Queen's Economics Department Working Paper No. 1259

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

Morten ÃŸrregaard 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 multivariate fractional time series models*

Morten Ørregaard Nielsen
Queen's University and CREATES
Email: mon@econ.queensu.ca

August 11, 2014

Abstract

This paper proves consistency and asymptotic normality for the conditional-sum-of-squares estimator, which is equivalent to the conditional maximum likelihood estimator, in multivariate fractional time series models. The model is parametric and quite general, and, in particular, encompasses the multivariate non-cointegrated fractional ARIMA model. The novelty of the consistency result, in particular, is that it applies to a multivariate model and to an arbitrarily large set of admissible parameter values, for which the objective function does not converge uniformly in probablity, 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.

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

JEL classification: C22, C32.

1 Introduction

This paper considers conditional-sum-of-squares (CSS) estimation of multivariate 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 (unconditional) Gaussian maximum likelihood estimator and being efficient under Gaussianity. However, Gaussianity is not assumed in this paper. Compared to (unconditional) maximum likelihood estimation, though, CSS estimation is computationally much simpler. For these reasons, the CSS estimator has been very widely applied in the literature, also for fractional time series models.

In the simplest case, the univariate fractional time series model is

$$\Delta_+^d X_t = \varepsilon_t, \tag{1}$$

*I am grateful to the editor, two anonymous referees, James Davidson, Søren Johansen, James MacKinnon, Lealand Morin, and participants at the second CREATES Long Memory Symposium for comments. This research was supported by the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation).

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

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

denoting the coefficients in the usual binomial expansion of $(1-z)^{-u}$, $\Gamma(\cdot)$ denoting the Gamma function, 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 therefore always well defined. 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) 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, and that no prior knowledge needs to be assumed about the value of d. Importantly, this implies that both stationary, nonstationary, and overdifferenced time series are permitted and that the range of admissible values of the fractional parameter can be arbitrarily large.

This paper proves consistency and asymptotic normality results for CSS estimators in multivariate fractional time series models. Although the CSS estimator has found widespread use in the literature, the conditions under which it is consistent in fractional time series models, and especially multivariate fractional time series models, are only recently beginning to be well understood, as explained below. Consistency results are of course 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 in the context of model (1), 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. 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 on 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 by, for example, 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), 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 approximate magnitude of d_0 one can (fractionally) difference the data, estimate d, and add back. See also Hualde & Robinson (2011, pp. 3153–3154) for additional discussion of these issues.

Only very recently, Hualde & Robinson (2011), Lieberman, Rosemarin & Rousseau (2012), and Johansen & Nielsen (2012*a*) have proven consistency for time domain estimators[1] in parametric fractional

[1] In 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, Velasco & Robinson (2000) for tapered Whittle estimation, and Shao (2010) for nonstationarity-extended Whittle estimation. While the latter is efficient under Gaussianity, it differs from CSS estimation by using a "type I" definition of fractional integration and by requiring $d_0 \neq \pm 1/2, \pm 3/2$, etc.

time series models for a large set of admissible values of d. Lieberman et al. (2012), however, considers only univariate stationary Gaussian processes, i.e. $d < 1/2$, and do not allow nonstationary or non-Gaussian processes. Johansen & Nielsen (2012*a*) applies methods that are in some respects similar to the ones in this paper, but considers an entirely different class of fractional vector autoregressive models that is not nested with the multivariate fractional time series models in this paper.

Dealing with the same model setup as the present paper, Hualde & Robinson (2011) give a consistency proof in the univariate case, using a method of proof that is rather different from the one used below. They also consider the multivariate case, but argue that their consistency proof does not straightforwardly extend to that case (see their discussion on pp. 3174–3176), and therefore provide an asymptotic distribution result assuming the existence of a $\sqrt{T}$-consistent initial estimator from which a single Newton step is taken.

On the other hand, this paper provides a full consistency proof for the CSS estimator in multivariate fractional time series models. The consistency proof in Section 3 below shares with the univariate proof of Hualde & Robinson (2011) the idea of analyzing the behavior of the objective function in distinct intervals. However, their proof for the neighborhood around $d - d_0 = -1/2$ requires splitting the critical interval $d - d_0 \in [-1/2 - \kappa_2, -1/2 + \kappa_3]$ into two separate intervals immediately to the left and right, respectively, of the critical point $d - d_0 = -1/2$, which the proof below does not. It is the interval around this critical point that is most delicate to analyze in the proof, since this is where uniform convergence of the objective function fails, and it therefore also poses the greatest challenge in the multivariate case. In particular, the inclusion of the critical interval around $d - d_0 = -1/2$ in the proof below is achieved by a truncation argument, making it possible to show that when $v = d - d_0 \in [-1/2 - \kappa_2, -1/2 + \kappa_3]$, 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 $(\kappa_3, T) \to (0, \infty)$.

