

Nielsen, Morten Ørregaard

Working Paper

Asymptotics for the conditional-sum-of-squares estimator in fractional time series models

Queen's Economics Department Working Paper, No. 1259

Provided in Cooperation with:

Queen's University, Department of Economics (QED)

Suggested Citation: Nielsen, Morten Ørregaard (2011) : Asymptotics for the conditional-sum-of-squares estimator in fractional time series models, Queen's Economics Department Working Paper, No. 1259, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:

<https://hdl.handle.net/10419/67828>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Queen's Economics Department Working Paper No. 1259

Asymptotics for the conditional-sum-of-squares estimator in fractional time series models

Morten Ørregaard Nielsen
Queen's University and CREATES

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2011

Asymptotics for the conditional-sum-of-squares estimator in fractional time series models*

Morten Ørregaard Nielsen[†]
Queen's University and CREATES

January 21, 2011

Abstract

This paper proves consistency and asymptotic normality for the conditional-sum-of-squares (CSS) estimator in fractional time series models. The models are parametric and quite general. The novelty of the consistency result is that it applies to an arbitrarily large set of admissible parameter values, for which the objective function does not converge uniformly in probability thus making the proof much more challenging than usual. The neighborhood around the critical point where uniform convergence fails is handled using a truncation argument. The only other consistency proof for such models that applies to an arbitrarily large set of admissible parameter values appears to be Hualde & Robinson (2010), who require all moments of the innovation process to exist. In contrast, the present proof requires only a few moments of the innovation process to be finite (four in the simplest case). Finally, all arguments, assumptions, and proofs in this paper are stated entirely in the time domain, which is somewhat remarkable for this literature.

Key words and phrases: Asymptotic normality, conditional-sum-of-squares estimator, consistency, fractional integration, fractional time series, likelihood inference, long memory, nonstationary, uniform convergence.

MSC 2010 subject classification: Primary 62F12; secondary 60G22, 62E20, 62M10.

JEL classification: C22.

*I am grateful to James Davidson, Søren Johansen, James MacKinnon, and Lealand Morin for comments, and to the Social Sciences and Humanities Research Council of Canada (SSHRC grant 410-2009-0183), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation) for financial support.

[†]Address: Department of Economics, Dunning Hall Room 307, Queen's University, 94 University Avenue, Kingston, Ontario K7L 3N6, Canada. Email: mon@econ.queensu.ca

1 Introduction

This paper considers conditional-sum-of-squares (CSS) estimation of fractional time series models. The CSS estimator is based on minimizing the sum of squared residuals, and was applied in classical work on ARIMA models by, e.g., Box & Jenkins (1970). In later work, CSS estimation was introduced for fractional time series models by Li & McLeod (1986) and Robinson (1994), in the latter case for hypothesis testing purposes. The CSS estimator has the anticipated advantage of having the same asymptotic normal distribution as the Gaussian maximum likelihood estimator and being efficient under Gaussianity. However, Gaussianity is not assumed in this paper. Compared to maximum likelihood estimation, though, CSS estimation is computationally much simpler.

In the simplest case, the fractional time series model is

$$\Delta_+^d X_t = \varepsilon_t, \quad (1)$$

where the operator Δ_+^d is given by $\Delta_+^d X_t = \Delta^d X_t 1_{\{t \geq 1\}} = \sum_{n=0}^{t-1} \pi_n(-d) X_{t-n}$ with

$$\pi_n(u) = \frac{\Gamma(u+n)}{\Gamma(u)\Gamma(n+1)} = \frac{u(u+1)\dots(u+n-1)}{n!}. \quad (2)$$

denoting the coefficients in the usual binomial expansion of $(1-z)^{-u}$ and $1_{\{A\}}$ denoting the indicator function of the event A . Note that $\Delta_+^d X_t$ only depends on X_t for $t \geq 1$ and is invertible on the sequences which are zero for $t \leq 0$. The inverse operator Δ_+^{-d} is given by $\Delta_+^{-d} X_t = \sum_{n=0}^{t-1} \pi_n(d) X_{t-n}$.

The definition of fractional integration applied in (1), and in the remainder of this paper, is the so-called “type II” fractional integration. While “type II” is certainly not the only type of fractional integration, it does have the desirable feature that the same definition is valid for any value of the fractional parameter, d . Importantly, this implies that both stationary and nonstationary time series are permitted and that the range of admissible values of the fractional parameter can be arbitrarily large.

Specifically, this paper proves consistency and asymptotic normality results for CSS estimators in fractional time series models. Consistency results are important in their own right and are also necessary prerequisites in any proof of asymptotic normality for implicitly defined estimators such as the CSS estimator. However, proofs of consistency have been avoided in the literature due to the non-uniform convergence of the objective function.

To illustrate the issue, let the true value of the fractional integration parameter be denoted by d_0 . Then the data generating process is $X_t = \Delta_+^{-d_0} \varepsilon_t$, which is found by inverting (1), and residuals defined as $\Delta_+^d X_t = \Delta_+^{d-d_0} \varepsilon_t$ appear in the (conditional) likelihood or in the CSS objective function; see details below. Now, when $d-d_0 > -1/2$, the residuals are stationary (except for the truncation in the definition of Δ_+), and a law of large numbers can be combined with standard methods to obtain uniform convergence in probability of the CSS objective function in any compact subset of

$d - d_0 > -1/2$. On the other hand, when $d - d_0 < -1/2$ the residuals are nonstationary and a functional central limit theorem applies under additional moment conditions. Furthermore, the rate of convergence of the CSS objective function is different in this case, compared to $d - d_0 > -1/2$. This change in behavior of the objective function around the critical point $d - d_0 = -1/2$ implies that the objective function does not converge uniformly in probability on a large parameter space, i.e. one that includes this point, thus making consistency proofs on a large parameter space much more challenging than usual.

These difficulties have previously been avoided, e.g., by restricting the range of admissible values to an interval of length less than one-half as in, among others, Fox & Taqqu (1986), Dahlhaus (1989), Giraitis & Surgailis (1990), Hosoya (1996), and more recently Robinson (2006). Other works, e.g. Li & McLeod (1986) and Beran (1995), basically assume consistency in application of the usual Taylor expansion of the score function to derive the asymptotic distribution, while Tanaka (1999) and Nielsen (2004) give local consistency proofs. Alternatively, with some prior knowledge of the magnitude of d_0 one can fractionally difference the data, estimate d , and add back. See also Hualde & Robinson (2010, pp. 2-3) for additional discussion of these issues.

Only very recently, Hualde & Robinson (2010), Johansen & Nielsen (2010a), and Lieberman, Rosemarin & Rousseau (2010) have proven consistency for time domain estimators in parametric fractional time series models¹ with a large set of admissible values of d . The latter paper, however, considers only stationary processes, i.e. the interval $d < 1/2$, and do not allow nonstationary processes. Johansen & Nielsen (2010a) applies methods similar to the ones in this paper, but deals with an entirely different model that is not nested with the fractional models in this paper. In independent and concurrent work dealing with the same model, Hualde & Robinson (2010) gives a consistency proof that requires all moments finite for the innovations even in the simplest model (1). On the other hand, the proof in this paper requires only four moments for the model (1), and two additional moments for each additional parameter to be estimated in more sophisticated models.

The results in this paper apply to an arbitrarily large admissible parameter set for d , and hence it includes values of d where $\Delta_+^d X_t$ is asymptotically stationary, nonstationary, and also critical in the sense that $d - d_0 \simeq -1/2$. In the proof, each of these parts of the parameter space needs to be dealt with separately. In particular, the inclusion of the neighborhood around $d - d_0 = -1/2$ in the proof is achieved by a truncation argument, making it possible to show that when $v = d - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa_2]$ for (κ_1, κ_2) sufficiently small, then the inverse of product moments of critical processes $\Delta_+^v \varepsilon_t$ is tight in v , and further that it is convergent uniformly to zero in probability for $(T, \kappa_2) \rightarrow (\infty, 0)$.

¹In frequency domain estimation, consistency results for admissible parameter intervals of lengths greater than one-half are more common, for example Robinson (1995) and Shimotsu & Phillips (2005) for semiparametric estimation and Velasco & Robinson (2000) for tapered Whittle estimation.

Another complication in CSS estimation is the truncation of the autoregressive representation, which is needed both in the definition of the residual to calculate the objective function and in the “type II” definition of fractional integration. The additional complications arising from the truncation of the autoregressive filter have often been ignored in the literature; notable exceptions are Robinson (2005) and Robinson (2006). The present paper contains a rigorous treatment of the effects of the truncation.

Finally, all arguments, assumptions, proofs, etc., in the present paper are stated entirely in the time domain. This is somewhat remarkable for this literature, and demonstrates that frequency domain methods are not in fact necessary, or more simple for that matter, in proofs of asymptotic properties for fractional time series models.

The remainder of the paper is structured as follows. In the next section a consistency result is presented for a minimal fractionally integrated model under very simple assumptions. Section 3 presents the corresponding result for a general fractional time series model which allows a wide range of short memory innovations. In section 4 the asymptotic distribution theory is given and section 5 concludes. The two consistency proofs are quite involved and are presented in sections 6 and 7, respectively, followed by two sections of auxiliary results used in the main proofs.

Some comments on notation: for a sequence of stochastic processes $X_T(s) \in \mathbb{R}$, $s = (s_1, \dots, s_m) \in S$, where S is a compact subset of m -dimensional Euclidean space, the notation $X_T \Rightarrow X$ or $X_T(s) \Rightarrow X(s)$ is used to indicate convergence in distribution of the sequence, either as continuous processes in $\mathcal{C}(S)$ or as cadlag processes in $\mathcal{D}(S)$, whereas $X_T(s) \xrightarrow{D} X(s)$ means convergence in distribution in \mathbb{R} for a fixed s . The Euclidean norm is denoted $|\cdot|$ and when $E|X|^q < \infty$ the L_q -norm is defined as $\|X\|_q = (E|X|^q)^{1/q}$.

2 First consistency result

In this section the simple model (1) is analyzed under the following assumptions on the errors ε_t and the true parameter value, which is denoted by subscript zero.

Assumption A *The errors ε_t are i.i.d. with $E\varepsilon_t = 0$, $E\varepsilon_t^2 = \sigma_0^2 > 0$, and $E|\varepsilon_t|^q < \infty$ for some $q \geq 4$.*

Assumption B *The true value satisfies $d_0 \in D = [d_1, d_2]$ with $-\infty < d_1 \leq d_2 < \infty$.*

Importantly, only four moments are assumed finite in Assumption A. On the contrary, Hualde & Robinson (2010) assume that arbitrarily many moments are finite in their Assumption A3, which requires that a functional central limit theorem holds for $\Delta_+^{d-d_0}\varepsilon_t$ for $d - d_0 \leq -1/2 - \eta$ for arbitrarily small $\eta > 0$. However, a necessary condition for such a functional central limit theorem is that $E|\varepsilon_t|^q < \infty$ for $q \geq \eta^{-1}$, see Johansen & Nielsen (2010c). Since $\eta > 0$ can be arbitrarily small, all moments of ε_t are required to exist.

Assumption B permits the length of the interval D of admissible values of d to be arbitrarily large. Specifically, the length of D is not limited to less than $1/2$ as in

most previous studies of fractional models that include proofs of consistency. Thus, under Assumption B, the model can simultaneously accommodate both nonstationary, stationary, and overdifferenced processes.

The conditional log-likelihood function is given by

$$L(d, \sigma^2) = -\frac{T}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^T (\Delta_+^d X_t)^2, \quad (3)$$

and the profile log-likelihood per observation (apart from a constant) is

$$L(d) = -\frac{1}{2} \log \left(T^{-1} \sum_{t=1}^T (\Delta_+^d X_t)^2 \right). \quad (4)$$

The conditional maximum likelihood estimator (CMLE) is found as

$$\hat{d} = \arg \max_{d \in D} L(d), \quad (5)$$

which is the same as minimizing the conditional sum of squared residuals, i.e.,

$$\hat{d} = \arg \min_{d \in D} RSS(d), \quad (6)$$

$$RSS(d) = T^{-1} \sum_{t=1}^T (\Delta_+^d X_t)^2. \quad (7)$$

The estimator (6) is usually denoted the conditional-sum-of-squares (CSS) estimator, although for both (5) and (6) truncated-sum-of-squares may be more appropriate as remarked by Hualde & Robinson (2010). The estimator is well-known from e.g. Li & McLeod (1986) and Beran (1995) in a fractional context and of course Box & Jenkins (1970) for non-fractional models.

