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

Tobias Hartl*[1,2], Rolf Tschernig[1], and Enzo Weber[1,2]

[1]University of Regensburg, 93053 Regensburg, Germany

[2]Institute for Employment Research (IAB), 90478 Nuremberg, Germany

December 2020

Abstract. Measures for the business cycle obtained from trend-cycle decompositions are puzzling, as they often are noisy, at odds with the NBER chronology, and not well in line with economic theory. We argue that these results are driven by the neglect of fractionally integrated trends in log US real GDP. To account for fractional integration we develop a generalization of trend-cycle decompositions that avoids prior assumptions about the long-run dynamic characteristics and treats the integration order as a random variable. The integration order is jointly estimated with the other model parameters via a quasi maximum likelihood estimator that is shown to be consistent and asymptotically normal. In addition, single-step estimators for the latent components that are identical to the Kalman filter and smoother but computationally superior are derived. We find that log US real GDP is integrated of order around 1.3, the resulting trend-cycle decomposition is in line with the NBER chronology, and the model well explains the puzzling results in the literature that result from model misspecification.

Keywords. output gap, trend-cycle decomposition, unobserved components, fractional lag operator, long memory, Kalman filter.

JEL-Classification. C22, C51, E32

*Corresponding author. E-Mail: tobias1.hartl@ur.de

The authors thank Lutz Arnold, Uwe Hassler, James Morley, Morten Ø. Nielsen, the participants of the econometric seminar in Nuremberg, the department seminar at the Christian Albrechts University Kiel, the DAGStat conference 2019 in Munich, the workshop on high-dimensional time series in economics and finance 2019 in Vienna, the department seminar at the University of Regensburg, the PhD seminar at the University of Regensburg, and the International Conference on Computational and Methodological Statistics 2020. Support through the project 356439312 financed by the German Research Foundation (DFG) is gratefully acknowledged. A previous version of this paper was circulated under the title "Fractional trends and cycles in macroeconomic time series"

1 Introduction

Measuring and extracting the business cycle plays a key role in applied research, as numerous macroeconomic models impose assumptions on the long- and short-run behavior of real output. To verify these assumptions, appropriate methods for the decomposition of time series into trend and cycle are of order and will be 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), results in the literature are puzzling. While empirical evidence supports a strong negative correlation of long- and short-run shocks, 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 together with a noisy cycle, thereby missing the NBER chronology and contradicting macroeconomic common sense.

Two conflicting solutions to the above puzzle have been proposed by Perron and Wada (2009) and Kamber et al. (2018). The former authors show that log US real GDP is well explained by a deterministic long-run component with a trend break after the first quarter of 1973, implying that the stochastic part of log US real GDP is $I(0)$. While the resulting decomposition is in line with the NBER chronology, it is at odds with the bulk of the cointegration literature that assumes stochastic long-run relationships among GDP and other macroeconomic series as the decomposition does not allow for long-run stochastic shocks. On the other hand, Kamber et al. (2018) obtain an economically plausible decomposition by restricting the variance-ratio of long- and short-run shocks to be small, thereby forcing the $I(1)$ long-run component to be smooth, which leaves additional dynamics to be captured by the cycle. This directly raises the question why the unrestricted decomposition of Morley et al. (2003), that nests the model of Kamber et al. (2018), yields different parameter estimates that are eliminated from the parameter space via the restriction on the variance-ratio but obviously correspond to a greater log likelihood.

In this paper, we will argue that all aforementioned features are artifacts generated by the presence of a smooth fractionally integrated long-run component in log US real GDP with an integration order greater than one. In that case the variance estimate of the long-run shocks of Morley et al. (2003) is upward-biased, as the additional persistence of the long-run component that is not captured by the $I(1)$ specification goes directly into the shocks. The upward-biased variance estimate yields an erratic long-run component together with a noisy cycle and a large variance-ratio of long- and short-run shocks. Conversely, if the variance-ratio is restricted to be small as in Kamber et al. (2018), the estimated long-run shocks become a fractionally integrated process to grasp the remaining persistence that is not captured by the $I(1)$ specification, thereby violating the white noise assumption. While the decomposition fits the NBER chronology, the violation of the white noise assumption yields inconsistent parameter estimates. Finally, as smooth fractional processes are well approximated by deterministic processes with structural breaks, this explains the findings of Perron and Wada (2009).

To examine the above hypothesis, we contribute the methodological literature by deriving a fractional UC model, where the long-run component is allowed to be fractionally integrated ($I(d)$), $d \in \mathbb{R}_+$, whereas the fractional lag operator $L_d = 1 - \Delta^d$ of Johansen (2008), that is defined in (4), enters the lag polynomial of the cyclical component. The model encompasses state-of-the-art integer-integrated UC models, requires only mild distributional assumptions on the long- and short-run shocks, and bears the advantage that no additional assumptions on d are required. It allows to jointly estimate the integration order together with the other model parameters via a quasi maximum likelihood estimator that is consistent and asymptotically normal. In addition, single-step estimators for trend and cycle that are identical to the Kalman filter and smoother are derived. While state-of-the-art non-parametric estimators for the integration order such as e.g. the exact local Whittle estimator Shimotsu and Phillips (2005) and the estimator of Geweke and Porter-Hudak (1983) are downward-biased when strong short-run variation is present (Sun and Phillips; 2004), our parametric setup allows to draw inference on the integration order by explicitly modelling the short-run variation via the cyclical component.

With the model at hand, we empirically examine the integration order of log US real GDP and the validity of the $I(1)$ assumption frequently made in the literature, the stochastic nature of the long-run component in comparison to the suggested deterministic form in Perron and Wada (2009), and the economic plausibility of the resulting trend-cycle decomposition. Our results substantiate the fractional hypothesis, as we find log US real GDP to be integrated of order around 1.3, the 95% confidence interval for the integration order excludes $d = 1$, and a glimpse at figure 3 verifies that the decomposition fits the NBER chronology well. Estimators for the integration order that are robust to structural breaks confirm the estimated integration order, while no evidence for structural breaks is found when fractional integration is allowed, implying that the long-run component of log US real GDP is in fact a fractionally integrated process rather than a spurious fractionally integrated process generated by structural breaks. From an economic perspective, our findings suggest to withdraw the predominant assumption that long-run shocks have only a contemporaneous effect on GDP growth and to rather interpret them as having a gradually decreasing impact over time.

The outline of the paper is as follows. Section 2 details the unobserved components puzzle and motivates the necessity of a fractional UC model to examine the long-run properties of log US real GDP. We demonstrates the contradicting results of Morley et al.

(2003), Perron and Wada (2009), and Kamber et al. (2018) are likely to result from model misspecification due to the presence of a neglected fractionally integrated long-run component. While we find comprehensive evidence for fractional integration in log US real GDP that is robust to the presence of structural breaks, no evidence for such breaks is found as soon as fractional integration is allowed for. Section 3 then derives the fractional UC model and relates it to the literature. Parameter estimation via a quasi maximum likelihood estimator, that is shown to be consistent and asymptotically normal, is considered in section 4. For the estimation of latent trend and cycle single-step formulas that are identical to Kalman filter and smoother are derived. Section 5 applies the model to decompose log US real GDP, while section 6 concludes. All proofs are contained in the appendix.

2 The unobserved components puzzle

In this section we give a brief overview of the puzzling results in the UC literature for log US real GDP and propose a potential solution to the puzzle. UC models assume that log real GDP y_t can be represented as the sum of a long-run component τ_t and a cyclical component c_t that are unobserved

$$y_t = \tau_t + c_t, \qquad \tau_t = \mu_0 + \mu_1 t + x_t, \qquad \Delta^d x_t = \eta_t, \qquad a(L)c_t = b(L)\varepsilon_t, \tag{1}$$

$t = 1, ..., n$, and the latent components are disentangled by imposing assumptions on their autocovariance functions. The long-run component τ_t is characterized by an autocovariance function that decays more slowly than with an exponential rate and captures the long-run dynamics of a time series, whereas the cycle component c_t will be $I(0)$ and accounts for transitory fluctuations of a series around its trend. While specifications with $d = 2$ exist (Clark; 1987; Oh and Zivot; 2006), the bulk of the literature for GDP assumes τ_t to follow a random walk with drift by setting $d = 1$, while c_t is a stationary and invertible ARMA process, and the underlying long- and short-run shocks η_t, ε_t, that generate τ_t and c_t, typically are Gaussian white noise and may be contemporaneously correlated.

As the model in (1) is not identified without further restrictions, the UC literature either restricts long- and short-run shocks to be uncorrelated (cf. e.g. Harvey; 1985), assumes $b(L) = 1$ as in the correlated UC model of Morley et al. (2003), or estimates τ_t, c_t based on a reduced form ARMA representation of Δy_t via the decomposition of Beveridge and Nelson (1981). As found by Morley et al. (2003), uncorrelatedness of short- and long-run shocks is likely to be violated for US GDP when $\tau_t \sim I(1)$, and their resulting decomposition is virtually identical to the Beveridge-Nelson decomposition.

However, as also summarized by Perron and Wada (2009, ch. 2), allowing for correlated shocks yields a decomposition that is at odds with economic common sense: Estimates from the correlated UC model imply a high variance for the long-run shocks together with a small variance for the short-run shocks, leading to an erratic long-run component that strongly fluctuates around a linear trend and leaving only little variation to the cycle, see Morley et al. (2003, fig. 3) and Perron and Wada (2009, fig. 1). Consequently, the estimated cycle behaves noisy and does not resemble the NBER chronology. In addition, the estimated components are not in line with multivariate UC models that find a rather smooth trend together with a cycle hitting the NBER recession periods, see for instance Harvey and Trimbur (2003), Basistha and Nelson (2007) and Harvey et al. (2007).

Two important solutions to these problems have recently been suggested. First, Perron and Wada (2009) argue that the correlated UC model misspecifies the long-run properties of log US real GDP, and τ_t should be modelled as a purely deterministic trend that allows for a change in slope after the first quarter of 1973. And second, Kamber et al. (2018) suggest to restrict the variance-ratio of long- and short-run shocks $\mathrm{Var}(\eta_t)/\mathrm{Var}(\varepsilon_t)$ to be small, thereby forcing the trend to become smooth. Both decompositions are well in line with the NBER chronology but have conflicting implications for the stochastic nature of GDP's long-run component.

If we assume the specification of Perron and Wada (2009) to be correct, then log US real GDP is $I(0)$ and fluctuates around a deterministic trend with a single break after 1973:1, which contradicts the bulk of the literature that finds GDP to be cointegrated with several macroeconomic aggregates. Instead, the model ascribes a deterministic nature to permanent shocks, while e.g. technological innovations are widely assumed to be of a stochastic nature and to exhibit a permanent impact. Conversely, if we believe the restricted UC model of Kamber et al. (2018), it remains an open question why the unrestricted model of Morley et al. (2003), that nests the restricted model, yields a different estimate and, thus, a greater log likelihood.

We will argue that the solution to this puzzle is that neither the model of Perron and Wada (2009) nor the model of Kamber et al. (2018) is correctly specified, as log US real GDP violates both the $I(0)$ assumption of the former and the $I(1)$ assumption of the latter. Our argument is based on the observation that a smooth fractionally integrated trend in log US real GDP with integration order greater than one, i.e. $\tau_t \sim I(d)$, $d \in \mathbb{R}_+$, $d > 1$, well explains the contradicting results in Morley et al. (2003), Perron and Wada (2009), and Kamber et al. (2018), and we provide evidence for this hypothesis in the following.

Given that log US real GDP is integrated of order greater than one, both the correlated UC model of Morley et al. (2003) and the corresponding Beveridge-Nelson decomposition will attribute the additional persistence that is not captured by the $I(1)$ specification to the long-run shocks η_t, so that their variance estimate will be upward-biased. This explains the erratic trend that is obtained in Morley et al. (2003) and leaves only small variation to the cyclical component. If the parameter space is then restricted to the region where the variance of the long-run shocks is small, as in Kamber et al. (2018), the underlying long-run shock estimates become fractionally integrated, as the additional persistence cannot be grasped by their variance parameter. This violates the white noise assumption, yields inconsistent estimates and explains why the model of Morley et al. (2003) that nests the restricted specification returns a different estimate together with a greater log likelihood. Our theoretical considerations are substantiated by the estimated long-run shocks of Kamber et al. (2018), for which figure 1 plots the autocorrelation function together with the smoothed periodogram.

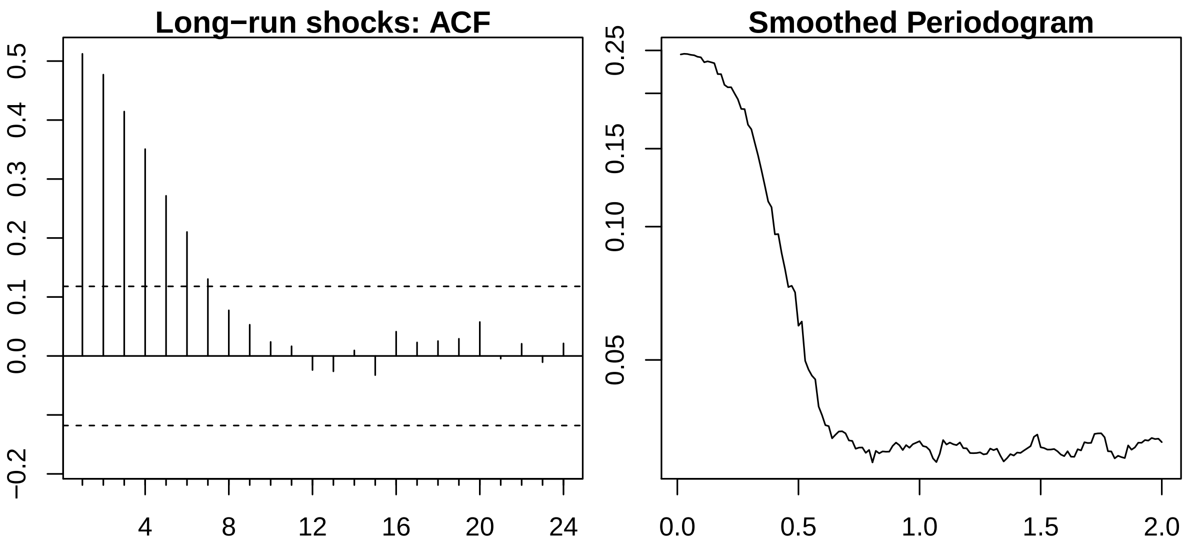

Figure 1: Estimated autocorrelations and smoothed periodogram for the long-run shock estimates in Kamber et al. (2018). The plots were generated based on the publicly available code from Kamber et al. (2018).

As becomes clear from figure 1, the long-run shocks are strongly correlated, and the white noise assumption thus is violated. The rather linear decay of the autocorrelation function and the maximum of the periodogram at the zero frequency indicate that the long-run shocks are fractionally integrated, which is substantiated by the exact local Whittle estimator (Shimotsu and Phillips; 2005) and the estimator of Geweke and Porter-Hudak (1983) that estimate an integration order of 0.2041 and 0.3352 for the long-run shocks. Consequently, the estimated long-run component of Kamber et al. (2018), which is an unweighted sum of the long-run shocks, is not $I(1)$ as implied by the model, but rather integrated of order around 1.3. Therefore, the model of Kamber et al. (2018) is misspecified, which explains why the unrestricted estimator in Morley et al. (2003) returns different parameter estimates and serves as evidence for the fractional hypothesis.

While the results in figure 1 clearly indicate a violation of the $I(1)$ assumption for log US real GDP and support the fractional hypothesis, they are not sufficient to prove that log US real GDP is indeed a fractionally integrated process. A hyperbolic decay of the autocovariance function together with a pole in the spectral density at frequency zero can be generated by both, a fractionally integrated process, and a short memory process contaminated by level shifts and deterministic trends. While the former is typically called a long memory process, the latter case is referred to as spurious long memory in the literature (see e.g. Sibbertsen; 2004). This brings us to the findings of Perron and Wada (2009), who argue that the long-run component of log US real GDP is a deterministic trend with structural breaks, thereby supporting the spurious long memory interpretation. While in particular smooth fractionally integrated processes are well approximated by deterministic processes with structural breaks and vice versa (Diebold and Inoue; 2001; Sibbertsen; 2004), they can be distinguished by accounting for low frequency contaminations in the estimation of the integration order and testing for $d = 0$ (Hou and Perron; 2014), and by utilizing the different slope of the periodogram of long memory and spurious long memory processes in the neighborhood of the zero frequency and testing against spurious long memory (Qu; 2011; Kruse; 2015). In addition, the number of breaks diverges as $n \to \infty$ when the long-run component of GDP is a fractionally integrated process, so that additional breaks should turn up when the model of Perron and Wada (2009) is estimated for a larger sample.

Figure 2 plots the trend-cycle decomposition as suggested by Perron and Wada (2009) and gives a first hint on the the stochastic nature of the long-run component. While the decomposition with a trend break after 1973:1 provides a reasonable explanation of the short- and long-run dynamics until the end of the 20th century, which is the sample size considered by Perron and Wada (2009), this is not the case for the 21st century. There, the estimated cycle is strictly positive until the Great Recession, from which on it decreases until the end of the sample, implying that the Great Recession never ended. Of course, the decomposition could be fixed by adding additional trend breaks. However, the incidence of additional breaks is a first sign for the presence of fractional trends.

The stochastic nature of long-run log GDP can ultimately be assessed up to a certain level of statistical significance by means of the modified local Whittle estimator for d as proposed by Hou and Perron (2014) that is robust to random level shifts among others. The estimator, which requires to take first differences of log GDP for the trend-breaks to become level shifts and for the integration order to be below 0.5, should return an estimate for d that is statistically indistinguishable from zero given that the long-run component of log GDP is deterministic. Conversely, a rejection of the hypothesis $H_0 : d = 0$ implies

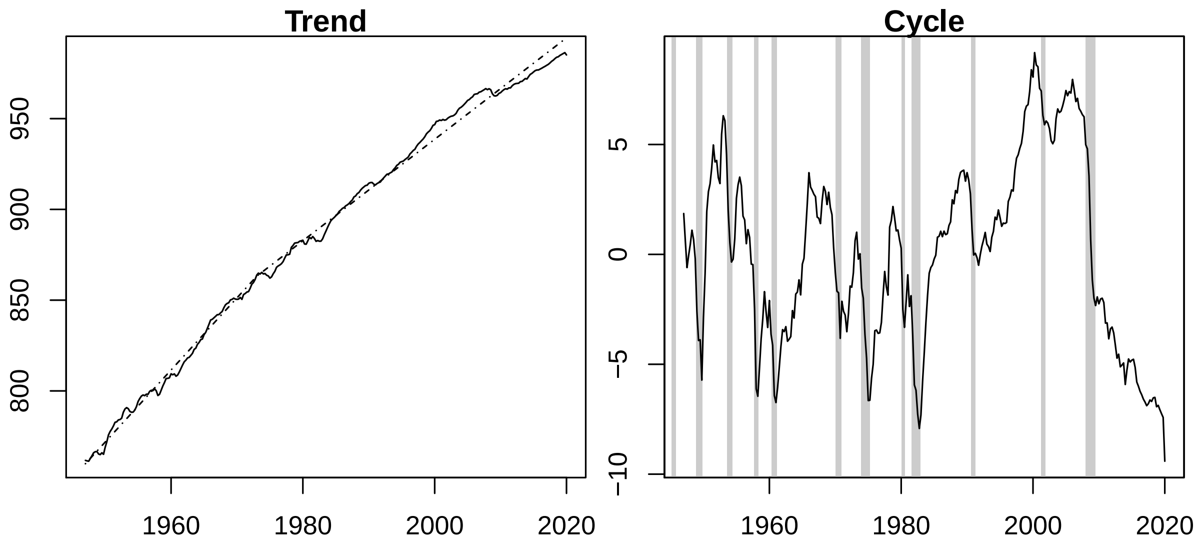

Figure 2: Trend-cycle decomposition according to Perron and Wada (2009). The left plot sketches log US real GDP (solid) together with the fitted values from the regression of log US real GDP on a linear trend with a break after 1973:1 (dashed), while the right plot shows the residuals. Shaded areas correspond to NBER recession periods.

that GDP is driven by a fractionally integrated long-run component given a certain level of statistical significance.

While the method of Hou and Perron (2014) allows to test against fractional integration, the test against spurious long memory of Qu (2011) and its modification of Kruse (2015) test for fractional integration under the null hypothesis and against spurious fractional integration generated by structural breaks under the alternative. The latter two tests utilize the different behavior of the two aforementioned processes over different frequency bands local to zero, do not require to specify the number of breaks or the break dates, and exhibit favorable size and power properties compared to other structural break tests proposed in the literature, see Qu (2011, ch. 6). The only difference in the tests of Qu (2011) and Kruse (2015) is that the latter takes fractional differences according to the estimated d of Hou and Perron (2014) which is shown to increase the power. If long-run GDP is driven by a fractionally integrated process, then the two tests should not reject the null hypothesis. Conversely, a rejection of the null hypothesis implies a that GDP is driven by spurious long memory generated by structural breaks, given a certain level of statistical significance.

Since all aforementioned methods are spectral-based, they require to choose the relevant number of frequencies m around the origin for the estimation of the periodogram, and we follow the authors' suggestions in choosing m. However, the results are robust to different choices of m.

For the first difference of log US real GDP the modified local Whittle estimator of Hou and Perron (2014) with $m = \lfloor n^{0.8} \rfloor$ yields an estimated integration order of 0.2753 (standard error: 0.0513). Since the estimator is asymptotically normally distributed, a 99% confidence interval for the integration order is $[0.1432, 0.4075]$ and clearly the hypothesis that GDP is $I(0)$ around a deterministic trend with structural breaks is rejected. Thus, with probability close to one log US real GDP is fractionally integrated, and the estimates support an integration order of around 1.3 for the levels. In addition, the spectral-based test of Qu (2011) against spurious fractional integration yields a test statistic of 0.8397, which is smaller than even the 10% critical value (1.022) as stated in Qu (2011). Thus, the test fails to reject the null hypothesis of long memory against the alternative of structural breaks on any conventional level of significance. The modified test of Kruse (2015) yields a test statistic of 0.2482 that is even smaller than in the non-modified test of Qu (2011) and has the same critical values. Thus, it again fails to reject the null hypothesis of fractional integration against the alternative of structural breaks. Consequently, while there is comprehensive evidence for log US real GDP being integrated of order around 1.3, there is no evidence for the presence of structural breaks as soon as fractional integration is allowed for.