An additional complication in CSS estimation in the multivariate p-dimensional case is that the objective function may involve processes whose fractional differencing parameters belong to different intervals (stationary, nonstationary, critical). This presents a significant challenge in the consistency proof. In particular, a variation of the truncation argument that is applied in the critical interval can be applied in this case as well, but only if the nonstationary and critical intervals are separated by a wedge, i.e., if the former interval is bounded from the right by $d - d_0 \leq -1/2 - \kappa_1$ for some $\kappa_1 > \kappa_2$. Although this appears to suggest that a small interval needs to be eliminated from the parameter space, that is in fact not the case. This separation of the nonstationary and critical intervals is achieved in the proof by a careful study of all three intervals individually, followed by a study of each situation where either the nonstationary or the critical interval is left out, and in each of these cases it is shown that the nonstationary and critical intervals may in fact overlap, i.e., $\kappa_1 < \kappa_2$ is allowed. This overlap is then used to prove that, for the general case with all intervals included, the nonstationary and critical intervals can be assumed to be separated by a wedge, such that $\kappa_1 > \kappa_2$ for the general proof. To illustrate the intuition, suppose there are two processes with fractional differencing parameters that are very close and close to the boundary between the nonstationary and critical intervals. Then the overlap between the two intervals implies that the two processes can be assumed to both belong to the nonstationary interval or both belong to the critical interval. In either case, the other interval is not needed for those processes which implies that one interval is left out and that situation has already been studied. In other words, only when these two processes have fractional differencing parameters that are quite distant is it necessary to assume that they are in different intervals, and hence the intervals can be assumed to be separated by a wedge, i.e. $\kappa_1 > \kappa_2$, in which case the truncation argument can be applied.

The remainder of the paper is structured as follows. In the next section the consistency result is presented for a multivariate fractional time series model which allows a wide range of short memory innovations. The consistency proof is quite involved and is presented in Section 3. In Section 4 the asymptotic distribution theory is given with proof in Section 5. Section 6 concludes, and the paper ends with two appendices of auxiliary results used in the main proofs. No empirical applications or finite sample simulations of the CSS estimation procedure are included because these can be found in, e.g., Nielsen & Frederiksen (2005), Hualde & Robinson (2011), and the references therein.

Some comments on notation: for a sequence of stochastic processes $X_T(s) \in \mathbb{R},\ s = (s_1, \ldots, s_m) \in S$, 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 $E|X|^r < \infty$ the L_r-norm is defined as $||X||_r = (E|X|^r)^{1/r}$. A function $f(x) : \mathbb{R}^q \to \mathbb{R}$ satisfies a Lipschitz condition of order α, or is in Lip(α), if there exists a finite constant $K > 0$ such that $|f(x_1) - f(x_2)| \leq K|x_1 - x_2|^\alpha$ for all $x_1, x_2 \in \mathbb{R}^q$. For a function $f(x) : \mathbb{R}^q \to \mathbb{R}$, the k'th order derivative is sometimes denoted $\frac{\partial^k f(x)}{\partial x^{(k)}}$, which is $q \times 1$ and $q \times q$ when $k = 1$ and $k = 2$, respectively. For higher-order derivatives this can be thought of as a matrix of high dimension, the exact form of which is not important. For any vector a, the i'th element is denoted $a^{(i)}$, while for any matrix A, $A^{(i,j)}$, $A^{(i\cdot)}$, and $A^{(\cdot i)}$ denote the (i, j)'th element, i'th row, and i'th column, respectively. Furthermore, $A > B$ ($A \geq B$) denotes that $A - B$ is positive (semi)definite, tr$\{A\}$ denotes the trace of A, $|A| = (\text{tr}\{A'A\})^{1/2}$ is the Euclidean norm of A, and det$\{A\}$ is the determinant of A. Finally, as a convention, it is assumed that $j^{-1} = 0$ for $j = 0$ in summations over j.

2 Consistency result

Let $X_t = (X_{1t}, \ldots, X_{pt})'$ be a p-dimensional time series and generalize the simple model (1) as follows:

$$X_t = \Lambda_+(d)^{-1} u_t \text{ and } u_t = A(L, \psi)\varepsilon_t, \tag{3}$$

where $d = (d_1, \ldots, d_p)'$, $\Lambda_+(d) = \text{diag}(\Delta_+^{d_1}, \ldots, \Delta_+^{d_p})$, ψ is a q-dimensional parameter vector and $A(z, \psi) = \sum_{n=0}^{\infty} A_n(\psi) z^n$ with $p \times p$ matrix coefficients, $A_n(\psi)$. The parametric form of the function $A(z, \psi)$ is assumed known. Model (3) generalizes model (1) to multivariate time series and to allow short memory dynamics (i.e., weak dependence) in u_t. Specifically, u_t is assumed to be a linear process governed by an underlying q-dimensional parameter vector. For example, u_t could be generated by a vector ARMA model or by the exponential spectrum model of Bloomfield (1973), which is somewhat popular in the fractional literature owing to the relatively simple covariance matrix formula it offers in this setting, see, e.g., Robinson (1994).

Model (3) is analyzed under the following assumptions on the errors ε_t and the true parameter values, which are denoted by subscript zero.

Assumption A *The p-dimensional errors ε_t are stationary and ergodic with finite fourth moments and satisfy $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Sigma$ almost surely, where $\mathcal{F}_t = \sigma(\{\varepsilon_s, s \leq t\})$ is the sigma-algebra of events generated by $\varepsilon_s, s \leq t$. Finally, the conditional (on $\mathcal{F}_{t-1}$) third and fourth moments of ε_t are finite and equal the unconditional moments.*

Assumption B *The true parameter values satisfy $\Sigma_0 > 0$ and $(d_0, \psi_0) \in D^p \times \Psi$, where $D = [\nabla_1, \nabla_2]$ with $-\infty < \nabla_1 \leq \nabla_2 < \infty$ and the set $\Psi \subseteq \mathbb{R}^q$ is convex and compact.*

Importantly, only four moments are assumed finite in Assumption A and Gaussianity is not assumed. The errors are assumed to be conditionally homoskedastic martingale differences, which is somewhat weaker than the independence and identical distribution assumption in Johansen & Nielsen (2012*a*). Furthermore, as in Hualde & Robinson (2011, Section 3), positive definiteness of Σ_0 rules out cointegration among the components of X_t. However, even though cointegration, which has been popular especially in recent empirical macroeconomics, is ruled out, the present model can still be applied to test a number of interesting hypotheses such as joint stationarity or $I(0)$-ness, in which case cointegration is not a concern.

