higher precision via the fractional UC model whenever $d_0 \neq 1$.

Table 2.4 compares the estimates for x_t and c_t for the fractional UC model and the $I(1)$ UC benchmark (which sets $d = 1$). As before, it contains the parameter estimates for the CSS estimator and the QML estimator. As can be seen, differences between the coefficients of determination are negligible. Strikingly, for $d_0 = 1$ the fractional UC model shows no loss in efficiency compared to the $I(1)$ UC model. For non-integer d_0, the coefficient of determination for x_t should not be interpreted for the $I(1)$ benchmark, as a high R_x^2 may also result from a spurious regression, and thus only the R_c^2 is considered. There, the fractional model clearly outperforms the benchmark model, especially when ν_0 is small. However, the R_c^2 is still relatively high for the $I(1)$ benchmarks, so that, at least for the setup considered, integer-integrated UC models are able to approximate the fractionally integrated trend well.

2.6.2 Fractional UC model with correlated innovations

To examine the estimation properties for the latent components and parameters of the fractional UC model when the long- and short-run innovations are allowed to be correlated, I modify (2.25) by allowing for a non-diagonal Q in

$$\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim \text{NID}(0, Q). \tag{2.26}$$

As before, the cyclical coefficients are set to $b_{1,0} = 1.6$, $b_{2,0} = -0.8$. Q_0 is parameterized as $\sigma_{\eta,0} = 1$, $\sigma_\epsilon = \nu_0 \in \{1, 5, 10\}$, which yields medium to strong cyclical fluctuations. I set $\sigma_{\eta\epsilon,0} = \rho_0\sqrt{\nu_0}$ with $\rho_0 = -0.2$, so that long- and short-run innovations are slightly negatively correlated. Starting values for the numerical optimization are set to $\theta_{start} = (d_{start}, \nu_{start}, \nu_{2,start}, b_{1_{start}}, b_{2_{start}})' = (1, 1, 0, 0.5, -0.5)'$, and to $(\nu_{start}, \nu_{2,start}, b_{1_{start}}, b_{2_{start}})' = (1, 0, 0.5, -0.5)'$ for the $I(1)$ UC model. For the fractional UC model, I only present estimation results for the CSS estimator. This is because the QML estimation in the correlated setup is computationally expensive. Furthermore, while the optimization is performed over ν_2, results are reported for the transformed $\rho = \nu_2/\sqrt{\nu}$, as the correlation is easier to interpret.

For the correlated fractional UC model, table 2.5 shows the RMSE and the median bias for the estimated integration orders via the CSS and the exact local Whittle estimator. As before, the RMSE decreases in n. While the fractional UC

model outperforms most of the Whittle estimates, the latter performs surprisingly well for a bandwidth choice of $\alpha = 0.65$ for $n = 100$, and $\alpha = 0.70$ for $n = 200$. However, for $n = 300$, all benchmarks are outperformed by the fractional UC model. As before, estimates for the fractional UC model show little to no bias for d_0, while the benchmarks are significantly perturbed by the cyclical dynamics.

Tables 2.6 and 2.7 show the RMSE and the median bias for ν_0, ρ_0, and the autoregressive parameters for the fractional UC model and the integer-integrated UC model. As in the uncorrelated case, the estimates for ν_0 have a large RMSE and are biased. However, the bias is more pronounced for the $I(1)$ benchmark, where the RMSE is also higher. More interestingly, the benchmark estimates for ν_0 are upward-biased whenever $d_0 < 1$, and downward-biased whenever $d_0 > 1$. Since $\nu_0 = \sigma^2_{\epsilon,0}/\sigma^2_{\eta,0}$ is the variance ratio of the innovations, this is natural: Whenever $d_0 < 1$, the random walk for a fixed σ^2_η has a faster diverging variance than the $I(d_0)$ process. To compensate for the slower rate of divergence of the $I(d_0)$ process, $\hat{\nu}$ must be upward-biased in the $I(1)$ model, and vice versa for $d_0 > 1$. For ρ_0, note that a similar pattern is visible for the CSS estimates. For $d_0 < 1$, the estimates for the correlation between long- and short-run shocks are upward-biased, and often positive. This is due to the upward-biased $\hat{\nu}$, which yields an estimate for the trend that is smoother than the true one. Thus, the cycle needs to account for the additional long-run fluctuations that are not captured by the smooth trend, which can be achieved by a positive estimate for the correlation coefficient. For $d_0 > 1$, the smoothed trend of the $I(1)$ model is more volatile than the true one, and the $I(1)$ UC model re-adjusts by estimating a downward-biased correlation coefficient, resulting in a more negative relation between trend and cycle than in the data-generating mechanism. Note that the potential for adjustment of the $I(1)$ model to fractionally integrated trends via the correlation parameter estimate is limited by the nature of the correlation $\rho \in [-1; 1]$, and thus corner solutions with $\hat{\rho} = -1$ can be expected when d_0 is greater than one, and with $\hat{\rho} = 1$ whenever d_0 is smaller than one. As before, there are little to no differences for the estimates of the autoregressive coefficients between the fractional model and the $I(1)$ model.

As for the uncorrelated models, table 2.8 compares the estimates for trend and cycle of the fractional and the $I(1)$ UC model via the coefficients of determination. As before, the performance is similar for $d_0 = 1$. For non-integer integration orders the fractional model yields better estimates for the cycle whenever ν_0 is small. For high ν_0, the coefficients of determination are similar.

2.7 Application

In this section, I apply the fractional UC model to log annual US CO2 emissions. Beyond estimates of the memory parameter, which may be of interest in their own right[4], I address the following research questions: (i) What is the trending behavior of US carbon 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 carbon 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 carbon emissions? Does it align with the business cycle, as results of Doda (2014) suggest? (iii) Is there evidence of a decoupling of economic activity and CO2 emissions (see Haberl et al.; 2020)? Does the decoupling affect the cycle, the trend, or both? (iv) Is there evidence for correlation between 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 to be gained from a fractional model compared to integer-integrated UC models?

Data on annual US carbon 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 millions of tons, is log-transformed to account for the exponential growth, and is shown in figure 2.1.

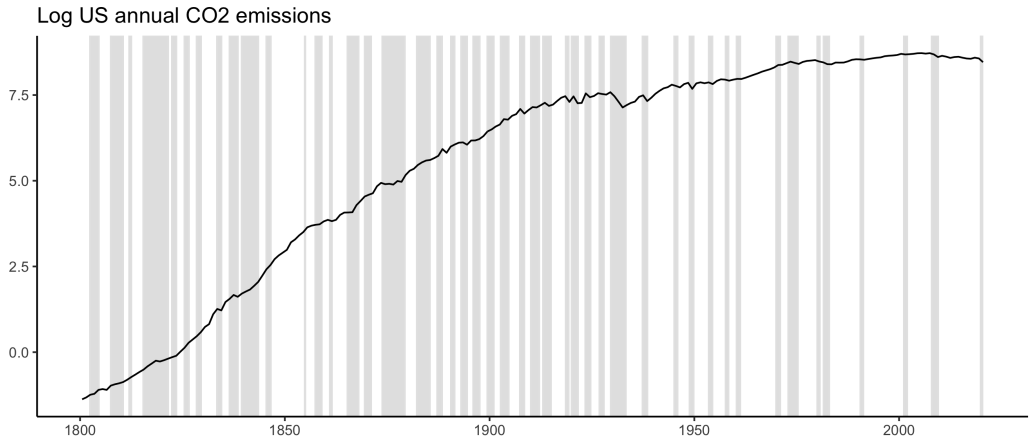

Figure 2.1: Log annual US carbon emissions from 1800 to 2020. Shaded areas correspond to US recession periods. Data stem from the Global Carbon Project and were collected by Ritchie et al. (2020).

From figure 2.1, it becomes apparent that log carbon emissions, at least for the

[4]To date, there is no consensus on how to appropriately model the long-run dynamics of carbon emissions, see Wagner (2008) for a discussion of model specification issues.

first halve of the sample, 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 of Shimotsu (2010), which includes 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 2.6 found the exact local Whittle estimator to be strongly 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 consistent with the UC literature. The resulting fractional UC model is thus given by

$$y_t = \mu_0 + \mu_1 t + x_t + c_t, \qquad \Delta_+^d x_t = \eta_t, \qquad \sum_{j=0}^{p} b_j c_{t-j} = \epsilon_t, \tag{2.27}$$

where $b_0 = 1$, and μ_0 and μ_1 account for a constant and a linear trend. Moreover, $\text{Var}(\eta_t, \epsilon_t)' = Q$. 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 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. The objective function of the CSS estimator is then minimized numerically for each of the 100 starting values, and the estimate corresponding to the smallest value of the objective function is chosen as the final estimate.

Table 2.9 shows the estimation results for $p \in \{0, 1, 2, 3, 4, 5\}$, along with the corresponding value of the objective function, for both uncorrelated and correlated innovations. As can be seen, for $p > 3$ both the estimates for the correlated and the uncorrelated model are relatively stable. The integration order is found to be around 1.75, indicating that trend carbon emissions are strongly persistent, non-mean-reverting, and clearly closer to a quadratic stochastic 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 suggests a converging effect 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. Since the specification with $p = 5$ autoregressive lags for the cyclical component and correlated innovations encompasses the other specifications in table 2.9, it is more robust to

model misspecification than the other specifications. It is selected as the preferred model and is examined in more detail below.

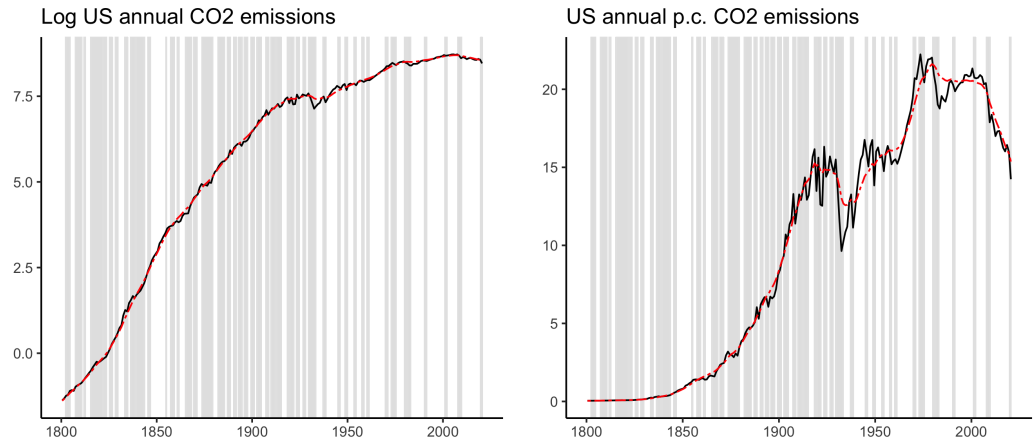

Figure 2.2: Trend CO2 emissions: The left plot sketches log annual US carbon emissions (black) together with the estimated trend $\hat{x}_t(y_{n:1}, \hat{\theta}) + \hat{\mu}_0 + \hat{\mu}_1 t$ (red). The right plot shows annual US carbon emissions per capita (measured in tons), together with the respective transformation of the trend estimate. Shaded areas correspond to US recession periods.

Figure 2.2 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 annual US carbon emissions. The left plot shows the series in logs, while the right plot displays annual US per capita carbon emissions in tons CO2. The fractional UC model estimates a smooth trend which is due to the relatively high estimate $\hat{\nu}$, 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 mark two turning points in per capita carbon 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 2.2, support the environmental Kuznets curve hypothesis.

Figure 2.3 shows the smoothed estimates for the cycle $\hat{c}_t(y_{n:1}, \hat{\theta})$. In line with the high estimate $\hat{\nu}$, the smoothed cyclical component exhibits rich dynamics and persistent behavior. Clearly, $\hat{c}_t(y_{n:1}, \hat{\theta})$ evolves along the business cycle, as sharp declines occur mostly during recession periods, while gradual increases in cyclical carbon emissions happen during periods of economic recovery and prosperity. The massive downturn during the Great Depression is particularly striking. Since the second half of the 20th century, the magnitude of pro-cyclical variation appears to have decreased. While the decoupling of economic activity and emissions may be true for the long-run behavior, as suggested by figure 2.2, figure 2.3 shows that cyclical

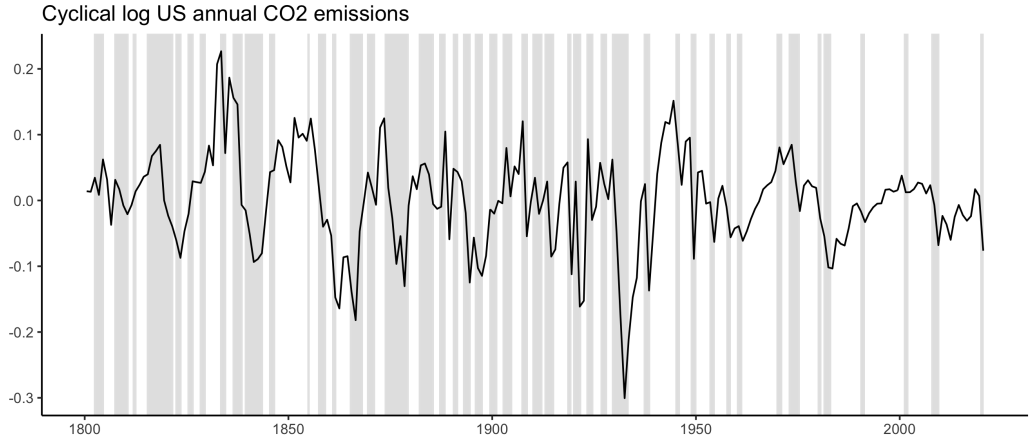

Figure 2.3: Estimated cyclical log annual US carbon emissions $\hat{c}_t(y_{n:1}, \hat{\theta})$. Shaded areas correspond to US recession periods.

emissions and the business cycle are still coupled. Finally, there is moderate positive correlation between long- and short-run innovations, as table 2.9 shows. One possible explanation, which is also supported by figures 2.2 and 2.3, is that recessions do not only lead to a decline in cyclical economic activity and thus in cyclical emissions. Instead, they may also have permanent effects on the economy, e.g. through the replacement of outdated technologies with newer ones, through a permanent reduction of the workforce, or through a transformation of energy production. The positive relationship between permanent and transitory shocks calls for further investigation, and I leave this open for future research.

Finally, I investigate to what extent the fractional UC model reveals new information about the trending and cyclical behavior of carbon emissions by comparing the above results to integer-integrated benchmark models. I consider a model in the spirit of Harvey (1985), in which x_t is assumed to be a random walk, c_t is an autoregressive process, and correlation between long- and short-run innovations is excluded. As a second model, I consider the correlated UC model of Morley et al. (2003), that in addition allows for correlated innovations. The third benchmark is the filter of Hodrick and Prescott (1997), which assumes x_t to be $I(2)$. The first two models are obtained by setting $d = 1$ in (2.27), while the HP filter is obtained by setting $t = n$, $d = 2$, $b(L, \varphi) = 1$ in (2.10), where $\nu = \sigma_\epsilon^2/\sigma_\eta^2$ is the tuning parameter of the HP filter, as also discussed in section 2.3.

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 initial-

ized as before. The starting values enter the numerical optimization of the quasi-likelihood, and the QML estimates corresponding to 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.

Table 2.10 contains the parameter estimates for both the uncorrelated and the correlated $I(1)$ UC model. For the latter, the estimates for the correlation coefficients converge to -1, so that 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 does not fully capture the long-run dynamics of log annual US carbon 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 parameterizations of table 2.10 evolve in a non-mean-reverting manner.

While the $I(1)$ specification is clearly at odds with the estimation results for d in table 2.9, the $I(2)$ trend assumption of the HP filter can be expected to better match the long-run dynamics of log annual US carbon emissions. Instead of estimating a parametric model, the HP filter requires setting a tuning parameter ν that penalizes the cyclical dynamics. As shown earlier, 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 ν, the more variation is attributed to the cyclical component. Following Ravn and Uhlig (2002), I set $\nu = 6.25$, which is typically chosen in the empirical literature for annual data, and was also set by Doda (2014) for decomposing log carbon emissions into trend and cycle via the HP filter. Figure 2.4 shows the estimated cycle from the HP filter along with the estimate from the fractional UC model with correlated innovations. The HP filter attributes less variation to the cyclical component than the fractional UC model. It lacks the persistent patterns of peaks and troughs, is comparatively noisy, and misses the cyclical patterns at the end of the sample. Thus, compared to both, $I(1)$ and $I(2)$ trend specifications, the fractional UC model offers additional insights into the permanent and cyclical dynamics of annual US carbon emissions.

2.8 Conclusion

This paper introduces a novel unobserved components model in which the trend component is specified as a type II fractionally integrated process. The model encompasses the bulk of unobserved components models in the literature, allows for

richer long-run dynamics beyond integer-integrated specifications, and for a data-dependent specification of the trend. Trend and cycle are estimated via the analytical solution to the optimization problem of the Kalman filter. The model allows for a joint estimation of the integration order and the other model parameters via the conditional sum-of-squares estimator, which is shown to be consistent and asymptotically normally distributed. For log annual US carbon emissions, the fractional unobserved components model reveals a smooth trend component starting to exhibit an inverted U-shape, together with a rich cyclical component that evolves along the business cycle.

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

2.A Appendix

2.A.1 Additional figures and tables

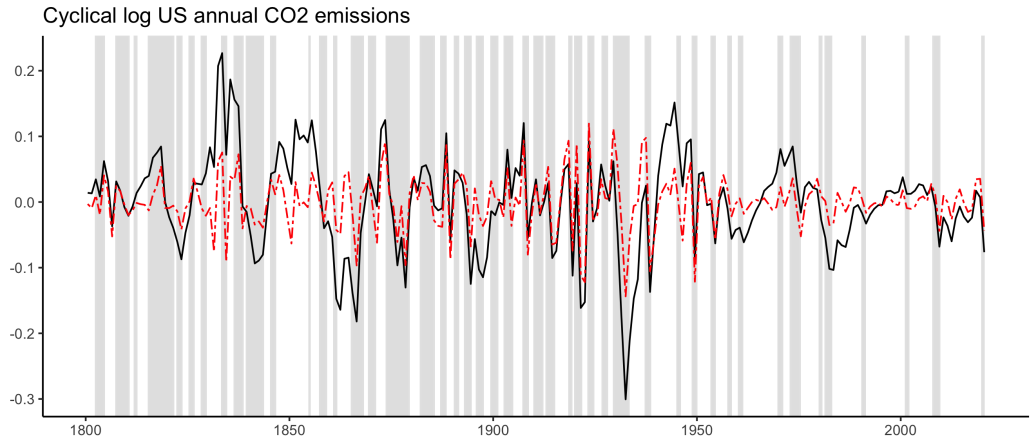

Figure 2.4: Estimated cyclical component of the HP filter with $\lambda = 6.25$ as suggested by Ravn and Uhlig (2002) (red, dashed), and of the fractional UC model with correlated innovations $\hat{c}_t(y_{n:1}, \hat{\theta})$ (black, solid). Shaded areas correspond to US recession periods.

			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 2.1: Root mean squared errors (RMSE) and median bias for the integration order estimates of the fractional UC model with uncorrelated innovations in subsection 2.6.1. The columns indicate the integration order estimates via the CSS estimator ($\hat{d}_{CSS}$), the QML estimator $\hat{d}_{QML}$, and the Whittle estimator of Shimotsu and Phillips (2005) with tuning parameter α ($\hat{d}^{EW}_{\alpha}$).

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}_{1_{QML}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}_{2_{QML}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
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 2.2: Root mean squared errors (RMSE) for the other parameter estimates of the fractional UC model with uncorrelated innovations in subsection 2.6.1. The different columns indicate the parameter estimates via the CSS estimator (subscript CSS) and the QML estimator (subscript QML) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}_{QML}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}_{1_{QML}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}_{2_{QML}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
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 2.3: Median bias for the other parameter estimates of the fractional UC model with uncorrelated innovations in subsection 2.6.1. The different columns indicate the parameter estimates via the CSS estimator (subscript CSS) and the QML estimator (subscript QML) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

			Trend				Cycle			
n	ν_0	d_0	R^2_{CSS}	R^2_{QML}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$	R^2_{CSS}	R^2_{QML}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$
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 2.4: Coefficient of determination from regressing true trend and cycle x_t and c_t on their respective estimates from the Kalman smoother for the uncorrelated UC models.

			RMSE						bias					
n	ν_0	d_0	$\hat{d}_{CSS}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$	$\hat{d}_{CSS}$	$\hat{d}^{EW}_{.50}$	$\hat{d}^{EW}_{.55}$	$\hat{d}^{EW}_{.60}$	$\hat{d}^{EW}_{.65}$	$\hat{d}^{EW}_{.70}$
100	1	.75	.332	.639	.578	.409	.240	.602	-.136	-.623	-.554	-.357	.065	.551
		1.00	.279	.668	.593	.432	.214	.440	-.050	-.636	-.560	-.385	-.050	.376
		1.25	.282	.631	.560	.420	.219	.309	-.015	-.589	-.520	-.374	-.115	.231
	5	.75	.379	.718	.679	.511	.274	.776	-.195	-.714	-.668	-.468	.074	.728
		1.00	.352	.867	.804	.623	.274	.569	-.142	-.854	-.788	-.591	-.120	.503
		1.25	.324	.894	.826	.664	.345	.385	-.111	-.871	-.803	-.637	-.267	.290
	10	.75	.410	.728	.694	.530	.279	.803	-.233	-.725	-.686	-.489	.075	.755
		1.00	.399	.918	.863	.681	.293	.587	-.225	-.910	-.851	-.651	-.141	.521
		1.25	.363	.988	.922	.754	.393	.394	-.192	-.970	-.903	-.729	-.317	.293
200	1	.75	.237	.614	.640	.567	.389	.154	-.077	-.600	-.631	-.555	-.370	.036
		1.00	.169	.589	.628	.551	.396	.145	-.019	-.562	-.609	-.534	-.377	-.065
		1.25	.185	.522	.572	.507	.376	.164	-.019	-.490	-.551	-.488	-.356	-.118
	5	.75	.335	.724	.734	.700	.525	.184	-.160	-.721	-.733	-.696	-.511	.038
		1.00	.288	.816	.848	.779	.606	.220	-.090	-.803	-.839	-.769	-.594	-.147
		1.25	.258	.782	.830	.765	.623	.309	-.068	-.762	-.815	-.752	-.611	-.277
	10	.75	.357	.738	.744	.722	.558	.188	-.190	-.737	-.743	-.720	-.544	.037
		1.00	.306	.887	.912	.853	.676	.243	-.144	-.879	-.906	-.845	-.665	-.172
		1.25	.288	.885	.928	.864	.716	.364	-.113	-.868	-.915	-.852	-.705	-.333
300	1	.75	.206	.505	.612	.605	.499	.224	-.041	-.483	-.602	-.596	-.489	-.200
		1.00	.142	.444	.575	.575	.479	.257	-.008	-.412	-.558	-.562	-.467	-.239
		1.25	.170	.375	.514	.525	.441	.260	-.009	-.335	-.494	-.510	-.428	-.243
	5	.75	.271	.675	.728	.727	.658	.317	-.104	-.669	-.726	-.725	-.653	-.294
		1.00	.221	.681	.799	.797	.701	.427	-.056	-.664	-.789	-.789	-.692	-.414
		1.25	.220	.612	.757	.766	.685	.475	-.043	-.590	-.744	-.755	-.675	-.465
	10	.75	.338	.710	.742	.742	.693	.339	-.173	-.706	-.741	-.742	-.690	-.316
		1.00	.286	.773	.874	.872	.780	.482	-.118	-.760	-.867	-.866	-.773	-.471
		1.25	.249	.716	.854	.860	.778	.556	-.072	-.698	-.841	-.850	-.769	-.547

Table 2.5: Root mean squared errors (RMSE) and median bias for the integration order estimates of the fractional UC model with correlated innovations in subsection 2.6.2. The different columns indicate the integration order estimates via the CSS estimator ($\hat{d}_{CSS}$) and the Whittle estimator of Shimotsu and Phillips (2005) with tuning parameter α ($\hat{d}^{EW}_{\alpha}$).

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{\rho}_{CSS}$	$\hat{\rho}^{I(1)}_{CSS}$	$\hat{\rho}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
100	1	.75	3.871	1.157	51.513	.821	1.105	1.092	.116	.142	.155	.118	.131	.150
		1.00	7.920	5.412	22.609	.871	.710	.733	.132	.150	.138	.134	.134	.135
		1.25	1.459	3.983	22.582	.790	.674	.613	.173	.335	.246	.163	.250	.221
	5	.75	18.050	42.321	256.167	.760	.915	.925	.098	.113	.110	.091	.107	.104
		1.00	22.585	22.485	131.863	.792	.648	.679	.106	.125	.108	.096	.114	.107
		1.25	23.109	12.890	74.154	.787	.584	.553	.134	.175	.125	.116	.137	.127
	10	.75	23.957	52.803	463.427	.719	.843	.878	.110	.101	.108	.093	.096	.097
		1.00	24.049	33.106	264.162	.763	.687	.733	.120	.104	.111	.103	.098	.112
		1.25	24.634	17.350	143.743	.785	.554	.574	.130	.113	.114	.106	.103	.112
200	1	.75	3.346	4.464	29.425	.805	1.166	1.159	.069	.073	.098	.072	.074	.105
		1.00	4.630	3.179	7.951	.784	.526	.545	.080	.076	.085	.080	.078	.095
		1.25	6.150	3.236	11.982	.660	.679	.604	.109	.210	.202	.100	.167	.182
	5	.75	22.720	46.967	247.966	.699	.947	.939	.057	.071	.069	.056	.067	.072
		1.00	24.703	2.206	89.326	.726	.478	.506	.063	.063	.067	.064	.062	.071
		1.25	25.059	6.482	23.934	.640	.542	.494	.078	.072	.071	.072	.065	.067
	10	.75	32.564	61.136	504.761	.663	.824	.849	.068	.072	.064	.063	.070	.069
		1.00	31.829	3.686	216.754	.708	.540	.565	.067	.071	.064	.065	.071	.068
		1.25	29.824	11.964	67.398	.668	.457	.459	.068	.079	.112	.068	.074	.081
300	1	.75	6.088	4.398	22.002	.830	1.180	1.175	.060	.058	.072	.064	.059	.076
		1.00	3.014	1.568	6.763	.732	.429	.437	.059	.058	.064	.060	.061	.072
		1.25	4.340	.627	6.691	.543	.697	.637	.093	.210	.166	.096	.159	.155
	5	.75	23.760	47.913	221.193	.684	.966	.964	.050	.050	.053	.053	.049	.055
		1.00	25.405	16.504	55.158	.668	.367	.389	.051	.054	.069	.054	.056	.076
		1.25	25.261	3.625	3.626	.543	.546	.517	.056	.058	.069	.055	.052	.076
	10	.75	34.951	65.558	507.699	.633	.818	.840	.051	.058	.051	.049	.064	.059
		1.00	35.037	24.214	141.516	.663	.412	.439	.051	.059	.060	.054	.064	.069
		1.25	32.436	7.875	7.929	.592	.439	.415	.053	.052	.059	.052	.047	.067

Table 2.6: Root mean squared errors (RMSE) for the other parameter estimates of the fractional UC model with correlated innovations in subsection 2.6.2. The different columns indicate the parameter estimates via the CSS estimator (subscript CSS) and the QML estimator (subscript QML) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

n	ν_0	d_0	$\hat{\nu}_{CSS}$	$\hat{\nu}^{I(1)}_{CSS}$	$\hat{\nu}^{I(1)}_{QML}$	$\hat{\rho}_{CSS}$	$\hat{\rho}^{I(1)}_{CSS}$	$\hat{\rho}^{I(1)}_{QML}$	$\hat{b}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{CSS}}$	$\hat{b}^{I(1)}_{1_{QML}}$	$\hat{b}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{CSS}}$	$\hat{b}^{I(1)}_{2_{QML}}$
100	1	.75	1.169	2.212	16.654	.049	.958	.954	.021	-.010	-.035	-.026	.000	.025
		1.00	2.089	1.118	4.404	.173	.194	.270	.025	-.027	-.023	-.030	.014	.016
		1.25	2.563	.446	3.796	.084	-.590	-.437	.039	-.178	-.099	-.040	.139	.111
	5	.75	4.735	25.868	141.005	-.280	.484	.486	.001	-.024	-.033	-.005	.015	.025
		1.00	6.973	8.246	4.913	-.176	.075	.153	.011	-.028	-.023	-.011	.018	.017
		1.25	6.659	-.547	9.397	-.090	-.446	-.331	.027	-.055	-.033	-.022	.036	.032
	10	.75	3.357	36.145	261.668	-.395	.243	.207	-.000	-.025	-.034	-.001	.018	.023
		1.00	3.299	13.866	88.703	-.333	.036	.099	.013	-.023	-.024	-.012	.015	.019
		1.25	3.359	-2.624	2.075	-.193	-.338	-.227	.021	-.030	-.024	-.013	.018	.019
200	1	.75	.940	1.037	6.674	.113	1.147	1.131	.004	.008	-.020	-.009	-.007	.022
		1.00	.868	.553	.909	.127	.140	.193	.005	-.005	-.009	-.008	.003	.009
		1.25	1.126	.097	1.207	.033	-.657	-.539	.017	-.119	-.093	-.015	.103	.105
	5	.75	7.841	32.742	141.739	-.202	.709	.655	-.009	-.009	-.019	.003	.007	.018
		1.00	8.594	7.090	22.113	-.131	.085	.126	-.000	-.007	-.007	-.003	.006	.007
		1.25	7.944	-2.668	-1.668	-.125	-.499	-.431	.009	-.025	-.021	-.008	.014	.015
	10	.75	9.857	48.797	308.955	-.319	.378	.283	-.008	-.011	-.018	.001	.009	.013
		1.00	9.145	13.135	62.278	-.259	.044	.080	-.001	-.010	-.007	-.002	.009	.005
		1.25	8.655	-5.645	-.215	-.159	-.384	-.306	.005	-.017	-.015	-.005	.009	.006
300	1	.75	1.670	.847	4.253	.194	1.168	1.164	.008	.011	-.013	-.013	-.008	.014
		1.00	.563	.305	.552	.108	.107	.148	.003	-.003	-.004	-.005	.003	.005
		1.25	.561	-.139	.293	-.001	-.689	-.604	.017	-.130	-.092	-.019	.104	.096
	5	.75	8.062	34.267	122.729	-.157	.807	.774	-.008	-.001	-.010	.001	.001	.010
		1.00	8.101	4.787	1.071	-.151	.048	.093	-.002	-.004	-.003	-.000	.003	.004
		1.25	8.131	-3.324	-3.090	-.130	-.525	-.446	.005	-.021	-.020	-.003	.010	.014
	10	.75	12.034	55.702	319.616	-.268	.468	.392	-.010	-.008	-.011	.004	.007	.008
		1.00	11.528	8.952	3.993	-.236	.036	.067	-.002	-.007	-.004	-.001	.006	.002
		1.25	11.500	-7.103	-6.749	-.156	-.402	-.354	.005	-.012	-.012	-.004	.003	.004

Table 2.7: Median bias for the other parameter estimates of the fractional UC model with correlated innovations in subsection 2.6.2. The different columns indicate the parameter estimates via the CSS estimator (subscript CSS) and the QML estimator (subscript QML) for the fractional UC model and the $I(1)$-integrated UC model (superscript $I(1)$).

			Trend			Cycle		
n	ν_0	d_0	R^2_{CSS}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$	R^2_{CSS}	$R^{I(1)^2}_{CSS}$	$R^{I(1)^2}_{QML}$
100	1	.75	.299	.344	.389	.771	.733	.768
		1.00	.607	.687	.678	.679	.715	.662
		1.25	.828	.822	.856	.584	.547	.461
	5	.75	.138	.207	.129	.895	.909	.828
		1.00	.417	.521	.392	.841	.873	.782
		1.25	.723	.766	.702	.784	.789	.687
	10	.75	.103	.182	.090	.919	.941	.840
		1.00	.320	.438	.283	.873	.911	.815
		1.25	.648	.718	.609	.826	.858	.748
200	1	.75	.445	.446	.500	.782	.743	.786
		1.00	.786	.833	.803	.707	.761	.669
		1.25	.937	.912	.930	.651	.586	.434
	5	.75	.224	.270	.176	.903	.921	.833
		1.00	.599	.695	.549	.852	.900	.784
		1.25	.877	.897	.847	.816	.818	.689
	10	.75	.163	.239	.104	.928	.952	.840
		1.00	.512	.612	.424	.885	.927	.814
		1.25	.829	.863	.779	.853	.867	.744
300	1	.75	.523	.513	.568	.782	.741	.786
		1.00	.855	.889	.858	.720	.771	.670
		1.25	.970	.955	.962	.678	.587	.447
	5	.75	.300	.325	.214	.904	.921	.830
		1.00	.716	.786	.643	.860	.902	.778
		1.25	.935	.940	.904	.837	.816	.675
	10	.75	.212	.285	.123	.933	.955	.839
		1.00	.626	.722	.520	.890	.931	.810
		1.25	.903	.925	.856	.869	.876	.736

Table 2.8: Coefficient of determination from regressing true trend and cycle x_t and c_t on their respective estimates from the Kalman smoother for the correlated UC models.

$Q(y,\hat{\theta})$	$\hat{d}$	$\hat{\nu}$	$\hat{\nu}_2$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\widehat{\text{Corr}(\eta_t,\epsilon_t)}$
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 2.9: 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 $\widehat{\text{Corr}(\eta_t,\epsilon_t)} = \hat{\nu}_2/\sqrt{\hat{\nu}}$.

$\log L(\psi)$	$Q(y,\hat{\psi})$	$\hat{\nu}$	$\hat{\nu}_2$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$	$\hat{b}_4$	$\hat{b}_5$	$\widehat{\text{Corr}(\eta_t,\epsilon_t)}$
-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 2.10: 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.

2.A.2 Proof of theorem 2.4.1

Proof of theorem 2.4.1. Theorem 2.4.1 holds if the objective function (2.16) satisfies a uniform weak law of large numbers (UWLLN), i.e. there exists a function $g_t(y_{t:1}) \geq 0$ such that for all $\theta_1, \theta_2 \in \Theta$, it holds that $|v_t^2(\theta_1) - v_t^2(\theta_2)| \leq g_t(y_{t:1})||\theta_1 - \theta_2||$, and both, $v_t(\theta)$ and $g_t(y_{t:1})$ satisfy a WLLN (Wooldridge; 1994, thm. 4.2). Since $v_t^2(\theta)$ is continuously differentiable, a natural choice for $g_t(y_{t:1})$ is the supremum of the absolute gradient, as follows from the mean value expansion of $v_t^2(\theta)$ about θ, see Newey (1991, cor. 2.2) and Wooldridge (1994, eqn. 4.4).

However, as can be seen from (2.15), uniform convergence of the objective function fails around the point $d = d_0 - 1/2$: Since y_t is $I(d_0)$, the d-th differences $\Delta_+^d y_{t+1} = \xi_{t+1}(d)$ as well as $S_d y_{t:1} = \xi_{t:1}(d)$ are $I(d_0 - d)$, and thus asymptotically stationary whenever $d > d_0 - 1/2$, otherwise non-stationary. Subsequently, I will show that the pointwise probability limit of $Q(y, \theta)$ is given by

$$\operatorname{plim}_{n\to\infty} Q(y,\theta) = \operatorname{plim}_{n\to\infty} \tilde{Q}(y,\theta) = \begin{cases} \mathrm{E}(\tilde{v}_t^2(\theta)) & \text{for } d - d_0 > -1/2, \\ \infty & \text{else,} \end{cases} \tag{2.28}$$

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

$$\tilde{v}_t(\theta) = \tilde{\xi}_t(d) + \sum_{j=1}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d) = \sum_{j=0}^{\infty} \tau_j(\theta)\tilde{\xi}_{t-j}(d), \tag{2.29}$$

generated by the untruncated fractional differencing polynomial Δ^d and the untruncated polynomial $b(L, \varphi) = \sum_{j=0}^{\infty} b_j(\varphi)L^j$. $\tilde{\xi}_t(d) = \Delta^{d-d_0}\eta_t + \Delta^d c_t$ is the untruncated residual, while the $\tau_j(\theta)$ stem from the ∞-vector $(\tau_1(\theta), \tau_2(\theta), \cdots) = \nu(b_1(\varphi) - \pi_1(d), b_2(\varphi) - \pi_2(d), \cdots)(B'_{\varphi,\infty}B_{\varphi,\infty} + \nu S'_{d,\infty}S_{d,\infty})^{-1}S'_{d,\infty}$, and $\tau_0(\theta) = 1$ as before. Note that the dependence of the $\tau_j(\theta)$ on t is resolved in (2.29) by letting the dimension of the t-dimensional coefficient vector go to infinity. Hence, while the truncated forecast errors in (2.15) are non-ergodic, the untruncated errors (2.29) 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 (2.28), I adapt the proof strategy of Nielsen (2015) for CSS estimation of ARFIMA models: I partition the parameter space for d into three compact subsets $D_1 = D_1(\kappa_1) = D \cap \{d : d - d_0 \leq -1/2 - \kappa_1\}$, $D_2 = D_2(\kappa_2, \kappa_3) = D \cap \{d : -1/2 - \kappa_2 \leq d - d_0 \leq -1/2 + \kappa_3\}$, and $D_3 = D_3(\kappa_3) = D \cap \{d : -1/2 + \kappa_3 \leq d - d_0\}$, for some constants $0 < \kappa_1 < \kappa_2 < \kappa_3 < 1/2$ to be determined later. Note that $\cup_{i=1}^3 D_i = D$. Within D_1 and D_3 convergence is uniform,

while within the overlapping D_2, which covers both stationary and non-stationary forecast errors, convergence is non-uniform. Denote the partitioned parameter spaces for θ as $\Theta_j = D_j \times \Sigma_\nu \times \Phi$, $j = 1, 2, 3$. Non-uniform convergence of (2.28) is then asymptotically ruled out by showing that for a given constant $K > 0$ there always exists a fixed $\bar{\kappa} > 0$ such that

$$\Pr\left(\inf_{d \in D \setminus D_3(\bar{\kappa}), \nu \in \Sigma_\nu, \varphi \in \Phi} Q(y, \theta) > K\right) \to 1 \quad \text{as } n \to \infty, \tag{2.30}$$

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

$$\sup_{\theta \in D_3(\kappa_3) \times \Sigma_\nu \times \Phi} \left| Q(y, \theta) - \mathrm{E}(\tilde{v}_{t+1}^2(\theta)) \right| \xrightarrow{p} 0, \qquad \text{as } n \to \infty, \tag{2.31}$$

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 (2.30) and (2.31) are well established for the CSS estimator in the ARFIMA literature, see Hualde and Robinson (2011) and Nielsen (2015), showing them to carry over to the fractional UC model requires some additional effort. Even within $\theta \in \Theta_3(\kappa_3)$, the forecast errors in (2.14) are not ergodic for two reasons: First, since the lag polynomial generated by the truncated fractional differencing polynomial Δ_+^d includes more lags as t increases, $\xi_t(d) = \Delta_+^{d-d_0}\eta_t + \Delta_+^d c_t$ are not ergodic. Second, the $\tau_j(\theta, t)$ in (2.15) depend on t. Consequently, also within $\Theta_3(\kappa_3)$ a WLLN for stationary and ergodic processes does not immediately apply. I tackle these problems by showing the expected difference between (2.15) and (2.29) to be

$$\mathrm{E}\left[(\tilde{v}_{t+1}(\theta) - v_{t+1}(\theta))^2\right] \to 0, \qquad \text{as } t \to \infty, \tag{2.32}$$

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

$$Q(y, \theta) = \tilde{Q}(y, \theta) + o_p(1) \xrightarrow{p} \mathrm{E}(\tilde{v}_t^2(\theta)), \qquad \text{as } n \to \infty. \tag{2.33}$$

(2.33) can be generalized to uniform convergence by showing that a WLLN also holds for the supremum of the absolute gradient, which yields (2.31). From (2.30)

and (2.31), theorem 2.4.1 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 (2.31) and (2.33) 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 (2.32), I first show that

$$\begin{aligned}\tilde{v}_{t+1}(\theta) - v_{t+1}(\theta) &= \sum_{j=0}^{t} \tau_j(\theta,t)\left(\tilde{\xi}_{t+1-j}(d) - \xi_{t+1-j}(d)\right) \\ &\quad + \sum_{j=t+1}^{\infty} \tau_j(\theta)\tilde{\xi}_{t+1-j}(d) + \sum_{j=0}^{t}\left(\tau_j(\theta) - \tau_j(\theta,t)\right)\tilde{\xi}_{t+1-j}(d) \\ &= \sum_{j=0}^{\infty} \phi_{\eta,j}(\theta,t)\eta_{t+1-j} + \sum_{j=0}^{\infty} \phi_{\epsilon,j}(\theta,t)\epsilon_{t+1-j},\end{aligned} \tag{2.34}$$

where $\phi_{\eta,j}(\theta,t)$ is $O((1+\log(t+1))^2(t+1)^{\max(-d+d_0,-\zeta)-1})$ for $j \leq t$, and $O((1+\log j)^3 j^{\max(-d+d_0,-\zeta)-1})$ for $j > t$, whereas $\phi_{\epsilon,j}(\theta,t)$ is $O((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1})$ for $j \leq t$, and $O((1+\log j)^4 j^{\max(-d,-\zeta)-1})$ for $j > t$. This can be verified by considering the three different terms in (2.34) 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

$$\begin{aligned}&\sum_{j=0}^{t} \tau_j(\theta,t)\left(\tilde{\xi}_{t+1-j}(d) - \xi_{t+1-j}(d)\right) \\ &\qquad = \sum_{j=t+1}^{\infty} \phi_{1,\eta,j}(\theta,t)\eta_{t+1-j} + \sum_{j=t+1}^{\infty} \phi_{1,\epsilon,j}(\theta,t)\epsilon_{t+1-j},\end{aligned} \tag{2.35}$$

where the coefficients are $\phi_{1,\eta,j}(\theta,t) = \sum_{k=0}^{t} \tau_k(\theta,t)\pi_{j-k}(d-d_0)$, and $\phi_{1,\epsilon,j}(\theta,t) = \sum_{k=0}^{t} \tau_k(\theta,t)\sum_{l=0}^{j-t-1} a_l(\varphi_0)\pi_{j-k-l}(d)$. Using Johansen and Nielsen (2010, lemma B.4), who show $\sum_{k=1}^{j-1} k^{\max(-d,-\zeta)-1}(j-k)^{-d+d_0-1} \leq K(1+\log j)j^{\max(-d+d_0,-\zeta)-1}$ for some finite constant $K > 0$, together with assumption 2.3, (2.68), lemma 2.A.2, and $j > t$, the coefficients in (2.35) are $\phi_{1,\eta,t} = O((1+\log j)^2 j^{\max(-d+d_0,-\zeta)-1})$, and $\phi_{1,\epsilon,t} = O((1+\log j)^3 j^{\max(-d,-\zeta)-1})$.

Next, consider the second term in (2.34)

$$\sum_{j=t+1}^{\infty} \tau_j(\theta)\tilde{\xi}_{t+1-j}(d) = \sum_{j=t+1}^{\infty} \eta_{t+1-j}\phi_{2,\eta,j}(\theta,t) + \sum_{j=t+1}^{\infty} \epsilon_{t+1-j}\phi_{2,\epsilon,j}(\theta,t), \tag{2.36}$$

with coefficients $\phi_{2,\epsilon,j}(\theta,t) = \sum_{k=0}^{j-t-1} \tau_{t+1+k}(\theta)\sum_{l=0}^{j-t-1-k} a_l(\varphi_0)\pi_{j-t-1-k-l}(d) =$

$O((1+\log j)^3 j^{\max(-d,-\zeta)-1})$, and $\phi_{2,\eta,j}(\theta,t) = \sum_{k=0}^{j-t-1} \pi_k(d-d_0)\tau_{j-k}(\theta) = O((1+\log j)^2 j^{\max(-d+d_0,-\zeta)-1})$ by assumption 2.3, lemma 2.A.1 and lemma 2.A.2.

For the third term in (2.34), by lemma 2.A.3

$$\begin{aligned}
&\sum_{j=0}^{t} (\tau_j(\theta) - \tau_j(\theta,t))\, \tilde{\xi}_{t+1-j}(d) \\
&= -\sum_{j=0}^{\infty} \eta_{t+1-j} \sum_{k=0}^{\min(j,t)} \pi_{j-k}(d-d_0) \sum_{m=t+1}^{\infty} r_{\tau,k,m}(\theta) \\
&\quad - \sum_{j=0}^{\infty} \epsilon_{t+1-j} \sum_{k=0}^{\min(j,t)} \left(\sum_{m=t+1}^{\infty} r_{\tau,k,m}(\theta) \right) \sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d) \\
&= \sum_{j=0}^{\infty} \phi_{3,\eta,j}(\theta,t)\eta_{t+1-j} + \sum_{j=0}^{\infty} \phi_{3,\epsilon,j}(\theta,t)\epsilon_{t+1-j}.
\end{aligned} \tag{2.37}$$

By lemma 2.A.3, $\sum_{m=t+1}^{\infty} r_{\tau,k,m}(\theta) = O((1+\log(t+1))^2 (t+1)^{\max(-d,-\zeta)-1})$, while $\pi_j(d-d_0) = O(j^{-d+d_0-1})$ and $\sum_{l=0}^{j-k} a_l(\varphi_0)\pi_{j-k-l}(d) = O((1+\log(j-k))(j-k)^{\max(-d,-\zeta)-1})$, see lemma 2.A.1 together with Johansen and Nielsen (2010, lemma B.4). Thus, it holds that $\phi_{3,\eta,j}(\theta,t) = -\sum_{k=0}^{\min(j,t)} \left(\sum_{m=t+1}^{\infty} r_{\tau,k,m}(\theta)\right) \pi_{j-k}(d-d_0)$ is $O\left((1+\log(t+1))^2 (t+1)^{\max(-d+d_0,-\zeta)-1}\right)$ for $j \leq t$, whereas for $j > t$ it is $O\left((1+\log j)^3 j^{\max(-d+d_0,-\zeta)-1}\right)$ since $d - d_0 > 1/2$ for all $\theta \in \Theta_3(\kappa_3)$. 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 $j \leq t$, and $O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right)$ for $j > t$. Together, (2.35), (2.36), (2.37) and the rates established below prove (2.34).

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

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

where the first equality follows by assumption 2.1, while the second follows from the convergence rates of $\phi_{\eta,j}(\theta,t)$, $\phi_{\epsilon,j}(\theta,t)$ as derived above, and the third equality follows from $\zeta > 0$ and $d - d_0 + 1/2 > \kappa_3 > 0$ for all $\theta \in \Theta_3(\kappa_3)$. (2.32) follows

directly. From the law of large numbers for stationary and ergodic processes, (2.33) follows immediately.