The fractional hypothesis that long-run log US real GDP is a fractionally integrated process with integration order greater than one thus does not only well explain the results of Morley et al. (2003), Perron and Wada (2009), and Kamber et al. (2018), but is also supported by the above test results against spurious fractional integration generated by structural breaks. As standard UC models do not capture fractionally integrated processes but restrict the integration order to a fixed integer that is assumed to be known, a new UC model that treats d as a continuous random variable and thus allows for fractional trends is of order and will be derived in the next section. This model encompasses the bulk of UC models in the literature, makes comparably weak assumptions on the underlying long- and short-run shocks, and thus allows for a more reliable decomposition of GDP into its long- and short-run component. In addition, it solves the problem of downward-biased estimates for d from non-parametric estimators as those of Shimotsu (2010) and Geweke and Porter-Hudak (1983) due to the small ratio $\operatorname{Var}(\eta_t)/\operatorname{Var}(\varepsilon_t)$, see Sun and Phillips (2004), by explicitly modelling the short-run dynamics.

3 A fractional trend-cycle decomposition

For the trend-cycle decomposition

$$y_t = \tau_t + c_t, \qquad t = 1, ..., n, \tag{2}$$

to be suitable to fractionally integrated processes y_t, we specify the long-run component τ_t as a combination of a linear deterministic process and a fractionally integrated series

$$\tau_t = \mu_0 + \mu_1 t + x_t, \qquad \Delta_+^d x_t = \eta_t, \tag{3}$$

where μ_0 and μ_1 are constants, $d \in \mathbb{R}_+$, and η_t are the long-run shocks that will be defined in (6) below. The specification (3) generalizes the UC literature, where the integration order of τ_t (and thus also of y_t) is typically restricted to $d = 1$ (e.g. Harvey; 1985; Morley et al.; 2003), implying that x_t is a random walk, or to $d = 2$ (e.g. Clark; 1987; Oh and Zivot; 2006). The fractional difference operator Δ^d is defined as

$$\Delta^d = (1-L)^d = \sum_{j=0}^{\infty} \pi_j(d) L^j, \qquad \pi_j(d) = \begin{cases} \frac{j-d-1}{j}\pi_{j-1}(d) & j = 1, 2, ..., \\ 1 & j = 0, \end{cases} \tag{4}$$

and a +-subscript denotes a truncation of an operator at $t \leq 0$, e.g. for an arbitrary process z_t, $\Delta_+^d z_t = \sum_{j=0}^{t-1} \pi_j(d) z_{t-j}$ (Johansen; 2008, def. 1). The fractional long-run component x_t adds flexibility to the weighting of past shocks for $d \in \mathbb{R}_+$ and nests the classic integer-integrated specifications for $d \in \mathbb{N}$. The memory parameter d determines the rate at which the autocovariance function of x_t decays, and a higher d implies a slower decay. For $d < 1$ x_t is mean-reverting, while $d \in [1, 2)$ yields the aggregate of a mean-reverting process. Throughout the paper, we adopt the type II definition of fractional integration (Marinucci and Robinson; 1999) that assumes deterministic starting values for all fractional processes, and, as a consequence, allows for a smooth treatment of the asymptotically stationary ($d < 0.5$) and the nonstationary ($d \geq 0.5$) case. Due to the type II definition of fractional integration the inverse of the fractional difference operator $\Delta_+^{-d} = (1-L)_+^{-d}$ exists for all d, so that we can write $x_t = \Delta_+^{-d}\eta_t = \sum_{j=0}^{t-1} \pi_j(-d)\eta_{t-j}$ with $\pi_j(-d)$ as given in (4). From this, it becomes clear that τ_t in (3) is a type II fractionally integrated process of order d generated by the long-run shocks $\eta_1, ..., \eta_t$, where the weights $\pi_j(-d)$ decrease in j if $d < 1$, are constant and equal to one if $d = 1$, and increase in j if $d > 1$.

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

$$\phi(L_d) c_t = \varepsilon_t, \tag{5}$$

where $\phi(L_d) = 1 - \phi_1 L_d - ... - \phi_p L_d^p$, $L_d = 1 - \Delta^d$ is the fractional lag operator (Johansen; 2008, eq. 2), and ε_t are the short-run shocks that will be defined in (6) below. For stability of the fractional lag polynomial $\phi(L_d)$ the condition of Johansen (2008, cor. 6) is required to hold. It implies that the roots of $|\phi(z)| = 0$ lie outside the image $\mathbb{C}_d$ of the unit disk under the mapping $z \mapsto 1-(1-z)^d$. In fractional models L_d plays the role of the standard lag operator $L_1 = L$, since $(1-L_d)x_t = \Delta^d x_t \sim I(0)$. While for an arbitrary 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 $L_d z_t = (1-(1-L_d))z_t = z_t - \Delta^d z_t$ subtracts an $I(-d)$ process. In addition, $L_d z_t = -\sum_{j=1}^{\infty} \pi_j(d) z_{t-j}$ is a weighted sum of past z_t, and hence L_d qualifies as a lag operator. By definition, the filter $\phi(L_d)$ preserves the integration order of a series since $d > 0$.

Turning to the long- and short-run shocks η_t, ε_t, we assume that they are white noise with finite third and fourth moments and a non-diagonal covariance matrix Q, implying

$$\mathrm{E}\begin{pmatrix}\eta_t\\ \varepsilon_t\end{pmatrix} = 0, \quad \mathrm{Var}\begin{pmatrix}\eta_t\\ \varepsilon_t\end{pmatrix} = Q = \begin{bmatrix}\sigma_\eta^2 & \sigma_{\eta\varepsilon}\\ \sigma_{\eta\varepsilon} & \sigma_\varepsilon^2\end{bmatrix}, \quad \mathrm{Cov}\left[\begin{pmatrix}\eta_t\\ \varepsilon_t\end{pmatrix}, \begin{pmatrix}\eta_{t-s}\\ \varepsilon_{t-s}\end{pmatrix}\right] = 0_{2,2}\ \forall s \neq 0. \quad (6)$$

This is somewhat more general than the bulk of the literature on UC models that requires the shocks to be Gaussian white noise (e.g. Morley et al.; 2003; Perron and Wada; 2009). The white noise assumption is required for the disturbances of the reduced form to aggregate to a moving average process via Grangers lemma (Granger and Morris; 1976), as will become clear in (8) and (18) below. Note that the model allows for contemporaneous correlation between trend and cycle innovations, $\rho = \mathrm{Corr}(\eta_t, \varepsilon_t) \neq 0$.

Our UC model in (2), (3), (5), and (6) is very general in terms of its long-run dynamic characteristics, as it nests the well-known framework of Harvey (1985) for $d = 1$, $\sigma_{\eta\varepsilon} = 0$, where the long-run component is a random walk with drift, and c_t is an autoregressive process of finite order. Correlated shocks as in Balke and Wohar (2002), Morley et al. (2003), and Weber (2011) are explicitly allowed. For $d = 2$, one obtains the double-drift model of Clark (1987), and a fractional plus noise decomposition as proposed in Harvey (2007) for stochastic volatility modelling is obtained by setting $d \in \mathbb{R}_+$, $p = 0$. As will become clear in section 5, the model can easily be enriched to encompass trend breaks as in Perron and Wada (2009).

One interesting property of the fractional UC model is that it can be interpreted as a generalization of the decomposition of Beveridge and Nelson (1981) to the fractional domain. To see this, plug (3) and (5) into (2) and take fractional differences. Then

$$\Delta_+^d (y_t - \mu_0 - \mu_1 t) = \eta_t + \Delta_+^d \phi(L_d)^{-1} \varepsilon_t, \quad (7)$$

where $\phi(L_d)^{-1}$ exists since $\phi(L_d)$ is stable. For an arbitrary z_t, define $\Delta_-^d z_t = \Delta^d z_t - \Delta_+^d z_t$ as the fraction of the polynomial Δ^d that is truncated away, and note that for an $I(0)$ process z_t, $\Delta_-^d z_t = \sum_{j=t}^{\infty} \pi_j(d) z_{t-j} = o_p(1)$ since $\pi_j(d) = O(j^{-d-1})$ (Hassler; 2018, eq. 5.25). We denote $r(t, d, \phi) = -\Delta_-^d \phi(L_d)^{-1} \varepsilon_t = o_p(1)$ and write $\Delta_+^d \phi(L_d)^{-1} \varepsilon_t = \Delta^d \phi(L_d)^{-1} \varepsilon_t +$

$r(t,d,\phi) = (1-L_d)\phi(L_d)^{-1}\varepsilon_t + r(t,d,\phi) = \theta^{\varepsilon}(L_d)\varepsilon_t + r(t,d,\phi)$, with $\theta^{\varepsilon}(L_d) = (1-L_d)\phi(L_d)^{-1}$ as an infinite MA polynomial in the fractional lag operator which follows since $\phi(L_d)$ is invertible. Plugging this result into (7) then yields an ARFIMA model in the fractional lag operator L_d

$$\Delta_+^d(y_t - \mu_0 - \mu_1 t) = \eta_t + \theta^{\varepsilon}(L_d)\varepsilon_t + r(t,d,\phi) = \theta^u(L_d)u_t + r(t,d,\phi), \tag{8}$$

where u_t is white noise with $\mathrm{Var}(u_t) = \sigma_u^2 = \sigma_\eta^2 + \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon}$, and $\theta^u(L_d) = 1 + \sum_{j=1}^{\infty}\theta_j^u L_d^j$. The last equality in (8) follows from the aggregation properties of MA processes in the fractional lag operator with white noise innovations, as shown in appendix C.1 where also a recursive formula for the θ_j^u is derived. For the particular case in (8), where a white noise process is added to an MA process in the fractional lag operator, it follows that $\theta_j^u = \theta_j^{\varepsilon}(\sigma_\varepsilon/\sigma_u)$, $j = 1, 2, ...$

From (8) the fractional Beveridge-Nelson decomposition follows directly

$$\Delta_+^d(y_t - \mu_0 - \mu_1 t) = \theta^u(L_d)u_t + r(t,d,\phi) = \theta^u(1)u_t - (1-L_d)\sum_{k=0}^{\infty} L_d^k u_t \sum_{j=k+1}^{\infty}\theta_j^u + r(t,d,\phi),$$

so that multiplication with Δ_+^{-d} yields the long- and short-run components

$$x_t^{BN} = \Delta_+^{-d}\theta^u(1)u_t = x_t, \qquad c_t^{BN} = -\sum_{k=0}^{\infty} L_d^k u_t \sum_{j=k+1}^{\infty}\theta_j^u = c_t, \tag{9}$$

where we use $\Delta_+^{-d} r(t,d,\phi) = 0$, since $r(t,d,\phi)$ only depends on ε_j, $j \le 0$, and thus all coefficients in Δ_+^{-d} attached to $r(t,d,\phi)$ are zero. While it was shown by Morley et al. (2003) that the UC model in (2), (3), and (5) has a Beveridge-Nelson decomposition for $d = 1$, this carries over directly to the fractional case. x_t^{BN} and c_t^{BN} are identical to the unobserved components in (3) and (5) for any d, which follows immediately from plugging $L_d = 1$ in (8) and multiplying with Δ_+^{-d}, since $\theta^{\varepsilon}(L_d) = (1-L_d)\phi(L_d)^{-1}$ is zero for $L_d = 1$ and $\Delta_+^{-d} r(t,d,\phi) = 0$. Consequently, the fractional trend-cycle decomposition generalizes the $I(1)$ Beveridge-Nelson decomposition to the class of ARFIMA models in the fractional lag operator.

4 Estimation

With the fractional UC model of section 3 at hand, we next consider the estimation of the unknown model parameter vector $\theta = (d, \phi_1, ..., \phi_p, \sigma_\eta^2, \sigma_{\eta\varepsilon}, \sigma_\varepsilon^2)'$ and the latent components x_t, c_t in (2), (3), (5), and (6).

In subsection 4.1 parameter estimation is carried out by means of the quasi maximum likelihood (QML) estimator based on the prediction error form of the UC model, which is in line with the methodological UC literature, see e.g. Durbin and Koopman (2012, ch. 7.2). To arrive at the prediction error form, we derive an estimator for the conditional expectations of the latent x_t, c_t, given the parameters θ and the data until the preceding period $y_1, ..., y_{t-1}$. Given all data available $y_1, ..., y_n$ and a parameter vector θ, we introduce a single-step estimator for the conditional expectations of x_t, c_t. The estimators for the conditional expectations of x_t, c_t are shown to be identical to the Kalman filter and smoother but are computationally more efficient for the fractional UC model. For empirical researchers, the results derived in subsection 4.1 are sufficient to fully estimate all unknown terms in the fractional UC model.

The remaining two subsections 4.2 and 4.3 are to show identification and to ensure that the QML estimator is consistent and asymptotically normal, implying that it produces reliable estimates in large samples, that standard inference is valid for hypothesis testing, and that the Kalman smoother is asymptotically the minimum variance linear unbiased estimator (MVLUE) for x_t, c_t, given the data $y_1, ..., y_n$.

The asymptotic results in subsection 4.3 are not only important for the fractional UC model itself, but also fill a gap in the literature on integer-integrated UC models that are nested in our setup. There, e.g. in Morley et al. (2017), it is argued that consistency of the QML estimator for integer-integrated UC models follows from the reduced form being nested in the class of ARMA models, where consistency of the QML estimator is well established (see e.g. Pötscher; 1991). However, this is not sufficient, as UC models have implicit restrictions on the parameters of the reduced form ARMA models. While of course the reduced form parameters of an integer-integrated UC model can be consistently estimated by a QML estimator for ARMA models, for the asymptotic properties of the QML-ARMA-estimator to carry over to the prediction-error-form QML estimator of UC models a continuous mapping theorem is required to hold. To the best of our knowledge, the literature on integer-integrated UC models lacks such a proof. Consequently, the results in subsection 4.3 also establish consistency and asymptotic normality for integer-integrated UC models, which so far has only been assumed, and give an analytical expression for the Fisher information matrix.

In the following, we denote the data-generating parameter vector $\theta_0 \in \Theta$. $\Theta = D \times \Phi \times \Omega$ is the parameter space with $D = \{d \in \mathbb{R}_+ : d_{min} \leq d \leq d_{max}\}$, $\Phi = \{\phi \in \mathbb{R}^p : \deg(\phi) \leq p, \phi(0) = 1\}$ and the stability condition as discussed below (5) holds, and $\Omega \subseteq \mathbb{R}^3$ with the parameter space Ω being bounded by the conditions $\sigma_\eta^2, \sigma_\varepsilon^2 \geq 0$, and $|\operatorname{Corr}(\eta_t, \varepsilon_t)| < 1$. We define $\mathcal{F}_t$ as the σ-field generated by the observable variables $y_1, ..., y_t$. The expected value operator $\mathrm{E}_\theta(z_t)$ of an arbitrary random variable z_t denotes that expectation is taken with respect to the distribution of z_t given θ, so that $\mathrm{E}_{\theta_0}(z_t) = \mathrm{E}(z_t)$. To simplify the proofs, we assume that y_t has been mean- and trend-adjusted and set $\mu_0 = \mu_1 = 0$, analogous to Hualde and Robinson (2011).

4.1 Estimation of latent components and model parameters

Parameter estimation in the UC literature is typically carried out by the (quasi) maximum likelihood estimator of the prediction error form (Durbin and Koopman; 2012, ch. 7.2), which is defined as the (quasi) likelihood of the residual $v_t(\theta) = y_t - \mathrm{E}_\theta(y_t|\mathcal{F}_{t-1})$ and requires an analytical expression for the conditional expectation. The latter is obtained from the Kalman filter, which, for any latent z_t and parameter vector θ, is a recursion for calculating $\mathrm{E}_\theta(z_t|\mathcal{F}_t)$, $\mathrm{E}_\theta(z_{t+1}|\mathcal{F}_t)$, and the corresponding conditional variances (Durbin and Koopman; 2012, ch. 4.3). In a second step, the Kalman smoother is utilized to estimate the latent components given all the data available $y_1, ..., y_n$, and the QML estimates for θ_0.

Extending this approach to the fractional UC model, we first derive analytical expressions for the conditional expectations $\mathrm{E}_\theta(x_t|\mathcal{F}_{t-1})$, $\mathrm{E}_\theta(c_t|\mathcal{F}_{t-1})$ that are identical to the Kalman filter but can be calculated in a single step. Given a parameter vector θ, this allows to calculate the prediction error $v_t(\theta) = y_t - \mathrm{E}_\theta(y_t|\mathcal{F}_{t-1}) = y_t - \mathrm{E}_\theta(x_t|\mathcal{F}_{t-1}) - \mathrm{E}_\theta(c_t|\mathcal{F}_{t-1})$. Based on $v_t(\theta)$, we construct the QML estimator $\hat{\theta}$. Furthermore, we derive analytical expressions for the conditional expectations $\mathrm{E}_\theta(x_t|\mathcal{F}_n)$, $\mathrm{E}_\theta(c_t|\mathcal{F}_n)$ that are identical to the Kalman smoother, but again only involve a single step in their calculation. To derive $\mathrm{E}_\theta(x_t|\mathcal{F}_{t-1})$, $\mathrm{E}_\theta(c_t|\mathcal{F}_{t-1})$ for the fractional UC model in (2), (3), (5), and (6), let $y_{1:t} = (y_1, ..., y_t)'$, $\eta_{1:t} = (\eta_1, ..., \eta_t)'$, and $\varepsilon_{1:t} = (\varepsilon_1, ..., \varepsilon_t)'$ denote the vectors collecting all variables until t, and define $\phi(L_d)^{-1} = \sum_{j=0}^{\infty} \omega_j L^j$ as the stable MA polynomial of ε_t. For an arbitrary z_t and a parameter vector θ, the Kalman filter recursively determines $z_{t+1|t} = \mathrm{E}_\theta(z_{t+1}|\mathcal{F}_t) = \mathrm{Cov}_\theta(z_{t+1}, y_{1:t}) \, \mathrm{Var}_\theta(y_{1:t})^{-1} y_{1:t}$, see Durbin and Koopman (2012, lemma 1). For our latent components as defined in (3) and (5) it follows from $x_{t+1} = \Delta_+^{-d}\eta_{t+1} = \sum_{j=0}^{t} \pi_j(-d)\eta_{t+1-j}$ and $c_{t+1} = \phi(L_d)^{-1}\varepsilon_{t+1} = \sum_{j=0}^{\infty} \omega_j \varepsilon_{t+1-j} = \sum_{j=0}^{t} \omega_j \varepsilon_{t+1-j} + o_p(1)$ that

$$x_{t+1|t}(\theta) = \sum_{j=1}^{t} \pi_j(-d) \, \mathrm{Cov}_\theta(\eta_{t+1-j}, y_{1:t}) \, \mathrm{Var}_\theta(y_{1:t})^{-1} y_{1:t}, \tag{10}$$

$$c_{t+1|t}(\theta) = \sum_{j=1}^{t} \omega_j \, \mathrm{Cov}_\theta(\varepsilon_{t+1-j}, y_{1:t}) \, \mathrm{Var}_\theta(y_{1:t})^{-1} y_{1:t} + o_p(1), \tag{11}$$

where the $o_p(1)$ term in (11) results from the truncation of the sum for indexes $j \leq 0$ where y_j is unobservable. As it is asymptotically negligible it will be omitted in the following. The terms $\mathrm{Cov}_\theta(\eta_{t-j}, y_{1:t})$, $\mathrm{Cov}_\theta(\varepsilon_{t-j}, y_{1:t})$, with $j \geq 0$, and $\mathrm{Var}_\theta(y_{1:t})$ solely depend on the parameter vector θ. From (7) we obtain $y_t = \Delta_+^{-d}\eta_t + \phi(L_d)^{-1}\varepsilon_t = \sum_{j=0}^{t-1}\pi_j(-d)\eta_{t-j} + \sum_{j=0}^{\infty}\omega_j\varepsilon_{t-j}$. Then $\mathrm{Cov}_\theta(y_t, \eta_{t-j}) = \pi_j(-d)\sigma_\eta^2 + \omega_j\sigma_{\eta\varepsilon}$, $\mathrm{Cov}_\theta(y_t, y_{t-j}) = \sum_{k=0}^{t-j-1}[\pi_k(-d)\pi_{k+j}(-d)\sigma_\eta^2 + (\omega_k\pi_{k+j}(-d) + \pi_k(-d)\omega_{k+j})\sigma_{\eta\varepsilon} + \omega_k\omega_{k+j}\sigma_\varepsilon^2] + o_p(1)$, and finally $\mathrm{Cov}_\theta(y_t, \varepsilon_{t-j}) = \pi_j(-d)\sigma_{\eta\varepsilon} + \omega_j\sigma_\varepsilon^2$, for all $j \geq 0$, so that

$$\mathrm{Var}\begin{pmatrix}\eta_{1:t}\\ \varepsilon_{1:t}\\ y_{1:t}\end{pmatrix} = \begin{bmatrix}\sigma_\eta^2 I & \sigma_{\eta\varepsilon}I & \Sigma_{\eta_{1:t}y_{1:t}}\\ \sigma_{\eta\varepsilon}I & \sigma_\varepsilon^2 I & \Sigma_{\varepsilon_{1:t}y_{1:t}}\\ \Sigma'_{\eta_{1:t}y_{1:t}} & \Sigma'_{\varepsilon_{1:t}y_{1:t}} & \Sigma_{y_{1:t}}\end{bmatrix}, \quad \mathrm{Var}\begin{pmatrix}\eta_{1:t}\\ \varepsilon_{1:t}\\ y_{1:n}\end{pmatrix} = \begin{bmatrix}\sigma_\eta^2 I & \sigma_{\eta\varepsilon}I & \Sigma_{\eta_{1:t}y_{1:n}}\\ \sigma_{\eta\varepsilon}I & \sigma_\varepsilon^2 I & \Sigma_{\varepsilon_{1:t}y_{1:n}}\\ \Sigma'_{\eta_{1:t}y_{1:n}} & \Sigma'_{\varepsilon_{1:t}y_{1:n}} & \Sigma_{y_{1:n}}\end{bmatrix},$$

where $\Sigma_{\eta_{1:t}y_{1:t}} = \mathrm{Cov}_\theta(\eta_{1:t}, y_{1:t})$, $\Sigma_{\eta_{1:t}y_{1:n}} = \mathrm{Cov}_\theta(\eta_{1:t}, y_{1:n})$, $\Sigma_{\varepsilon_{1:t}y_{1:t}} = \mathrm{Cov}_\theta(\varepsilon_{1:t}, y_{1:t})$, $\Sigma_{\varepsilon_{1:t}y_{1:n}} = \mathrm{Cov}_\theta(\varepsilon_{1:t}, y_{1:n})$, $\Sigma_{y_{1:t}} = \mathrm{Var}_\theta(y_{1:t})$, and $\Sigma_{y_{1:n}} = \mathrm{Var}_\theta(y_{1:n})$ with entries as given above. Let $e_j(t)$ be a t-dimensional unit vector with a one at column j and zeros elsewhere. Then the estimators for the conditional expectations as defined in (10) and (11) are

$$x_{t+1|t}(\theta) = \sum_{j=1}^{t}\pi_j(-d)e_{t+1-j}(t)\Sigma_{\eta_{1:t}y_{1:t}}\Sigma_{y_{1:t}}^{-1}y_{1:t}, \tag{12}$$

$$c_{t+1|t}(\theta) = \sum_{j=1}^{t}\omega_j e_{t+1-j}(t)\Sigma_{\varepsilon_{1:t}y_{1:t}}\Sigma_{y_{1:t}}^{-1}y_{1:t}, \tag{13}$$

and can easily be calculated with θ, $y_{1:t}$ at hand. They are identical to the Kalman filter but do not involve any recursions (Durbin and Koopman; 2012, ch. 4.1).