The cointegrated case is analyzed in Johansen & Nielsen (2012*a*) using a different model that is not nested with (3).

In Assumption B the short memory parameters ψ are assumed to be in a compact and convex subset of $\mathbb{R}^q$. More importantly, Assumption B permits the length of the interval $D = [\triangledown_1, \triangledown_2]$ of admissible values of d_i to be arbitrarily large. Specifically, the length of D is not limited to less than 1/2 as in most previous studies of fractional time series models that include proofs of consistency. Thus, under Assumption B, the model can simultaneously accommodate both nonstationary, (asymptotically) stationary, and overdifferenced processes.

The following condition is imposed on the linear filter $A(z, \psi)$ and the associated coefficients:

Assumption C *For all $\psi \in \Psi$ and all z in the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$ it holds that:*

(i) $A_0(\psi) = I_p$ *and* $\det\{A(z, \psi)\}$ *is bounded and bounded away from zero.*

(ii) Each element of $A(e^{i\lambda}, \psi)$ *is* $2 + \max(s, 0)$ *times differentiable in* λ *with* $2 + \max(2, 0)$*'th derivative in* $\mathrm{Lip}(\xi)$ *for any* $\xi > 0$ *and* s *defined as the integer part of* $\min(p - 1, \triangledown_2 - \triangledown_1 - 3/2)$.

(iii) $A(z, \psi) = \sum_{n=0}^{\infty} A_n(\psi) z^n$ *is continuously differentiable in* ψ *and the derivatives* $\dot{A}_n^{(i,j)}(\psi) = \frac{\partial A_n^{(i,j)}(\psi)}{\partial \psi}$ *satisfy* $\sum_{n=0}^{\infty} |\dot{A}_n^{(i,j)}(\psi)| < \infty$.

Assumption C(i) ensures invertibility of u_t in (3). Under this assumption the function $B(z, \psi) = A(z, \psi)^{-1} = \sum_{n=0}^{\infty} B_n(\psi) z^n$ is well-defined by its power series expansion for $|z| \leq 1 + \delta$ for some $\delta > 0$, and has $\det\{B(z, \psi)\}$ bounded and bounded away from zero on the complex unit disk. Under Assumption C the $p \times p$ matrix coefficients $A_n(\psi)$ and $B_n(\psi)$ satisfy

$$|A_n(\psi)| = O(n^{-2-\max(s,0)-\xi}) \text{ and } |B_n(\psi)| = O(n^{-2-\max(s,0)-\xi}) \text{ uniformly in } \psi \in \Psi, \tag{4}$$

see Zygmund (2003, pp. 46 and 71). In contrast, under Hualde & Robinson's (2011) univariate Assumption A(ii) the required rate is only $O(n^{-1-\zeta})$ for $\zeta > 1/2$. Assumption C(ii) is the only assumption that differs from those in Hualde & Robinson (2011). The stronger rate required for the multivariate model in this paper illustrates an interesting trade-off: allowing a higher dimensional model with a large parameter space for the fractional parameters (i.e., a larger/positive s) requires more smoothness of the linear coefficients. In any case, Assumption C is easily satisfied by the Bloomfield model or by stationary and invertible ARMA processes due to the exponential decay of their linear representation coefficients.

Thus, letting $\theta = (d, \psi) \in D^p \times \Psi = \Theta$ and letting $B_+(L, \psi) X_t = B(L, \psi) X_t 1_{\{t \geq 1\}} = \sum_{n=0}^{t-1} B_n(\psi) X_{t-n}$ denote the truncated filter, the residuals are defined as

$$\varepsilon_t(\theta) = B_+(L, \psi) \Lambda_+(d) X_t, \tag{5}$$

and the classical least squares or CSS estimator is found by minimizing the sum of squared residuals, i.e.,

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \det\{R(\theta)\}, \tag{6}$$

$$R(\theta) = T^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta) \varepsilon_t(\theta)'. \tag{7}$$

The estimator (6) is well-known from, e.g., Li & McLeod (1986) and Beran (1995) in a (univariate) fractional context and of course Box & Jenkins (1970) for non-fractional models. Motivation for (6) comes from the fact that (the trace of) (7) is proportional to the exponent in the conditional Gaussian (quasi-)likelihood function, and (6) is of course equivalent to the conditional (quasi-)maximum likelihood estimator.

Note the truncation of the autoregressive representation of X_t in calculating the residual in (5), which is inherent to CSS estimation, and presents an additional challenge that is non-trivial in the context of fractional models. For discussion of this issue, see e.g. Robinson (2005) and Robinson (2006), where detailed treatments of the consequences of the truncation are given. 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. The present paper contains a rigorous treatment of the effects of the truncation.