The first main result is the following.

Theorem 1 *Suppose X_t is generated by model (1) with errors satisfying Assumption A and true value satisfying Assumption B, and let the estimator \hat{d} be defined by either (5) or equivalently (6). Then $\hat{d} \xrightarrow{P} d_0$ as $T \rightarrow \infty$.*

3 Second consistency result

Now let the model be given by

$$X_t = \Delta_+^{-d} u_t, \quad (8)$$

$$u_t = a(L, \psi) \varepsilon_t, \quad (9)$$

where ψ is a p -dimensional parameter vector and

$$a(z, \psi) = \sum_{n=0}^{\infty} a_n(\psi) z^n. \quad (10)$$

The model (8)-(9) generalizes model (1) to allow short-run dynamics in u_t . Specifically, u_t is assumed to be a linear process governed by an underlying p -dimensional parameter vector. For example, u_t could be generated by an ARMA model or by the exponential spectrum model of Bloomfield (1973), which is somewhat popular in the fractional literature owing to the neat covariance matrix formula it offers in this setting, see Robinson (1994).

The earlier Assumptions A and B are extended to:

Assumption A' *Assumption A is satisfied with $q \geq 2p + 4$.*

Assumption B' *The true values satisfy $(d_0, \psi_0) \in D \times \Psi$, where D is defined in Assumption B and the set $\Psi \subseteq \mathbb{R}^p$ is convex and compact.*

The moment condition in Assumption A' is stronger than the one in Assumption A, but is needed to apply the moment-bound tightness arguments to obtain uniform convergence of the objective function in the consistency proof. Note, though, that still only a limited number of moments are assumed finite unlike in Hualde & Robinson (2010). The parameter space for d in Assumption B' is the same as in Assumption B, and the short-run parameters ψ are assumed to be in a compact and convex subset of \mathbb{R}^p .

The following regularity condition is imposed on the coefficients of the linear filter $a(z, \psi)$:

Assumption C

(i) *For all $\psi \in \Psi$ it holds that $a_0(\psi) = 1$ and the function $a(z, \psi) = \sum_{n=0}^{\infty} a_n(\psi)z^n$ is bounded and bounded away from zero on the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$.*

(ii) *For all $\psi \in \Psi$ the coefficients $a_n(\psi)$ satisfy $\sum_{n=0}^{\infty} n|a_n(\psi)| < \infty$.*

(iii) *For all z the function $a(z, \psi) = \sum_{n=0}^{\infty} a_n(\psi)z^n$ is continuously differentiable in ψ and the derivatives $\dot{a}_n(\psi) = \frac{\partial a_n(\psi)}{\partial \psi}$ satisfy $\sum_{n=0}^{\infty} |\dot{a}_n(\psi)| < \infty$ for all $\psi \in \Psi$.*

Assumption C(i) ensures invertibility of the model (9). Under this assumption the function $b(z, \psi) = a(z, \psi)^{-1}$ is well-defined by its power series expansion for $|z| \leq 1 + \delta$ for some $\delta > 0$, and is also bounded and bounded away from zero on the complex unit disk. Thus, letting $b(z, \psi) = \sum_{n=0}^{\infty} b_n(\psi)z^n$ and defining the residuals

$$\varepsilon_t(d, \psi) = \sum_{n=0}^{t-1} b_n(\psi) \Delta_+^d X_{t-n}, \quad (11)$$

the CSS estimator (which is equivalent to the CMLE) is found as

$$(\hat{d}, \hat{\psi}) = \arg \min_{(d, \psi) \in D \times \Psi} RSS(d, \psi), \quad (12)$$

$$RSS(d, \psi) = T^{-1} \sum_{t=1}^T \varepsilon_t(d, \psi)^2. \quad (13)$$

Note the truncation of the autoregressive representation of X_t in calculating the residual in (11), which has often been ignored in the literature. Notable exceptions are Robinson (2005) and Robinson (2006), who include detailed treatments of the consequences of the truncation. However, Robinson (2005) does not consider consistency for the estimation of d , and the consistency proof in Robinson (2006) restricts the length of the interval D to be less than $1/2$.

Assumptions C(ii)-(iii) are easily satisfied by the Bloomfield model or by stationary and invertible ARMA processes due to the exponential decay of their linear representation coefficients. Instead of Assumption C(ii), Hualde & Robinson (2010) make the assumption that $a(e^{i\lambda}, \psi)$ is differentiable in λ with derivative in $Lip(\zeta)$ for $\zeta > 1/2$. Assumption C(ii) would follow if, e.g., $a(e^{i\lambda}, \psi)$ were twice differentiable in λ with second order derivative in $Lip(\zeta)$ for $\zeta > 0$, in which case $a_n(\psi) = O(n^{-2-\zeta})$, see Zygmund (2003, p. 71).

Finally, the following identification condition will also be needed:

Assumption D For all $\psi \in \Psi \setminus \{\psi_0\}$ it holds that $a(z, \psi) \neq a(z, \psi_0)$ on a subset of $\{z \in \mathbb{C} : |z| = 1\}$ of positive measure.

Assumption D is satisfied, for example, by all stationary and invertible ARMA processes whose AR and MA polynomials are not both overspecified.

The main result of this section is stated in the following theorem.

Theorem 2 Suppose X_t is generated by model (8)-(9) satisfying Assumptions A', B', C, and D, and let the estimator $(\hat{d}, \hat{\psi})$ be defined by (12). Then $(\hat{d}, \hat{\psi}) \xrightarrow{P} (d_0, \psi_0)$ as $T \rightarrow \infty$.

4 Asymptotic distribution theory

For the simple model (1) no additional assumptions are needed to prove asymptotic normality, except having the true value lie in the interior of the parameter space.

Theorem 3 Under the assumptions of Theorem 1 and, in addition, $d_0 \in \text{int}(D)$,

$$\sqrt{T}(\hat{d} - d_0) \xrightarrow{D} N(0, 6/\pi^2) \text{ as } T \rightarrow \infty.$$

Proof. By consistency of \hat{d} , the asymptotic distribution theory for the CSS estimator is obtained from the usual Taylor series expansion of the score function. That is,

$$0 = T^{1/2} \frac{\partial RSS(\hat{d})}{\partial d} = T^{1/2} \frac{\partial RSS(d_0)}{\partial d} + T^{1/2}(\hat{d} - d_0) \frac{\partial^2 RSS(\bar{d})}{\partial d^2}, \quad (14)$$

where \bar{d} is an intermediate value satisfying $|\bar{d} - d_0| \leq |\hat{d} - d_0|$.

Using (16) and the simple fact that $\pi_0(d) = 1$ for all d , the normalized score function evaluated at the true value is

$$T^{1/2} \frac{\partial RSS(d_0)}{\partial d} = 2T^{-1/2} \sum_{t=1}^T \varepsilon_t y_{1,t-1},$$

where $y_{k,t-1} = \sum_{n=1}^{t-1} \frac{\partial^k \pi_n(d_0-d)}{\partial d^k} \Big|_{d=d_0} \varepsilon_{t-n}$, $k = 1, 2$. Furthermore, $\frac{\partial \pi_n(u)}{\partial d} \Big|_{u=0} = n^{-1}$ for $n \geq 1$ such that $y_{1,t-1} = -\sum_{n=1}^{t-1} n^{-1} \varepsilon_{t-n}$. Define also the stationary and ergodic (untruncated) process $z_{1,t-1} = -\sum_{n=1}^{\infty} n^{-1} \varepsilon_{t-n}$, which is measurable with respect to the sigma-algebra $\mathcal{F}_{t-1} = \sigma(\{\varepsilon_1, \dots, \varepsilon_{t-1}\})$ such that $\varepsilon_t z_{1,t-1}$ is a martingale difference sequence with respect to the filtration \mathcal{F}_{t-1} . By the law of large numbers for stationary and ergodic processes the sum of conditional variances is $T^{-1} \sum_{t=1}^T E(\varepsilon_t^2 z_{1,t-1}^2 | \mathcal{F}_{t-1}) = \sigma_0^2 T^{-1} \sum_{t=1}^T z_{1,t-1}^2 \xrightarrow{P} \sigma_0^4 \pi^2 / 6$ because $\sum_{n=1}^{\infty} n^{-2} = \pi^2 / 6$. The Lindeberg condition is satisfied since $\varepsilon_t z_{1,t-1}$ is stationary with finite variance. It therefore follows from the central limit theorem for martingales, e.g. Hall & Heyde (1980, chp. 3), that $T^{-1/2} \sum_{t=1}^T \varepsilon_t z_{1,t-1} \xrightarrow{D} N(0, \sigma_0^4 \pi^2 / 6)$. Noting that $y_{1,t-1} - z_{1,t-1} = \sum_{n=t}^{\infty} n^{-1} \varepsilon_{t-n}$ and $E(y_{1,t-1} - z_{1,t-1})^2 = \sigma_0^2 \sum_{n=t}^{\infty} n^{-2} \leq ct^{-1}$, it follows that

$$T^{1/2} \frac{\partial RSS(d_0)}{\partial d} = 2T^{-1/2} \sum_{t=1}^T \varepsilon_t z_{1,t-1} + o_P(1) \xrightarrow{D} N(0, 4\sigma_0^4 \pi^2 / 6).$$

Because the second derivative is tight in a neighborhood of d_0 (by the criterion (21) using (46) of Lemma 5) and \hat{d} is consistent (by Theorem 1) it follows from Lemma A.3 of Johansen & Nielsen (2010b) that the second derivative in (14) can be evaluated at d_0 . Hence,

$$\frac{\partial^2 RSS(d_0)}{\partial d^2} = 2T^{-1} \sum_{t=1}^T \varepsilon_t y_{2,t-1} + 2T^{-1} \sum_{t=1}^T y_{1,t-1}^2,$$

where $\varepsilon_t y_{2,t-1}$ is a martingale difference sequence such that the first term on the right-hand side is $O_P(T^{-1/2})$. For the second term, apply the same methods as above to obtain $2T^{-1} \sum_{t=1}^T y_{1,t-1}^2 = 2T^{-1} \sum_{t=1}^T z_{1,t-1}^2 + o_P(1) \xrightarrow{P} 2\sigma_0^2 \pi^2 / 6$, which proves the result. ■

For the general model (8)-(9), the smoothness conditions on the coefficients need to be strengthened and an additional condition is needed to ensure that the asymptotic variance matrix of the estimator is well-defined. As above, these conditions are easily satisfied, e.g. by the Bloomfield model or by stationary and invertible ARMA processes.

Assumption E For all z , $a(z, \psi) = \sum_{n=0}^{\infty} a_n(\psi) z^n$ is three times differentiable in ψ and the derivatives $\frac{\partial^k a_n(\psi)}{\partial \psi^{(k)}}$ satisfy $\sum_{n=0}^{\infty} |\frac{\partial^k a_n(\psi)}{\partial \psi^{(k)}}| < \infty$ for all $\psi \in \Psi$ and $k = 1, 2, 3$.

Assumption F The matrix

$$\Sigma_0 = \begin{bmatrix} \pi^2/6 & -\sum_{n=1}^{\infty} \gamma_n(\psi_0)' / n \\ -\sum_{n=1}^{\infty} \gamma_n(\psi_0) / n & \sum_{n=1}^{\infty} \gamma_n(\psi_0) \gamma_n(\psi_0)' \end{bmatrix}$$

is non-singular, where $\gamma_n(\psi) = \sum_{m=0}^{n-1} a_m(\psi) \dot{b}_{n-m}(\psi)$.