The Kalman smoother recursively calculates $\mathrm{E}_\theta(x_t|\mathcal{F}_n)$, $\mathrm{E}_\theta(c_t|\mathcal{F}_n)$, for which an analytical expression follows directly from (12) and (13)

$$x_{t|n}(\theta) = \sum_{j=0}^{t-1}\pi_j(-d)e_{t-j}(t)\Sigma_{\eta_{1:t}y_{1:n}}\Sigma_{y_{1:n}}^{-1}y_{1:n}, \tag{14}$$

$$c_{t|n}(\theta) = \sum_{j=0}^{t-1}\omega_j e_{t-j}(t)\Sigma_{\varepsilon_{1:t}y_{1:n}}\Sigma_{y_{1:n}}^{-1}y_{1:n}. \tag{15}$$

The single-step calculation of (12), (13), (14), and (15) somewhat differs from the standard computation of the Kalman filter and smoother in the literature, where a model is cast in state space form to sequentially calculate the Kalman recursions, as described e.g. in Durbin and Koopman (2012, ch. 4.3 and 4.4). While both approaches yield identical results and only differ in their computation, the natural reason for the recursive calculation of the Kalman filter is to avoid the computationally intensive inversion of the $n \times n$ matrix $\Sigma_{y_{1:n}}$. However, computational gains from the recursive calculation of the Kalman filter crucially depend on the dimension of the state vector, which itself depends on the dynamic specification of the latent components. Since x_t and c_t each require a state vector that is at least of dimension $n-1$ when cast in state space form, we found the single-step calculation of the Kalman filter and smoother to be much faster for the fractional UC model.

From (12) and (13), an objective function for the QML estimator $\hat{\theta}$ can be constructed via the prediction error

$$v_t(\theta) = y_t - x_{t|t-1}(\theta) - c_{t|t-1}(\theta) = y_t - \mathrm{E}_\theta(y_t|\mathcal{F}_{t-1}), \tag{16}$$

and is given by

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} l_t(\theta), \qquad l_t(\theta) = -\frac{1}{2}\log\sigma_v^2 - \frac{1}{2\sigma_v^2}v_t^2(\theta). \tag{17}$$

As will be shown in subsection 4.3, $\sigma_v^2 = \sigma_\eta^2 + \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon} = \sigma_u^2$, so that the QML estimator (17) is fully specified. With the single-step formulas for the Kalman filter in (12) and (13) at hand, the model parameters can be estimated via the QML estimator in (17), while the latent components are estimated via the single-step Kalman smoother given $\hat{\theta}$, see (14) and (15).

However, for the QML estimator in (17) to asymptotically yield reliable estimates, it remains to be shown that $\hat{\theta} \xrightarrow{p} \theta_0$ as n diverges, implying that the QML estimator is consistent. In addition, asymptotic normality of the QML estimator needs to be shown for standard inference to be valid for hypothesis testing. Since the objective function of the QML estimator in (17) is quite inconvenient to establish the asymptotic properties, we tackle them based on a reduced form representation of the fractional UC model in the following two subsections. As they are quite technical, we briefly summarize the key findings and their implications.

In subsection 4.2, we show that the fractional UC model is uniquely identified for any degree p of the polynomial $\phi(L_d)$ whenever $d \neq 1$, and for any $p > 1$ when $d = 1$, implying that the objective function in (17) has a unique maximum. Subsection 4.3 then demonstrates that the reduced form prediction error as given in (24) below is identical to the Kalman-filter-based prediction error in (16), and also the parameter spaces are identical. Therefore, the Kalman-filter-based QML estimator $\hat{\theta}$ in (17) and the reduced-form-based QML estimator as given in (25) are identical, implying that their asymptotic properties will also be identical. Consistency and asymptotic normality are then established in theorems 4.1 and 4.2. Consequently, (14) and (15) asymptotically are the minimum variance linear unbiased estimators for the conditional expectations of x_{t+1} and c_{t+1} given $y_1, ..., y_n$ when the Kalman smoother is evaluated at $\hat{\theta}$, see Durbin and Koopman (2012, lemma 2).

4.2 Identification

To show identification of the fractional UC model, we first derive the reduced form of (2), (3), (5), and (6), which is an ARFIMA model in the fractional lag operator. To see this, multiply (7) with $\phi(L_d)$ to obtain

$$\phi(L_d)\Delta_+^d y_t = \phi(L_d)\eta_t + \Delta_+^d \varepsilon_t = \phi(L_d)\eta_t + (1 - L_d)\varepsilon_t - \Delta_-^d \varepsilon_t = \psi(L_d)u_t - \Delta_-^d \varepsilon_t, \quad (18)$$

where $\psi(L_d) = \phi(L_d)\theta^u(L_d)$ is a moving average polynomial of infinite order that results from the aggregation of $\phi(L_d)\eta_t + (1-L_d)\varepsilon_t$. Its existence together with a recursive formula for the coefficients ψ_j is shown in appendix C.1. As before, u_t holds the disturbances of the reduced form and is white noise with $\mathrm{Var}(u_t) = \sigma_u^2 = \sigma_\eta^2 + \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon}$, which follows from Granger and Morris (1976, p. 248f) for contemporaneously dependent white noise processes ε_t, η_t.

While aggregating MA processes in the standard lag operator yields an MA process whose lag length equals the maximum lag order of its aggregates, this does not hold in general for the aggregation of MA processes in the fractional lag operator L_d, since $L_d^i u_t$, $L_d^j u_t$ are not independent for $i, j > 1$, $i \neq j$. Only for $p = 1$ equation (18) asymptotically becomes an ARFIMA$(1, d, 1)$ model in the fractional lag operator, since $\eta_t + \varepsilon_t$ and $L_d(\phi_1 \eta_t + \varepsilon_t)$ are independent. For $d \in \mathbb{N}$ the model in (18) nests the integer-integrated ARIMA models. Due to the inclusion of the fractional lag operator (18) differs from the standard ARFIMA model. Nonetheless, (18) exhibits an ARFIMA representation in L as $\phi(L_d)$, $\psi(L_d)$ can be written as stable polynomials in the standard lag operator L, see (21) and (22) below.

Since y_t is unobserved for $t \leq 0$, an estimator for θ will necessarily be based on a truncated representation of (18) where all random variables indexed by $t \leq 0$ will be set to zero. As before, we denote this truncation with a +-subscript and define $u_t(\theta)$ as the residual given θ. (18) then becomes

$$\phi_+(L_d)\Delta_+^d y_t = \psi_+(L_d)u_t(\theta). \quad (19)$$

While $d, \phi_1, ..., \phi_p$ are directly identified from the reduced form (19), it remains to be shown that $\sigma_\eta^2, \sigma_{\eta\varepsilon}, \sigma_\varepsilon^2$ can be uniquely recovered from (19) by matching the autocovariance functions of $\psi_+(L_d)u_t(\theta)$ and $\phi_+(L_d)\eta_t + (1 - L_d)_+\varepsilon_t$.

To match the autocovariance functions, we rewrite $\phi(L_d)$, $\psi(L_d)$ as polynomials in the standard lag operator L. With regard to the first polynomial one has

$$\begin{aligned}\phi(L_d) &= 1-\sum_{j=1}^{p}\phi_j(1-\Delta^d)^j = 1-\sum_{j=1}^{p}\phi_j\sum_{k=0}^{j}\binom{j}{k}(-1)^k\Delta^{dk}\\ &= (1-\phi_1-...-\phi_p)-\sum_{j=1}^{p}(-1)^j\sum_{k=j}^{p}\phi_k\binom{k}{j}\sum_{i=0}^{\infty}\pi_i(dj)L^i\\ &= 1-\sum_{j=1}^{p}(-1)^j\sum_{k=j}^{p}\phi_k\binom{k}{j}\sum_{i=1}^{\infty}\pi_i(dj)L^i, \end{aligned}\tag{20}$$

where the second step uses the binomial theorem, the third step rearranges the sums, and the fourth step utilizes $\pi_0(b) = 0$ for all b, so that $\sum_{j=1}^{p}(-1)^j\sum_{k=j}^{p}\phi_k\binom{k}{j}\pi_0(dj) = \sum_{j=1}^{p}(-1)^j\sum_{k=j}^{p}\phi_k\binom{k}{j} = \phi_1+...+\phi_p$. It then follows from (20) that

$$\phi(L_d) = \tilde{\phi}(L) = 1-\sum_{l=1}^{\infty}\tilde{\phi}_l L^l, \qquad \tilde{\phi}_l = \sum_{j=1}^{p}(-1)^j\pi_l(dj)\sum_{k=j}^{p}\phi_k\binom{k}{j}, \tag{21}$$

and analog for $\psi(L_d)$

$$\psi(L_d) = \tilde{\psi}(L) = 1+\sum_{l=1}^{\infty}\tilde{\psi}_l L^l, \qquad \tilde{\psi}_l = \sum_{j=1}^{p}\pi_l(dj)\sum_{k=j}^{p}\psi_k\binom{k}{j}. \tag{22}$$

From (19), (21), and (22) it follows that

$$\mathrm{Cov}(\phi_+(L_d)u_t, \phi_+(L_d)u_{t-k}) = \sigma_u^2\gamma_k^u(t) = \sigma_\eta^2\gamma_k^\eta(t)+\sigma_\varepsilon^2\gamma_k^\varepsilon(t)-\sigma_{\eta\varepsilon}\gamma_k^{\eta\varepsilon}(t),$$

with $\gamma_k^u(t) = \sum_{j=0}^{t-k-1}\tilde{\psi}_j\tilde{\psi}_{j+k}$, $\gamma_k^\eta(t) = \sum_{j=0}^{t-k-1}\tilde{\phi}_j\tilde{\phi}_{j+k}$, $\gamma_k^\varepsilon(t) = \sum_{j=0}^{t-k-1}\pi_j(d)\pi_{j+k}(d)$, and $\gamma_k^{\eta\varepsilon}(t) = \sum_{j=0}^{t-k-1}\left(\tilde{\phi}_j\pi_{j+k}(d)+\tilde{\phi}_{j+k}\pi_j(d)\right)$ where $\tilde{\phi}_0 = -1$. Thus, the model is identified if A in

$$\begin{pmatrix}\gamma_0^u(t)\\ \gamma_1^u(t)\\ \gamma_2^u(t)\end{pmatrix}\sigma_u^2 = A\begin{pmatrix}\sigma_\eta^2\\ \sigma_{\eta\varepsilon}\\ \sigma_\varepsilon^2\end{pmatrix}, \qquad A = \begin{bmatrix}\gamma_0^\eta(t) & \gamma_0^{\eta\varepsilon}(t) & \gamma_0^\varepsilon(t)\\ \gamma_1^\eta(t) & \gamma_1^{\eta\varepsilon}(t) & \gamma_1^\varepsilon(t)\\ \gamma_2^\eta(t) & \gamma_2^{\eta\varepsilon}(t) & \gamma_2^\varepsilon(t)\end{bmatrix},$$

has full rank, so that it is invertible and σ_η^2, $\sigma_{\eta\varepsilon}$, σ_ε^2 can be uniquely recovered from $\gamma_0^u(t)$, $\gamma_1^u(t)$, $\gamma_2^u(t)$. Note that A only has reduced rank if $c\tilde{\phi}_j = \pi_j(d)$ for any constant c and all j or if one of the three rows in A is zero. The former case is excluded by the definition of $\phi(L_d)$ as otherwise $\phi(L_d)$ would not generate an $I(0)$ cycle. For the latter case, note that for any $d \in \mathbb{R}_+ \setminus \{1\}$, $\pi_j(d) \neq 0$ for all $j = 0, 1, 2$, so that $\gamma_k^\varepsilon, \gamma_k^{\eta\varepsilon} \neq 0$ for $d \neq 1$, $k = 0, 1, 2$. Thus, for $d \neq 1$ the model is identified for any degree of the polynomial

$\phi(L_d)$. A special case occurs for $d = 1$, where the fractional UC model becomes an $I(1)$ UC model as considered in Morley et al. (2003). There, $\gamma_2^{\varepsilon}(t) = 0$, since all $\pi_j(1) = 0$ for $j > 1$, so that for the model to be identified it is required that $\gamma_2^{\eta}(t) \neq 0$, which holds for all $p \geq 2$. Thus, the model is identified for all $p \geq 0$ if $d \neq 1$, and identified for all $p \geq 2$ if $d = 1$. Contrary to the $I(2)$-model of Oh et al. (2008) with three shocks that requires $p \geq 4$, our model is identified for any $p \geq 0$ when $d = 2$, which follows from $\pi_j(2) \neq 0$ for $j \leq 2$. Consequently, the fractional UC model allows for a parsimonious parametrization compared to the correlated UC model of Morley et al. (2003) and can be applied to processes with $p < 2$ whenever $d \neq 1$.

4.3 Consistency and asymptotic normality

Having shown that the model is identified, we next derive the asymptotic properties of the QML estimator based on the reduced form that will be shown to be identical to the Kalman-filter-based QML estimator $\hat{\theta}$ in (17). To derive the likelihood function, we solve the reduced form equation (19) for the residual $u_t(\theta)$ and obtain

$$u_t(\theta) = \psi_+(L_d)^{-1}\phi_+(L_d)\Delta_+^d y_t, \tag{23}$$

where $u_t(\theta)$ is the reduced form prediction error that depends on the parameter vector θ and $u_t(\theta_0) = u_t$. $u_t(\theta)$ qualifies as the prediction error since the one-step ahead prediction of y_t, given the information until the preceding period $\mathcal{F}_{t-1}$, yields an error

$$\begin{aligned} y_t - \mathrm{E}_\theta(y_t|\mathcal{F}_{t-1}) &= y_t - \mathrm{E}_\theta\left[-(\psi_+(L_d)^{-1}\phi_+(L_d)\Delta_+^d - 1)y_t|\mathcal{F}_{t-1}\right] \\ &= y_t + (\psi_+(L_d)^{-1}\phi_+(L_d)\Delta_+^d - 1)y_t = u_t(\theta), \end{aligned} \tag{24}$$

where in the second step it was used that $(\psi_+(L_d)^{-1}\phi_+(L_d)\Delta_+^d - 1)y_t$ is $\mathcal{F}_{t-1}$-measurable since $\psi(0) = \phi(0) = 1$ so that y_t cancels out and the term only consists of lagged $y_1, ..., y_{t-1}$ that are contained in $\mathcal{F}_{t-1}$. From (24) it follows directly that the reduced form prediction error $u_t(\theta)$ is identical to the Kalman-filter-based prediction error $v_t(\theta)$ given in (16), since $u_t(\theta) = y_t - \mathrm{E}_\theta(y_t|\mathcal{F}_{t-1}) = y_t - x_{t|t-1} - c_{t|t-1} = v_t(\theta)$.

In line with the literature on unobserved components models, we define the objective function of the maximum likelihood estimator as the quasi log likelihood of the prediction error (23) for our model in (2), (3), (5), and (6)

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} l_t(\theta), \qquad l_t(\theta) = -\frac{1}{2}\log\sigma_u^2 - \frac{1}{2\sigma_u^2}u_t^2(\theta), \tag{25}$$

where $l_t(\theta)$ is the profile log likelihood of the normal distribution and $\sigma_u^2 = \sigma_\eta^2 + \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon}$. Since the parameter space Θ in (25) is identical to the parameter space as considered in (17), and since $u_t(\theta) = v_t(\theta)$, it follows directly the two estimators in (17) and (25) are identical, with (25) being more convenient for the analysis of the asymptotic properties of the QML estimator.

The asymptotic properties of the QML estimator are summarized in theorems 4.1 and 4.2, where the former theorem establishes consistency and the latter considers asymptotic normality. The proofs are contained in appendix C.2, and a brief explanation of each proof is given below the respective theorem.

Theorem 4.1. *The maximum likelihood estimator for θ_0 of the fractional UC model in* (2)*,* (3)*,* (5)*, and* (6) *is consistent, $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$.*

In the proof of 4.1 we face the following challenges. First, $u_t(\theta)$ as given in (23) is integrated of order $I(d_0 - d)$, and thus is stationary for $d_0 - d < 0.5$, and nonstationary for $d_0 - d \geq 0.5$. While in the former case it can be shown that a uniform weak law of large numbers applies to the objective function, yielding uniform convergence in probability on any compact subset of $d_0 - d < 0.5$, this does not hold for the nonstationary case. There, under additional assumptions a functional central limit theorem applies to the objective function, and the rate of convergence of the QML estimator can be shown to depend on the integration order $d_0 - d$. Since the asymptotic behavior of the objective function changes around $d_0 - d = 0.5$, the objective function does not uniformly converge in probability on the set of admissible values for d. Fortunately, for a broad class of multivariate ARFIMA processes, that also contains our fractional UC model, it was shown by Nielsen (2015) that the relevant parameter space reduces to the region $d_0 - d < 0.5$ asymptotically, where $u_t(\theta)$ is stationary. Thus, we solve the problem of different limiting behavior of the objective function in the stationary and the nonstationary region by showing that the results of Nielsen (2015) carry over to the fractional UC model, so that attention can be restricted to the case where $d_0 - d < 0.5$.

The second challenge is to show that the objective function converges uniformly on Θ given that $d_0 - d < 0.5$, which implies that the objective function satisfies a uniform weak law of large numbers (UWLLN). While pointwise convergence in probability of the objective function follows directly from the weak law of large numbers for stationary processes, uniform convergence is more difficult to establish. Following Wooldridge (1994, th. 4.2), pointwise convergence can be strengthened to uniform convergence if the partial derivatives of the objective function w.r.t. θ satisfy a weak law of large numbers, implying that they are dominated uniformly by a random variable that is $O_p(1)$ (Newey; 1991, cor. 2.2). By carefully studying the partial derivatives of $l_t(\theta)$ we show that a weak law of large numbers applies, so that a UWLLN holds for the objective function. Given that the model is identified as shown in section 4.2, consistency for the QML estimator follows directly (Wooldridge; 1994, th. 4.3).

With the consistency result at hand, we are able to derive the asymptotic distribution of the QML estimator, where theorem 4.2 summarizes the results.

Theorem 4.2. *The maximum likelihood estimator for θ_0 of the fractional UC model in* (2), (3), (5), *and* (6) *is asymptotically normal,* $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega_0^{-1})$ *as* $n \to \infty$, *and* Ω_0 *is the Fisher information matrix.*

The proof is contained in appendix C.2 and is carried out by considering a Taylor expansion of the score function at θ_0. We first show that a central limit theorem applies to the score function at θ_0. Next, to evaluate the Hessian matrix in the Taylor expansion at θ_0, we prove that a UWLLN holds for the Hessian matrix by showing that the the Hessian matrix itself and its first partial derivatives satisfy a weak law of large numbers. From Wooldridge (1994, th. 4.2) it then follows that the QML estimator $\hat{\theta}$ is asymptotically normally distributed with the asymptotic variance being equal to the inverse Fisher information matrix, see (70).

5 Fractional trends and cycles in US GDP

With the fractional UC model of section 3 and estimators for the unknown parameters and latent components as derived in section 4 at hand, we can now estimate a fractional trend-cycle decomposition for log US real GDP, where several advantages of the model directly become apparent.

First, as already discussed in section 2, state-of-the-art UC models as suggested by Balke and Wohar (2002), Morley et al. (2003), Perron and Wada (2009), and Kamber et al. (2018) either yield implausible trend-cycle decompositions that contradict the NBER chronology, or are misspecified in terms of the long-run dynamic properties, so that the business cycle estimate (i.e. the estimated cyclical component $\hat{c}_{t|n}$) is unreliable. The fractional UC model takes into account the evidence for fractional integration in log US real GDP contained in section 2 and thus allows for a reliable estimation of the business cycle, particularly as it nests the aforementioned models. Second, results from the fractional UC model extend the empirical literature by providing new insights on the data-generating mechanism of log US real GDP. In particular, treating the integration order d as a random variable allows to draw inference on the persistence of long-run shocks. As the model encompasses integer-integrated UC models with autoregressive cycles, point hypotheses like $d = 1$ can be tested to see whether UC models as e.g. those of Harvey (1985), Balke and Wohar (2002), Morley et al. (2003), and Kamber et al. (2018) match the long-run dynamics of GDP. Third, the model allows to investigate whether stochastic UC models with correlated long- and short-run shocks are able to explain the NBER chronology. If so, then implausible estimates from integer-integrated correlated UC models are caused by a violation of the $I(1)$ hypothesis. Fourth, estimates from the fractional UC model can be compared to those of a model with a deterministic long-run component and structural breaks as suggested by Perron and Wada (2009). While the latter is clearly misspecified for log US real GDP, as shown in section 2, it may still provide a good approximation to the fractional long-run component of log US real GDP based on a simple model. And fifth, from a methodological perspective the endogenous treatment of the integration order neither requires assumptions about the persistence of a series nor prior unit root testing or differencing.

While decomposing GDP is clearly the key application of UC models, trend-cycle decompositions of several other macroeconomic aggregates have been considered in the literature. The interested reader is referred to appendix B, where applications of the fractional UC model for industrial production, investment, consumption, and employment are considered that highlight the benefits of our flexible, data-driven method to treat permanent and transitory components in macroeconomic applications.

The data on seasonally and inflation adjusted US GDP was downloaded from the Federal Reserve Bank of St. Louis (mnemonic: GDPC1), is in quarterly frequency, spans from 1947:1 to 2020:1, and was log-transformed. It was trend- and mean-adjusted based on the exact local Whittle estimator of Shimotsu (2010), which allows for deterministic trends and yields an estimate $\tilde{d}$ for the integration order. Estimates for μ_0, μ_1 in (3) were then obtained via the least squares regression

$$\Delta_+^{\tilde{d}} \log GDP_t = \mu_0 \Delta_+^{\tilde{d}} 1 + \mu_1 \Delta_+^{\tilde{d}} t + error_t, \qquad t = 1, ..., n. \tag{26}$$

The trend-adjusted series $y_t = \log GDP_t - \hat{\mu}_0 - \hat{\mu}_1 t$ then enters the UC model as the observable variable.