Finally, the following identification condition will also be needed:

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

Assumption D is identical to Assumption A1(i) in Hualde & Robinson (2011) and 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 1 *Suppose X_t is generated by model (3) and satisfies Assumptions A–D, and let $(\hat{d},\hat{\psi})$ be defined by (6). Then $(\hat{d},\hat{\psi}) \overset{P}{\to} (d_0,\psi_0)$ as $T \to \infty$.*

3 Proof of Theorem 1

The residual in (5) is $\varepsilon_t(\theta) = B_+(L,\psi)\Lambda_+(d-d_0)u_t$, and clearly the convergence properties of $R(\theta)$ in (7) depend on the vector $d - d_0$. Let the deterministic function $r(\theta)$ denote the pointwise probability limit of $\det\{R(\theta)\}$, shown subsequently to be given by

$$r(\theta) = \begin{cases} \det\{E(\eta_t(\theta)\eta_t(\theta)')\} & \text{if } d_k - d_{0k} > -1/2 \text{ for } k = 1,\ldots,p, \\ \infty & \text{otherwise,} \end{cases} \tag{8}$$

where the untruncated process $\eta_t(\theta) = B(L,\psi)\Lambda(d-d_0)u_t$ applies the untruncated filter $\Lambda(d) = \text{diag}(\Delta^{d_1},\ldots,\Delta^{d_p})$ and is well-defined when $d_k - d_{0k} > -1/2$ for $k = 1,\ldots,p$. In the latter case, $\eta_t(\theta)$ can be represented in the following convenient way,

$$\begin{aligned} \eta_t(\theta) &= \sum_{n=0}^{\infty} B_n(\psi)\Lambda(d-d_0)u_{t-n} \\ &= \sum_{n=0}^{\infty}\sum_{k=1}^{p} B_n^{(\cdot k)}(\psi)\sum_{j=0}^{\infty}\pi_j(d_{0k}-d_k)u_{t-j-n}^{(k)} \\ &= \sum_{k=1}^{p}\sum_{j=0}^{\infty}\pi_j(d_{0k}-d_k)\sum_{n=0}^{\infty}B_n^{(\cdot k)}(\psi)u_{t-j-n}^{(k)} \\ &= \sum_{k=1}^{p}\Delta^{d_k-d_{0k}}e_{kt}(\psi), \end{aligned} \tag{9}$$

where the linear processes $e_{kt}(\psi) = C_k(L,\psi)\varepsilon_t$, $k = 1,\ldots,p$, are defined using

$$C_k(z,\psi) = B^{(\cdot k)}(z,\psi)A^{(k\cdot)}(z,\psi_0) = \sum_{n=0}^{\infty} C_{kn}(\psi)z^n. \tag{10}$$

In Lemma 1, presented below, a similar representation is given for $\varepsilon_t(\theta)$.

From Assumption C the $p \times p$ matrix coefficients $C_{kn}(\psi)$ in (10) satisfy

$$|C_{kn}(\psi)| = O(n^{-2-\max(s,0)-\xi}) \text{ uniformly in } \psi \in \Psi. \tag{11}$$

In the following, the Beveridge-Nelson decomposition,

$$e_{kt}(\psi) = C_k(L,\psi)\varepsilon_t = C_k^*(\psi)\varepsilon_t + \Delta \sum_{n=0}^{\infty} \tilde{C}_{kn}(\psi)\varepsilon_{t-n}, \tag{12}$$

will be applied repeatedly, where $C_k^*(\psi) = \sum_{n=0}^{\infty} C_{kn}(\psi)$ is finite and $\tilde{C}_{kn}(\psi) = -\sum_{m=n+1}^{\infty} C_{km}(\psi) = O(n^{-1-\max(s,0)-\xi})$ uniformly in $\psi \in \Psi$, see (11) and also Phillips & Solo (1992, Lemma 2.1).

Conforming with (8), the parameter space $D = [\triangledown_1, \triangledown_2]$ for d_i is partitioned into three compact subsets, $D_1 = D_1(\kappa_1) = D \cap \{d_i : d_i - d_{0i} \leq -1/2 - \kappa_1\}$, $D_2 = D_2(\kappa_2, \kappa_3) = D \cap \{d_i : -1/2 - \kappa_2 \leq d_i - d_{0i} \leq -1/2 + \kappa_3\}$, and $D_3 = D_3(\kappa_3) = D \cap \{d_i : d_i - d_{0i} \geq -1/2 + \kappa_3\}$, for some constants $0 < \kappa_1 < \kappa_2 < \kappa_3 < 1/2$, to be determined later. Note that $\cup_{i=1}^3 D_i = D$ and that there is an overlap between D_1 and D_2 because $\kappa_1 < \kappa_2$, which is used explicitly in Section 3.3.7. Special care is taken with respect to D_2, where the convergence of the objective function is non-uniform, as evident in (8), see Section 3.3.1.

Clearly, $\theta_0 \in D_3^p \times \Theta$ and if $\triangledown_1 > \max_i d_{0i} - 1/2$ then the choice $\kappa_3 = \triangledown_1 - \max_i d_{0i} + 1/2 > 0$ implies that D_1 and D_2 are empty for all $i = 1, \ldots, p$, in which case the proof is easily simplified accordingly.

The proof of Theorem 1 proceeds by showing the following results. For any $K > 0$ there exists a (fixed) $\bar{\kappa}_3 > 0$ such that

$$P\left(\inf_{d \in D^p \setminus D_3(\bar{\kappa}_3)^p, \psi \in \Psi} \det\{R(\theta)\} > K\right) \to 1 \text{ as } T \to \infty. \tag{13}$$

This implies that $P(\hat{\theta} \in D_3(\bar{\kappa}_3)^p \times \Psi) \to 1$ as $T \to \infty$, so that the relevant parameter space is reduced to $\Theta_3(\bar{\kappa}_3) = D_3(\bar{\kappa}_3)^p \times \Psi$. From Theorem 5.7 of van der Vaart (1998) the desired result then follows if, for any fixed $\kappa_3 \in (0, 1/2)$,

$$\sup_{\theta \in \Theta_3(\kappa_3)} |\det\{R(\theta)\} - \det\{r(\theta)\}| \overset{P}{\to} 0 \text{ as } T \to \infty, \tag{14}$$

$$\inf_{\theta \in \Theta_3(\kappa_3) \cap \{\theta : |\theta - \theta_0| \geq \delta\}} \det\{r(\theta)\} > \det\{r(\theta_0)\} \text{ for all } \delta > 0. \tag{15}$$

The first condition entails uniform convergence of the objective function on Θ_3, and the second condition ensures that the optimum of the limit function is uniquely attained at the true value.