Theorem 4 Under the assumptions of Theorem 2, Assumptions E-F, and $(d_0, \psi_0) \in \text{int}(D \times \Psi)$,

$$\sqrt{T}((\hat{d}, \hat{\psi}) - (d_0, \psi_0)) \xrightarrow{D} N(0, \Sigma_0^{-1}) \text{ as } T \rightarrow \infty.$$

Proof. Let $\theta = (d, \psi)$. Apply the Taylor series expansion

$$0 = T^{1/2} \frac{\partial RSS(\hat{\theta})}{\partial \theta} = T^{1/2} \frac{\partial RSS(\theta_0)}{\partial \theta} + T^{1/2} (\hat{\theta} - \theta_0) \frac{\partial^2 RSS(\bar{\theta})}{\partial \theta \partial \theta'} \quad (15)$$

and examine the score and second derivative.

The normalized score is

$$T^{1/2} \frac{\partial RSS(\theta_0)}{\partial \theta} = 2T^{-1/2} \sum_{t=1}^T \varepsilon_t(d_0, \psi_0) y_{1,t-1} \text{ with } y_{1,t-1} = \frac{\partial}{\partial \theta} \varepsilon_t(d, \psi)|_{\theta=\theta_0}.$$

Define also

$$S_T = 2T^{-1/2} \sum_{t=1}^T \varepsilon_t z_{1,t-1} \text{ with } z_{1,t-1} = \frac{\partial}{\partial \theta} \Delta^{d-d_0} e_t(\psi)|_{\theta=\theta_0}.$$

It is shown in Robinson (2006, pp. 135-136) that $T^{1/2} \frac{\partial RSS(\theta_0)}{\partial \theta} - S_T = o_P(1)$ under the assumptions of Theorem 4. The first element of $z_{1,t-1}$ is $-\sum_{n=1}^{\infty} n^{-1} \varepsilon_{t-n}$ and the remaining p elements are given by $\sum_{n=1}^{\infty} \dot{b}_n(\psi_0) u_{t-n} = \sum_{n=1}^{\infty} \gamma_n(\psi_0) \varepsilon_{t-n}$. By Assumption A, $\nu_t = \varepsilon_t z_{1,t-1}$ is a martingale difference sequence with respect to \mathcal{F}_{t-1} and by the law of large numbers for stationary and ergodic processes the sum of conditional variances is $T^{-1} \sum_{t=1}^T E(\nu_t \nu_t' | \mathcal{F}_{t-1}) = \sigma_0^2 T^{-1} \sum_{t=1}^T z_{1,t-1} z_{1,t-1}' \xrightarrow{P} \sigma_0^4 \Sigma_0$. It follows from the central limit theorem for martingales that $S_T \xrightarrow{D} N(0, 4\sigma_0^4 \Sigma_0)$ and therefore also $T^{1/2} \frac{\partial RSS(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, 4\sigma_0^4 \Sigma_0)$.

The second derivative in (15) can be evaluated at the true value for the same reason as in the proof of Theorem 3 because it is tight by the criterion (33) using (51) of Lemma 6. Furthermore, from the argument in Robinson (2006, pp. 135-136) it is enough to consider $H_T = 2T^{-1} \sum_{t=1}^T \varepsilon_t z_{2,t-1} + 2T^{-1} \sum_{t=1}^T z_{1,t-1} z_{1,t-1}'$, where $z_{2,t-1} = \frac{\partial^2}{\partial \theta \partial \theta'} \Delta^{d-d_0} e_t(\psi)|_{\theta=\theta_0}$. Because $\varepsilon_t z_{2,t-1}$ is a martingale difference sequence with respect to \mathcal{F}_{t-1} the first term is $O_P(T^{-1/2})$. By the law of large numbers, it follows that $2T^{-1} \sum_{t=1}^T z_{1,t-1} z_{1,t-1}' \xrightarrow{P} 2\sigma_0^2 \Sigma_0$, which proves the result. ■

Theorems 3 and 4 prove the anticipated result that the CSS estimator has the same asymptotic distribution as the maximum likelihood estimator. For similar asymptotic distribution results for the CSS estimator, see also Beran (1995), Tanaka (1999), Nielsen (2004), and Hualde & Robinson (2010). An important consequence of these results is that the CSS estimator is efficient under Gaussianity, c.f. Fox & Taqqu (1986), Dahlhaus (1989). However, the asymptotic normality results in Theorems 3 and 4 are valid much more generally because Gaussianity is not assumed in this paper.

5 Conclusion

This paper has proven consistency and asymptotic normality for the CSS estimator in general parametric fractional time series models. The consistency proof is important in its own right and also because it is a required prerequisite for a proof of asymptotic normality. The only other proof of consistency for the CSS estimator in fractional time series models which applies to a large set of admissible parameter values is that of Hualde & Robinson (2010), and the proof in this paper distinguishes itself from their proof by requiring only a finite number of moments of the innovation process.

Finally, a multivariate extension of the results in this paper should be straightforward requiring additional and more complicated notation, but does not change the strategy and method of proof. In fact, Johansen & Nielsen (2010a) applies methods similar to the ones in this paper to prove consistency in a multivariate fractionally cointegrated model, although their model is not nested with – and indeed is entirely different from – the models considered in this paper. This is in contrast with Hualde & Robinson (2010), whose consistency proof applies only to the univariate case (see their discussion on p. 19 and p. 21) – and requires all moments finite.

6 Proof of Theorem 1

Defining the residual $\varepsilon_t(d) = \Delta_+^d X_t = \Delta_+^{d-d_0} \varepsilon_t$, the objective function can be rewritten as

$$RSS(d) = T^{-1} \sum_{t=1}^T \varepsilon_t(d)^2, \quad (16)$$

and it is clear that the convergence properties of $RSS(d)$ will depend on $d - d_0$. Let the deterministic function $r(d)$ denote the pointwise probability limit of $RSS(d)$, shown subsequently to be given by

$$r(d) = \begin{cases} E(\Delta^{d-d_0} \varepsilon_t)^2 & \text{if } d - d_0 > -1/2, \\ \infty & \text{if } d - d_0 \leq -1/2. \end{cases} \quad (17)$$

According to (17) the interval D is partitioned into three disjoint compact subintervals $D_1 = D_1(\kappa_1) = D \cap \{d : d - d_0 \leq -1/2 - \kappa_1\}$, $D_2 = D_2(\kappa_1, \kappa_2) = D \cap \{-1/2 - \kappa_1 \leq d - d_0 \leq -1/2 + \kappa_2\}$, and $D_3 = D_3(\kappa_2) = D \cap \{d : d - d_0 \geq -1/2 + \kappa_2\}$, for some constants $\kappa_i \in (0, 1/2)$, $i = 1, 2$, to be determined later. Here, special care is taken with respect to D_2 , where the convergence of the objective function is non-uniform.

First, it is shown that for any $K > 0$ there exists a $\kappa_2 = \kappa_2(K)$ such that

$$\Pr(\inf_{d \in D_1 \cup D_2} RSS(d) > K) \rightarrow 1 \text{ as } T \rightarrow \infty. \quad (18)$$

This implies that $\Pr(\hat{d} \in D_1 \cup D_2) \rightarrow 0$ as $T \rightarrow \infty$ and hence $\Pr(\hat{d} \in D_3) \rightarrow 1$ as $T \rightarrow \infty$, so that the relevant parameter space is reduced to $D_3(\kappa_2)$. From Theorem

5.7 of van der Vaart (1998) the desired result then follows if

$$\sup_{d \in D_3} |RSS(d) - r(d)| \xrightarrow{P} 0 \text{ as } T \rightarrow \infty, \quad (19)$$

$$\inf_{d \in D_3 \cap \{d: |d-d_0| \geq \delta\}} r(d) > r(d_0) \text{ for all } \delta > 0. \quad (20)$$

The first condition entails uniform convergence of the objective function on D_3 , and the second condition ensures that the optimum of the limit function is uniquely attained at the true value.

To show uniform convergence it is sufficient to prove pointwise convergence as $T \rightarrow \infty$ and tightness (or stochastic equicontinuity) of the process as a function of d . Tightness is proven using the moment condition in Billingsley (1968, Theorem 12.3), which requires showing that $RSS(d_0)$ is tight and that

$$\|RSS(d_1) - RSS(d_2)\|_2 \leq c|d_1 - d_2| \quad (21)$$

for some constant $c > 0$ that does not depend on T , d_1 , or d_2 .

6.1 Convergence on $D_1(\kappa_1)$

If $d \in D_1(\kappa_1)$ then $\varepsilon_t(d)$ is nonstationary and the functional central limit theorem of Marinucci & Robinson (2000) applies if also $E|\varepsilon_t|^q < \infty$ for some $q > \max(2, (d_0 - d - 1/2)^{-1})$. Thus, as $T \rightarrow \infty$,

$$T^{d-d_0+1/2} \Delta_+^{d-d_0} \varepsilon_{[Tr]} \Rightarrow W_{d_0-d-1}(r) = \Gamma(d_0 - d)^{-1} \int_0^r (r-s)^{d_0-d-1} dW(s) \text{ in } \mathcal{D}[0, 1], \quad (22)$$

where W denotes Brownian motion generated by ε_t and W_{d_0-d-1} is the corresponding fractional Brownian motion of type II. By the continuous mapping theorem, as $T \rightarrow \infty$,

$$T^{2(d-d_0)+1} RSS(d) \xrightarrow{D} \int_0^1 W_{d_0-d-1}(r)^2 dr \quad (23)$$

for fixed d , which shows the pointwise limit.

To show that $T^{2(d-d_0)+1} RSS(d)$ is tight on $d \in D_1$ note that the pointwise convergence implies tightness for any fixed $d \in D_1$. Tightness of $T^{2(d-d_0)+1} RSS(d)$ then follows upon verifying the tightness condition (21), which is satisfied by (45) of Lemma 5. Hence the convergence in (23) is strengthened to

$$T^{2(d-d_0)+1} RSS(d) \Rightarrow \int_0^1 W_{d_0-d-1}(r)^2 dr \text{ in } \mathcal{C}(D_1).$$

By the continuous mapping theorem applied to the $\inf_{d \in D_1}$ mapping, which is continuous because D_1 is compact, it then holds that

$$\inf_{d \in D_1} T^{2(d-d_0)+1} RSS(d) \xrightarrow{D} \inf_{d \in D_1} \int_0^1 W_{d_0-d-1}(r)^2 dr,$$

which is positive almost surely. It follows that, for any $K > 0$, $\Pr(\inf_{d \in D_1} RSS(d) > K) \rightarrow 1$ because $2(d - d_0) + 1 \leq -2\kappa_1 < 0$.

It remains only to discuss the choice of κ_1 . In order for (22) to hold (pointwise) for all $d \in D_1(\kappa_1)$ it is necessary to assume $q > \max(2, \kappa_1^{-1})$ moments. Thus, to avoid a very strong moment condition, κ_1 cannot be too small, and specifically it is necessary to choose $\kappa_1 > 1/4$ in view of Assumption A. On the contrary, in the consistency proof in Hualde & Robinson (2010), their analysis of the intervals around $d - d_0 = -1/2$ requires them to choose κ_1 arbitrarily small, thus needing arbitrarily many moments of ε_t , c.f. Johansen & Nielsen (2010c).

6.2 Convergence on $D_2(\kappa_1, \kappa_2)$

Let $v = d - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa_2]$ and decompose $\Delta_+^v \varepsilon_t$ as

$$\Delta_+^v \varepsilon_t = \sum_{n=0}^{N-1} \pi_n(-v) \varepsilon_{t-n} + \sum_{n=N}^{t-1} \pi_n(-v) \varepsilon_{t-n} = w_{1t} + w_{2t}$$

for some $N \geq 1$ to be determined. It then holds that

$$\begin{aligned} T^{-1} \sum_{t=1}^T (\Delta_+^v \varepsilon_t)^2 &= T^{-1} \sum_{t=1}^T w_{1t}^2 + T^{-1} \sum_{t=1}^T w_{2t}^2 + 2T^{-1} \sum_{t=1}^T w_{1t} w_{2t} \\ &\geq T^{-1} \sum_{t=1}^T w_{1t}^2 + 2T^{-1} \sum_{t=1}^T w_{1t} w_{2t}, \end{aligned} \quad (24)$$

where the second term multiplied by $(\log T)^{-2} T^{1/2 - \kappa_1} N^{-1/2 - \kappa_1}$ is tight and $O_P(1)$ by (52) and (53) of Lemma 7. For the first term, (54) and (55) of Lemma 7 show that

$$(\log T)^{-2} T^{1/2} N^{-1/2 - 2\kappa_1} \left(T^{-1} \sum_{t=1}^T w_{1t}^2 - \sigma_0^2 T^{-1} (T - N) F_N(v) \right) = O_P(1)$$

and that the left-hand side is tight, where $F_N(v) = \sum_{n=0}^N \pi_n(-v)^2$. Setting $N = T^\alpha$ with $0 < \alpha < \min(\frac{1/2 - \kappa_1}{1/2 + \kappa_1}, \frac{1/2}{1/2 + 2\kappa_1})$, and hence $\kappa_1 < 1/2$, thus shows that the right-hand side of (24) minus $\sigma_0^2 T^{-1} (T - N) F_N(v)$ converges uniformly in probability to zero as $T \rightarrow \infty$.