For the QML estimation of θ_0 via (17) we draw 100 combinations of starting values from uniform distributions with appropriate support and maximize the quasi log likelihood of the fractional UC model via the Nelder-Mead algorithm. The lag order is set to $p = 1$ as this minimizes both, the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) for $p \in \{1, ..., 8\}$, where the effective number of observations is held fixed according to Ng and Perron (2005). However, we also report estimation results for $p = 4$ to show that both, the estimates for θ_0 and the decomposition into x_t and c_t are robust to the choice of higher p.

	$p=1$		$p=4$	
	estimate	std. error	estimate	std. error
$\hat{d}$	1.3365	0.1360	1.2697	0.1577
$\hat{\sigma}_\eta^2$	0.1193	0.1726	0.1578	0.1774
$\hat{\sigma}_{\eta\varepsilon}$	-0.4021	0.3861	-0.4193	0.3224
$\hat{\sigma}_\varepsilon^2$	1.4757	0.6334	1.3944	0.4714
$\hat{\phi}_1$	0.8417	0.0449	0.9302	0.0987
$\hat{\phi}_2$			-0.0005	0.0592
$\hat{\phi}_3$			-0.0597	0.0391
$\hat{\phi}_4$			-0.0052	0.0195
$l_t(\hat{\theta})$	-370.3803		-368.9246	
AIC	735.7807		744.3060	
BIC	754.1815		773.7473	
$\widehat{\text{Corr}(\eta_t, \varepsilon_t)}$	-0.9584		-0.8938	

Table 1: Estimation results for the fractional UC model in (2), (3), (5), and (6) for log US real GDP via the QML estimator (17) for $p = 1$ and $p = 4$.

Table 1 displays the estimation results for $p = 1$ and $p = 4$ of the fractional UC model in (2), (3), (5), and (6) via the QML estimator in (17), together with the log likelihoods, the model selection criteria, and the estimated correlation between η_t and ε_t. The results show that log GDP is integrated of order around 1.3, and the 95% confidence interval for the $p = 1$ model is $[1.0700, 1.6030]$, indicating that the $I(1)$ hypothesis is likely violated. The estimated integration order $\hat{d} = 1.3365$ implies that a long-run shock on GDP growth has not only a contemporaneous impact, as imposed in the $I(1)$ model, but retains 33.65% of its initial impact after one quarter, 22.49% after two quarters, and 14.61% after one year. It converges to zero at a hyperbolic rate, so that after four years 5.95% and after ten years 3.20% of its initial impact remains. This suggests that long-run shocks have a gradually decreasing impact on GDP growth, which we find intuitive, as e.g. technological innovations are not adapted by the whole economy at a fixed point in time (as an $I(1)$ assumption would suggest), but rather successively, fostering the interpretation of GDP growth as a stationary, mean-reverting fractionally integrated process of order around 0.3.

In line with the $I(1)$ UC literature, we find that log US GDP is well explained by a small number of cyclical lags. While Morley et al. (2003) choose $p = 2$, which is the smallest number of lags they consider as their model is not identified for $p = 1$, see subsection 4.2, we find that $p = 1$ minimizes both, the AIC and the BIC. Additional lags do not improve the overall fit of the model on a large scale, as their coefficients are rather small and insignificant on a 5% level, see table 1. Thus, the $p = 1$ specification seems to be well suited in terms of a minimal representation for log US GDP. Setting $p = 1$ is supported by figure 6, which shows that the estimated cyclical components for the fractional UC model with $p = 1$ and $p = 4$ are almost identical, and that neither the prediction error, nor the estimates for η_t, ε_t display significant autocorrelation for $p = 1$. For the coefficients $\tilde{\phi}(L)$ in the standard lag operator our estimates imply $\hat{\tilde{\phi}}_1 = 1.1249$, $\hat{\tilde{\phi}}_2 = -0.1893$, $\hat{\tilde{\phi}}_3 = -0.0419$, $\hat{\tilde{\phi}}_4 = -0.0174$ and all remaining coefficients are smaller than 0.01 in absolute value and converge to zero rapidly. The sign-change after the first coefficient illustrates that the fractional lag operator is capable of generating an oscillating behavior, for which standard AR models require at least two lags.

Turning to $\hat{Q}$, the QML estimator returns a relatively small variance estimate for the long-run shocks $\hat{\sigma}_\eta^2$, while the short-run shock variance is estimated to be comparably large. This suggests that the long-run component of output is relatively smooth, and a larger $\hat{\sigma}_\varepsilon^2$ yields richer cyclical dynamics, as explained e.g. in Kamber et al. (2018) who explicitly restrict the ratio $\sigma_\varepsilon/\sigma_\eta$ to be large. Furthermore, the strong negative correlation between long- and short-run shocks confirms the findings of Morley et al. (2003).

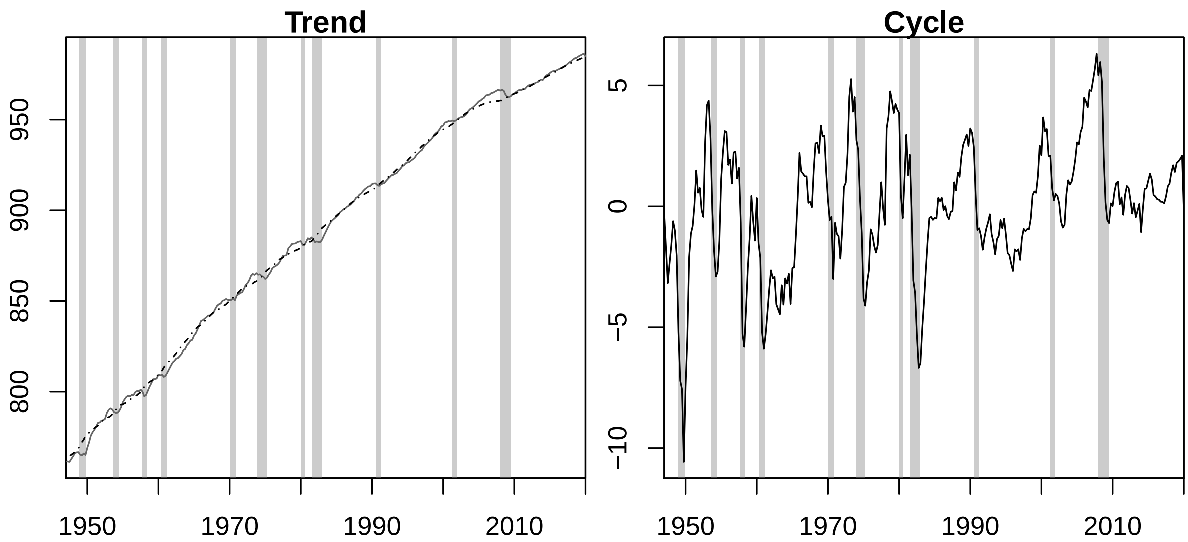

Figure 3: Trend-cycle decomposition for log US real GDP with correlated innovations. The left plot sketches the trend component estimate $\tau_{t|n}$ from the fractional UC model in black, dashed, together with the time series for log US real GDP in gray, solid. The plot on the right-hand side shows the estimated cyclical component $c_{t|n}$ Estimates were carried out via the single-step Kalman smoother as discussed in section 4.1, while parameter estimates are reported in table 1. Shaded areas correspond to NBER recession periods.

Figure 3 plots the trend-cycle decomposition from the fractional UC model, where estimates for x_t, c_t are obtained from the single-step Kalman smoother as discussed in section 4.1 and the parameter estimates $\hat{\theta}$ as given in table 1 for $p = 1$ are plugged in for θ. Due to the small $\hat{\sigma}_\eta^2$ the decomposition yields a smooth trend that is rather unaffected by the recession periods. The estimated cyclical component displays a persistent behavior and has the shape of an asymmetric sinus curve. It gradually increases in periods of economic recovery and prosperity, until it sharply drops during the shaded recession periods, which is in line with economic common sense. Similar cycle estimates are obtained from the nonlinear regime-switching UC-FP-UR model of Morley and Piger (2012). Thus, the parsimonious parametrization of the fractional UC model together with its ability to resemble nonlinear dynamics foster its generality.

From figure 3 it becomes clear that the fractional UC model solves the problem of obtaining implausible cycle estimates in the integer-integrated UC literature, where estimates for x_t behave erratic, while estimates for c_t are rather noisy, see Morley et al. (2003, fig. 3). The underlying reason is that, given log GDP is integrated of order around 1.3, forcing the long-run component to be $I(1)$ upward-biases the estimate $\hat{\sigma}_\eta$, as the additional persistence that is not captured by the $I(1)$ specification goes into the estimates for the long-run innovations η_t. As UC decompositions are fully model-based, the cyclical component needs to re-adjust for the erratic long-run shocks, yielding a noisy cycle that does not follow a clear path in periods of economic upswing. In contrast, the fractional UC model fully captures the persistence of log US GDP, allowing σ_η^2 to be small, which yields a smooth trend estimate together with a cycle that is in line with economic common sense. These findings are consistent with the work of Kamber et al. (2018), who obtain plausible cycle estimates when restricting the ratio of $\sigma_\eta/\sigma_\varepsilon$ to be small.

Finally, we comment on the possibility of structural breaks in log US real GDP. The test results in section 2 already provide comprehensive evidence for GDP's long-run component being a fractionally integrated process, while there is no evidence for the presence of structural breaks. We substantiate these findings by including a trend break after 1973:1 into (26), so that log US real GDP is first adjusted for a constant and a linear trend with a structural break in slope after 1973:1. Estimation is carried out as before, where again 100 combinations of starting values are drawn for the QML estimator in (17) and $p = 1$. Estimation results are contained in table 2 in appendix A and are close to the results reported in table 1 for the fractional UC model. The integration order is again estimated to be around 1.3, a small ratio of $\hat{\sigma}_\eta^2/\hat{\sigma}_\varepsilon^2$ is found, and correlation between long- and short-run shocks is negative, but considerably smaller as in the purely fractional model. Naturally the estimated variance of the long-run shocks decreases compared to table 1, as the inclusion of a trend break after 1973:1 explains some variation of the long-run component. Since the trend break captures some dynamics of the fractional long-run component and is uncorrelated with the short-run shocks, this explains the smaller correlation. However, the log likelihood when $\sigma_\eta^2 = 0$ is imposed falls by 119 points, and the standard error of d remains small, implying that the parameter is well identified via the long-run dynamics of log US real GDP, which would not be the case if $\sigma_\eta^2 = 0$.

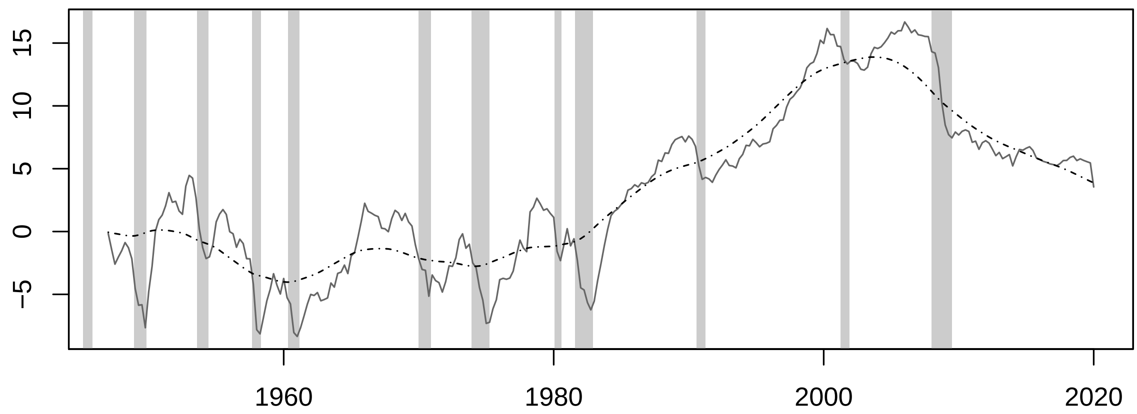

Figure 4: Fractional trend-cycle decomposition with structural breaks. The figure shows the estimate for x_t from the fractional UC model with a linear (deterministic) trend break after 1973:1 in black, dashed, together with trend-adjusted log US real GDP in gray, solid. Estimates were carried out via the single-step Kalman smoother as discussed in section 4.1, while parameter estimates are reported in table 2 in appendix A. Shaded areas correspond to NBER recession periods.

Figure 4 sketches the estimated stochastic long-run component of the fractional UC model with a structural break after 1973:1, showing that the inclusion of a deterministic trend break after 1973:1 reduces the variation of the estimated stochastic long-run component considerably during the first half of our sample and explains the good fit of the decomposition of Perron and Wada (2009), as they consider a sample until 1998:2. However, the approximation becomes worse for the second half of the sample. From the 1981 – 1982 recession on, the estimate for x_t gradually increases despite the inclusion of a trend break. Growth slows down before the early 1990s recession and recovers in the aftermath. Around the change of the millennium the estimated long-run component becomes flat, until it tends to decrease from 2005 on. After the Great Recession another change in growth is visible. Consequently, the fractional long-run component survives the inclusion of a trend break after 1973:1, and the estimated integration order is hardly affected.

From both, the test results in section 2 and the inclusion of the Perron and Wada (2009) trend break, it follows that there is no evidence for GDP being driven by structural breaks in the linear trend, while there is comprehensive evidence for the fractional hypothesis. However, the fractional long-run component may still be well-approximated by a deterministic trend with trend breaks, particularly due to its smooth nature (Diebold and Inoue; 2001; Sibbertsen; 2004). While such an approximation does not allow to draw inference on the persistence of long-run shocks, it bears the advantage of providing a rough approximation to the business cycle via the simple model in spirit of Perron and Wada (2009)

$$\log GDP_t = \mu_0 + \sum_{j=1}^{q} \mathbb{1}(t > t_j^*)\mu_j(t - t_j^*) + c_t, \qquad a(L)c_t = e_t \tag{27}$$

where $\mu_0, ..., \mu_q$ are constants, $\mathbb{1}(t > t_j^*)$ is an indicator function that takes the value 1 if $t > t_j^*$, else zero, t_j^* are the points in time after which trend breaks occur, $t_1^* = 0$ accounts for an overall trend, c_t is the cycle that is modelled as an AR(p) process, and e_t is white noise.

The unknown terms in (27) are estimated in a two-step approach, where we first estimate $\mu_0, ..., \mu_q$ by regressing log GDP on the deterministic terms and next fit an AR model to the residuals. The lag order for c_t is chosen via the AIC. Since the break dates t_j^*, $j = 2, ..., q$ are unknown, all $\binom{n-1}{q-1}$ possible combinations of break dates are estimated in an endogenous break date search, and the combination with the greatest log likelihood is chosen. We consider q up to order 4, which requires to estimate $\binom{292}{3} = 4,106,980$ different specifications. A higher q is thus computationally infeasible, which immediately illustrates the limitations of the approximation in (27). Estimation results for $q = 2, 3, 4$ are contained in table 3 in appendix A. Both AIC and the more conservative BIC clearly suggest $q = 4$, which is the maximum number of trends considered. The estimated AR coefficients are in line with the fractional UC results in table 1.

Figure 5 plots the trend-cycle decompostions from the fractional UC model and the deterministic long-run specification in (27). As can be seen, the estimated long-run components largely overlap, and the obtained measures for the business cycle are similar. Differences occur in the 1960s, where the fractional UC model estimates a gradual recovery from the 1960-1961 recession, while the model with structural breaks implies a rather steep recovery directly after the recession, and before the Great Recession, where the fractional model finds a stronger cyclical upswing. Both deviations are due to the estimated break points $\hat{t}_1^* =$ 1963:3, $\hat{t}_2^* =$ 1966:1, and $\hat{t}_3^* =$ 2005:1, around which the linear approximation becomes imprecise.

While the linear trend break model of Perron and Wada (2009) provides a rough approximation to GDP's long-run component, its limitations directly become apparent from figure 5. While for a closer approximation further break points are required, the maximum number of break points is limited to three, as e.g. four breaks would require to estimate $\binom{292}{4} = 296,729,305$ different combinations of break points, which is computationally infeasible. Consequently, approximating the long-run component via a deterministic process with trend breaks as in (27) gives a rough idea of the business cycle but should only be

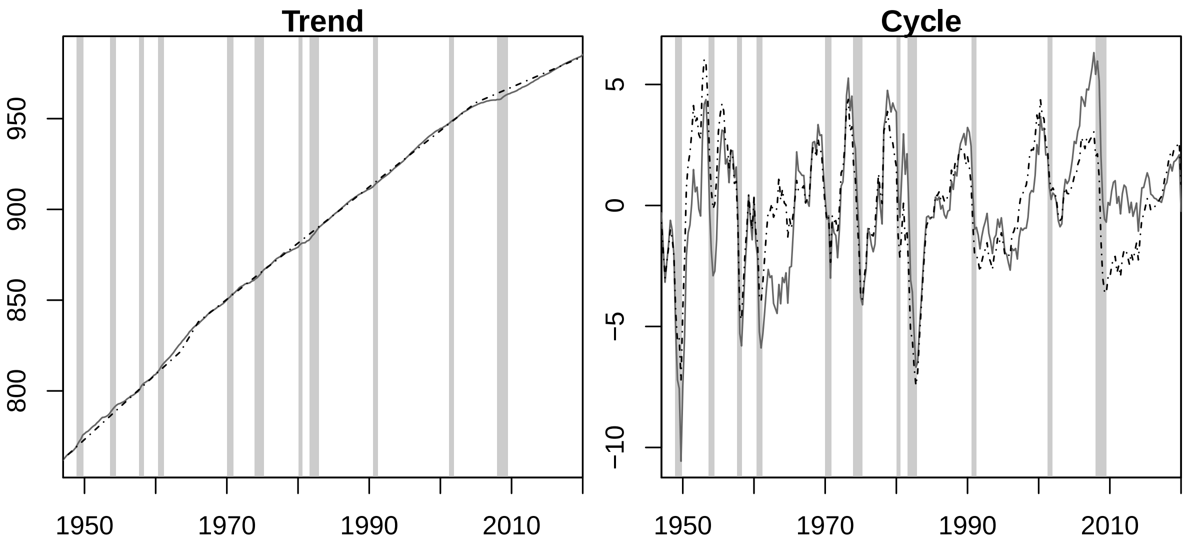

Figure 5: Trend-cycle decompositions from the fractional UC model (gray, solid) and the model with deterministic long-run components and structural breaks in (27) (black, dashed). Parameter estimates are reported in table 1 and table 3. Estimated break points are $\hat{t}_1^* = 1963{:}3$, $\hat{t}_2^* = 1966{:}1$, and $\hat{t}_3^* = 2005{:}1$. Shaded areas correspond to NBER recession periods.

considered as a first step in empirical research, while for a more precise measure one should clearly favor the fractional UC model.

A brief summary of the above results and their implications concludes this section. First, since the fractional UC model encompasses state-of-the-art integer-integrated UC models with stochastic long-run components and autoregressive cycles (for instance Harvey; 1985; Clark; 1987; Balke and Wohar; 2002; Morley et al.; 2003; Kamber et al.; 2018), and since the confidence interval for d does not contain the points $d = 1$ and $d = 2$, integer-integrated UC models are likely misspecified for log US real GDP. Second, a deterministic long-run specification as suggested by Perron and Wada (2009) serves as a good approximation to GDP's long-run component but is limited by the maximum number of trend breaks that are computationally feasible. In addition, no evidence for the presence of structural breaks is found one fractional integration is taken into account. Third, the solution to the unobserved components puzzle for log US real GDP lies in the presence of a yet neglected fractionally integrated long-run component that upward-biases the long-run shock variance of the correlated UC model of Morley et al. (2003), violates the white noise hypothesis for the model of Kamber et al. (2018), and is well approximated by a deterministic trend with structural breaks by Perron and Wada (2009). Once the fractional long-run component is taken into account, the resulting trend-cycle decomposition is well in line with the NBER chronology, no remaining autocorrelation in the estimated long- and short-run shocks is found, and the estimated integration order confirms the nonparametric estimates in section 2. And fourth, the violation of the $I(1)$ assumption for log US real GDP has implications beyond the UC literature. Contrary to the predominant interpretation that long-run shocks have only a contemporaneous impact on GDP growth they should be rather interpreted as having a gradually decreasing effect over time. In addition, the results call into question several models where GDP is treated as an $I(1)$ process.

6 Conclusion

To examine whether the puzzling results in the unobserved components literature were driven by the presence of a latent, fractionally integrated long-run component in log US real GDP, we generalized unobserved components models to the fractionally integrated processes. For the estimation of the model parameters we derived a quasi maximum likelihood estimator that was shown to be consistent and asymptotically normal. Estimators for the latent components that are identical to the Kalman filter and smoother but computationally superior for fractional models were derived.

For log US real GDP, we obtained an estimated integration order of around 1.3, implying that long-run shocks have a gradually decreasing impact on GDP growth over time. The resulting trend-cycle decomposition is well in line with the NBER chronology and explains the contradicting results of Morley et al. (2003), Perron and Wada (2009), and Kamber et al. (2018) well.

The model offers a variety of opportunities for future research. It may be generalized to the multivariate case, where fractional trends of different persistence with correlated innovations are allowed. A multivariate fractional trend-cycle model would then allow to estimate common fractional trends of cointegrated variables and test for polynomial cointegration. Furthermore, inferential methods that test for the number of common trends or the equality of integration orders could be established. As shown in Diebold and Inoue (2001), fractionally integrated processes and structural breaks are related, since the former class of processes can produce level shifts and since structural breaks can be misinterpreted as $I(d)$ processes. Hence, combining both concepts, e.g. in a fractional UC model with regime switching, can be a fruitful challenge for future research.

To applied researchers, the model offers a flexible data-driven method to treat permanent and transitory components in macroeconomic and financial applications. It provides a solution for many issues of model specification that caused uncertainty and debates about realistic trend-cycle decompositions and estimation of recessions. Based on that, also the interaction of trends and cycles can be analyzed.

A Graphs and tables

	estimate	std. error
$\hat{d}$	1.3528	0.0689
$\hat{\sigma}_\eta^2$	0.0753	0.0469
$\hat{\sigma}_{\eta\varepsilon}^2$	-0.0710	0.0234
$\hat{\sigma}_\varepsilon^2$	0.7959	0.0188
$\hat{\phi}_1$	0.8727	0.0368
$l_t(\hat{\theta})$	-370.7755	
$\widehat{\mathrm{Corr}(\eta_t, \varepsilon_t)}$	-0.2899	

Table 2: Robustness check: Estimation results for the fractional UC model in (2), (3), (5), and (6) for log US real GDP via the QML estimator (17) for $p = 1$ with a trend break after 1973:1, as suggested by Perron and Wada (2009).