The proofs of (14), (15), and (13) are given in Sections 3.1, 3.2, and 3.3, respectively. The most delicate part is justifying (13) because in (13) the integration orders of the different processes involved may fall into different subsets. Hence this presents the main complication for the multivariate case relative to the univariate case.

Before proceding to show (13)–(15), the following lemma gives a representation of $\varepsilon_t(\theta)$ similar to (9) and also analyzes the effect of the truncation in the residual in the definition of $R(\theta)$. The result is given for D_2 and D_3, while the corresponding result for D_1 is more conveniently given in Section 3.3.2, see Lemma 2.

Lemma 1 *With the notation of this section it holds that*

$$\varepsilon_t(\theta) = \eta_t^+(\theta) + \sum_{k=1}^{p} r_{kt}(d_k, \psi), \tag{16}$$

$$\eta_t^+(\theta) = \sum_{k=1}^{p} \Delta_+^{d_k - d_{0k}} e_{kt}(\psi), \tag{17}$$

where, under the assumptions of Theorem 1,

$$\sup_{d_k \in D_2 \cup D_3, \psi \in \Psi} |T^{-1} \sum_{t=1}^{T} r_{kt}(d_k, \psi) r_{kt}(d_k, \psi)'| \overset{P}{\to} 0 \text{ as } T \to \infty. \tag{18}$$

Proof. *Proof of (16)–(17)*: As in (9), $\varepsilon_t(\theta) = B_+(L, \psi)\Lambda_+(d - d_0)u_t$ is

$$\begin{aligned}
\varepsilon_t(\theta) &= \sum_{n=0}^{t-1} \sum_{k=1}^{p} B_n^{(\cdot k)}(\psi) \sum_{j=0}^{t-n-1} \pi_j(d_{0k} - d_k) u_{t-j-n}^{(k)} \\
&= \sum_{k=1}^{p} \sum_{j=0}^{t-1} \pi_j(d_{0k} - d_k) \sum_{n=0}^{t-j-1} B_n^{(\cdot k)}(\psi) u_{t-j-n}^{(k)} \\
&= \sum_{k=1}^{p} \Delta_+^{d_k - d_{0k}} e_{kt}(\psi) - \sum_{k=1}^{p} \sum_{j=0}^{t-1} \pi_j(d_{0k} - d_k) \sum_{n=t-j}^{\infty} B_n^{(\cdot k)}(\psi) u_{t-j-n}^{(k)}
\end{aligned}$$

such that $r_{kt}(d_k, \psi) = -\sum_{j=0}^{t-1} \pi_j(d_{0k} - d_k) \sum_{n=t-j}^{\infty} B_n^{(\cdot k)}(\psi) u_{t-j-n}^{(k)}$.

Proof of (18): First note that $r_{kt}(d_k, \psi) = \sum_{m=t}^{\infty} \phi_{tmk} u_{t-m}^{(k)}$, where $\phi_{tmk} = -\sum_{j=0}^{t-1} \pi_j(d_{0k} - d_k) B_{m-j}^{(\cdot k)}(\psi)$ satisfies, see (4) and Lemmas A.1 and A.2,

$$\begin{aligned}
\sup_{\psi \in \Psi} \sum_{m=t}^{\infty} |\phi_{tmk}| &\leq c \sum_{m=t}^{\infty} \sum_{j=1}^{t-1} j^{d_{0k} - d_k - 1} (m - j)^{-2-\xi} \\
&\leq c \sum_{j=1}^{t-1} j^{d_{0k} - d_k - 1} (t - j)^{-1-\xi} \\
&\leq c (\log t) t^{\max(d_{0k} - d_k, -\xi) - 1}.
\end{aligned}$$

Because $|u_{t-m}^{(k)} u_{t-n}^{(k)}| = O_P(1)$ uniformly in t, n, m, the sum of squares of $r_{kt}(d_k, \psi)$ satisfies the bound

$$\begin{aligned}
\sup_{\psi \in \Psi} \sum_{t=1}^{T} |r_{kt}(d_k, \psi) r_{kt}(d_k, \psi)'| &= \sup_{\psi \in \Psi} \sum_{t=1}^{T} \sum_{m=t}^{\infty} \sum_{n=t}^{\infty} |\phi_{tmk} u_{t-m}^{(k)} u_{t-n}^{(k)} \phi_{tnk}'| \\
&\leq \sup_{\psi \in \Psi} \sum_{t=1}^{T} \left(\sum_{m=t}^{\infty} |\phi_{tmk}| \right)^2 O_P(1) \\
&= O_P\left(\sum_{t=1}^{T} (\log t)^2 t^{2\max(d_{0k} - d_k, -\xi) - 2} \right) \\
&= O_P((\log T)^2 T^{2\max(d_{0k} - d_k - 1/2, -\xi - 1/2, 0)})
\end{aligned}$$

so that

$$\begin{aligned}
\sup_{d_k \in D_2 \cup D_3, \psi \in \Psi} |T^{-1} \sum_{t=1}^{T} r_{kt}(d_k, \psi) r_{kt}(d_k, \psi)'| &= O_P(\sup_{d_k \in D_2 \cup D_3} (\log T)^2 T^{-1+2\max(d_{0k} - d_k - 1/2, -\xi - 1/2, 0)}) \\
&= O_P((\log T)^2 T^{2\kappa_2 - 1}).
\end{aligned}$$