6.3 Proof of (18): $\Pr(\inf_{d \in D_1 \cup D_2} RSS(d) > K) \rightarrow 1$ as $T \rightarrow \infty$

The problem is to show that, for any $K > 0, \eta > 0$, there exists a κ_2 and a T_0 such that

$$\Pr\left(\inf_{d \in D_1 \cup D_2} RSS(d) < K\right) \leq \eta$$

for all $T \geq T_0$. Since

$$\inf_{d \in D_1 \cup D_2} RSS(d) \leq \sum_{j=1}^2 \inf_{d \in D_j} RSS(d),$$

the two sets D_1 and D_2 can be considered in turn.

First consider the interval $D_1(\kappa_1)$ with $\kappa_1 = 1/3 > 1/4$, and define $\bar{D}_1 = D_1(1/3)$. It holds from section 6.1 that $\Pr(\inf_{d \in \bar{D}_1} RSS(d) > K) \rightarrow 1$ as $T \rightarrow \infty$, i.e., for any $K > 0, \eta > 0$, there exists a T_1 such that

$$\Pr(\inf_{d \in \bar{D}_1} RSS(d) < K) \leq \eta/2$$

for all $T \geq T_1$.

Second, having already fixed $\kappa_1 = 1/3$, consider $D_2(1/3, \kappa_2)$. From section 6.2 with $\kappa_1 = 1/3$ and $\alpha = 1/6$,

$$RSS(d) \geq \sigma_0^2 T^{-1} (T - T^{1/6}) F_{T^{1/6}}(d - d_0) + \mu_T(d),$$

where $\mu_T(d) \xrightarrow{P} 0$ as $T \rightarrow \infty$ uniformly in $d \in D_2(1/3, \kappa_2)$. From Lemma 4(b), $F_{T^{1/6}}(d - d_0) \geq 1 + c(1 - T^{-\kappa_2/6})/(\kappa_2/6) \rightarrow \infty$ as $(\kappa_2, T) \rightarrow (0, \infty)$. It follows that for any $K > 0, \eta > 0$, there exists $\bar{\kappa}_2$ (small) and T_2 such that

$$\Pr(\inf_{d \in \bar{D}_2} RSS(d) < K) \leq \eta/2$$

for all $T \geq T_2$ and $\bar{D}_2 = D_2(1/3, \bar{\kappa}_2)$. Thus, κ_1 needs to be relatively large (and fixed) to allow the weakest possible moment condition in the analysis of D_1 , whereas κ_2 needs to be very small ($\rightarrow 0$) in the analysis of D_2 . This illustrates the importance of the asymmetry of the interval D_2 around $d - d_0 = -1/2$.

Combining these results, for any $K > 0, \eta > 0$,

$$\Pr(\inf_{d \in \bar{D}_1 \cup \bar{D}_2} RSS(d) < K) \leq \sum_{j=1}^2 \Pr(\inf_{d \in \bar{D}_j} RSS(d) < K) \leq \sum_{j=1}^2 \eta/2 = \eta$$

for all $T \geq \max(T_1, T_2) = T_0$, which proves (18) for $\kappa_2 = \bar{\kappa}_2$.

6.4 Convergence on $D_3(\kappa_2)$ and proof of (19): uniform convergence

If $d \in D_3$ then $\varepsilon_t(d)$ is asymptotically stationary for any $\kappa_2 > 0$. In this case, define also the stationary and ergodic process $\eta_t(d) = \Delta^{d-d_0} \varepsilon_t$ (without truncation) and note that $\eta_t(d) - \varepsilon_t(d) = \sum_{n=t}^{\infty} \pi_n(d_0 - d) \varepsilon_{t-n}$ with $E(\eta_t(d) - \varepsilon_t(d))^2 = \sigma_0^2 \sum_{n=t}^{\infty} \pi_n(d_0 - d)^2 \leq c \sum_{n=t}^{\infty} n^{2(d_0-d)-2} \rightarrow 0$. It follows from the law of large numbers for stationary and ergodic processes that

$$RSS(d) = T^{-1} \sum_{t=1}^T \eta_t(d)^2 + o_P(1) \xrightarrow{P} E(\Delta^{d-d_0} \varepsilon_t)^2 \text{ as } T \rightarrow \infty, \quad (25)$$

which shows the pointwise limit in probability.

Because $RSS(d_0) = T^{-1} \sum_{t=1}^T \varepsilon_t^2 \xrightarrow{P} \sigma_0^2$ as $T \rightarrow \infty$ by a law of large numbers, it is tight. Thus, tightness of $RSS(d)$ follows from condition (21) which is satisfied by (46) of Lemma 5.

Since this result holds for any κ_2 it holds specifically for $\kappa_2 = \bar{\kappa}_2$ and therefore the convergence in probability is uniform on $d \in \bar{D}_3 = D_3(\bar{\kappa}_2)$. This proves (19).

6.5 Proof of (20): uniqueness

Since $r(d_0) = \sigma_0^2$ it is sufficient to prove that $\inf_{d \in D_3 \cap \{d: |d-d_0| \geq \delta\}} E(\Delta^{d-d_0} \varepsilon_t)^2 > \sigma_0^2$ for all $\delta > 0$. The variance of $\Delta^{d-d_0} \varepsilon_t = \sum_{n=0}^{\infty} \pi_n(d_0 - d) \varepsilon_{t-n}$ is $E(\Delta^{d-d_0} \varepsilon_t)^2 = \sigma_0^2 \sum_{n=0}^{\infty} \pi_n(d_0 - d)^2$. Because $\sum_{n=0}^{\infty} \pi_n(d_0 - d)^2 = 1 + \sum_{n=1}^{\infty} \pi_n(d_0 - d)^2 \geq 1$ with equality if and only if $d = d_0$, this proves (20) by continuity of $\pi_n(\cdot)$ and compactness of D_3 .

7 Proof of Theorem 2

The proof of this theorem follows the same general line of reasoning as the proof of Theorem 1 with the appropriate additional arguments to deal with the added complexity from the extra parameters in the linear process innovations. Let $\theta = (d, \psi) \in D \times \Psi = \Theta$, where the parameter space Θ is partitioned into the subsets $\Theta_j = D_j \times \Psi$ for $j = 1, 2, 3$, where D_j is defined as in the proof of Theorem 1. The objective function and its pointwise limit are

$$RSS(\theta) = T^{-1} \sum_{t=1}^T \varepsilon_t(\theta)^2, \quad (26)$$

$$r(\theta) = \begin{cases} E(\Delta^{d-d_0} e_t(\psi))^2 & \text{if } d - d_0 > -1/2, \\ \infty & \text{if } d - d_0 \leq -1/2, \end{cases} \quad (27)$$

where the residual is given by $\varepsilon_t(\theta) = \sum_{n=0}^{t-1} b_n(\psi) \Delta_+^{d-d_0} u_{t-n}$ and the untruncated process $e_t(\psi) = c(L, \psi) \varepsilon_t$ is defined using

$$c(z, \psi) = b(z, \psi) a(z, \psi_0) = \frac{a(z, \psi_0)}{a(z, \psi)} = \sum_{n=0}^{\infty} c_n(\psi) z^n. \quad (28)$$

From Assumption C the coefficients $c_n(\psi)$ satisfy

$$\sum_{n=0}^{\infty} n |c_n(\psi)| < \infty \text{ uniformly in } \psi \in \Psi. \quad (29)$$

As in the proof of Theorem 1 it is shown subsequently that for any $K > 0$ there exists a $\kappa_2 = \kappa_2(K)$ such that

$$\Pr\left(\inf_{\theta \in \Theta_1 \cup \Theta_2} RSS(\theta) > K\right) \rightarrow 1 \text{ as } T \rightarrow \infty, \quad (30)$$

which implies that $\Pr(\hat{\theta} \in \Theta_3) \rightarrow 1$ as $T \rightarrow \infty$. From Theorem 5.7 of van der Vaart (1998) the desired result follows if

$$\sup_{\theta \in \Theta_3} |RSS(\theta) - r(\theta)| \xrightarrow{P} 0 \text{ as } T \rightarrow \infty, \quad (31)$$

$$\inf_{\theta \in \Theta_3 \cap \{\theta: |\theta - \theta_0| \geq \delta\}} r(\theta) > r(\theta_0) \text{ for all } \delta > 0. \quad (32)$$

Tightness is now proven using the multiparameter version of (21) given in Kallenberg (2001, Corollary 16.9), which requires showing that $RSS(\theta_0)$ is tight and that

$$\|RSS(\theta_1) - RSS(\theta_2)\|_{p+2} \leq c|\theta_1 - \theta_2| \quad (33)$$

for some constant $c > 0$ that does not depend on T , θ_1 , or θ_2 .

The following lemma shows that the problem can be simplified by considering the sum of squares of $\Delta_+^{d-d_0}e_t(\psi)$ rather than $\varepsilon_t(\theta)$ in the analysis of $RSS(\theta)$. This is the important lemma that analyzes the effect of the truncation in the residual in the definition of $RSS(\theta)$. Note that this result requires $\kappa_1 < 1/4$ as opposed to $\kappa_1 > 1/4$ in the moment condition in section 6.1.

Lemma 1 *With the notation of this section and under the assumptions of Theorem 2 and $\kappa_1 < 1/4$ it holds that, as $T \rightarrow \infty$,*

$$\sup_{\theta \in \Theta_1} \left| T^{2(d-d_0)} \sum_{t=1}^T \varepsilon_t(\theta)^2 - T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0}e_t(\psi))^2 \right| \xrightarrow{P} 0, \quad (34)$$

$$\sup_{\theta \in \Theta_2 \cup \Theta_3} \left| T^{-1} \sum_{t=1}^T \varepsilon_t(\theta)^2 - T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0}e_t(\psi))^2 \right| \xrightarrow{P} 0. \quad (35)$$

Proof. First it is shown that

$$\sup_{\psi \in \Psi} \sum_{t=1}^T (\varepsilon_t(\theta) - \Delta_+^{d-d_0}e_t(\psi))^2 = O_P((\log T)^2 T^{2 \max(d_0-d-1/2, 0)}). \quad (36)$$

Note that

$$\varepsilon_t(\theta) - \Delta_+^{d-d_0}e_t(\psi) = \sum_{j=0}^{t-1} \sum_{n=t-j}^{\infty} \pi_j(d_0-d)b_n(\psi)u_{t-j-n} = \sum_{m=t}^{\infty} \phi_{tm}u_{t-m},$$

where $\phi_{tm} = \sum_{k=0}^{\min(m-t, t-1)} \pi_k(d_0-d)b_{m-k}(\psi)$ satisfies, see Lemma 3 and Assumption C,

$$\begin{aligned} \sum_{m=t}^{\infty} \sup_{\psi \in \Psi} |\phi_{tm}| &\leq c \sum_{m=t}^{\infty} \sum_{k=1}^{\min(m-t, t-1)} k^{d_0-d-1} (m-k)^{-2} \\ &\leq c \sum_{k=1}^{t-1} k^{d_0-d-1} (t-k)^{-1} \leq c(1 + \log t) t^{\max(d_0-d-1, -1)}. \end{aligned}$$