	$q = 2$		$q = 3$		$q = 4$	
	estimate	std. error	estimate	std. error	estimate	std. error
$\hat{\mu}_0$	765.6046	0.4897	759.3243	0.5039	761.5529	0.5735
$\hat{\mu}_1$	0.8543	0.0036	0.9811	0.0077	0.8974	0.0141
$\hat{\mu}_2$	-0.3962	0.0147	-0.2013	0.0105	0.7404	0.0758
$\hat{\mu}_3$			-0.3554	0.0152	-0.8635	0.0670
$\hat{\mu}_4$					-0.3517	0.0141
$\hat{a}_1$	1.2720	0.0593	1.2191	0.0584	1.1958	0.0583
$\hat{a}_2$	-0.1528	0.0948	-0.1480	0.0927	-0.1348	0.0916
$\hat{a}_3$	-0.2533	0.0948	-0.1721	0.0583	-0.1829	0.0582
$\hat{a}_4$	0.0873	0.0589				
AIC	729.6638		721.8059		715.6252	
BIC	755.4250		747.5671		745.0665	
$\hat{t}_1^*$	2000:3		1969:2		1963:3	
$\hat{t}_2^*$			2004:4		1966:1	
$\hat{t}_3^*$					2005:1	

Table 3: Estimation results for the approximate model with a purely deterministic long-run component and structural breaks in (27).

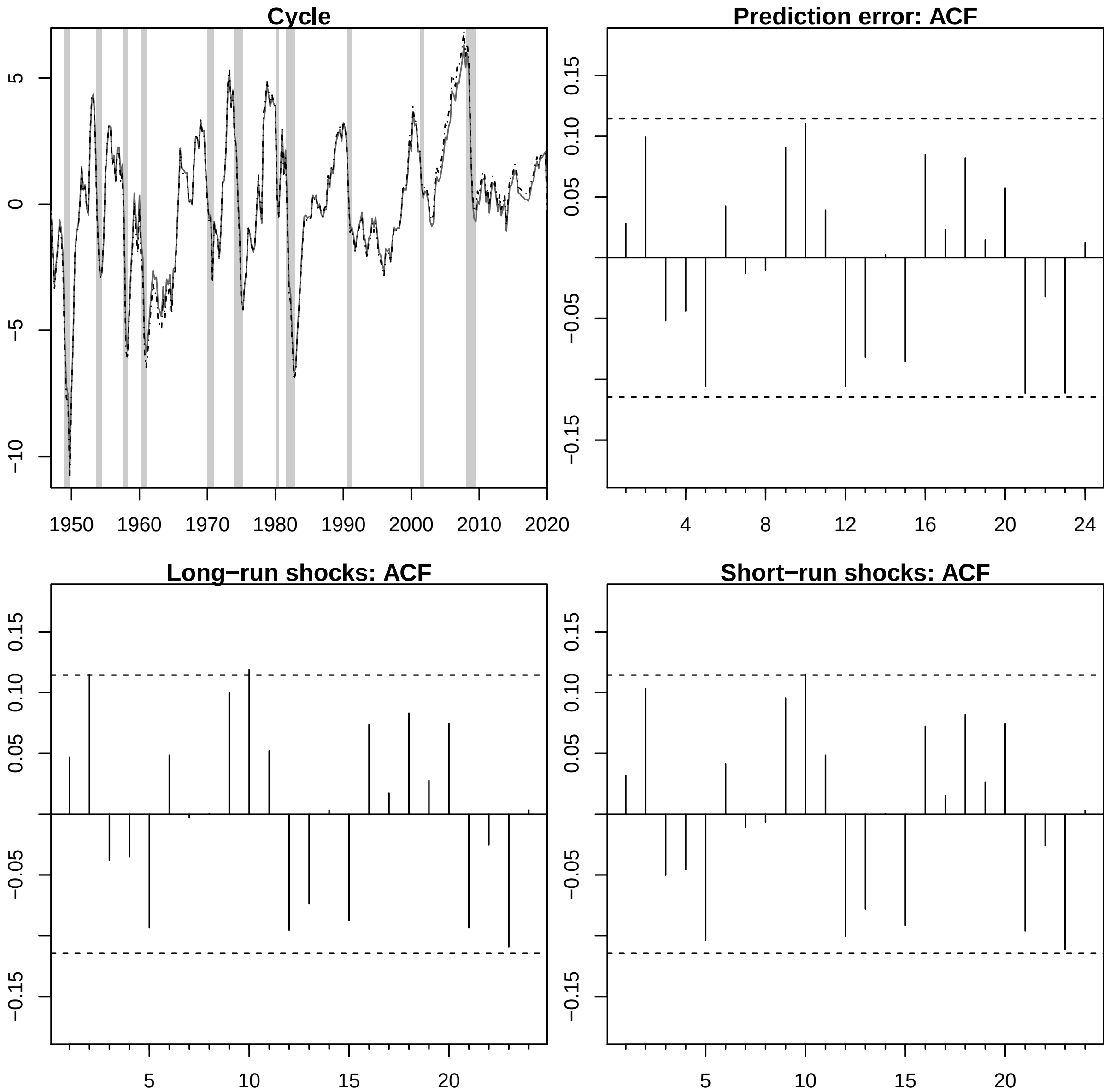

Figure 6: Robustness check: Additional lags and autocorrelation. The upper left plot sketches the estimated cyclical components for the fractional UC models with $p = 1$ (grey, solid) and $p = 4$ (black, dashed). The upper-right plot displays the estimated autocorrelations of the prediction errors for $p = 1$, together with a 95% confidence interval, while the lower-left plot sketches the estimated autocorrelations for $\eta_{t|n}$ from the Kalman smoother and the lower-right plot for $\varepsilon_{t|n}$.

B Further applications

While decomposing log US real GDP into trend and cycle as considered in section 5 is the key application of UC models, the literature has also considered the decomposition of other macroeconomic aggregates, including industrial production (Clark; 1987; Weber; 2011), investment (Harvey and Trimbur; 2003), consumption (Morley; 2007), and employment (Koopman et al.; 2012; Klinger and Weber; 2016). In this appendix, we illustrate the fractional trend-cycle decomposition for the three additional variables. This allows to draw inference on the persistence of long-run shocks and to test the hypotheses about d in the UC literature for the three variables of interest. Furthermore, we shed light on the cyclical dynamics and the economic plausibility of the resulting decomposition.

Our data consists of log US industrial production index (mnemonic: INDPRO), log US real gross private domestic investment (mnemonic: GPDIC1), log real personal consumption expenditures (mnemonic: PCECC96), and log total nonfarm employees (mnemonic: PAYEMS) and was downloaded from the St. Louis FED. It is in quarterly frequency, spans from 1947:1 to 2020:1, and is seasonally adjusted. Deterministic terms have been eliminated from the data via (26), estimation of the model parameters θ was carried out as in section 5, and the lag order of the cyclical polynomial p was chosen via the AIC.

	Ind. Production		Investment		Consumption		Employees	
	estimate	std. err.	estimate	std. err.	estimate	std. err.	estimate	std. err.
$\hat{d}$	1.3799	0.1273	1.3208	0.0749	0.8791	0.0654	1.2392	0.0231
$\hat{\sigma}_\eta^2$	0.3014	0.3125	0.3785	0.4373	4.2553	1.5741	2.4026	0.1350
$\hat{\sigma}_{\eta\varepsilon}$	-0.9448	0.8044	-1.6685	1.4466	-5.5326	1.8024	-2.5074	0.1598
$\hat{\sigma}_\varepsilon^2$	3.9189	1.3516	23.8628	3.2176	7.3518	2.0192	2.7522	0.2472
$\hat{\phi}_1$	0.9892	0.0785	0.8036	0.0454	1.0086	0.0245	1.9843	0.0453
$\hat{\phi}_2$	-0.2134	0.0696			0.0797	0.0309	-1.4436	0.0674
$\hat{\phi}_3$	0.0921	0.0331			-0.1600	0.0407	0.6302	0.0414
$\hat{\phi}_4$	-0.0701	0.0274					-0.2299	0.0125
$\hat{\phi}_5$							-0.0968	0.0147
$\hat{\phi}_6$							0.2051	0.0061
$\hat{\phi}_7$							-0.0802	0.0041
$l_t(\hat{\theta})$	-539.3758		-854.5616		-331.5174		-120.9329	
AIC	1082.2300		1665.7794		667.5800		263.9920	
BIC	1111.6714		1684.1803		693.3412		304.4739	
$\hat{\rho}$	-0.8694		-0.5519		-0.9891		-0.9751	

Table 4: Estimation results for the fractional UC model in (2), (3), (5), and (6) for log US industrial production, log US real gross private domestic investment, log real personal consumption expenditures, and log all employees (total, nonfarm) via the QML estimator (17). p was chosen via the AIC. $\hat{\rho} = \widehat{\mathrm{Corr}(\eta_t, \varepsilon_t)}$.

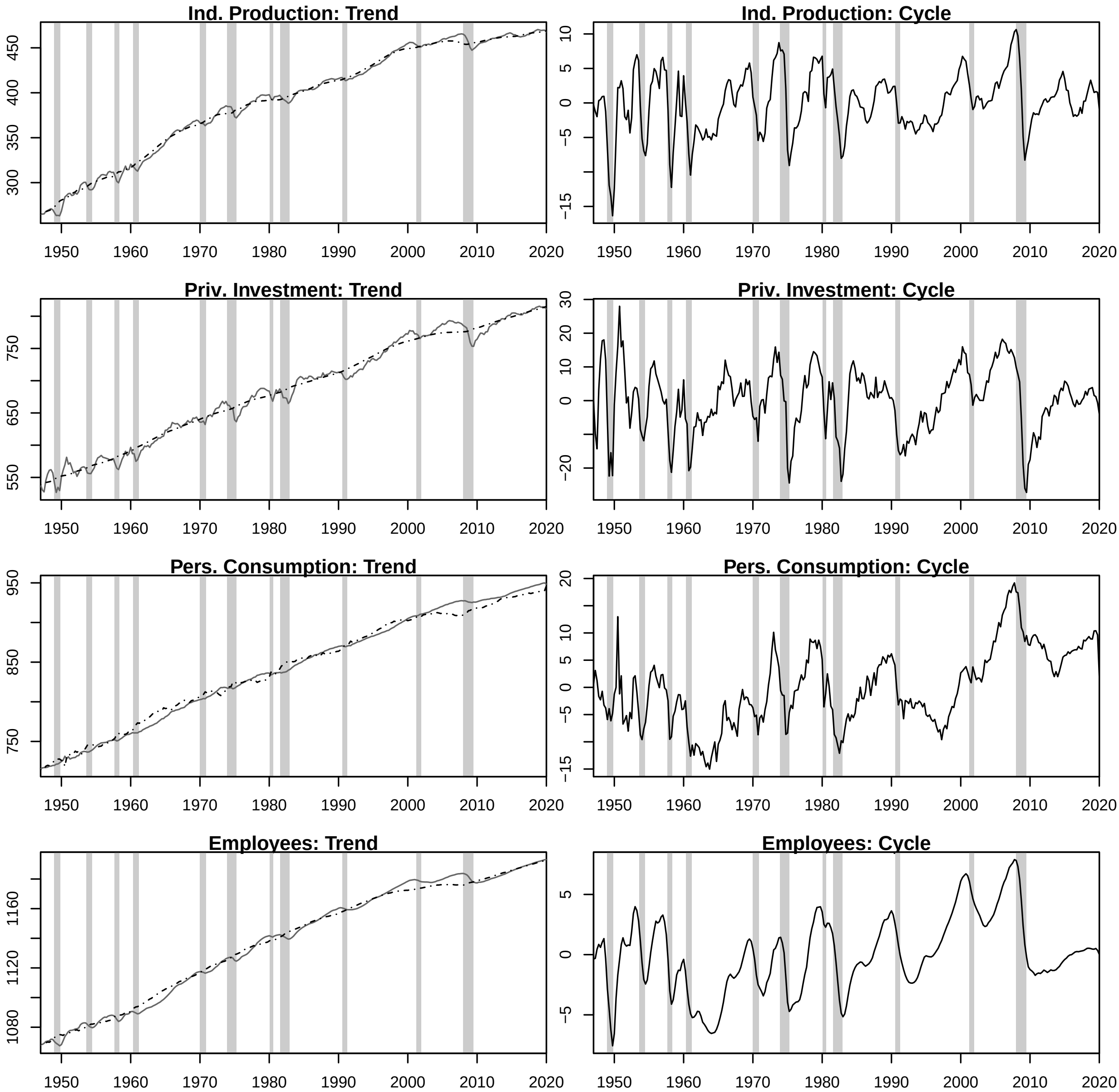

Figure 7: Trend-cycle decomposition for log US industrial production, log US real gross private domestic investment, log real personal consumption expenditures, and log all employees (total, nonfarm) with correlated innovations. The left plots sketch the trend component estimates $\tau_{t|n}$ from the fractional UC model in black, dashed, together with the observable time series in gray, solid. The plots on the right-hand side show the estimated cyclical components $c_{t|n}$ Estimates were carried out via the single-step Kalman smoother as discussed in section 4.1, while parameter estimates are reported in table 4. Shaded areas correspond to NBER recession periods.

Estimation results are contained in table 4. While for industrial production, private investment and total employees a 95% confidence interval for d does not contain the random walk case with $d = 1$, the 95% confidence interval for private consumption is

$[0.7509, 1.0074]$, so that $d = 1$ lies on the boarder of the interval. Consequently, integer-integrated UC models are likely to be misspecified for the former three series, while for consumption we cannot show a violation of the $I(1)$ hypothesis on a 95% level of significance.

Figure 7 plots the three trend-cycle decompositions for the variables of interest and illustrates that all decompositions are well in line with the NBER chronology. For industrial production we obtain a decomposition hitting NBER recession periods, while $I(1)$ UC models frequently produce implausible cycles for industrial production that increase during recessions, particularly when p is chosen to be small (Weber; 2011). The decomposition for private investment finds a smooth trend component together with strong cyclical variation, which especially becomes apparent during the Great Recession, where a strong cyclical upswing before the recession is followed by a pronounced slump of the cycle. Private consumption is found to exhibit less persistent long-run shocks compared to the other series, together with a strong cyclical component. Finally, long- and short-run shocks to total employees exhibit similar estimates for their variances. While the model finds a reduction of long-run growth from the 21st century on, it attributes job creation before the dotcom bust and the Great Recession to the transitory component.

C Mathematical appendix

C.1 Univariate moving average representation of aggregated model

We consider the aggregation of two moving average (MA) processes in the lag operator L_d with generic lag polynomials $h(L_d)$ and $\tilde{h}(L_d)$ of order q and $\tilde{q}$, respectively,

$$z_t = h(L_d)\eta_t + \tilde{h}(L_d)\varepsilon_t, \tag{28}$$

with the white noise processes

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

In what follows, set $p = \max(q, \tilde{q})$ and let $h_i = 0$ for all $i > q$, $\tilde{h}_i = 0$ for all $i > \tilde{q}$. We first derive the MA representation in the standard lag operator $L = L_1$. Next we derive the MA representation in the fractional lag operator L_d which is in general not of finite order. To rewrite (28) in the conventional lag operator L define

$$L_d^k = (1 - \Delta^d)^k = \sum_{i=k}^{\infty} \varsigma_{k,i}(d) L^i,$$

insert it into (28), and rearrange terms

$$\begin{aligned} z_t &= \eta_t + \varepsilon_t + \sum_{k=1}^{p} \left(h_k \sum_{i=k}^{\infty} \varsigma_{k,i}(d)\eta_{t-i} + \tilde{h}_k \sum_{i=k}^{\infty} \varsigma_{k,i}(d)\varepsilon_{t-i} \right) \\ &= \eta_t + \varepsilon_t + \sum_{k=1}^{p} \sum_{i=k}^{\infty} \varsigma_{k,i}(d) \left(h_k \eta_{t-i} + \tilde{h}_k \varepsilon_{t-i} \right). \end{aligned}$$

Redefining the sum indexes we obtain

$$z_t = \eta_t + \varepsilon_t + \sum_{l=1}^{\infty} \eta_{t-l} \sum_{k=1}^{l} \varsigma_{k,l}(d) h_k + \sum_{l=1}^{\infty} \varepsilon_{t-l} \sum_{k=1}^{l} \varsigma_{k,l}(d) \tilde{h}_k \tag{29}$$

$$= \sum_{l=0}^{\infty} g_l \, \eta_{t-l} + \sum_{l=0}^{\infty} \tilde{g}_l \, \varepsilon_{t-l}, \tag{30}$$

with $g_0 = \tilde{g}_0 = 1$ and $g_l = \sum_{k=1}^{l} \varsigma_{k,l}(d) h_k$ and $\tilde{g}_l = \sum_{k=1}^{l} \varsigma_{k,l}(d) \tilde{h}_k$, $l = 1, 2, \ldots, \infty$. Note that both moving average processes are of order ∞. If (30) can be aggregated, there exists a univariate moving average process

$$z_t = c(L)u_t, \quad u_t \sim \text{i.i.d.}(0, \sigma_u^2).$$

To compute the coefficients c_i, note that $\text{Cov}(z_t, c_l u_{t-l}) = \text{Cov}(z_t, g_l \eta_{t-l} + \tilde{g}_l \varepsilon_{t-l})$, which gives

$$c_l^2 \sigma_u^2 = g_l^2 \sigma_\eta^2 + \tilde{g}_l^2 \sigma_\varepsilon^2 + 2 g_l \tilde{g}_l \sigma_{\eta\varepsilon}, \quad l = 0, 1, \dots. \tag{31}$$

To make the dependence of c_l^2 on the parameters of the fractional moving average polynomials explicit insert g_l and $\tilde{g}_l$ into (31). This delivers for $l \geq 1$

$$\begin{aligned} c_l^2 \sigma_u^2 &= \left(\sum_{k=1}^{l} \varsigma_{k,l}(d) h_k\right)^2 \sigma_\eta^2 + \left(\sum_{k=1}^{l} \varsigma_{k,l}(d) \tilde{h}_k\right)^2 \sigma_\varepsilon^2 + 2 \left(\sum_{k=1}^{l} \varsigma_{k,l}(d) h_k\right) \left(\sum_{k=1}^{l} \varsigma_{k,l}(d) \tilde{h}_k\right) \sigma_{\eta\varepsilon} \\ &= \sum_{k=1}^{l} \sum_{i=1}^{l} \varsigma_{k,l}(d) \varsigma_{i,l}(d) \left(h_k h_i \sigma_\eta^2 + \tilde{h}_k \tilde{h}_i \sigma_\varepsilon^2 + 2 \sigma_{\eta\varepsilon} h_k \tilde{h}_i\right), \end{aligned} \tag{32}$$

with $c_0 = 1$, $\sigma_u^2 = \sigma_\eta^2 + \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon}$. Solving for c_l yields the MA coefficients for u_t.
Next we derive the univariate moving average representation in the fractional lag operator which is typically of infinite order

$$z_t = \psi(L_d) u_t. \tag{33}$$

If (33) exists, then it can be rewritten similarly to (29) in the standard lag operator as

$$z_t = u_t + \sum_{l=1}^{\infty} u_{t-l} \left(\sum_{k=1}^{l} \varsigma_{k,l}(d) \psi_k\right).$$

For such a representation to exist, there must exist parameters ψ_i, $i = 1, \dots, q_u$ such that

$$c_l = \sum_{k=1}^{l} \varsigma_{k,l}(d) \psi_k, \quad l = 1, 2, \dots,$$

while (31) holds. Solving for ψ_l delivers

$$\psi_l = \frac{c_l - \sum_{k=1}^{l-1} \varsigma_{k,l}(d) \psi_k}{\varsigma_{l,l}(d)}. \tag{34}$$

Obviously, the order of the moving average polynomial in the fractional lag operator would only be of finite order q_u if

$$c_l = \sum_{k=1}^{l-1} \varsigma_{k,l}(d)\psi_k, \quad l > q_u. \tag{35}$$

In general this is not the case. In order to represent the ψ_l, $l = 1, ..., q_u$, in terms of the parameters h_j, $j = 1, ..., q$, and $\tilde{h}_k$, $k = 1, ..., \tilde{q}$, of the moving average polynomials in L_d, one inserts (32) into (34) and obtains

$$\psi_l = \frac{\sqrt{\sum_{k=1}^{l} \sum_{i=1}^{l} \varsigma_{k,l}(d)\varsigma_{i,l}(d) \left(h_k h_i \sigma_\eta^2 + \tilde{h}_k \tilde{h}_i \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon} h_k \tilde{h}_i \right)}/\sigma_u - \sum_{k=1}^{l-1} \varsigma_{k,l}(d)\psi_k}{\varsigma_{l,l}(d)}. \tag{36}$$

Since only $\psi_1, ..., \psi_{l-1}$ enter (36), ψ_l can be calculated recursively, where the first coefficient is $\psi_1 = \sigma_u^{-1}\sqrt{h_1^2\sigma_\eta^2 + \tilde{h}_1^2\sigma_\varepsilon^2 + 2h_1\tilde{h}_1\sigma_{\eta\varepsilon}}$ and $\sigma_u = \sqrt{\sigma_\eta^2 + \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon}}$.

A special case occurs when one of the two MA processes in (28) is purely white noise, e.g. $\tilde{h}(L_d) = 1$. Then the square root in (36) becomes $\sqrt{\sum_{k,i=1}^{l} \varsigma_{k,l}(d)\varsigma_{i,l}(d) h_k h_i \sigma_\eta^2} = \sqrt{\left(\sum_{k=1}^{l} \varsigma_{k,l}(d) h_k \sigma_\eta\right)^2} = \sum_{k=1}^{l} \varsigma_{k,l}(d) h_k \sigma_\eta$, and the recursion in (36) yields $\psi_l = h_l(\sigma_\eta/\sigma_u)$.