(2.33) 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 (2.31) holds, so that the objective function satisfies a UWLLN within the stationary region of the parameter space $\Theta_3(\kappa_3)$. The gradient of the objective function is given by

$$\begin{aligned} \frac{\partial Q(y,\theta)}{\partial \theta_{(l)}} &= \frac{2}{n}\sum_{t=1}^{n} v_t(\theta)\frac{\partial v_t(\theta)}{\partial \theta_{(l)}}, \\ \frac{\partial v_t(\theta)}{\partial \theta_{(l)}} &= \sum_{j=1}^{t-1}\frac{\partial \tau_j(\theta,t)}{\partial \theta_{(l)}}\xi_{t-j}(d) + \sum_{j=0}^{t-1}\tau_j(\theta,t)\frac{\partial \xi_{t-j}(d)}{\partial \theta_{(l)}}, \end{aligned} \tag{2.38}$$

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

$$\begin{aligned} &\sup_{(d_1,d_2,\nu,\varphi)\in\tilde{\Theta}}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\frac{\partial^k \Delta_+^{d_1}}{\partial d_1^k}\sum_{m=0}^{\infty}\tilde{\tau}_{i,m}(\theta)z_{i,t-m}(\theta)\right]\left[\frac{\partial^l \Delta_+^{d_2}}{\partial d_2^l}\sum_{m=0}^{\infty}\tilde{\tau}_{j,m}(\theta)z_{j,t-m}(\theta)\right]\right| \\ &\qquad = \begin{cases} O_p(1) & \text{for } a > 0, \\ O_p((\log n)^{1+k+l}n^{-a}) & \text{for } a \leq 0, \end{cases} \end{aligned} \tag{2.39}$$

$i,j = 1,2$, $k,l = 1,2,...$, as shown by Nielsen (2015, lemma B.3). Now, note that by lemmas 2.A.2 and 2.A.4 both the coefficients $\tau_j(\theta,t)$ and their partial derivatives satisfy the absolute summability condition, i.e. $\sum_{j=0}^{t-1}|\tau_j(\theta,t)| < \infty$ and $\sum_{j=0}^{t-1}|\partial\tau_j(\theta,t)/\partial\theta_{(l)}| < \infty$ for all $\theta_{(l)}$ and uniformly in $\theta \in \Theta$. In addition, by assumption 2.3, 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 (2.38) satisfy the conditions for (2.39) with $a > 0$. Thus, by (2.39), 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, (2.33) holds uniformly in $\theta\in\Theta_3(\kappa_3)$. As this holds for any κ_3, this proves (2.31).

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 (2.16), together with (2.15), it holds that

$$\begin{aligned} Q(y,\theta) =& \frac{1}{n}\sum_{t=1}^{n}\left[\sum_{j=0}^{t-1}\tau_j(\theta,t)\xi_{t-j}(d)\right]^2 \geq \frac{1}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right)^2 \\ &+ \frac{2}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right)\left(\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\right), \end{aligned} \tag{2.40}$$

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 (2.40), by lemma 2.A.2 $\sum_{j=0}^{t-1}|\tau_j(\theta,t)| < \infty$, and by assumption 2.3 and lemma 2.A.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 (2.39)

$$\sup_{\theta\in\Theta_2(\kappa_2,\kappa_3)}\left|\frac{1}{n}\sum_{t=1}^{n}\left[\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right]\left[\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\right]\right| = O_p(1). \tag{2.41}$$

Next, consider the first term in (2.40), for which one has by lemma 2.A.3

$$\begin{aligned} \Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j} &= \Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta)\eta_{t-j} + \Delta_+^{d-d_0}\sum_{j=1}^{t-1}\left(\sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta)\right)\eta_{t-j} \\ &= \Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j} + r_{\eta,t}(\theta), \end{aligned} \tag{2.42}$$

where

$$\begin{aligned} r_{\eta,t}(\theta) &= -\Delta_+^{d-d_0}\sum_{j=t}^{\infty}\tau_j(\theta)\eta_{t-j} + \Delta_+^{d-d_0}\sum_{j=1}^{t-1}\eta_{t-j}\sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta) \\ &= \Delta_+^{d-d_0}\sum_{j=1}^{\infty}\alpha_j\eta_{t-j}, \end{aligned} \tag{2.43}$$

and $\alpha_j = \sum_{i=t+1}^{\infty}r_{\tau,j,i}(\theta)$ for $j<t$ and $\alpha_j = -\tau_j(\theta)$ for $j\geq t$. By lemmas 2.69 and 2.A.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 $r_{\eta,t}(\theta)$

$$r_{\eta,t}(\theta) = \Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j + \Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}, \qquad \alpha_j^* = -\sum_{i=j+1}^{\infty}\alpha_i, \tag{2.44}$$

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 (2.42)

$$\Delta_+^{d-d_0}\sum_{j=0}^{\infty}\tau_j(\theta)\eta_{t-j} = \Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta) + \Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}, \tag{2.45}$$

where $\tau_j^*(\theta) = -\sum_{i=j+1}^{\infty}\tau_i(\theta)$, and $\sum_{j=0}^{\infty}\tau_j(\theta) = O(1)$ by lemma 2.69. By (2.42), (2.44), and (2.45), it follows for the first term in (2.40) that

$$\frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\Big)^2 \geq \frac{1}{n}\sum_{t=1}^{n}\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)^2 \tag{2.46}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)\Big(\Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j\Big)\right] \tag{2.47}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}\Big)\right] \tag{2.48}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\Big)\right] \tag{2.49}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}\Big)\Big(\Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j\Big)\right] \tag{2.50}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0+1}\sum_{j=0}^{\infty}\tau_j^*(\theta)\eta_{t-j}\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\Big)\right] \tag{2.51}$$

$$+\frac{2}{n}\sum_{t=1}^{n}\left[\Big(\Delta_+^{d-d_0}\eta_{t-1}\sum_{j=1}^{\infty}\alpha_j\Big)\Big(\Delta_+^{d-d_0+1}\sum_{j=1}^{\infty}\alpha_j^*\eta_{t-j}\Big)\right]. \tag{2.52}$$

From (2.39), it immediately follows that (2.48) to (2.52) 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 2.3, it follows that (2.46) asymptotically dominates (2.47), so that the rate of convergence of (2.40) will depend solely on (2.46). The asymptotic probability limit of the first

term (2.46) 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 (2.46)

$$\frac{1}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\eta_t \sum_{j=0}^{\infty}\tau_j(\theta)\right)^2 \geq \frac{1}{n}\sum_{t=N+1}^{n}\left(w_t^2 + 2w_t u_t\right). \tag{2.53}$$

As shown by Nielsen (2015, p. 164), for some κ satisfying $\max(\kappa_2, \kappa_3) \leq \kappa < 1/2$, setting $N = n^\alpha$ with $0 < \alpha < \min\left(\frac{1/2-\kappa}{1/2+\kappa}, \frac{1/2}{1/2+2\kappa}\right)$, it holds by Nielsen (2015, eqn. B.4 in lemma B.2) that $n^{-1}\sum_{t=n^\alpha+1}^{n} w_t u_t \xrightarrow{p} 0$ 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 (2.53) satisfies

$$\sup_{\theta\in\Theta_2(\kappa,\kappa)}\left|\frac{1}{n}\sum_{t=n^\alpha+1}^{n} w_t^2 - \sigma_{\eta,0}^2\Big(\sum_{j=0}^{\infty}\tau_j(\theta)\Big)^2 \sum_{j=0}^{n^\alpha-1}\pi_j^2(d-d_0)\right| \xrightarrow{p} 0, \tag{2.54}$$

as $n \to \infty$, and by Nielsen (2015, lemma A.3) the latter sum is bounded from below by $\sum_{j=0}^{n^\alpha-1}\pi_j^2(d-d_0) \geq 1 + K\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}$ for some $K > 0$. The limit of the fraction $\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3}$ is discussed by Nielsen (2015, p. 165): It increases in n from zero (for $n = 2$) to $1/(2\kappa_3)$ as $n \to \infty$, and decreases in κ_3 from $\alpha\log(n-1)$ for $\kappa_3 = 0$ to zero for $\kappa_3 \to 1/2$. Consequently $\frac{1-(n-1)^{-2\alpha\kappa_3}}{2\kappa_3} \to \infty$ as $(n,\kappa_3) \to (\infty, 0)$. This, together with (2.46), (2.53), and (2.54) 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 (2.40), (2.41), and (2.42) 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

$$\Pr\left(\inf_{d\in D_2(\kappa_2,\bar{\kappa}_3),\nu\in\Sigma_\nu,\varphi\in\Phi} Q(y,\theta) > K\right) \geq 1-\delta, \quad \text{for all } T \geq T_2, \tag{2.55}$$

and (2.55) holds for any $\kappa_2 \in (0, 1/2)$.

Convergence on $\Theta_1(\kappa_1)$ Finally, consider the non-stationary subset $\Theta_1(\kappa_1) = D_1(\kappa_1) \times \Sigma_\nu \times \Phi$. Starting again with (2.40) above, the second term in (2.40), by the same argument with respect to absolute summability of the coefficients as for (2.41), is now

$$\frac{1}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\sum_{j=0}^{t-1}\tau_j(\theta,t)\eta_{t-j}\right)\left(\Delta_+^{d}\sum_{j=0}^{t-1}\tau_j(\theta,t)c_{t-j}\right) = O_p\left(1+\log(n)n^{d_0-d-1}\right), \tag{2.56}$$

for all $\theta \in \Theta_1(\kappa_1)$ by (2.39) 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 (2.40) 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 (2.40) into $\Delta_+^{d-d_0} \sum_{j=0}^{\infty} \tau_j(\theta) \eta_{t-j}$ and $r_{\eta,t}(\theta)$ in (2.42) and (2.43) above also applies in $\Theta_1(\kappa_1)$. Consequently, the Beveridge-Nelson decompositions in (2.44) and (2.45) also hold for $\theta \in \Theta_1(\kappa_1)$. Again, the decomposition in (2.46) to (2.52) applies, however the terms in (2.48) to (2.52) will not necessarily be $O_p(1)$, since $d - d_0$ is no longer bounded from above by -1 or by -2. However, as will become clear, the first term (2.46) asymptotically dominates all other terms in (2.47) to (2.52) 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 (2.46). It follows from the Cauchy-Schwarz inequality that

$$n^{2(d-d_0)} \sum_{t=1}^{n} \left(\Delta_+^{d-d_0} \eta_t \sum_{j=0}^{\infty} \tau_j(\theta) \right)^2 \geq \left(n^{d-d_0-1/2} \sum_{t=1}^{n} \Delta_+^{d-d_0} \eta_t \sum_{j=0}^{\infty} \tau_j(\theta) \right)^2, \quad (2.57)$$

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

$$r_n(\theta) = n^{d-d_0-1/2} \sum_{t=1}^{n} \Delta_+^{d-d_0} \eta_t \sum_{j=0}^{\infty} \tau_j(\theta) = n^{d-d_0-1/2} \Delta_+^{d-d_0-1} \eta_n \sum_{j=0}^{\infty} \tau_j(\theta) \quad (2.58)$$

follows if assumptions A(i) to A(iv) of Hosoya (2005) hold. Since $0 < \sum_{j=0}^{\infty} |\tau_j(\theta)| < \infty$ and $\mathrm{E}(\eta_j | \mathcal{F}_t) = 0$ for all $j > t$, as well as $\mathrm{E}(\eta_j \eta_k | \mathcal{F}_t) - \mathrm{E}(\eta_j \eta_k) = 0$ for $j, k > t$ by assumption 2.1, 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 2.1. 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 $n \to \infty$

$$n^{d-d_0-1/2}\Delta_+^{d-d_0-1}\eta_{\lfloor nr \rfloor}\sum_{j=0}^{\infty}\tau_j(\theta) \Rightarrow W_{d_0-d}(r) \quad \text{in } \mathcal{D}[0,1], \tag{2.59}$$

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 $\eta_t \sum_{j=0}^{\infty}\tau_j(\theta)$. (2.59) is equivalent to Nielsen (2015, eqn. 30) for the univariate case. From (2.59) it follows that $r_n(\theta) \xrightarrow{d} r(\theta) = W_{d_0-d}(1)$ for fixed $d \in D_1(\kappa_1)$. Pointwise convergence $r_n(\theta)$ can be generalized to uniform convergence in $D_1(\kappa_1)$ if $r_n(\theta)$ is tight (stochastically equicontinuous) as a function of θ on $\theta \in \Theta_1(\kappa_1)$. Since the parameters φ, ν only enter $r_n(\theta)$ through $\sum_{j=0}^{\infty}\tau_j(\theta)$, it is sufficient for tightness of $r_n(\theta)$ in θ that $n^{d-d_0-1/2}\Delta_+^{d-d_0-1}\eta_n$ is tight in $(d-d_0)$. As in Nielsen (2015, pp. 169f), tightness in $(d-d_0)$ can be shown using the moment condition in Billingsley (1968, thm. 12.3) which requires to show that $r_n(\theta)$ is tight for a fixed $d-d_0$ and that $|n^{d_1-1/2}\Delta_+^{d_1-1}\eta_n - n^{d_2-1/2}\Delta_+^{d_2-1}\eta_n| \leq K|d_1-d_2|$ 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 (2.40), for which a lower bound is given by the expressions (2.46) to (2.52), note that by (2.57) the first term (2.46) is bounded from below (when scaled appropriately) by

$$\inf_{\theta\in\Theta_1(\kappa_1)} \frac{1}{n}\sum_{t=1}^{n}\left(\Delta_+^{d-d_0}\eta_t\sum_{j=0}^{\infty}\tau_j(\theta)\right)^2 \geq n^{2(d_0-d-1/2)}\inf_{\theta\in\Theta_1(\kappa_1)} r_n(\theta)^2. \tag{2.60}$$

The probability limits of (2.48) to (2.52) can be derived by (2.39) 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 (2.48) to (2.52) 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 (2.46). For (2.47), note that $\sum_{j=1}^{\infty}\alpha_j = O((1+\log t)^2 t^{\max(-d,-\zeta)})$, while $\sum_{j=0}^{\infty}\tau_j(\theta)$ is bounded away from zero by assumption 2.3. Consequently, (2.47) will also diverge at a slower rate than (2.46). Finally, as already shown in (2.56), the second term in (2.40) is $O_p\left(1+\log(n)n^{d_0-d-1}\right)$ and thus is also dominated by (2.46). It follows that the rate of divergence of the objective function is determined by the first term in (2.40) and

is given by the divergence rate of (2.46). This, together with (2.60), yields

$$\inf_{\theta\in\Theta_1(\kappa_1)} Q(y,\theta) \geq n^{2(d_0-d-1/2)} \inf_{\theta\in\Theta_1(\kappa_1)} r_n(\theta)^2 \geq n^{2\kappa_1} \inf_{\theta\in\Theta_1(\kappa_1)} r_n(\theta)^2 \tag{2.61}$$

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

$$\Pr\left(\inf_{d\in D_1(\kappa_1),\nu\in\Sigma_\nu,\varphi\in\Phi} \frac{1}{n} Q(y,\theta) > K\right) \to 1, \quad \text{as } T \to \infty. \tag{2.62}$$

Together, (2.55) and (2.62) prove (2.30). □

2.A.3 Proof of theorem 2.4.2

Proof of theorem 2.4.2. Since $\hat{\theta}$ is consistent, see theorem 2.4.1, the asymptotic distribution theory can be derived based on the Taylor series expansion of the score function as usual

$$0 = \sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\hat{\theta}} = \sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} + \sqrt{n}\frac{\partial^2 Q(y,\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\bar{\theta}} \left(\hat{\theta}-\theta_0\right), \tag{2.63}$$

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

$$\sqrt{n}\frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} = \frac{2}{\sqrt{n}}\sum_{t=1}^{n} v_t(\theta_0)\frac{\partial v_t(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}, \tag{2.64}$$

with $v_t(\theta)$ denoting the prediction error as defined in (2.14) and (2.15), and its partial derivative as given in (2.38). Denote the normalized, untruncated score

$$\sqrt{n}\frac{\partial \tilde{Q}(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} = \frac{2}{\sqrt{n}}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)\frac{\partial \tilde{v}_t(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}, \tag{2.65}$$

with $\tilde{v}_t(\theta)$ as defined in (2.29). As shown in lemma 2.A.6, the difference between truncated and untruncated score is asymptotically negligible. Therefore it is sufficient to consider the distribution of the latter. By assumption 2.5, the untruncated prediction error $\tilde{v}_t(\theta_0)$ is a stationary MDS when adapted to $\mathcal{F}_t^{\tilde{\xi}} = \sigma(\tilde{\xi}_s, s \leq t)$. Thus, for (2.65) a central limit theorem can be shown to apply following Nielsen (2015, p. 175): By the Cramér-Wold device it is sufficient to show that for any $q+2$-dimensional vector μ, $\mu'\sqrt{n}\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}\big|_{\theta=\theta_0} = \sqrt{n}\sum_{i=1}^{q+2}\mu_{(i)}\left(\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}\big|_{\theta=\theta_0}\right)_{(i)} =$

$\frac{2}{\sqrt{n}}\sum_{i=1}^{q+2}\mu_{(i)}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{1,t}+\tilde{h}_{2,t})_{(i)} \xrightarrow{d} \mathrm{N}(0,4\sigma_{v,0}^2\mu'\Omega_0\mu)$ as $n\to\infty$, with $\tilde{h}_{1,t}=\sum_{j=1}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$, as well as $\tilde{h}_{2,t}=\sum_{j=0}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$. As $\tilde{h}_{1,t}$ and $\tilde{h}_{2,t}$ are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, $\nu_t=\sum_{i=1}^{q+2}\mu_{(i)}\tilde{v}_t(\theta_0)(\tilde{h}_{1,t}+\tilde{h}_{2,t})_{(i)}$ together with $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS. Thus, by the law of large numbers for stationary and ergodic processes, it holds that

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

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

Next, consider the second derivatives in (2.63). By Johansen and Nielsen (2010, lemma A.3), the Hessian matrix in (2.63) 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 2.4.1, $\nu\in\Sigma_\nu$ as defined in section 2.4, and $\varphi\in N_\delta(\varphi_0)$ as defined in assumptions 2.2 and 2.4, by a random variable $B_n=O_p(1)$, see Newey (1991, cor. 2.2). This holds by lemma 2.A.7. Therefore, the second derivative in (2.63) can be evaluated at the true value θ_0

$$\begin{aligned}\frac{\partial^2 Q(y,\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\Big|_{\theta=\theta_0}=&\frac{2}{n}\sum_{t=1}^{n}\frac{\partial v_t(\theta)}{\partial\theta_{(k)}}\Big|_{\theta=\theta_0}\frac{\partial v_t(\theta)}{\partial\theta_{(l)}}\Big|_{\theta=\theta_0}\\&+\frac{2}{n}\sum_{t=1}^{n}v_t(\theta_0)\frac{\partial^2 v_t(\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\Big|_{\theta=\theta_0},\end{aligned}\tag{2.66}$$

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

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

From the law of large numbers for stationary and ergodic processes, it then holds for the first term in (2.66) 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 2.A.9 the second term in (2.66) is $\frac{2}{n}\sum_{t=1}^{n}v_t(\theta_0)\frac{\partial^2 v_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}=\frac{2}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\frac{\partial^2\tilde{v}_t(\theta)}{\partial\theta\partial\theta'}\Big|_{\theta=\theta_0}+o_p(1)$. As $(\tilde{v}_t(\theta_0),\mathcal{F}_t^{\tilde{\xi}})$ is a stationary MDS, while the sec-

ond partial derivatives are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, it holds that $\frac{2}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\frac{\partial^2\tilde{v}_t(\theta)}{\partial\theta\partial\theta'}\big|_{\theta=\theta_0} = o_p(1)$. Taken together, this implies for (2.66) that

$$\left.\frac{\partial^2 Q(y,\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\right|_{\theta=\theta_0} = \frac{2}{n}\sum_{t=1}^{n}\left.\frac{\partial\tilde{v}_t(\theta)}{\partial\theta_{(k)}}\right|_{\theta=\theta_0}\left.\frac{\partial\tilde{v}_t(\theta)}{\partial\theta_{(l)}}\right|_{\theta=\theta_0} + o_p(1). \tag{2.67}$$

Finally, from the law of large numbers, it follows that $\frac{\partial^2 Q(y,\theta)}{\partial\theta_{(k)}\partial\theta_{(l)}}\big|_{\theta=\theta_0} \xrightarrow{p} 2\Omega_{0_{(k,l)}}$. Thus, solving (2.63) for $\sqrt{n}(\hat{\theta}-\theta_0)$ yields the desired result

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

□

2.A.4 Additional lemmas

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

Lemma 2.A.1 (Convergence rates of $\pi_j(d)$, $b_j(\varphi)$, and related vector and matrix entries)**.** *It holds that*

$$\pi_j(d) = O(j^{-d-1}), \tag{2.68}$$

$$b_j(\varphi) = O(j^{-\zeta-1}), \tag{2.69}$$

$$(B'_{\varphi,t}B_{\varphi,t})_{(i,j)} = \begin{cases} O(|i-j|^{-\zeta-1}) & \text{for } i\neq j, \\ O(1) & \text{for } i=j, \end{cases} \tag{2.70}$$

$$(S'_{d,t}S_{d,t})_{(i,j)} = \begin{cases} O(|i-j|^{-d-1}) & \text{for } i\neq j, \\ O(1) & \text{for } i=j, \end{cases} \tag{2.71}$$

$$(B'_{\varphi,t}B_{\varphi,t})^{-1}_{(i,j)} = \begin{cases} O(|i-j|^{-\zeta-1}) & \text{for } i\neq j, \\ O(1) & \text{for } i=j, \end{cases} \tag{2.72}$$

$$(B_{\varphi,t}B_{\varphi,t}+\nu S'_{d,t}S_{d,t})^{-1}_{(i,j)} = \begin{cases} O(|i-j|^{\max(-d,-\zeta)-1}) & \text{for } i\neq j, \\ O(1) & \text{for } i=j, \end{cases} \tag{2.73}$$

$$(B'_{\varphi,t}\beta_t)_{(j)} = O((t-j+1)^{-\zeta-1}), \tag{2.74}$$

$$(S'_{d,t}s_t)_{(j)} = O((t-j+1)^{-d-1}), \tag{2.75}$$

with $\pi_j(d)$ as defined in (2.3)*, $b_j(\varphi)$ as defined below assumption 2.3, $B_{\varphi,t}$ and $S_{d,t}$*

as defined in (2.5)*, and* $\beta_t' = (b_t(\varphi)\cdots b_1(\varphi))$, $s_t' = (\pi_t(d)\cdots\pi_1(d))$.

Proof of Lemma 2.A.1. (2.68) follows by Johansen and Nielsen (2010, lemma B.3) while (2.69) follows by assumption 2.3. (2.70) follows from (2.69) 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 (2.71) is analogous and follows from (2.68), as $(S_{d,t}'S_{d,t})_{(i,j)} = \sum_{k=0}^{\min(i,j)-1} \pi_k(d)\pi_{k+|i-j|}(d) = O(|i-j|^{-d-1})$ for $i \neq j$, $(S_{d,t}'S_{d,t})_{(i,i)} = O(1)$.
To derive the convergence rates for the entries of $(B_{\varphi,t}'B_{\varphi,t})^{-1}$ and $(B_{\varphi,t}'B_{\varphi,t} + \nu S_{d,t}'S_{d,t})^{-1}$ in (2.72) and (2.73), 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[5] $T_t(f_1)$ and $T_t(f_2)$ with symbols $f_1(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_1(j)e^{i\lambda j}$, $\gamma_1(j) = \sum_{k=0}^{\infty} b_k(\varphi)b_{k+j}(\varphi)$, $f_2(\lambda) = (2\pi)^{-1}\sum_{j=0}^{\infty}\gamma_2(j)e^{i\lambda j}$, $\gamma_2(j) = \sum_{k=0}^{\infty}[b_k(\varphi)b_{k+j}(\varphi) + \nu\pi_k(d)\pi_{k+j}(d)]$, where $\gamma_1(j) = O(j^{-\zeta-1})$ and $\gamma_2(j) = O(j^{\max(-d,-\zeta)-1})$ as $j \to \infty$. Consequently, $(B_{\varphi,t}'B_{\varphi,t})^{-1}$ and $(B_{\varphi,t}'B_{\varphi,t}+\nu S_{d,t}'S_{d,t})^{-1}$ converge to the Toeplitz matrices $T_t(1/f_1)$ and $T_t(1/f_2)$ that exist by assumption 2.3. 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 2.3. 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 (2.72) and (2.73).
(2.74) follows immediately from (2.69), 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 (2.75) follows immediately from (2.68) 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 2.A.2 (Convergence rates of $\tau_j(\theta,t)$)**.** *For the coefficients* $\tau_j(\theta,t)$ *as defined in* (2.15) *and below, it holds that*

$$\tau_j(\theta,t) = O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right). \tag{2.76}$$

Proof of Lemma 2.A.2. To prove (2.76), consider $\tau_j(\theta,t)$ as defined in (2.15) and

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

below

$$\tau_j(\theta,t) = \nu \sum_{k=1}^{t} \Big[\Big(b_1(\varphi) - \pi_1(d) \quad \cdots \quad b_t(\varphi) - \pi_t(d) \Big) \times (B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1} \Big]_{(k)} S_{d,t_{(j,k)}}. \tag{2.77}$$

The left term in (2.77) is

$$\begin{aligned} &\Big[\Big(b_1(\varphi) - \pi_1(d) \quad \cdots \quad b_t(\varphi) - \pi_t(d) \Big) (B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1} \Big]_{(k)} \\ &= (b_k(\varphi) - \pi_k(d)) (B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1}_{(k,k)} \\ &+ \sum_{i=1}^{k-1} (b_i(\varphi) - \pi_i(d)) (B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1}_{(i,k)} \\ &+ \sum_{i=k+1}^{t} (b_i(\varphi) - \pi_i(d)) (B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1}_{(i,k)}. \end{aligned} \tag{2.78}$$

Note that $\pi_k(d) = O(k^{-d-1})$, $b_k(\varphi) = O(k^{-\zeta-1})$, $(B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1}_{(k,k)} = O(1)$, and $(B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1}_{(i,k)} = O(|i-k|^{\max(-d,-\zeta)-1})$ for $i \neq k$ by (2.68), (2.69), and (2.73). Thus, the first term in (2.78) is $O\left(k^{\max(-d,-\zeta)-1}\right)$, while the second term is $\sum_{i=1}^{k-1} O\left(i^{\max(-d,-\zeta)-1}(k-i)^{\max(-d,-\zeta)-1}\right) = O\left((1+\log k) k^{\max(-d,-\zeta)-1}\right)$, where the last equality follows from Johansen and Nielsen (2010, lemma B.4), who show that $\sum_{i=1}^{k-1} i^{\max(-d,-\zeta)-1}(k-i)^{\max(-d,-\zeta)-1} = O((1+\log k) k^{\max(-d,-\zeta)-1})$. Again using (2.69) and (2.73), the third term in (2.78) can be shown to be bounded by $\sum_{i=k+1}^{t} O\left(i^{\max(-d,-\zeta)-1}(i-k)^{\max(-d,-\zeta)-1}\right) = O((k+1)^{\max(-d,-\zeta)-1} \sum_{i=k+1}^{t} (i-k)^{\max(-d,-\zeta)-1}) = O((k+1)^{\max(-d,-\zeta)-1})$. Therefore

$$\begin{aligned} &\Big[\Big(b_1(\varphi) - \pi_1(d) \quad \cdots \quad b_t(\varphi) - \pi_t(d) \Big) (B'_{\varphi,t} B_{\varphi,t} + \nu S'_{d,t} S_{d,t})^{-1} \Big]_{(k)} \\ &= O\left((1+\log k) k^{\max(-d,-\zeta)-1} \right). \end{aligned} \tag{2.79}$$

By plugging (2.79) into (2.77) and using (2.5) together with (2.68), one obtains

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

$$= O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\Big) + O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\sum_{k=1}^{t-j}k^{-d-1}\Big)$$
$$= O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right), \tag{2.80}$$

since $\sum_{k=1}^{t-j} k^{-d-1} = O(1)$ for all $d > 0$. This proves (2.76).

□

Lemma 2.A.3 (Convergence of $\tau_j(\theta,t)$ as $t\to\infty$)**.** *For the coefficients $\tau_j(\theta,t)$ as defined in* (2.15) *and below, it holds that*

$$\tau_j(\theta,t) = \tau_j(\theta,t+1) + r_{\tau,j,t+1}(\theta), \tag{2.81}$$

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 2.A.3. To prove (2.81), I study the impact of an increase from t to $t+1$ on $\tau_j(\theta,t+1) = \nu[(b_1(\varphi)-\pi_1(d)\cdots b_{t+1}(\varphi)-\pi_{t+1}(d))(B'_{\varphi,t+1}B_{\varphi,t+1}+\nu S'_{d,t+1}S_{d,t+1})^{-1}S'_{d,t+1}]_{(j)}$. Denote

$$B_{\varphi,t+1} = \begin{bmatrix} B_{\varphi,t} & \beta_t \\ 0_{1\times t} & 1 \end{bmatrix}, \qquad S_{d,t+1} = \begin{bmatrix} S_{d,t} & s_t \\ 0_{1\times t} & 1 \end{bmatrix}, \tag{2.82}$$

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

$$\Xi_{t+1}(\theta) = \begin{bmatrix} \Xi_t(\theta)+R_1 & R_2 \\ R_2' & R_3 \end{bmatrix}, \tag{2.83}$$

with the block entries

$$R_3 = [(1+\beta_t'\beta_t+\nu+\nu s_t's_t) - (\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t)]^{-1},$$
$$R_2 = -R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t),$$
$$R_1 = R_3\Xi_t(\theta)(B'_{\varphi,t}\beta_t+\nu S'_{d,t}s_t)(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta).$$

Clearly $R_3 = O(1)$, since by (2.73), (2.74) and (2.75)

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

$$+ O((t+1-j)^{\max(-d,-\zeta)-1}) + O\Big(\sum_{i=1}^{t-j}(t+1-i-j)^{\max(-d,-\zeta)-1} i^{\max(-d,-\zeta)-1}\Big)$$
$$= O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right), \tag{2.84}$$

and again by (2.74) and (2.75)

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

which is $O(1)$. This, together with $1+\beta_t'\beta_t+\nu+\nu s_t' s_t = \sum_{j=0}^{t} b_j^2(\varphi)+\nu\sum_{j=0}^{t}\pi_j^2(d) = O(1)$, yields $R_3^{-1} = O(1)$. Furthermore, R_3^{-1} is bounded away from zero, as $\Xi_t(\theta)^{-1}$ is regular by assumption 2.3. Next, consider R_2, for which by (2.84) it follows that $R_{2_{(j)}} = O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$. Finally, for R_1, by (2.84) it follows that $R_{1_{(i,j)}} = O\big((1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big)$.

Next, consider the vector

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

where $R_4 = (b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2 + (b_{t+1}(\varphi)-\pi_{t+1}(d))R_3$. By (2.68) and (2.69), it holds for the terms in R_4 that $[b_{t+1}(\varphi)-\pi_{t+1}(d)]R_3 = O((t+1)^{\max(-d,-\zeta)-1})$, and $(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_2 = O\big(\sum_{j=1}^{t} j^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big) = O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Thus $R_4 = O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Analogously, for the other terms in the above vector, one has $[(b_{t+1}(\varphi)-\pi_{t+1}(d))R_2']_{(j)} = O\big((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\big)$, and $[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))R_1]_{(j)} = O\big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\sum_{i=1}^{t}(1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1} i^{\max(-d,-\zeta)-1}\big) = O\big((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}(1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}\big)$. Therefore, for $j=1,...,t$, the whole term $\tau_j(\theta,t+1)$ is

$$\begin{aligned}\tau_j(\theta,t+1) &= \nu\left((b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S_{d,t}' + R_5'\right)_{(j)} \\ &= \tau_j(\theta,t) + \nu R_{5_{(j)}},\end{aligned} \tag{2.85}$$

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

For R_5

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

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

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

Hence, $R_{5_{(j)}} = O\big((1+\log(t+1))^2(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^2(t+1-j)^{\max(-d,-\zeta)-1}\big)$. This completes the proof of (2.81).

□

Lemma 2.A.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* (2.15) *and below, it holds that*

$$\frac{\partial \tau_j(\theta,t)}{\partial d} = O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \tag{2.86}$$

$$\frac{\partial \tau_j(\theta,t)}{\partial \nu} = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \tag{2.87}$$

$$\frac{\partial \tau_j(\theta,t)}{\partial \varphi_{(l)}} = O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \tag{2.88}$$

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

Proof of Lemma 2.A.4. Denote $\dot{\pi}_j(d) = \partial \pi_j(d)/\partial d = O((1+\log j) j^{-d-1})$, see Johansen and Nielsen (2010, lemma B.3), and $\dot{b}_j(\varphi_{(l)}) = \partial b_j(\varphi)/\partial \varphi_{(l)} = O(j^{-\zeta-1})$ by

assumption 2.3. Furthermore, denote the partial derivatives of $S_{d,t}$ and $B_{\varphi,t}$ as

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

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

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

$$[\dot{S}'_{d,t}S_{d,t}]_{(i,j)} = \begin{cases} \sum_{k=1}^{i-1} \dot{\pi}_k(d)\pi_{k+j-i}(d) = O((1+j-i)^{-d-1}) & \text{if } i \leq j, \\ \sum_{k=0}^{j-1} \pi_k(d)\dot{\pi}_{k+i-j}(d) = O((1+\log(i-j))(i-j)^{-d-1}) & \text{if } i > j. \end{cases} \tag{2.89}$$

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

$$[\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}]_{(i,j)} = \begin{cases} \sum_{k=1}^{i-1} \dot{b}_k(\varphi_{(l)})b_{k+j-i}(\varphi) = O((1+j-i)^{-\zeta-1}) & \text{if } i \leq j, \\ \sum_{k=0}^{j-1} b_k(\varphi)\dot{b}_{k+i-j}(\varphi_{(l)}) = O((i-j)^{-\zeta-1}) & \text{if } i > j. \end{cases} \tag{2.90}$$

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

$$\begin{aligned} \frac{\partial \tau_j(\theta,t)}{\partial d} = &-\nu^2 \big[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d)) \\ &\times \Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)S'_{d,t}\big]_{(j)} \\ &+ \nu[(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d))\Xi_t(\theta)\dot{S}'_{d,t}]_{(j)} \\ &- \nu[(\dot{\pi}_1(d) \cdots \dot{\pi}_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)}. \end{aligned} \tag{2.91}$$

For the first term, note that by (2.89) $[\dot{S}'_{d,t}S_{d,t} + S'_{d,t}\dot{S}_{d,t}]_{(i,j)} = [\dot{S}'_{d,t}S_{d,t}]_{(i,j)} + [\dot{S}'_{d,t}S_{d,t}]_{(j,i)} = O((1+\log|i-j|)|i-j|^{-d-1})$ 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 (2.79) it follows for the first terms in (2.91) that

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

$$
\begin{aligned}
=&O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right)\\
&+O\Big(\sum_{i=1}^{j-1}(1+\log i)i^{\max(-d,-\zeta)-1}(1+\log(j-i))(j-i)^{-d-1}\Big)\\
&+O\Big(\sum_{i=j+1}^{t}(1+\log i)i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\Big)\\
=&O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \qquad (2.92)
\end{aligned}
$$

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 (2.91), by (2.73) and (2.92)

$$
\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)]_{(j)}\\
&\quad=O\Big((1+\log j)^3 j^{\max(-d,-\zeta)-1}\Big)\\
&\qquad+O\Big(\sum_{i=1}^{j-1}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\Big)\\
&\qquad+O\Big(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{\max(-d,-\zeta)-1}\Big)\\
&\qquad=O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \qquad (2.93)
\end{aligned}
$$

where the proof is the same as for (2.92) besides the additional log-factor. Adding the last term, it follows by (2.68) and (2.93) that

$$
\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)S'_{d,t}]_{(j)}\\
&=\sum_{i=j}^{t}[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{S}'_{d,t}S_{d,t}+S'_{d,t}\dot{S}_{d,t})\Xi_t(\theta)]_{(i)}\pi_{i-j}(d)\\
&=O\Big((1+\log j)^4 j^{\max(-d,-\zeta)-1}\Big)+O\Big(\sum_{i=j+1}^{t}(1+\log i)^4 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\Big)\\
&=O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right), \qquad (2.94)
\end{aligned}
$$

where the second equality uses $\pi_0(d)=1$ to obtain the first term, while the last equality uses $\sum_{i=1}^{t-j} i^{-d-1} = O(1)$, which holds for all $d>0$. Consequently, the first term in (2.91) is bounded by $O\left((1+\log j)^4 j^{\max(-d,-\zeta)-1}\right)$. Turning to the second

term in (2.91), by (2.79)

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)\dot{S}'_{d,t}]_{(j)}\\
&=\sum_{i=j+1}^{t}[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)]_{(i)}\dot{\pi}_{i-j}(d)\\
&=O\Big(\sum_{i=j+1}^{t}(1+\log i)i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\Big)\\
&=O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right), \qquad (2.95)
\end{aligned}$$

where the last equality follows from $\sum_{i=1}^{t-j}(1+\log i)i^{-d-1}=O(1)$ for all $d>0$. By an analogous proof, the third term in (2.91) is

$$\begin{aligned}
[(\dot{\pi}_1(d)\cdots\dot{\pi}_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)} &= \sum_{i=j}^{t}[(\dot{\pi}_1(d)\cdots\dot{\pi}_t(d))\Xi_t(\theta)]_{(i)}\pi_{i-j}(d) \qquad (2.96)\\
=&O\Big((1+\log j)^2 j^{\max(-d,-\zeta)-1}\Big)\\
&+O\Big(\sum_{i=j+1}^{t}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(1+\log(i-j))(i-j)^{-d-1}\Big)\\
=&\,O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right). \qquad (2.97)
\end{aligned}$$

Together, (2.94), (2.95), and (2.96) yield (2.86).

To prove (2.87), consider the partial derivatives $\partial\tau_j(\theta,t)/\partial\nu$, for which

$$\begin{aligned}
\frac{\partial\tau_j(\theta,t)}{\partial\nu} &= [(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}]_{(j)} \qquad (2.98)\\
&\quad-\nu[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)S'_{d,t}]_{(j)}. \qquad (2.99)
\end{aligned}$$

By (2.80) the first term (2.98) is $O\big((1+\log j)j^{\max(-d,-\zeta)-1}\big)$, while by (2.71) and (2.79), it holds for the second term (2.99) that

$$\begin{aligned}
[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}]_{(j)} =& O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\Big)\\
&+O\Big(\sum_{i=1}^{j-1}(1+\log i)i^{\max(-d,-\zeta)-1}(j-i)^{-d-1}\Big)\\
&+O\Big(\sum_{i=j+1}^{t}(1+\log i)i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\Big)\\
=&O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right), \qquad (2.100)
\end{aligned}$$

and the proof is analogous to (2.92) besides one log-factor. Furthermore, by (2.73) and (2.100)

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)]_{(j)}\\
&\quad=O\Big((1+\log j)^2 j^{\max(-d,-\zeta)-1}\Big)\\
&\qquad+O\Big(\sum_{i=1}^{j-1}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\Big)\\
&\qquad+O\Big(\sum_{i=j+1}^{t}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(i-j)^{\max(-d,-\zeta)-1}\Big)\\
&\quad=O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \qquad (2.101)
\end{aligned}$$

where again the proof is analogous to (2.93) besides one log-factor. From (2.68) and (2.101) it then follows for (2.99) that

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)S'_{d,t}S_{d,t}\Xi_t(\theta)S'_{d,t}]_{(j)}\\
&=O\Big((1+\log j)^3 j^{\max(-d,-\zeta)-1}\Big)+O\Big(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\Big)\\
&=O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right), \qquad (2.102)
\end{aligned}$$

and the proof can be carried out the same way as (2.94). Thus, (2.87) holds.

Turning to (2.88), consider the partial derivatives $\partial\tau_j(\theta,t)/\partial\varphi_{(l)}$, where

$$\frac{\partial\tau_j(\theta,t)}{\partial\varphi_{(l)}}=\nu[(\dot b_1(\varphi_{(l)})\cdots\dot b_t(\varphi_{(l)}))\Xi_t(\theta)S'_{d,t}]_{(j)} \qquad (2.103)$$

$$\begin{aligned}
&-\nu[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)\\
&\times(\dot B'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot B_{\varphi_{(l)},t})\Xi_t(\theta)S'_{d,t}]_{(j)}.
\end{aligned} \qquad (2.104)$$

By assumption 2.3, 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 (2.103), analogously to (2.79)

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

and, analogously to (2.80)

$$[(\dot b_1(\varphi_{(l)})\cdots\dot b_t(\varphi_{(l)}))\Xi_t(\theta)S_{d,t}]_{(j)}=O\left((1+\log j)j^{\max(-d,-\zeta)-1}\right), \qquad (2.105)$$

so that (2.105) determines the rate of (2.103). Next, consider (2.104), for which one

has by (2.79) and (2.90)

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})]_{(j)}\\
&\quad=O\Big((1+\log j)j^{\max(-d,-\zeta)-1}\Big)\\
&\qquad+O\Big(\sum_{i=1}^{j-1}(1+\log i)i^{\max(-d,-\zeta)-1}(j-i)^{-\zeta-1}\Big)\\
&\qquad+O\Big(\sum_{i=j+1}^{t}(1+\log i)i^{\max(-d,-\zeta)-1}(i-j)^{-\zeta-1}\Big)\\
&\qquad=O\left((1+\log j)^2 j^{\max(-d,-\zeta)-1}\right),
\end{aligned}\tag{2.106}$$

where the proof is identical to (2.92). By the same proof as for (2.93), by (2.73) and (2.106)

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta)]_{(j)}\\
&\quad=O\Big((1+\log j)^2 j^{\max(-d,-\zeta)-1}\Big)\\
&\qquad+O\Big(\sum_{i=1}^{j-1}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(j-i)^{\max(-d,-\zeta)-1}\Big)\\
&\qquad+O\Big(\sum_{i=j+1}^{t}(1+\log i)^2 i^{\max(-d,-\zeta)-1}(i-j)^{\max(-d,-\zeta)-1}\Big)\\
&\qquad=O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right).
\end{aligned}\tag{2.107}$$

Finally, again by using the same proof as for (2.94), by (2.68) and (2.106)

$$\begin{aligned}
&[(b_1(\varphi)-\pi_1(d)\cdots b_t(\varphi)-\pi_t(d))\Xi_t(\theta)(\dot{B}'_{\varphi_{(l)},t}B_{\varphi,t}+B'_{\varphi,t}\dot{B}_{\varphi_{(l)},t})\Xi_t(\theta)S'_{d,t}]_{(j)}\\
&=O\Big((1+\log j)^3 j^{\max(-d,-\zeta)-1}\Big)+O\Big(\sum_{i=j+1}^{t}(1+\log i)^3 i^{\max(-d,-\zeta)-1}(i-j)^{-d-1}\Big)\\
&=O\left((1+\log j)^3 j^{\max(-d,-\zeta)-1}\right).
\end{aligned}\tag{2.108}$$

Together, (2.105) and (2.108) yield (2.88). □

Lemma 2.A.5 (Convergence of the partial derivatives of $\tau_j(\theta,t)$ to $\tau_j(\theta)$)**.** *For the partial derivatives of $\tau_j(\theta,t)$, it holds that*

$$\frac{\partial\tau_j(\theta,t)}{\partial\theta}\Big|_{\theta=\theta_0}-\frac{\partial\tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}=\sum_{k=t+1}^{\infty}\frac{\partial r_{\tau,j,k}(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}=O\left((1+\log t)^5 t^{\max(-d_0-\zeta)-1}\right),\tag{2.109}$$

with $r_{\tau,j,k}(\theta)$ as given in lemma 2.A.3.