The left-hand side of (36) is

$$\begin{aligned} \sup_{\psi \in \Psi} \sum_{t=1}^T \sum_{m=t}^{\infty} \sum_{n=t}^{\infty} \phi_{tm} \phi_{tn} u_{t-n} u_{t-m} &= O_P \left(\sum_{t=1}^T \left(\sum_{m=t}^{\infty} \sup_{\psi \in \Psi} |\phi_{tm}| \right)^2 \right) \\ &= O_P \left(\sum_{t=1}^T (1 + \log t)^2 t^{2 \max(d_0-d-1, -1)} \right) \\ &= O_P \left((\log T)^2 T^{2 \max(d_0-d-1/2, 0)} \right). \end{aligned}$$

To prove (34) and (35) write

$$\begin{aligned} \sum_{t=1}^T \varepsilon_t(\theta)^2 - \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 &= \sum_{t=1}^T (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))^2 \\ &\quad + 2 \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi)) (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi)). \end{aligned}$$

From (36) it holds that, as $T \rightarrow \infty$,

$$\begin{aligned} \sup_{\theta \in \Theta_1} |T^{2(d-d_0)} \sum_{t=1}^T (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))^2| &= O_P(T^{-1}(\log T)^2) \xrightarrow{P} 0, \\ \sup_{\theta \in \Theta_2 \cup \Theta_3} |T^{-1} \sum_{t=1}^T (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))^2| &= O_P((\log T)^2 T^{2\kappa_1-1}) \xrightarrow{P} 0. \end{aligned}$$

Next, by the Cauchy-Schwarz inequality,

$$\begin{aligned} &\left| \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi)) (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi)) \right| \\ &\leq \left(\sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 \right)^{1/2} \left(\sum_{t=1}^T (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))^2 \right)^{1/2}, \end{aligned}$$

which proves (34) because $\sup_{\theta \in \Theta_1} T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 = O_P(1)$ by (50) of Lemma 6. To prove (35) note that $\sup_{\theta \in \Theta_3} T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 = O_P(1)$ by (51) of Lemma 6, which shows the result for $\theta \in \Theta_3$. Finally, for $\theta \in \Theta_2$,

$$\begin{aligned} &\sup_{\theta \in \Theta_2} |T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi)) (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))| \\ &\leq \left(\sup_{\theta \in \Theta_2} T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 \right)^{1/2} \left(\sup_{\theta \in \Theta_2} T^{-1} \sum_{t=1}^T (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))^2 \right)^{1/2} \\ &= O_P(T^{2\kappa_1} (\log T)^2 T^{2\kappa_1-1}) \end{aligned}$$

by (36) and Lemma 6. This completes the proof because $\kappa_1 < 1/4$. ■

7.1 Convergence on $\Theta_1(\kappa_1)$

First of all, if $\theta \in \Theta_1(\kappa_1)$ then $\varepsilon_t(\theta)$ is nonstationary, and by Lemma 1 the difference between $T^{2(d-d_0)+1} RSS(\theta)$ and $T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2$ is negligible in probability uniformly in $\theta \in \Theta_1$, so it suffices to consider the latter product moment. By (29) the functional central limit theorem of Marinucci & Robinson (2000) applies to $\Delta_+^{d-d_0} e_t(\psi)$

for all $\theta \in \Theta_1$ under the moment condition $E|\varepsilon_t|^q < \infty$ for some $q > \max(2, \kappa_1^{-1})$. By the continuous mapping theorem,

$$T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 \xrightarrow{D} \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 \int_0^1 W_{d_0-d-1}(r)^2 dr \quad (37)$$

as $T \rightarrow \infty$. The tightness condition (33) follows from (50) of Lemma 6 with $m = p+1$, which can be applied by Assumption A' and because $\sum_{n=0}^{\infty} |c_n(\psi)| < \infty$ uniformly in $\psi \in \Psi$ by (29). Thus, (37) can be strengthened to

$$T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 \Rightarrow \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 \int_0^1 W_{d_0-d-1}(r)^2 dr \text{ in } \mathcal{C}(\Theta_1),$$

so that by the continuous mapping theorem,

$$\inf_{\theta \in \Theta_1} T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 \xrightarrow{D} \inf_{\theta \in \Theta_1} \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 \int_0^1 W_{d_0-d-1}(r)^2 dr.$$

Since $2(d-d_0)+1 \leq -2\kappa_1$, $\inf_{d \in D_1} \int_0^1 W_{d_0-d-1}(r)^2 dr > 0$ almost surely, and $\inf_{\psi \in \Psi} (\sum_{n=0}^{\infty} c_n(\psi))^2 > 0$ by Assumption C, it holds that, for any $K > 0$, $\Pr(\inf_{\theta \in \Theta_1} RSS(\theta) > K) \rightarrow 1$.

7.2 Convergence on $\Theta_2(\kappa_1, \kappa_2)$

First note that by Lemma 1 it suffices to prove the result for $T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2$ assuming $\kappa_1 < 1/4$.

Because of (29) the Beveridge-Nelson decomposition

$$e_t(\psi) = c(L, \psi) \varepsilon_t = \left(\sum_{n=0}^{\infty} c_n(\psi) \right) \varepsilon_t + \Delta \sum_{n=0}^{\infty} \tilde{c}_n(\psi) \varepsilon_{t-n} \quad (38)$$

holds with $0 < \sum_{n=0}^{\infty} c_n(\psi) < \infty$ and $\tilde{c}_n(\psi) = -\sum_{k=n+1}^{\infty} c_k(\psi)$ satisfying $\sum_{n=0}^{\infty} |\tilde{c}_n(\psi)| < \infty$ uniformly in $\psi \in \Psi$, see Phillips & Solo (1992, Lemma 2.1). Letting $v = d - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa_2]$ and applying the decomposition (38) it holds that

$$\Delta_+^v e_t(\psi) = \left(\sum_{n=0}^{\infty} c_n(\psi) \right) \Delta_+^v \varepsilon_t + \Delta_+^{1+v} \sum_{n=0}^{\infty} \tilde{c}_n(\psi) \varepsilon_{t-n},$$

such that the relevant product moment is

$$\begin{aligned} T^{-1} \sum_{t=1}^T (\Delta_+^v e_t(\psi))^2 &\geq \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 T^{-1} \sum_{t=1}^T (\Delta_+^v \varepsilon_t)^2 \\ &\quad + 2 \left(\sum_{n=0}^{\infty} c_n(\psi) \right) T^{-1} \sum_{t=1}^T (\Delta_+^v \varepsilon_t) \left(\Delta_+^{1+v} \sum_{n=0}^{\infty} \tilde{c}_n(\psi) \varepsilon_{t-n} \right) \\ &= \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 R_{1T}(v) + 2 \left(\sum_{n=0}^{\infty} c_n(\psi) \right) R_{2T}(v, \psi). \end{aligned}$$

The term $R_{1T}(v)$ is analyzed in section 6.2 from which it follows that $R_{1T}(v) - \sigma_0^2 T^{-1}(T - N)F_N(v) \xrightarrow{P} 0$ as $T \rightarrow \infty$ uniformly in $d \in D_2$. Since $0 < (\sum_{n=0}^{\infty} c_n(\psi))^2 < \infty$ uniformly in $\psi \in \Psi$ it follows that

$$\left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 R_{1T}(v) - \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 \sigma_0^2 T^{-1}(T - N)F_N(v) \xrightarrow{P} 0$$

as $T \rightarrow \infty$ uniformly in $\theta \in \Theta_2$. For the second term, $R_{2T}(v, \psi)$, apply Lemma 8 and combine this with the fact that $\sum_{n=0}^{\infty} c_n(\psi)$ is bounded and bounded away from zero uniformly in $\psi \in \Psi$ to conclude that $2(\sum_{n=0}^{\infty} c_n(\psi))R_{2T}(v, \psi) = O_P(1)$ and is tight in $\theta \in \Theta_2$.

7.3 Proof of (30): $\Pr(\inf_{\theta \in \Theta_1 \cup \Theta_2} RSS(\theta) > K) \rightarrow 1$ as $T \rightarrow \infty$

In light of the results of sections 7.1-7.2, the arguments in section 6.3 can be used, changing D_j to $\Theta_j, j = 1, 2, 3$, and noting that $\kappa_1 < 1/4$ is needed for Lemma 1. Thus, setting $\kappa_1 = 1/5$ requires $q > 5$ moments in section 7.1, but that is satisfied by Assumption A'. In the analysis of Θ_2 it is also necessary to remark that $\inf_{\psi \in \Psi} (\sum_{n=0}^{\infty} c_n(\psi))^2 > 0$ by Assumption C.

7.4 Convergence on $\Theta_3(\kappa_2)$ and proof of (31): uniform convergence

Again, by Lemma 1, it suffices to prove the result for $T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2$. As in section 6.4, define the stationary and ergodic process $\eta_t(\theta) = \Delta_+^{d-d_0} e_t(\psi)$ and note that $\eta_t(\theta) - \Delta_+^{d-d_0} e_t(\psi) = \sum_{n=t}^{\infty} \pi_n(d_0 - d) e_{t-n}(\psi)$ with

$$\begin{aligned} E(\eta_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))^2 &= \sigma_0^2 \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} c_{m_1}(\psi) c_{m_2}(\psi) \sum_{n=t}^{\infty} \pi_n(d_0 - d) \pi_{n+m_1-m_2}(d_0 - d) \\ &\leq c \sum_{m_1=0}^{\infty} |c_{m_1}(\psi)| \sum_{m_2=0}^{\infty} |c_{m_2}(\psi)| \sum_{n=t}^{\infty} n^{d_0-d-1} (n + m_1 - m_2)^{d_0-d-1} \\ &\leq c \sum_{m_1=0}^{\infty} |c_{m_1}(\psi)| \sum_{m_2=0}^{\infty} |c_{m_2}(\psi)| \sum_{n=t}^{\infty} n^{2(d_0-d-1)} \leq ct^{2(d_0-d)-1} \rightarrow 0 \end{aligned}$$

using (29) and Lemma 3. It follows from the law of large numbers that

$$RSS(\theta) = T^{-1} \sum_{t=1}^T \eta_t(\theta)^2 + o_P(1) \xrightarrow{P} E(\Delta_+^{d-d_0} e_t(\psi))^2 \quad (39)$$

as $T \rightarrow \infty$, which shows the pointwise limit in probability.

Because $RSS(\theta_0) = T^{-1} \sum_{t=1}^T \varepsilon_t^2 + o_P(1) \xrightarrow{P} \sigma_0^2$ as $T \rightarrow \infty$ by (39), it is tight. Thus, tightness of $RSS(\theta)$ follows from condition (33) which is satisfied by (51) of Lemma 6.

This result holds for any κ_2 and specifically for $\kappa_2 = \bar{\kappa}_2$ chosen in the analysis of Θ_2 in section 7.3, c.f. section 6.3. Therefore the convergence in probability is uniform on $\theta \in \Theta_3(\bar{\kappa}_2)$, which proves (31).

7.5 Proof of (32): uniqueness

As in the proof of (20) for Theorem 1, it is sufficient to prove that

$$\inf_{\theta \in \Theta_3 \cap \{\theta: |\theta - \theta_0| \geq \delta\}} E(\Delta^{d-d_0} e_t(\psi))^2 > \sigma_0^2 \text{ for all } \delta > 0.$$

Since $\Delta^{d-d_0} e_t(\psi) = \sum_{k=0}^{\infty} \tau_k(d, \psi) \varepsilon_{t-k}$, where $\tau_k(d, \psi) = \sum_{m=0}^k \pi_m(d_0 - d) c_{k-m}(\psi)$ it holds that $E(\Delta^{d-d_0} e_t(\psi))^2 = \sigma_0^2 (1 + \sum_{k=1}^{\infty} \tau_k(d, \psi)^2) \geq \sigma_0^2$. By Assumption D, the equality holds if and only if $\theta = \theta_0$, which proves the result.

8 Inequalities

This section presents some useful inequalities that are used both in the proofs of the main theorems and in proofs of variation bounds in the next section. First note the following properties of the L_q -norm,

$$\|X + Y\|_q \leq \|X\|_q + \|Y\|_q, \quad \|XY\|_q \leq \|X\|_{2q} \|Y\|_{2q}, \quad \|X\|_q \leq \|X\|_r \text{ for } 1 \leq q \leq r. \quad (40)$$

The first inequality states that $\|\cdot\|_q$ is a norm (triangle inequality), the second follows from the Cauchy-Schwarz inequality, and the third follows from Jensen's inequality.