C.2 Asymptotic properties of the maximum likelihood estimator

Proof of theorem 4.1. To prove consistency of the quasi maximum likelihood (QML) estimator for $\theta = (d, \phi_1, ..., \phi_p, \sigma^2_\eta, \sigma_{\eta\varepsilon}, \sigma^2_\varepsilon)$ of model (2), (3), (5), and (6) as given in (25)

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} l_t(\theta), \qquad l_t(\theta) = -\frac{1}{2} \log \sigma^2_u - \frac{1}{2\sigma^2_u} u^2_t(\theta),$$

we proceed as follows. For the parameter space of $d \in D = [d_{min}, d_{max}]$ we denote $D^*(\kappa) = D \cap \{d : d_0 - d \leq 1/2 - \kappa\}$, $0 < \kappa < 1/2$ as the region where $u_t(\theta)$ is stationary. In a first step, we show that our model is nested in the class of ARFIMA processes considered by Nielsen (2015), for which it is proven that given any constant $K > 0$, there exists a fixed $\bar{\kappa} > 0$, such that

$$Pr\left(\inf_{d \in D \backslash D^*(\bar{\kappa}) \cap \theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} u^2_t(\theta) > K\right) \to 1 \quad \text{as } n \to \infty, \tag{37}$$

which implies $Pr(\hat{d} \in D^*(\bar{\kappa}) \cap \theta \in \Theta) \to 1$ as $n \to \infty$. Thus, the relevant parameter space asymptotically reduces to the stationary region $\Theta^*(\bar{\kappa}) = \{\theta | \theta \in \Theta, d \in D^*(\bar{\kappa})\}$.

From the results of Nielsen (2015, eq. 8) it then follows that within the stationary region the sum of squared residuals satisfies a weak law of large numbers (WLLN), while it diverges in probability for $d_0 - d \geq 1/2$

$$\text{plim}_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} u^2_t(\theta) = \begin{cases} \mathrm{E}(\tilde{u}^2_t(\theta)) & \text{if } d_0 - d < 1/2, \\ \infty & \text{otherwise}, \end{cases} \tag{38}$$

where $\tilde{u}_t(\theta) = \tilde{\psi}(L)^{-1} \tilde{\phi}(L) \Delta^d y_t$ is the untruncated residual generated by the untruncated polynomials $\tilde{\phi}(L) = 1 - \sum_{j=1}^{\infty} \tilde{\phi}_j L^j = 1 - \sum_{j=1}^{p} \phi_j L^j_d$, $\Delta^d = \sum_{j=0}^{\infty} \pi_j(d) L^j$, and $\tilde{\psi}(L) = 1 + \sum_{j=1}^{\infty} \tilde{\psi}_j L^j$ with coefficients as defined in (21) and (40).

The second and more delicate step is then to show that the objective function satisfies a uniform weak law of large numbers (UWLLN), i.e. there exists a function $c_t(\theta) \geq 0$ such that for all $\theta_1, \theta_2 \in \Theta^*(\kappa)$ it holds that $|l_t(\theta_1) - l_t(\theta_2)| \leq c_t(\theta) ||\theta_1 - \theta_2||$, and $c_t(\theta)$ satisfies a WLLN (Wooldridge; 1994, th. 4.2). Since $l_t(\theta)$ is continuously differentiable, a natural choice for $c_t(\theta)$ that satisfies the above inequality is the supremum of the gradient, which follows from the mean value expansion of $l_t(\theta)$ about θ (Wooldridge; 1994, eq. 4.4). Thus, it remains to be shown that the supremum of the gradient satisfies a WLLN, i.e.

$$\sup_{\theta \in \Theta^*(\kappa)} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} l_t(\theta) \right| = O_p(1), \tag{39}$$

for any fixed $\kappa \in (0, 1/2)$. From (39) the UWLLN for the objective function follows, see Newey (1991, cor. 2.2) and Wooldridge (1994, th. 4.2). Since the model is identified, as shown in section 4.2, consistency of the QML estimator then follows directly, see Wooldridge (1994, th. 4.3).

Conditions (37) and (38) were shown to hold by Nielsen (2015, eq. 8 and 13) for a general class of ARFIMA processes where the disturbances are conditionally homoskedastic martingale difference sequences with finite moments up to order four, the parameter space is convex and compact, the AR and MA polynomials are stationary and invertible, and the model is identified. Since the class of ARFIMA processes considered there nests the fractional UC model, it follows directly that (37) and (38) are satisfied for our model. Nonetheless, consistency of the QML estimator does not follow directly from the proofs of Nielsen (2015), since first, a different estimator is considered there, and second, our model imposes restrictions on the parameters of the model considered by Nielsen (2015). Consequently, (39) needs to be shown to hold for consistency.

To verify (39) a convenient representation of the MA polynomial $\psi(L_d) = \tilde{\psi}(L) = 1 + \sum_{j=1}^{\infty} \tilde{\psi}_j L^j$ in the standard lag operator L, that links the coefficients $\tilde{\psi}_j$ to the parameters in θ, is required. For the AR polynomial $\phi(L_d)$ a representation in the standard lag operator L is already given in (21). For the MA polynomial it follows from (18) that $\tilde{\phi}(L)\eta_t + (1 - L_d)\varepsilon_t = (1 - \sum_{l=1}^{\infty} \tilde{\phi}_l L^l)\eta_t + \sum_{l=0}^{\infty} \pi_l(d) L^l \varepsilon_t = (1 + \sum_{l=1}^{\infty} \psi_l L_d^l) u_t = (1 + \sum_{l=1}^{\infty} \tilde{\psi}_l L^l) u_t$, so that by matching the autocovariance functions of the white noise processes $\tilde{\phi}(L)\eta_t + \Delta^d \varepsilon_t$ and $\tilde{\psi}(L) u_t$ one obtains

$$\tilde{\psi}(L) = 1 + \sum_{l=1}^{\infty} \tilde{\psi}_l L^l, \qquad \tilde{\psi}_l = \frac{1}{\sigma_u} \sqrt{\tilde{\phi}_l^2 \sigma_\eta^2 + \pi_l^2(d) \sigma_\varepsilon^2 - 2\tilde{\phi}_l \pi_l(d) \sigma_{\eta\varepsilon}}, \tag{40}$$

for all $l > 0$, and $\sigma_u^2 = \sigma_\eta^2 + \sigma_\varepsilon^2 + 2\sigma_{\eta\varepsilon}$.

Using (21) and (40), the objective function in (25) can be written as

$$l_t(\theta) = -\frac{1}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} \left(\tilde{\psi}_+(L)^{-1} \tilde{\phi}_+(L) \Delta_+^d y_t \right)^2, \tag{41}$$

from which the partial derivatives are of interest to verify (39).

As will become clear, the crucial part in establishing (39) is to show that

$$\sup_{\theta \in \Theta^*(\kappa)} \frac{1}{n} \sum_{t=1}^{n} u_t(\theta) \frac{\partial u_t(\theta)}{\partial \theta_i} = O_p(1),$$

where θ_i is the i-th entry in θ. Here, it will be helpful to note that for a white noise process u_t, MA weights $\sum_{j=0}^{\infty} |m_{i,j}(\theta)| < \infty$, $i = 1, 2$, and the set $\tilde{\Theta} = \{\theta | \theta \in \Theta, d - d_0 > -1/2\}$, the product moments satisfy

$$\sup_{\theta\in\tilde{\Theta}}\left|n^{-1}\sum_{t=1}^{n}\left[\frac{\partial^k\Delta_+^{d-d_0}}{\partial d^k}\sum_{j=0}^{\infty}m_{1,j}(\theta)u_{t-j}\right]\left[\frac{\partial^l\Delta_+^{d-d_0}}{\partial d^l}\sum_{j=0}^{\infty}m_{2,j}(\theta)u_{t-j}\right]\right| = O_p(1), \tag{42}$$

for $k,l\geq 0$ as shown in Nielsen (2015, lemma B.3). Since $u_t(\theta)$ as defined in (23) satisfies the absolute summability condition of its coefficients for (42) when $\theta\in\Theta^*(\kappa)$ since $\tilde{\psi}_+(L)^{-1}$, $\tilde{\phi}_+(L)$ are stable, it remains to be shown that absolute summability holds for the partial derivatives.

Starting with the partial derivatives w.r.t. the variance parameters, one has

$$\begin{aligned}
\frac{\partial l_t(\theta)}{\partial\sigma_i^2} &= -\frac{1}{2\sigma_u^2}+\frac{1}{2\sigma_u^4}u_t^2(\theta)+\frac{1}{\sigma_u^2}u_t(\theta)\tilde{\psi}_+(L)^{-2}\frac{\partial\tilde{\psi}_+(L)}{\partial\sigma_i^2}\tilde{\phi}_+(L)\Delta_+^d y_t\\
&= -\frac{1}{2\sigma_u^2}+\frac{1}{2\sigma_u^4}u_t^2(\theta)+\frac{1}{\sigma_u^2}u_t(\theta)\tilde{\psi}_+(L)^{-1}\frac{\partial\tilde{\psi}_+(L)}{\partial\sigma_i^2}u_t(\theta), &(43)\\
\frac{\partial l_t(\theta)}{\partial\sigma_{\eta\varepsilon}} &= -\frac{1}{\sigma_u^2}+\frac{1}{\sigma_u^4}u_t^2(\theta)+\frac{1}{\sigma_u^2}u_t(\theta)\tilde{\psi}_+(L)^{-1}\frac{\partial\tilde{\psi}_+(L)}{\partial\sigma_{\eta\varepsilon}}u_t(\theta), &(44)
\end{aligned}$$

where $i=\eta,\varepsilon$, and $u_t^2(\theta)=O_p(1)$ for $\theta\in\Theta^*(\kappa)$ due to the stationary nature of $u_t(\theta)$. For (43) and (44) to satisfy (42) it is sufficient to show that the coefficients in $\partial\tilde{\psi}(L)/\partial\theta_i=\sum_{j=1}^{\infty}(\partial\tilde{\psi}_jL^j)/\partial\theta_i$ are absolutely summable for all $\theta_i=\sigma_\eta^2,\sigma_\varepsilon^2,\sigma_{\eta\varepsilon}$, since the absolute infinite sum of partial derivatives dominates the absolute finite sum of partial derivatives of the truncated polynomial $\tilde{\psi}_+(L)$. The partial derivatives of the MA coefficients $\tilde{\psi}_j$, $j=1,...,\infty$, in (40) are

$$\begin{aligned}
\frac{\partial\tilde{\psi}_j}{\partial\sigma_\eta^2} &= \frac{-1}{2\sigma_u^3}\sqrt{\tilde{\phi}_j^2\sigma_\eta^2+\pi_j^2(d)\sigma_\varepsilon^2-2\tilde{\phi}_j\pi_j(d)\sigma_{\eta\varepsilon}}+\frac{\tilde{\phi}_j^2}{2\sigma_u}\left(\tilde{\phi}_j^2\sigma_\eta^2+\pi_j^2(d)\sigma_\varepsilon^2-2\tilde{\phi}_j\pi_j(d)\sigma_{\eta\varepsilon}\right)^{-1/2}\\
&= -\frac{\tilde{\psi}_j}{2\sigma_u^2}+\frac{\tilde{\phi}_j^2}{2\tilde{\psi}_j\sigma_u^2}, &(45)\\
\frac{\partial\tilde{\psi}_j}{\partial\sigma_\varepsilon^2} &= \frac{-1}{2\sigma_u^3}\sqrt{\tilde{\phi}_j^2\sigma_\eta^2+\pi_j^2(d)\sigma_\varepsilon^2-2\tilde{\phi}_j\pi_j(d)\sigma_{\eta\varepsilon}}+\frac{\pi_j^2(d)}{2\sigma_u}\left(\tilde{\phi}_j^2\sigma_\eta^2+\pi_j^2(d)\sigma_\varepsilon^2-2\tilde{\phi}_j\pi_j(d)\sigma_{\eta\varepsilon}\right)^{-1/2}\\
&= -\frac{\tilde{\psi}_j}{2\sigma_u^2}+\frac{\pi_j^2(d)}{2\tilde{\psi}_j\sigma_u^2}, &(46)\\
\frac{\partial\tilde{\psi}_j}{\partial\sigma_{\eta\varepsilon}} &= \frac{-1}{\sigma_u^3}\sqrt{\tilde{\phi}_j^2\sigma_\eta^2+\pi_j^2(d)\sigma_\varepsilon^2-2\tilde{\phi}_j\pi_j(d)\sigma_{\eta\varepsilon}}-\frac{\tilde{\phi}_j\pi_j(d)}{\sigma_u}\left(\tilde{\phi}_j^2\sigma_\eta^2+\pi_j^2(d)\sigma_\varepsilon^2-2\tilde{\phi}_j\pi_j(d)\sigma_{\eta\varepsilon}\right)^{-1/2}\\
&= -\frac{\tilde{\psi}_j}{\sigma_u^2}-\frac{\tilde{\phi}_j\pi_j(d)}{\tilde{\psi}_j\sigma_u^2}. &(47)
\end{aligned}$$

This yields the partial sums

$$\frac{\partial \tilde{\psi}(L)}{\partial \sigma_\eta^2} = \frac{1}{2\sigma_u^2}\sum_{j=1}^{\infty}\left(-\tilde{\psi}_j + \frac{\tilde{\phi}_j^2}{\tilde{\psi}_j}\right)L^j = \frac{1}{2\sigma_u^2}\left[-\tilde{\psi}(L) + 1 + \sum_{j=1}^{\infty}\frac{\tilde{\phi}_j^2}{\tilde{\psi}_j}L^j\right],$$

$$\frac{\partial \tilde{\psi}(L)}{\partial \sigma_\varepsilon^2} = \frac{1}{2\sigma_u^2}\sum_{j=1}^{\infty}\left(-\tilde{\psi}_j + \frac{\pi_j^2(d)}{\tilde{\psi}_j}\right)L^j = \frac{1}{2\sigma_u^2}\left[-\tilde{\psi}(L) + 1 + \sum_{j=1}^{\infty}\frac{\pi_j^2(d)}{\tilde{\psi}_j}L^j\right],$$

$$\frac{\partial \tilde{\psi}(L)}{\partial \sigma_{\eta\varepsilon}} = \frac{1}{\sigma_u^2}\sum_{j=1}^{\infty}\left(-\tilde{\psi}_j - \frac{\tilde{\phi}_j\pi_j(d)}{\tilde{\psi}_j}\right)L^j = \frac{1}{\sigma_u^2}\left[-\tilde{\psi}(L) + 1 - \sum_{j=1}^{\infty}\frac{\tilde{\phi}_j\pi_j(d)}{\tilde{\psi}_j}L^j\right],$$

so that (43) and (44) are

$$\frac{\partial l_t(\theta)}{\partial \sigma_\eta^2} = -\frac{1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}u_t(\theta)\tilde{\psi}_+(L)^{-1}\left(1 + \sum_{j=1}^{\infty}\frac{\tilde{\phi}_j^2}{\tilde{\psi}_j}L^j\right)_+ u_t(\theta), \tag{48}$$

$$\frac{\partial l_t(\theta)}{\partial \sigma_\varepsilon^2} = -\frac{1}{2\sigma_u^2} + \frac{1}{2\sigma_u^4}u_t(\theta)\tilde{\psi}_+(L)^{-1}\left(1 + \sum_{j=1}^{\infty}\frac{\pi_j^2(d)}{\tilde{\psi}_j}L^j\right)_+ u_t(\theta), \tag{49}$$

$$\frac{\partial l_t(\theta)}{\partial \sigma_{\eta\varepsilon}} = -\frac{1}{\sigma_u^2} + \frac{1}{\sigma_u^4}u_t(\theta)\tilde{\psi}_+(L)^{-1}\left(1 - \sum_{j=1}^{\infty}\frac{\tilde{\phi}_j\pi_j(d)}{\tilde{\psi}_j}L^j\right)_+ u_t(\theta). \tag{50}$$

Since $\tilde{\psi}_+(L)^{-1}$ is a stable polynomial, for (48), (49), and (50) to satisfy (42) it is sufficient to show that the coefficients in parentheses are summable in absolute value. Since the absolute sum of truncated coefficients is bounded by the absolute sum of untruncated coefficients, we only prove the latter term to be finite. Here it will be useful to note that $\pi_j(d) \sim j^{-d-1}/\Gamma(-d)$ as $j \to \infty$ (Hassler; 2018, eq. 5.25), so that $\pi_j(d) = O(j^{-d-1})$ and $\tilde{\phi}_j = O(j^{-d-1})$ which can be seen from (21) since $\sum_{k=j}^{p}\phi_k\binom{k}{j} = O(1)$ for all finite p and the finite sum $\sum_{l=1}^{p}\pi_j(dl)$ is asymptotically dominated by $\pi_j(d) = O(j^{-d-1})$. Finally, $\tilde{\psi}_j^{-1} = O(j^{d+1})$ which follows directly from plugging the above results into the inverse of $\tilde{\psi}_j$ in (40). This implies that the untruncated polynomials $\left(1 + \sum_{j=1}^{\infty}\tilde{\phi}_j^2/\tilde{\psi}_jL^j\right)$, $\left(1 + \sum_{j=1}^{\infty}\pi_j^2(d)/\tilde{\psi}_jL^j\right)$, $\left(1 - \sum_{j=1}^{\infty}\tilde{\phi}_j\pi_j(d)/\tilde{\psi}_jL^j\right)$ are absolutely summable, as $1 + \sum_{j=1}^{\infty}O(j^{-d-1}) < \infty$. Consequently, the coefficients of the three polynomials in parentheses in (48), (49), and (50) are absolutely summable, and thus the partial derivatives of $u_t(\theta)$ satisfy the absolute summability condition for (42) when $\theta \in \Theta^*(\kappa)$. From this it follows directly that (42) holds for the partial derivatives in (48), (49), and (50). Hence, a WLLN follows

$$\sup_{\theta\in\Theta^*(\kappa)}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta_i}l_t(\theta)\right| = O_p(1), \qquad \theta_i = \sigma_\eta^2, \sigma_\varepsilon^2, \sigma_{\eta\varepsilon}. \tag{51}$$

Next, we consider the partial derivatives of the objective function (41) w.r.t. the AR coefficients $\phi_1, ..., \phi_p$. Since $\tilde{\phi}_+(L) = \phi_+(L_d) = \left(1 - \sum_{j=1}^p \phi_j L_d^j\right)_+$, the partial derivative w.r.t. ϕ_j is

$$\frac{\partial l_t(\theta)}{\partial \phi_j} = \frac{1}{\sigma_u^2} u_t(\theta) \left[\tilde{\psi}_+(L)^{-1} L_{d_+}^j \Delta_+^d y_t + \phi_+(L_d) \tilde{\psi}_+(L)^{-2} \frac{\partial \tilde{\psi}_+(L)}{\partial \phi_j} \Delta_+^d y_t\right]. \tag{52}$$

The first term $\tilde{\psi}_+(L)^{-1} L_{d_+}^j \Delta_+^d y_t$ satisfies absolute summability as required for (42) when $\theta \in \Theta^*(\kappa)$, since $\tilde{\psi}_+(L)^{-1}$ is the inverse of a stable MA polynomial, L_d^j does not affect the memory of a process, and $\Delta_+^d y_t$ is stationary for $d \in D^*(\kappa)$. For the derivative of $\tilde{\psi}_+(L)$ in the second term it holds that the absolute sum of the coefficients in $\partial \tilde{\psi}(L)/\partial \phi_j = \sum_{l=1}^\infty (\partial \tilde{\psi}_l / \partial \phi_j) L^l$ is an upper bound for the absolute sum of the coefficients in $\partial \tilde{\psi}_+(L)/\partial \phi_j$. Showing $\sum_{l=1}^\infty |\partial \tilde{\psi}_l / \partial \phi_j| < \infty$ is thus sufficient for (42) to hold for the latter term in (52). The partial derivatives of the coefficients $\tilde{\psi}_l$, $l = 1, ..., \infty$, in (40) are

$$\begin{aligned} \frac{\partial \tilde{\psi}_l}{\partial \phi_j} &= \frac{1}{\sigma_u} \left(\tilde{\phi}_l^2 \sigma_\eta^2 + \pi_l^2(d) \sigma_\varepsilon^2 - 2\tilde{\phi}_l \pi_l(d) \sigma_{\eta\varepsilon}\right)^{-1/2} \left(\tilde{\phi}_l \sigma_\eta^2 - \pi_l(d) \sigma_{\eta\varepsilon}\right) \frac{\partial \tilde{\phi}_l}{\partial \phi_j} \\ &= \frac{1}{\sigma_u^2} \tilde{\psi}_l^{-1} \left(\tilde{\phi}_l \sigma_\eta^2 - \pi_l(d) \sigma_{\eta\varepsilon}\right) \frac{\partial \tilde{\phi}_l}{\partial \phi_j}, \end{aligned} \tag{53}$$

where the derivative of $\tilde{\phi}_l$ defined in (21) is

$$\frac{\partial \tilde{\phi}_l}{\partial \phi_j} = \frac{\partial}{\partial \phi_j} \sum_{i=1}^p (-1)^i \pi_l(di) \sum_{k=i}^p \phi_k \binom{k}{i} = \sum_{i=1}^j (-1)^i \pi_l(di) \binom{j}{i} = b_{\phi_j, l}, \tag{54}$$

and $b_{\phi_j,l} = O(l^{-d-1})$ since the sum in (54) is dominated by $\pi_l(d)$ and $j \leq p$ is finite. Plugging this result into (53) yields the partial derivative of $\tilde{\psi}(L)$ that is given by the polynomial

$$\frac{\partial \tilde{\psi}(L)}{\partial \phi_j} = \sum_{l=1}^\infty a_{\phi_j,l} L^l, \quad a_{\phi_j,l} = \frac{1}{\sigma_u^2} \tilde{\psi}_l^{-1} \left(\tilde{\phi}_l \sigma_\eta^2 - \pi_l(d) \sigma_{\eta\varepsilon}\right) \sum_{i=1}^j (-1)^i \pi_l(di) \binom{j}{i}, \tag{55}$$

where $a_{\phi_j,l}$ is $O(l^{-d-1})$, since the finite sum $\sum_{i=1}^j (-1)^i \pi_l(di) \binom{j}{i}$ is asymptotically dominated by $\pi_l(d)$ for all $j \leq p$, and $\tilde{\phi}_l = O(l^{-d-1})$, $\pi_l(d) = O(l^{-d-1})$, and $\tilde{\psi}_l^{-1} = O(l^{d+1})$ as noted before. Thus, the coefficients of the partial derivatives of $\tilde{\psi}(L)$ in (55) are absolutely summable $\sum_{l=1}^\infty |a_{\phi_j,l}| = \sum_{l=1}^\infty O(l^{-d-1}) < \infty$. As they are the upper bound for the absolute sum of $\partial \tilde{\psi}_+(L)/\partial \phi_j$, the coefficients of the latter are also absolutely summable. Since $\phi_+(L_d)$, $\tilde{\psi}_+(L)^{-2}$ in (52) are stable polynomials, and since $\Delta_+^d y_t$ is stationary for $\theta \in \Theta^*(\kappa)$, the second term $\phi_+(L_d) \tilde{\psi}_+(L)^{-2} \frac{\partial \tilde{\psi}_+(L)}{\partial \phi_j} \Delta_+^d y_t$ in (52) satisfies the absolute summability condition required for (42), so that (42) holds for (52). Consequently, a WLLN follows

$$\sup_{\theta \in \Theta^*(\kappa)} \Big| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial}{\partial \phi_i} l_t(\theta) \Big| = O_p(1), \qquad i = 1, ..., p. \tag{56}$$

Finally, the partial derivative w.r.t. the integration order d is

$$\frac{\partial l_t(\theta)}{\partial d} = \frac{u_t(\theta)}{\sigma_u^2} \left[\tilde{\psi}_+(L)^{-1} \frac{\partial \tilde{\psi}_+(L)}{\partial d} u_t(\theta) - \tilde{\psi}_+(L)^{-1} \frac{\partial \phi_+(L_d)}{\partial d} \Delta_+^d y_t - \tilde{\psi}_+(L)^{-1} \phi_+(L_d) \frac{\partial \Delta_+^d}{\partial d} y_t \right], \tag{57}$$

for which it has to be shown that (42) holds.