Proof of lemma 2.A.5. From (2.85) and below $r_{\tau,j,t+1}(\theta) = -\nu R5_{(j)}$, where

$$\begin{aligned} R5_{(j)} =& [(b_{t+1}(\varphi) - \pi_{t+1}(d)) \left(R_2' S_{d,t}' + R_3 s_t'\right)]_{(j)} \\ &+ [(b_1(\varphi) - \pi_1(d) \cdots b_t(\varphi) - \pi_t(d)) \left(R_2 s_t' + R_1 S_{d,t}'\right)]_{(j)}, \end{aligned}$$

and with $B_{\varphi,t}$ and $S_{d,t}$ as defined in (2.5), $\beta_t' = (b_t(\varphi) \cdots b_1(\varphi))$, $s_t' = (\pi_t(d) \cdots \pi_1(d))$ as given in lemma 2.A.1, and R_1, R_2, R_3 as stated below (2.83). The partial derivative of $R5_{(j)}$ w.r.t. the l-th entry $\theta_{(l)}$ is thus given by

$$\frac{\partial R5_{(j)}}{\partial \theta_{(l)}} = \left[\frac{\partial (b_{t+1}(\varphi) - \pi_{t+1}(d))}{\partial \theta_{(l)}} \left(R_2' S_{d,t}' + R_3 s_t'\right)\right]_{(j)} \tag{2.110}$$

$$+ \left[\left(\frac{\partial (b_1(\varphi) - \pi_1(d))}{\partial \theta_{(l)}} \cdots \frac{\partial (b_t(\varphi) - \pi_t(d))}{\partial \theta_{(l)}}\right) \left(R_2 s_t' + R_1 S_{d,t}'\right)\right]_{(j)} \tag{2.111}$$

$$+ \left[(b_{t+1}(\varphi) - \pi_{t+1}(d)) \left(R_2' \frac{\partial S_{d,t}'}{\partial \theta_{(l)}} + R_3 \frac{\partial s_t'}{\partial \theta_{(l)}}\right)\right]_{(j)} \tag{2.112}$$

$$+ \left[((b_1(\varphi) - \pi_1(d)) \cdots (b_t(\varphi) - \pi_t(d))) \left(R_2 \frac{\partial s_t'}{\partial \theta_{(l)}} + R_1 \frac{\partial S_{d,t}'}{\partial \theta_{(l)}}\right)\right]_{(j)} \tag{2.113}$$

$$+ \left[(b_{t+1}(\varphi) - \pi_{t+1}(d)) \left(\frac{\partial R_2'}{\partial \theta_{(l)}} S_{d,t}' + \frac{\partial R_3}{\partial \theta_{(l)}} s_t'\right)\right]_{(j)} \tag{2.114}$$

$$+ \left[((b_1(\varphi) - \pi_1(d)) \cdots (b_t(\varphi) - \pi_t(d))) \left(\frac{\partial R_2}{\partial \theta_{(l)}} s_t' + \frac{\partial R_1}{\partial \theta_{(l)}} S_{d,t}'\right)\right]_{(j)}. \tag{2.115}$$

As noted in the proof of lemma 2.A.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 2.3. Thus, the convergence rates of (2.110) and (2.111) can be derived analogously to the proof of lemma 2.A.3. This yields that (2.110) 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 (2.111) 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 2.A.3. Thus, it holds that (2.112) is $O((t + 1)^{\max(-d,-\zeta)-1}(1 + \log(t + 1 - j))^3(t + 1 - j)^{\max(-d,-\zeta)-1})$, while (2.113) 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 (2.114) and (2.115), note that $R_3 = O(1)$ as shown in (2.84) and below. Since $\beta_t'(\partial \beta_t/\partial \theta_{(l)})$,

$s_t'(\partial s_t/\partial\theta_{(l)})$, $s_t's_t$, $(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)\partial(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})'/\partial\theta_{(l)}$, and $(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})(\partial\Xi_t(\theta)/\partial\theta_{(l)})(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})'$ are $O(1)$, it follows that

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

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

$$\frac{\partial R_{2_{(j)}}}{\partial\theta_{(l)}} = -\frac{\partial R_3}{\partial\theta_{(l)}}\left[(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\Xi_t(\theta)\right]_{(j)} \tag{2.116}$$

$$- R_3\left[(\beta_t'B_{\varphi,t}+\nu s_t'S_{d,t})\frac{\partial\Xi_t(\theta)}{\partial\theta_{(l)}}\right]_{(j)} \tag{2.117}$$

$$\begin{aligned}- R_3\Big[\big(&\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}}+\frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t}+\frac{\partial\nu}{\partial\theta_{(l)}}s_t'S_{d,t}\\ &+\nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t}+\nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\big)\Xi_t(\theta)\Big]_{(j)},\end{aligned} \tag{2.118}$$

where the first term (2.116) is $O\left((1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1}\right)$ by (2.84) and by $\partial R_3/\partial\theta_{(l)}=O(1)$. For the term (2.117), 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 (2.84). Together with $\partial\Xi_t(\theta)/\partial\theta_{(l)} = -\Xi_t(\theta)[(\partial/\partial\theta_{(l)})(B_{\varphi,t}'B_{\varphi,t}+\nu S_{d,t}'S_{d,t})]\Xi_t(\theta)$, (2.89) and (2.90), it follows that

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

Finally, using (2.73), one obtains

$$\begin{aligned}&\left\{(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)\left[\frac{\partial}{\partial\theta_{(l)}}\left(B_{\varphi,t}' B_{\varphi,t} + \nu S_{d,t}' S_{d,t}\right)\right]\Xi_t(\theta)\right\}_{(j)} \\ &= O\left((1+\log(t+1-j))^4 (t+1-j)^{\max(-d,-\zeta)-1}\right),\end{aligned} \tag{2.119}$$

which yields the binding rate of convergence for the second term in (2.116). For (2.118)

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

by lemma 2.A.1. Hence, using (2.73) yields an upper bound for (2.118)

$$\begin{aligned}&\left[\left(\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}} + \frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t} + \frac{\partial\nu}{\partial\theta_{(l)}}s_t' S_{d,t} + \nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t} + \nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\right)\Xi_t(\theta)\right]_{(j)} \\ &= O\left((1+\log(t+1-j))^2 (t+1-j)^{\max(-d,-\zeta)-1}\right).\end{aligned} \tag{2.120}$$

Together, the rates of convergence of (2.116) and (2.118) yield

$$\frac{\partial R_{2_{(j)}}}{\partial\theta_{(l)}} = O\left((1+\log(t+1-j))^3 (t+1-j)^{\max(-d,-\zeta)-1}\right). \tag{2.121}$$

For the partial derivatives of R_1, note that

$$\frac{\partial R_{1_{(i,j)}}}{\partial\theta_{(l)}} = -\frac{\partial R_{2_{(i)}}}{\partial\theta_{(l)}}\left[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\Xi_t(\theta)\right]_{(j)} \tag{2.122}$$

$$- R_{2_{(i)}}\left[(\beta_t' B_{\varphi,t} + \nu s_t' S_{d,t})\frac{\partial\Xi_t(\theta)}{\partial\theta_{(l)}}\right]_{(j)} \tag{2.123}$$

$$\begin{aligned}-R_{2_{(i)}}\Big[\Big(&\beta_t'\frac{\partial B_{\varphi,t}}{\partial\theta_{(l)}} + \frac{\partial\beta_t'}{\partial\theta_{(l)}}B_{\varphi,t} + \frac{\partial\nu}{\partial\theta_{(l)}}s_t' S_{d,t} \\ &+ \nu\frac{\partial s_t'}{\partial\theta_{(l)}}S_{d,t} + \nu s_t'\frac{\partial S_{d,t}}{\partial\theta_{(l)}}\Big)\Xi_t(\theta)\Big]_{(j)}.\end{aligned} \tag{2.124}$$

From (2.84) and (2.121), the first term in (2.122) is $O((1+\log(t+1-i))^4(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))(t+1-j)^{\max(-d,-\zeta)-1})$. Similarly, using (2.119) and the convergence rate of $R_{2_{(i)}}$ as derived in the proof of lemma 2.A.3, the second term in (2.123) is $O((1+\log(t+1-i))(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1})$. By (2.120), it follows that (2.124) is $O((1+\log(t+1-i))(t+$

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

$$\begin{aligned}\frac{\partial R1_{(i,j)}}{\partial\theta_{(l)}} = O\Big(&(1+\log(t+1-i))^4(t+1-i)^{\max(-d,-\zeta)-1}\\ &\times(1+\log(t+1-j))^4(t+1-j)^{\max(-d,-\zeta)-1}\Big).\end{aligned} \tag{2.125}$$

With (2.121) at hand, it follows directly for (2.114) that

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

By (2.68) and (2.69), it follows that (2.114) is $O\big((t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\big)$. For (2.115), it follows from (2.121) and (2.125) that $\left(\frac{\partial R_2}{\partial\theta_{(l)}}s_t' + \frac{\partial R_1}{\partial\theta_{(l)}}S_{d,t}'\right)_{(i,j)} = O\big((1+\log(t+1-i))^4(t+1-i)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\big)$. Again using (2.68) and (2.69), it thus follows that (2.115) is $O\big((1+\log(t+1))^5(t+1)^{\max(-d,-\zeta)-1}(1+\log(t+1-j))^5(t+1-j)^{\max(-d,-\zeta)-1}\big)$. Together, this implies for (2.109) that

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

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

Lemma 2.A.6. *For the truncated score function as given in* (2.64)*, and the untruncated score function as given in* (2.65)*, it holds for all* $\theta\in\Theta_3(\kappa_3)$ *that*

$$\sqrt{n}\left[\frac{\partial\tilde{Q}(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} - \frac{\partial Q(y,\theta)}{\partial\theta}\bigg|_{\theta=\theta_0}\right] = o_p(1). \tag{2.126}$$

Proof of lemma 2.A.6. For brevity, define $h_{1,t} = \sum_{j=1}^{t-1}\frac{\partial\tau_j(\theta,t)}{\partial\theta}\Big|_{\theta=\theta_0}\xi_{t-j}(d_0)$, $\tilde{h}_{1,t} = \sum_{j=1}^{\infty}\frac{\partial\tau_j(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0)$, as well as $h_{2,t} = \sum_{j=0}^{t-1}\tau_j(\theta_0,t)\frac{\partial\xi_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$, and $\tilde{h}_{2,t} = \sum_{j=0}^{\infty}\tau_j(\theta_0)\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\Big|_{\theta=\theta_0}$. Then plugging (2.64), (2.65) into (2.126) and using (2.38) yields

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

$$
\begin{aligned}
&= \frac{2}{\sqrt{n}} \left[\sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{1,t} - h_{1,t}) + \sum_{t=1}^{n} h_{1,t} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0) \right) \right] \\
&+ \frac{2}{\sqrt{n}} \left[\sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{2,t} - h_{2,t}) + \sum_{t=1}^{n} h_{2,t} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0) \right) \right],
\end{aligned} \tag{2.127}
$$

so that it remains to be shown that all four terms in (2.127) are $o_p(1)$.

For the proofs it will be very useful to note that $\tilde{v}_t(\theta_0)$ adapted to the filtration $\mathcal{F}_t^{\tilde{\xi}} = \sigma(\tilde{\xi}_s, s \leq t)$ is a stationary martingale difference sequence (MDS), as explained in the proof of theorem 2.4.2. Note in addition that all $\tilde{h}_{1,t}$, $\tilde{h}_{2,t}$ are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, as $\tau_0 = \pi_0 = 1$ are invariant w.r.t. θ.

Starting with the first term of (2.127), by plugging in $h_{1,t}$ and $\tilde{h}_{1,t}$

$$
\begin{aligned}
&\frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{1,t} - h_{1,t}) \\
&\quad = \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=1}^{t-1} \frac{\partial \tau_j(\theta, t)}{\partial \theta}\bigg|_{\theta=\theta_0} \left(\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0) \right) && (2.128) \\
&\quad + \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=1}^{t-1} \left(\frac{\partial \tau_j(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta, t)}{\partial \theta}\bigg|_{\theta=\theta_0} \right) \tilde{\xi}_{t-j}(d_0) && (2.129) \\
&\quad + \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=t}^{\infty} \frac{\partial \tau_j(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0} \tilde{\xi}_{t-j}(d_0). && (2.130)
\end{aligned}
$$

As $\sum_{j=t}^{\infty} \frac{\partial \tau_j(\theta)}{\partial \theta}\big|_{\theta=\theta_0} \tilde{\xi}_{t-j}(d_0)$ is $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable, $\tilde{v}_t(\theta_0) \sum_{j=t}^{\infty} \frac{\partial \tau_j(\theta)}{\partial \theta}\big|_{\theta=\theta_0} \tilde{\xi}_{t-j}(d_0)$ is also a MDS. Since $\frac{\partial \tau_j(\theta)}{\partial \theta}\big|_{\theta=\theta_0} = O((1+\log j)^4 j^{\max(-d_0,-\zeta)-1})$, see lemma 2.A.4, it follows that (2.130) is $o_p(1)$. In (2.129), $\tilde{v}_t(\theta_0) \sum_{j=1}^{t-1} \left(\frac{\partial \tau_j(\theta)}{\partial \theta}\big|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta,t)}{\partial \theta}\big|_{\theta=\theta_0} \right) \tilde{\xi}_{t-j}(d_0)$ adapted to $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS, while the sum $\sum_{j=1}^{t-1} \left(\frac{\partial \tau_j(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} - \frac{\partial \tau_j(\theta,t)}{\partial \theta}\Big|_{\theta=\theta_0} \right) \tilde{\xi}_{t-j}(d_0) = O_p((1+\log t)^5 t^{\max(-d_0,-\zeta)})$ by lemma 2.A.5. Hence (2.129) is $o_p(1)$. For (2.128), note that by assumption 2.1

$$
\begin{aligned}
&\mathrm{E}\left\{ \left[\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=1}^{t-1} \frac{\partial \tau_j(\theta, t)}{\partial \theta}\bigg|_{\theta=\theta_0} \left(\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0) \right) \right]^2 \right\} \\
&\quad = \mathrm{E}\Bigg[\sum_{s,t=1}^{n} \Big(\sum_{j=0}^{\infty} \eta^2_{\min(s,t)-j} \tau_j(\theta_0) \tau_{j+|t-s|}(\theta_0) \Big) \\
&\qquad \times \sum_{j=0}^{\infty} \epsilon^2_{-j} \Big(\sum_{k=0}^{t-1} \frac{\partial \tau_k(\theta, t)}{\partial \theta}\bigg|_{\theta=\theta_0} \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0) \Big) && (2.131) \\
&\qquad\quad \times \left(\sum_{k=0}^{s-1} \frac{\partial \tau_k(\theta, s)}{\partial \theta'}\bigg|_{\theta=\theta_0} \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0) \right) \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\Bigg(\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\Bigg)\\
&\qquad\times\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\Bigg]
\end{aligned}
\tag{2.132}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\Bigg(\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\\
&\qquad\times\Bigg(\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{s-1}\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\bigg|_{\theta=\theta_0}\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\Bigg].
\end{aligned}
\tag{2.133}
$$

For (2.131), 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 2.A.2, and $\sum_{k=0}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta}\big|_{\theta=\theta_0}\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0) = O\big((1+\log(t+j))^6(t+j)^{\max(-d_0,-\zeta)-1}\big)$, see (2.68) together with lemma 2.A.4. This yields the upper bound for (2.131)

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

Similarly, for the second term (2.132), by (2.68) and lemma 2.A.2 it holds that

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

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

Furthermore, by lemma 2.A.4

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

so that by the same proof as for (2.131), it holds that (2.132) is also $O(1)$.

By (2.68) and lemmas 2.A.2 and 2.A.4, the third term (2.133) is bounded from above by

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

As all three terms (2.131) to (2.133) are $O(1)$, it follows directly by the scaling that (2.128) is $o_p(1)$. Now, since (2.128) to (2.130) are $o_p(1)$, the first term in (2.127) is also $o_p(1)$.

Next, consider the third term in (2.127). I plug in $h_{2,t}$ and $\tilde{h}_{2,t}$ which gives

$$\begin{aligned}
&\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{2,t}-h_{2,t})\\
&=\frac{2}{\sqrt{n}}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left(\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}-\frac{\partial\xi_{t-j}(d)}{\partial\theta}\bigg|_{\theta=\theta_0}\right) \qquad (2.134)
\end{aligned}$$

$$+ \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=0}^{t-1} \left(\tau_j(\theta_0) - \tau_j(\theta_0, t)\right) \frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\bigg|_{\theta=\theta_0} \tag{2.135}$$

$$+ \frac{2}{\sqrt{n}} \sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=t}^{\infty} \tau_j(\theta_0) \frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\bigg|_{\theta=\theta_0}. \tag{2.136}$$

For (2.136), note that $(\tilde{v}_t(\theta_0), \mathcal{F}_t^{\tilde{\xi}})$ is a stationary MDS, and $\sum_{j=t}^{\infty} \tau_j(\theta_0) \frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\big|_{\theta=\theta_0}$ is $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. Since $\partial \tilde{\xi}_{t-i}(d)/\partial\theta$ is $O_p(1)$ for all $d > d_0 - 1/2$, it follows by lemma 2.A.2 that $\sum_{j=t}^{\infty} \tau_j(\theta_0) \frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\big|_{\theta=\theta_0} = O_p((1+\log t) t^{\max(-d_0, -\zeta)})$, and thus (2.136) is $o_p(1)$.

For (2.135), note that $\tilde{v}_t(\theta_0) \sum_{j=0}^{t-1} \left(\tau_j(\theta_0) - \tau_j(\theta_0, t)\right) \frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\big|_{\theta=\theta_0}$ together with $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS. Furthermore, by lemma 2.A.3, it holds that $\tau_j(\theta_0) - \tau_j(\theta_0, t) = O((1+\log t)^2 t^{\max(-d_0, -\zeta)-1})$. Since the partial derivatives of $\tilde{\xi}_t(d)$ are bounded in probability, $\sum_{j=0}^{t-1} \left(\tau_j(\theta_0) - \tau_j(\theta_0, t)\right) \frac{\partial \tilde{\xi}_{t-j}(d)}{\partial \theta}\big|_{\theta=\theta_0} = O_p((1+\log t)^2 t^{\max(-d_0, -\zeta)})$. Therefore, (2.135) is $o_p(1)$.

For (2.134), I use $\frac{\partial \pi_j(d-d_0)}{\partial d}\big|_{d=d_0} = -j^{-1}$ as shown by Robinson (2006, pp. 135-136) and Hualde and Robinson (2011, p. 3170). Thus, the partial derivative in (2.134) w.r.t. d is

$$\frac{\partial \tilde{\xi}_t(\theta)}{\partial d}\bigg|_{\theta=\theta_0} - \frac{\partial \xi_t(\theta)}{\partial d}\bigg|_{\theta=\theta_0} = -\sum_{j=t}^{\infty} j^{-1} \eta_{t-j} + \sum_{j=0}^{\infty} \epsilon_{-j} \sum_{k=0}^{j} \frac{\partial \pi_{t+j-k}(d)}{\partial d}\bigg|_{\theta=\theta_0} a_k(\varphi_0). \tag{2.137}$$

As the partial derivatives w.r.t. all other entries in θ are zero, by assumption 2.1 it is sufficient to consider

$$\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{j=0}^{t-1} \tau_j(\theta_0, t) \left(\frac{\partial \tilde{\xi}_{t-j}(d)}{\partial d}\bigg|_{\theta=\theta_0} - \frac{\partial \xi_{t-j}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right)\right]^2\right\} \\
&= \sum_{s,t=1}^{n} \mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1} \eta^2_{\min(s,t)-j} \tau_j(\theta_0) \tau_{j+|t-s|}(\theta_0)\right] \\
&\qquad \times \mathrm{E}\left[\sum_{j=0}^{\infty} \eta^2_{-j} \left(\sum_{k=0}^{t-1} \frac{\tau_k(\theta_0, t)}{t+j-k}\right)\left(\sum_{k=0}^{s-1} \frac{\tau_k(\theta_0, s)}{s+j-k}\right)\right. \\
&\qquad\quad + \sum_{j=0}^{\infty} \epsilon^2_{-j} \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j} a_l(\varphi_0) \frac{\partial \pi_{j+t-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right) \\
&\qquad\quad \left. \times \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j} a_l(\varphi_0) \frac{\partial \pi_{j+s-k-l}(d)}{\partial d}\bigg|_{\theta=\theta_0}\right)\right]
\end{aligned} \tag{2.138}$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0)\sum_{l=0}^{j+|t-s|-k}a_l(\varphi_0)\pi_{j+|t-s|-k-l}(d_0)\right)\Bigg]\\
&\quad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\frac{\tau_k(\theta_0,t)}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\frac{\tau_k(\theta_0,s)}{s+j-k}\right)\right.\\
&\qquad+\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\left.\frac{\partial\pi_{j+t-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\\
&\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\left.\frac{\partial\pi_{j+s-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\right]
\end{aligned}
\tag{2.139}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg\{\Bigg[\sum_{j=t}^{\infty}\eta^2_{t-j}\tau_j(\theta_0)\sum_{k=0}^{t-1}\frac{-\tau_k(\theta_0,t)}{j-k}\\
&\qquad+\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\Bigg]\\
&\quad\Bigg[\sum_{j=s}^{\infty}\eta^2_{s-j}\tau_j(\theta_0)\sum_{k=0}^{s-1}\frac{-\tau_k(\theta_0,s)}{j-k}\\
&\qquad+\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\qquad\left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\left.\frac{\partial\pi_{j-k-l}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\Bigg]\Bigg\}.
\end{aligned}
\tag{2.140}
$$

For (2.138), note the first expectation is $\sigma^2_{\eta,0}\sum_{j=0}^{\min(s,t)-1}\tau_j(\theta_0)\tau_{j+|t-s|}(\theta_0)=O(|t-s|^{\max(-d_0,-\zeta)-1})$ for all $t\neq s$, and $O(1)$ for $t=s$, see lemma 2.A.2. For the other terms it holds that $\mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\frac{1}{t+j-k}\right)\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\frac{1}{s+j-k}\right)\right]\leq K\sum_{j=0}^{\infty}(1+\log(t+j))^2(t+j)^{-1}(1+\log(s+j))^2(s+j)^{-1}$, together with

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

by lemma 2.A.2. It follows that (2.138) is bounded from above by

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

for $0<\kappa<1/2$, since $\sum_{j=0}^{\infty}(s+j)^{-2}=O(s^{-1})$, see Chan and Palma (1998, lemma 3.2), and, as the logarithm is dominated by its powers, $\sum_{j=0}^{\infty}(1+\log(s+j))^2(s+j)^{-2}=O(s^{-1+\kappa})$ for all $0<\kappa<1/2$. For (2.139), by lemmas 2.A.1 and 2.A.2, the first expectation is bounded by

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

for all $t\neq s$, and is $O(1)$ for $t=s$. Hence, by the same proof as for (2.138) the second term (2.139) is also $O(n^{\kappa})$, $0<\kappa<1/2$. For the third term (2.140) one has by lemma 2.A.2

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

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

which is $O(1)$, and thus all terms (2.138) to (2.140) are $O(n^\kappa)$. As (2.134) is appropriately scaled, it follows that (2.134) is $o_p(1)$ and thus the third term in (2.127) is $o_p(1)$.

Next, consider the second term in (2.127) that can be decomposed into

$$
\begin{aligned}
&\frac{2}{\sqrt{n}} \sum_{t=1}^{n} h_{1,t} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0) \right) \\
&\quad = \frac{2}{\sqrt{n}} \sum_{t=1}^{n} h_{1,t} \sum_{j=0}^{t-1} (\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0)) \tau_j(\theta_0, t) \\
&\quad + \frac{2}{\sqrt{n}} \sum_{t=0}^{n} h_{1,t} \sum_{j=1}^{t-1} (\tau_j(\theta_0) - \tau_j(\theta_0, t)) \tilde{\xi}_{t-j}(d_0) \\
&\quad + \frac{2}{\sqrt{n}} \sum_{t=1}^{n} h_{1,t} \sum_{j=t}^{\infty} \tau_j(\theta_0) \tilde{\xi}_{t-j}(d_0).
\end{aligned}
\tag{2.141}
$$

For the first term in (2.141), note that by assumption 2.1

$$
\begin{aligned}
&\mathrm{E}\left\{ \left[\sum_{t=1}^{n} h_{1,t} \sum_{j=0}^{t-1} (\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0)) \tau_j(\theta_0, t) \right]^2 \right\} \\
&= \sum_{s,t=1}^{n} \mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1} \left. \frac{\partial \tau_j(\theta, \min(s,t))}{\partial \theta} \right|_{\theta=\theta_0} \left. \frac{\partial \tau_{j+|t-s|}(\theta, \max(s,t))}{\partial \theta'} \right|_{\theta=\theta_0} \eta^2_{\min(s,t)-j} \right] \\
&\qquad \times \mathrm{E}\left[\sum_{j=0}^{\infty} \epsilon^2_{-j} \left(\sum_{k=0}^{t-1} \tau_k(\theta_0, t) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+t-k-l}(d_0) \right) \right. \\
&\qquad\qquad \left. \times \left(\sum_{k=0}^{s-1} \tau_k(\theta_0, s) \sum_{l=0}^{j} a_l(\varphi_0) \pi_{j+s-k-l}(d_0) \right) \right]
\end{aligned}
\tag{2.142}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\left.\frac{\partial\tau_k(\theta,\min(s,t))}{\partial\theta}\right|_{\theta=\theta_0}\sum_{l=0}^{j-k}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{j+|t-s|}\left.\frac{\partial\tau_k(\theta,\max(s,t))}{\partial\theta'}\right|_{\theta=\theta_0}\sum_{l=0}^{j+|t-s|-k}\pi_l(d_0)a_{j+|t-s|-k-l}(\varphi_0)\right)\Bigg]\\
&\qquad\times\mathrm{E}\Bigg[\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\\
&\qquad\qquad\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\Bigg]
\end{aligned}
\tag{2.143}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\Bigg(\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{t-1}\left.\frac{\partial\tau_k(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,t-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\\
&\qquad\times\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{s-1}\left.\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,s-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\\
&\qquad\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg)\Bigg].
\end{aligned}
\tag{2.144}
$$

For (2.142), one has for all $t\neq s$

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

by lemma 2.A.4, and $O(1)$ for $t=s$. Furthermore, for (2.143), the first term is bounded by

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

by lemmas 2.A.1 and 2.A.4 for $t\neq s$, and $O(1)$ otherwise. In addition, for both (2.142) and (2.143), by lemmas 2.A.1 and 2.A.2 the other remaining term is bounded

by

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

Consequently, both (2.142) and (2.143) are $\sum_{s,t=1}^{n} O\big((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1}|t-s|^{\max(-d_0,-\zeta)-1}\big) = O(1)$. Finally, by lemmas 2.A.1, 2.A.2, and 2.A.4, (2.144) is

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

Thus, the first term in (2.141) is $o_p(1)$. For the second term in (2.141), note that by lemma 2.A.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 (2.141)

$$\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^{n}h_{1,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0)\right]^2\right\}\\
&=\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\eta_{\min(s,t)-j}^2\left.\frac{\partial\tau_j(\theta,\min(s,t))}{\partial\theta}\right|_{\theta=\theta_0}\left.\frac{\partial\tau_{j+|t-s|}(\theta,\max(s,t))}{\partial\theta'}\right|_{\theta=\theta_0}\right]\\
&\quad\times\mathrm{E}\left[\sum_{j=0}^{\infty}\eta_{-j}^2\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)+\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{j}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\\
&\quad\left.\times\left(\sum_{k=0}^{j}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right]
\end{aligned}\tag{2.145}$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\left.\frac{\partial\tau_k(\theta,\min(s,t))}{\partial\theta}\right|_{\theta=\theta_0}\sum_{l=0}^{j-k}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\\
&\quad\times\left(\sum_{k=0}^{j+|t-s|}\left.\frac{\partial\tau_k(\theta,\max(s,t))}{\partial\theta'}\right|_{\theta=\theta_0}\sum_{l=0}^{j+|t-s|-k}\pi_l(d_0)a_{j+|t-s|-k-l}(\varphi_0)\right)\Bigg]\\
&\quad\times\mathrm{E}\Bigg[\sum_{j=0}^{\infty}\eta^2_{-j}\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)+\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{j}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\\
&\quad\times\left(\sum_{k=0}^{j}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\Bigg]
\end{aligned}
\tag{2.146}
$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\left(\sum_{j=t}^{\infty}\epsilon^2_{t-j}\left(\sum_{k=0}^{t-1}\left.\frac{\partial\tau_k(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,t-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\right.\\
&\quad\left.\times\left(\sum_{k=0}^{j-t}\tau_{j+k}(\theta_0)\sum_{l=0}^{j-t-k}a_l(\varphi_0)\pi_{j-t-k-l}(d_0)\right)\right)\\
&\quad\times\left(\sum_{j=s}^{\infty}\epsilon^2_{s-j}\left(\sum_{k=0}^{s-1}\left.\frac{\partial\tau_k(\theta,s)}{\partial\theta'}\right|_{\theta=\theta_0}\sum_{l=0}^{\min(j-k,s-1)}\pi_l(d_0)a_{j-k-l}(\varphi_0)\right)\right.\\
&\quad\left.\times\left(\sum_{k=0}^{j-s}\tau_{j+k}(\theta_0)\sum_{l=0}^{j-s-k}a_l(\varphi_0)\pi_{j-s-k-l}(d_0)\right)\right)\Bigg].
\end{aligned}
\tag{2.147}
$$

For (2.145) and (2.146), it holds that

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

and $\mathrm{E}\left[\sum_{j=0}^{\infty}\eta^2_{-j}\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)\right]=O((1+\log t)t^{-\min(d_0,\zeta)}(1+\log s)s^{-\min(d_0,\zeta)-1})$. Thus, analogously to (2.142) and (2.143), expressions (2.145) and (2.146) are $O(1)$. Also analogously to (2.144), by lemmas 2.A.1, 2.A.2, and 2.A.4, (2.147) is bounded from above by

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

Therefore, also the third term in (2.141) is $o_p(1)$. It follows that the second term in

(2.127) is $o_p(1)$. Finally, consider the last term in (2.127)

$$\begin{aligned}
&\frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
&\quad = \frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0))\tau_j(\theta_0, t) \\
&\qquad + \frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\sum_{j=1}^{t-1}(\tau_j(\theta_0) - \tau_j(\theta_0, t))\tilde{\xi}_{t-j}(d_0) \\
&\qquad + \frac{2}{\sqrt{n}}\sum_{t=1}^{n} h_{2,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0).
\end{aligned} \tag{2.148}$$

For the first term in (2.148), by assumption 2.1 it holds that

$$\begin{aligned}
&\mathrm{E}\left\{\left[\sum_{t=1}^{n}\left(\sum_{j=0}^{t-1}\tau_j(\theta_0, t)\left.\frac{\partial\xi_{t-j}(d)}{\partial d}\right|_{\theta=\theta_0}\right)\sum_{j=0}^{t-1}(\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0))\tau_j(\theta_0, t)\right]^2\right\} \\
&= \sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1}\eta^2_{\min(s,t)-j}\left(\sum_{k=1}^{j}\frac{1}{k}\tau_{j-k}(\theta_0, \min(s,t))\right)\right. \\
&\qquad\qquad \left.\times\left(\sum_{k=1}^{j+|t-s|}\frac{1}{k}\tau_{j+|t-s|-k}(\theta_0, \max(s,t))\right)\right] \\
&\qquad \times\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0, t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\right. \\
&\qquad\qquad \left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0, s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\right]
\end{aligned} \tag{2.149}$$

$$\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon^2_{\min(s,t)-j}\left(\sum_{k=0}^{j}\tau_k(\theta_0, \min(s,t))\sum_{l=0}^{j-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\right)\right. \\
&\qquad \left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0, \max(s,t))\sum_{l=0}^{j+|t-s|-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j+|t-s|-k-l}(\varphi_0)\right)\right] \\
&\qquad \times\mathrm{E}\left[\sum_{j=0}^{\infty}\epsilon^2_{-j}\left(\sum_{k=0}^{t-1}\tau_k(\theta_0, t)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+t-k-l}(d_0)\right)\right. \\
&\qquad\qquad \left.\times\left(\sum_{k=0}^{s-1}\tau_k(\theta_0, s)\sum_{l=0}^{j}a_l(\varphi_0)\pi_{j+s-k-l}(d_0)\right)\right]
\end{aligned} \tag{2.150}$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\Bigg(\sum_{j=t}^{\infty}\epsilon_{t-j}^{2}\Bigg(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{t-1-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\Bigg)\\
&\qquad\times\Bigg(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{j-t}a_l(\varphi_0)\pi_{j-k-l}(d_0)\Bigg)\Bigg)\\
&\quad\times\Bigg(\sum_{j=s}^{\infty}\epsilon_{s-j}^{2}\Bigg(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{s-1-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\Bigg)\\
&\qquad\times\Bigg(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{j-s}a_l(\varphi_0)\pi_{j-k-l}(d_0)\Bigg)\Bigg)\Bigg],
\end{aligned}
\tag{2.151}
$$

while all other partial derivatives of $\xi_{t-j}(d)$ (i.e. those w.r.t. all other entries except d) are zero. By lemma 2.A.2, the first term in (2.149) is

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

for $t\neq s$, and $O(1)$ otherwise. In addition, by lemmas 2.A.1 and 2.A.2 it holds that the first term of (2.150) is

$$
\begin{aligned}
&\mathrm{E}\Bigg[\sum_{j=0}^{\min(s,t)-1}\epsilon_{\min(s,t)-j}^{2}\Bigg(\sum_{k=0}^{j}\tau_k(\theta_0,\min(s,t))\sum_{l=0}^{j-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\Bigg)\\
&\qquad\times\Bigg(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0,\max(s,t))\sum_{l=0}^{j+|t-s|-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j+|t-s|-k-l}(\varphi_0)\Bigg)\Bigg]\\
&\qquad=O(|t-s|^{\max(-d_0,-\zeta)-1}),
\end{aligned}
\tag{2.152}
$$

for $t\neq s$, and $O(1)$ otherwise. The second term in (2.149) and (2.150) is

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

Thus, analogously to (2.142), (2.143), (2.145) and (2.146), it holds that (2.149) and

(2.150) are $O(1)$. Finally, (2.151) is bounded from above by

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

Hence, the first term in (2.148) is $o_p(1)$. For the second term in (2.148), by lemma 2.A.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 (2.141), 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 (2.141)

$$\mathrm{E}\left\{\left[\sum_{t=1}^{n} h_{2,t}\sum_{j=t}^{\infty}\tau_j(\theta_0)\tilde{\xi}_{t-j}(d_0)\right]^2\right\}$$
$$= \sum_{s,t=1}^{n} \mathrm{E}\left[\sum_{j=1}^{\min(s,t)-1} \eta_{\min(s,t)-j}^2 \left(\sum_{k=1}^{j}\frac{1}{k}\tau_{j-k}(\theta_0,\min(s,t))\right)\right.$$
$$\left. \times \left(\sum_{k=1}^{j+|t-s|}\frac{1}{k}\tau_{j+|t-s|-k}(\theta_0,\max(s,t))\right)\right]$$
$$\times \mathrm{E}\left[\sum_{j=0}^{\infty}\eta_{-j}^2\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)\right. \tag{2.153}$$
$$+\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{j}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)$$
$$\left.\times\left(\sum_{k=0}^{j}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right]$$
$$+\sum_{s,t=1}^{n}\mathrm{E}\left[\sum_{j=0}^{\min(s,t)-1}\epsilon_{\min(s,t)-j}^2\left(\sum_{k=0}^{j}\tau_k(\theta_0,\min(s,t))\sum_{l=0}^{j-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\right)\right.$$
$$\left.\times\left(\sum_{k=0}^{j+|t-s|}\tau_k(\theta_0,\max(s,t))\sum_{l=0}^{j+|t-s|-k}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j+|t-s|-k-l}(\varphi_0)\right)\right. \tag{2.154}$$
$$\left.\times\mathrm{E}\left[\sum_{j=0}^{\infty}\eta_{-j}^2\tau_{t+j}(\theta_0)\tau_{s+j}(\theta_0)+\sum_{j=0}^{\infty}\epsilon_{-j}^2\left(\sum_{k=0}^{j}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right.\right.$$
$$\left.\left.\times\left(\sum_{k=0}^{j}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-k}a_l(\varphi_0)\pi_{j-k-l}(d_0)\right)\right]\right.$$

$$
\begin{aligned}
&+\sum_{s,t=1}^{n}\mathrm{E}\Bigg[\Bigg(\sum_{j=t}^{\infty}\epsilon_{t-j}^{2}\Bigg(\sum_{k=0}^{t-1}\tau_k(\theta_0,t)\sum_{l=0}^{t-k-1}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\Bigg)\\
&\qquad\times\Bigg(\sum_{k=0}^{j-t}\tau_{t+k}(\theta_0)\sum_{l=0}^{j-t-k}a_l(\varphi_0)\pi_{j-t-k-l}(d_0)\Bigg)\Bigg)\\
&\qquad\times\Bigg(\sum_{j=s}^{\infty}\epsilon_{s-j}^{2}\Bigg(\sum_{k=0}^{s-1}\tau_k(\theta_0,s)\sum_{l=0}^{s-k-1}\left.\frac{\partial\pi_l(d)}{\partial d}\right|_{\theta=\theta_0}a_{j-k-l}(\varphi_0)\Bigg)\\
&\qquad\times\Bigg(\sum_{k=0}^{j-s}\tau_{s+k}(\theta_0)\sum_{l=0}^{j-s-k}a_l(\varphi_0)\pi_{j-s-k-l}(d_0)\Bigg)\Bigg)\Bigg].
\end{aligned}
\tag{2.155}
$$

As noted above, the first expected value in (2.153) is $O(|t-s|^{-1})$ for $s \neq t$, else $O(1)$. For the second term (2.154), note that the first expectation is $O(|t-s|^{\max(-d_0,-\zeta)-1})$ for $s \neq t$, else $O(1)$, see (2.152). Furthermore, as shown below (2.147), the second expectation in (2.153) and (2.154) is $O((1+\log t)^3 t^{\max(-d_0,-\zeta)}(1+\log s)^3 s^{\max(-d_0,-\zeta)-1})$, and thus (2.153) and (2.154) are $O(1)$. Finally, the last term (2.155) is $O(1)$, and the proof is identical to (2.151). Thus, also the third term in (2.148) is $o_p(1)$. This shows that (2.127) is $o_p(1)$ and completes the proof. □

Lemma 2.A.7 (Boundedness of third partial derivatives of $Q(y,\theta)$)**.** *For $d \in D_3$ as defined in the proof of theorem 2.4.1, $\nu \in \Sigma_\nu$ as defined in section 2.4, and $\varphi \in N_\delta(\varphi_0)$ as defined in assumptions 2.2 and 2.4, the third partial derivatives of the objective function* (2.16) *are uniformly dominated by some random variable B_n that is $O_p(1)$,*

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

Proof of lemma 2.A.7. The third partial derivatives are

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

for $k,l,m = 1,...,q+2$, with $\partial v_t(\theta)/(\partial\theta_{(k)})$ in (2.38),

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

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

Boundedness in probability of the third partial derivatives then follows from (2.39) 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 (2.39). As can be seen from lemma 2.A.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 2.A.1, those for the first partial derivatives are contained in the proof of lemma 2.A.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 2.4. Based on them, the convergence rates of the following matrices are obtained

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

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

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

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

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

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

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

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

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

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

for $k,l,m = 1,2,...,q+2$. As becomes apparent, the partial derivatives just add a log-term to the convergence rates that is always dominated by its powers and thus does not affect the convergence of the partial derivatives. It follows that the first, second and third partial derivatives of $\tau_j(\theta,t)$ are absolutely summable in j and thus satisfy the condition for (2.39). By (2.39), $B_n = \sup_{d\in D_3,\nu\in\Sigma_\nu,\varphi\in N_\delta(\varphi_0)} \left|\frac{\partial^3 Q(y,\theta)}{\partial\theta^{(3)}}\right| = O_p(1)$. □

Lemma 2.A.8. *For the partial derivatives of $v_t(\theta)$, it holds that*

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

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

Proof of lemma 2.A.8. Consider

$$\left.\frac{\partial\tilde{v}_t(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial v_t(\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \sum_{j=1}^{t-1}\left.\frac{\partial\tau_j(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\left[\tilde{\xi}_{t-j}(d_0) - \xi_{t-j}(d_0)\right] \tag{2.156}$$

$$+\sum_{j=1}^{t-1}\left[\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\tau_j(\theta,t)}{\partial\theta}\right|_{\theta=\theta_0}\right]\tilde{\xi}_{t-j}(d_0) + \sum_{j=t}^{\infty}\left.\frac{\partial\tau_j(\theta)}{\partial\theta}\right|_{\theta=\theta_0}\tilde{\xi}_{t-j}(d_0) \tag{2.157}$$

$$+\sum_{j=0}^{t-1}\tau_j(\theta_0,t)\left[\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0} - \left.\frac{\partial\xi_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0}\right] \tag{2.158}$$

$$+\sum_{j=1}^{t-1}\left[\tau_j(\theta_0) - \tau_j(\theta_0,t)\right]\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0} + \sum_{j=t}^{\infty}\tau_j(\theta_0)\left.\frac{\partial\tilde{\xi}_{t-j}(d)}{\partial\theta}\right|_{\theta=\theta_0}. \tag{2.159}$$

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 (2.68), lemma 2.A.4, and assumption 2.2, the term (2.156) 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 2.A.5, (2.68), and assumption 2.3, the first term in (2.157) is

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

For the second term in (2.157), by lemma 2.A.4, (2.68), and assumption 2.3

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

Note that (2.158), (2.159) are non-zero only for the derivative w.r.t. d. For (2.158), it holds that $\left.\frac{\partial\pi_j(d-d_0)}{\partial d}\right|_{d=d_0} = -j^{-1}$, see Robinson (2006, pp. 135-136). Thus

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

by lemma 2.A.2, Johansen and Nielsen (2010, lemma B.3), and assumption 2.3. For the first term in (2.159), by lemmas 2.A.2, 2.A.3, Johansen and Nielsen (2010, lemma B.3), and assumption 2.3

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

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

while for the second term in (2.159), by lemma 2.A.2, Johansen and Nielsen (2010, lemma B.3), and assumption 2.3

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

Together, the results above prove lemma 2.A.8. □

Lemma 2.A.9. *For $v_t(\theta)$ as defined and (2.15) and $\tilde{v}_t(\theta)$ as defined in (2.29), 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 2.A.9. The proof is analogous to the proof of lemma 2.A.6 and thus is only summarized briefly. It will be helpful to note that there exists a constant $0<K<\infty$ such that

$$
\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}=O\left((1+\log k)^K k^{\max(-d,-\zeta)-1}\right), \tag{2.160}
$$

$$
\frac{\partial^2\tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}-\frac{\partial^2\tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}=O\left((1+\log t)^K t^{\max(-d,-\zeta)-1}\right). \tag{2.161}
$$

(2.160) can be seen directly from the proof of lemma 2.A.4, as the second partial derivatives only add a log-factor to the convergence rates in lemma 2.A.4. (2.161) can be shown analogously to the proof of lemma 2.A.5, where again the second partial derivatives only add a log-factor to the convergence rates in lemma 2.A.5.

To simplify the notation, define $h_{3,t_{(i,j)}} = \sum_{k=1}^{t-1} \frac{\partial^2 \tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0} \xi_{t-k}(d_0)$, $h_{4,t_{(i,j)}} = \sum_{k=1}^{t-1} \tau_k(\theta_0,t) \frac{\partial^2 \xi_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, $h_{5,t_{(i,j)}} = \sum_{k=1}^{t-1} \frac{\partial \tau_k(\theta,t)}{\partial\theta_{(i)}}\big|_{\theta=\theta_0} \frac{\partial \xi_{t-k}(d)}{\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, as well as $\tilde{h}_{3,t_{(i,j)}} = \sum_{k=1}^{\infty} \frac{\partial^2 \tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0} \tilde{\xi}_{t-k}(d_0)$, $\tilde{h}_{4,t_{(i,j)}} = \sum_{k=1}^{\infty} \tau_k(\theta_0) \frac{\partial^2 \tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$, $\tilde{h}_{5,t_{(i,j)}} = \sum_{k=1}^{\infty} \frac{\partial \tau_k(\theta)}{\partial\theta_{(i)}}\big|_{\theta=\theta_0} \frac{\partial \tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\big|_{\theta=\theta_0}$. The term of interest then can be written as

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \frac{\partial^2 \tilde{v}_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} - \frac{1}{n}\sum_{t=1}^{n} v_t(\theta_0) \frac{\partial^2 v_t(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \\
&= \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \left(\tilde{h}_{3,t_{(i,j)}} - h_{3,t_{(i,j)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{3,t_{(i,j)}} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
&+ \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \left(\tilde{h}_{4,t_{(i,j)}} - h_{4,t_{(i,j)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{4,t_{(i,j)}} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
&+ \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \left(\tilde{h}_{5,t_{(i,j)}} - h_{5,t_{(i,j)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{5,t_{(i,j)}} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
&+ \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \left(\tilde{h}_{5,t_{(j,i)}} - h_{5,t_{(j,i)}}\right) + \frac{1}{n}\sum_{t=1}^{n} h_{5,t_{(j,i)}} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right),
\end{aligned}
\tag{2.162}
$$

and thus the different terms in (2.162) can be considered separately and will be shown to be $o_p(1)$. Note that $\tilde{v}_t(\theta_0)$ adapted to the filtration $\mathcal{F}_t^{\tilde{\xi}}$ is a MDS as explained in the proof of theorem 2.4.2, while $\tilde{h}_{3,t_{(i,j)}}$, $\tilde{h}_{4,t_{(i,j)}}$, $\tilde{h}_{5,t_{(i,j)}}$ are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. Starting with the first term in (2.162), by plugging in $\tilde{h}_{3,t_{(i,j)}}$, $h_{3,t_{(i,j)}}$

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0)(\tilde{h}_{3,t_{(i,j)}} - h_{3,t_{(i,j)}}) \\
&= \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{k=1}^{t-1} \frac{\partial^2 \tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \left(\tilde{\xi}_{t-k}(d_0) - \xi_{t-k}(d_0)\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{k=1}^{t-1} \left(\frac{\partial^2 \tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} - \frac{\partial^2 \tau_k(\theta,t)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \right) \tilde{\xi}_{t-k}(d_0) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n} \tilde{v}_t(\theta_0) \sum_{k=t}^{\infty} \frac{\partial^2 \tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \tilde{\xi}_{t-k}(d_0).
\end{aligned}
\tag{2.163}
$$

The latter two terms in (2.163) are MDS when adapted to $\mathcal{F}_t^{\tilde{\xi}}$, as $(\tilde{v}_t(\theta_0), \mathcal{F}_t^{\tilde{\xi}})$ is a stationary MDS and as the other terms are $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. By (2.160) and (2.161), $\sum_{k=t}^{\infty} \frac{\partial^2 \tau_k(\theta)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0} \tilde{\xi}_{t-k}(d_0)$ 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 (2.163) are also $o_p(1)$. In contrast, the first term in (2.163) is not a MDS. However, by the same proof as for (2.128) (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, (2.163) is $o_p(1)$. For the third term in (2.162), by plugging in $\tilde{h}_{4,t_{(i,j)}}$, $h_{4,t_{(i,j)}}$

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{4,t_{(i,j)}} - h_{4,t_{(i,j)}}) \\
&\quad= \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\left(\tau_k(\theta_0) - \tau_k(\theta_0,t)\right)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \\
&\quad+ \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\tau_k(\theta_0,t)\left(\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}} - \frac{\partial^2\xi_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\right)\bigg|_{\theta=\theta_0} \\
&\quad+ \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=t}^{\infty}\tau_k(\theta_0)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\bigg|_{\theta=\theta_0},
\end{aligned} \tag{2.164}$$

where the first and third term are MDS when adapted to $\mathcal{F}_t^{\tilde{\xi}}$, as $\tilde{v}_t(\theta_0)$ is a MDS and the remaining term is $\mathcal{F}_{t-1}^{\tilde{\xi}}$-measurable. The third term is $o_p(1)$, because $\sum_{k=t}^{\infty}\tau_k(\theta_0)\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$ is $o_p(1)$ by lemma 2.A.2, and by Hualde and Robinson (2011, lemma 4). The first term is $o_p(1)$ since $(\tau_k(\theta_0) - \tau_k(\theta_0,t))\frac{\partial^2\tilde{\xi}_{t-k}(d)}{\partial\theta_{(i)}\partial\theta_{(j)}}\big|_{\theta=\theta_0}$ is $o_p(1)$ by lemma 2.A.3. The second term can be shown to be $o_p(1)$ analogously to (2.134) 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 (2.162), similarly to (2.163) and (2.164)

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)(\tilde{h}_{5,t_{(i,j)}} - h_{5,t_{(i,j)}}) \\
&\quad= \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=t}^{\infty}\frac{\partial\tau_k(\theta_0)}{\partial\theta_{(i)}}\bigg|_{\theta=\theta_0}\frac{\partial\tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\bigg|_{\theta=\theta_0} \\
&\quad+ \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\frac{\partial\tau_k(\theta,t)}{\partial\theta_{(i)}}\bigg|_{\theta=\theta_0}\left(\frac{\partial\tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}} - \frac{\partial\xi_{t-k}(d)}{\partial\theta_{(j)}}\right)\bigg|_{\theta=\theta_0} \\
&\quad+ \frac{1}{n}\sum_{t=1}^{n}\tilde{v}_t(\theta_0)\sum_{k=1}^{t-1}\left(\frac{\partial\tau_k(\theta)}{\partial\theta_{(i)}} - \frac{\partial\tau_k(\theta,t)}{\partial\theta_{(i)}}\right)\bigg|_{\theta=\theta_0}\frac{\partial\tilde{\xi}_{t-k}(d)}{\partial\theta_{(j)}}\bigg|_{\theta=\theta_0},
\end{aligned} \tag{2.165}$$

where the first and third term are MDS as before. The first term is $o_p(1)$ by lemma 2.A.4, while the third term is $o_p(1)$ by lemma 2.A.5. The second term can be shown to be $o_p(1)$ analogously to (2.134) using (2.137), as the partial derivatives of $\tau_k(\theta,t)$ only add a log-factor to the convergence rates, see lemma 2.A.4. Thus, (2.165) is

also $o_p(1)$. The second, fourth and sixth term in (2.162) can be written as

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \left(\tilde{v}_t(\theta_0) - v_t(\theta_0)\right) \\
=&\frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=0}^{t-1} (\tilde{\xi}_{t-k}(d_0) - \xi_{t-k}(d_0))\tau_k(\theta_0, t) \\
+&\frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=1}^{t-1} (\tau_k(\theta_0) - \tau_k(\theta_0, t))\tilde{\xi}_{t-k}(d_0) \\
+&\frac{1}{n}\sum_{t=1}^{n} h_{l,t_{(i,j)}} \sum_{k=t}^{\infty} \tau_k(\theta_0)\tilde{\xi}_{t-k}(d_0),
\end{aligned} \tag{2.166}$$

with $l = 3, 4, 5$. For $l = 3$, (2.166) only differs from (2.141) 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 (2.160). For $l = 4$, (2.166) is almost identical to (2.148), 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 (2.148) (Hualde and Robinson; 2011, lemma 4). For $l = 5$, (2.166) is again almost identical to (2.148) 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 2.A.4 the derivative again only adds a log-factor to the convergence rate of $\tau_k(\theta, t)$. Thus, it follows directly from (2.141) and (2.148), together with (2.160) and Hualde and Robinson (2011, lemma 4), that (2.166) is $o_p(1)$. The two remaining terms in (2.162) are $o_p(1)$ by (2.165) and (2.166), as i, j can be interchanged. This completes the proof.