■

3.1 Convergence on $\Theta_3(\kappa_3)$ and proof of (14)

First of all, if $\theta \in \Theta_3(\kappa_3)$ for any $\kappa_3 > 0$ then $\varepsilon_t(\theta)$ is (asymptotically) stationary. By Lemma 1 the difference between $R(\theta)$ and $S(\theta) = T^{-1}\sum_{t=1}^{T}\eta_t^+(\theta)\eta_t^+(\theta)'$ is negligible in probability uniformly in $d \in D_3(\kappa_3)^p, \psi \in \Psi$, so it suffices to consider $S(\theta)$. The difference between $\eta_t^+(\theta)$ and the stationary and ergodic process $\eta_t(\theta)$ (without truncation), see (9) and (17), is

$$\eta_t(\theta) - \eta_t^+(\theta) = \sum_{k=1}^{p}\sum_{n=t}^{\infty}\pi_n(d_{0k}-d_k)e_{k,t-n}(\psi) = \sum_{n=t}^{\infty}\tau_n(\theta)\varepsilon_{t-n}, \tag{19}$$

where $\tau_n(\theta) = \sum_{k=1}^{p}\sum_{m=0}^{n}\pi_m(d_{0k}-d_k)C_{k,n-m}(\psi)$ and $\sup_{\theta\in\Theta_3}|\tau_n(\theta)| \le c(1+\log n)n^{\max(-1/2-\kappa_3,-2-\xi)} \le c(1+\log n)n^{-1/2-\kappa_3}$ by (11) and Lemma A.2. It follows that

$$\begin{aligned} E(\eta_t(\theta)-\eta_t^+(\theta))(\eta_t(\theta)-\eta_t^+(\theta))' &= \sum_{n=t}^{\infty}\tau_n(\theta)\Sigma_0\tau_n(\theta)' \\ &\le c\sum_{n=t}^{\infty}(1+\log n)^2 n^{-1-2\kappa_3} \le c(1+\log t)^2 t^{-2\kappa_3} \to 0 \end{aligned}$$

for all $\theta \in \Theta_3$ (pointwise). From the law of large numbers for stationary and ergodic processes it then holds that

$$S(\theta) = T^{-1}\sum_{t=1}^{T}\eta_t(\theta)\eta_t(\theta)' + o_P(1) \xrightarrow{P} E(\eta_t(\theta)\eta_t(\theta)') \text{ as } T \to \infty, \tag{20}$$

which shows the pointwise limit in probability, see (8).

The result (20) can be strengthened to uniform convergence in probability by showing that $S(\theta)$ is stochastically equicontinuous (or tight). From Billingsley (1968, Problem 6.6) this holds if $S^{(i,j)}(\theta)$ is stochastically equicontinuous for each $i,j = 1,\ldots,p$, and from Newey (1991, Corollary 2.2) this holds if the derivative of $T^{-1}\sum_{t=1}^{T}\eta_t^{+(i)}(\theta)\eta_t^{+(j)}(\theta) = \sum_{k,l=1}^{p}T^{-1}\sum_{t=1}^{T}(\Delta_+^{d_k-d_{0k}}e_{kt}^{(i)}(\psi))(\Delta_+^{d_l-d_{0l}}e_{lt}^{(j)}(\psi))$ is dominated uniformly in $\theta \in \Theta_3$ by a random variable $B_T = O_P(1)$. From Lemma B.3 with $u_1 = d_k - d_{0k} \ge -1/2+\kappa_3, u_2 = d_l - d_{0l} \ge -1/2+\kappa_3, a = 2\kappa_3$, and $\tilde{\Psi} = \Psi$ (noting that only summability of the linear coefficients is assumed in Lemma B.3 and this is satisfied uniformly on Ψ by the derivatives of $C_{kn}(\psi)$ by Assumption C(iii)), it holds that $B_T = \sup_{\theta\in\Theta_3}|\frac{\partial S^{(i,j)}(\theta)}{\partial\theta}| = O_P(1)$, showing that $S(\theta)$ is stochastically equicontinuous on Θ_3 and hence that (20) holds uniformly in $\theta \in \Theta_3$. Because the result holds for any κ_3 it proves (14).

3.2 Proof of (15)

If $A > 0$ and $B \ge 0$ then $\det\{A+B\} \ge \det\{A\}$ with equality if and only if $B = 0$. Since $\det\{r(\theta_0)\} = \det\{\Sigma_0\}$ it is therefore sufficient to prove that

$$\inf_{\theta\in\Theta_3(\kappa_3)\cap\{\theta:|\theta-\theta_0|\ge\delta\}} E(\eta_t(\theta)\eta_t(\theta)') - \Sigma_0 \ge 0 \text{ (and } \neq 0) \text{ for all } \delta > 0 \text{ and all } \kappa_3 \in (0,1/2).$$