For $\theta \in \Theta^*(\kappa)$, $\tilde{\psi}_+(L)^{-1}\phi_+(L_d)\frac{\partial \Delta_+^d}{\partial d} y_t = \tilde{\psi}_+(L)^{-1}\phi_+(L_d)\frac{\partial \Delta_+^{d-d_0}}{\partial d}\Delta_+^{d_0} y_t$ in (57) trivially satisfies the conditions for (42) with $k = 0$, $l = 1$, as $\Delta_+^{d_0} y_t$ is a stationary moving average process with white noise shocks u_t, see (8), and $\tilde{\psi}_+(L)^{-1}$, $\phi_+(L_d)$ are stable polynomials.

For the two remaining terms in (57) it will be helpful to note that the partial derivative $(\partial \Delta^d/\partial d) = \sum_{j=1}^{\infty} (\partial \pi_j(d)/\partial d) L^j$, where an analytical expression for the derivative of $\pi_j(b)$ given that b is negative was derived in Hartl et al. (2020). To meet this requirement, we use $\pi_j(d) = \pi_j(d-1) - \pi_{j-1}(d-1) = -(d/j)\pi_{j-1}(d-1)$ (Hassler; 2018, eq. 5.22) which, by iteration, yields $\pi_j(d) = \pi_{j-s}(d-s)\prod_{k=0}^{s-1}(k-d)/(j-k)$ for all $j \geq s$. Now, set $s = \lceil d \rceil$, so that $d - s \leq 0$. Then, from Hartl et al. (2020) $\partial \pi_{j-s}(d-s)/\partial d = \pi_{j-s}(d-s)\sum_{k=0}^{j-s-1}(d-s-k)^{-1}$, $j \geq s$, and

$$\begin{aligned}
\frac{\partial \pi_j(d)}{\partial d} &= \frac{\partial}{\partial d}\left(\pi_{j-s}(d-s)\prod_{k=0}^{s-1}\frac{k-d}{j-k}\right) \\
&= -\pi_{j-s}(d-s)\sum_{l=0}^{s-1}\frac{1}{j-l}\prod_{k=0,k\neq l}^{s-1}\frac{k-d}{j-k} + \frac{\partial \pi_{j-s}(d-s)}{\partial d}\prod_{k=0}^{s-1}\frac{k-d}{j-k} \\
&= \pi_{j-s}(d-s)\prod_{k=0}^{s-1}\frac{k-d}{j-k}\sum_{l=0}^{s-1}(d-l)^{-1} + \pi_{j-s}(d-s)\prod_{k=0}^{s-1}\frac{k-d}{j-k}\sum_{l=0}^{j-s-1}(d-s-l)^{-1} \\
&= \pi_j(d)\sum_{l=0}^{s-1}\frac{1}{d-l} + \pi_j(d)\sum_{l=s}^{j-1}\frac{1}{d-l} = \pi_j(d)\sum_{k=0}^{j-1}(d-k)^{-1} = O(j^{-d-1}(1+\log j)),
\end{aligned} \tag{58}$$

for all $j \geq s$ which results from the sum being bounded by the limit of an harmonic series that is $O(1 + \log j)$, and $\pi_j(d) = O(j^{-d-1})$. The same result was derived in Johansen and Nielsen (2010) and Nielsen (2015) using a different proof.

With a limit for the partial derivative of $\pi_j(d)$ as given in (58) at hand, we can derive the limits of the two remaining polynomials in (57). Starting with $\partial\phi_+(L_d)/\partial d = \partial\tilde{\phi}_+(L)/\partial d$, of which the absolute sum of coefficients is bounded by the absolute sum of the untruncated polynomial $\partial\tilde{\phi}(L)/\partial d = -\sum_{j=1}^{\infty}(\partial\tilde{\phi}_j/\partial d)L^j$, one has

$$\frac{\partial\tilde{\phi}_j}{\partial d} = \sum_{l=1}^{p}(-1)^l\frac{\partial\pi_j(dl)}{\partial d}\sum_{k=l}^{p}\phi_k\binom{k}{l} = \sum_{l=1}^{p}(-1)^l\pi_j(dl)\left(\sum_{k=0}^{j-1}\frac{l}{dl-k}\right)\sum_{k=l}^{p}\phi_k\binom{k}{l} = b_{d,j}, \tag{59}$$

where, as noted before, $\sum_{k=l}^{p}\phi_k\binom{k}{l} = O(1)$ since p is finite, $\sum_{k=0}^{j-1}\frac{l}{dl-k} = O(1+\log j)$ since l is bounded by p and the sum is bounded by the limit of a harmonic series that is $O(1+\log j)$, and the $\pi_j(dl)$ are bounded by $\pi_j(d) = O(j^{-d-1})$, so that the whole term in (59) is $b_{d,j} = O(j^{-d-1}(1+\log j))$. Thus, $\partial\tilde{\phi}(L)/\partial d = \sum_{j=1}^{\infty}O(j^{-d-1}(1+\log j))$, and since $\log j$ is always dominated by any j^b with $b<0$, the coefficients of the partial derivative are absolutely summable $\sum_{j=1}^{\infty}O(j^{-d-1}(1+\log j)) < \infty$. The absolute sum of coefficients of the untruncated polynomial is an upper bound for the truncated polynomial, and thus the second term in parentheses in (57) satisfies the absolute summability condition for (42).

The third and final component in (57) to be studied is the partial derivative $\partial\tilde{\psi}_+(L)/\partial d$, where again the absolute sum of the coefficients $\partial\tilde{\psi}(L)/\partial d = \sum_{j=1}^{\infty}(\partial\tilde{\psi}_j/\partial d)L^j$ provides an upper bound. Via (40) one has

$$\frac{\partial\tilde{\psi}_j}{\partial d} = \frac{1}{\sigma_u^2}\tilde{\psi}_j^{-1}\left[\left(\tilde{\phi}_j\sigma_\eta^2 - \pi_j(d)\sigma_{\eta\varepsilon}\right)\frac{\partial\tilde{\phi}_j}{\partial d} + \left(\pi_j(d)\sigma_\varepsilon^2 - \tilde{\phi}_j\sigma_{\eta\varepsilon}\right)\frac{\partial\pi_j(d)}{\partial d}\right] = a_{d,j}. \tag{60}$$

As shown above, for $\theta \in \Theta^*(\kappa)$, $\tilde{\psi}_j^{-1} = O(j^{d+1})$, $\tilde{\phi}_j = O(j^{-d-1})$, $\pi_j(d) = O(j^{-d-1})$, while $\partial\tilde{\phi}_j/\partial d = O(j^{-d-1}(1+\log j))$ as shown in (59) and below, as well as $\partial\pi_j(d)/\partial d = O(j^{-d-1}(1+\log j))$ as shown in (58). Consequently, partial derivative in (60) is $a_{d,j} = O(j^{-d-1}(1+\log j))$, so that the coefficients of the partial derivative of $\tilde{\psi}(L)$ are absolutely summable $\sum_{j=1}^{\infty}|a_{d,j}| = \sum_{j=1}^{\infty}O(j^{-d-1}(1+\log j)) < \infty$. This implies absolute summability of the coefficients in $\partial\tilde{\psi}(L)_+/\partial d$ and thus the last term in (57) also satisfies the condition for (42).

As we have shown, (42) holds for all terms in (57), so that a WLLN follows for the partial derivative of the objective function w.r.t. d

$$\sup_{\theta\in\Theta^*(\kappa)}\left|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial d}l_t(\theta)\right| = O_p(1). \tag{61}$$

As shown in (51), (56), and (61), all partial derivatives of the objective function satisfy a WLLN, so that (39) holds. This generalizes the pointwise convergence of the objective function to weak convergence, implying that a UWLLN holds for the objective function. Since the model is identified, consistency for the QML estimator follows directly $\hat{\theta} \xrightarrow{p} \theta_0$ as $n \to \infty$, see Wooldridge (1994, th. 4.3).

$\square$

Proof of theorem 4.2. As shown in theorem 4.1, the QML estimator $\hat{\theta}$ is consistent, so that the asymptotic distribution can be obtained by applying a Taylor expansion of the score function at θ_0

$$0 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\theta)}{\partial \theta}\Big|_{\theta=\hat{\theta}} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}\Big|_{\theta=\bar{\theta}} (\hat{\theta} - \theta_0), \tag{62}$$

where $\bar{\theta}$ satisfies $|\bar{\theta}_i - \theta_{0,i}| \leq |\hat{\theta}_i - \theta_{0,i}|$ for all $i = 1, ..., p+4$, and $p+4$ is the dimension of θ. The normalized score at $\theta = \theta_0$ is given by

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial l_t(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} = \frac{-1}{\sqrt{n}\sigma_{u,0}^2} \sum_{t=1}^{n} \left\{ u_t \frac{\partial u_t(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} + \frac{1}{2} \frac{\partial \sigma_u^2}{\partial \theta} \left(1 - \frac{u_t^2}{\sigma_{u,0}^2} \right) \right\} = S_n + o_p(1),$$

with

$$S_n = \frac{-1}{\sqrt{n}\sigma_{u,0}^2} \sum_{t=1}^{n} \left\{ u_t \frac{\partial \tilde{u}_t(\theta)}{\partial \theta}\Big|_{\theta=\theta_0} + \frac{1}{2} \frac{\partial \sigma_u^2}{\partial \theta} \left(1 - \frac{u_t^2}{\sigma_{u,0}^2} \right) \right\}, \tag{63}$$

where the second equality is shown to hold in Robinson (2006, pp. 135-136) and Nielsen (2015, pp. 174-175), and $\tilde{u}_t(\theta)$ is the untruncated residual as defined in (38) and below. Furthermore, define $S_n^{(i)}$ as the i-th entry of S_n that holds the partial derivative w.r.t. θ_i.

To establish asymptotic normality of the QML estimator, we first show that a central limit theorem holds for the score function at θ_0. Next, we prove that a UWLLN holds for the Hessian matrix by showing that the Hessian matrix and its first partial derivatives satisfy a WLLN (Wooldridge; 1994, th. 4.2). This allows to evaluate the Hessian matrix in (62) at θ_0. Then it follows that the QML estimator $\hat{\theta}$ is asymptotically normally distributed, and the asymptotic variance follows from the inverse Fisher information matrix.

Following Nielsen (2015, p. 175) a central limit theorem for the score function is obtained by using the Cramér-Wold device. Thus, it has to be shown that for any $p+4$-vector μ, it holds that $\mu' S_n = \sum_{i=1}^{p+4} \mu_i S_n^{(i)} \xrightarrow{d} N(0, \mu' \Omega_0 \mu)$. Now, given the σ-field $\tilde{\mathcal{F}}_t = \sigma(\{u_s, s \leq t\})$ generated by the white noise process u_t and its lags, it is easy to see that in (63), $u_t \frac{\partial \tilde{u}_t(\theta)}{\partial \theta}\big|_{\theta=\theta_0}$ adapted to $\tilde{\mathcal{F}}_t$ is a stationary martingale difference sequence (MDS) since u_t is white noise, the partial derivatives are $\tilde{\mathcal{F}}_{t-1}$-measurable, and the coefficients of the partial derivatives are absolutely summable, as shown in the proof of theorem

4.1. In addition, the second term in (63), $\frac{1}{2}\partial\sigma_u^2/\partial\theta\left(1-u_t^2/\sigma_{u,0}^2\right)$ is a stationary MDS, as $\mathrm{E}(u_t^2)=\sigma_0^2$. Since we assume finite third and fourth moments, ν_t as given in

$$\nu_t=\sum_{i=1}^{p+4}(\nu_{1,i,t}+\nu_{2,i,t}),\quad \nu_{1,i,t}=\frac{\mu_i}{\sigma_{u,0}^2}u_t\frac{\partial\tilde{u}_t(\theta)}{\partial\theta_i}\Big|_{\theta=\theta_0},\quad \nu_{2,i,t}=\frac{\mu_i}{2\sigma_{u,0}^2}\frac{\partial\sigma_u^2}{\partial\theta_i}\left(1-\frac{u_t^2}{\sigma_{u,0}^2}\right),$$

adapted to $\tilde{\mathcal{F}}_t$ is a stationary MDS, and $\mu' S_n=-n^{-1/2}\sum_{t=1}^n\nu_t$.

As in Nielsen (2015, p. 175), the sum of conditional variances for $\mu' S_n$ with S_n as given in (63) is then

$$\frac{1}{n}\sum_{t=1}^n\mathrm{E}(\nu_t^2|\tilde{\mathcal{F}}_{t-1})=\frac{1}{n}\sum_{t=1}^n\sum_{i,j=1}^{p+4}\mathrm{E}\left[\nu_{1,i,t}\nu_{1,j,t}+\nu_{1,i,t}\nu_{2,j,t}+\nu_{2,i,t}\nu_{1,j,t}+\nu_{2,i,t}\nu_{2,j,t}|\tilde{\mathcal{F}}_{t-1}\right].\tag{64}$$

Since $\mathrm{E}[u_t^2|\tilde{\mathcal{F}}]=\sigma_{u,0}^2$, $\mathrm{E}[u_t\left(1-u_t^2/\sigma_{u,0}^2\right)|\tilde{\mathcal{F}}_{t-1}]=-\gamma_{u,0}\sigma_{u,0}$ where $\gamma_{u,0}$ is the skewness of u_t that is finite by assumption, and since $\mathrm{E}[\left(1-u_t^2/\sigma_{u,0}^2\right)^2|\tilde{\mathcal{F}}_{t-1}]=\mathrm{E}[u_t^4/\sigma_{u,0}^4|\tilde{\mathcal{F}}_{t-1}]-1=\kappa_{u,0}-1$ where $\kappa_{u,0}$ is the kurtosis of u_t that is finite by assumption, we have

$$\mathrm{E}\left[\nu_{1,i,t}\nu_{1,j,t}|\tilde{\mathcal{F}}_{t-1}\right]=\frac{\mu_i\mu_j}{\sigma_{u,0}^2}\frac{\partial\tilde{u}_t(\theta)}{\partial\theta_i}\Big|_{\theta=\theta_0}\frac{\partial\tilde{u}_t(\theta)}{\partial\theta_j}\Big|_{\theta=\theta_0},\tag{65}$$

$$\mathrm{E}\left[\nu_{2,i,t}\nu_{2,j,t}|\tilde{\mathcal{F}}_{t-1}\right]=\frac{\mu_i\mu_j}{4\sigma_{u,0}^4}\frac{\partial\sigma_u^2}{\partial\theta_i}\frac{\partial\sigma_u^2}{\partial\theta_j}(\kappa_{u,0}-1),\tag{66}$$

$$\mathrm{E}\left[\nu_{1,i,t}\nu_{2,j,t}|\tilde{\mathcal{F}}_{t-1}\right]=-\frac{\mu_i\mu_j}{2\sigma_{u,0}^3}\frac{\partial\tilde{u}_t(\theta)}{\partial\theta_i}\Big|_{\theta=\theta_0}\frac{\partial\sigma_u^2}{\partial\theta_j}\gamma_{u,0}.\tag{67}$$

Note that in (67) $\mathrm{E}\left\{\mathrm{E}\left[\nu_{1,i,t}\nu_{2,j,t}|\tilde{\mathcal{F}}_{t-1}\right]\right\}=0$ since $\mathrm{E}[\partial\tilde{u}_t(\theta)/\theta_i|_{\theta=\theta_0}]=0$ for all $i=1,...,p+4$ and all other terms are purely deterministic. Furthermore (66) is purely deterministic, since the partial derivatives are zero if $\theta_i\notin\{\sigma_\eta^2,\sigma_{\eta\varepsilon},\sigma_\varepsilon^2\}$, equal to one if $\theta_i\in\{\sigma_\eta^2,\sigma_\varepsilon^2\}$, and equal to two if $\theta_i=\sigma_{\eta\varepsilon}$. Thus, from the law of large numbers for stationary and ergodic processes, it follows that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n\mathrm{E}(\nu_t^2|\tilde{\mathcal{F}}_{t-1})&\xrightarrow{p}\sum_{i,j=1}^{p+4}\mathrm{E}\left\{\mathrm{E}\left[\nu_{1,i,t}\nu_{1,j,t}+\nu_{2,i,t}\nu_{2,j,t}|\tilde{\mathcal{F}}_{t-1}\right]\right\}\\&=\sum_{i,j=1}^{p+4}\frac{\mu_i\mu_j}{\sigma_{u,0}^2}\left\{\mathrm{E}\left[\frac{\partial\tilde{u}_t(\theta)}{\partial\theta_i}\Big|_{\theta=\theta_0}\frac{\partial\tilde{u}_t(\theta)}{\partial\theta_j}\Big|_{\theta=\theta_0}\right]+\frac{\partial\sigma_u^2}{\partial\theta_i}\frac{\partial\sigma_u^2}{\partial\theta_j}\frac{\kappa_{u,0}-1}{4\sigma_{u,0}^2}\right\}.\end{aligned}\tag{68}$$

To obtain an expression for the expected values in (68), we evaluate the partial derivatives of $\tilde{u}_t(\theta)$ at $\theta=\theta_0$, where we use that $(\partial/\partial d)\Delta^d\big|_{\theta=\theta_0}=(\partial/\partial d)\Delta^{d-d_0}\big|_{\theta=\theta_0}\Delta^{d_0}=\Delta^{d_0}\sum_{l=1}^\infty l^{-1}L^l$, see Hartl et al. (2020). The partial derivatives for $\tilde{u}_t(\theta)$ then follow directly from the proof of theorem 4.1 and are summarized in the following, where an expression for $a_{d,l,0}$ is given in (60), for $b_{d,l,0}$ in (59), and for $a_{\phi_j,l,0}$ in (55), and all coefficients are evaluated at $\theta = \theta_0$, which we denote with a zero in the subscript. Hence

$$\begin{aligned}
\frac{\partial \tilde{u}_t(\theta)}{\partial d}\Big|_{\theta=\theta_0} &= \left(\sum_{l=1}^{\infty} l^{-1} L^l - \tilde{\psi}_0(L)^{-1} \sum_{l=1}^{\infty} a_{d,l,0} L^l - \tilde{\phi}_0(L)^{-1} \sum_{l=1}^{\infty} b_{d,l,0} L^l \right) u_t = m_{d,0}(L) u_t, \\
\frac{\partial \tilde{u}_t(\theta)}{\partial \phi_j}\Big|_{\theta=\theta_0} &= - \left(\tilde{\phi}_0(L)^{-1} (1-\Delta^{d_0})^j + \tilde{\psi}_0(L)^{-1} \sum_{l=1}^{\infty} a_{\phi_j,l,0} L^l \right) u_t = m_{\phi_j,0}(L) u_t, \\
\frac{\partial \tilde{u}_t(\theta)}{\partial \sigma_\eta^2}\Big|_{\theta=\theta_0} &= \frac{\tilde{\psi}_0(L)^{-1}}{2\sigma_{u,0}^2} \sum_{l=1}^{\infty} \left(\tilde{\psi}_{l,0} - \frac{\tilde{\phi}_{l,0}^2}{\tilde{\psi}_{l,0}} \right) L^l u_t = m_{\sigma_\eta^2,0}(L) u_t, \\
\frac{\partial \tilde{u}_t(\theta)}{\partial \sigma_{\eta\varepsilon}}\Big|_{\theta=\theta_0} &= \frac{\tilde{\psi}_0(L)^{-1}}{\sigma_{u,0}^2} \sum_{l=1}^{\infty} \left(\tilde{\psi}_{l,0} + \frac{\tilde{\phi}_{l,0} \pi_l(d_0)}{\tilde{\psi}_{l,0}} \right) L^l u_t = m_{\sigma_{\eta\varepsilon},0}(L) u_t, \\
\frac{\partial \tilde{u}_t(\theta)}{\partial \sigma_\varepsilon^2}\Big|_{\theta=\theta_0} &= \frac{\tilde{\psi}_0(L)^{-1}}{2\sigma_{u,0}^2} \sum_{l=1}^{\infty} \left(\tilde{\psi}_{l,0} - \frac{\pi_l^2(d_0)}{\tilde{\psi}_{l,0}} \right) L^l u_t = m_{\sigma_\varepsilon^2,0}(L) u_t.
\end{aligned}$$

Thus, for the expected value in (68)

$$\mathrm{E}\left[\frac{\partial \tilde{u}_t(\theta)}{\partial \theta_i}\Big|_{\theta=\theta_0} \frac{\partial \tilde{u}_t(\theta)}{\partial \theta_j}\Big|_{\theta=\theta_0} \right] = \sigma_{u,0}^2 \sum_{l=1}^{\infty} m_{\theta_i,l,0} m_{\theta_j,l,0}, \tag{69}$$

and $\sum_{l=1}^{\infty} |m_{\theta_i,l,0} m_{\theta_j,l,0}| < \infty$ for all $i, j = 1, ..., p+4$, which follows directly for $i, j = 2, ..., p+4$, since all coefficients in the polynomials are $O(l^{-d-1})$. For the partial derivative w.r.t. d, note that $a_{d,l,0}$, $b_{d,l,0}$ are $O(l^{-d-1}(1 + \log l))$, and $\sum_{l=1}^{\infty} l^{-1}$ is a harmonic series, so that $\mathrm{Var}(\sum_{l=1}^{\infty} l^{-1} u_{t-l}) = \sigma_{u,0}^2 \sum_{l=1}^{\infty} l^{-2}$ is bounded by the limit of the Riemann zeta function $\zeta(s)$ with $s = 2$. Consequently in (68), $n^{-1} \sum_{t=1}^{n} \mathrm{E}(\nu_t^2 | \tilde{\mathcal{F}}_{t-1}) \xrightarrow{p} \mu_i \mu_j \Omega_0^{(i,j)}$, where

$$\Omega_0^{(i,j)} = \begin{cases} \sum_{l=1}^{\infty} m_{\theta_i,l,0} m_{\theta_j,l,0} + \frac{\partial \sigma_u^2}{\partial \theta_i} \frac{\partial \sigma_u^2}{\partial \theta_j} \frac{\kappa_{u,0} - 1}{4\sigma_{u,0}^4} & \text{if both } \theta_i, \theta_j \in \{\sigma_\eta^2, \sigma_{\eta\varepsilon}, \sigma_\varepsilon^2\}, \\ \sum_{l=1}^{\infty} m_{\theta_i,l,0} m_{\theta_j,l,0} & \text{else.} \end{cases} \tag{70}$$

Since ν_t is stationary, a central limit theorem for MDS (cf. e.g. Davidson; 2000, th. 6.2.3) applies so that $S_n \xrightarrow{d} N(0, \Omega_0)$ with entries of Ω_0 as given in (70).