Lemma 2 *Let ε_t be i.i.d. with mean zero and $E|\varepsilon_t|^{2q} < \infty$ and define $Z = \sum_{j=0}^{\infty} \xi_j \varepsilon_j$ for some coefficients ξ_j for which $\sum_{j=0}^{\infty} \xi_j^2 < \infty$. Then*

$$\|Z\|_{2q} \leq c_q \|Z\|_2, \quad (41)$$

where the constant c_q does not depend on the coefficients ξ_j .

Proof. See Lemma B.1 of Johansen & Nielsen (2010b). ■

Lemma 3 *Uniformly in $-u_0 \leq \tilde{u} \leq u \leq u_0$ and for $j \geq 1$ it holds that*

$$|\pi_j(u)| \leq c j^{u-1}, \quad (42)$$

$$|\pi_j(u) - \pi_j(\tilde{u})| \leq c(u - \tilde{u})(1 + \log j) j^{u-1}, \quad (43)$$

where the constants do not depend on u, \tilde{u} , or j .

Uniformly in $-\delta_0 \leq v + 1/2 \leq \delta_0$ and for $j \geq 1$ it holds that

$$\pi_j(-v) \geq c j^{-v-1}, \quad (44)$$

where the constant does not depend on v or j .

Proof. The results (42) and (43) are in Lemma B.3 of Johansen & Nielsen (2010b). To prove (44) let $u = -v \in [1/2 - \delta_0, 1/2 + \delta_0]$ and apply Stirling's formula,

$$\pi_j(u) = \frac{\Gamma(u+j)}{\Gamma(u)\Gamma(j+1)} = \frac{1}{\Gamma(u)} j^{u-1} (1 + \epsilon(u, j)),$$

where $\sup_{1/2-\delta_0 \leq u \leq 1/2+\delta_0} |\epsilon(u, j)| \rightarrow 0$ as $j \rightarrow \infty$. This proves the result and shows that the constant can be chosen to depend only on δ_0 . ■

Lemma 4 Let $\xi_M(u, v, k) = \max_{1 \leq n, m \leq M} \sum_{t=\max(n, m)}^M |\zeta_{t-n}(-u, k) \zeta_{t-m}(-v, k)|$ for coefficients $\zeta_n(u, k)$ satisfying $\zeta_0(u, k) = 1$ and $\zeta_n(u, k) \leq c(\log n)^k n^{u-1}$ for $n \geq 1$, where $c > 0$ does not depend on u, k , or n , and let $F_M(v) = \sum_{n=0}^M \pi_n(-v)^2$. Then:

(a) Uniformly for $\min(v + 1, u + 1, u + v + 1) \geq a$ it holds that

$$\xi_M(u, v, k) \leq \begin{cases} c(1 + \log M)^{1+2k} M^{-a} & \text{if } a \leq 0, \\ c & \text{if } a > 0, \end{cases}$$

where $c > 0$ does not depend on u, v , or M .

(b) For $a \geq 0$ and $v \leq -1/2 + a$,

$$F_M(v) \geq 1 + c \frac{1 - M^{-2a}}{2a},$$

where $c > 0$ does not depend on a, v , or M .

Proof. (a) Only the terms with $t \geq \max(n, m) + 1$ are considered, those with $t = n$ or $t = m$ being simpler. First, for $a \leq 0$,

$$\begin{aligned} \sum_{t=\max(n, m)+1}^M |\zeta_{t-n}(-u, k) \zeta_{t-m}(-v, k)| &\leq c(1 + \log M)^{2k} \sum_{t=\max(n, m)+1}^M (t - n)^{-u-1} (t - m)^{-v-1} \\ &\leq c(1 + \log M)^{2k} \sum_{t=\max(n, m)+1}^M (t - \max(n, m))^{-u-v-2}, \end{aligned}$$

where

$$\sum_{t=\max(n, m)+1}^M (t - \max(n, m))^{-u-v-2} \leq c(1 + \log M) M^{-a}$$

because $-u - v - 2 \geq -a - 1 \geq -1$. Next, for $a > 0$ bound the factors $(1 + \log(t - n))^k (t - n)^{-a/3}$ and $(1 + \log(t - m))^k (t - m)^{-a/3}$ by a constant and apply the inequality $\sum_{t=\max(n, m)+1}^M (t - \max(n, m))^{-a+2a/3-1} \leq c$ for $a > 0$.

(b) Using (44) of Lemma 3 and the inequality $\sum_{n=1}^M n^{-u-1} \geq ca^{-1}(1 - M^{-a})$ for $u \leq a$ it holds that

$$F_M(v) \geq 1 + c \sum_{n=1}^M n^{-2v-2} \geq 1 + c \frac{1 - M^{-2a}}{2a}.$$

■

9 Variation bounds

This section contains a series of lemmas that are used to verify tightness conditions for the processes in the previous sections. The first lemma deals with product moments of nonstationary and stationary processes, and considers tightness conditions for the

fractional parameter, d . The next lemma extends these results to include tightness in the short-run parameters, ψ . The third lemma covers product moments of processes that are nearly stationary, and the final lemma covers mixed product moments with both stationary and nearly stationary processes.

Lemma 5 *Let $Z_t = \sum_{n=0}^{\infty} \xi_n \varepsilon_{t-n}$ be a stationary linear process with $\sum_{n=0}^{\infty} |\xi_n| < \infty$ and ε_t i.i.d. with mean zero, variance $\sigma^2 > 0$, and $E|\varepsilon_t|^{2m+2} < \infty$, and define $Q_T(u) = T^{-1} \sum_{t=1}^T (\Delta_+^u Z_t)^2$. Then:*

(a) *Uniformly in $v_0 \leq \tilde{u} \leq u \leq u_0 < -1/2$,*

$$\|T^{2u+1}Q_T(u)\|_{m+1} \leq c \text{ and } \|T^{2u+1}Q_T(u) - T^{2\tilde{u}+1}Q_T(\tilde{u})\|_{m+1} \leq c|u - \tilde{u}| \quad (45)$$

for $m \geq 1$, where the constants do not depend on u, \tilde{u} , or T .

(b) *Uniformly in $-1/2 < v_0 \leq \tilde{u} \leq u \leq u_0$,*

$$\left\| \frac{\partial^k Q_T(u)}{\partial u^k} \right\|_{m+1} \leq c \text{ and } \left\| \frac{\partial^k Q_T(u)}{\partial u^k} - \frac{\partial^k Q_T(\tilde{u})}{\partial \tilde{u}^k} \right\|_{m+1} \leq c|u - \tilde{u}| \quad (46)$$

for $m \geq 1$ and $k = 0, 1, 2$, where the constants do not depend on u, \tilde{u} , or T .

Proof. *Proof of (a):* Apply the decomposition

$$Q_T(u) - Q_T(\tilde{u}) = T^{-1} \sum_{t=1}^T \Delta_+^u Z_t (\Delta_+^u Z_t - \Delta_+^{\tilde{u}} Z_t) + T^{-1} \sum_{t=1}^T \Delta_+^{\tilde{u}} Z_t (\Delta_+^u Z_t - \Delta_+^{\tilde{u}} Z_t)$$

and use the properties (40) and Lemma 2 to find

$$\begin{aligned} & \|T^{2u+1}Q_T(u) - T^{2\tilde{u}+1}Q_T(\tilde{u})\|_{m+1} \quad (47) \\ & \leq T^{-1} \sum_{t=1}^T \|T^{u+1/2} \Delta_+^u Z_t\|_{2m+2} \|T^{u+1/2} \Delta_+^u Z_t - T^{\tilde{u}+1/2} \Delta_+^{\tilde{u}} Z_t\|_{2m+2} \\ & \quad + T^{-1} \sum_{t=1}^T \|T^{\tilde{u}+1/2} \Delta_+^{\tilde{u}} Z_t\|_{2m+2} \|T^{u+1/2} \Delta_+^u Z_t - T^{\tilde{u}+1/2} \Delta_+^{\tilde{u}} Z_t\|_{2m+2} \\ & \leq T^{-1} \sum_{t=1}^T \|T^{u+1/2} \Delta_+^u Z_t\|_2 \|T^{u+1/2} \Delta_+^u Z_t - T^{\tilde{u}+1/2} \Delta_+^{\tilde{u}} Z_t\|_2 \\ & \quad + T^{-1} \sum_{t=1}^T \|T^{\tilde{u}+1/2} \Delta_+^{\tilde{u}} Z_t\|_2 \|T^{u+1/2} \Delta_+^u Z_t - T^{\tilde{u}+1/2} \Delta_+^{\tilde{u}} Z_t\|_2 \end{aligned}$$

In the same way, $\|T^{2u+1}Q_T(u)\|_{m+1} \leq T^{-1} \sum_{t=1}^T \|T^{u+1/2} \Delta_+^u Z_t\|_2^2$. The result now follows from Johansen & Nielsen (2010b, Lemma C.3), which states that

$$\|T^{u+1/2} \Delta_+^u Z_t\|_2 \leq c \text{ and } \|T^{u+1/2} \Delta_+^u Z_t - T^{\tilde{u}+1/2} \Delta_+^{\tilde{u}} Z_t\|_2 \leq c|u - \tilde{u}| \quad (48)$$

uniformly in $v_0 \leq \tilde{u} \leq u \leq u_0 < -1/2$.

Proof of (b): Apply the same decomposition and proof as for part (a), but instead of (48) use Johansen & Nielsen (2010b, Lemma C.4), which states that

$$\left\| \frac{\partial^k}{\partial u^k} \Delta_+^u Z_t \right\|_2 \leq c \text{ and } \left\| \frac{\partial^k}{\partial u^k} \Delta_+^u Z_t - \frac{\partial^k}{\partial \tilde{u}^k} \Delta_+^{\tilde{u}} Z_t \right\|_2 \leq c|u - \tilde{u}| \quad (49)$$

uniformly in $-1/2 < v_0 \leq \tilde{u} \leq u \leq u_0$. ■

Lemma 6 *Let $Z_t(\psi) = \sum_{n=0}^{\infty} \xi_n(\psi) \varepsilon_{t-n}$ be a stationary linear process with $\sum_{n=0}^{\infty} |\xi_n(\psi)| < \infty$ uniformly in ψ and let ε_t be i.i.d. with mean zero, variance $\sigma^2 > 0$, and $E|\varepsilon_t|^{2m+2} < \infty$. Furthermore, suppose the coefficients $\xi_n(\psi)$ are differentiable with derivatives $\dot{\xi}_n(\psi) = \frac{\partial \xi_n(\psi)}{\partial \psi}$ satisfying $\sum_{n=0}^{\infty} |\dot{\xi}_n(\psi)| < \infty$ uniformly in ψ , and define $Q_T(u, \psi) = T^{-1} \sum_{t=1}^T (\Delta_+^u Z_t(\psi))^2$. Then:*

(a) *Uniformly in $(\psi, \tilde{\psi}) \in \Psi \times \Psi$ and $v_0 \leq \tilde{u} \leq u \leq u_0 < -1/2$,*

$$\|T^{2u+1} Q_T(u, \psi)\|_{m+1} \leq c \text{ and } \|T^{2u+1} Q_T(u, \psi) - T^{2\tilde{u}+1} Q_T(\tilde{u}, \tilde{\psi})\|_{m+1} \leq c|(u, \psi) - (\tilde{u}, \tilde{\psi})| \quad (50)$$

for $m \geq 1$, where the constants do not depend on $u, \tilde{u}, \psi, \tilde{\psi}$, or T .