□

Chapter 3

Solving the unobserved components puzzle: a fractional approach to measuring the business cycle

3.1 Introduction

Measuring the business cycle plays a key role in applied research, as many macroeconomic models make assumptions about the long- and short-run behavior of real output. In order to verify these assumptions, appropriate methods for decomposing time series into trend and cycle are necessary and are considered in this paper.

For log US real GDP, which is the main application of trend-cycle decompositions (see e.g. Harvey; 1985; Morley et al.; 2003; Morley and Piger; 2012), the results in the literature are puzzling. While empirical evidence supports a strong negative correlation between long- and short-run innovations, both the correlated unobserved components (UC) model as proposed by Balke and Wohar (2002) and Morley et al. (2003), and the decomposition of Beveridge and Nelson (1981), estimate a volatile long-run component along with a noisy cycle, thereby missing the NBER chronology and contradicting macroeconomic common sense.

Since the above models do not provide plausible estimates of the business cycle, some empirical researchers prefer trend-cycle models in the spirit of Hodrick and Prescott (1997), which assume a trend component integrated of order two for log GDP. Although this specification is at odds with both, economic theory and empirical evidence, modeling the trend of log GDP as an $I(2)$ process yields an estimated

variance of the long-run innovations close to zero, making the estimated trend component very smooth. Any dynamics of log GDP away from the smooth trend are attributed to the cycle, yielding rich cyclical oscillations. As a similar solution but with an $I(1)$ specification for the trend, Kamber et al. (2018) suggest fixing the correlated $I(1)$ UC model by restricting the parameter space to the region where the variance-ratio of long- and short-run innovations is small. This forces the $I(1)$ long-run component to be smooth, leaving additional dynamics to be captured by the cycle. However, since the constrained parameter space is a subspace of the parameter space of the correlated $I(1)$ UC model, the question arises as to why the unconstrained optimization yields different parameter estimates corresponding to a higher log likelihood. In summary, both assuming an $I(2)$ trend for GDP, as well as restricting the parameter space of the correlated $I(1)$ UC model to the region where the variance-ratio of long- and short-run innovations is small, clearly yields a misspecified model for log US real GDP, but at the same time provides an economically plausible estimate of the business cycle.

We argue that the puzzling business cycle estimate of the correlated $I(1)$ UC model is an artifact generated by the presence of a smooth fractionally integrated long-run component in log US real GDP with an integration order greater than one but less than two. Misspecifying the integration order upward-biases the variance estimate for the long-run innovations in the correlated $I(1)$ UC model, as the additional memory that is not captured by the $I(1)$ trend feeds directly into the estimated long-run innovations. As a glance at figure 3.2 shows, the periodogram of the long-run innovations has a peak at the origin, indicating that the estimated innovations have long memory. The upwardly biased variance estimate leads to an erratic estimate of the trend component, along with a noisy cycle that adjusts for the fluctuations in the trend. Conversely, misspecifying the trend as an $I(2)$ process produces estimates of the long-run innovations that are anti-persistent, a trend that is too smooth, and attributes too much of the variation in GDP to the cycle.

We contribute to the literature by revisiting the puzzling estimates for the business cycle using the fractional UC model that was derived in chapter 2 of this thesis. Like traditional models, the fractional UC model allows to decompose an observable variable into trend and cycle, but the stochastic part of the trend is modeled as a fractionally integrated process of order $d \in \mathbb{R}_+$, denoted as $x_t \sim I(d)$. It encompasses the above $I(1)$ and $I(2)$ UC models as special cases. Since d can take any value on the real line, the model seamlessly links integer-integrated UC models and allows for intermediate solutions. Parameter estimation is performed by the quasi-maximum

likelihood (QML) estimator, where the integration order is estimated jointly with the other model parameters. This provides a data-driven solution to the specification of the memory of the trend. In addition, the model allows inference about the appropriate specification of the long-run component, and in particular the hypotheses that d equals one or two can be tested. For the cyclical component c_t, we consider several specifications, including both rich and parsimonious ARMA models as well as replacing the traditional lag operator with the fractional lag operator of Granger (1986). Furthermore, we include a deterministic trend component that is a polynomial of order b, $b \in \mathbb{R}_+$, as a generalization of the traditional linear trend. Estimates for trend and cycle are obtained from the analytical solution of the Kalman filter and smoother that was derived in chapter 2 of this thesis.

For empirical researchers, we provide guidance on the appropriate specification of UC models for arguably their most important use case: For log US real GDP, we investigate whether an integer-integrated trend component is appropriate, or whether a generalization to fractionally integrated processes better captures the long-run dynamics. We consider several different specifications for the cyclical component and select the most appropriate specification by minimizing the Bayesian information criterion (BIC). And we comment on whether a deterministic polynomial trend improves the fit as compared to the traditional linear deterministic specification. For the preferred specification in terms of the BIC, we estimate an integration order of about 1.30 for log US real GDP, while rejecting both the $d = 1$ and $d = 2$ hypotheses. This is consistent with the finding that smoothed long-run innovations from $I(1)$ UC models exhibit long memory, while those of $I(2)$ UC models are anti-persistent. Our results suggest a very parsimonious parameterization of the cycle for GDP, and the fractional lag operator improves the fit relative to the traditional lag operator. For the deterministic trend component, our results show that a general polynomial trend has little to no advantage over a linear trend.

The resulting estimate of the business cycle from the selected fractional UC model is well in line with economic common sense: As a glance at figure 3.3 reveals, the estimated cyclical component exhibits the same turning points as the theory-based output gap measure of the US Congressional Budget Office (CBO). It rises gradually in periods of economic recovery and prosperity, and falls sharply during the NBER recession periods. Moreover, we obtain new insights about the business cycle in addition to those of the CBO, most notably an overheating economy in the run-up to the Great Recession as also found by Barigozzi and Luciani (2021) using a high-dimensional dynamic factor model with many macroeconomic indicators as inputs.

Thus, our specification of the fractional UC model can serve as a complement to the theory-based CBO model. Finally, despite the generality of the fractional UC model, we find that long- and short-run innovations are (almost) perfectly correlated. While this aligns well to the empirical results from integer-integrated UC models (Morley et al.; 2003; Iwata and Li; 2015), it implies that long- and short-run innovations cannot be identified separately for US GDP.

The paper is organized as follows. Section 3.2 details the unobserved components puzzle and motivates the need for a fractional UC model, which is introduced in section 3.3. Section 3.4 outlines parameter estimation via the QML estimator, and details the estimation of trend and cycle via the Kalman filter and smoother. Section 3.5 applies the model to log US real GDP, while section 3.6 concludes. Robustness checks and additional figures and tables are included in the appendix.

3.2 The unobserved components puzzle

The UC literature builds on a simple model that decomposes an observable time series $\{y_t\}_{t=1}^n$ into unobserved trend τ_t and cycle c_t

$$y_t = \tau_t + c_t. \tag{3.1}$$

Trend and cycle are disentangled by their different spectra: The cyclical component c_t is a mean zero stationary process that is expected to capture the transitory fluctuations of y_t and is interpreted as the business cycle for y_t being log GDP. The long-run component τ_t, on the other hand, has an autocovariance function that decays more slowly than with an exponential rate, is allowed to be non-stationary, and is expected to capture the long-run dynamics of GDP. The UC literature specifies τ_t as the sum of a stochastic trend component and some deterministic terms, usually a constant plus a linear deterministic trend

$$\tau_t = \mu_0 + \mu_1 t + x_t, \qquad \Delta^d x_t = \eta_t.$$

The difference operator is defined as $\Delta^d = (1-L)^d$, where L is the lag operator, and $d \in \mathbb{N}$. It takes the d-th difference of x_t, and thus $x_t \sim I(d)$ is a stochastic trend with memory d.

The cyclical component is modeled as a mean zero stationary autoregressive process

$$b(L)c_t = \varepsilon_t,$$

and long- and short-run innovations are assumed to be Gaussian white noise with covariance matrix Q

$$\begin{pmatrix} \eta_t \\ \varepsilon_t \end{pmatrix} \sim \text{NID}(0, Q), \qquad Q = \begin{bmatrix} \sigma_\eta^2 & \sigma_{\eta\varepsilon} \\ \sigma_{\eta\varepsilon} & \sigma_\varepsilon^2 \end{bmatrix}.$$

To date, the literature has come up with a variety of different specifications for log GDP that are encompassed by the above UC model. Key differences are the specification of the trend memory d, and whether long- and short-run innovations are allowed to be correlated. The first generation of UC models builds on the seminal work of Harvey (1985) and Clark (1987): Models in the spirit of Harvey (1985) assume that the stochastic trend is a random walk and thus set $d = 1$. They will be called **UC-I(1)** models. In contrast, the double-drift model of Clark (1987) assumes that x_t is a quadratic stochastic trend, setting $d = 2$. Models in the spirit of Clark (1987) will be referred to as **UC-I(2)** models. Both classes of models specify c_t as a stationary, finite, zero mean autoregressive process and force Q to be diagonal.

The second generation of UC models was introduced by Balke and Wohar (2002) and Morley et al. (2003), and modifies the first generation models to allow for contemporaneously correlated long- and short-run innovations. For log US real GDP, Morley et al. (2003) provide extensive evidence that long- and short-run innovations are strongly negatively correlated. Since both Balke and Wohar (2002) and Morley et al. (2003) specify x_t to be a random walk, their models will be called **UC-I(1)-corr** models. Analogously, Oh et al. (2008) generalize the UC-I(2) model to non-diagonal Q. Their model will be referred to as the **UC-I(2)-corr** model.

Besides the first and second generation of UC models, less structural frameworks are proposed by Hodrick and Prescott (1997) and Kamber et al. (2018): The filter of Hodrick and Prescott (1997) assumes $x_t \sim I(2)$, implicitly treats c_t as white noise, and assumes a constant signal-to-noise ratio between long- and short-run innovations that is determined by the tuning parameter of the filter, see the discussion in chapter 2 of this thesis. It is hereafter referred to as the **HP-filter**, and is a special case of the Clark (1987) model, where the parameter space for Q is constrained by setting the variance-ratio of long- and short-run innovations to a certain value. The counterpart for $x_t \sim I(1)$ is derived by Kamber et al. (2018) and is called the **BN-filter** in what follows. Like the HP-filter, it also restricts the parameter space for Q by forcing the variance-ratio of long- and short-run innovations into a certain region. Both HP- and BN-filter are popular among empirical researchers, because the restricted parameter space typically forces the variance of long-run innovations to be small, yielding a

smooth stochastic trend x_t, while at the same time forcing the variance of short-run innovations to be comparatively large, thus attributing rich dynamics to the cyclical component c_t. A summary of all models and their underlying restrictions in terms of (3.1) is given in table 3.1.

	x_t	c_t	Q
UC-I(1)	$\Delta x_t = \eta_t$	$b(L)c_t = \varepsilon_t$	Q diagonal
UC-I(2)	$\Delta^2 x_t = \eta_t$	$b(L)c_t = \varepsilon_t$	Q diagonal
UC-I(1)-corr	$\Delta x_t = \eta_t$	$b(L)c_t = \varepsilon_t$	Q non-diagonal
UC-I(2)-corr	$\Delta^2 x_t = \eta_t$	$b(L)c_t = \varepsilon_t$	Q non-diagonal
BN-filter	$\Delta x_t = \eta_t$	$b(L)c_t = \varepsilon_t$	$\sigma_\eta^2/\sigma_\varepsilon^2$ small
HP-filter	$\Delta^2 x_t = \eta_t$	$c_t = \varepsilon_t$	$\sigma_\eta^2/\sigma_\varepsilon^2$ small

Table 3.1: Restrictions from the different UC models on trend, cycle, and innovations covariance matrix.

As becomes clear from table 3.1, the parameter spaces of the UC-I(1) model and the BN-filter are subspaces of the parameter space of the UC-I(1)-corr model, while the parameter spaces of the UC-I(2) model and the HP-filter are subspaces of the parameter space of the UC-I(2)-corr model. Thus, for the restricted models to be correctly specified for log US real GDP, it is necessary that the globally optimal parameter combinations within the encompassing parameter spaces of the UC-I(1)-corr and the UC-I(2)-corr model are contained in the respective restricted parameter spaces. If that holds, the maximum likelihood estimator for the underlying model parameters has the same optimum for the restricted and the encompassing models. Consequently, also the decomposition into trend and cycle is identical. However, as will be shown below, for log US real GDP, this is neither the case for models with $d = 1$, nor for models with $d = 2$. It is noteworthy that while the restricted models have a smaller log likelihood than their unrestricted counterparts, they yield a decomposition into trend and cycle that is more in line with economic common sense and, in particular, with the NBER chronology. This is what we call the unobserved components puzzle.

To learn more about an appropriate UC model for log US real GDP, we compare the trend-cycle decompositions and the corresponding parameter estimates for the UC models in table 3.1. The data on seasonally adjusted US real GDP come from the Federal Reserve Bank of St. Louis[1], are in quarterly frequency, cover the period from 1947Q1 to 2022Q4 (except for the BN-filter, as explained below), and are log transformed. Parameter estimates and filtered trend and cyclical components for the UC-I(1), the UC-I(2), the UC-I(1)-corr, and the UC-I(2)-corr models are obtained

[1]The series can be downloaded here: https://fred.stlouisfed.org/series/GDPC1

as described in the aforementioned literature: First, Q and $b(L)$ are estimated by maximum likelihood (ML), where the likelihood is based on the prediction error of the Kalman filter (see Harvey; 1989, ch. 3 and 4). The order of the AR polynomials is determined by the Bayesian information criterion, is equal to two for all models, and the chosen lag length is consistent with the literature. Estimates for x_t and c_t are obtained from the Kalman smoother, see Harvey (1989, ch. 3.6). Results for the HP-filter are obtained by setting the tuning parameter to 1600, as usual for quarterly data. Results for the BN-filter are those of Kamber et al. (2018), and thus only consider data up to 2016Q4. We have attempted to update their results with the most recent data, but were unable to reproduce as pronounced cyclical behavior as in Kamber et al. (2018), instead obtaining results very close to those of Morley et al. (2003). Figure 3.1 plots the estimated cyclical component of log US real GDP for the six models, while table 3.3 presents the parameter estimates for the parametric UC models.

Starting with the $I(1)$ models, note that all three decompositions are consistent with the estimates in Harvey and Jäger (1993, fig. 2), Morley et al. (2003, fig. 3), and Kamber et al. (2018, fig. 3). The UC-I(1) model provides an estimate for the business cycle that takes the form of an asymmetric sine curve: It falls sharply during the recession periods and gradually recovers in the aftermath. Surprisingly, the estimated cycle becomes less pronounced when correlation between long- and short-run innovations is allowed for: The UC-I(1)-corr model estimates a rather noisy cycle with no clear pro-cyclical pattern during the economic recovery periods. Instead, the estimated business cycle is characterized by sharp increases just before the recessionary periods, followed by a sharp downturn at the recessions. When the variance-ratio between long- and short-run shocks is restricted to be small, as for the BN-filter, the resulting business cycle estimate again exhibits a more pronounced cyclical behavior than in the UC-I(1) model, but it retains some of the noisy fluctuations of the correlated model. Results from multivariate UC models, such as those of Harvey and Trimbur (2003), Basistha and Nelson (2007), Harvey et al. (2007), and Barigozzi and Luciani (2021), support the pronounced cyclical behavior of the two restricted models rather than the noisy dynamics of the UC-I(1)-corr model.

Turning to the $I(2)$ models, note that the decompositions are again consistent with the literature, see Clark (1987, fig. IVb), Hodrick and Prescott (1997, fig. 1). The UC-I(2) and the UC-I(2)-corr models yield a similar log likelihood and produce almost identical estimates for the cyclical component, although the correlation coefficient is strongly negative whenever it is allowed to deviate from zero. In

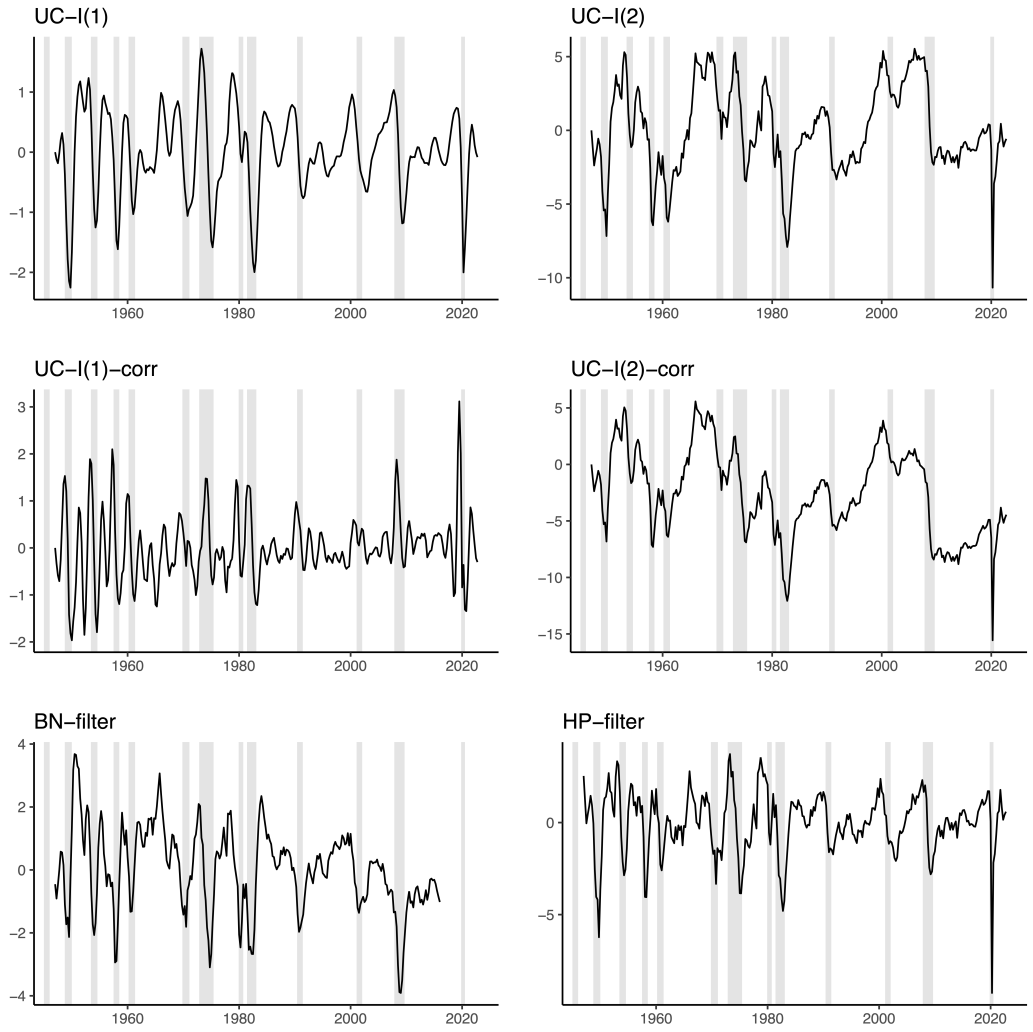

Figure 3.1: Estimated cyclical components $\hat{c}_t$ for the UC-I(1), the UC-I(2), the UC-I(1)-corr, and the UC-I(2)-corr model, as well as for the BN- and HP-filter. The first four models specify c_t as an AR(2) process as suggested by the literature and by the Bayesian information criterion. Estimates for the cyclical components there are obtained from the Kalman smoother. Results for the BN-filter are taken from Kamber et al. (2018). For the HP-filter the tuning parameter is set to 1600, as suggested by the literature. Shaded areas correspond to US recession periods as reported by the NBER.

contrast, the HP-filter attributes comparatively less variation to the cyclical component. Compared to the $I(1)$ models, the cyclical patterns appear to be much more path-dependent, taking longer to return to their mean. This is clearly due to the $I(2)$-specification of the trend, which forces a smaller parameter estimate for σ_η^2 to capture the long-run dynamics of GDP. As a result, the estimated trend becomes smoother, leaving more variation to be captured by the cycle compared to the $I(1)$

models.

Figure 3.1 immediately raises the question of which specification, if any, is appropriate for log US real GDP. The obvious first step in narrowing down the model choice is to look at the long-run component, which is the core difference between the $I(1)$ and $I(2)$ models. If log US real GDP is indeed $I(2)$, the Kalman filter under-differences the observable variable in the $I(1)$ models, and the remaining memory goes into the smoothed long-run shocks. The latter become long-range dependent, and the corresponding estimates for σ_η^2 in the $I(1)$ models are upward-biased to capture the additional variation that is caused by under-differencing of y_t. Conversely, if log US real GDP is $I(1)$, then the $I(2)$ models over-difference y_t, and the smoothed long-run shocks become anti-persistent, while the estimates for σ_η^2 in the $I(2)$ models are downward-biased. A simple way to detect both long-range dependence and anti-persistence is to look at the periodogram of the smoothed long-run innovations: Long-range dependent processes allocate much spectral density at the low frequencies, while anti-persistent processes have little or no spectral mass at the origin. The smoothed periodograms for the smoothed long-run innovations η_t of the different models considered are sketched in figure 3.2.

As can be seen in figure 3.2, the periodogram for the long-run innovations of both $I(1)$ and $I(2)$ trend specifications is anything but flat. For the $I(1)$ models, the innovations appear to be long-range dependent, as the periodograms peak at the origin. This suggests that there is memory left in the smoothed long-run innovations, resulting in a trend that is integrated of order $d > 1$ for log US real GDP. The opposite holds for the UC-I(2) and the UC-I(2)-corr models, where the corresponding periodograms show little or no spectral mass at the origin, indicating that the long-run innovations are anti-persistent. This suggests that log US real GDP is integrated of order $d < 2$. Therefore, figure 3.2 indicates that a fractionally integrated trend of order $1 < d < 2$ may be more appropriate for log US real GDP.[2]

We can further examine this hypothesis by estimating the integration order of the smoothed long-run innovations. Table 3.4 summarizes the integration order estimates from the exact local Whittle estimator of Shimotsu and Phillips (2005) as well as from the estimator of Geweke and Porter-Hudak (1983) for different bandwidth choices. For the two parametric $I(1)$ UC models, the estimates for the integration order of the long-run innovations fall into the interval $[0.09; 0.70]$, while for the two parametric $I(2)$ UC models they fall into $[-1; -0.05]$. Taken together, the integration order estimates support the hypothesis that trend GDP is integrated of order

[2]For readers unfamiliar with fractionally integrated processes, Hassler (2019) provides a good introduction.

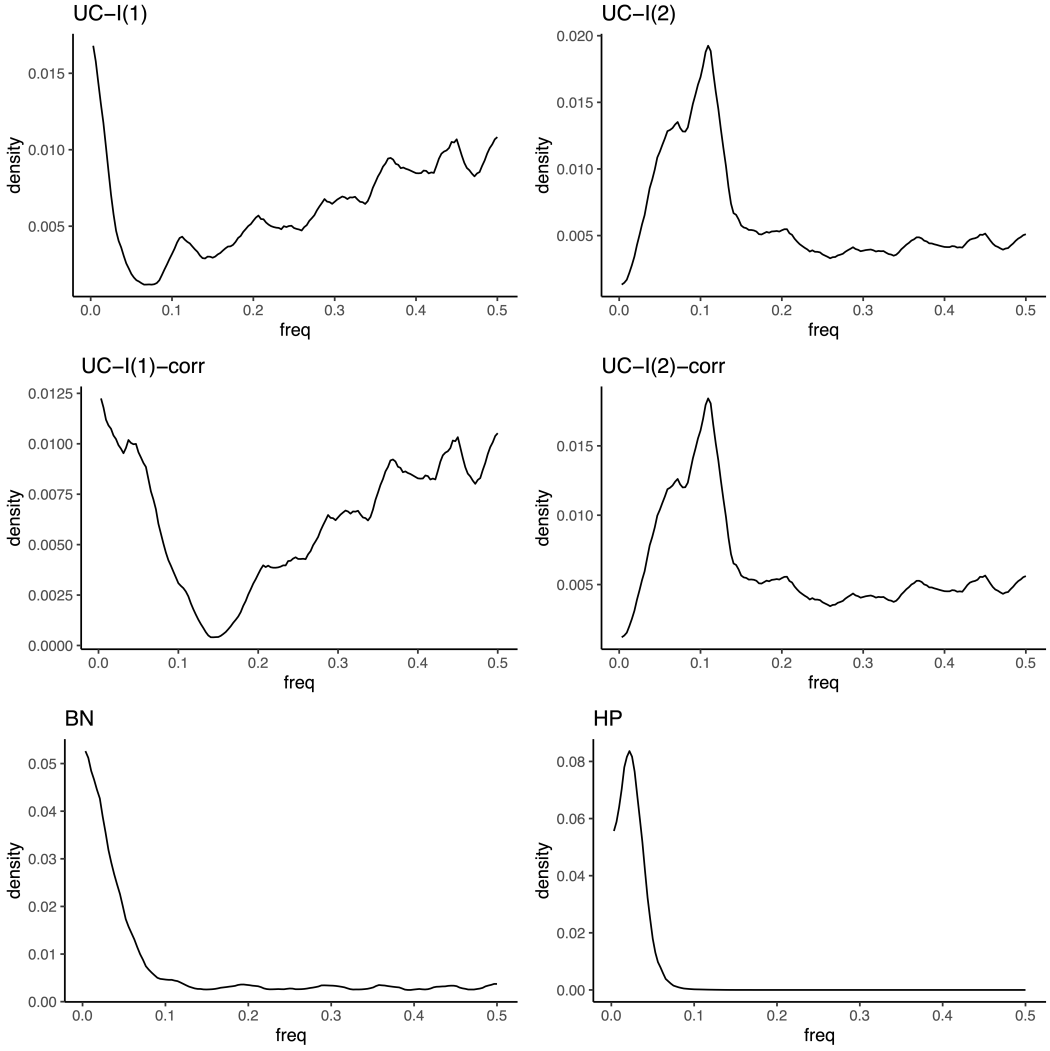

Figure 3.2: Smoothed periodogram for the smoothed long-run innovations for the UC-I(1), the UC-I(2), the UC-I(1)-corr, and the UC-I(2)-corr model, as well as for the BN- and HP-filter.

somewhere between one and two.

The fractional hypothesis also explains the puzzling results in figure 3.1: If the stochastic trend of GDP is integrated of order greater than one, the UC-I(1) and the UC-I(1)-corr model attribute the additional persistence, that is not captured by the $I(1)$ specification, to the long-run shocks η_t, resulting in an upwardly biased estimate of σ_η^2. This bias forces the Kalman filter to attribute additional short-run fluctuations to the trend, resulting in a rather noisy estimate. While the UC-I(1) model prohibits correlation between the long- and short-run innovations, the UC-I(1)-corr model allows the cyclical shocks to adjust for noisy long-run innovations. Therefore, the noisy behavior of the estimated trend innovations spills over directly

into the estimated cyclical innovations, producing an erratic cyclical estimate as shown in figure 3.1. By excluding the region of the parameter space where the variance-ratio between long- and short-run innovations becomes large, the BN-filter appears robust to under-differencing, as the constrained parameter space prohibits a large variance estimate for the trend innovations. However, strong long-range dependence in the long-run innovations lowers the likelihood of the BN-filter that is constructed under the assumption that long- and short-run innovations are white noise.

To finally assess the question on whether log US real GDP is indeed integrated of order between one and two, the next section develops a fractional UC model in which the integration order d of the stochastic trend is no longer restricted to positive integers. Instead, allowing for $d \in \mathbb{R}_+$ seamlessly links $I(1)$ and $I(2)$ UC models and allows for intermediate solutions. The memory parameter d enters the model as an unknown parameter to be estimated, allowing the specification of the trend to be chosen in a data-driven manner.

3.3 The fractional unobserved components model

In order for the trend-cycle decomposition (3.1) to be suitable for fractionally integrated processes, we specify the trend τ_t as the sum of a type II fractionally integrated process of order d, and a deterministic trend μ_t to be determined later

$$\tau_t = \mu_t + x_t, \qquad \Delta_+^d x_t = \eta_t. \tag{3.2}$$

The integration order $d \in D = \{d \in \mathbb{R} | 1/2 < d_{min} \leq d \leq d_{max} < \infty\}$ determines the memory of the trend, while η_t are the long-run shocks to be defined in (3.6) below. The lower bound $d_{min} > 1/2$ is to ensure that x_t is a long-run component, while d_{max} can be arbitrarily large and is required to keep the parameter space bounded. The fractional difference operator Δ_+^d depends only on the integration order d and controls the memory of x_t. Without the +-subscript, it exhibits a polynomial expansion of order infinite in the lag operator L

$$\Delta^d = (1-L)^d = \sum_{j=0}^{\infty} \pi_j(d) L^j, \qquad \pi_j(d) = \begin{cases} \frac{j-d-1}{j} \pi_{j-1}(d) & j = 1, 2, ..., \\ 1 & j = 0. \end{cases} \tag{3.3}$$

The +-subscript denotes the truncation of an operator at $t \leq 0$, $\Delta_+^d x_t = \Delta^d x_t \mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} \pi_j(d) x_{t-j}$, where $\mathbb{1}(t \geq 1)$ is the indicator function taking the value one

for positive subscripts of x_{t-j}, otherwise zero. The use of the truncated fractional difference operator reflects the type II definition of fractionally integrated processes (Marinucci and Robinson; 1999), and is required for non-stationary trends ($d > 1/2$).

The stochastic trend specification in (3.2) encompasses the integer-integrated specifications in the UC literature as considered in the previous section: Setting $d = 1$ makes x_t a random walk as considered by Harvey (1985), Balke and Wohar (2002), Morley et al. (2003), and Kamber et al. (2018). For $d = 2$ one obtains the quadratic stochastic trend specification of Clark (1987), Hodrick and Prescott (1997), and Oh et al. (2008). Furthermore, general integer-integrated trends as studied by Burman and Shumway (2009) are contained for $d \in \mathbb{N}$. Allowing for non-integer d seamlessly links the integer-integrated models and allows for more general patterns of long-run dynamics with memory between the integer-integrated cases: An integration order $1/2 < d < 1$ yields a trend that is asymptotically non-stationary but (conditionally) mean-reverting, while $1 < d < 2$ yields a trend with more memory than the random walk but less than the quadratic stochastic trend. Since the model treats d as an unknown parameter to be estimated, it allows d to be determined in a data-driven manner and does not rely on strong prior assumptions about d. While the empirical macroeconomic literature has, to the best of our knowledge, so far only considered UC models with integer integration orders, stationary long memory models (i.e. $d < 1/2$) are popular in the field of realized volatility modeling, see Ray and Tsay (2000), Chen and Hurvich (2006), and Varneskov and Perron (2018). In chapter 2 of this thesis, the asymptotic estimation theory for the general class of fractional unobserved components models was derived, also allowing for $d \geq 1/2$. We build on this work by studying explicit specifications for trend and cycle, and tailor them to be suitable for log GDP.

For the deterministic term μ_t, we consider a polynomial trend of order b together with an intercept

$$\mu_t = \mu_0 + \mu_1 t^b. \tag{3.4}$$

The exponent $b \in B = \{b \in \mathbb{R} | 0 < b_{min} \leq b \leq b_{max} < \infty\}$ determines the shape of the deterministic component. While $b = 1$ yields a linear trend as typically assumed in the UC literature for log GDP, setting $b = d$ yields a deterministic trend of the same order as the stochastic long-run component. The latter is equivalent to including a non-zero constant in $\Delta_+^d x_t$. Alternatively, b can also be treated as an additional parameter to be estimated, allowing the order of the deterministic trend to be determined by the data.

Turning to the transitory component, we allow for an ARMA(p, q) process in the fractional lag operator L_δ

$$\phi(L_\delta, \varphi)c_t = \frac{a(L_\delta)}{m(L_\delta)}c_t = \varepsilon_t, \tag{3.5}$$

where $a(L_\delta) = 1 - \sum_{j=1}^{p} a_j L_\delta^j$ is a stable AR polynomial in the fractional lag operator L_δ as introduced by Granger (1986), $b(L_\delta) = 1 + \sum_{j=1}^{q} m_j L_\delta^j$ is an invertible MA polynomial in the fractional lag operator L_δ, $\phi(L_\delta, \varphi) = 1 - \sum_{j=1}^{\infty} \phi_j(\varphi) L_\delta^j$, $L_\delta = 1 - \Delta_+^\delta$ is the fractional lag operator with $\delta \in C = \{\delta \in \mathbb{R} | 0 < \delta_{min} \leq \delta \leq \delta_{max} < \infty\}$, and $\varphi = (\delta, a_1, ..., a_p, m_1, ..., m_q)$ holds the relevant parameters governing $\phi(L_\delta, \varphi)$. ε_t are the short-run innovations to be defined in (3.6) below. For stability of the fractional lag polynomials $a(L_\delta)$ and $m(L_\delta)$, the condition of Johansen (2008, cor. 6) is required to hold and is imposed in what follows. It implies that the roots of $|a(z)| = 0$ and $|m(z)| = 0$ lie outside the image $\mathbb{C}_\delta$ of the unit disk under the mapping $z \mapsto 1 - (1 - z)^\delta$. It follows immediately that $\phi(L_\delta, \varphi)c_t$ is stationary for all feasible φ. In fractional models L_δ plays the role of the standard lag operator $L_1 = L$, since $(1 - L_\delta) = \Delta_+^\delta$. While, for an arbitrary $I(0)$ process z_t, the standard lag operator $Lz_t = (1 - (1 - L))z_t = z_t - \Delta z_t$ subtracts an $I(-1)$ process from z_t, the fractional lag operator $L_\delta z_t = (1 - (1 - L_\delta))z_t = z_t - \Delta_+^\delta z_t$ subtracts an $I(-\delta)$ process from z_t. In addition, $L_\delta z_t = -\sum_{j=1}^{t-1} \pi_j(\delta) z_{t-j}$ is a weighted sum of past z_t, and thus L_δ qualifies as a lag operator. By definition, the polynomial $\phi(L_\delta, \varphi)$ preserves the integration order of a series since $\delta > 0$.

Turning to the long- and short-run shocks η_t, ε_t, we assume that they are mean-zero stationary and ergodic martingale difference sequences when adapted to their joint natural filtration $\mathcal{F}_t = \sigma\left((\eta_s, \varepsilon_s), s \leq t\right)$, and their autocovariance functions are assumed to be absolutely summable. Conditional of $\mathcal{F}_{t-1}$, their third and fourth moments are assumed to be finite and equal to their unconditional moments. (η_t, ε_t) may exhibit a non-diagonal covariance matrix Q, which implies that

$$\begin{aligned} \mathrm{E}\begin{pmatrix}\eta_t \\ \varepsilon_t\end{pmatrix} = 0, \quad \mathrm{Var}\begin{pmatrix}\eta_t \\ \varepsilon_t\end{pmatrix} = Q = \begin{bmatrix}\sigma_\eta^2 & \sigma_{\eta\varepsilon} \\ \sigma_{\eta\varepsilon} & \sigma_\varepsilon^2\end{bmatrix}, \\ \mathrm{Cov}\left[\begin{pmatrix}\eta_t \\ \varepsilon_t\end{pmatrix}, \begin{pmatrix}\eta_{t-s} \\ \varepsilon_{t-s}\end{pmatrix}\right] = 0_{2,2} \ \forall s \neq 0. \end{aligned} \tag{3.6}$$

The assumptions about (η_t, ε_t) are identical to those in chapter 2 of this thesis, and are necessary for consistency and asymptotic normality of the parameter estimates as discussed in section 3.4. They are somewhat more general than most of the literature

on UC models, which assumes the shocks to be Gaussian white noise (e.g. Clark; 1987; Morley et al.; 2003).

Since different parameterizations of the model will be considered later in the application, it will be helpful to refer to the model in (3.1), (3.2), (3.4), (3.5), and (3.6) as a UC(d, b, δ, ρ) model. It follows immediately from this convention that the model nests the different UC models in the previous section: The model of Harvey (1985) is a UC(1, 1, 1, 0) model, while allowing for correlated shocks as in Balke and Wohar (2002), Morley et al. (2003), and Weber (2011) yields a UC(1, 1, 1, ρ) model. The double-drift model of Clark (1987) is a UC(2, 1, 1, 0) model, which is generalized to allow for correlated innovations by Oh et al. (2008).

A key property of integer-integrated UC models is that they encompass the decomposition of Beveridge and Nelson (1981) for $\rho = -1$, see Proietti (2004, 2006), and Oh et al. (2008). This carries over to the fractional UC(d, b, d, -1) model, which can be interpreted as a generalization of the Beveridge-Nelson decomposition to the fractional domain. To see this, assume $\rho = -1$, substitute (3.2), (3.4), and (3.5) into (3.1), replace $\varepsilon_t = -(\sigma_\varepsilon/\sigma_\eta)\eta_t$, and take fractional differences. Then

$$\Delta_+^d(y_t - \mu_t) = \eta_t - \Delta_+^d \phi(L_d, \varphi)^{-1} \frac{\sigma_\varepsilon}{\sigma_\eta}\eta_t = \eta_t - (1 - L_d)\psi(L_d, \varphi)\frac{\sigma_\varepsilon}{\sigma_\eta}\varepsilon_t, \tag{3.7}$$

where $\phi(L_d, \varphi)^{-1} = \psi(L_d, \varphi) = 1 + \psi_1 L_d + \psi_2 L_d^2 + ...$, with $\psi_1(\varphi) = \phi_1(\varphi)$, $\psi_j(\varphi) = \phi_j(\varphi) + \sum_{k=1}^{j-1} \psi_k(\varphi)\phi_{j-k}(\varphi)$, $j \geq 2$. $\psi(L_d, \varphi)$ exists since $\phi(L_d, \varphi)$ is stable. By the aggregation properties of white noise processes (Granger and Morris; 1976, p. 248f), (3.7) is an ARFIMA model in the fractional lag operator L_d

$$\Delta_+^d(y_t - \mu_t) = \tilde{\psi}(L_d, \varphi)u_t, \qquad u_t \sim \text{WN}\left(0, (\sigma_\eta - \sigma_\varepsilon)^2\right), \tag{3.8}$$

where $\tilde{\psi}(L_d, \varphi) = 1 + \tilde{\psi}_1(\varphi)L_d + \tilde{\psi}_2(\varphi)L_d^2 + ...$, $\tilde{\psi}_0 = 1$, $\tilde{\psi}_1(\varphi) = \sigma_\varepsilon(1 - \psi_1(\varphi))/(\sigma_\eta - \sigma_\varepsilon)$, $\tilde{\psi}_j(\varphi) = \sigma_\varepsilon(\psi_{j-1}(\varphi) - \psi_j(\varphi))/(\sigma_\eta - \sigma_\varepsilon)$ for all $j \geq 2$, and $u_t = \eta_t(1 - \sigma_\varepsilon/\sigma_\eta)$. The Beveridge-Nelson decomposition of (3.8) follows from noting that

$$\tilde{\psi}(L_d, \varphi)u_t = \tilde{\psi}(1, \varphi)u_t - (1 - L_d)\sum_{k=0} L_d^k u_t \sum_{j=k+1} \tilde{\psi}_j(\varphi), \tag{3.9}$$

where

$$\begin{aligned}\tilde{\psi}(1,\varphi)u_t &= \left[1 + \frac{\sigma_\varepsilon}{\sigma_\eta - \sigma_\varepsilon}(1-\psi_1(\varphi)) + \frac{\sigma_\varepsilon}{\sigma_\eta - \sigma_\varepsilon}\sum_{j=2}(\psi_{j-1}(\varphi) - \psi_j(\varphi))\right] u_t \\ &= \left(1 + \frac{\sigma_\varepsilon}{\sigma_\eta - \sigma_\varepsilon}\right) u_t = \eta_t.\end{aligned}$$

Consequently, $\tilde{\psi}(1,\varphi)u_t$ in (3.9) is the fractionally differenced trend from the Beveridge-Nelson decomposition and equals the long-run innovations from the UC(d, b, d, -1) model, while $-(1-L_d)\sum_{k=0} L_d^k u_t \sum_{j=k+1}\tilde{\psi}_j(\varphi)$ is the fractionally differenced cycle from the Beveridge-Nelson decomposition and equals $\Delta_+^d c_t$ of the UC(d, b, d, -1) model. Un-taking fractional differences then generalizes the Beveridge-Nelson decomposition to fractionally integrated processes, where

$$x_t^{BN} = \Delta_+^{-d}\tilde{\psi}(1,\varphi)u_t = \Delta_+^{-d}\eta_t = x_t, \qquad c_t^{BN} = -\sum_{k=0} L_d^k u_t \sum_{j=k+1}\tilde{\psi}_j(\varphi) = c_t.$$

3.4 Estimation

Having introduced the fractional UC model, we now turn to the estimation of the model parameters and the latent components. For this, let $\theta = (d, b, \delta, \varphi', \sigma_\eta^2, \sigma_{\eta\varepsilon}, \sigma_\varepsilon^2)'$ denote the vector collecting all model parameters of the fractional UC model, and let $\theta_0 = (d_0, b_0, \delta_0, \varphi_0', \sigma_{\eta,0}^2, \sigma_{\eta\varepsilon,0}, \sigma_{\varepsilon,0}^2)'$ denote the true parameters of the data-generating mechanism to be estimated. The parameters of the fractional UC model are estimated by the quasi-maximum likelihood (QML) estimator as derived in section 2 of this thesis. Trend and cycle are estimated by the Kalman filter and smoother, for which an analytical solution was derived in chapter 2 that is computationally superior to the Kalman recursions for fractional UC models. In the following, we summarize the main results of chapter 2 about the estimation of fractional UC models and discuss identification.

In subsection 3.4.1, we first show that our fractional UC model in (3.1), (3.2), (3.4), and (3.5) is a state space model. Therefore, the Kalman recursions can be applied and allow to filter, predict, and smooth the unobserved x_t and c_t, which is the core of subsection 3.4.2. The filtered values are the projections of x_t and c_t onto the space of $y_1, ..., y_t$, i.e. the data observable at period t, again conditional on some realization of the parameter vector θ. Based on (3.2) and (3.5), the one-step ahead predictions for x_{t+1} and c_{t+1} (given $y_1, ..., y_t$, θ) can be obtained, which is referred to as the prediction step of the Kalman filter in the state space literature. Estimates

for x_t and c_t conditional on the full sample information $y_1, ..., y_n$ are referred to as smoothed values and are obtained by projecting x_t and c_t onto the space of $y_1, ..., y_n$, conditional on some realization of the parameter vector θ. Subtracting the one-step-ahead predictions for x_{t+1} and c_{t+1} from y_{t+1} yields the prediction error, based on which the quasi-likelihood function for parameter estimation can be set up, as done in subsection 3.4.3. There, we also discuss the asymptotic theory for the QML estimator of fractional UC models as derived in chapter 2 of this thesis, as well as the identification of the fractional UC model.

In the following, collect $x_{t:1} = (x_t, ..., x_1)'$, $c_{t:1} = (c_t, ..., c_1)'$, $\mu_{t:1} = (\mu_t, ..., \mu_1)'$, and $y_{t:1} = (y_t, ..., y_1)'$ in the respective t-vectors.

3.4.1 State space form

The state space form is a special case of the more general model considered in chapter 2 of this thesis, and is set up analogously. Define $\tilde{\phi}(L, \varphi) = 1 - \sum_{j=1}^{\infty} \tilde{\phi}_j(\varphi) L^j = 1 - \sum_{j=1}^{p} \phi_j L_d^j$ as the representation of $\phi(L_\delta, \varphi)$ in the standard lag operator L, and denote $\tilde{\phi}_+(L, \varphi) c_t = \tilde{\phi}(L, \varphi) c_t \mathbb{1}(t \geq 1) = c_t - \sum_{j=1}^{t-1} \tilde{\phi}_j(\varphi) c_{t-j}$ as the truncated $\tilde{\phi}(L, \varphi)$ polynomial, where $\mathbb{1}(t \geq 1)$ is the indicator function, which takes the value one for positive subscripts of c_{t-j}, otherwise zero. The truncation takes into account that y_t is only observable for positive t, and thus parameter and unobserved components estimation can only be carried out for the observable $\{y_t\}_{t=1}^n$. The state space representation of the fractional UC model is

$$y_t = \mu_t + Z\alpha_t, \tag{3.10}$$

$$\alpha_t = T\alpha_{t-1} + R\zeta_t, \tag{3.11}$$

where the state vector can be partitioned into $\alpha_t = (\alpha_t^{(x)'}, \alpha_t^{(c)'})'$, with $(n-1)$-vectors $\alpha_t^{(x)} = (x_t, x_{t-1}, ..., x_{t-n+2})'$ for the stochastic trend, and $\alpha_t^{(c)} = (c_t, c_{t-1}, ..., c_{t-n+2})'$ for the cycle. The observation matrix $Z = (Z^{(x)}, Z^{(c)})$ consists of the $(n-1)$-dimensional row vectors $Z^{(x)} = (1, 0, ..., 0)$ and $Z^{(c)} = (1, 0, ..., 0)$, which pick the first entry of $\alpha_t^{(x)}$ and $\alpha_t^{(c)}$. The transition equation (3.11) is specified via $T = \text{diag}(T^{(x)}, T^{(c)})$, $R = \text{diag}(R^{(x)}, R^{(c)})$, where

$$T^{(x)} = \begin{bmatrix} -\pi_1(d) & -\pi_2(d) & \cdots & -\pi_{n-1}(d) \\ 1 & & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix}, \quad T^{(c)} = \begin{bmatrix} \tilde{\phi}_1(\varphi) & \tilde{\phi}_2(\varphi) & \cdots & \tilde{\phi}_{n-1}(\varphi) \\ 1 & & & 0 \\ \vdots & \ddots & & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix},$$

and $R^{(x)} = (1, 0, ..., 0)'$, $R^{(c)} = (1, 0, ..., 0)'$ are $(n-1)$-vectors picking the respective entries of $\zeta_t = (\eta_t, \varepsilon_t)'$, and $\text{Var}(\zeta_t) = Q$. Note that whenever $\tilde{\phi}(L, \varphi)$ is a polynomial of order $r < n - 1$ (e.g. for $\delta = 1$, $p < n$, $q = 0$), there exists a minimal state space representation where $\alpha_t^{(c)}, Z^{(c)}, R^{(c)}$ are vectors of dimension r, and $T^{(c)}$ is $r \times r$. The same holds whenever d is an integer, since x_t then admits a d-dimensional state space representation. The system is initialized deterministically with $x_j, c_j = 0$ for all $j \leq 0$, and the deterministic terms are placed directly in the observations equation (3.10), as this ensures stabilisability of the model (see Harvey; 1989, ch. 4.2.5). They are estimated using the GLS estimator as discussed at the end of subsection 3.4.2.