For asymptotic normality of the QML estimator it remains to be shown that the Hessian matrix satisfies a UWLLN (Wooldridge; 1994, th. 4.4), which holds if a WLLN applies to the Hessian matrix and

$$\sup_{\theta \in \Theta^*(\kappa)} \left| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^3}{\partial \theta^3} l_t(\theta) \right| = O_p(1), \tag{71}$$

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

A WLLN holds for the Hessian matrix

$$H_t(\theta) = \frac{\partial^2 l_t(\theta)}{\partial\theta\partial\theta'} = -\frac{1}{\sigma_u^2}\left[\frac{\partial u_t(\theta)}{\partial\theta}\frac{\partial u_t(\theta)}{\partial\theta'} + u_t(\theta)\frac{\partial^2 u_t(\theta)}{\partial\theta\partial\theta'}\right] + R_{\sigma_u,t}, \tag{72}$$

$$R_{\sigma_u,t} = \frac{\partial\sigma_u^2}{\partial\theta}\left[\frac{1}{2\sigma_u^4}\frac{\partial\sigma_u^2}{\partial\theta'} - \frac{1}{\sigma_u^6}u_t^2(\theta)\frac{\partial\sigma_u^2}{\partial\theta'} + \frac{2}{\sigma_u^4}u_t(\theta)\frac{\partial u_t(\theta)}{\partial\theta'}\right],$$

if (72) satisfies the absolute summability condition for (42). $R_{\sigma_u,t}$ holds the additional terms for the partial derivatives w.r.t. $\sigma_\eta^2, \sigma_{\eta\varepsilon}, \sigma_\varepsilon^2$ and satisfies the absolute summability condition for (42) since the first partial derivatives were shown to be absolutely summable in the proof of theorem 4.1. The same argument applies to $\frac{\partial u_t(\theta)}{\partial\theta}\frac{\partial u_t(\theta)}{\partial\theta'}$ in (72), so that it remains to be shown that absolute summability also holds for $u_t(\theta)\frac{\partial^2 u_t(\theta)}{\partial\theta\partial\theta'}$. Note that the coefficients in $\partial^2\phi_+(L_d)/(\partial\theta\partial\theta') = -\partial^2/(\partial\theta\partial\theta')\sum_{j=1}^p \phi_j(1-\Delta_+^d)^j$ are absolutely summable, since $d > 0$ and $\partial^k\Delta_+^d/\partial d^k = (\partial^k/\partial d^k)(\sum_{j=1}^\infty \pi_j(d)L^j)_+ = [\sum_{j=1}^\infty O((1+\log j)^k j^{-d-1})L^j]_+$, see Nielsen (2015, lemma A.1), is absolutely summable. Furthermore, it follows from the proof of theorem 4.1 that the coefficients of the products $(\partial\tilde{\psi}_+(L)^{-1}/\partial\theta)(\partial\phi_+(L_d)/\partial\theta')$, $(\partial\tilde{\psi}_+(L)^{-1}/\partial\theta)(\partial\Delta_+^d/\partial\theta')$, $(\partial\phi_+(L_d)/\partial\theta)(\partial\Delta_+^d/\partial\theta')$ are absolutely summable, so that for (72) to satisfy the absolute summability condition for (42), absolute summability remains to be shown for the coefficients $\partial^2\tilde{\psi}_+(L)^{-1}/(\partial\theta\partial\theta') = 2\tilde{\psi}_+(L)^{-3}(\partial\tilde{\psi}_+(L)/\partial\theta)(\partial\tilde{\psi}_+(L)/\partial\theta') - \tilde{\psi}_+(L)^{-2}(\partial^2\tilde{\psi}_+(L)/(\partial\theta\partial\theta'))$. Since $\partial\tilde{\psi}_+(L)^{-1}/\partial\theta$ satisfies the absolute summability condition, as shown in the proof of theorem 4.1, and since $\tilde{\psi}_+(L)$ is a stable moving average polynomial, it is sufficient to prove that $\partial^2\tilde{\psi}(L)/(\partial\theta\partial\theta') = \sum_{j=1}^\infty \left|\partial^2\tilde{\psi}_j/(\partial\theta_k\partial\theta_l)\right| < \infty$ for all $k, l = 1, ..., p+4$.

We collect the second partial derivatives of $\tilde{\psi}_j$ in a matrix M_j

$$\frac{\partial^2\tilde{\psi}_j}{\partial\theta\partial\theta'} = \frac{\partial}{\partial\theta}\begin{pmatrix} a_{d,j} & a_{\phi_1,j} & \cdots & a_{\phi_p,j} & \frac{\tilde{\phi}_j^2-\tilde{\psi}_j^2}{2\tilde{\psi}_j\sigma_u^2} & \frac{-\tilde{\phi}_j\pi_j(d)-\tilde{\psi}_j^2}{\tilde{\psi}_j\sigma_u^2} & \frac{\pi_j(d)^2-\tilde{\psi}_j^2}{2\tilde{\psi}_j\sigma_u^2}\end{pmatrix} = M_j.$$

The entries $M_j^{(k,l)}$, with $k, l = 1, ..., p+4$, are summarized below. Convergence rates follow from $\tilde{\phi}_j = O(j^{-d-1})$, $\tilde{\psi}_j = O(j^{-d-1})$, $\tilde{\psi}_j^{-1} = O(j^{1+d})$, and $\pi_j(d) = O(j^{-d-1})$, as stated in the proof of theorem 4.1. Furthermore $a_{\phi_k,j} = O(j^{-d-1})$ as shown in (55), $b_{\phi_k,j} = O(j^{-d-1})$ as shown in (54), $a_{d,j} = O(j^{-d-1}(1+\log j))$ as shown in (60), $b_{d,j} = O(j^{-d-1}(1+\log j))$ as shown in (59), and $\partial^k\pi_j(d)/\partial d^k = O(j^{-d-1}(1+\log j)^k)$ as given in Nielsen (2015, lemma A.1). Finally, note that $\partial b_{d,j}/\partial d = \partial^2\tilde{\phi}_j/\partial d^2 = \sum_{l=1}^p(-1)^l(\partial^2\pi_j(dl)/\partial d^2)\sum_{k=l}^p\phi_k\binom{k}{l} = O(j^{-d-1}(1+\log j)^2)$ and $\partial b_{\phi_l,j}/\partial d = \sum_{i=1}^l(-1)^i(\partial\pi_j(di)/\partial d)\binom{l}{i} = O(j^{-d-1}(1+\log j))$. In the formulas below, denote $k, l = 1, ..., p$ as indices for the p coefficients $\phi_1, ..., \phi_p$. Then

$$\begin{aligned}
M_j^{(1,1)} =& -\tilde{\psi}_j^{-1} a_{d,j}^2 + \frac{1}{\sigma_u^2 \tilde{\psi}_j} \Big[\left(b_{d,j}\sigma_\eta^2 - \frac{\partial \pi_j(d)}{\partial d}\sigma_{\eta\varepsilon} \right) b_{d,j} + \left(\tilde{\phi}_j \sigma_\eta^2 - \pi_j(d)\sigma_{\eta\varepsilon} \right) \frac{\partial^2 \tilde{\phi}_j}{\partial d^2} \\
&+ \left(\frac{\partial \pi_j(d)}{\partial d}\sigma_\varepsilon^2 - b_{d,j}\sigma_{\eta\varepsilon} \right) \frac{\partial \pi_j(d)}{\partial d} + \left(\pi_j(d)\sigma_\varepsilon^2 - \tilde{\phi}_j \sigma_{\eta\varepsilon} \right) \frac{\partial^2 \pi_j(d)}{\partial d^2} \Big] \\
=& O(j^{-d-1}(1+\log j)^2), \qquad (73) \\
M_j^{(1,l+1)} =& -\tilde{\psi}_j^{-1} a_{\phi_l,j} a_{d,j} + \frac{1}{\sigma_u^2 \tilde{\psi}_j} \left[b_{\phi_l,j} b_{d,j} \sigma_\eta^2 + (\tilde{\phi}_j \sigma_\eta^2 - \pi_j(d)\sigma_{\eta\varepsilon}) \frac{\partial b_{d,j}}{\partial \phi_l} - b_{\phi_l,j} \frac{\partial \pi_j(d)}{\partial d}\sigma_{\eta\varepsilon} \right] \\
=& O(j^{-d-1}(1+\log j)), \qquad (74) \\
M_j^{(1,p+2)} =& -\frac{a_{d,j}}{2\sigma_u^2} \left(1 + \frac{\tilde{\phi}_j^2}{\tilde{\psi}_j^2} \right) + \frac{\tilde{\phi}_j}{\sigma_u^2 \tilde{\psi}_j} b_{d,j} = O(j^{-d-1}(1+\log j)), \qquad (75) \\
M_j^{(1,p+3)} =& \frac{-a_{d,j}}{\sigma_u^2} \left(1 - \frac{\tilde{\phi}_j \pi_j(d)}{\tilde{\psi}_j^2} \right) - \frac{1}{\sigma_u^2 \tilde{\psi}_j} \left(\pi_j(d) b_{d,j} + \tilde{\phi}_j \frac{\partial \pi_j(d)}{\partial d} \right) = O(j^{-d-1}(1+\log j)), \\
& \qquad (76) \\
M_j^{(1,p+4)} =& -\frac{a_{d,j}}{2\sigma_u^2} \left(1 + \frac{\pi_j^2(d)}{\tilde{\psi}_j^2} \right) + \frac{\pi_j(d)}{\sigma_u^2 \tilde{\psi}_j} \frac{\partial \pi_j(d)}{\partial d} = O(j^{-d-1}(1+\log j)), \qquad (77) \\
M_j^{(1+k,1+l)} =& -\tilde{\psi}_j^{-1} a_{\phi_l,j} a_{\phi_k,j} + \frac{\sigma_\eta^2 b_{\phi_l,j} b_{\phi_k,j}}{\sigma_u^2 \tilde{\psi}_j} = O(j^{-d-1}), \qquad (78) \\
M_j^{(1+k,2+p)} =& -\frac{a_{\phi_k,j}}{2\sigma_u^2} \left(1 + \frac{\tilde{\phi}_j^2}{\tilde{\psi}_j^2} \right) + \frac{\tilde{\phi}_j b_{\phi_k,j}}{\tilde{\psi}_j \sigma_u^2} = O(j^{-d-1}), \qquad (79) \\
M_j^{(1+k,3+p)} =& -\frac{a_{\phi_k,j}}{\sigma_u^2} \left(1 - \frac{\tilde{\phi}_j \pi_j(d)}{\tilde{\psi}_j^2} \right) - \frac{\pi_j(d) b_{\phi_k,j}}{\tilde{\psi}_j \sigma_u^2} = O(j^{-d-1}), \qquad (80) \\
M_j^{(1+k,4+p)} =& -\frac{a_{\phi_k,j}}{2\sigma_u^2} \left(1 + \frac{\pi_j^2(d)}{\tilde{\psi}_j^2} \right) = O(j^{-d-1}), \qquad (81) \\
M_j^{(2+p,2+p)} =& \frac{3\tilde{\psi}_j^4 - 2\tilde{\phi}_j^2 \tilde{\psi}_j^2 - \tilde{\phi}_j^4}{4\tilde{\psi}_j^3 \sigma_u^4} = O(j^{-d-1}), \qquad (82) \\
M_j^{(2+p,3+p)} =& \frac{3\tilde{\psi}_j^4 + \tilde{\psi}_j^2 \tilde{\phi}_j \pi_j(d) - \tilde{\psi}_j^2 \tilde{\phi}_j^2 + \tilde{\phi}_j^3 \pi_j(d)}{2\tilde{\psi}_j^3 \sigma_u^4} = O(j^{-d-1}), \qquad (83) \\
M_j^{(2+p,4+p)} =& \frac{3\tilde{\psi}_j^4 - \tilde{\phi}_j^2 \tilde{\psi}_j^2 - \pi_j^2(d) \tilde{\psi}_j^2 - \tilde{\phi}_j^2 \pi_j^2(d)}{4\tilde{\psi}_j^3 \sigma_u^4} = O(j^{-d-1}), \qquad (84) \\
M^{(3+p,3+p)} =& \frac{3\tilde{\psi}_j^4 + 2\tilde{\phi}_j \pi_j(d) \tilde{\psi}_j^2 - \tilde{\phi}_j^2 \pi_j^2(d)}{\tilde{\psi}_j^3 \sigma_u^4} = O(j^{-d-1}), \qquad (85) \\
M^{(3+p,4+p)} =& \frac{3\tilde{\psi}_j^4 + \tilde{\phi}_j \pi_j(d) \tilde{\psi}_j^2 - \pi_j^2(d) \tilde{\psi}_j^2 + \tilde{\phi}_j \pi_j^3(d)}{2\tilde{\psi}_j^3 \sigma_u^4} = O(j^{-d-1}), \qquad (86)
\end{aligned}$$

$$M_j^{(4+p,4+p)} = \frac{3\tilde{\psi}_j^4 - 2\pi_j^2(d)\tilde{\psi}_j^2 - \pi_j^4(d)}{4\tilde{\psi}_j^3\sigma_u^4} = O(j^{-d-1}), \tag{87}$$

and thus $M_j = O(j^{-d-1}(1+\log j)^2)$, so that $\sum_{j=1}^{\infty} |M_j| < \infty$. Consequently, the coefficients of $\partial^2 u_t(\theta)/(\partial\theta\partial\theta')$ are absolutely summable, so that (42) holds for the Hessian matrix. Hence, the Hessian matrix satisfies a WLLN.

Finally, for (71) to hold, it remains to be shown that $\partial H_t(\theta)/\partial\theta$ satisfies the absolute summability condition for (42), with $H_t(\theta)$ given in (72). In addition to the third partial derivatives of $u_t(\theta)$, $\partial H_t(\theta)/\partial\theta$ depends on the cross products of first and second partial derivatives that have already been shown to satisfy the conditions for (42). Thus, it remains to be shown that absolute summability holds for $u_t(\partial^3 u_t(\theta)/\partial\theta^3)$.

For the same reason as above, the coefficients in $\partial^3\phi_+(L_d)/(\partial\theta^3) = -\partial^3/(\partial\theta^3)\sum_{j=1}^{p}\phi_j(1-\Delta_+^d)^j$ are absolutely summable, since $d > 0$ and $\partial^k\Delta_+^d/\partial d^k = (\partial^k/\partial d^k)(\sum_{j=1}^{\infty}\pi_j(d)L^j)_+ = [\sum_{j=1}^{\infty} O((1+\log j)^k j^{-d-1})L^j]_+$, see Nielsen (2015, lemma A.1), is absolutely summable. Furthermore, it follows from the proof of theorem 4.1 and from the properties of $M_j^{(k,l)}$ as stated above that the coefficients of the cross products $(\partial^2\tilde{\psi}_+(L)^{-1}/\partial\theta\partial\theta')(\partial\phi_+(L_d)/\partial\theta)$, $(\partial^2\tilde{\psi}_+(L)^{-1}/\partial\theta\partial\theta')(\partial\Delta_+^d/\partial\theta)$, $(\partial\tilde{\psi}_+(L)^{-1}/\partial\theta')(\partial^2\Delta_+^d/\partial\theta\partial\theta')$, $(\partial^2\phi_+(L_d)/\partial\theta\partial\theta')(\partial\Delta_+^d/\partial\theta)$, $(\partial\phi_+(L_d)/\partial\theta')(\partial^2\Delta_+^d/\partial\theta\partial\theta')$, and $(\partial\tilde{\psi}_+(L)^{-1}/\partial\theta')(\partial^2\phi_+(L_d)/\partial\theta\partial\theta')$, are absolutely summable, so that for the third partial derivatives of the objective function to satisfy the absolute summability condition for (42), it remains to be shown that the coefficients in $\partial^3\tilde{\psi}_+(L)^{-1}/(\partial\theta^3)$ are absolutely summable. Finally, since absolute summability holds for $\partial\tilde{\psi}_+(L)/(\partial\theta)$, $\partial^2\tilde{\psi}_+(L)/(\partial\theta\partial\theta')$, and since $\tilde{\psi}_+(L)$ is a stable moving average polynomial, it is sufficient to prove that $\sum_{j=1}^{\infty}\left|\partial^3\tilde{\psi}_j/(\partial\theta_k\partial\theta_l\partial\theta_m)\right| = \sum_{j=1}^{\infty}\left|\partial M_j^{(k,l)}/\partial\theta_m\right| < \infty$ for all $k, l, m = 1, ..., p+4$.

In the following, we will make use of $\partial a_{\phi_k,j}/\partial\theta = O(j^{-d-1}(1+\log j))$ and $\partial b_{\phi_k,j}/\partial\theta = O(j^{-d-1}(1+\log j))$, which is easy to see from (54) and (55), as the partial derivatives w.r.t. d are $O(j^{-d-1}(1+\log j))$, all other partial derivatives of $b_{\phi_k,j}$ are zero, and those of $a_{\phi_k,j}$ are $O(j^{-d-1})$. Furthermore $\partial a_{d,j}/\partial\theta = O(j^{-d-1}(1+\log j)^2)$ and $\partial b_{d,j}/\partial\theta = O(j^{-d-1}(1+\log j)^2)$, as the partial derivatives w.r.t. d are $O(j^{-d-1}(1+\log j)^2)$, and all others are $O(j^{-d-1}(1+\log j))$, which can be seen from (59) and (60). Finally, since $\partial^k\pi_j(d)/\partial d^k = O(j^{-d-1}(1+\log j)^k)$, it follows that $\partial^2 b_{\phi_l,j}/(\partial\theta\partial\theta') = O(j^{-d-1}(1+\log j)^2)$, which can be seen directly from (54), and $\partial^2 b_{d,j}/(\partial\theta\partial\theta') = O(j^{-d-1}(1+\log j)^3)$, which can be seen from (59).

With these limits at hand, from (78) – (87) is easy to see that $\partial M_j^{(k,l)}/\partial\theta = O(j^{-d-1}(1+\log j))$ for $k, l > 1$, since $\partial M_j^{(k,l)}/\partial d = O(j^{-d-1}(1+\log j))$ adds a log factor to the convergence rate, while all other partial derivatives of (78) – (87) are $O(j^{-d-1})$. Furthermore, from (74) – (77) it follows that $\partial M_j^{(1,k)}/\partial\theta = O(j^{-d-1}(1+\log j)^2)$ for $k > 1$, since the partial derivative w.r.t. d adds a log factor, while all other partial derivatives preserve the limiting behavior. Finally, for $\partial M_j^{(1,1)}/\partial\theta$ it follows from the limiting behavior of $\partial^2 b_{d,j}/(\partial\theta\partial\theta')$ and $\partial^3\pi_j(d)/\partial d^3$ that $\partial M_j^{(1,1)}/\partial\theta = O(j^{-d-1}(1+\log j)^3)$. Since the log factor is always dominated by j^{-d-1}, the coefficients of the partial derivatives of M_j are absolutely summable, so that (42) holds. Thus, a WLLN applies to the third partial derivatives of the objective function and (71) holds respectively. From this, it follows that pointwise convergence of the Hessian matrix can be generalized to uniform convergence, so that the QML estimator converges in distribution $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,\Omega_0^{-1})$ with Ω_0 as given in (70). This completes the proof. $\square$

References

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

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.

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

Davidson, J. (2000). *Econometric Theory*, Blackwell Publishers.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching, *Journal of Econometrics* **105**(1): 131–159.

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

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.

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

Hartl, T., Tschernig, R. and Weber, E. (2020). Fractional trends in unobserved components models, *arXiv:2005.03988*, arXiv.org.
URL: *https://arxiv.org/pdf/2005.03988.pdf*

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

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

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. (2018). *Time Series Analysis with Long Memory in View*, Wiley Series in Probability and Statistics, Wiley.

Hou, J. and Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations, *Journal of Econometrics* **182**(2): 309–328.

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

Johansen, S. (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.

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.

Klinger, S. and Weber, E. (2016). Decomposing Beveridge curve dynamics by correlated unobserved components, *Oxford Bulletin of Economics and Statistics* **78**(6): 877–894. **URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1111/obes.12135*

Koopman, S. J., Ooms, M. and Hindrayanto, A. I. W. (2012). A multivariate periodic unobserved components time series analysis for sectoral US employment, *in* W. R. Bell, S. H. Holan and T. McElroy (eds), *Economic Time Series: Modeling and Seasonality*, Chapman and Hall/CRC Press, pp. 3–35.

Kruse, R. (2015). A modified test against spurious long memory, *Economics Letters* **135**: 34–38.

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

Morley, J. C. (2007). The slow adjustment of aggregate consumption to permanent income, *Journal of Money, Credit and Banking* **39**(2–3): 615–638.

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., Panovska, I. B. and Sinclair, T. M. (2017). Testing stationarity with unobserved-components models, *Macroeconomic Dynamics* **21**: 160–182.

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.

Ng, S. and Perron, P. (2005). A note on the selection of time series models, *Oxford Bulletin of Economics and Statistics* **67**(1): 115–134.

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

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

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

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.

Pötscher, B. (1991). Noninvertibility and pseudo-maximum likelihood estimation of misspecified ARMA models, *Econometric Theory* **7**(4): 435–449.

Qu, Z. (2011). A test against spurious long memory, *Journal of Business & Economic Statistics* **29**(3): 423–438.

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 *ISM Lecture Notes-Monograph Series*, pp. 130–137.

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

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

Sibbertsen, P. (2004). Long memory versus structural breaks: An overview, *Statistical Papers* **45**: 465–515.

Sun, Y. and Phillips, P. C. B. (2004). Understanding the Fisher equation, *Journal of Applied Econometrics* **19**(7): 869–886.

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.