(b) *Suppose the coefficients $\xi_n(\psi)$ are $k+1$ times differentiable with derivatives $\frac{\partial^{j+1} \xi_n(\psi)}{\partial \psi^{(j+1)}}$ satisfying $\sum_{n=0}^{\infty} \left| \frac{\partial^{j+1} \xi_n(\psi)}{\partial \psi^{(j+1)}} \right| < \infty$ uniformly in ψ for $j \leq k$. Uniformly in $(\psi, \tilde{\psi}) \in \Psi \times \Psi$ and $-1/2 < v_0 \leq \tilde{u} \leq u \leq u_0$, for $i, j, k = 0, 1, 2$ and $i + j = k$,*

$$\left\| \frac{\partial^k Q_T(u, \psi)}{\partial u^i \partial \psi^{(j)}} \right\|_{m+1} \leq c \text{ and } \left\| \frac{\partial^k Q_T(u, \psi)}{\partial u^i \partial \psi^{(j)}} - \frac{\partial^k Q_T(\tilde{u}, \tilde{\psi})}{\partial \tilde{u}^i \partial \tilde{\psi}^{(j)}} \right\|_{m+1} \leq c|(u, \psi) - (\tilde{u}, \tilde{\psi})| \quad (51)$$

for $m \geq 1$, where the constants do not depend on $u, \tilde{u}, \psi, \tilde{\psi}$, or T .

Proof. *Proof of (a):* The first result follows directly from (48) because $\sum_{n=0}^{\infty} |\xi_n(\psi)| < \infty$ uniformly in ψ . To prove the second result, decompose

$$\begin{aligned} \|T^{2u+1} Q_T(u, \psi) - T^{2\tilde{u}+1} Q_T(\tilde{u}, \tilde{\psi})\|_{m+1} &\leq \|T^{2u+1} Q_T(u, \psi) - T^{2\tilde{u}+1} Q_T(\tilde{u}, \psi)\|_{m+1} \\ &\quad + \|T^{2\tilde{u}+1} Q_T(\tilde{u}, \psi) - T^{2\tilde{u}+1} Q_T(\tilde{u}, \tilde{\psi})\|_{m+1} \end{aligned}$$

by the triangle inequality, see (40), so that the increments in u and ψ can be considered separately. The first term (increment in u) is bounded by $c|u - \tilde{u}|$ using the decomposition (47) and application of (48). For the second term (increment in ψ) note that $Q_T(\tilde{u}, \psi)$ is differentiable in ψ in the convex set Ψ and apply the mean value theorem,

$$T^{2\tilde{u}+1} Q_T(\tilde{u}, \psi) - T^{2\tilde{u}+1} Q_T(\tilde{u}, \tilde{\psi}) = T^{2\tilde{u}+1} |\psi - \tilde{\psi}| \frac{\partial Q_T(\tilde{u}, \bar{\psi})}{\partial \psi},$$

where $\bar{\psi}$ is an intermediate value between ψ and $\tilde{\psi}$. The derivative is

$$\frac{\partial Q_T(\tilde{u}, \bar{\psi})}{\partial \psi} = 2T^{-1} \sum_{t=1}^T (\Delta_+^{\tilde{u}} Z_t(\bar{\psi})) (\Delta_+^{\tilde{u}} \dot{Z}_t(\bar{\psi})),$$

where $\dot{Z}_t(\psi) = \sum_{n=0}^{\infty} \dot{\xi}_n(\psi) \varepsilon_{t-n}$. Because $\sum_{n=0}^{\infty} |\dot{\xi}_n(\psi)| < \infty$, the first result in (48) applies to fractional differences of $\dot{Z}_t(\psi)$ such that

$$\|T^{2\tilde{u}+1} \frac{\partial Q_T(\tilde{u}, \bar{\psi})}{\partial \psi}\|_{m+1} \leq T^{-1} \sum_{t=1}^T \|T^{\tilde{u}+1/2} \Delta_+^u Z_t(\bar{\psi})\|_{2m+2} \|T^{\tilde{u}+1/2} \Delta_+^u \dot{Z}_t(\bar{\psi})\|_{2m+2} \leq c$$

using also (40) and Lemma 2. Thus $\|T^{2\tilde{u}+1} Q_T(\tilde{u}, \psi) - T^{2\tilde{u}+1} Q_T(\tilde{u}, \tilde{\psi})\|_{m+1} \leq c|\psi - \tilde{\psi}|$.

Proof of (b): If $k = 0$ the same proof can be applied as for part (a), but using (49) instead of (48). The derivatives with respect to u do not change the proof because they are included in (49). Derivatives with respect to ψ also do not change the proof since the coefficients $\frac{\partial^{j+1} \xi_n(\psi)}{\partial \psi^{(j+1)}}$ are absolutely summable uniformly in $\psi \in \Psi$. ■

Lemma 7 Let $w_{1t} = w_{1t}(v) = \sum_{n=0}^{N-1} \pi_n(-v) \varepsilon_{t-n}$ and $w_{2t} = w_{2t}(v) = \sum_{n=N}^{t-1} \pi_n(-v) \varepsilon_{t-n}$, where ε_t is i.i.d. with mean zero, variance $\sigma^2 > 0$, and $E|\varepsilon_t|^4 < \infty$, and define the product moments $Q_{12NT}(v) = T^{-1} \sum_{t=N+1}^T w_{1t} w_{2t}$ and $Q_{11NT}(v) = T^{-1} \sum_{t=N+1}^T w_{1t}^2 - \sigma_0^2 T^{-1} (T-N) F_N(v)$. Then it holds that, uniformly in $-1/2 - \kappa_1 \leq \tilde{v} \leq v \leq -1/2 + \kappa_2$,

$$\|Q_{12NT}(v)\|_2 \leq c(\log T) T^{-1/2+\kappa_1} N^{1/2+\kappa_1}, \quad (52)$$

$$\|Q_{12NT}(v) - Q_{12NT}(\tilde{v})\|_2 \leq c|v - \tilde{v}| (\log T)^2 T^{-1/2+\kappa_1} N^{1/2+\kappa_1}, \quad (53)$$

$$\|Q_{11NT}(v)\|_2 \leq c(\log T) T^{-1/2} N^{1/2+2\kappa_1}, \quad (54)$$

$$\|Q_{11NT}(v) - Q_{11NT}(\tilde{v})\|_2 \leq c|v - \tilde{v}| (\log T)^2 T^{-1/2} N^{1/2+2\kappa_1}, \quad (55)$$

where the constants do not depend on v, \tilde{v} , or T .

Proof. *Proof of (52):* First evaluate

$$EQ_{12NT}(v)^2 = T^{-2} E \prod_{k=1}^2 \sum_{t_k=N+1}^T \sum_{n_k=0}^{N-1} \sum_{m_k=N}^{t_k-1} \pi_{n_k}(-v) \pi_{m_k}(-v) \varepsilon_{t_k-n_k} \varepsilon_{t_k-m_k}.$$

The term $E(\prod_{k=1}^2 \varepsilon_{t_k-n_k} \varepsilon_{t_k-m_k})$ is non-zero only if the subscripts are equal in pairs. However, $n_k < N \leq m_k$ such that $t_k - n_k > t_k - m_k$ for $k = 1, 2$. There are two possibilities. First, $t_1 - n_1 = t_2 - m_2$ and $t_1 - m_1 = t_2 - n_2$ implies $m_2 - n_1 = t_2 - t_1 = n_2 - m_1$, but because $m_2 - n_1 > 0$ and $n_2 - m_1 < 0$ this cannot happen. Second, when $t_1 - n_1 = t_2 - n_2$ and $t_1 - m_1 = t_2 - m_2$ eliminate $n_2 = t_2 - t_1 + n_1$ and $m_2 = t_2 - t_1 + m_1$. In this case the contribution is

$$\sigma_0^4 T^{-2} \sum_{t_1=N+1}^T \sum_{t_2=N+1}^T \sum_{n_1=0}^{N-1} \sum_{m_1=N}^{t_1-1} \pi_{n_1}(-v) \pi_{t_2-t_1+n_1}(-v) \pi_{m_1}(-v) \pi_{t_2-t_1+m_1}(-v),$$

but $\sum_{n_1=0}^{N-1} \pi_{n_1}(-v) \pi_{t_2-t_1+n_1}(-v) \leq \xi_N(v, v, 0)$ and $\sum_{m_1=N}^{t_1-1} \pi_{m_1}(-v) \pi_{t_2-t_1+m_1}(-v) \leq \xi_T(v, v, 0)$ by (42) of Lemma 3, so the contribution is bounded by

$$cT^{-2} \sum_{t_1=N+1}^T \sum_{t_2=N+1}^T \xi_N(v, v, 0) \xi_T(v, v, 0).$$

The result now follows because $|t_2 - t_1| = |n_2 - n_1| \leq N$ and using Lemma 4(a).

Proof of (53): Next consider $\|Q_{12NT}(v) - Q_{12NT}(\tilde{v})\|_2$ which is bounded by

$$\|T^{-1} \sum_{t=N+1}^T (w_{1t}(v) - w_{1t}(\tilde{v}))w_{2t}(v)\|_2 + \|T^{-1} \sum_{t=N+1}^T w_{1t}(\tilde{v})(w_{2t}(v) - w_{2t}(\tilde{v}))\|_2.$$

For the first term write $w_{1t}(v) - w_{1t}(\tilde{v}) = \sum_{n=0}^{N-1} (\pi_n(-v) - \pi_n(-\tilde{v}))\varepsilon_{t-n} = (v - \tilde{v}) \sum_{n=0}^{N-1} \zeta_n(-v, 1)\varepsilon_{t-n}$, see (43) of Lemma 3 and Lemma 4(a). Now apply the same proof as for (52), noting that only a log-factor is added. The same proof can be used for the second term.

Proof of (54): The expectation is

$$\begin{aligned} EQ_{11NT}(v) &= T^{-1} \sum_{t=N+1}^T \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} \pi_{n_1}(-v)\pi_{n_2}(-v)E(\varepsilon_{t-n_1}\varepsilon_{t-n_2}) - \sigma_0^2 T^{-1}(T-N)F_N(v) \\ &= \sigma_0^2 T^{-1} \sum_{t=N+1}^T \sum_{n=0}^{N-1} \pi_n(-v)^2 - \sigma_0^2 T^{-1}(T-N)F_N(v) = 0. \end{aligned} \quad (56)$$

The second moment is

$$\begin{aligned} EQ_{11NT}(v)^2 &= E(T^{-1} \sum_{t=N+1}^T w_{1t}^2)^2 + \sigma_0^4 T^{-2}(T-N)^2 F_N(v)^2 \\ &\quad - 2E(T^{-1} \sum_{t=N+1}^T w_{1t}^2)\sigma_0^2 T^{-1}(T-N)F_N(v) \\ &= E(T^{-1} \sum_{t=N+1}^T w_{1t}^2)^2 - \sigma_0^4 T^{-2}(T-N)^2 F_N(v)^2 \end{aligned} \quad (57)$$

using (56). Now,

$$E(T^{-1} \sum_{t=N+1}^T w_{1t}^2)^2 = T^{-2} E \prod_{k=1}^2 \sum_{t_k=N+1}^T \sum_{n_k=0}^{N-1} \sum_{m_k=0}^{N-1} \pi_{n_k}(-v)\pi_{m_k}(-v)\varepsilon_{t_k-n_k}\varepsilon_{t_k-m_k},$$

where again the subscripts in $\prod_{k=1}^2 \varepsilon_{t_k-n_k}\varepsilon_{t_k-m_k}$ have to be equal in pairs. There are three cases.

1) Suppose first that $t_k - n_k = t_k - m_k$ for $k = 1, 2$, i.e. $n_k = m_k$. In this case the contribution is

$$\sigma_0^4 T^{-2} \prod_{k=1}^2 \sum_{t_k=N+1}^T \sum_{n_k=0}^{N-1} \pi_{n_k}(-v)^2 = \sigma_0^4 T^{-2}(T-N)^2 F_N(v)^2,$$

which cancels with the second term of (57).

2) If $t_1 - n_1 = t_2 - n_2$ and $t_1 - m_1 = t_2 - m_2$ the contribution is

$$\sigma_0^4 T^{-2} \sum_{t_1=N+1}^T \sum_{t_2=N+1}^T \left(\sum_{n=0}^{N-1} \pi_n(-v) \pi_{t_2-t_1+n}(-v) \right)^2 \leq c T^{-2} \sum_{t_1=N+1}^T \sum_{t_2=N+1}^T \xi_N(v, v, 0)^2,$$

which with the restriction $|t_2 - t_1| = |n_2 - n_1| \leq N$ is bounded by $c T^{-1} N \xi_N(v, v, 0)^2 \leq c(\log T)^2 T^{-1} N^{1+4\kappa_1}$ by Lemma 4(a).