3.4.2 Filtering and smoothing

By the state space representation of the fractional UC model in (3.10) and (3.11), the Kalman recursions can be used to filter, predict, and smooth the latent x_t and c_t (see e.g. Harvey; 1989, ch. 3). However, as argued in chapter 2 of this thesis, instead of computing the filtered states recursively via the Kalman filter, one can also derive an analytical solution to the optimization problem of the Kalman filter. Both approaches yield the identical filtered components, however for the fractional UC model the analytical solution is computationally much simpler, see the discussion in chapter 2 of this thesis. Given the high dimension of the state vector when x_t is a fractionally integrated trend (see subsection 3.4.1), reducing the computational complexity is an important issue for fractional UC models, and thus we briefly outline the analytical solution to the Kalman filter optimization problem below. To arrive at the analytical solution, consider the optimization problem of the Kalman filter, which is obtained by minimizing the concentrated joint quasi-log likelihood of $\{(\eta_j, \varepsilon_j)'\}_{j=1}^t$

$$\begin{aligned}\hat{x}_{t:1}(y_{t:1}, \theta) = \arg\min_{x_{t:1}} \frac{1}{t} \sum_{j=1}^{t} \left[\begin{pmatrix} \eta_j & \varepsilon_j \end{pmatrix} Q^{-1} \begin{pmatrix} \eta_j \\ \varepsilon_j \end{pmatrix} \right] = \arg\min_{x_{t:1}} \frac{1}{t} \frac{1}{\sigma_\eta^2 \sigma_\varepsilon^2 - \sigma_{\eta\varepsilon}^2} \\ \times \sum_{j=1}^{t} \Big[\sigma_\varepsilon^2 (\Delta_+^d x_j)^2 - 2\sigma_{\eta\epsilon} \Delta_+^d x_j \tilde{\phi}_+(L, \varphi)(y_j - \mu_j - x_j) \\ + \sigma_\eta^2 (\tilde{\phi}_+(L, \varphi)(y_j - \mu_j - x_j))^2 \Big],\end{aligned}$$

where the second equality follows from inserting $\eta_j = \Delta_+^d x_j$ via (3.2), as well as $\varepsilon_j = \tilde{\phi}_+(L, \varphi)(y_j - \mu_j - x_j)$ via (3.1) and (3.5). For a matrix representation of the optimization problem, introduce the $t \times t$ difference matrix $S_{d,t}$, and the $t \times t$

coefficient matrix $B_{\varphi,t}$ analogously to (2.5) in chapter 2

$$\begin{aligned} S_{d,t} &= \begin{bmatrix} \pi_0(d) & \pi_1(d) & \cdots & \pi_{t-1}(d) \\ 0 & \pi_0(d) & \cdots & \pi_{t-2}(d) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_0(d) \end{bmatrix}, \\ B_{\varphi,t} &= \begin{bmatrix} 1 & -\tilde{\phi}_1(\varphi) & \cdots & -\tilde{\phi}_{t-1}(\varphi) \\ 0 & 1 & \cdots & -\tilde{\phi}_{t-2}(\varphi) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \end{aligned} \tag{3.12}$$

Then $S_{d,t}x_{t:1} = (\Delta_+^d x_t, ..., \Delta_+^d x_1)'$, and $B_{\varphi,t}c_{t:1} = (\tilde{\phi}_+(L,\varphi)c_t, ..., \tilde{\phi}_+(L,\varphi)c_1)'$. Omitting the constant fraction, the optimization problem for $x_{t:1}$ becomes

$$\begin{aligned} \hat{x}_{t:1}(y_{t:1},\theta) = \arg\min_{x_{t:1}} \frac{1}{t}\Big[&\sigma_\epsilon^2 x_{t:1}' S_{d,t}' S_{d,t} x_{t:1} - 2\sigma_{\eta\epsilon}(y_{t:1} - x_{t:1} - \mu_{t:1})' B_{\varphi,t}' S_{d,t} x_{t:1} \\ &+ \sigma_\eta^2 (y_{t:1} - x_{t:1} - \mu_{t:1})' B_{\varphi,t}' B_{\varphi,t} (y_{t:1} - x_{t:1} - \mu_{t:1})\Big], \end{aligned}$$

which yields the analytical solution to the optimization problem of the Kalman filter as derived in chapter 2 of this thesis

$$\begin{aligned} \hat{x}_{t:1}(y_{t:1},\theta) = &\left[\sigma_\eta^2 B_{\varphi,t}' B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}' B_{\varphi,t} + B_{\varphi,t}' S_{d,t}) + \sigma_\epsilon^2 S_{d,t}' S_{d,t}\right]^{-1} \\ &\times \left(\sigma_\eta^2 B_{\varphi,t}' B_{\varphi,t} + \sigma_{\eta\epsilon} S_{d,t}' B_{\varphi,t}\right)(y_{t:1} - \mu_{t:1}). \end{aligned} \tag{3.13}$$

Either analogously, or by using $\hat{c}_{t:1}(y_{t:1},\theta) = y_{t:1} - \mu_{t:1} - \hat{x}_{t:1}(y_{t:1},\theta)$, the filtered cycle is

$$\begin{aligned} \hat{c}_{t:1}(y_{t:1},\theta) = &\left[\sigma_\eta^2 B_{\varphi,t}' B_{\varphi,t} + \sigma_{\eta\epsilon}(S_{d,t}' B_{\varphi,t} + B_{\varphi,t}' S_{d,t}) + \sigma_\epsilon^2 S_{d,t}' S_{d,t}\right]^{-1} \\ &\times \left(\sigma_\epsilon^2 S_{d,t}' S_{d,t} + \sigma_{\eta\epsilon} B_{\varphi,t}' S_{d,t}\right)(y_{t:1} - \mu_{t:1}). \end{aligned} \tag{3.14}$$

The one-step ahead predictions for trend and cycle are then obtained by rolling the transition equations for trend and cycle (3.2) and (3.5) one period ahead, i.e.

$$\hat{x}_{t+1}(y_{t:1},\theta) = -\begin{pmatrix} \pi_1(d) & \cdots & \pi_t(d) \end{pmatrix} \hat{x}_{t:1}(y_{t:1},\theta), \tag{3.15}$$

$$\hat{c}_{t+1}(y_{t:1},\theta) = \begin{pmatrix} \tilde{\phi}_1(\varphi) & \cdots & \tilde{\phi}_t(\varphi) \end{pmatrix} \hat{c}_{t:1}(y_{t:1},\theta). \tag{3.16}$$

Finally, the prediction error is obtained by subtracting the one-step ahead predictions

$$v_{t+1}(\theta) = y_{t+1} - \mu_{t+1} - \hat{x}_{t+1}(y_{t:1}, \theta) - \hat{c}_{t+1}(y_{t:1}, \theta). \tag{3.17}$$

In practice, computing (3.17) as well as the $\hat{x}_{t+1}(y_{t:1}, \theta)$ and $\hat{c}_{t+1}(y_{t:1}, \theta)$ requires knowledge about $\mu_{(t+1):1}$, which is unobservable. However, since the Kalman filter is a linear operation, we can write (3.17) as $v_{t+1}(\theta) = F(\theta, t+1)(y_{(t+1):1} - \mu_{(t+1):1})$, where $F(\theta, t+1)y_{(t+1):1}$ would be the prediction error of the Kalman filter if the deterministic terms were zero. Since $v_{t+1}(\theta) = F(\theta, t+1)y_{(t+1):1} - F(\theta, t+1)\mu_{(t+1):1}$, the same filter can be applied separately to the observations $y_{(t+1):1}$ and the deterministic terms $\mu_{(t+1):1}$, see Harvey (1989, ch. 3.4.2). An estimate for the coefficients μ_0 and μ_1 in (3.4) is then obtained by regressing $F(\theta, t+1)y_{(t+1):1}$ on $F(\theta, t+1)M_{(t+1):1}$, where $M_{(t+1):1}$ is the $(t+1) \times 2$ regressor matrix of $\mu_{(t+1):1} = M_{(t+1):1}(\mu_0, \mu_1)'$, and the resulting estimator for μ_0, μ_1 is the GLS estimator.

3.4.3 Parameter estimation and identification

To estimate θ_0, the QML estimator is set up based on the prediction error $v_{t+1}(\theta)$ as defined in the previous subsection. Let $\sigma^2_{v_t}(\theta)$ denote the (hypothetical) variance of the prediction error $v_t(\theta)$ for a given parameter vector θ. $\sigma^2_{v_t}(\theta)$ depends only on θ, is independent of $y_1, ..., y_n$, and can be calculated recursively via the Kalman recursions for the prediction error variance as given in Harvey (1989, ch. 3.2). Since the state space model in (3.10) and (3.11) is both detectable and stabilizable, it follows that the Kalman recursions for $\sigma^2_{v_t}(\theta)$ converge to the steady state value $\sigma^2_v(\theta)$ at an exponential rate. Typically, only a few iterations are required until $\sigma^2_{v_t}(\theta)$ is sufficiently close to its steady state, where the Kalman recursions can be terminated and $\sigma^2_{v_t}(\theta)$ can be assumed to be constant from then on. The QML estimator for θ_0 is then constructed based on the prediction error $v_t(\theta)$ and is given by

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta} \log L(\theta), \\ \log L(\theta) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n} \log \sigma^2_{v_t} - \frac{1}{2}\sum_{t=1}^{n} \frac{v_t^2(\theta)}{\sigma^2_{v_t}}. \end{aligned} \tag{3.18}$$

The asymptotic theory for the QML estimator of fractional UC models was derived in chapter 2 of this thesis and carries over to (3.18) upon verification of assumptions 1 to 5 of chapter 2. The assumptions we make in section 3.3 about the long- and short-run innovations as well as about the parameters are identical to assump-

tions 1 and 2 of chapter 2 and are thus satisfied. Assumption 3 requires that the cyclical polynomial $\phi(L_\delta, \varphi)$ in (3.5) is stable with stable partial derivatives and is satisfied for any stationary and invertible ARMA polynomial, both in the standard lag operator L, as well as in the fractional lag operator L_δ (under the additional condition of Johansen (2008, cor. 6) as imposed below (3.5)). Thus, assumptions 1 to 3 of chapter 2 hold, so that the QML estimator for θ_0 is consistent. Furthermore, (3.18) is asymptotically normally distributed under the additional assumptions 4 and 5 of chapter 2: Assumption 4 strengthens the smoothness assumption on $\phi(L_\delta)$ and is again satisfied for all stationary and invertible ARMA polynomials. Assumption 5 essentially requires that the Kalman filter asymptotically becomes the best predictor for y_t given $\mathcal{F}_{t-1}$ as $t \to \infty$, forcing the prediction error to converge to a martingale difference sequence when adapted to the filtration of all past y_s, $s < t$. While assumption 5 follows immediately when long- and short-run innovations are assumed to be Gaussian, it cannot be verified for non-Gaussian shocks. Therefore, when interpreting the standard errors in our application, we assume assumption 5 of chapter 2 to be satisfied.

Finally, it should be noted that the results on consistency and asymptotic normality as derived in chapter 2 are conditional on the model being identified. While identification is a crucial problem in the traditional UC literature that is discussed among others by Morley et al. (2003), Oh et al. (2008), and Trenkler and Weber (2016), it is less of an issue for the fractional UC model as will become clear.

To illustrate that the fractional UC model is identified under comparatively weak conditions as compared to traditional UC models, let $\tilde{a}(L, \varphi) = 1 - \sum_{j=1}^{\infty} \tilde{a}_j(\varphi) L^j = a(L_\delta) = 1 - \sum_{j=1}^{\infty} a_j L_\delta^j$ and $\tilde{m}(L, \varphi) = 1 + \sum_{j=1}^{\infty} \tilde{m}_j(\varphi) L^j = m(L_\delta) = 1 - \sum_{j=1}^{\infty} m_j L_\delta^j$ denote the cyclical AR and MA polynomials of (3.5) in the standard lag operator L, and note that $\tilde{a}_j(\varphi) = a_j$ as well as $\tilde{m}_j(\varphi) = m_j$ for all $j = 1, 2, ...$ whenever $\delta = 1$. Moreover, let $\tilde{a}_+(L, \varphi) = \tilde{a}(L, \varphi)\mathbb{1}(t \geq 1)$ and $\tilde{b}_+(L, \varphi) = \tilde{b}(L, \varphi)\mathbb{1}(t \geq 1)$ denote the truncated polynomials that take into account that y_t is only observable for $t \geq 1$. Then the (truncated) reduced form of the fractional UC model is

$$\tilde{a}_+(L, \varphi)\Delta_+^d(y_t - \mu_t) = \tilde{a}_+(L, \varphi)\eta_t + \tilde{m}_+(L, \varphi)\Delta_+^d \varepsilon_t = b_+(L, \theta)\epsilon_t, \tag{3.19}$$

which is obtained by plugging (3.2) and (3.5) into (3.1), taking fractional differences, and multiplying both sides by $\tilde{a}_+(L, \varphi)$. By the aggregation properties of moving average processes (see Granger and Morris; 1976, p. 248f), (3.19) equals the moving average process $b_+(L, \theta)\epsilon_t = \sum_{j=0}^{t-1} b_j(L, \theta)\epsilon_{t-j}$, $\mathrm{Var}(\epsilon_t) = \sigma_\epsilon^2$, whose order depends on d, as well as on φ. Obviously, d, δ, $a_1, ..., a_p$ are identified based on the observ-

able left-hand side of (3.19), and thus identifiability crucially depends on whether $m_1, ..., m_q$, and σ_η^2, $\sigma_{\eta\varepsilon}$, σ_ε^2 can be recovered from the right-hand side of (3.19). For this purpose, let $\gamma_j(\theta) = \text{Cov}(\tilde{a}_+(L,\varphi)\Delta_+^d(y_t - \mu_t), \tilde{a}_+(L,\varphi)\Delta_+^d(y_{t-j} - \mu_{t-j}))$ denote the j-th autocovariance of (3.19), so that for all $j < t$

$$\begin{aligned}
\gamma_j(\theta) &= \sigma_\eta^2 \sum_{k=0}^{t-j-1} \tilde{a}_k(\varphi)\tilde{a}_{k+j}(\varphi) \\
&\quad + \sigma_\varepsilon^2 \sum_{k=0}^{t-j-1} \left(\sum_{l=0}^{j+k} \tilde{m}_{j+k-l}(\varphi)\pi_l(d) \right) \left(\sum_{l=0}^{k} \tilde{m}_{k-l}(\varphi)\pi_l(d) \right) \\
&\quad + \sigma_{\eta\varepsilon} \Bigg[\sum_{k=0}^{t-j-1} \tilde{a}_k(\varphi) \sum_{l=0}^{j+k} \tilde{m}_{j+k-l}(\varphi)\pi_l(d) \\
&\qquad\qquad + \sum_{k=0}^{t-j-1} \tilde{a}_{j+k}(\varphi) \sum_{l=0}^{k} \tilde{m}_{k-l}(\varphi)\pi_l(d) \Bigg] \\
&= \sigma_\epsilon^2 \sum_{k=0}^{t-j-1} b_k(\theta) b_{k+j}(\theta).
\end{aligned} \tag{3.20}$$

Obviously, for $d \notin \mathbb{N}$, the autocovariance $\gamma_j(\theta) \neq 0$ for all $j = 0, 1, ..., t-1$. Since $t = 1, ..., n$, indexes the observable y_t, we have $n+1$ equations to identify the $(q+3)$ parameters $(m_1, ..., m_q, \sigma_\eta^2, \sigma_{\eta\varepsilon}, \sigma_\varepsilon^2)$. It only remains to be checked whether the contribution of the different parameters $(m_1, ..., m_q, \sigma_\eta^2, \sigma_{\eta\varepsilon}, \sigma_\varepsilon^2)$ to $\gamma_j(\theta)$ is nonzero for $j = 0, 1, ..., q+3$.

Let $d \notin \mathbb{N}$. Then σ_ε^2 enters (3.20) only via the term in the second row of (3.20), which is non-zero for all $j < t$. Moreover, $\sigma_{\eta\varepsilon}$ enters (3.20) via the term in the third and fourth row of (3.20), which is again non-zero for all $j < t$. σ_η^2 enters (3.20) via the term in the first row of (3.20), which is non-zero for $j = 0$, and for all $j = 1, ..., \tilde{p}$, where $\tilde{p}$ is the number of non-zero $\tilde{a}_1(\varphi), ..., \tilde{a}_{\tilde{p}}(\varphi) \neq 0$. Finally, $m_1, ..., m_q$ enter (3.20) via the term in the second row and the term in the third and fourth row of (3.20), both of which are non-zero for all $j < t$. Consequently, for non-integer d, the contribution of $m_1, ..., m_q$ and σ_ε^2, $\sigma_{\eta\varepsilon}$ to $\gamma_j(\theta)$ is non-zero for all $j < t$, while the contribution of σ_η^2 to $\gamma_j(\theta)$ is non-zero at least for $j = 0$. It follows that for non-integer d, the fractional UC model is identified for any reasonable choice of the lag orders of the cyclical ARMA polynomials $p \geq 0$, $q \geq 0$. However, for high q, the identification of the MA parameters becomes weak because $\pi_l(d) = O(l^{-d-1})$ converges to zero quickly.

In contrast, integer-integrated UC models cannot identify such parsimoniously parameterized models, nor models with rich moving average dynamics. To see this,

note that for $\delta = 1$, $d \in \mathbb{N}$, it holds that $a_j = \tilde{a}_j(\varphi)$, $a_j = 0$ for all $j > p$, and $m_j = \tilde{m}_j(\varphi)$, $m_j = 0$ for all $j > q$. Thus, in (3.19) the polynomial $\tilde{a}_+(L, \varphi)$ is of order p, while $\tilde{m}_+(L, \varphi)(1 - L)^d_+$ is of order $q + d$, and thus $b_+(L, \varphi, \theta)$ is of order $\max(p, q + d)$. Consequently, $\gamma_j(\theta) = 0$ for all $j > \max(p, q + d)$, which allows to identify $\max(p, q + d) + 1$ parameters. On the other hand, the structural model has $q+3$ parameters still to be identified, and so identification requires $\max(p, q+d)+1 \geq q + 3$. For $d = 1$, it immediately follows that the correlated UC model is neither identified for $p < 2$, nor for $p < q + 2$.

3.5 Fractional trends and cycles in US GDP

With the fractional UC model at hand, we revisit the puzzling results for the trend-cycle decomposition of US GDP from traditional UC models as summarized in section 3.2. As will become clear, our new model provides additional insights regarding the specification of trend and cycle, and explains the puzzling estimates for the business cycle in the literature:

First, while traditional UC models require to specify the integration order d prior to estimation, d enters the fractional UC model as a parameter to be estimated. Therefore, we provide evidence on the memory of log GDP and draw inference on the appropriate specification of the trend. Specifically, we test the hypotheses that $d_0 = 1$ or $d_0 = 2$. If both hypotheses are rejected, an intermediate solution may better explain the long-run dynamics of US GDP.

Second, besides the stochastic long-run dynamics, we also investigate the specification of the deterministic long-run component μ_t for log GDP. In addition to the traditional linear trend component, we check whether the explanatory power of the model is improved when the constant is placed in the state equation for x_t, yielding a deterministic polynomial trend of order d. Moreover, we check whether allowing for a polynomial trend of order $b \in \mathbb{R}_+$ improves the explanatory power of the fractional UC model for log GDP. This is related to Perron and Wada (2009), who argue that the long-run component of log GDP evolves as a deterministic, nonlinear trend rather than a non-stationary stochastic trend. By letting a deterministic polynomial trend compete against a fractional stochastic trend, we investigate empirically whether the long-run component of log GDP is rather deterministic, or stochastic, or a combination of both.

Third, we shed light on the appropriate specification of the cyclical component. As shown in subsection 3.4.3, the fractional UC model is identified under much weaker conditions than traditional UC models. Therefore, both richer and more

parsimonious parametrizations of c_t can be considered. In particular, we allow for a cyclical ARMA(p,q) polynomial up to orders $p, q \leq 4$, and investigate whether richer cyclical dynamics provide an economically plausible estimate for the cycle. We also examine the fit of the fractional UC model when the standard lag operator is replaced by a fractional lag operator.

Fourth, our results allow to draw inference on the correlation between long- and short-run innovations. While the literature typically finds a correlation coefficient close to -1 once correlated trend and cycle innovations are allowed for, we investigate whether this still holds when the integration order is allowed to deviate from one or two.

In the following, we treat the observable y_t as log seasonally adjusted real US GDP. As in section 3.2, the data for y_t come from the Federal Reserve Bank of St. Louis,[3] are in quarterly frequency, cover the period from 1947Q1 to 2022Q4, and are log-transformed.

3.5.1 Model specification and estimation

We consider several specifications for the different components of the fractional UC model: The deterministic component μ_t in (3.4) either consists of an intercept plus a linear time trend ($b = 1$), as is common in the UC literature for GDP, or it assumes an intercept plus a polynomial trend of order d (i.e. $b = d$), which is equivalent to allowing for a non-zero intercept in the second equation of (3.2) and thus for a drift in x_t. Yet another specification allows for an intercept plus a polynomial trend of order $b \in \mathbb{R}_+$, where b is estimated jointly with the other parameters, thus adding additional flexibility to the deterministic trend component. With respect to the cycle, we parameterize (3.5) either as an ARMA(p, q) process in the standard lag operator L (i.e. $\delta = 1$), where $p = 1, ..., 4$, $q = 0, ..., 4$ orders are considered. As an alternative, we replace L by the fractional lag operator L_d, where the memory parameter takes the integration order of x_t as its value (i.e. $\delta = d$). For the latter specification, we consider lag orders $p = 1, ..., 4$ for the cyclical AR polynomial. Moreover, all specifications include a dummy variable as an additional regressor that takes the value one in the second quarter of 2020, and zero otherwise: Due to the COVID pandemic, US GDP fell by 8.9 percent in the second quarter of 2020, the largest single-quarter contraction in more than 70 years. As a glimpse on figure 3.4 reveals, not controlling for the sudden drop of GDP would result in two large outliers for the prediction errors in quarters two and three of 2020, with the sum

[3]The series can be downloaded here: https://fred.stlouisfed.org/series/GDPC1

of squared prediction errors for these two quarters almost leveling that of all other quarters. In all trend-cycle decompositions, the regression coefficient of the dummy is attributed to the cyclical component.

Parameter estimation of θ_0 via the QML estimator as described in subsection 3.4.3 is carried out as follows: First, for each specification 100 combinations of starting values are drawn from uniform distributions with appropriate support.[4] For each vector of starting values, the quasi-log likelihood is maximized by the BFGS algorithm, and the estimate with the highest likelihood value is selected as the final estimate for θ_0. Note also that by the type II definition of the fractional trend, pre-sample observations are treated as zero, and thus the prediction errors at the beginning of the sample should be treated with caution. The UC literature typically deals with the estimation uncertainty at the beginning of the sample by diffusely initializing the state vector, i.e. setting the initial variance of the state vector to an arbitrarily high value. This reduces the contribution of the first few prediction errors to the objective function to nearly zero, so that they have little effect on the estimates (see e.g. Harvey; 1989, ch. 3.3.4). We take a similar approach by simply excluding the first 40 prediction errors when calculating the quasi-log likelihood. In this way, we eliminate the estimation uncertainty at the beginning of the sample, but at the same time avoid the computationally intensive Kalman recursions for the diffuse initialization of the state covariance.[5]

3.5.2 Estimation results

Tables 3.5 to 3.8 present the estimation results for the different parameterizations of the fractional UC model, along with the model selection criteria and the estimated correlations between η_t and ε_t. Table 3.5 reports the results for fractional UC(d, b, d, ρ) models, where c_t is an autoregressive process in the fractional lag operator L_d, $a(L_d)c_t = \varepsilon_t$. In contrast, tables 3.6 to 3.8 present the estimation results for UC models where the cyclical component is modeled as an ARMA(p, q) process, with $p = 1, ..., 4$, $q = 0, ..., 4$. Specifically, table 3.6 considers the UC(d, 1, 1, ρ) model where the deterministic trend is assumed to be linear, $\mu_t = \mu_0 + \mu_1 t$, while table 3.7 displays the results for the UC(d, d, 1, ρ) model with a polynomial deterministic trend of order d, $\mu_t = \mu_0 + \mu_1 t^d$. Table 3.8 shows the results for the UC(d, b, 1, ρ) model that allows for a polynomial deterministic trend of order b, $\mu_t = \mu_0 + \mu_1 t^b$, where b is estimated jointly with the other parameters via the QML estimator.

[4] In particular, we draw from $d, b \in [1/2; 2]$.

[5] However, the estimation results are robust to diffuse initialization of the state vector instead of dropping the first 40 observations (= observations from 1947Q1 to 1956Q4).

Estimation results for the UC(d, 1, d, ρ) model in table 3.5, which includes a linear trend in μ_t and the fractional lag operator for c_t, indicate an integration order of about 1.30 for log US real GDP. Once the linear deterministic trend is replaced by the polynomial $\mu_t = \mu_0 + \mu_1 t^d$, the estimate for d_0 becomes less than one whenever $p \geq 2$. As a glimpse on figure 3.3 shows, log GDP evolves slightly concave in the long run, which is well approximated by $\mu_1 t^d$ for $d < 1$. Consequently, the QML estimator produces a smaller estimate for d_0 to account for the concave evolution of log GDP via the deterministic component. Note, however, that the likelihood of the UC(d, 1, d, ρ) model is always greater than the likelihood of the UC(d, d, d, ρ) model, suggesting that a linear deterministic trend is more appropriate for log US GDP. Further evidence comes from the UC(d, b, d, ρ) model in table 3.5: Once b is allowed to deviate from d, the deterministic component becomes concave, while the stochastic trend is estimated to be integrated of order greater than one, supporting the results of the UC(d, 1, d, ρ) model. All three models find a small variance-ratio of long- and short-run innovations, which is in contrast to the results from integer-integrated UC models, see section 3.2 and table 3.3. Furthermore, within the set of models with fractional lag operator L_d, the information criteria favor the UC(d, 1, d, ρ) specification. Obviously, placing the constant into the second equation of (3.2) downward-biases the estimate of the memory of log GDP, thus reducing the likelihood. Allowing for $d \neq b \neq 1$ again yields an estimated integration order greater than one, along with an estimate for b_0 slightly below one, but the additional parameter increases the information criteria. Moreover, table 3.5 provides evidence against the narrative of Perron and Wada (2009) that the long-run dynamics of GDP are purely deterministic and driven by a nonlinear trend: If there were no stochastic long-run dynamics in log US GDP, the UC(d, b, d, ρ) estimate for the variance of the long-run innovations should either be indistinguishable from zero, or the estimate for d_0 should go to its lower bound of 1/2, such that the trend component captures at least the very persistent cyclical dynamics. However, the estimates for the UC(d, b, d, ρ) model indicate an integration order greater than one, together with a non-zero variance of the long-run innovations.

Turning to the estimates for UC models with the standard lag operator L for c_t in tables 3.6 to 3.8, it is striking that for all three specifications of the deterministic component, the BIC always favors the most parsimonious model, i.e. where c_t is an ARMA$(1, 0)$. As discussed at the end of subsection 3.4.3, an advantage of the fractional UC model is that it remains identified for all $d \neq 1$ even when c_t is an ARMA$(1, 0)$, whereas the $I(1)$ correlated UC model requires at least two autoregres-

sive lags to identify all parameters. On the other hand, results from the BIC show that additional AR and MA lags do not improve the fit on a large scale. Thus, while in principle the fractional UC model allows for rich parameterizations of the cycle, such specifications are not supported by the data. Iwata and Li (2015) argue that even an AR(2) specification for the cyclical component is likely to be overparameterized for $I(1)$ UC models of US GDP, and this seems to hold for fractional UC models as well. Another notable result from tables 3.6 to 3.8 is that all three models yield an estimate of d_0 that is less than unity for almost all parameterizations of the cycle, along with comparatively high estimates of $\sigma^2_{\eta,0}$. Consequently, they produce a rather volatile estimate for the trend component. Of all the specifications, the BIC favors the UC(d, d, 1, ρ) model with a single autoregressive lag for the cycle. Since the more general UC(d, b, 1, ρ) model yields estimates for b_0 that are close to those for d_0, the additional parameter does not improve the fit of the model by much, which explains the higher BIC. However, all models with ARMA cycles are clearly outperformed in terms of the BIC by the UC(d, 1, d, ρ) model with fractional lag operator L_d .

Regardless of the specification of the deterministic component μ_t or the use of the fractional lag operator for the cycle, all estimates in tables 3.5 to 3.8 converge to the corner solution where $\hat{\rho} = \widehat{\text{Corr}}(\eta_t, \varepsilon_t) = -1$. Therefore, the Hessian matrix of the QML estimator is nearly singular, making the estimated standard errors in tables 3.5 to 3.8 unreliable. While an estimate $\hat{\rho}$ close to -1 is frequently obtained in the empirical literature once correlation between long- and short-run innovations is accounted for (see e.g. Morley et al.; 2003; Iwata and Li; 2015), the result ultimately implies that long- and short-run shocks cannot be identified separately for log US real GDP, even when rich cyclical ARMA dynamics are considered. This raises the question whether UC models in general are able to distinguish between long- and short-run innovations, or whether the estimated correlation coefficient is an artifact generated by the model and estimation procedure, rather than the data. We address this question in appendix 3.A.1, where we provide evidence that the estimated correlation coefficient of (almost) -1 is a feature of the data, not of the model or the estimation procedure.

Among all fractional UC models, the UC(d, 1, d, ρ) model with a single autoregressive lag for c_t in the fractional lag operator (i.e. $p = 1$) minimizes the Bayesian information criterion. Therefore, it is treated as the preferred fractional UC specification in the following. The stochastic long-run component of GDP is estimated to be integrated of order around 1.30, which is quite far from the integer-integrated

specifications as considered in section 3.2. However, testing the hypotheses that $d_0 = 1$ and $d_0 = 2$ using the t-test is problematic, because the standard errors reported in table 3.5 are likely to suffer from the corner solution of the QML estimator. Fortunately, the likelihood ratio test provides a solution.

The right columns of table 3.3 show the estimation results when the integration order is fixed to one or two. They were estimated analogously to the fractional UC models by excluding the first 40 prediction errors from the quasi-likelihood and by including a dummy for the second quarter of 2020 to account for the outlier generated by the COVID pandemic. Both the correlated $I(1)$ and the correlated $I(2)$ UC models exhibit a log likelihood of about -316, while the log likelihood of the UC(d, 1, d, ρ) model with two autoregressive lags (as for the models under the null hypothesis) is about -301. The latter includes only the single additional parameter d, which is set to either one or two in the integer-integrated models. The test statistic of the likelihood ratio test is about 30 for both models. Consequently, the test rejects both the hypotheses that $d_0 = 1$ and $d_0 = 2$ at any conventional level of significance.

3.5.3 Trend-cycle decomposition

The estimated integration order $\hat{d} = 1.30$ implies that a long-run shock on GDP growth (i.e. the first difference of log GDP) not only has a contemporaneous effect, as imposed in the $I(1)$ model, but evolves as a mean-reverting fractionally integrated process of order around 0.3. A long-run shock then retains 30% of its initial impact after one quarter, 20% after two quarters, and 12% after one year. It converges to zero at a hyperbolic rate, leaving 5% of its initial impact after four years and 3% after ten years. A possible economic interpretation is that long-run shocks, such as technological innovations, are not adapted by the whole economy at a fixed point in time, but rather successively.

Turning to the cycle, we find that a parsimonious parameterization of c_t is suggested by the BIC, which is in line with the findings of Morley et al. (2003). Additional lags do not significantly improve the overall fit, as their coefficients are small and insignificant at the 5% level (to see this, compare the likelihood ratios of table 3.5). As argued by Iwata and Li (2015), a small p in the data-generating mechanism of log GDP complicates the separation of long- and short-run innovations, which may also explain the estimated correlation coefficient. From the results in table 3.5,

it follows for c_t that

$$\hat{c}_t = 0.91 L_{1.30} \hat{c}_t = 1.18\hat{c}_{t-1} - 0.18\hat{c}_{t-2} - 0.04\hat{c}_{t-3} - 0.02\hat{c}_{t-4} - ...,$$

where the sign-change after the first coefficient illustrates that the fractional lag operator is able to generate oscillatory behavior, for which standard AR models require at least two lags. The parameter estimates for the cycle indicate strong persistence, which is intuitive for the business cycle. However, the roots of the autoregressive polynomial are all outside the unit circle, with the smallest root being 1.03. Moreover, the estimated variance-ratio between long- and short-run innovations is $\hat{\sigma}_\eta^2 / \hat{\sigma}_\varepsilon^2 = 0.11$, indicating that much of the variation in log GDP is due to the cycle, while the stochastic trend is comparatively smooth. This is in contrast to the results for the correlated $I(1)$ UC model, see section 3.2.

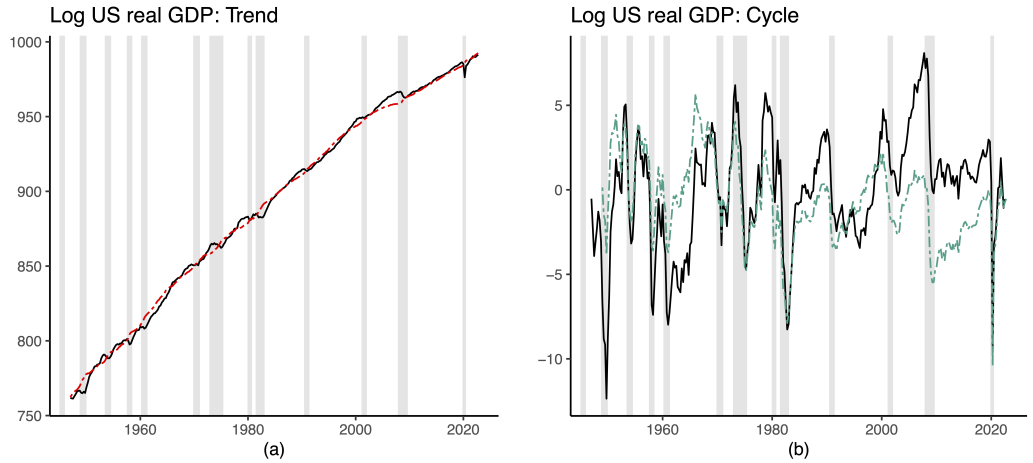

Figure 3.3: Trend-cycle decomposition for log US real GDP from the UC(d, 1, d, ρ) model with $p = 1$ lags for c_t. The left plot (a) sketches the smoothed trend component $\tau_t = \mu_t + x_t$ from the fractional UC model in red, dashed, together with log US real GDP in black, solid. The right plot (b) shows the smoothed cyclical component c_t (black, solid) from the same fractional UC model, where c_t is an autoregressive process in the fractional lag operator L_d, together with the output gap estimate from the US Congressional Budget Office (blue, dashed). Shaded areas correspond to NBER recession periods.

Figure 3.3 plots the trend-cycle decomposition for the fractional UC(d, 1, d, ρ) model, where the estimates for x_t, c_t are obtained from the single-step Kalman smoother as discussed in subsection 3.4.2. Due to the small $\hat{\sigma}_\eta^2$, the decomposition yields a smooth trend with little short-run variation. Furthermore, the Kalman smoother attributes much of the variation in log GDP to the cyclical component due to the relatively high $\hat{\sigma}_\varepsilon^2$. The smoothed cyclical component exhibits a persistent be-

havior and has the shape of an asymmetric sine curve. It rises gradually in periods of economic recovery and upswing, and falls sharply during recession periods. Similar estimates are obtained from the nonlinear regime-switching UC-FP-UR model of Morley and Piger (2012), which promotes the generality of the parsimonious fractional UC model. Moreover, the estimate for c_t shows similar pro-cyclical dynamics as the US Congressional Budget Office (CBO)[6] estimate for the output gap that is sketched in blue. This is striking, because the CBO estimate is based on macroeconomic theory: It models potential GDP by a Cobb-Douglas production function with labor, capital, and total factor productivity as inputs, and ties changes in output to changes in unemployment using Okun's law. In contrast, our results are fully data-driven and thus complement the CBO's theory-based results. While our estimate of the cycle coincides with the output gap estimate of the CBO in terms of the general patterns as well as the key turning points, there are some interesting differences: The fractional UC model reveals a persistent overheating of the US economy in the run-up to the Great Recession, which is a feature not detected by the CBO estimate. However, our finding aligns well to the output gap measure of Barigozzi and Luciani (2021), which is based on a dynamic factor model with macroeconomic indicators as inputs, as well as to Borio et al. (2017), who argue that credit growth was a key driver of the overheating of the US economy in the run-up to the Great Recession. Another striking result concerns the COVID-19 recession, where neither our estimate nor the CBO's output gap shows any signs of overheating.

From figure 3.3 it becomes clear that the fractional UC model solves the problem of obtaining implausible cycle estimates in the integer-integrated UC literature. The solution to the UC puzzle is that, given that log GDP is integrated of order around 1.3, forcing the long-run component to be $I(1)$ upward-biases the estimate $\hat{\sigma}_\eta^2$, as the additional memory that is not captured by the $I(1)$ specification goes into the estimates for the long-run innovations η_t. To adjust for the erratic behavior of the long-run innovations, the estimate for the cycle becomes noisy. In contrast, the fractional UC model fully captures the memory of log US GDP, which allows $\hat{\sigma}_\eta^2$ to be small, and thus yields a smooth trend estimate along with a persistent cyclical component that hits all NBER recession periods. These results are consistent with the work of Kamber et al. (2018), who obtain plausible cycle estimates when restricting the variance-ratio of long- and short-run innovations to be small.

[6]The series can be downloaded here: https://fred.stlouisfed.org/graph/?g=f1cZ

3.5.4 Model diagnostics

A few diagnostic checks for the UC(d, 1, d, ρ) model are in order: Figure 3.4 plots the estimated autocorrelation function of the prediction error of the Kalman filter $v_t(\hat{\theta})$ as given in (3.17), the smoothed periodogram for the smoothed long-run innovations η_t, the prediction error $v_t(\hat{\theta})$ itself, and the prediction error for the UC(d, 1, d, ρ) model without a dummy variable for the second quarter of 2020. The estimated autocorrelation function shows that there is little autocorrelation left in the prediction errors, with only slightly significant autocorrelation at lag five. As the cyclical dynamics of log GDP are well captured by the parsimonious UC(d, 1, d, ρ) model with only a single lag for the cyclical component, little to no additional benefit can be expected from models with richer parameterizations for the cyclical component. This is consistent with the estimation results in tables 3.5 to 3.8, where richer models yield only small improvements in terms of the likelihood. Moreover, the estimated (almost) perfect correlation between long- and short-run innovations can be expected to persist even in richer models.

For the smoothed long-run innovations η_t, the periodogram in figure 3.4 reveals a spike at frequency 0.1, which refers to a period of ten quarters and is due to the slightly significant autocorrelation at lag 5. However, neither a peak, nor a zero at the origin of the periodogram can be spotted. Thus, contrary to the $I(1)$ UC models in figure 3.2, allowing for a fractionally integrated trend removes the long memory from the smoothed long-run innovations, and at the same time ensures them not to exhibit intermediate memory as for $I(2)$ UC models.

For the prediction errors of the UC(d, 1, d, ρ) model with a dummy for the first quarter of the COVID pandemic (i.e. 2020Q2), the bottom-left panel of figure 3.4 shows no large outliers. However, starting in the mid 1980s, the prediction errors appear to have a lower variance compared to the first half of the sample. For UC models with an $I(1)$ specification for the trend, a generalization that accounts for the structural break in the mid 1980s is proposed by Weber (2011). It would be interesting to see whether controlling for a structural break in the fractional UC model can reveal additional details about the parameters, in particular the correlation of the innovations, as well as about the dynamics of the trend and the cycle.

To illustrate the importance of the COVID-19 dummy, the bottom-right panel shows the prediction error of the UC(d, 1, d, ρ) model without the dummy in the second quarter of 2020. When the dummy is omitted from the model, the prediction errors show a strong negative outlier in 2020Q2, followed by a spike in the next quarter. Since the objective function of the QML estimator minimizes the sum

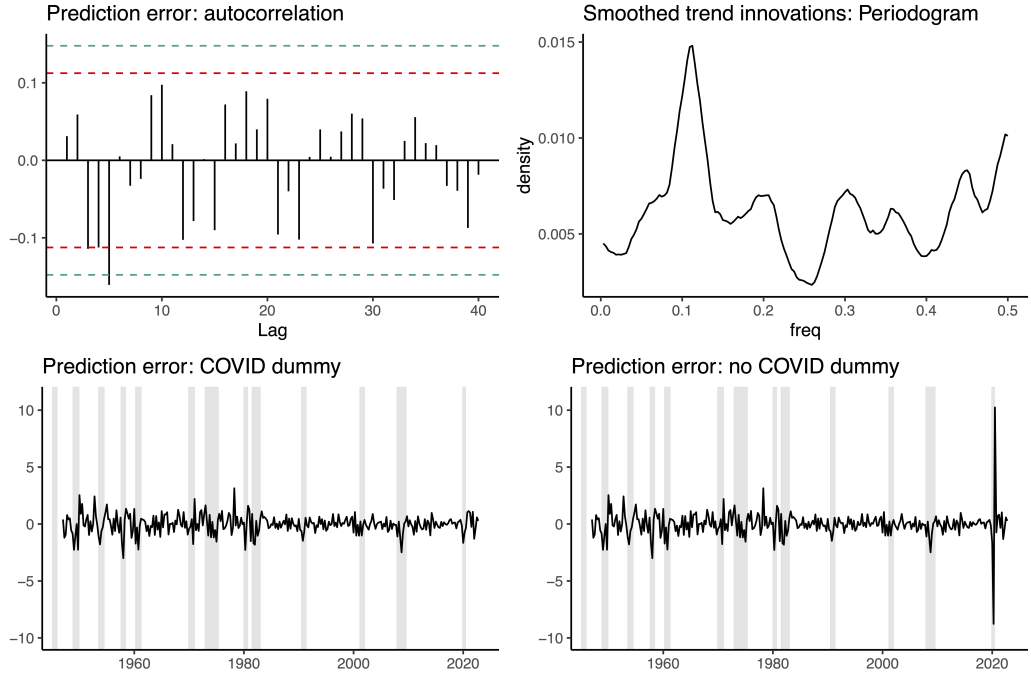

Figure 3.4: Diagnostic checks: For the UC(d, 1, d, ρ) model with $p = 1$ autoregressive lags, the upper-left panel displays the estimated autocorrelation function of the prediction errors, together with 5% (red) and 1% (blue) confidence bands, the upper-right panel shows the smoothed periodogram for the smoothed long-run innovations η_t, the bottom-left panel plots the prediction errors, while the bottom-right panel shows the respective prediction errors of the UC(d, 1, d, ρ) without a dummy in the second quarter of 2020. Shaded areas highlight NBER recession periods.

of squared prediction errors, omitting the dummy would assign a strong weight to the two respective quarters, which is likely to deteriorate the estimation of the parameters and the latent components. Therefore, any UC model that includes the period of the COVID pandemic needs to somehow adjust for the strong outliers, and including a single dummy in 2020Q2 obviously does a good job.

Finally, we take a closer look at the corner solution of the QML estimator: If long- and short-run innovations are perfectly correlated, we can write $\varepsilon_t = \beta\eta_t$, where $\beta = \sigma_{\eta\varepsilon}/\sigma_\eta^2$ is the regression coefficient from regressing ε_t on η_t. The UC(d, 1, d, $\rho = -1$) model can then be written as

$$a(L_d)\Delta_+^d(y_t - \mu_t) = a(L_d)\eta_t + (1 - L_d)\beta\eta_t = c(L_d)v_t, \tag{3.21}$$

where $v_t = (1+\beta)\eta_t$ with $\text{Var}(v_t) = (1+\beta)^2\sigma_\eta^2$, and $c(L_d) = \sum_{j=0}^p c_j L_d^j$, with $c_0 = 1$, $c_1 = -(a_1 + \beta)/(1 + \beta)$, and $c_j = -a_j/(1 + \beta)$ for all $2 \leq j \leq p$. (3.21) is a single-source-of-error model that admits an ARFIMA representation in the fractional lag

operator L_d. For the parameter estimates of the chosen UC(d, 1, d, ρ) model with $p = 1$, the corresponding coefficients in (3.21) can be calculated from the results in table 3.5. These are $\hat{\beta} = -2.99$, $\hat{c}_1 = -1.04$, and $\widehat{\text{Var}}(v_t) = 0.57$. However, (3.21) can also be estimated directly via the QML estimator by solving (3.21) for v_t and optimizing over $\theta_{SSE} = (d, a_1, \sigma_\eta^2, \beta)'$, which yields $\hat{\theta}_{SSE} = (1.32, 0.90, 0.16, -3.00)'$. The estimates are almost identical to the results from the fractional UC model. In addition, the log likelihood is -303.69 and thus is very close to the likelihood of the fractional UC model.

3.6 Conclusion

In this paper, we revisited the puzzling estimates for the business cycle generated by traditional, integer-integrated unobserved components models. Our hypothesis was that the long-run dynamics of log GDP are captured neither by UC models with an $I(1)$ trend component, nor by those with an $I(2)$ component. Instead, the periodograms of the smoothed long-run innovations for integer-integrated UC models suggested an integration order somewhere between the $I(1)$ and the $I(2)$ specifications.

To test whether an intermediate solution for the memory of the trend solves the unobserved components puzzle, we revisited the puzzling results using the fractional UC model that was derived in chapter 2 of this thesis, which models the trend as a fractionally integrated process of order d, nesting the traditional, integer-integrated specifications. Since d is estimated jointly with the other model parameters, the fractional UC model provides a data-driven solution to the specification of the long-run dynamics in UC models.

Our estimation results indicate an integration order of about 1.3, implying that the trend of log GDP is more persistent than assumed by $I(1)$ UC models, but less persistent than assumed by $I(2)$ UC models. The likelihood ratio test rejects both, an $I(1)$ trend and an $I(2)$ trend in log GDP, indicating that integer-integrated UC models are misspecified. In contrast to integer-integrated UC models, the trend-cycle decomposition from the fractional UC model hits all NBER recession periods, identifies the same turning points as the output gap reported by the US Congressional Budget Office, and reveals some additional details, e.g. an overheating economy in the run-up to the Great Recession.

While the estimates for trend and cycle are very different from the traditional literature, our estimates for the correlation between long- and short-run innovations are well in line with the literature: For almost all parameterizations considered, we

estimate an (almost) perfect correlation between long- and short-run innovations. Thus, for GDP, long- and short-run innovations cannot be structurally identified by the fractional UC model.

The model offers a variety of opportunities for future research. First, to separately identify long- and short-run innovations, a multivariate generalization of the fractional UC model could be considered. This would allow to model GDP jointly with other economic variables that may exhibit a more pronounced cyclical behavior. Second, fractional trends of different persistence could be incorporated, which would allow to decompose time series into components of different memory.

For applied researchers, the model provides a flexible, data-driven method for treating permanent and transitory components in macroeconomic and financial applications. It provides a solution to many model specification issues that have caused uncertainty and debate about realistic trend-cycle decompositions and the estimation of recessions.

3.A Appendix

3.A.1 The corner solution of the QML estimator

In this appendix, we investigate the corner solution of the QML estimator as found in section 3.5, i.e. the estimated correlation of -1 between long- and short-run innovations. We want to ensure that the estimated (almost) perfect correlation is a feature of the data on log US real GDP, and not a result of weak identification or an artifact generated by the fractional UC model. First, to rule out weak identification, we analyze the log likelihood graphically, and check whether it is flat around the optimum. If so, then the QML estimator is likely to suffer from weak identification, and numerical optimization routines may converge to the boundary of the parameter space where they are terminated. Second, we screen out corner solutions by considering a constrained optimization problem that penalizes strong correlation between long- and short-run innovations. If the constrained optimization yields a similar value for the log likelihood, this would also indicate weak identification. Conversely, a much smaller likelihood would indicate that the (almost) perfect correlation is a feature of the data. Third, to see whether fractional UC models can in principle identify a correlation coefficient other than ± 1, we perform a small Monte Carlo study, where all parameters are equal to the estimates for the preferred UC(d, 1, d, ρ) model in section 3.5 with a single lag for the cyclical component, except the correlation coefficient that is set to -0.8 to mimic strong, but not perfect, correlation. We then examine whether the QML estimator is able to find the true correlation coefficient, or whether it also converges to the corner solution.