The variance of $\eta_t(\theta) = B(L,\psi)\Lambda(d-d_0)u_t = \sum_{n=0}^{\infty}\tau_n(\theta)\varepsilon_{t-n}$ is $E(\eta_t(\theta)\eta_t(\theta)') = \sum_{n=0}^{\infty}\tau_n(\theta)\Sigma_0\tau_n(\theta)'$, where $\tau_n(\theta) = \sum_{k=1}^{p}\sum_{m=0}^{n}\pi_m(d_{0k}-d_k)C_{k,n-m}(\psi)$ with $\tau_0(\theta) = \sum_{k=1}^{p}\pi_0(d_{0k}-d_k)B_0^{(\cdot k)}(\psi)A_0^{(k\cdot)}(\psi_0) = \sum_{k=1}^{p}I_p^{(\cdot k)}I_p^{(k\cdot)} = I_p$ for all $\theta \in \Theta_3$ by Assumption C. It follows that $\sum_{n=0}^{\infty}\tau_n(\theta)\Sigma_0\tau_n(\theta)' = \Sigma_0 + \sum_{n=1}^{\infty}\tau_n(\theta)\Sigma_0\tau_n(\theta)'$, where the last term is zero if and only if $\theta = \theta_0$ by Assumption D. Hence, $E(\eta_t(\theta)\eta_t(\theta)') - \Sigma_0 = \sum_{n=1}^{\infty}\tau_n(\theta)\Sigma_0\tau_n(\theta)' \ge 0$ for all $\theta \neq \theta_0$, which proves (15) by continuity of $\tau_n(\cdot)$ and compactness of Θ_3.

3.3 Proof of (13)

The proof of (13) is structured as follows. First, the proof is given for the case with $\theta \in \Theta_2 = D_2^p \times \Psi$ in Section 3.3.1. Then the proof for the case with $\theta \in \Theta_1 = D_1^p \times \Psi$ is given in Section 3.3.2. In Section 3.3.3 some notation is introduced for the proof in the general case. Sections 3.3.4–3.3.6 consider cases with processes in at most two different subsets. In Section 3.3.7 an overlap argument is used to show that a wedge can be inserted between κ_1 and κ_2, which will subsequently be used in Section 3.3.8 to prove (13) in the general case. In each of Sections 3.3.5–3.3.8, the processes involved may fall into different subsets.

3.3.1 Convergence on $\Theta_2(\kappa_2, \kappa_3)$

First note that, by (18) of Lemma 1, it suffices to prove the result for $S^+(\theta) = T^{-1}\sum_{t=1}^T \eta_t^+(\theta)\eta_t^+(\theta)'$. Using the notation $\varepsilon_{kt}^*(\psi) = C_k^*(\psi)\varepsilon_t$, the Beveridge-Nelson decomposition (12) shows that the product moment $S^+(\theta)$ can be decomposed as

$$S^+(\theta) \geq T^{-1}\sum_{k,l=1}^p \sum_{t=1}^T (\Delta_+^{d_k-d_{0k}}\varepsilon_{kt}^*(\psi))(\Delta_+^{d_l-d_{0l}}\varepsilon_{lt}^*(\psi))' \tag{21}$$

$$+T^{-1}\sum_{k,l=1}^p \sum_{t=1}^T (\Delta_+^{d_k-d_{0k}}\varepsilon_{kt}^*(\psi))(\Delta_+^{d_l-d_{0l}+1}\sum_{n=0}^\infty \tilde{C}_{ln}(\psi)\varepsilon_{t-n})' \tag{22}$$

$$+T^{-1}\sum_{k,l=1}^p \sum_{t=1}^T (\Delta_+^{d_k-d_{0k}+1}\sum_{n=0}^\infty \tilde{C}_{kn}(\psi)\varepsilon_{t-n})(\Delta_+^{d_l-d_{0l}}\varepsilon_{lt}^*(\psi))'. \tag{23}$$

The (i,j)'th element of (23) is $T^{-1}\sum_{k,l=1}^p\sum_{t=1}^T(\Delta_+^{d_k-d_{0k}+1}\sum_{n=0}^\infty \tilde{C}_{kn}^{(i\cdot)}(\psi)\varepsilon_{t-n})(\Delta_+^{d_l-d_{0l}}C_l^{*(j\cdot)}(\psi)\varepsilon_t)' = O_P(1)$ uniformly in $\theta \in \Theta_2(\kappa,\kappa)$ for any $0 < \kappa < 1/2$ by Lemma B.3 with $u_1 = d_k - d_{0k} + 1 \geq 1/2 - \kappa, u_2 = d_l - d_{0l} \geq -1/2 - \kappa$ such that $u_1 + u_2 \geq -2\kappa$ and $a = \min(1/2-\kappa, 1-2\kappa) > 0$. The proof for (22) is identical.

Let the right-hand side of (21) be denoted $S_1^+(d,\psi) = \sum_{k,l=1}^p T^{-1}\sum_{t=1}^T(\Delta_+^{d_k-d_{0k}}\varepsilon_{kt}^*(\psi))(\Delta_+^{d_l-d_{0l}}\varepsilon_{lt}^*(\psi))'$ and define the p-vectors $w_{kt} = \sum_{n=0}^{N-1}\pi_n(d_{0k}-d_k)\varepsilon_{k,t-n}^*(\psi)$ and $v_{kt} = \sum_{n=N}^{t-1}\pi_n(d_{0k}-d_k)\varepsilon_{k,t-n}^*(\psi)$. To analyze $S_1^+(d,\psi)$, decompose $\Delta_+^{d_k-d_{0k}}\varepsilon_{kt}^*(\psi)$ as

$$\Delta_+^{d_k-d_{0k}}\varepsilon_{kt}^*(\psi) = \sum_{n=0}^{t-1}\pi_n(d_{0k}-d_k)\varepsilon_{k,t-n}^*(\psi) = w_{kt} + v_{kt}, \quad t \geq N+1,$$