3) If $t_1 - n_1 = t_2 - m_2$ and $t_1 - m_1 = t_2 - n_2$ the contribution is

$$\begin{aligned} & \sigma_0^4 T^{-2} \sum_{t_1=N+1}^T \sum_{t_2=N+1}^T \sum_{n_1=0}^{N-1} \sum_{m_1=0}^{N-1} \pi_{n_1}(-v) \pi_{t_2-t_1+m_1}(-v) \pi_{m_1}(-v) \pi_{t_2-t_2+n_1}(-v) \\ & \leq c(\log T)^2 T^{-1} N^{1+4\kappa_1} \end{aligned}$$

in the same way as in case 2).

Proof of (55): Apply the same decomposition as in the proof of (53) and then use the same proof as for (54) with an extra log-factor. ■

Lemma 8 *Let $Z_t(\psi) = \sum_{n=0}^{\infty} \xi_n(\psi) \varepsilon_{t-n}$ be a stationary linear process with $\sum_{n=0}^{\infty} |\xi_n(\psi)| < \infty$ uniformly in ψ and let ε_t be i.i.d. with mean zero, variance $\sigma^2 > 0$, and $E|\varepsilon_t|^{2m+2} < \infty$. Furthermore, suppose the coefficients $\xi_n(\psi)$ are differentiable with derivatives $\dot{\xi}_n(\psi) = \frac{\partial \xi_n(\psi)}{\partial \psi}$ satisfying $\sum_{n=0}^{\infty} |\dot{\xi}_n(\psi)| < \infty$ uniformly in ψ , and define $Q_T(v, \psi) = T^{-1} \sum_{t=1}^T (\Delta_+^v \varepsilon_t) (\Delta_+^{1+v} Z_t(\psi))$. Then, for $\kappa_i \in (0, 1/2)$, $i = 1, 2$, it holds uniformly in $(\psi, \tilde{\psi}) \in \Psi \times \Psi$ and $-1/2 - \kappa_1 \leq \tilde{v} \leq v \leq -1/2 + \kappa_2$ that*

$$\|Q_T(v, \psi)\|_{m+1} \leq c \text{ and } \|Q_T(v, \psi) - Q_T(\tilde{v}, \tilde{\psi})\|_{m+1} \leq c|(v, \psi) - (\tilde{v}, \tilde{\psi})| \quad (58)$$

for $m \geq 1$, where the constants do not depend on $v, \tilde{v}, \psi, \tilde{\psi}$, or T .

Proof. First consider

$$\begin{aligned} & E \left(T^{-1} \sum_{t=1}^T \sum_{n=1}^t \sum_{l=1}^t \pi_{t-n}(-v) \pi_{t-l}(-v-1) \varepsilon_n Z_l(\psi) \right)^{m+1} \\ & = T^{-m-1} \sum_{(1)} \left(\prod_{k=1}^{m+1} \pi_{t_k-n_k}(-v) \pi_{t_k-l_k}(-v-1) \right) E \left(\prod_{k=1}^{m+1} \varepsilon_{n_k} Z_{l_k}(\psi) \right), \end{aligned}$$

where the summation $\sum_{(1)}$ is over $1 \leq n_k, l_k \leq t_k \leq T$, $k = 1, 2, \dots, m+1$. This, on the other hand, is bounded by

$$\begin{aligned} & \left(\prod_{k=1}^{m+1} \max_{n_k, l_k} \sum_{t_k=\max(n_k, l_k)}^T |\pi_{t_k-n_k}(-v) \pi_{t_k-l_k}(-v-1)| \right) T^{-m-1} \sum_{(2)} |E \prod_{k=1}^{m+1} \varepsilon_{n_k} Z_{l_k}(\psi)| \\ & \leq \xi_T(v, v+1, 0)^{m+1} T^{-m-1} \sum_{(2)} |E \prod_{k=1}^{m+1} \varepsilon_{n_k} Z_{l_k}(\psi)|, \end{aligned} \quad (59)$$

where the summation $\sum_{(2)}$ is over $1 \leq n_k, m_k \leq T, k = 1, 2, \dots, m + 1$.

Because $\min(v + 1, v + 2, 2v + 2) \geq 1/2 - \kappa_1 > 0$, it holds by Lemma 4(a) that $\xi_T(v, v + 1, 0) \leq c$, where $c > 0$ does not depend on T or v , and the first result in (58) will follow upon showing that the second factor in (59) is bounded. Thus,

$$\sum_{(2)} E \left(\prod_{k=1}^{m+1} \varepsilon_{n_k} Z_{l_k}(\psi) \right) = \sum_{(3)} \left(\prod_{k=1}^{m+1} \xi_{j_k}(\psi) \right) E \left(\prod_{k=1}^{m+1} \varepsilon_{n_k} \varepsilon_{l_k - j_k} \right).$$

The sum $\sum_{(3)}$ extends over $1 \leq n_k, l_k \leq T, 0 \leq j_k \leq \infty, k = 1, 2, \dots, m + 1$, but most terms are zero because $E\varepsilon_t = 0$. There is a non-zero contribution if the subscripts $n_k, l_k - j_k$ are equal in pairs, triples, or more. For each pair, there is one constraint, for each triple there are two constraints, etc., and thus there are fewest constraints – and hence the largest number of terms – when the subscripts are equal in pairs. This imposes $m + 1$ constraints on the n_k, l_k summations. Next, because $\sum_{j=0}^{\infty} |\xi_j(\psi)| < \infty$ uniformly in $\psi \in \Psi$, the $m + 1$ summations over the indices $j_k, k = 1, 2, \dots, m + 1$, are finite and bounded by a constant that does not depend on ψ . This leaves only $m + 1$ summations over indices n_k or l_k which results in at most T^{m+1} terms and shows that $T^{-m-1} \sum_{(2)} E(\prod_{k=1}^{m+1} \varepsilon_{n_k} Z_{l_k}(\psi))$ is bounded by a coefficient that does not depend on T or ψ .

To prove the second result apply the triangle inequality,

$$\|Q_T(v, \psi) - Q_T(\tilde{v}, \tilde{\psi})\|_{m+1} \leq \|Q_T(v, \psi) - Q_T(\tilde{v}, \psi)\|_{m+1} + \|Q_T(\tilde{v}, \psi) - Q_T(\tilde{v}, \tilde{\psi})\|_{m+1}.$$

For the first term (increment in v), apply the same proof as above but replace either $\pi_{t-n}(-v)$ by $\pi_{t-n}(-v) - \pi_{t-n}(-\tilde{v}) = \zeta_{t-n}(-v, 1)(v - \tilde{v})$ or $\pi_{t-l}(-v - 1)$ by $\pi_{t-l}(-v - 1) - \pi_{t-l}(-\tilde{v} - 1) = \zeta_{t-l}(-v - 1, 1)(v - \tilde{v})$, see (43), and use Lemma 4(a) with $\xi_T(v, v + 1, 1) \leq c$. For the second term (increment in ψ), note that $Q_T(\tilde{v}, \psi)$ is differentiable in ψ in the convex set Ψ and apply the mean value theorem,

$$Q_T(\tilde{v}, \psi) - Q_T(\tilde{v}, \tilde{\psi}) = |\psi - \tilde{\psi}| \frac{\partial Q_T(\tilde{v}, \bar{\psi})}{\partial \psi},$$

where $\bar{\psi}$ is an intermediate value between ψ and $\tilde{\psi}$. The derivative is

$$\frac{\partial Q_T(\tilde{v}, \bar{\psi})}{\partial \psi} = T^{-1} \sum_{t=1}^T (\Delta_+^v \varepsilon_t) (\Delta_+^{1+v} \dot{Z}_t(\psi)),$$

where $\dot{Z}_t(\psi) = \sum_{n=0}^{\infty} \dot{\xi}_n(\psi) \varepsilon_{t-n}$. Because $\sum_{n=0}^{\infty} |\dot{\xi}_n(\psi)| < \infty$ uniformly in $\psi \in \Psi$, the same proof as above can be applied again with $\dot{\xi}_n(\psi)$ replacing $\xi_n(\psi)$. ■

References

Beran, J. (1995), ‘Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models’, *Journal of the Royal Statistical Society Series B* **57**, 659–672.

- Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.
- Bloomfield, P. (1973), ‘An exponential model for the spectrum of a scalar time series’, *Biometrika* **60**, 217–226.
- Box, G. E. P. & Jenkins, G. M. (1970), *Time Series Analysis, Forecasting, and Control*, Holden-Day, San Francisco.
- Dahlhaus, R. (1989), ‘Efficient parameter estimation for self-similar processes’, *Annals of Statistics* **17**, 1749–1766.
- Fox, R. & Taqqu, M. S. (1986), ‘Large-sample properties of parameter estimates for strongly dependent stationary gaussian series’, *Annals of Statistics* **14**, 517–532.
- Giraitis, L. & Surgailis, D. (1990), ‘A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle’s estimate’, *Probability Theory and Related Fields* **86**, 87–104.
- Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, Academic Press, New York.
- Hosoya, Y. (1996), ‘The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence’, *Journal of Econometrics* **73**, 217–236.
- Hualde, J. & Robinson, P. M. (2010), ‘Gaussian pseudo-maximum likelihood estimation of fractional time series models’. Working Paper, Universidad Publica de Navarra.
- Johansen, S. & Nielsen, M. Ø. (2010a), ‘Likelihood inference for a fractionally cointegrated vector autoregressive model’. QED Working Paper 1237, Queen’s University.
- Johansen, S. & Nielsen, M. Ø. (2010b), ‘Likelihood inference for a nonstationary fractional autoregressive model’, *Journal of Econometrics* **158**, 51–66.
- Johansen, S. & Nielsen, M. Ø. (2010c), ‘A necessary moment condition for the fractional functional central limit theorem’. QED Working Paper 1244, Queen’s University.
- Kallenberg, O. (2001), *Foundations of Modern Probability Theory*, 2nd edn, Springer, New York.
- Li, W. K. & McLeod, A. I. (1986), ‘Fractional time series modelling’, *Biometrika* **73**, 217–221.
- Lieberman, O., Rosemarin, R. & Rousseau, J. (2010), ‘Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes’. Forthcoming in *Econometric Theory*.

- Marinucci, D. & Robinson, P. M. (2000), ‘Weak convergence of multivariate fractional processes’, *Stochastic Processes and their Applications* **86**, 103–120.
- Nielsen, M. Ø. (2004), ‘Efficient likelihood inference in nonstationary univariate models’, *Econometric Theory* **20**, 116–146.
- Phillips, P. C. B. & Solo, V. (1992), ‘Asymptotics for linear processes’, *Annals of Statistics* **20**, 971–1001.
- Robinson, P. M. (1994), ‘Efficient tests of nonstationary hypotheses’, *Journal of the American Statistical Association* **89**, 1420–1437.
- Robinson, P. M. (1995), ‘Gaussian semiparametric estimation of long range dependence’, *Annals of Statistics* **23**, 1630–1661.
- Robinson, P. M. (2005), ‘Efficiency improvements in inference on stationary and nonstationary fractional time series’, *Annals of Statistics* **33**, 1800–1842.
- Robinson, P. M. (2006), Conditional-sum-of-squares estimation of models for stationary time series with long memory, in H.-C. Ho, C.-K. Ing & T. L. Lai, eds, ‘Time Series and Related Topics: In Memory of Ching-Zong Wei’, IMS Lecture Notes–Monograph Series Vol. 52, pp. 130–137.
- Shimotsu, K. & Phillips, P. C. B. (2005), ‘Exact local Whittle estimation of fractional integration’, *Annals of Statistics* **33**, 1890–1933.
- Tanaka, K. (1999), ‘The nonstationary fractional unit root’, *Econometric Theory* **15**, 549–582.
- van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- Velasco, C. & Robinson, P. M. (2000), ‘Whittle pseudo-maximum likelihood estimation for nonstationary time series’, *Journal of the American Statistical Association* **95**, 1229–1243.
- Zygmund, A. (2003), *Trigonometric Series*, third edn, Cambridge University Press, Cambridge.