Figure 3.5 plots the negative log likelihood of the UC(d, 1, d, ρ) model with a single autoregressive lag for the cyclical component around its optimum in table 3.5: Each plot shows the negative log likelihood for a two-dimensional grid of two parameter combinations in θ, with all other parameters held fixed at their estimated values. Thus, each plot shows how the likelihood changes if we vary over two parameters in θ, holding all other parameters fixed. As can be seen, the likelihood is hump-shaped for the combinations (d, a_1), (d, σ_η^2), $(d, \sigma_\varepsilon^2)$, (a_1, σ_η^2), and $(a_1, \sigma_\varepsilon^2)$, indicating that these parameters are well identified. As to be expected, the plot for $(\sigma_\eta^2, \sigma_\varepsilon^2)$ shows a ridge under perfect correlation. Most interestingly, the likelihood is steep in the direction of ρ around -1: For $\rho > -.995$, the negative log likelihood is greater than 315, while at the optimum it is about 301. Thus, the objective function is anything but flat in the direction $\rho \to -1$, which rules out that the estimated (almost) perfect correlation is an artifact generated by the numerical optimization procedure.

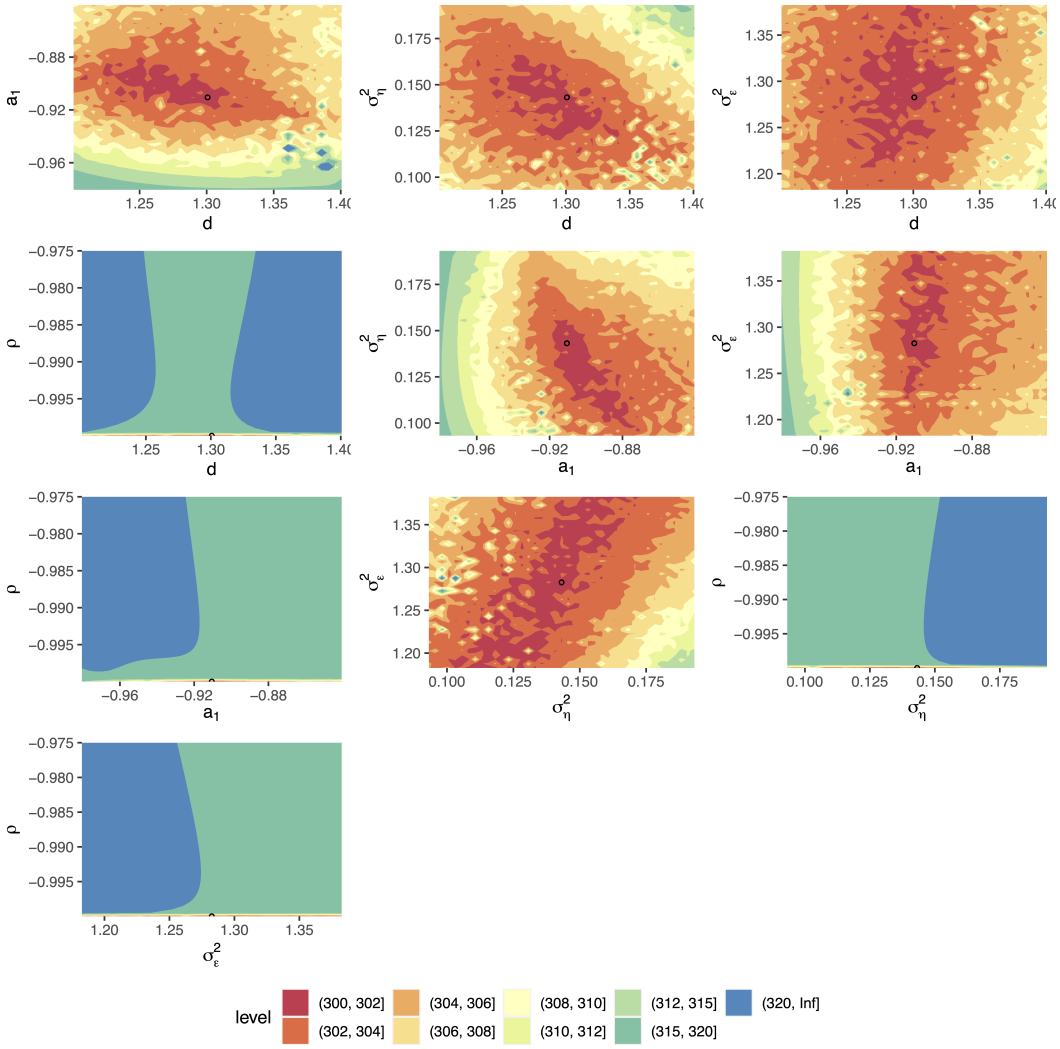

Figure 3.5: Contour plot of the negative log likelihood for the UC(d, 1, d, ρ) model with a single autoregressive lag for the cyclical component. The black dot visualizes the parameter estimate of the QML estimator as also reported in table 3.5. All parameters except those on the (x, y)-axis are held fixed at their point estimates as reported in table 3.5.

Next, we reconsider the optimization problem of section 3.5, but constrain the parameter space: We penalize corner solutions for the correlation coefficient by adding a penalty term to the log likelihood once $|\rho| > 0.99$ that is

$$Penalty(\theta) = \begin{cases} \log \frac{1-|\rho|}{0.01} & \text{if } |\rho| > 0.99, \\ 0 & \text{else.} \end{cases}$$

By design, the penalty term is zero for $|\rho| = 0.99$, and goes to $-\infty$ for $|\rho| \to 1$. Thus,the region of the parameter space for Θ is reduced by excluding (almost) perfect

correlation of long- and short-run innovations.

Estimation results for the constrained optimization are contained in tables 3.9 to 3.12 for the same parameterizations of fractional UC models as for the unconstrained optimization in section 3.5. As can be seen from the four tables, numerical optimization of the constrained likelihood now yields an estimated correlation of -0.99 for almost all parameterizations. Thus, constraining the optimization problem does not yield an optimum where the correlation is significantly less pronounced. For the preferred UC(d, 1, d, ρ) model with a single autoregressive lag for the cycle, the parameter estimates are similar to the unconstrained results, however the log likelihood is much smaller. This is well illustrated by the plots in figure 3.5 with ρ on the ordinate, where the new optimum of the constrained optimization problem now falls into the green region. In addition, the constrained results often show a comparatively high estimate for $\sigma^2_{\eta,0}$, together with a variance-ratio of trend and cycle innovations that is estimated to be greater than one. Consequently, a lot of variation is attributed to the trend, making it erratic, while little variation is left to be captured by the cycle. Thus, constraining the optimization problem to exclude (almost) perfectly correlated innovations not only reduces the likelihood, but also yields economically implausible estimates for trend and cycle.

Finally, we check whether the estimated (almost) perfect correlation is an artifact generated by the QML estimator for fractional UC models. For this purpose, we conduct a Monte Carlo study, where we simulate y_t by drawing 1000 replications from the distribution that is generated under $\theta_0 = (1.30, 0.14, -0.34, 1.28, 0.91)'$ for the selected UC(d, 1, d, ρ) model. θ_0 is equal to the QML estimates for the UC(d, 1, d, ρ) model with a single autoregressive lag for the cyclical component as reported in table 3.5, with the only exception that the covariance parameter is set such that $\rho_0 = -0.80$. Thus, we study a data-generating mechanism in which the innovations are highly, but not perfectly, correlated. We then examine whether the QML estimator is able to reliably estimate the covariance parameter, or whether it again converges to the corner solution where the correlation is perfectly negative. The latter would imply that the corner solution for US GDP is not a property of the data, but rather an artifact that is generated by the model and the estimation procedure.

Table 3.2 illustrates the results from the Monte Carlo study. The correlation coefficient itself is not estimated directly (since the optimization is conducted over $\sigma_{\eta\varepsilon}$), however it can be computed for each estimate. As table 3.2 shows, the mean estimate for ρ_0 is slightly upward-biased. Since ρ_0 is close to the lower bound of the

	θ_0	mean	$q_{.1}$	$q_{.2}$	$q_{.25}$	$q_{.5}$	$q_{.75}$	$q_{.8}$	$q_{.9}$	rMSE	bias	median bias
d	1.30	1.26	1.15	1.20	1.21	1.26	1.32	1.33	1.37	0.11	-0.04	-0.04
σ_η^2	0.14	0.20	0.01	0.02	0.03	0.08	0.14	0.16	0.25	0.76	0.06	-0.06
$\sigma_{\eta\varepsilon}$	-0.34	-0.28	-0.45	-0.33	-0.31	-0.17	-0.06	-0.04	0.03	0.77	0.06	0.17
σ_ε^2	1.28	0.87	0.45	0.55	0.59	0.74	0.95	1.06	1.20	0.87	-0.41	-0.54
a_1	0.91	0.88	0.77	0.82	0.84	0.90	0.94	0.95	0.98	0.10	-0.03	-0.01
ρ	-0.80	-0.59	-1.00	-0.94	-0.91	-0.80	-0.51	-0.39	0.29	0.56	0.21	-0.00

Table 3.2: Simulation results for the UC(d, 1, d, ρ) model. The table shows the mean estimate (mean), the respective quantiles ($q_.$), the root mean squared error (rMSE), the mean bias (bias) and the median bias (bias) for the QML estimator of θ_0 in a Monte Carlo study with 1000 replications.

support for ρ, this is not surprising. As can be seen from the quantiles reported in table 3.2, at least 10% of all estimates are at the boundary of the parameter space, where $\hat{\rho} = -1$ is estimated. However, the 20%-quantile is already away from the corner solution. The median estimate equals ρ_0 itself, so that the QML estimator appears to be median-unbiased, which is reassuring. In sum, the simulation shows that the QML estimator can identify correlations that differ from the corner solution in section 3.5, and thus the (almost) perfect correlation for log US real GDP appears to be generated by the data, and not by the estimation procedure or the fractional UC model.

3.A.2 Additional figures and tables

	(a)				(b)			
	$I(1)$	$I(1)$-corr	$I(2)$	$I(2)$-cor	$I(1)$	$I(1)$-corr	$I(2)$	$I(2)$-cor
σ_η^2	1.06	1.73	0.00	0.00	0.33	1.60	0.00	0.00
	(0.15)	(0.42)	(0.00)	(0.00)	(0.10)	(0.31)	(0.00)	(0.00)
$\sigma_{\eta\varepsilon}$		-0.53		0.03		-0.89		-0.02
		(0.32)		(0.03)		(0.27)		(0.00)
σ_ε^2	0.14	0.32	1.23	1.23	0.25	0.67	0.63	0.65
	(0.12)	(0.27)	(0.10)	(0.10)	(0.12)	(0.22)	(0.04)	(0.05)
b_1	1.51	1.15	1.07	1.09	1.61	0.67	1.29	1.30
	(0.16)	(0.29)	(0.06)	(0.06)	(0.12)	(0.07)	(0.05)	(0.06)
b_2	-0.64	-0.65	-0.12	-0.12	-0.61	-0.23	-0.32	-0.30
	(0.17)	(0.20)	(0.06)	(0.06)	(0.12)	(0.02)	(0.05)	(0.05)
$-\log L(\theta)$	469.60	467.56	468.98	468.58	317.88	316.81	319.13	316.23
ρ		-0.71		1.00		-0.86		-1.00
AIC	947.19	945.11	945.96	947.16	643.77	643.62	646.25	642.46
BIC	962.06	963.70	960.83	965.75	658.63	662.20	661.12	661.05

Table 3.3: Parameter estimates for the integer-integrated UC models. Standard errors are denoted in parentheses and are obtained from the numerical Hessian matrix. Columns (a) contain the quasi-maximum likelihood estimates for the diffusely initialized state vector, while columns (b) omit the first 40 prediction errors from the likelihood, avoid the diffuse initialization of the objective function, and allow for a dummy in the second quarter of 2020.

	$I(1)$	$I(1)$ corr.	$I(2)$	$I(2)$ corr.	$I(1)$ BN filter	$I(2)$ HP filter
$\hat{d}^{EW}_{.5}$	0.53	0.09	-1.00	-1.00	-0.10	0.29
$\hat{d}^{GPH}_{.5}$	0.70	0.17	-0.58	-0.58	0.34	-0.03
$\hat{d}^{EW}_{.55}$	0.56	0.09	-0.76	-0.77	0.48	0.97
$\hat{d}^{GPH}_{.55}$	0.65	0.13	-0.63	-0.63	0.41	0.43
$\hat{d}^{EW}_{.6}$	0.44	0.17	-0.65	-0.66	0.60	1.55
$\hat{d}^{GPH}_{.6}$	0.53	0.19	-0.60	-0.61	0.49	0.85
$\hat{d}^{EW}_{.65}$	0.30	0.35	-0.52	-0.54	0.66	1.98
$\hat{d}^{GPH}_{.65}$	0.36	0.43	-0.47	-0.48	0.54	1.29
$\hat{d}^{EW}_{.7}$	0.22	0.55	-0.31	-0.33	0.65	2.00
$\hat{d}^{GPH}_{.7}$	0.22	0.53	-0.32	-0.34	0.56	1.61
$\hat{d}^{EW}_{.75}$	0.14	0.32	-0.05	-0.08	0.61	2.00
$\hat{d}^{GPH}_{.75}$	0.10	0.37	-0.18	-0.20	0.53	1.89

Table 3.4: Estimates for the memory parameter of the smoothed long-run innovations for the UC-I(1)-, the UC-I(2)-, the UC-I(1)-corr, and the UC-I(2)-corr model, as well as for the BN- and HP-filter. Memory parameter estimates stem from the exact local Whittle estimator of Shimotsu and Phillips (2005), denoted as $\hat{d}^{EW}_{.}$, and the estimator of Geweke and Porter-Hudak (1983), denoted as $\hat{d}^{GPH}_{.}$, and the subscript indexes the bandwidth $\alpha \in \{.5, ..., .75\}$.

	UC(d, 1, d, ρ)				UC(d, d, d, ρ)				UC(d, b, d, ρ)			
p	1	2	3	4	1	2	3	4	1	2	3	4
d	1.30	1.30	1.30	1.27	1.22	0.90	0.92	0.88	1.31	1.30	1.04	1.04
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
b									0.93	0.90	0.92	0.92
									(0.00)	(0.00)	(0.00)	(0.04)
σ_η^2	0.14	0.14	0.14	0.14	1.26	3.43	2.32	5.11	0.09	0.08	0.86	0.86
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.07)	(0.00)	(0.00)	(0.00)	(0.00)
$\sigma_{\eta\varepsilon}$	-0.43	-0.43	-0.43	-0.43	-2.17	-3.29	-3.49	-6.79	-0.31	-0.30	-1.55	-1.55
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.09)	(0.00)	(0.00)	(0.00)	(0.01)
σ_ε^2	1.28	1.28	1.28	1.28	3.73	3.16	5.24	9.01	1.11	1.11	2.79	2.79
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.04)	(0.12)	(0.00)	(0.00)	(0.00)	(0.01)
a_1	0.91	0.91	0.91	0.91	0.96	1.46	1.07	1.07	0.91	0.90	1.05	1.05
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
a_2		0.00	0.00	-0.01		-0.50	-0.06	-0.05		0.00	-0.05	-0.05
		(0.00)	(0.00)	(0.00)		(0.04)	(0.01)	(0.01)		(0.00)	(0.00)	(0.01)
a_3			0.00	0.00			-0.08	-0.07			-0.07	-0.07
			(0.00)	(0.00)			(0.01)	(0.01)			(0.00)	(0.01)
a_4				-0.01				-0.02				0.00
				(0.00)				(0.02)				(0.01)
$-\log L(\theta)$	301.01	300.98	301.66	300.27	308.72	306.51	301.40	301.03	304.26	303.99	301.39	301.24
ρ	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
μ_0	760.51	760.46	760.52	760.34	765.15	760.42	760.64	759.97	760.92	760.62	760.75	760.73
μ_1	0.89	0.89	0.88	0.89	0.27	1.23	1.07	1.26	1.05	1.16	1.07	1.07
μ_2	-8.02	-8.03	-8.04	-8.03	-8.07	-7.99	-8.05	-8.05	-8.04	-8.04	-8.04	-8.04
AIC	612.02	613.96	617.33	616.55	627.45	625.03	616.80	618.06	620.51	621.98	618.78	620.48
BIC	629.90	635.41	642.36	645.16	645.33	646.48	641.83	646.67	641.97	647.01	647.38	652.67

Table 3.5: Estimation results for the UC(d, b, d, ρ) model. The table shows the quasi-maximum likelihood estimates for the fractional UC model with a polynomial deterministic trend of order b, $\mu_t = \mu_0 + \mu_1 t^b$, where b is either set to unity (left columns), or equals d (center columns), or is estimated jointly with the other parameters via QML (right columns). Further components are an AR(p) cyclical component in the fractional lag operator L_d (i.e. , $a(L_d)c_t = \varepsilon_t$, and thus $\delta = d$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. μ_2 denotes the regression coefficient for the dummy in 2020Q2 as described at the beginning of section 3.5. Standard errors are denoted in parentheses and are obtained from the inverted numerical Hessian matrix.

$(p;q)$	(1;0)	(2;0)	(3;0)	(4;0)	(1;1)	(2;1)	(3;1)	(4;1)	(1;2)	(2;2)	(3;2)	(4;2)	(1;3)	(2;3)	(3;3)	(4;3)	(1;4)	(2;4)	(3;4)	(4;4)
d	0.86	0.77	0.66	0.67	0.72	0.72	0.66	0.67	0.72	0.72	0.96	1.03	0.73	0.94	0.79	0.98	0.73	0.96	1.26	1.04
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.02)	(0.07)	(0.01)	(0.00)	(0.00)	(0.00)
σ_η^2	4.49	11.01	18.43	18.50	16.27	16.26	18.44	18.50	16.26	16.26	1.83	1.90	16.19	2.08	9.30	1.90	16.24	2.57	0.25	1.32
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.48)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.16)	(0.20)	(0.21)	(0.00)	(0.00)	(0.00)
$\sigma_{\eta\varepsilon}$	-6.08	-13.50	-21.67	-21.75	-19.46	-19.46	-21.68	-21.75	-19.45	-19.46	-2.87	-2.95	-19.36	-3.18	-11.59	-2.94	-19.42	-3.79	-0.62	-2.20
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.59)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.19)	(0.29)	(0.24)	(0.00)	(0.00)	(0.00)
σ_ε^2	8.23	16.56	25.48	25.57	23.27	23.27	25.49	25.57	23.27	23.27	4.48	4.57	23.17	4.87	14.44	4.57	23.22	5.58	1.57	3.66
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.71)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)	(0.22)	(0.43)	(0.27)	(0.00)	(0.00)	(0.00)
a_1	0.91	0.84	0.73	0.73	0.89	0.89	0.73	0.73	0.89	0.89	0.57	0.85	0.90	1.30	0.67	1.51	0.90	0.39	-0.46	1.74
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.03)	(0.08)	(0.01)	(0.00)	(0.00)	(0.00)
a_2		0.07	0.10	0.10		0.00	0.10	0.10		0.00	0.86	-0.74		-0.36	-0.67	-1.54		0.48	0.44	-1.81
		(0.00)	(0.00)	(0.00)		(0.00)	(0.01)	(0.00)		(0.00)	(0.00)	(0.03)		(0.00)	(0.04)	(0.06)		(0.00)	(0.00)	(0.00)
a_3			0.02	0.02			0.02	0.02			-0.51	0.91			0.81	1.17			0.82	1.42
			(0.00)	(0.00)			(0.01)	(0.00)			(0.00)	(0.02)			(0.03)	(0.08)			(0.00)	(0.00)
a_4				0.01				0.01				-0.12				-0.22				-0.40
				(0.00)				(0.00)				(0.02)				(0.05)				(0.00)
m_1					-0.11	-0.11	0.00	0.00	-0.11	-0.11	0.45	0.22	-0.10	-0.29	0.19	-0.49	-0.10	0.63	1.65	-0.65
					(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.04)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
m_2									0.00	0.00	-0.43	0.90	0.00	0.04	0.89	1.01	0.00	0.14	1.40	1.08
									(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
m_3													0.00	0.00	-0.04	-0.15	0.00	0.08	0.47	-0.31
													(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
m_4																	0.01	0.05	0.20	0.02
																	(0.01)	(0.00)	(0.00)	(0.00)
$-\log L(\theta)$	304.31	303.13	302.40	302.15	302.65	302.57	302.39	302.03	302.61	302.57	302.04	300.77	302.51	302.24	299.46	297.82	302.18	301.81	298.73	295.08
ρ	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
μ_0	761.59	762.03	761.96	761.88	761.94	762.00	761.95	761.88	761.95	762.00	761.95	761.47	761.93	761.97	761.92	763.51	761.97	761.99	761.49	763.44
μ_1	0.79	0.78	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.77	0.79	0.83	0.77	0.78	0.80	0.71	0.77	0.79	0.84	0.74
μ_2	-8.02	-8.05	-8.07	-8.08	-8.05	-8.05	-8.07	-8.06	-8.05	-8.05	-8.08	-8.10	-8.06	-8.04	-8.14	-8.12	-8.07	-8.14	-8.32	-8.20
AIC	618.62	618.25	618.81	620.31	617.30	619.15	620.77	622.06	619.21	621.15	622.08	621.54	621.02	622.48	618.91	617.64	622.37	623.62	619.46	614.16
BIC	636.50	639.71	643.84	648.92	638.75	644.18	649.38	654.25	644.25	649.75	654.27	657.30	649.63	654.66	654.67	656.97	654.55	659.38	658.79	657.07

Table 3.6: Estimation results for the UC(d, 1, 1, ρ) model. The table shows the QML estimates for the fractional UC model with a linear deterministic trend $\mu_t = \mu_0 + \mu_1 t$ (i.e. $b = 1$), an ARMA(p, q) cyclical component in the standard lag operator L, $a(L)c_t = m(L)\varepsilon_t$ (i.e. $\delta = 1$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. We consider ARMA polynomials up to order $p = 1, ..., 4$, $q = 0, ..., 4$, and μ_2 denotes the regression coefficient for the dummy in 2020Q2 as described at the beginning of section 3.5. Standard errors are denoted in parentheses and are obtained from the numerical Hessian matrix.

$(p;q)$	(1;0)	(2;0)	(3;0)	(4;0)	(1;1)	(2;1)	(3;1)	(4;1)	(1;2)	(2;2)	(3;2)	(4;2)	(1;3)	(2;3)	(3;3)	(4;3)	(1;4)	(2;4)	(3;4)	(4;4)
d	0.89	0.89	0.91	0.93	0.90	0.87	0.91	0.93	0.91	0.91	0.91	0.96	0.91	0.91	0.94	0.96	0.93	0.94	0.94	0.93
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ_η^2	10.74	10.76	3.82	1.95	10.74	9.33	3.82	1.92	3.22	3.35	3.82	1.31	3.25	3.35	1.22	1.31	1.76	2.74	4.31	4.69
	(0.03)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.06)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\sigma_{\eta\varepsilon}$	-13.39	-13.41	-5.38	-2.97	-13.39	-11.67	-5.38	-2.94	-4.54	-4.75	-5.38	-2.19	-4.58	-4.75	-2.04	-2.19	-2.79	-4.01	-5.88	-6.37
	(0.04)	(0.00)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.09)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ_ε^2	16.70	16.72	7.56	4.53	16.69	14.59	7.56	4.49	6.41	6.74	7.56	3.64	6.46	6.74	3.43	3.64	4.40	5.88	8.02	8.64
	(0.05)	(0.00)	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.16)	(0.05)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
a_1	0.94	0.94	0.97	1.01	0.94	0.20	0.97	1.01	0.92	1.27	0.97	1.35	0.92	1.27	0.74	1.35	0.88	0.59	0.67	-0.01
	(0.00)	(0.00)	(0.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
a_2		0.00	0.00	-0.04		0.70	0.00	-0.03		-0.33	0.00	-1.29		-0.33	0.41	-1.29		0.29	-0.67	-0.18
		(0.00)	(0.01)	(0.00)		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.03)		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)
a_3			-0.05	-0.05			-0.05	-0.05			-0.05	0.93			-0.27	0.93			0.85	0.50
			(0.12)	(0.00)			(0.00)	(0.00)			(0.00)	(0.02)			(0.00)	(0.00)			(0.00)	(0.00)
a_4				-0.01				-0.01				-0.09				-0.09				0.45
				(0.00)				(0.00)				(0.02)				(0.00)				(0.00)
m_1					0.00	0.73	0.00	-0.01	0.05	-0.30	0.00	-0.33	0.05	-0.29	0.27	-0.33	0.13	0.42	0.33	1.00
					(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
m_2									0.05	0.04	0.00	0.94	0.05	0.04	-0.16	0.94	0.11	0.11	0.99	1.17
									(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
m_3													0.02	0.00	0.00	0.00	0.05	0.06	0.09	0.60
													(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
m_4																	0.04	0.04	0.05	0.09
																	(0.00)	(0.00)	(0.00)	(0.00)
$-\log L(\theta)$	303.66	304.48	301.84	301.16	308.63	303.07	301.63	300.88	302.20	301.75	301.70	297.90	301.76	301.72	300.47	297.90	301.45	300.66	297.01	296.24
ρ	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
μ_0	760.25	760.09	760.53	760.65	758.13	760.41	760.34	760.64	760.58	760.54	760.45	763.34	760.51	760.54	760.75	763.34	760.86	761.00	760.94	760.77
μ_1	1.15	1.18	1.12	1.06	1.41	1.28	1.14	1.06	1.12	1.14	1.13	0.86	1.14	1.13	1.01	0.86	1.03	1.02	1.02	1.04
μ_2	-8.01	-8.00	-8.04	-8.01	-8.00	-8.08	-8.04	-8.04	-8.04	-8.06	-8.04	-8.21	-8.07	-8.06	-8.08	-8.20	-8.08	-8.12	-8.14	-8.30
AIC	617.32	620.96	617.68	618.32	629.25	620.14	619.26	619.76	618.41	619.50	621.40	615.81	619.52	621.43	620.94	617.80	620.91	621.33	616.02	616.48
BIC	635.20	642.41	642.72	646.93	650.71	645.17	647.86	651.94	643.44	648.11	653.58	651.57	648.13	653.62	656.70	657.14	653.09	657.09	655.35	659.39

Table 3.7: Estimation results for the UC(d, d, 1, ρ) model. The table shows the QML estimates for the fractional UC model with a polynomial deterministic trend of order d, $\mu_t = \mu_0 + \mu_1 t^d$ (i.e. $b = d$), an ARMA(p, q) cyclical component in the standard lag operator L, $a(L)c_t = m(L)\varepsilon_t$ (i.e. $\delta = 1$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. We consider ARMA polynomials up to order $p = 1, ..., 4$, $q = 0, ..., 4$, and μ_2 denotes the regression coefficient for the dummy in 2020Q2 as described at the beginning of section 3.5. Standard errors are denoted in parentheses and are obtained from the numerical Hessian.

$(p;q)$	(1;0)	(2;0)	(3;0)	(4;0)	(1;1)	(2;1)	(3;1)	(4;1)	(1;2)	(2;2)	(3;2)	(4;2)	(1;3)	(2;3)	(3;3)	(4;3)	(1;4)	(2;4)	(3;4)	(4;4)
d	0.84	0.77	1.02	1.01	0.72	0.77	1.00	0.93	0.72	1.06	1.03	0.93	0.72	0.78	0.78	0.93	1.07	1.26	1.07	0.91
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
b	0.92	0.93	0.91	0.92	0.92	0.93	0.91	0.93	0.92	0.93	0.92	0.93	0.92	0.95	0.95	0.93	0.92	0.59	0.91	0.93
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)	(0.03)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ^2_η	4.83	10.79	1.04	1.04	14.68	10.79	1.04	3.47	14.68	0.71	0.81	3.46	14.68	27.40	27.39	3.46	0.69	3.10	0.72	3.49
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.77)	(0.00)	(0.00)	(0.08)	(1.09)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\sigma^2_{\eta\varepsilon}$	-6.53	-13.29	-1.80	-1.79	-17.60	-13.29	-1.79	-4.84	-17.60	-1.36	-1.50	-4.84	-17.60	-31.45	-31.45	-4.84	-1.32	-4.46	-1.37	-4.87
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.92)	(0.00)	(0.00)	(0.10)	(1.23)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
σ^2_ε	8.82	16.36	3.12	3.11	21.09	16.36	3.11	6.77	21.09	2.60	2.81	6.76	21.09	36.11	36.10	6.76	2.54	6.42	2.59	6.80
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(1.10)	(0.00)	(0.00)	(0.12)	(1.39)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
a_1	0.91	0.84	1.05	1.06	0.90	0.84	1.05	0.11	0.90	1.34	0.76	0.11	0.90	0.53	0.53	0.11	0.90	-0.32	0.39	0.10
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
a_2		0.07	-0.05	-0.05		0.07	-0.05	0.88		-0.40	0.57	0.88		0.32	0.32	0.88		-0.80	-0.31	0.89
		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.02)		(0.00)	(0.00)	(0.00)		(0.00)	(0.00)	(0.00)
a_3			-0.08	-0.08			-0.08	-0.08			-0.41	-0.08			0.00	-0.08			0.72	-0.09
			(0.00)	(0.00)			(0.00)	(0.00)			(0.00)	(0.03)			(0.00)	(0.00)			(0.00)	(0.00)
a_4				-0.01				-0.04				-0.04				-0.04				-0.04
				(0.00)				(0.00)				(0.02)				(0.00)				(0.00)
m_1					-0.11	0.00	0.00	0.97	-0.11	-0.24	0.31	0.97	-0.11	0.45	0.45	0.97	0.21	0.95	0.73	0.97
					(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
m_2									0.00	0.05	-0.29	0.00	0.00	-0.01	-0.01	0.00	0.14	1.36	1.04	-0.01
									(0.01)	(0.00)	(0.00)	(0.03)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
m_3													0.00	0.05	0.05	0.00	0.07	0.68	0.29	0.03
													(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
m_4																	0.05	0.21	0.14	0.03
																	(0.00)	(0.00)	(0.00)	(0.00)
$-\log L(\theta)$	304.14	302.35	301.84	301.56	302.35	302.34	301.83	294.42	302.35	301.67	301.53	294.41	302.35	301.26	301.39	294.40	301.85	299.60	299.14	293.69
ρ	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
μ_0	760.50	760.65	760.56	760.60	760.52	760.65	760.58	761.21	760.52	760.84	760.60	761.20	760.52	761.07	761.07	761.20	760.84	755.34	760.74	761.19
μ_1	1.10	1.08	1.12	1.10	1.10	1.08	1.11	1.04	1.10	1.06	1.10	1.04	1.10	0.94	0.95	1.04	1.09	4.81	1.12	1.04
μ_2	-7.99	-8.05	-8.07	-8.06	-8.06	-8.05	-8.05	-8.07	-8.06	-8.04	-8.10	-8.07	-8.06	-8.05	-8.05	-8.07	-8.07	-8.32	-8.25	-8.10
AIC	620.28	618.70	619.69	621.11	618.70	620.68	621.65	608.83	620.70	621.33	623.06	610.83	622.70	622.51	624.78	612.81	623.70	621.20	622.28	613.37
BIC	641.73	643.73	648.29	653.30	643.73	649.29	653.84	644.59	649.30	653.52	658.82	650.16	654.88	658.27	664.12	655.72	659.46	660.54	665.19	659.86

Table 3.8: Estimation results for the UC(d, b, 1, ρ) model. The table shows the QML estimates for the fractional UC model with a polynomial deterministic trend of order b, $\mu_t = \mu_0 + \mu_1 t^b$, where $b \in B$ is estimated jointly with the other parameters via quasi-maximum likelihood. Further components are an ARMA(p, q) cyclical component in the standard lag operator L, $a(L)c_t = m(L)\varepsilon_t$ (i.e. $\delta = 1$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. We consider ARMA polynomials up to order $p = 1, ..., 4$, $q = 0, ..., 4$, and μ_2 denotes the regression coefficient for the dummy in 2020Q2 as described at the beginning of section 3.5. Standard errors are denoted in parentheses and are obtained from the inverted numerical Hessian matrix.

	UC(d, 1, d, ρ)				UC(d, d, d, ρ)				UC(d, b, d, ρ)			
p	1	2	3	4	1	2	3	4	1	2	3	4
d	1.33	1.03	1.03	1.21	0.79	0.81	0.82	0.81	1.29	0.83	0.96	0.90
	(0.06)	(0.16)	(0.04)	(0.06)	(0.08)	(0.05)	(0.14)	(0.19)	(0.01)	(0.05)	(0.10)	(0.13)
b									0.84	0.81	0.82	0.81
									(0.04)	(0.02)	(0.07)	(0.11)
σ_η^2	0.13	1.30	4.32	0.49	5.62	2.13	1.05	2.11	0.01	2.13	0.35	1.47
	(0.07)	(1.07)	(0.01)	(0.15)	(3.57)	(0.46)	(0.11)	(3.31)	(0.02)	(0.96)	(0.20)	(1.37)
$\sigma_{\eta\varepsilon}$	-0.36	-0.80	-5.76	0.07	-5.49	-1.96	-1.82	-1.99	-0.10	-1.95	-0.80	-1.29
	(0.14)	(1.11)	(0.09)	(0.06)	(2.98)	(0.43)	(0.16)	(3.40)	(0.11)	(0.92)	(0.40)	(1.51)
σ_ε^2	1.05	0.69	7.85	0.01	5.48	1.84	3.21	1.92	0.83	1.83	1.88	1.15
	(0.22)	(1.11)	(0.27)	(0.02)	(2.20)	(0.39)	(0.23)	(2.99)	(0.27)	(0.86)	(0.76)	(1.55)
a_1	0.81	1.11	1.07	1.48	0.62	1.64	1.24	1.31	0.96	1.63	1.18	1.35
	(0.00)	(0.52)	(0.13)	(0.24)	(0.11)	(0.37)	(0.13)	(0.33)	(0.15)	(0.48)	(0.32)	(0.19)
a_2		-0.46	-0.03	-0.17		-0.66	-0.10	0.37		-0.65	-0.09	0.29
		(0.31)	(0.07)	(0.28)		(0.30)	(0.20)	(0.13)		(0.39)	(0.43)	(0.15)
a_3			-0.05	-0.77			-0.16	-1.09			-0.11	-1.06
			(0.02)	(0.20)			(0.31)	(0.87)			(0.47)	(0.46)
a_4				0.41				0.40				0.41
				(0.22)				(0.50)				(0.25)
$-\log L(\theta)$	317.44	316.76	316.19	315.37	316.25	313.71	313.01	312.38	315.11	313.71	312.63	312.05
ρ	-0.99	-0.85	-0.99	0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99
μ_0	760.90	760.97	761.41	761.45	759.03	758.73	760.34	758.84	760.41	758.72	760.41	759.22
μ_1	0.72	0.75	0.85	0.76	2.37	2.06	2.01	2.07	1.73	2.07	1.94	2.14
μ_2	-8.01	-8.05	-8.05	-8.11	-8.02	-8.03	-8.06	-8.08	-8.05	-8.03	-8.06	-8.09
AIC	644.87	645.51	646.37	646.73	642.50	639.43	640.02	640.76	642.23	641.42	641.26	642.10
BIC	662.75	666.97	671.41	675.34	660.38	660.88	665.05	669.37	663.68	666.45	669.87	674.28

Table 3.9: Estimation results for the constrained UC(d, b, d, ρ) model where $|\rho| \leq .99$ is imposed. The table shows the QML estimates for the fractional UC model with a polynomial deterministic trend of order b, $\mu_t = \mu_0 + \mu_1 t^b$, where b is either set to unity (left columns), or equals d (center columns), or is estimated jointly with the other parameters (right columns). Further components are an AR(p) cyclical component in the fractional lag operator L_d (i.e. , $a(L_d)c_t = \varepsilon_t$, and thus $\delta = d$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. μ_2 denotes the regression coefficient for the dummy in 2020Q2 as described at the beginning of section 3.5. Standard errors are denoted in parentheses and are obtained from the inverted numerical Hessian matrix.

$(p;q)$	(1;0)	(2;0)	(3;0)	(4;0)	(1;1)	(2;1)	(3;1)	(4;1)	(1;2)	(2;2)	(3;2)	(4;2)	(1;3)	(2;3)	(3;3)	(4;3)	(1;4)	(2;4)	(3;4)	(4;4)
d	0.86	1.19	1.19	1.19	0.98	1.22	1.22	1.22	1.07	1.22	1.22	1.19	0.95	1.07	1.01	1.10	1.05	1.26	1.26	1.10
	(0.13)	(0.05)	(0.06)	(0.04)	(0.05)	(0.06)	(0.07)	(0.00)	(0.06)	(0.55)	(0.00)	(0.01)	(0.16)	(0.12)	(0.00)	(0.11)	(0.15)	(0.00)	(0.00)	(0.03)
σ_η^2	3.70	0.75	0.75	0.96	1.88	0.74	0.74	0.74	1.18	0.74	0.74	0.84	2.21	1.16	1.67	1.18	1.23	0.61	0.61	1.18
	(2.73)	(0.07)	(0.07)	(0.00)	(0.25)	(0.08)	(0.12)	(0.05)	(0.31)	(0.97)	(0.08)	(0.05)	(0.18)	(0.48)	(0.01)	(0.48)	(0.15)	(0.03)	(0.03)	(0.13)
$\sigma_{\eta\varepsilon}^2$	-2.21	-0.06	-0.06	-0.18	-1.70	-0.06	-0.06	-0.06	-0.43	-0.06	-0.06	-0.43	-1.67	-0.99	-1.51	-0.98	-1.24	-0.79	-0.79	-0.98
	(2.28)	(0.02)	(0.02)	(0.00)	(0.23)	(0.02)	(0.06)	(0.02)	(0.03)	(0.57)	(0.03)	(0.03)	(0.12)	(0.43)	(0.00)	(0.56)	(0.21)	(0.05)	(0.05)	(0.11)
σ_ε^2	1.34	0.01	0.01	0.04	1.57	0.01	0.01	0.01	0.23	0.01	0.01	0.23	1.39	0.86	1.39	0.83	1.37	1.06	1.06	0.83
	(1.94)	(0.00)	(0.00)	(0.00)	(0.22)	(0.00)	(0.01)	(0.00)	(0.02)	(0.09)	(0.00)	(0.02)	(0.08)	(0.38)	(0.00)	(0.62)	(0.28)	(0.06)	(0.06)	(0.08)
a_1	0.55	0.41	0.41	0.45	0.34	0.38	0.38	0.38	-0.97	0.38	0.38	-1.56	-0.57	1.04	1.35	-0.84	-0.33	-0.34	-0.34	-0.84
	(0.08)	(0.02)	(0.18)	(0.06)	(0.05)	(0.03)	(2.13)	(0.09)	(0.76)	(0.12)	(0.00)	(0.08)	(0.23)	(0.03)	(0.01)	(0.23)	(0.14)	(0.01)	(0.01)	(0.04)
a_2		-0.93	-0.93	0.03		-0.91	-0.91	-0.91		-0.91	-0.91	-0.75		-0.93	-1.26	-0.85		-0.82	-0.82	-0.85
		(0.02)	(0.06)	(0.03)		(0.03)	(0.81)	(0.13)		(0.07)	(0.01)	(0.10)		(0.03)	(0.02)	(0.29)		(0.02)	(0.02)	(0.03)
a_3			0.00	-0.36			0.00	0.00			0.00	0.36			0.30	-0.46			0.00	-0.46
			(0.19)	(0.01)			(1.96)	(0.06)			(0.04)	(0.16)			(0.01)	(0.23)			(0.02)	(0.05)
a_4				0.87				0.00				0.30				0.10				0.10
				(0.03)				(0.03)				(0.22)				(0.19)				(0.01)
m_1					0.63	0.40	0.40	0.40	1.35	0.40	0.40	1.50	1.51	-0.19	-0.36	1.29	1.32	0.96	0.96	1.29
					(0.04)	(0.29)	(2.49)	(0.01)	(1.31)	(5.17)	(0.00)	(0.16)	(0.30)	(0.12)	(0.01)	(0.09)	(0.56)	(0.04)	(0.04)	(0.00)
m_2									0.41	0.00	0.00	0.92	0.98	0.11	0.26	1.25	1.27	1.45	1.45	1.25
									(1.09)	(7.97)	(0.10)	(0.00)	(0.04)	(0.12)	(0.01)	(0.09)	(0.48)	(0.06)	(0.06)	(0.06)
m_3													0.34	0.81	0.65	0.93	1.05	0.71	0.71	0.93
													(0.12)	(0.13)	(0.02)	(0.06)	(0.47)	(0.02)	(0.02)	(0.03)
m_4																	0.63	0.26	0.26	0.00
																	(0.88)	(0.06)	(0.06)	(0.02)
$-\log L(\theta)$	317.48	313.28	313.28	312.43	316.87	312.27	312.27	312.26	316.54	312.27	312.27	311.62	315.04	310.82	310.17	308.19	314.22	308.39	308.39	308.19
ρ	-0.99	-0.99	-0.99	-0.97	-0.99	-0.99	-0.99	-0.99	-0.81	-0.99	-0.99	-0.99	-0.95	-0.99	-0.99	-0.99	-0.96	-0.99	-0.99	-0.99
μ_0	761.41	762.42	762.42	762.56	761.42	762.90	762.90	762.91	761.49	762.90	762.90	761.10	761.32	761.56	761.38	761.49	761.70	760.54	760.54	761.49
μ_1	0.76	0.76	0.76	0.75	0.73	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.70	0.71	0.80	0.72	0.84	0.84	0.80
μ_2	-8.04	-8.21	-8.21	-8.27	-8.06	-8.22	-8.22	-8.22	-8.06	-8.22	-8.22	-8.26	-8.15	-8.12	-8.12	-8.23	-8.19	-8.23	-8.23	-8.23
AIC	644.97	638.56	640.56	640.86	645.75	638.53	640.53	642.52	647.08	640.53	642.53	643.25	646.07	639.63	640.35	638.39	646.43	636.78	638.78	640.39
BIC	662.85	660.02	665.59	669.47	667.20	663.56	669.14	674.71	672.12	669.14	674.72	679.01	674.68	671.82	676.11	677.72	678.61	672.54	678.12	683.30

Table 3.10: Estimation results for the constrained UC(d, 1, 1, ρ) model where $|\rho| \leq .99$ is imposed. The table shows the QML estimates for the fractional UC model with a linear deterministic trend $\mu_t = \mu_0 + \mu_1 t$ (i.e. $b = 1$), an ARMA(p, q) cyclical component in the standard lag operator L, $a(L)c_t = m(L)\varepsilon_t$ (i.e. $\delta = 1$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. We consider ARMA polynomials up to order $p = 1, ..., 4$, $q = 0, ..., 4$, and μ_2 denotes the regression coefficient for the dummy in 2020Q2. Standard errors are denoted in parentheses and are obtained from the numerical Hessian.

$(p;q)$	(1;0)	(2;0)	(3;0)	(4;0)	(1;1)	(2;1)	(3;1)	(4;1)	(1;2)	(2;2)	(3;2)	(4;2)	(1;3)	(2;3)	(3;3)	(4;3)	(1;4)	(2;4)	(3;4)	(4;4)
d	0.76	0.82	0.83	1.25	0.78	0.82	0.79	1.25	0.82	0.79	0.78	0.78	0.83	0.78	0.78	0.84	0.83	0.82	0.78	0.79
	(0.06)	(0.18)	(0.00)	(0.06)	(0.06)	(0.20)	(0.11)	(0.06)	(0.14)	(0.06)	(0.12)	(0.14)	(0.05)	(0.00)	(0.15)	(0.00)	(0.47)	(0.01)	(0.00)	(0.00)
σ_η^2	5.04	1.21	0.91	0.76	5.00	0.92	0.31	0.76	1.33	4.41	4.23	4.23	1.09	3.71	4.23	1.74	0.99	3.14	0.07	0.35
	(1.99)	(1.52)	(0.06)	(0.09)	(4.25)	(1.28)	(0.00)	(0.06)	(0.28)	(4.07)	(2.94)	(3.84)	(1.58)	(0.18)	(4.19)	(0.00)	(0.20)	(0.48)	(0.00)	(0.02)
$\sigma_{\eta\varepsilon}$	-3.37	-1.02	-1.63	-0.07	-6.48	-0.75	-0.17	-0.07	-2.18	-5.82	-2.70	-2.70	-1.31	-2.58	-2.70	-2.11	-1.19	-3.26	-0.15	-0.78
	(2.01)	(1.35)	(0.07)	(0.03)	(4.84)	(1.13)	(0.05)	(0.01)	(0.41)	(4.65)	(2.21)	(2.88)	(1.70)	(0.26)	(3.13)	(0.00)	(0.23)	(0.50)	(0.00)	(0.04)
σ_ε^2	2.29	0.87	2.97	0.01	8.56	0.62	0.54	0.01	3.66	7.83	1.76	1.76	1.61	1.87	1.76	2.60	1.46	3.45	0.82	1.77
	(2.12)	(1.18)	(0.07)	(0.00)	(5.50)	(0.97)	(0.01)	(0.00)	(0.58)	(5.27)	(1.55)	(2.00)	(1.79)	(0.28)	(2.17)	(0.00)	(0.26)	(0.50)	(0.00)	(0.08)
a_1	0.47	1.54	1.03	0.29	0.98	1.54	0.52	0.14	0.97	0.00	0.15	0.15	0.97	-0.81	0.15	-0.14	0.96	-0.35	0.65	0.01
	(0.06)	(0.30)	(0.26)	(0.36)	(0.01)	(0.39)	(0.77)	(0.22)	(0.36)	(0.04)	(0.10)	(0.12)	(0.01)	(0.07)	(0.25)	(0.01)	(0.02)	(0.07)	(0.01)	(0.02)
a_2		-0.56	0.01	-0.07		-0.55	0.94	-0.15		0.95	-0.76	-0.76		-0.44	-0.76	1.06		-0.86	-0.58	-0.10
		(0.11)	(0.23)	(0.22)		(0.28)	(0.80)	(0.12)		(0.04)	(0.06)	(0.08)		(0.14)	(0.07)	(0.00)		(0.03)	(0.00)	(0.00)
a_3			-0.07	-0.48			-0.49	-0.61			0.37	0.37			0.37	0.37			0.87	0.48
			(0.21)	(0.22)			(1.02)	(0.13)			(0.08)	(0.08)			(0.16)	(0.00)			(0.01)	(0.02)
a_4				0.82				0.65				0.00				-0.53				0.49
				(0.34)				(0.22)				(0.05)				(0.00)				(0.02)
m_1					-0.06	0.13	0.92	0.48	0.03	0.93	0.33	0.33	0.50	1.20	0.33	0.70	0.30	0.81	0.57	1.07
					(0.05)	(0.23)	(0.09)	(0.19)	(0.15)	(0.06)	(0.09)	(0.10)	(0.33)	(0.06)	(0.24)	(0.02)	(0.06)	(0.09)	(0.01)	(0.06)
m_2									0.05	-0.04	0.86	0.86	0.01	0.86	0.86	-0.76	0.45	1.25	1.25	1.34
									(0.22)	(0.05)	(0.10)	(0.13)	(0.06)	(0.14)	(0.18)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
m_3													0.13	0.31	0.00	-0.94	-0.15	0.50	0.39	0.83
													(0.11)	(0.04)	(0.15)	(0.00)	(0.13)	(0.08)	(0.01)	(0.06)
m_4																	0.18	0.16	0.17	0.20
																	(0.10)	(0.01)	(0.00)	(0.04)
$-\log L(\theta)$	314.92	313.47	312.90	311.20	314.37	313.46	312.69	312.21	313.12	312.64	310.06	310.06	312.83	311.93	310.06	308.93	312.26	309.51	307.18	299.23
ρ	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.42	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.98	-0.99	-0.99	-0.99	-0.99	-0.65	-0.99
μ_0	759.75	758.99	760.42	764.04	759.97	759.04	759.87	765.76	760.36	760.02	759.50	759.50	760.23	760.24	759.50	759.02	760.41	758.65	759.14	758.84
μ_1	2.76	2.01	1.93	0.19	2.38	2.01	2.41	0.18	1.98	2.27	2.51	2.51	1.94	2.43	2.51	1.84	1.87	2.08	2.54	2.36
μ_2	-8.05	-8.03	-8.05	-8.20	-8.04	-8.03	-8.07	-8.17	-8.05	-8.07	-8.16	-8.16	-8.06	-8.22	-8.16	-8.08	-8.08	-8.20	-8.20	-8.30
AIC	639.85	638.95	639.81	638.39	640.74	640.92	641.39	642.42	640.24	641.28	638.12	640.12	641.65	641.87	640.12	639.87	642.52	639.02	636.35	622.46
BIC	657.73	660.40	664.84	667.00	662.20	665.95	669.99	674.60	665.27	669.88	670.31	675.88	670.26	674.05	675.88	679.20	674.70	674.78	675.69	665.37

Table 3.11: Estimation results for the constrained UC(d, d, 1, ρ) model where $|\rho| \leq .99$ is imposed. The table shows the QML estimates for the fractional UC model with a polynomial deterministic trend of order d, $\mu_t = \mu_0 + \mu_1 t^d$ (i.e. $b = d$), an ARMA(p, q) cyclical component in the standard lag operator L, $a(L)c_t = m(L)\varepsilon_t$ (i.e. $\delta = 1$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. We consider ARMA polynomials up to order $p = 1, ..., 4$, $q = 0, ..., 4$, and μ_2 denotes the regression coefficient for the dummy in 2020Q2. Standard errors are denoted in parentheses and are obtained from the numerical Hessian.

$(p;q)$	(1;0)	(2;0)	(3;0)	(4;0)	(1;1)	(2;1)	(3;1)	(4;1)	(1;2)	(2;2)	(3;2)	(4;2)	(1;3)	(2;3)	(3;3)	(4;3)	(1;4)	(2;4)	(3;4)	(4;4)
d	0.67	0.81	0.70	0.86	0.64	0.62	0.79	1.24	0.74	0.66	0.50	0.50	0.80	0.66	0.50	1.28	1.02	1.18	0.50	1.05
	(0.09)	(0.06)	(0.38)	(0.14)	(0.38)	(0.12)	(0.55)	(0.23)	(0.07)	(0.03)	(0.21)	(0.01)	(0.11)	(0.03)	(0.00)	(0.04)	(0.55)	(0.26)	(0.00)	(0.28)
b	0.82	0.82	0.82	0.81	0.83	0.77	0.82	0.83	0.83	0.83	0.73	0.73	0.81	0.83	0.73	0.89	0.83	0.73	0.73	0.79
	(0.05)	(0.05)	(0.03)	(0.00)	(0.07)	(0.02)	(0.18)	(0.03)	(0.02)	(0.03)	(0.08)	(0.01)	(0.08)	(0.03)	(0.00)	(0.07)	(0.03)	(0.09)	(0.00)	(0.05)
σ^2_η	6.07	1.19	3.22	0.85	5.75	12.73	1.53	0.02	2.67	4.64	0.27	0.27	1.38	4.64	0.27	0.32	0.20	0.70	0.27	0.12
	(6.14)	(0.36)	(2.74)	(0.06)	(7.12)	(1.64)	(4.23)	(0.09)	(0.92)	(0.17)	(0.32)	(0.00)	(0.50)	(0.17)	(0.00)	(0.11)	(1.33)	(0.73)	(0.00)	(0.20)
$\sigma^2_{\eta\varepsilon}$	-6.19	-1.01	-3.22	-0.64	-7.24	-12.34	-1.44	-0.14	-3.83	-6.04	-0.65	-0.65	-1.41	-6.04	-0.65	-0.75	-0.55	-0.89	-0.65	-0.38
	(5.85)	(0.33)	(2.78)	(0.06)	(5.62)	(1.71)	(4.20)	(0.31)	(1.12)	(0.23)	(0.55)	(0.00)	(0.52)	(0.22)	(0.00)	(0.17)	(2.46)	(0.72)	(0.00)	(0.39)
σ^2_ε	6.44	0.87	3.28	0.49	9.31	12.21	1.38	0.87	5.59	8.02	1.62	1.62	1.47	8.02	1.62	1.77	1.51	1.14	1.62	1.19
	(5.30)	(0.29)	(2.76)	(0.05)	(4.74)	(1.75)	(3.81)	(0.54)	(1.26)	(0.29)	(0.78)	(0.01)	(0.53)	(0.28)	(0.00)	(0.20)	(3.60)	(0.67)	(0.00)	(0.63)
a_1	0.97	1.54	0.90	1.39	0.96	-0.46	0.55	0.30	0.96	-0.02	0.72	0.72	0.96	-0.02	0.72	1.03	0.96	-0.33	0.72	0.07
	(0.01)	(0.30)	(0.25)	(0.22)	(0.02)	(0.27)	(0.06)	(0.10)	(0.60)	(0.09)	(0.06)	(0.01)	(0.58)	(0.08)	(0.00)	(0.00)	(0.02)	(0.05)	(0.00)	(0.25)
a_2		-0.55	0.35	0.22		0.50	0.87	1.04		0.94	-0.54	-0.54		0.94	-0.54	-0.91		-0.86	-0.54	-0.12
		(0.47)	(0.07)	(0.23)		(0.12)	(0.10)	(0.07)		(0.07)	(0.06)	(0.00)		(0.07)	(0.00)	(0.03)		(0.04)	(0.00)	(0.22)
a_3			-0.29	-1.02			-0.46	-0.30			0.78	0.78			0.78	0.99			0.78	0.52
			(0.44)	(0.26)			(0.18)	(0.08)			(0.05)	(0.01)			(0.01)	(0.06)			(0.00)	(0.16)
a_4				0.41				-0.09				0.00				-0.38				0.44
				(0.43)				(0.06)				(0.01)				(0.04)				(0.19)
m_1					-0.15	0.97	0.77	0.96	-0.05	0.86	0.20	0.20	-0.02	0.86	0.20	0.00	0.22	0.90	0.20	1.15
					(0.24)	(0.38)	(0.00)	(0.04)	(0.27)	(0.08)	(0.04)	(0.00)	(0.07)	(0.07)	(0.00)	(0.00)	(0.39)	(0.17)	(0.00)	(0.18)
m_2									0.03	-0.11	0.94	0.94	0.14	-0.11	0.94	0.91	0.16	1.38	0.94	1.49
									(0.34)	(0.16)	(0.04)	(0.02)	(0.02)	(0.15)	(0.01)	(0.02)	(0.24)	(0.18)	(0.01)	(0.11)
m_3													0.62	0.00	0.00	-0.26	0.06	0.64	0.00	0.95
													(0.11)	(0.04)	(0.01)	(0.02)	(0.11)	(0.21)	(0.01)	(0.15)
m_4																	0.06	0.22	0.00	0.30
																	(0.09)	(0.11)	(0.01)	(0.05)
$-\log L(\theta)$	314.47	313.47	312.60	311.53	313.23	313.19	311.83	311.27	313.04	311.83	304.74	304.74	312.74	311.83	304.74	304.04	312.02	305.46	304.74	296.64
ρ	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99
μ_0	760.18	758.99	759.29	759.01	760.33	758.51	758.62	760.36	760.40	760.30	758.06	758.06	757.77	760.30	758.06	758.83	760.47	757.65	758.06	758.57
μ_1	1.98	2.01	2.02	2.11	1.94	2.61	2.04	1.90	1.93	1.94	3.31	3.31	2.16	1.94	3.31	1.29	1.87	3.03	3.31	2.28
μ_2	-8.02	-8.03	-8.08	-8.09	-8.06	-8.08	-8.08	-8.08	-8.05	-8.08	-8.24	-8.24	-8.06	-8.08	-8.24	-8.24	-8.08	-8.25	-8.24	-8.32
AIC	640.94	640.94	641.21	641.05	640.46	642.37	641.67	642.53	642.08	641.66	629.48	631.48	643.47	643.66	631.48	632.09	644.05	632.92	633.48	619.28
BIC	662.40	665.97	669.82	673.23	665.50	670.98	673.85	678.29	670.69	673.85	665.24	670.81	675.65	679.42	670.81	675.00	679.81	672.25	676.39	665.76

Table 3.12: Estimation results for the constrained UC(d, b, 1, ρ) model where $|\rho| \leq .99$ is imposed. The table shows the QML estimates for the fractional UC model with a polynomial deterministic trend of order b, $\mu_t = \mu_0 + \mu_1 t^b$, where $b \in B$ is estimated via QML. Further components are an ARMA(p, q) cyclical component in the standard lag operator L, $a(L)c_t = m(L)\varepsilon_t$ (i.e. $\delta = 1$), a fractional trend $\Delta_+^d x_t = \eta_t$, and correlated innovations $\rho = \text{Corr}(\eta_t, \varepsilon_t)$ as described in section 3.3. We consider ARMA polynomials up to order $p = 1, ..., 4$, $q = 0, ..., 4$, and μ_2 denotes the regression coefficient for the dummy in 2020Q2. Standard errors are in parentheses and are obtained from the numerical Hessian.

Chapter 4

Macroeconomic forecasting with fractional factor models

4.1 Introduction

At least since the seminal work of Chamberlain and Rothschild (1983), Forni et al. (2000), and Stock and Watson (2002), factor models have become an important tool for economic analysis and forecasting. They are particularly popular in fields where strong cross sectional dependencies and large data sets are present, such as macroeconomics and finance, because they handle covariation in the cross section efficiently by condensing it into a typically small number of common latent factors. In contrast to the cross section, little attention has been paid to strong dependence of the factors in the time domain: While economic data are frequently found to be highly persistent, often non-stationary, and to exhibit long memory, the vast majority of factor models assume the factors to be stationary. By allowing for factors with different memory, this paper investigates whether combining fractional integration techniques and factor models improves the forecast performance for macroeconomic data.

A major drawback of most factor models is the inefficient use of longitudinal information. As a simple example, consider a macroeconomic panel in which some variables are found to have high integration orders around two (e.g. prices, money, and credit), others have integration orders close to one (e.g. economic output, income, and employment), and still others are stationary (e.g. interest rates). Suppose the panel is driven by a set of common factors with heterogeneous integration orders, where the factors with high memory affect only the non-stationary variables and impose cointegration relations, while the factors with low memory may affect the

entire panel. In order to apply traditional factor models, which require stationary data, to such data, the data are typically first tested for unit roots and differenced if necessary. However, since the integration order of an observable variable equals the highest integration order of all factors with non-zero loadings, all factors with lower memory are likely to be over-differenced when the data are pre-processed. Model selection criteria and model specification tests for the number of factors may then miss the over-differenced additional factors, as their corresponding eigenvalues tend to zero.

The problem of potentially non-stationary factors has been addressed by allowing for unit roots in the factors (see, e.g., Bai; 2004; Banerjee and Marcellino; 2009; Eickmeier; 2009; Chang et al.; 2009; Banerjee et al.; 2014, 2016; Barigozzi et al.; 2021). While this clearly improves the suitability of factor models for various applications in macroeconomics and finance, these models come at the cost of requiring prior assumptions about the integration orders of the factors, and typically all factors (and thus all observable variables) are assumed to be either $I(1)$ or $I(0)$. This, in turn, limits the model to factors (and data) with either perfect or short memory, reduces statistical inference about the integration orders of data and factors to prior unit root testing, and hinders a data-driven estimation of the integration orders together with the other model parameters. Misspecifying the integration orders of the observable variables and factors may bias factor estimates, lead to incorrect inference about the number of common factors, and is likely to deteriorate the forecast performance. Moreover, treating integration orders as (known) integers ignores the non-standard behavior of many economic series that are fractionally integrated, as well as the uncertainty about the true integration orders. The latter points are particularly important in macroeconomics, where the literature has provided extensive evidence for long memory and fractional cointegration in the data (see e.g. Hassler and Wolters; 1995; Baillie; 1996; Gil-Alaña and Robinson; 1997; Tschernig et al.; 2013).

While some generalizations of factor models to fractionally integrated processes exist, the literature has so far mostly considered semiparametric models. Morana (2004) suggests a frequency domain principal components estimator that allows for long memory, while Luciani and Veredas (2015), Cheung (2022), and Ergemen (2023) estimate fractionally integrated factors by a principal components estimator based on a data set in first differences. Recent parametric models have been designed with a different focus, or are much more restrictive: Ergemen and Velasco (2017) and Ergemen (2019) focus on eliminating common fractional factors, while Mesters et al.

(2016) restrict the memory of the common factors to the stationary region. In a setup closest to the one considered in this paper, Hartl and Jucknewitz (2021) introduce a parametric fractionally integrated factor model that allows for both, stationary and non-stationary factors. They decompose a panel of observable data into groups of purely fractionally integrated factors with different integration orders, plus some stationary factors that exhibit an ARMA structure. The model is then applied to analyze and forecast realized covariance matrices.

Building on the model of Hartl and Jucknewitz (2021), this paper aims to provide insights on whether fractional integration techniques have merit for at least a relevant fraction of the numerous and heterogeneous macroeconomic variables typically under study. To this end, I search for a suitable factor model formulation that decomposes a panel of macroeconomic data with heterogeneous integration orders into common factors and idiosyncratic errors. The factors may exhibit long memory and thus generate cointegration relations among the observable variables, and the memory may differ across the factors. Specifically, I study three different models that incorporate long memory into a parametric factor model setup: The first model introduces ARFIMA processes to the non-stationary factor model setup of Barigozzi et al. (2021), thus allowing for more general patterns of persistence than the usual integer-integrated specifications. Building on the work of Hartl and Jucknewitz (2021), the second model distinguishes between purely fractionally integrated factors, which determine the long-run behavior, and $I(0)$ factors, which model common cyclical dynamics of the data. The third model generalizes the pre-differencing of the data: Instead of taking first or second differences, the data enter the model in fractional differences, with the exact level of differencing determined within the model. Factors are then estimated based on the fractionally differenced series.

All models are cast in state space form so that factors, loadings, and integration orders can be estimated jointly by a combination of the Kalman recursions and maximum likelihood. To arrive at a computationally feasible formulation of the state space model that keeps the dimension of the state vector manageable, I approximate the fractional differencing polynomial by small ARMA polynomials as suggested by Hartl and Jucknewitz (2022). For a given integration order, the ARMA coefficients are fitted beforehand and are smoothed over a sequence of integration orders to obtain a continuous function that maps from the integration order to the ARMA coefficients. Thus, optimization can be carried out over the integration order parameters. The model parameters are then estimated by means of the maximum likelihood estimator, where the expectation-maximization algorithm as derived by

Hartl and Jucknewitz (2022) is used within the numerical optimization procedure. The factors are estimated using the Kalman filter and smoother. Starting values for the parametric optimization are obtained via principal components.

Just like the usual factor models, the fractional factor models are applicable to high-dimensional data, but bear several advantages: They allow joint modeling of data of different memory, and do not rely on prior assumptions about the memory of the data. Moreover, they allow for joint estimation of factors, loadings, and integration orders instead of the multi-step estimation procedure of semiparametric models. They capture cointegration through the common fractionally integrated factors, and are more robust to over-differencing.

The forecast performance of the fractional factor models is studied in a pseudo out-of-sample forecast experiment for the high-dimensional US macroeconomic data set of McCracken and Ng (2016). I provide a guided choice among the different models by considering the forecast performance for 112 macroeconomic variables. Ultimately, I find comprehensive evidence that an adequate combination of fractional integration techniques and factor models can significantly improve forecasts relative to standard factor models and other benchmarks.

The remainder of the paper is organized as follows. Section 4.2 details the construction of fractional factor models. Parameter optimization and factor estimation are discussed in section 4.3. Section 4.4 compares the forecast performance of the fractional factor models with various benchmarks in a pseudo out-of-sample forecast experiment, and section 4.5 concludes.

4.2 Fractional factor models

To begin with, consider the following factor model for possibly fractionally integrated data

$$y_t = f(\chi_t) + u_t, \qquad t = 1, ..., T, \tag{4.1}$$

where $y_t = (y_{1,t}, ..., y_{N,t})'$ holds the observable data, $f(\chi_t)$ is the common component, and $u_t = (u_{1,t}, ..., u_{N,t})'$ are the idiosyncratic disturbances. The common component $f(\chi_t)$ is driven by r common factors $\chi_t = (\chi_{1,t}, ..., \chi_{r,t})'$ that account for common short- and long-run dynamics among the y_t, while u_t is purely idiosyncratic and has a diagonal covariance matrix.

The observable $y_{i,t} \sim I(d_i^*)$ are fractionally integrated of type II with integration orders d_i^*, $d_i^* \in D = \{b \in \mathbb{R} | 0 \leq b \leq d_{max} < \infty\}$, for all $i = 1, ..., N$, and d_i^* may vary

across $i = 1, ..., N$. D rules out anti-persistence by ensuring the integration orders to be non-negative, while the upper bound d_{max} may be arbitrarily large, however it forces the parameter space for the integration orders to be bounded. An integration order d_i^* implies that the fractional difference of a series is $I(0)$, i.e. $\Delta_+^{d_i^*} y_{i,t} \sim I(0)$, $i = 1, ..., N$.

The fractional difference operator Δ_+^b depends only on the integration order b. Without subscript, it exhibits a polynomial expansion in the lag operator L of order infinite

$$\Delta^b = (1-L)^b = \sum_{j=0}^{\infty} \pi_j(b) L^j, \qquad \pi_j(b) = \begin{cases} \frac{j-b-1}{j} \pi_{j-1}(b) & j = 1, 2, ..., \\ 1 & j = 0, \end{cases} \tag{4.2}$$

where the $\pi_j(b)$ are defined recursively. The +-subscript of Δ_+^b denotes the truncation of an operator at $t \leq 0$, e.g. for an arbitrary process z_t, $\Delta_+^b z_t = \Delta^b z_t \mathbb{1}(t \geq 1) = \sum_{j=0}^{t-1} \pi_j(b) z_{t-j}$, where $\mathbb{1}(t \geq 1)$ is the indicator function that takes the value one for positive subscripts of z_{t-j}, else zero. The use of the truncated fractional difference operator reflects the type II definition of fractional integration (Marinucci and Robinson; 1999). It is required to treat the asymptotically stationary case ($b < 1/2$) alongside the non-stationary case ($b > 1/2$).

Note that traditional factor models as considered by Forni et al. (2000), Bai and Ng (2002), and Stock and Watson (2002) among others assume $d_i^* = 0$ for all $i = 1, ..., N$, while the non-stationary factor models of Bai (2004), Banerjee and Marcellino (2009), Eickmeier (2009), Banerjee et al. (2014), Banerjee et al. (2016), and Barigozzi et al. (2021) allow for $d_i^* \in \{0; 1\}$ for all $i = 1, ..., N$. Allowing for integration orders $d_i^* \in D$ includes intermediate solutions between the integer-integrated factor models, such as long-range dependent but mean-reverting processes for $0 < d_i^* < 1$, and processes that are more persistent than random walks but less persistent than quadratic stochastic trends for $1 < d_i^* < 2$. Thus, extending the parameter space for the integration orders to the real line links the integer-integrated specifications seamlessly. Due to the type II definition, the inverse fractional difference $\Delta_+^{-b} z_t$ is well defined for all $b \in D$.

The key question addressed in the remainder of this section is how to specify a factor model of the form (4.1) when the y_t are allowed to be fractionally integrated. The dynamic specification of χ_t is crucial, as it must take into account the strong persistence and possibly non-stationarity of the y_t. In addition, the functional relation between χ_t and y_t has so far been left open, and is of key importance as it determines the cointegrating properties of the model. In the search for an appropriate specifica-

tion of the general factor model in (4.1), the next three subsections introduce three different candidate models that allow for fractionally integrated y_t. Subsection 4.2.1 generalizes $I(1)$ factor models (cf. e.g. Barigozzi et al.; 2021) to non-integer integration orders by allowing for autoregressive fractionally integrated factors. Subsection 4.2.2 decomposes the factors into purely fractionally integrated components that determine the long-run behavior of the y_t, and short memory components that reflect the short-run behavior of the y_t, and builds on the model of Hartl and Jucknewitz (2021). Furthermore, subsection 4.2.3 generalizes the pre-differencing of standard factor models to fractional differencing.

4.2.1 Dynamic fractional factor models

Consider a simple multivariate unobserved components model

$$y_t = \Lambda f_t + u_t, \qquad t = 1, ..., T, \tag{4.3}$$

where $f(\chi_t) = \Lambda f_t$ in (4.1), $f_t = (f_{1,t}, ..., f_{r,t})'$ holds the r common factors, Λ is a $N \times r$ matrix of factor loadings that is assumed to have full column rank, and the errors u_t account for idiosyncratic dynamics. The latent factors are assumed to follow r fractionally integrated autoregressive processes

$$B_j(L)\Delta_+^{d_j} f_{j,t} = \zeta_{j,t}, \qquad j = 1, ..., r, \tag{4.4}$$

where $B_j(L) = 1 - \sum_{k=1}^{p} B_{j,k}L^k$ is a stable lag polynomial. For the pervasive shocks that drive f_t, it is assumed that $(\zeta_{1,t}, ..., \zeta_{r,t})' = \zeta_t \sim \text{NID}(0, Q)$, where Q is diagonal. A matrix formulation of (4.4) follows directly by defining $d = (d_1, ..., d_r)'$, the matrix polynomials $D(d) = \text{diag}(\Delta_+^{d_1}, ..., \Delta_+^{d_r})$ and $B(L) = \text{diag}(B_1(L), ..., B_r(L))$, such that $B(L)D(d)f_t = \zeta_t$.

The errors $u_{i,t}$ are assumed to be mutually independent and are allowed to be autocorrelated

$$\rho_i(L)u_{i,t} = \xi_{i,t}, \qquad \xi_{i,t} \sim \text{NID}(0, \sigma_{\xi_i}^2), \qquad i = 1, ..., N, \tag{4.5}$$

where $\rho_i(L) = 1 - \sum_{k=1}^{p_i} \rho_{i,k}L^k$ is a stable lag polynomial.

The model may explain various degrees of common persistence that characterize the data by common components with long memory. For $d_1 = ... = d_r = 0$, the model nests the approximate dynamic factor model of Stock and Watson (2002), while $d_j \in \{0, 1\}$, $j = 1, ..., r$, yields a nonstationary dynamic factor model with

$I(1)$ factors as considered by Barigozzi et al. (2021). Therefore, the model can be interpreted as a fractional generalization that neither requires prior differencing of the data, nor prior assumptions about the integration orders.

4.2.2 Dynamic orthogonal fractional components

A more parsimonious factor model is proposed by Hartl and Jucknewitz (2021). Their model distinguishes between r_1 purely fractionally integrated factors $f_t^{(1)} = (f_{1,t}^{(1)}, ..., f_{r_1,t}^{(1)})'$, that establish cointegration relations among the y_t, and r_2 stationary autoregressive components $f_t^{(2)} = (f_{1,t}^{(2)}, ..., f_{r_2,t}^{(2)})'$, that account for common short-run behavior. I consider a slight modification that allows for autocorrelated idiosyncratic errors. The general framework for the dynamic orthogonal fractional components model is then given by

$$y_t = \begin{bmatrix} \Lambda^{(1)} & \Lambda^{(2)} \end{bmatrix} \begin{pmatrix} f_t^{(1)} \\ f_t^{(2)} \end{pmatrix} + u_t, \qquad t = 1, ..., T, \tag{4.6}$$

$$\Delta_+^{d_j} f_{j,t}^{(1)} = \zeta_{j,t}^{(1)}, \qquad j = 1, ..., r_1, \tag{4.7}$$

$$B_j^{(2)}(L) f_{j,t}^{(2)} = \zeta_{j,t}^{(2)}, \qquad j = 1, ..., r_2, \tag{4.8}$$

$$\rho_i(L) u_{i,t} = \xi_{i,t}, \qquad i = 1, ..., N, \tag{4.9}$$

for all $t = 1, ..., T$ and $r = r_1 + r_2 \leq N$. As before, $\rho_i(L) = 1 - \sum_{k=1}^{p_i} \rho_{i,k} L^k$ is a stable polynomial, and the N idiosyncratic shocks $\xi_t = (\xi_{1,t}, ..., \xi_{N,t})'$ are assumed to be independent Gaussian white noise processes $\xi_{i,t} \sim \text{NID}(0, \sigma_{\xi_i}^2)$, $i = 1, ..., N$. For the pervasive shocks $\zeta_t^{(1)} = (\zeta_{1,t}^{(1)}, ..., \zeta_{r_1,t}^{(1)})'$, $\zeta_t^{(2)} = (\zeta_{1,t}^{(2)}, ..., \zeta_{r_2,t}^{(2)})'$, it is assumed that $\text{vec}(\zeta_t^{(1)}, \zeta_t^{(2)}) \sim \text{NID}(0, Q)$ where Q is diagonal. In addition, the errors u_t are assumed to be independent of the factors f_t.

Define $B^{(2)}(L) = \text{diag}(B_1^{(2)}(L), ..., B_{r_2}^{(2)}(L))$, $D^{(1)}(d) = \text{diag}(\Delta_+^{d_1}, ..., \Delta_+^{d_{r_1}})$. Then, it follows immediately that the model is nested in the setup of subsection 4.2.1 for $f_t = \text{vec}(f_t^{(1)}, f_t^{(2)})$, $B(L) = \text{diag}(I, B^{(2)}(L))$, and $D(d) = \text{diag}(D^{(1)}(d), I)$. In terms of (4.1) the model specifies $f(\chi_t) = \Lambda^{(1)} f_t^{(1)} + \Lambda^{(2)} f_t^{(2)}$.

Note that the Gaussian white noise assumption on ζ_t together with Q being diagonal yields r orthogonal factors f_t. Moreover, since u_t, ζ_t are assumed to be independent, any correlation among the variables in y_t stems from the common long- and short-run components $f_t^{(1)}$ and $f_t^{(2)}$.

4.2.3 Dynamic factor models in fractional differences

The third model takes fractional differences of the observable variables to arrive at a short memory model, where all components are at most $I(0)$. The model differs from those in subsections 4.2.1 and 4.2.2 in that it eliminates fractional integration from the factors. Define

$$\Delta_+^{d_i^*} y_{i,t} = \Lambda_i f_t + \xi_{i,t}, \qquad t = 1, ..., T, \quad i = 1, ..., N, \tag{4.10}$$

$$B_j(L) f_{j,t} = \zeta_{j,t}, \qquad j = 1, ..., r, \tag{4.11}$$

where Λ_i is a r-dimensional row vector holding the loadings for $y_{i,t}$. As before, letting $y_t = (y_{1,t}, ..., y_{N,t})'$ denote the observable variables, $\Lambda = [\Lambda_1', ..., \Lambda_N']'$ the factor loadings, $f_t = (f_{1,t}, ..., f_{r,t})'$ the r latent factors, and $\xi_t = (\xi_{1,t}, ..., \xi_{N,t})'$ the idiosyncratic disturbances, (4.10) can be written as

$$D(d^*) y_t = \Lambda f_t + \xi_t, \tag{4.12}$$

where $D(d^*) = \text{diag}(\Delta_+^{d_1^*}, ..., \Delta_+^{d_N^*})$ is the $N \times N$ differencing matrix for the integration orders $d^* = (d_1^*, ..., d_N^*)'$. Contrary to the models of subsections 4.2.1 and 4.2.2, the factor model is now set up based on the pre-differenced y_t. In the notation of (4.1), the dynamic factor model in fractional differences specifies $f(\chi_t) = D(-d^*)\Lambda f_t$ and $u_t = D(-d^*)\xi_t$. Thus, the common component can be obtained by taking inverse fractional differences, that is by multiplying Λf_t with $D(-d^*)$.

By defining $B(L) = \text{diag}(B_1(L), ..., B_r(L))$ as in subsections 4.2.1 and 4.2.2, the factors f_t can be written as a diagonal VAR, $B(L) f_t = \zeta_t$, where $\zeta_t = (\zeta_{1,t}, ..., \zeta_{r,t})'$. The idiosyncratic and pervasive shocks are assumed to be orthogonal and to follow independent Gaussian white noise processes $\xi_t \sim \text{NID}(0, H)$ and $\zeta_t \sim \text{NID}(0, Q)$.

Taking fractional differences prior to estimating a factor model generalizes the pre-differencing of standard factor models to the fractional domain. In fractional differences, the model is an approximate dynamic factor model.

4.3 Estimation

This section discusses both, the estimation of the latent factors and of the model parameters for the fractional factor models as introduced in section 4.2. The factors are estimated by the means of the Kalman recursions, which requires to cast the fractional factor models in state space form. One particular challenge is to arrive at a computationally feasible state space representation of the fractional components: As

(4.2) illustrates, the fractional differencing polynomial admits a polynomial expansion of order infinite that is truncated at lag T by the type II definition of fractional integration. Thus, an exact state space representation of a single fractionally integrated factor requires a state vector of dimension $T-1$, which makes the Kalman recursions computationally infeasible even for moderate T. Therefore, subsection 4.3.1 introduces an approximation to the fractional differencing polynomial by using small ARMA polynomials. As noted by Hartl and Jucknewitz (2022), small ARMA polynomials are able to resemble the dynamics of the fractional differencing polynomial well for both integration orders $b < 1/2$ and $b > 1/2$, and keep the dimension of the state vector manageable. Subsection 4.3.2 then details the state space representation of the three fractional factor models.

Estimates for the factors via the Kalman recursions are obtained conditional on a parameter vector that contains the model parameters, i.e. the integration orders, the factor loadings, the autoregressive coefficients, and the variances of pervasive and idiosyncratic shocks. Since the true parameter vector is unobservable, it is estimated by maximum likelihood. To maximize the likelihood, I use the expectation-maximization (EM) algorithm, which was derived for fractional factor models by Hartl and Jucknewitz (2022), and is briefly described in subsection 4.3.3

Finally, as starting values are required for maximum likelihood estimation, subsection 4.3.4 discusses how to determine a suitable initial estimate for the parameter vector.

4.3.1 Approximations for the fractional differencing polynomial

The literature has considered a variety of approximations for long memory processes: For an arbitrary process $\Delta_+^b z_t = e_t$ with e_t white noise, Palma (2007, section 4.2) suggests to truncate the autoregressive representation of the fractional differencing polynomial after a certain lag m, i.e. $\sum_{j=0}^{m} \pi_j(b) z_{t-j} \approx e_t$, whereas Chan and Palma (1998) suggest to truncate the MA representation, i.e. $z_t \approx \sum_{j=0}^{m} \pi_j(-b) e_{t-j}$. In a simulation study, Hartl and Jucknewitz (2022) show that fitting small ARMA(v,w) models with $v, w \in \{3,4\}$ to approximate the fractional differencing polynomial Δ_+^{-b} clearly outperforms AR and MA approximations, and yields an approximation error that is hardly visible even for large T and non-stationary integration orders.

To illustrate the idea of approximating the fractional differencing polynomial by an ARMA polynomial, consider again the generic process $z_t = \Delta_+^{-b} e_t$ where e_t is

standardized white noise. An ARMA approximation for z_t is then given by

$$\tilde{z}_t = \left[\frac{1 + m_1 L + ... + m_w L^w}{1 - a_1 L - ... - a_v L^v}\right]_+ e_t = \sum_{j=0}^{t-1} \tilde{\pi}_j(\varphi) e_{t-j},$$

for finite v, w, where $\varphi = (a_1, ..., a_v, m_1, ..., m_w)'$, and all coefficients in φ must be made functionally dependent on b to approximate z_t by $\tilde{z}_t$. To achieve the latter, note that the approximation error is $\tilde{z}_t - z_t = \sum_{j=0}^{t-1} \left[\tilde{\pi}_j(\varphi) - \pi_j(-b)\right] e_{t-j}$, for given t, b, φ, so that the mean squared error (MSE) is $\mathrm{E}\left[(\tilde{z}_t - z_t)^2\right] = \sum_{j=0}^{t-1} \left[\tilde{\pi}_j(\varphi) - \pi_j(-b)\right]^2$. Averaging over all $t = 1, ..., T$ yields the objective function for a given b that is minimized to obtain an estimate for φ

$$\hat{\varphi}_T(b) = \arg\min_{\varphi} \mathrm{MSE}_T^b(\varphi), \qquad \mathrm{MSE}_T^b(\varphi) = \frac{1}{T} \sum_{t=1}^{T} \sum_{j=0}^{t-1} \left[\tilde{\pi}_j(\varphi) - \pi_j(-b)\right]^2. \quad (4.13)$$

Consequently, for a given b and the sample size T, (4.13) yields the optimal ARMA coefficients to approximate the fractional differencing polynomial in terms of the MSE. To obtain a smooth function that maps from b to the respective ARMA coefficients, the optimization (4.13) is carried out over a reasonable grid of b.[1] Next, the ARMA coefficients are smoothed over the grid for b using cubic regression splines. This yields a continuous, differentiable function $\varphi_T(b)$ that maps from b to the respective ARMA coefficients. Thus, optimization can be carried out over the integration order parameters. Further technical details and several simulation studies are contained in Hartl and Jucknewitz (2022). For the purely fractional factors of subsection 4.2.2 I use ARMA$(4, 4)$ polynomials to approximate the fractional differencing polynomials, as suggested by Hartl and Jucknewitz (2022). For the autoregressive fractionally integrated factors of subsection 4.2.1, the approximation quality of ARMA polynomials is not clear, and using ARMA$(4, 4)$ polynomials was found to deteriorate the estimates for the autoregressive coefficients. Therefore, I choose pure AR(5) polynomials to approximate the fractional differencing polynomial for the model in subsection 4.2.1.

Note that the main reason for using ARMA polynomials as approximations to the fractional differencing polynomial is to reduce the computational burdens of the Kalman recursions: While the exact state space representation of a single fractionally integrated process $z_T = \Delta_+^{-b} e_T$ requires a state vector of dimension $T - 1$, ARMA$(4, 4)$ approximations can be represented by a state vector of dimension five, as will become clear in what follows.

[1] I use $b \in [-0.5, 2.2]$

4.3.2 State space representation of fractional factor models

With a computationally feasible approximation of the fractional differencing polynomial at hand, the fractional factor models can be cast in state space form. The general form of a state space model for an N-dimensional vector of observable variables $\tilde{y}_t$ is

$$\tilde{y}_t = Z\alpha_t + \xi_t, \qquad \alpha_{t+1} = F\alpha_t + R\zeta_{t+1}, \tag{4.14}$$

where the first equation is termed the measurement equation. It maps the state vector α_t to the observable variables $\tilde{y}_t$ by the system matrix Z, while ξ_t accounts for serially uncorrelated disturbances with mean zero and diagonal covariance matrix $H = \text{Var}(\xi_t)$. The second equation is a first-order Markov process and is called the transition equation. It determines the development of the system via the transition matrix F. ζ_t are the systematic shocks that feed into the transition equation and equal the innovations of the factors. To uniquely identify the factor loadings, the variance of the factor innovations is set to unity, i.e. $Q = \text{Var}(\zeta_t) = I$. The system matrices Z, F, R, as well as the states α_t differ for the three fractional factor models and are derived separately in what follows.

Dynamic fractional factor models To begin with, consider the autoregressive fractionally integrated factors of subsection 4.2.1. As discussed in the previous subsection, the fractional differencing polynomial there is approximated by an AR polynomial, where the coefficients are made functionally dependent on the integration order. Letting $a(L, -b) \approx \Delta_+^b$ denote the respective AR polynomial for integration order b and fixed T. Then, the factors of subsection 4.2.1 are approximated by

$$\zeta_{j,t} = B_j(L)\Delta_+^{d_j} f_{j,t} \approx B_j(L)a(L, -d_j)_+\tilde{f}_{j,t}, \qquad j = 1, ..., r, \quad t = 1, ..., T.$$

To arrive at a matrix representation, define the matrix polynomial $A(L, -d) = I - \sum_{j=1}^{v} A_j(-d)L^j$, $A_j(-d) = \text{diag}(a_j(-d_1), ..., a_j(-d_r))$, $j = 1, ..., v$. Then, one has $B(L)A(L, -d) = \sum_{k=0}^{p+v}\sum_{l=0}^{k} B_l A_{k-l}(-d)L^k$ where $A_0(-d) = B_0 = -I$, $A_l(-d) = 0\ \forall l > v$, and $B_l = 0\ \forall l > p$.

Next, note that by (4.5) the idiosyncratic errors u_t are allowed to be autocorrelated. As suggested by Jungbacker and Koopman (2015), the model can be adjusted for idiosyncratic autocorrelation by manipulating the measurement equation, i.e. by

defining

$$\tilde{y}_{i,t} = y_{i,t} - \sum_{j=1}^{p_i} \rho_{i,j} y_{i,t-j}, \qquad \forall i = 1, ..., N. \tag{4.15}$$

Again, for a matrix representation of (4.15), collect $\tilde{y}_t = (\tilde{y}_{1,t}, ..., \tilde{y}_{N,t})'$ and define $\Psi_j = \text{diag}(\rho_{1,j}, ..., \rho_{N,j})$ with $\rho_{i,j} = 0$ for all $j > p_i$. Then $\tilde{y}_t = y_t - \sum_{j=1}^{\max(p_i)} \Psi_j y_{t-j}$.

With an appropriate matrix representation for the idiosyncratic and systematic terms at hand, the system matrices for the model in (4.3), (4.4), and (4.5) can be defined: For the transition matrix, let $s = \max(p + v, \max(p_i) + 1)$, such that

$$F = \begin{bmatrix} B_1 + A_1(-d) & \cdots & -\sum_{l=0}^{s-1} B_l A_{s-1-l}(-d) & -\sum_{l=0}^{s} B_l A_{s-l}(-d) \\ I & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix},$$

where F is $(sr \times sr)$ and depends on the parameters in d and $B(L)$. For the measurement equation, let

$$Z = \begin{bmatrix} \Lambda & -\Psi_1 \Lambda & \cdots & -\Psi_{s-1} \Lambda \end{bmatrix},$$

and thus Z is $(N \times sr)$ and depends on the parameters in Λ and $\rho_i(L)$, $i = 1, ..., N$. Furthermore, let $R = [I_r, 0]'$ be a $(sr \times r)$ matrix that allows for non-zero innovations in the first r rows of the transition equation, while $\alpha_t = (\tilde{f}_t', ..., \tilde{f}_{t-s+1}')'$ is a vector of dimension sr that holds the factors. To identify the r factors, the first r rows of Λ are restricted to be lower triangular.

Dynamic orthogonal fractional components Next, consider the model in subsection 4.2.2. Again as discussed in subsection 4.3.1, the fractional differencing polynomial is approximated by an ARMA polynomial, where the respective approximation is given by $f_t^{(1)} \approx [M(L,d)A(L,d)^{-1}]_+ \zeta_t^{(1)} = \tilde{f}_t^{(1)}$. The matrix ARMA polynomials are $M(L,d) = I + M_1(d)L + ... + M_w(d)L^w$, $M_j(d) = \text{diag}(m_j(d_1), ..., m_j(d_{r_1}))$, $A(L,d) = I - A_1(d)L - ... - A_v(d)L^v$, $A_j(d) = \text{diag}(a_j(d_1), ..., a_j(d_{r_1}))$, and $M_j(d) = 0$ $\forall j > w$, $A_j(d) = 0$ $\forall j > v$.

As before, autocorrelation in the idiosyncratic errors u_t is eliminated by transforming $\tilde{y}_t = y_t - \sum_{j=1}^{\max(p_i)} \Psi_j y_{t-j} = \Psi(L) y_t$, with coefficients Ψ_j as defined below

(4.15). Multiplication of (4.6) by $\Psi(L)$ yields

$$\Psi(L)y_t = \tilde{y}_t = \Psi(L)\begin{bmatrix}\Lambda^{(1)} & \Lambda^{(2)}\end{bmatrix}\begin{pmatrix} f_t^{(1)} \\ f_t^{(2)}, \end{pmatrix} + \xi_t, \tag{4.16}$$

where $\Psi(L)\Lambda^{(1)}f_t^{(1)} \approx \Psi(L)\Lambda^{(1)}[M(L,d)A(L,d)^{-1}]_+\zeta_t^{(1)} = \Psi(L)\Lambda^{(1)}\tilde{f}_t^{(1)}$.

To arrive at the state space representation (4.14), partition the system matrices into $F = \text{diag}(F^{(1)}, F^{(2)})$, $Z = \begin{bmatrix}Z^{(1)} & Z^{(2)}\end{bmatrix}$, and $R = \text{diag}(R^{(1)}, R^{(2)})$, where the superscript (1) refers to $f_t^{(1)}$, while the superscript (2) refers to $f_t^{(2)}$. Starting with the approximate fractionally integrated factors, for a minimal representation define the r vector $\tilde{\mu}_t = M(L,d)_+^{-1}\tilde{f}_t^{(1)}$, such that $A(L,d)_+\tilde{\mu}_t = \zeta_t^{(1)}$, and let $s_1 = \max(v, w + \max(p_i) + 1)$. Next, place $\tilde{\mu}_t$ in the state vector $\alpha_t^{(1)} = (\tilde{\mu}_t', ..., \tilde{\mu}_{t-s_1+1}')'$, such that multiplication with the MA polynomial yields $\begin{bmatrix}I & M_1(d) & \cdots & M_{s_1-1}(d)\end{bmatrix}\alpha_t^{(1)} = \tilde{f}_t^{(1)}$. Moreover, $\tilde{\mu}_{t+1} = \begin{bmatrix}A_1(d) & \cdots & A_{s_1}(d)\end{bmatrix}\alpha_t^{(1)} + \zeta_t^{(1)}$, which defines the $(s_1 r_1 \times s_1 r_1)$ dimensional transition matrix

$$F^{(1)} = \begin{bmatrix} A_1(d) & \cdots & A_{s_1-1}(d) & A_{s_1}(d) \\ I & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix}.$$

In the measurement equation, (4.16) needs to be taken into account, which yields $\Psi(L)\Lambda^{(1)}\tilde{f}_t^{(1)} = \Psi(L)\Lambda^{(1)}\begin{bmatrix}I & M_1(d) & \cdots & M_{u_1-1}(d)\end{bmatrix}\alpha_t^{(1)}$, and thus defines the $(N \times s_1 r_1)$ matrix

$$Z^{(1)} = \begin{bmatrix}\Lambda^{(1)} & -\sum_{k=0}^{1}\Psi_k\Lambda^{(1)}M_{1-k}(d) & \cdots & -\sum_{k=0}^{s_1-1}\Psi_k\Lambda^{(1)}M_{s_1-1-k}(d)\end{bmatrix},$$

where $\Psi_0 = -I$, and $Z^{(1)}$ solely depends on $\Lambda^{(1)}$, d, and the $\rho_{i,j}$. As before, $R^{(1)} = [I_{r_1}, 0]'$ is a $(s_1 r_1 \times r_1)$ matrix allowing the first r_1 rows of the transition equation to be influenced by $\zeta_t^{(1)}$.

Turning to the stationary autoregressive factors, it follows directly from (4.8) that for $\alpha_t^{(2)} = (f_t^{(2)'}, ..., f_{t-s_2+1}^{(2)'})'$ and $s_2 = \max(p, \max(p_i) + 1)$

$$F^{(2)} = \begin{bmatrix} B_1^{(2)} & \cdots & B_{s_2-1}^{(2)} & B_{s_2}^{(2)} \\ I & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix},$$

where $F^{(2)}$ is $(s_2 r_2 \times s_2 r_2)$ dimensional, and $B_j^{(2)} = 0 \; \forall j > p$. For the measurement equation, one has the $(N \times s_2 r_2)$ matrix

$$Z^{(2)} = \begin{bmatrix} \Lambda^{(2)} & -\Psi_1 \Lambda^{(2)} & \cdots & -\Psi_{s_2-1} \Lambda^{(2)} \end{bmatrix},$$

and $R^{(2)} = [I_{r_2}, 0]'$ is $(s_2 r_2 \times r_2)$. Finally, $\alpha_t = (\alpha_t^{(1)'}, \alpha_t^{(2)'})'$.

To identify the stationary autoregressive factors, a lower triangular structure is imposed on the first r_2 rows of $\Lambda^{(2)}$. The purely fractionally integrated factors are identified by their spectrum whenever $d_1 \neq d_2 \neq ... \neq d_{r_1}$. For blocks of identical memory within $f_t^{(1)}$, similar identifying restrictions have to be imposed on the respective loadings.

Dynamic factor models in fractional differences Since the factors of the third model (4.10) are stationary autoregressive processes, a state space representation as in (4.14) follows immediately by defining $\tilde{y}_t = (\Delta_+^{d_1^*} y_{1,t}, ..., \Delta_+^{d_N^*} y_{N,t})'$. The factors enter the state vector directly, whereas their AR coefficients in (4.11) are contained in F. Furthermore, the factor loadings enter the $(N \times rp)$ matrix Z, and R is again a $(rp \times r)$ selection matrix

$$\alpha_t = \begin{pmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{pmatrix}, \qquad F = \begin{bmatrix} B_1 & \cdots & B_{p-1} & B_p \\ I & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & 0 \end{bmatrix}, \qquad Z = \begin{bmatrix} \Lambda' \\ 0 \\ \vdots \\ 0 \end{bmatrix}', \qquad R = \begin{bmatrix} I_r \\ 0 \end{bmatrix}.$$

For identification of the factors, the first r rows of Λ are again restricted to be lower triangular.

4.3.3 Parameter estimation

Turning to the estimation of the model parameters, collect the unknown parameters in d, Λ, $B_1, ..., B_p$, $\rho_{1,1}, ..., \rho_{N,p_N}$, and H, that enter the system matrices of the state space model F, Z, and H, in a parameter vector θ. θ is estimated following the suggestions of Hartl and Jucknewitz (2022), who derive an EM algorithm for maximum likelihood estimation of fractional factor models. The EM algorithm bears the advantage of being relatively robust to starting values and converges rapidly towards the neighborhood of the optimum of the likelihood (Quah and Sargent; 1993; Doz et al.; 2012; Jungbacker and Koopman; 2015). However, the EM algorithm is found to be relatively slow around the optimum. Therefore, I switch to gradient-

based optimization routines with an analytical solution to the score function of the likelihood after a certain number of iterations. The whole parameter estimation procedure adopts the approach of Hartl and Jucknewitz (2022) and thus is only summarized briefly in what follows.

The EM algorithm is based on the expected complete data likelihood as given in Hartl and Jucknewitz (2022, eqn. (9)), and consists of product moments of the state vector, the observable y_t, and the measurement and transition disturbances ξ_t and ζ_t, as well as of the system matrices $Q = I$, H, R, F, and Z. In the expectation step, the product moments of α_t, y_t, ξ_t and ζ_t are computed via the Kalman filter and smoother given some realization of the parameter vector $\theta_{\{j\}}$. Next, the maximization step maximizes the expected complete data likelihood given the product moments from the expectation step and yields an updated estimate $\theta_{\{j+1\}}$. The procedure repeats until a certain level of convergence or a certain number of iterations is reached. Next, the resulting parameter estimates from the EM algorithm are used as starting values for gradient-based likelihood maximization via the BFGS algorithm, which uses the analytical solution for the score vector of Hartl and Jucknewitz (2022).

However, the EM algorithm requires an initial vector $\theta_{\{0\}}$ as starting value for the first run of the expectation step. Therefore, the next subsection details how to determine an initial guess for $\theta_{\{0\}}$ via the principal components estimator.

4.3.4 Starting values for parameter optimization

As shown by Zhang et al. (2019), fractionally integrated factors can be estimated consistently by the non-parametric method of principal components (PC) given that the idiosyncratic disturbances are stationary.[2] Therefore, initial estimates for the factors are obtained via principal components. Based on them, the model parameters can be obtained, which is discussed separately for the three different fractional factor models in section 4.2.

Dynamic fractional factor models The common components of the model in section 4.2.1 are assumed to follow r independent autoregressive fractionally integrated processes, and their correlation is zero for all leads and lags. To ensure the

[2]Note that, although stationarity of the idiosyncratic terms is assumed for all three models considered in this paper, this assumption is a very strong one and is likely to be violated for various applications. Whenever the idiosyncratic terms are non-stationary, PC are inconsistent. However, under a violation of the stationarity assumption the factor loadings can still be estimated consistently via PC when the data is pre-differenced such that the differenced idiosyncratic terms are stationary. The factors are then obtained by projecting the data onto the space spanned by the loadings, see Barigozzi et al. (2021) and Cheung (2022).

latter, the PC estimates are rotated via the method of Matteson and Tsay (2011) to obtain dynamic orthogonal components. The parameters in (4.4) are estimated by maximizing the likelihood function for a fractionally integrated VAR (see Nielsen; 2004) which yields estimates for $d_1, ..., d_r,\ B_1, ..., B_p$.[3] For some data sets the assumption of orthogonal factors may be violated. Then, the diagonal assumption on $B(L)$ can be dropped, which does not affect the identification of the fractional factor VAR but increases the number of parameters to be estimated. Factor loadings Λ in (4.3) are estimated by ordinary least squares (OLS).

Dynamic orthogonal fractional components To derive an estimator for the dynamic parameters of the model in section 4.2.2, one first needs to distinguish between the space spanned by the purely fractional factors and the stationary autoregressive components. The two factor subspaces of $f_t^{(1)}$ and $f_t^{(2)}$ are identified up to a rotation by estimating the fractional cointegration subspace and its orthogonal complement via the semiparametric method of Chen and Hurvich (2006), who use eigenvectors of an averaged periodogram matrix of the first m Fourier frequencies to estimate the fractional cointegration subspace. Orthogonal series within the fractional and non-fractional factors are then obtained by applying the decorrelation method of Matteson and Tsay (2011). The resulting fractional and non-fractional factor estimates are denoted as $\hat{f}_t^{(1)}$ and $\hat{f}_t^{(2)}$ respectively.

Factor loadings Λ in (4.6) and AR coefficients in (4.8) are estimated by OLS. Estimates for the integration orders of the common components in (4.7) are obtained by maximizing the likelihood of the r_1 ARFIMA(0, d_j, 0) processes, $j = 1, ..., r_1$.

Dynamic factor models in fractional differences Due to the stationary representation of the model in section 4.2.3 the PC estimator of Bai and Ng (2002) is directly applicable. The factors are again decorrelated by the means of dynamic orthogonal components of Matteson and Tsay (2011). As before, if the factors do not admit a dynamically orthogonal representation, the assumption of a diagonal factor VAR can be dropped and replaced by a non-diagonal VAR. The autoregressive coefficients for the r common factors in (4.11) together with their factor loadings in (4.10) are estimated by OLS.

Autocorrelated idiosyncratic terms An estimate for the idiosyncratic errors is obtained via $\hat{u}_t = y_t - \hat{\Lambda}\hat{f}_t$. Since the errors are assumed to follow N indepen-

[3]Note that this is the same as fitting an autoregressive fractionally integrated model to each of the r factors separately.

dent autoregressive processes, the AR parameters are estimated via OLS. From the residuals, an estimate for $\sigma^2_{\xi_i}$ is obtained via $\hat{\sigma}^2_{\xi_i} = T^{-1} \sum_{t=1}^{T} \hat{\xi}^2_{i,t}$, $i = 1, ..., N$.

4.4 Macroeconomic forecasting

Having discussed the estimation of the common factors together with the unknown parameters for the three fractionally integrated factor models in subsections 4.2.1–4.2.3, I next investigate their forecast performance when neither the DGP, nor the starting values, nor the number of factors, are known to the researcher. The underlying data set is the so-called FRED-MD by McCracken and Ng (2016). It consists of 112 macroeconomic variables, spans from January 1960 to December 2016, and is in monthly frequency. Subsection 4.4.1 outlines the forecast design and the model specifications, while empirical results are presented in subsection 4.4.2.

4.4.1 Forecast design and model specification

Starting with the forecast design, the forecast performance of the three different factor models is evaluated in a pseudo out-of-sample forecast experiment using a recursive window scheme. Forecasts are made for horizons $h = 1, ..., 12$, where the first forecast period is January 2000, whereas the last is December 2016, leading to 204 forecasts for 112 variables and 12 horizons. Forecast performance is then evaluated based on the mean squared prediction error (MSPE).

For the first forecast (January 2000), starting values for parameter estimation are obtained as described in subsection 4.3.4, while all subsequent periods use the optimized parameters from the preceding step as starting values. However, I also report estimates via the semiparametric approach as described in subsection 4.3.4 for all periods to evaluate the relative performance of the parametric models in comparison to the semiparametric counterparts. To distinguish between the parametric and the semiparametric models, the former are denoted as **KF** for Kalman filter, while the latter are denoted as **PC** for principal components. Abbreviations for the three fractional factor models are: Dynamic fractional factor model (**DFFM**) in subsection 4.2.1, dynamic orthogonal fractional components (**DOFC**) in subsection 4.2.2, and dynamic factor model in fractional differences (**DFFD**) in subsection 4.2.3.

To also compare the forecast performance with competing, non-fractional models, I include forecasts for four different benchmark models: The first benchmark is an autoregressive model (**AR**) where the AR lag order is chosen via the Akaike Information Criterion for each $y_{i,t}$. Moreover, two approximate dynamic factor mod-

els in the spirit of Stock and Watson (2002) are considered as further benchmarks. Both are estimated via principal components based on a pre-differenced data set. The second benchmark is denoted as (**PC**) and has the factor model representation $\Delta^{k_i} y_{i,t+h} = \Lambda_i f_{t+h} + \xi_{i,t+h}$, $\phi(L) f_{t+h} = \zeta_{t+h}$, where $\xi_{i,t}$, ζ_t are mutually independent and white noise, and k_i is an integer that is taken from McCracken and Ng (2016). The third model adds lagged dependent variables to the approximate dynamic factor model. It is given by $\phi(L) f_{t+h} = \zeta_{t+h}$, $c_i(L)\Delta^{k_i} y_{i,t+h} = \Lambda_i f_{t+h} + \xi_{i,t+h}$, where $\xi_{i,t}$, $\zeta_{j,t}$ are again mutually independent and white noise. It is denote it as **PCAR**. Finally, the last benchmark is the so-called factor-augmented error-correction model (**FECM**) of Banerjee and Marcellino (2009), which separates the observable variables into two groups $y = (y^{(1)\prime}, y^{(2)\prime})'$ and shrinks the latter group via principal components to seven factors $\hat{f}$, where the number of factors was chosen by the $PC_{(p3)}$ criterion of Bai and Ng (2002).[4] A vector error-correction model is then estimated for $(y^{(1)\prime}, \hat{f}')'$. Details on the forecast properties are found in Banerjee et al. (2014). Since we only obtain predictions for $y^{(1)}$, the **FECM** results are only reported in tables 4.2 and 4.3.

Model specification is chosen based on the data set from January 1960 to December 1999. To draw inference on an appropriate specification of the three different fractional factor models semiparametric methods are used: As no information criterion on the number of factors in a fractionally integrated setup is available, I first estimate the fractional integration orders of all observable variables via the exact local Whittle estimator of Shimotsu (2010), where I account for an intercept and a linear time trend, and the bandwidth is set to 1/2.[5] Next, the data are fractionally differenced according to their estimated integration orders, and the number of factors is determined by the $PC_{(p3)}$ criterion of Bai and Ng (2002), which suggests to include seven common factors.[6]

For the dynamic orthogonal fractional components model of subsection 4.2.2, the numbers of long and short memory factors r_1 and r_2 remain to be determined. A possible grouping of factors with equal integration orders is carried out as follows: First, I estimate the factor loadings for seven factors based on the data set in fractional differences as before. Next, I project the data in levels onto the space spanned by the factor loadings, which yields estimates for the seven factors in levels. Using the method of Matteson and Tsay (2011), I rotate the factors such that they are dynami-

[4] The criteria of Bai and Ng (2002) were evaluated for the data from January 1960 to December 1999.

[5] Using a higher bandwidth increases the risk of over-differencing and thus makes it more likely to underestimate the number of factors. Therefore, a comparably small bandwidth is selected.

[6] The other criteria of Bai and Ng (2002) either also find seven common factors, or slightly fewer.

cally orthogonal, i.e. uncorrelated for all leads and lags. To determine whether some factors exhibit the same integration order, the methods derived by Robinson and Yajima (2002) with the modification to possibly non-stationary integration orders by Nielsen and Shimotsu (2007) are applied: After estimating the integration order of each factor by the exact local Whittle estimator, they allow to sequentially test for the existence of $j = 1, 2, ..., 7$ groups of identical integration orders within the seven factor estimates. The sequential test terminates if for some j^* the null hypothesis of within-group equality of the integration order is not rejected. To jointly test within-group equality of the integration orders for a given grouping (i.e. not only testing for equal integration orders within one group, but within all groups together), I use the Wald test as proposed by Nielsen and Shimotsu (2007). For the exact local Whittle estimator, I again set the bandwidth to 1/2. The procedure suggests four different groups of factors, where the first three all contain a single factor whose integration order significantly differs from zero. The fourth group consists of four factors whose integration order cannot be significantly distinguished from zero. Hence, the latter are treated as short-range dependent and are assumed to belong to $f_t^{(2)}$, which yields $r_1 = 3$ and $r_2 = 4$.

For the factor model of subsection 4.2.1, I determine the AR lag order by the Bayesian information criterion (BIC): As before, factor estimates are obtained by estimating the factor loadings based on the fractionally differenced data and projecting the data in levels on the space spanned by the factor loadings. Next, they are rotated to become dynamically orthogonal by the method of Matteson and Tsay (2011). The lag order of the AR polynomials is then determined by estimating a diagonal VAR for the seven factors and choosing the lag order that minimizes the BIC. The procedure for the model in subsection 4.2.2 is identical, except that only the four short memory factors are used. For the model in subsection 4.2.3, the same procedure applies, except that the factors are directly estimated based on the data in fractional differences. For the dynamic fractional factor model in subsection 4.2.1 and the dynamic factor model in fractional differences in subsection 4.2.3, the BIC suggests a single lag for the AR polynomial of the factors, while for the dynamic orthogonal fractional components model in subsection 4.2.2, two lags minimize the BIC. Finally, for the models in subsections 4.2.1 and 4.2.2, I allow for a single autoregressive lag in the lag polynomial of the idiosyncratic term u_t. This, on the one hand, allows for autocorrelation in the idiosyncratic component, but keeps the dimension of the parameter space somewhat manageable on the other.

Besides the factors and autoregressive idiosyncratic components, the fractional

factor models as well as the benchmarks include deterministic components: All models allow for an intercept and a linear time trend for each observable variable. Furthermore, an observable variable is log-transformed whenever suggested by McCracken and Ng (2016).

A few caveats in terms of model specification are of order: First, note that the number of factors is determined based on the model selection criteria of Bai and Ng (2002) for the data set in fractional differences. If now an observable variable is driven by factors of different memory, then its integration order equals the highest integration order of all factors that load on the variable. Taking fractional differences thus over-differences those factors with a comparably low memory, which makes it difficult to identify them via the criteria of Bai and Ng (2002). Therefore, it is very likely that the number of overall factors is underestimated. As an alternative to the criteria of Bai and Ng (2002), one could also estimate the different factor models for several numbers of factors and use a likelihood-based information criterion like the BIC to determine the number of factors. Second, while the factors are grouped into different groups of equal memory, the observable variables are not. The number of non-zero factor loadings could be reduced by also grouping the observable variables into blocks of equal memory, and imposing a block-triangular structure on the factor loadings. This would restrict the more persistent factors to only load on those observable variables with high memory, and would reduce the number of loadings to be estimated. Third, allowing for different lag lengths among the factors with AR dynamics is likely to further increase the forecast performance. And last, allowing for only a single lag in the polynomials of the idiosyncratic terms is very restrictive. Addressing these caveats may further improve the forecast performance of the fractional factor model and is necessary whenever one aims for a structural analysis of the data. However, as will become clear in the next subsection, the fractional factor models are able to significantly improve the forecast accuracy compared to the four benchmarks, although there is room for improvement in terms of model specification.

4.4.2 Forecast results

For a given forecast horizon $h = 1, ..., 12$, table 4.1 shows how often each specification leads to the smallest MSPE for all 112 variables. Hence, it illustrates how frequently fractional factor models are able to outperform the benchmarks, i.e. autoregressive models and principal components of integer differences. To draw inference on the extent of forecast improvement, tables 4.2 and 4.3 report the relative MSPE (in relation to the AR benchmark) for twelve selected variables and for $h = 1, 2, 3, 6, 9,$ and 12.

Consequently, they also show how large the forecast accuracy fluctuates for each specification and highlight the robustness of the forecast results when a model is not chosen to be the best one.

	Benchmarks			DFFM		DOFC		DFFD	
Horizon	AR	PC	PCAR	PC	KF	PC	KF	PC	KF
1	14	10	16	7	0	3	26	20	16
2	14	12	11	8	0	1	26	24	16
3	14	10	6	6	1	5	28	23	19
4	17	9	7	11	2	4	25	20	17
5	15	11	7	10	1	5	27	22	14
6	17	8	5	10	6	6	23	19	18
7	15	9	3	10	4	7	26	18	20
8	14	8	3	10	12	7	20	19	19
9	15	8	3	10	10	7	22	17	20
10	15	8	3	11	14	7	16	15	23
11	15	8	3	13	15	7	13	15	23
12	14	7	3	10	17	9	13	16	23

Table 4.1: Frequency of smallest MSPE: The table shows how often, for a given forecast horizon h, a specification led to the smallest mean squared prediction error of all models.

As can be seen from table 4.1, fractional factor models tend to outperform autoregressive models, pre-differenced principal components models and mixtures of these two model classes. Over all 1344 forecasts, the benchmarks only exhibit a smaller MSPE than the fractional factor models in 357 cases (26.6%). Hence, for the remaining 987 forecasts (73.4 %) the smallest MSPE is achieved by one of the six fractional factor models. Among the fractional factor models, the dynamic orthogonal fractional components model in state space form produces the best predictions in terms of the MSPE for forecast horizons up to 9 months most frequently.

The DFFD models complement the predictive power of fractional factor models. They frequently yield the smallest MSPE whenever the DOFC-KF specification is not the best predictor in terms of the MSPE. Furthermore, principal components are found to yield a small MSPE at least for smaller forecast horizons when the data is in fractional differences, however they are frequently beaten by the state space formulation of the DOFC model. For higher forecast horizons, the forecast performance of the DFFD-KF model improves, leading to the highest number of best predictions in terms of the MSPE for $h = 10, 11, 12$.

The DFFM specification performs comparably poor in terms of the MSPE: While principal components are beaten by the DFFD model in terms of the frequencies of smallest MSPEs, the Kalman filter-based estimates appear particularly weak for

small h, where they only yield the smallest MSPE for a handful of forecasts. For higher horizons the DFFM-KF improves, however it is again outperformed by the DFFD-KF in terms of the frequency of smallest MSPEs.

More details about the forecast performance of fractional factor models can be identified by having a closer look at tables 4.2 and 4.3 that visualize the relative MSPEs for selected forecast horizons h and selected variables. The latter were chosen because they represent the full breadth of the macroeconomic data set as well as the full spectrum of integration orders (according to the exact local Whittle estimator with a constant and a linear time trend): With an estimated integration order smaller unity, average weekly overtime hours in the manufacturing businesses, the federal funds rate, and the US / UK foreign exchange rate have comparatively low memory, while the consumer price index, personal consumption index, and average hourly earnings have comparably high memory with an estimated integration order greater 3/2. The remaining variables exhibit an estimated integration order somewhere between the high and low cases.

As can be seen from tables 4.2 and 4.3, relative gains in forecast performance from the fractional factor models can be substantial: In many cases, fractional factor models can reduce the MSPE by more than 25% relative to the AR benchmark. For some variables, the MSPE is cut by half when fractional factor models are used, and reductions of more than 80% are possible. Within the class of fractional factor models, the DOFC-KF specification often results in the smallest MSPE, and at the same time does not show any large outliers in terms of a very high MSPE: For $h = 1, 2, 3$, the most accurate predictions for the consumer price index, personal consumption index and average hourly earnings stem from the DOFC-KF specification, which reduces the MSPE relative to the AR benchmark by more than 50%. In addition, the DOFC-KF specification exhibits the smallest MSPE for the St. Louis adjusted monetary base, for total reserves of depository institutions, and for the S&P500 frequently. The stable forecast performance of the DOFC-KF model is illustrated by the fact that its largest relative MSPE is 1.29, whereas its smallest relative MSPE is 0.17.

Another model that frequently produces a comparatively small MSPE is the DFFD model. For the industrial production index, the DFFD-PC specification exhibits the smallest MSPE for any forecast horizon. In addition, the DFFD-KF specification produces accurate predictions for the S&P500, average hourly earnings and the US / UK foreign exchange rate. Furthermore, its forecast performance is almost as stable as the DOFC-KF.

	Benchmarks				DFFM		DOFC		DFFD	
	AR	PC	PCAR	FECM	PC	KF	PC	KF	PC	KF
Horizon h = 1										
INDPRO	1.00	0.92	0.92	1.22	1.60	3.26	1.96	1.06	0.90	0.99
UNRATE	1.00	0.96	0.95	0.88	1.06	2.67	1.32	0.91	0.89	0.88
AWOTMAN	1.00	0.85	0.90	0.92	1.25	2.02	1.40	0.91	0.93	0.96
HOUST	1.00	0.71	0.70	0.99	0.98	1.36	0.94	0.87	1.09	1.07
AMBSL	1.00	0.86	1.17	0.69	0.77	2.62	0.77	0.81	2.78	3.15
TOTRESNS	1.00	0.71	1.45	0.69	0.66	2.28	0.70	0.69	4.84	5.14
S.P.500	1.00	1.08	1.08	1.01	1.04	2.67	1.24	1.02	1.06	0.99
FEDFUNDS	1.00	2.51	2.36	2.64	1.15	4.43	1.17	1.21	3.53	1.25
EXUSUKx	1.00	1.18	1.08	1.08	1.11	2.75	1.17	1.06	1.12	1.10
CPIAUCSL	1.00	0.64	0.96	0.45	1.77	2.01	1.58	0.41	0.49	0.49
PCEPI	1.00	0.69	0.98	0.50	3.32	1.98	2.70	0.41	0.47	0.47
CES0600000008	1.00	0.88	1.14	0.48	2.50	1.00	3.06	0.30	0.48	0.41
Horizon h = 2										
INDPRO	1.00	0.90	0.90	1.35	2.10	2.38	2.63	1.15	0.81	0.99
UNRATE	1.00	1.03	0.98	0.79	1.06	2.10	1.62	0.96	0.83	0.86
AWOTMAN	1.00	0.81	0.92	0.80	1.22	1.69	1.50	0.86	0.96	1.02
HOUST	1.00	0.71	0.71	1.01	0.99	1.18	0.95	0.99	1.02	1.00
AMBSL	1.00	0.79	1.27	0.86	0.76	1.25	0.80	0.72	1.52	1.65
TOTRESNS	1.00	0.69	1.56	0.75	0.67	1.17	0.75	0.65	2.45	2.51
S.P.500	1.00	1.14	1.16	1.18	1.05	1.55	1.30	1.02	1.11	0.98
FEDFUNDS	1.00	1.66	1.79	2.66	0.91	2.08	0.92	0.95	2.40	1.07
EXUSUKx	1.00	1.20	1.13	1.15	1.18	1.71	1.22	1.05	1.11	1.08
CPIAUCSL	1.00	0.61	0.98	0.54	1.76	0.50	1.62	0.42	0.60	0.55
PCEPI	1.00	0.62	0.99	0.56	3.24	0.45	2.66	0.39	0.52	0.48
CES0600000008	1.00	0.99	1.16	0.37	2.16	0.64	3.26	0.21	0.34	0.28
Horizon h = 3										
INDPRO	1.00	1.04	1.04	1.50	2.41	2.47	2.93	1.28	0.81	1.00
UNRATE	1.00	1.15	1.08	0.82	1.11	2.21	1.76	1.04	0.82	0.88
AWOTMAN	1.00	0.79	0.95	0.76	1.10	1.56	1.47	0.84	1.03	1.02
HOUST	1.00	0.70	0.70	1.01	0.88	1.11	0.85	0.95	1.09	0.99
AMBSL	1.00	0.77	1.41	0.97	0.74	0.95	0.81	0.66	1.16	1.22
TOTRESNS	1.00	0.73	1.62	0.81	0.69	0.94	0.78	0.64	1.81	1.81
S.P.500	1.00	1.25	1.27	1.30	1.08	1.45	1.35	1.02	1.18	1.00
FEDFUNDS	1.00	1.31	1.41	2.66	0.85	1.54	0.82	0.91	1.92	1.03
EXUSUKx	1.00	1.26	1.18	1.21	1.22	1.55	1.29	1.07	1.12	1.08
CPIAUCSL	1.00	0.57	1.01	0.54	1.71	0.47	1.63	0.38	0.63	0.56
PCEPI	1.00	0.60	1.01	0.58	3.12	0.43	2.72	0.36	0.53	0.47
CES0600000008	1.00	1.13	1.18	0.40	1.79	0.56	2.95	0.18	0.27	0.22

Table 4.2: Selected relative mean squared prediction errors for h=1, 2, and 3. Variable codes are INDPRO: industrial production index; UNRATE: unemployment rate; AWOTMAN: average weekly overtime hours in the manufacturing business; HOUST: housing starts; AMBSL: St. Louis adjusted monetary base; TOTRESNS: total reserves of depository institutions; S.P.500: S&P500 index; FEDFUNDS: effective federal funds rate; EXUSUKx: US / UK foreign exchange rate; CPIAUCSL: consumer price index; PCEPI: personal consumption index; CES0600000008: average hourly earnings

	Benchmarks				DFFM		DOFC		DFFD	
	AR	PC	PCAR	FECM	PC	KF	PC	KF	PC	KF
Horizon h = 6										
INDPRO	1.00	1.27	1.27	1.59	2.10	1.82	2.55	1.29	0.95	1.08
UNRATE	1.00	1.58	1.46	1.11	1.25	2.04	1.86	1.25	1.05	1.06
AWOTMAN	1.00	0.92	1.12	0.76	1.03	1.33	1.40	0.87	1.17	1.13
HOUST	1.00	0.63	0.63	1.02	0.77	0.88	0.78	0.82	0.98	0.87
AMBSL	1.00	0.89	1.87	0.87	0.62	0.68	0.76	0.55	0.89	0.90
TOTRESNS	1.00	0.95	1.62	0.75	0.63	0.68	0.76	0.57	1.18	1.16
S.P.500	1.00	1.50	1.53	1.40	1.09	1.21	1.47	1.00	1.19	1.01
FEDFUNDS	1.00	1.28	1.33	2.70	0.90	1.21	0.76	0.91	1.36	1.06
EXUSUKx	1.00	1.31	1.30	1.46	1.23	1.18	1.35	1.03	1.06	1.00
CPIAUCSL	1.00	0.57	1.09	0.46	1.49	0.26	1.65	0.31	0.63	0.54
PCEPI	1.00	0.59	1.06	0.54	2.69	0.25	2.85	0.34	0.52	0.44
CES0600000008	1.00	1.82	1.27	0.54	0.86	0.50	2.38	0.17	0.21	0.16
Horizon h = 9										
INDPRO	1.00	1.43	1.43	1.85	1.83	1.48	2.30	1.25	1.00	1.12
UNRATE	1.00	1.88	1.77	1.48	1.23	1.76	1.65	1.26	1.17	1.12
AWOTMAN	1.00	1.11	1.33	0.74	1.01	1.20	1.34	0.91	1.21	1.14
HOUST	1.00	0.60	0.61	1.02	0.72	0.78	0.73	0.76	0.94	0.85
AMBSL	1.00	1.31	3.30	0.91	0.59	0.59	0.79	0.51	0.81	0.80
TOTRESNS	1.00	1.54	1.74	0.73	0.61	0.60	0.76	0.53	0.97	0.94
S.P.500	1.00	1.76	1.79	1.50	1.09	1.13	1.54	0.99	1.19	1.01
FEDFUNDS	1.00	1.41	1.50	2.55	0.93	1.16	0.79	0.96	1.18	1.07
EXUSUKx	1.00	1.40	1.41	1.78	1.29	1.08	1.42	1.01	1.05	0.98
CPIAUCSL	1.00	0.64	1.13	0.43	1.32	0.21	1.56	0.31	0.60	0.51
PCEPI	1.00	0.65	1.10	0.53	2.40	0.21	2.70	0.37	0.50	0.42
CES0600000008	1.00	2.65	1.36	0.60	0.49	0.46	1.81	0.19	0.16	0.12
Horizon h = 12										
INDPRO	1.00	1.57	1.58	2.00	1.70	1.28	2.20	1.20	0.99	1.11
UNRATE	1.00	2.20	2.09	1.68	1.16	1.50	1.45	1.22	1.18	1.12
AWOTMAN	1.00	1.26	1.50	0.78	0.99	1.12	1.27	0.94	1.17	1.10
HOUST	1.00	0.61	0.61	1.11	0.72	0.74	0.71	0.74	0.94	0.86
AMBSL	1.00	1.84	6.01	0.77	0.52	0.52	0.75	0.45	0.70	0.69
TOTRESNS	1.00	2.36	1.93	0.63	0.55	0.53	0.70	0.48	0.80	0.77
S.P.500	1.00	2.00	2.03	1.57	1.07	1.09	1.59	0.98	1.20	1.00
FEDFUNDS	1.00	1.54	1.65	2.39	0.95	1.11	0.83	1.00	1.10	1.06
EXUSUKx	1.00	1.53	1.58	2.01	1.36	1.05	1.48	0.99	1.05	0.98
CPIAUCSL	1.00	0.79	1.18	0.40	1.14	0.15	1.43	0.34	0.54	0.46
PCEPI	1.00	0.76	1.14	0.51	2.08	0.16	2.47	0.42	0.46	0.39
CES0600000008	1.00	4.01	1.47	0.65	0.36	0.45	1.52	0.23	0.15	0.11

Table 4.3: Selected relative mean squared prediction errors for h=6, 9, and 12. Variable codes are INDPRO: industrial production index; UNRATE: unemployment rate; AWOTMAN: average weekly overtime hours in the manufacturing business; HOUST: housing starts; AMBSL: St. Louis adjusted monetary base; TOTRESNS: total reserves of depository institutions; S.P.500: S&P500 index; FEDFUNDS: effective federal funds rate; EXUSUKx: US / UK foreign exchange rate; CPIAUCSL: consumer price index; PCEPI: personal consumption index; CES0600000008: average hourly earnings

Finally, predictions from the DFFM model, which serves as the most general framework (as it nests the two other fractional factor models), exhibit large fluctuations in terms of the MSPE. Nonetheless, for higher forecast horizons, the DFFM-KF model produces accurate forecasts for the consumer price and personal consumption index. Interestingly, for the variables with highest memory, i.e. the consumer price index, the personal consumption index, and the average hourly earnings index, the DFFM-KF model and the DOFC-KF model exhibit the smallest MSPE for almost any horizon. A possible explanation is that these are the only models where the fractional factors enter in levels.

Note that the difference between the benchmark PC model and the DFFD-PC specification is the pre-differencing. The two models coincide in terms of their performance relative to the AR benchmark. The advantages over the AR model are therefore likely to result from cross-sectional dependencies that are detected by the common factors. In addition, the smaller MSPEs of the DFFD-PC model can be explained by the sensitivity of standard PC methods to spurious coefficients whenever there is autocorrelation left in the data (e.g. due to over- or under-differencing), as Franses and Janssens (2019) argue.

Note further that the forecast performance of the DOFC-KF model is similar to the DFFD model for many variables and horizons. Both models have in common that they allow the data to be fractionally integrated, however the former explicitly models fractional cointegration relations by common fractional factors, whereas the latter eliminates the memory by pre-differencing the data. However, there are cases where gains from the DOFC-KF specification relative to the DFFD model can be substantial, especially in situations where the latter produces a relative MSPE > 1. Consider e.g. the forecasts for the adjusted monetary base (AMBSL) and the total reserves of depository institutions (TOTRESNS) in tables 4.2 and 4.3, where the DOFC-KF and the FECM model perform well, wheres the DFFD-KF model yields large MSPEs. While the former two models take cointegration into account, the DFFD-KF model eliminates long-run components by prior differencing and is likely to produce over-differenced short-run components. Hence, the DOFC-KF model may exhibit an advantage over the DFFD-KF model whenever strong cointegration relations among the variables are apparent, and whenever additive short-run factors are present that are over-differenced by the DFFD-KF model.

Finally, to examine the forecast behavior when the economy is hit by a large shock, I take a closer look at the performance of the fractional factor models during the Great Recession. Figure 4.1 sketches the three-step ahead predictions for the

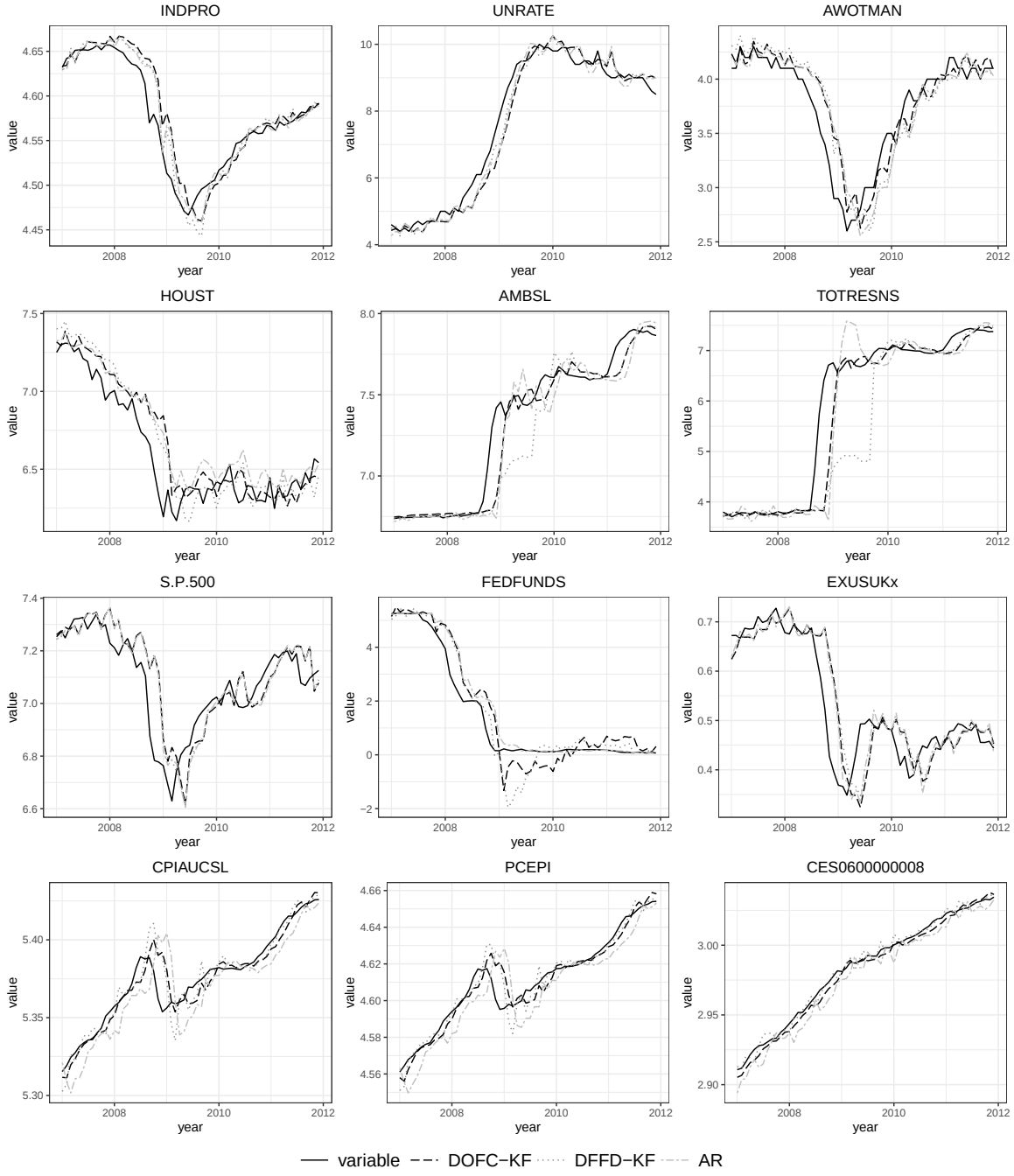

Figure 4.1: Forecast performance of the DOFC-KF, DFFD-KF and AR model during the Great Recession for $h = 3$. Variable codes are INDPRO: industrial production index; UNRATE: unemployment rate; AWOTMAN: average weekly overtime hours in the manufacturing business; HOUST: housing starts; AMBSL: St. Louis adjusted monetary base; TOTRESNS: total reserves of depository institutions; S.P.500: S&P500 index; FEDFUNDS: effective federal funds rate; EXUSUKx: US / UK foreign exchange rate; CPIAUCSL: consumer price index; PCEPI: personal consumption index; CES0600000008: average hourly earnings

twelve selected variables and the two best performing fractional factor models together with the AR benchmark from January 2007 to December 2011. As the graphs show, the forecast performance of the fractional factor models is not systematically disrupted by the Great Recession in comparison to the AR benchmark. Instead, the forecasts converge towards the realizations of the observable variables rapidly after the crisis. The DOFC-KF forecasts seem to be the least affected by the large shock, as they converge faster towards the actual realizations of the predicted variables than the other forecasts. Furthermore, the AR and DFFD-KF predictions for the adjusted monetary base and total reserves of depository institutions seem to be biased by the crisis until the end of 2009, which substantiates the relative robustness of the DOFC-KF specification.

4.5 Conclusion

This paper considered three different fractional factor models for macroeconomic forecasting. In a pseudo out-of-sample forecast experiment for the high-dimensional data set of McCracken and Ng (2016), it was shown that the fractional factor models are able to improve the forecast accuracy substantially, even when there is room for improvement in terms of model specification (see the discussion at the end of section 4.4.1). Especially the dynamic orthogonal fractional components model of Hartl and Jucknewitz (2021) and a dynamic factor model for the data in fractional differences are promising, as they showed significant forecast improvements in comparison to standard approximate dynamic factor models for a variety of macroeconomic variables.

Building on the results, future research could improve the model specification by carrying out a structural analysis of the FRED-MD data set: Determining the number of factors in a fractionally integrated panel is challenging, and may be addressed by means of likelihood-based information criteria, by (fractional) cointegration tests, or by a new information criterion that is robust to fractional integration in the data. Moreover, future research could examine whether a combination of the DOFC model in state space form and a factor model in fractional differences can further improve the predictive power of fractional factor models. In addition, one could combine principal components and the Kalman filter analogously to Bräuning and Koopman (2014): By partitioning the data set into two groups, where the former contains the variables to be predicted, while the latter is shrunk via principal components, one would greatly reduce the number of parameters and speed up the estimation. Moreover, fractional factor models could be used to explore common trends and cycles

in macroeconomic data, thus generalizing the work of Barigozzi and Luciani (2021) to fractionally integrated processes. Finally, the forecast performance of fractional factor models should also be examined for data other than the FRED-MD.

Chapter 5

Conclusion

The present thesis contributes to the methodological literature on unobserved components and factor models by generalizing either a single or multiple common stochastic trends to account for long memory. The practical benefits of fractional unobserved components and factor models are demonstrated in empirical applications to climate and macroeconomic data.

In chapter 2, long memory is incorporated into unobserved components models by modeling the stochastic trend component as a fractionally integrated process. To deal with the computational burden imposed by the high-dimensional state vector associated with the state space representation of fractional unobserved components models, an analytical solution to the optimization problem of the Kalman filter is derived. Furthermore, for a prototypical fractional unobserved components model, the asymptotic estimation theory for the conditional sum-of-squares estimator is derived under relatively mild assumptions as compared to the unobserved components literature. The results are then shown to carry over to more complex models with deterministic components and correlated long- and short-run innovations, as well as to the quasi-maximum likelihood estimator. For US carbon emissions, the fractional unobserved components model provides new insights on the memory of trend emissions, on the relationship between cyclical carbon emissions and the business cycle, and on the interaction between long- and short-run innovations.

Building on the methodological results for univariate fractional unobserved components models in chapter 2, future research could consider multivariate fractional unobserved components models, where multiple observable variables are driven by common, fractionally integrated components. The results in chapter 2 may provide a starting point for assessing the asymptotic theory, however additional difficulties can be expected when cointegration is allowed for. For empirical researchers, the fractional unobserved components model offers a flexible, data-driven solution to the specification of the trend component in unobserved components models.

Chapter 3 applies the fractional unobserved components model to revisit the puzzling estimates for the business cycle obtained from traditional, integer-integrated unobserved components models. It provides evidence that the puzzling results in the literature are an artifact generated by the presence of a smooth fractionally integrated trend in log US real GDP with an integration order greater than one but less than two. The long-run component of log GDP is found to be well captured by a fractionally integrated trend with an integration order of 1.30, suggesting that integer-integrated models for log GDP are misspecified. The resulting trend-cycle decomposition of log GDP yields a trend estimate that is smooth, along with a cyclical component that is consistent with the NBER chronology.

While the estimates for trend and cycle are very different from those of traditional unobserved components models, the estimated correlation between long- and short-run innovations is (almost) -1, which is also often found for integer-integrated correlated unobserved components models. Consequently, long- and short-run innovations cannot be structurally identified by the fractional model. This calls for further investigation, e.g. by a bivariate fractional unobserved components model that adds an additional variable with a more pronounced cyclical behavior to the setup. In addition, models that allow for a break in the covariance matrix of long- and short-run innovations to account for the Great Moderation could be considered.

In chapter 4, three different parametric factor models are considered that allow the factors to exhibit long memory. The forecast performance of the different fractional factor models is studied and compared to a variety of benchmarks in a pseudo out-of-sample forecast experiment using the macroeconomic data set of McCracken and Ng (2016). Among the three fractional factor models, it is found that the fractional components model of Hartl and Jucknewitz (2021) can significantly improve the forecast accuracy relative to traditional approximate dynamic factor models, with reductions of the mean squared prediction error of more than 50% possible. Moreover, instead of taking first or second differences of the observable data, factor models in which the observable data is fractionally pre-differenced are also shown to often yield a smaller mean squared prediction error.

However, there is still room for improvement in terms of model specification: Determining the number of factors in a fractionally integrated panel is challenging, and is done based on the criteria of Bai and Ng (2002) for the data set in fractional differences. A more structural analysis of the data set of McCracken and Ng (2016) may identify additional factors that can further improve the forecast performance. Moreover, the number of factor loadings may be reduced by shrinking small loadings

to zero. Finally, the data set of McCracken and Ng (2016) is one out of many high-dimensional panels with strong dependencies in the cross sections, and examining the forecast performance of fractional factor models for different data sets may provide additional insights.

Bibliography

Bai, J. (2004). Estimating cross-section common stochastic trends in nonstationary panel data, *Journal of Econometrics* **122**(1): 137–183.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.

Bai, J. and Ng, S. (2004). A panic attack on unit roots and cointegration, *Econometrica* **72**(4): 1127–1177.

Bai, J. and Ng, S. (2008). Large dimensional factor analysis, *Foundations and Trends in Econometrics* **3**(2): 89–163.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics, *Journal of Econometrics* **73**(1): 5–59.

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.

Banerjee, A. and Marcellino, M. (2009). Factor-augmented error correction models, *in* J. Castle and N. Shephard (eds), *The Methodology and Practice of Econometrics: A Festschrift for David Hendry*, Oxford University Press, Oxford, pp. 227–254.

Banerjee, A., Marcellino, M. and Masten, I. (2014). Forecasting with factor-augmented error correction models, *International Journal of Forecasting* **30**(3): 589–612.

Banerjee, A., Marcellino, M. and Masten, I. (2016). An overview of the factor-augmented error-correction model, *in* E. Hillebrand and S. J. Koopman (eds), *Dynamic Factor Models*, Emerald Group Publishing Limited, Bingley, pp. 3–41.

Barigozzi, M., Lippi, M. and Luciani, M. (2021). Large-dimensional dynamic factor models: Estimation of impulse-response functions with I(1) cointegrated factors, *Journal of Econometrics* **221**(2): 455–482.

Barigozzi, M. and Luciani, M. (2021). Measuring the output gap using large datasets, *The Review of Economics and Statistics* (forthcoming).

Basistha, A. and Nelson, C. R. (2007). New measures of the output gap based on the forward-looking new Keynesian Phillips curve, *Journal of Monetary Economics* **54**(2): 498–511.

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.

Borio, C., Disyatat, P. and Juselius, M. (2017). Rethinking potential output: Embedding information about the financial cycle, *Oxford Economic Papers* **69**(3): 655–677.

Bräuning, F. and Koopman, S. J. (2014). Forecasting macroeconomic variables using collapsed dynamic factor analysis, *International Journal of Forecasting* **30**(3): 572–584.

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.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* **51**(5): 1281–1304.

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

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

Chang, Y., 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.

Cheung, Y. L. (2022). Long memory factor model: On estimation of factor memories, *Journal of Business & Economic Statistics* **40**(2): 756–769.

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.

Doz, C., Giannone, D. and Reichlin, L. (2012). A quasi maximum likelihood approach for large approximate dynamic factor models, *The Review of Economics and Statistics* **94**(4): 1014–1024.

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.

Eickmeier, S. (2009). Comovements and heterogeneity in the euro area analyzed in a non-stationary dynamic factor model, *Journal of Applied Econometrics* **24**(6): 933–959.

Ergemen, Y. E. (2019). System estimation of panel data models under long-range dependence, *Journal of Business & Economic Statistics* **37**(1): 13–26.

Ergemen, Y. E. (2023). Parametric estimation of long memory in factor models, *Journal of Econometrics* (forthcoming).

Ergemen, Y. E. and Velasco, C. (2017). Estimation of fractionally integrated panels with fixed effects and cross-section dependence, *Journal of Econometrics* **196**(2): 248–258.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation, *The Review of Economics and Statistics* **82**(4): 540–554.

Franses, P. H. and Janssens, E. (2019). Spurious principal components, *Applied Economics Letters* **26**(1): 37–39.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis* **4**(4): 221–238.

Gil-Alaña, L. A. and Robinson, P. M. (1997). Testing of unit root and other non-stationary hypotheses in macroeconomic time series, *Journal of Econometrics* **80**(2): 241–268.

Granger, C. W. J. (1986). Developments in the study of cointegrated economic variables, *Oxford Bulletin of Economics and Statistics* **48**(3): 213–228.

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 & 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. (2021). Multivariate fractional components analysis, *Journal of Financial Econometrics* (forthcoming).

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

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

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

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

Harvey, A. C. and Jäger, A. (1993). Detrending, stylized facts and the business cycle, *Journal of Applied Econometrics* **8**(3): 231–247.

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

Harvey, A. C. and Trimbur, T. M. (2003). General model-based filters for extracting cycles and trends in economic time series, *The Review of Economics and Statistics* **85**(2): 244–255.

Harvey, A. C., Trimbur, T. M. and Dijk, H. K. V. (2007). Trends and cycles in economic time series: A Bayesian approach, *Journal of Econometrics* **140**(2): 618–649.

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

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: International evidence, *Journal of Business & Economic Statistics* **13**(1): 37–45.

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.

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

Iwata, S. and Li, H. (2015). What are the differences in trend cycle decompositions by Beveridge and Nelson and by unobserved components models?, *Econometric Reviews* **34**(1–2): 146–173.

Johansen, S. (2008). A representation theory for a class of vector autoregressive models for fractional processes, *Econometric Theory* **24**(3): 651–676.

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

Jungbacker, B. and Koopman, S. J. (2015). Likelihood-based dynamic factor analysis for measurement and forecasting, *Econometrics Journal* **18**: C1–C21.

Kamber, G., Morley, J. C. and Wong, B. (2018). Intuitive and reliable estimates of the output gap from a Beveridge-Nelson filter, *The Review of Economics and Statistics* **100**(3): 550–566.

Luciani, M. and Veredas, D. (2015). Estimating and forecasting large panels of volatilities with approximate dynamic factor models, *Journal of Forecasting* **34**: 163–176.

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

Matteson, D. S. and Tsay, R. S. (2011). Dynamic orthogonal components for multivariate time series, *Journal of the American Statistical Association* **106**(496): 1450–1463.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research, *Journal of Business & Economic Statistics* **34**(4): 574–589.

Mesters, G., Koopman, S. J. and Ooms, M. (2016). Monte Carlo maximum likelihood estimation for generalized long-memory time series models, *Econometric Reviews* **35**(4): 659–687.

Morana, C. (2004). Frequency domain principal components estimation of fractionally cointegrated processes, *Applied Economics Letters* **11**(13): 837–842.

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.

Morley, J. C. and Piger, J. (2012). The asymmetric business cycle, *The Review of Economics and Statistics* **94**(1): 208–221.

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

Nielsen, M. Ø. (2004). Efficient inference in multivariate fractionally integrated time series models, *Econometrics Journal* **7**: 63–97.

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.

Nielsen, M. Ø. and Shimotsu, K. (2007). Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach, *Journal of Econometrics* **141**(2): 574–596.

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

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

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

Proietti, T. (2004). Unobserved components models with correlated disturbances, *Statistical Methods and Applications* **12**(3): 277–292.

Proietti, T. (2006). Trend–cycle decompositions with correlated components, *Econometric Reviews* **25**(1): 61–84.

Quah, D. and Sargent, T. J. (1993). A dynamic index model for large cross sections, *in* J. H. Stock and M. W. Watson (eds), *Business Cycles, Indicators and Forecasting*, University of Chicago Press, Chicago.

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.

Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems, *Journal of Econometrics* **106**(2): 217–241.

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

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

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* **20**(2): 147–162.

Trenkler, C. and Weber, E. (2016). On the identification of multivariate correlated unobserved components models, *Economics Letters* **138**: 15–18.

Tschernig, R., Weber, E. and Weigand, R. (2013). Long-run identification in a fractionally integrated system, *Journal of Business & Economic Statistics* **31**(4): 438–450.

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.

Wagner, M. (2008). The carbon Kuznets curve: A cloudy picture emitted by bad econometrics?, *Resource and Energy Economics* **30**(3): 388–408.

Weber, E. (2011). Analyzing U.S. output and the Great Moderation by simultaneous unobserved components, *Journal of Money, Credit and Banking* **43**(8): 1579–1597.

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.

Zhang, R., Robinson, P. M. and Yao, Q. (2019). Identifying cointegration by eigenanalysis, *Journal of the American Statistical Association* **114**(526): 916–927.

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