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

$$S_1^+(d,\psi) \geq \sum_{k,l=1}^p T^{-1}\sum_{t=N+1}^T (\Delta_+^{d_k-d_{0k}}\varepsilon_{kt}^*(\psi))(\Delta_+^{d_l-d_{0l}}\varepsilon_{lt}^*(\psi))'$$

$$\geq T^{-1}\sum_{t=N+1}^T\sum_{k,l=1}^p w_{kt}w_{lt}' + T^{-1}\sum_{t=N+1}^T\sum_{k,l=1}^p w_{kt}v_{lt}' + T^{-1}\sum_{t=N+1}^T\sum_{k,l=1}^p v_{kt}w_{lt}'. \tag{24}$$

Setting $N = T^\alpha$ with $0 < \alpha < \min(\frac{1/2-\kappa}{1/2+\kappa}, \frac{1/2}{1/2+2\kappa})$ for some κ satisfying $\max(\kappa_2,\kappa_3) \leq \kappa < 1/2$, noting that such an α exists because $0 < \max(\kappa_2,\kappa_3) < 1/2$, it follows from (47) of Lemma B.2 that the second and third terms on the right-hand side of (24) converge in probability to zero uniformly in $\theta \in \Theta_2(\kappa,\kappa) \supseteq \Theta_2(\kappa_2,\kappa_3)$ and that the first term on the right-hand side of (24) satisfies

$$\sup_{\theta\in\Theta_2(\kappa,\kappa)}\left|T^{-1}\sum_{t=T^\alpha+1}^T\sum_{k,l=1}^p w_{kt}w_{lt}' - \sum_{k,l=1}^p C_k^*(\psi)\Sigma_0 C_l^*(\psi)' F_{T^\alpha}^{(k,l)}(d-d_0)\right| \overset{P}{\to} 0 \text{ as } T\to\infty,$$

where $F_N^{(k,l)}(d-d_0)=\sum_{n=0}^{N-1}\pi_n(d_{0k}-d_k)\pi_n(d_{0l}-d_l)$, see Lemma A.3.

Next, using (10), write $\sum_{k=1}^{p}C_k^*(\psi)\pi_n(d_{0k}-d_k)=B(1,\psi)H_n(d)A(1,\psi_0)$, where $H_n(d)=\mathrm{diag}(\pi_n(d_{01}-d_1),\ldots,\pi_n(d_{0p}-d_p))$, so that

$$\sum_{k,l=1}^{p}C_k^*(\psi)\Sigma_0C_l^*(\psi)'F_N^{(k,l)}(d-d_0)=\sum_{n=0}^{N-1}B(1,\psi)H_n(d)\Gamma_0H_n(d)B(1,\psi)'$$

with $\Gamma_0=A(1,\psi_0)\Sigma_0A(1,\psi_0)'$. The (k,l)'th element of $\sum_{n=0}^{N-1}H_n(d)\Gamma_0H_n(d)$ is $\Gamma_0^{(k,l)}\sum_{n=0}^{N-1}\pi_n(d_{0k}-d_k)\pi_n(d_{0l}-d_l)=\Gamma_0^{(k,l)}F_N^{(k,l)}(d-d_0)$ so that

$$\sum_{n=0}^{N-1}H_n(d)\Gamma_0H_n(d)=\Gamma_0\odot F_N(d-d_0),$$

where $\odot$ denotes the elementwise (Hadamard) product and $F_N(u)$ is the $p\times p$ matrix with (k,l)'th element $F_N^{(k,l)}(u)$. Hence,

$$\begin{aligned}\det\{\textstyle\sum_{k,l=1}^{p}C_k^*(\psi)\Sigma_0C_l^*(\psi)'F_N^{(k,l)}(d-d_0)\} &= \det\{B(1,\psi)(\Gamma_0\odot F_N(d-d_0))B(1,\psi)'\}\\ &= \det\{B(1,\psi)\}^2\det\{\Gamma_0\odot F_N(d-d_0)\}\\ &\geq \det\{B(1,\psi)\}^2\det\{\Gamma_0\}|F_N^{(1,1)}(d-d_0)|\cdots|F_N^{(p,p)}(d-d_0)|\end{aligned}$$

by Oppenheim's Inequality.

Thus,

$$S_1^+(d,\psi)\geq\sum_{k,l=1}^{p}C_k^*(\psi)\Sigma_0C_l^*(\psi)'F_{T^\alpha}^{(k,l)}(d-d_0)+\mu_{1T}(\theta),$$

where $\mu_{1T}(\theta)\overset{P}{\to}0$ as $T\to\infty$ uniformly in $\theta\in\Theta_2(\kappa,\kappa)\supseteq\Theta_2(\kappa_2,\kappa_3)$ and it follows that

$$\det\{S^+(\theta)\}\geq\det\{B(1,\psi)\}^2\det\{A(1,\psi_0)\}^2\det\{\Sigma_0\}|F_N^{(1,1)}(d-d_0)|\cdots|F_N^{(p,p)}(d-d_0)|+\mu_{2T}(\theta),$$

where $\mu_{2T}(\theta)=O_P(1)$ as $T\to\infty$ uniformly in $\theta\in\Theta_2(\kappa,\kappa)\supseteq\Theta_2(\kappa_2,\kappa_3)$. From Lemma A.3,

$$F_{T^\alpha}^{(k,k)}(d-d_0)\geq1+c\frac{1-(T-1)^{-2\alpha\kappa_3}}{2\kappa_3}$$

for $d\in D_2(\kappa_2,\kappa_3)^p$ and $k=1,\ldots,p$. The factor $(2\kappa_3)^{-1}(1-(T-1)^{-2\alpha\kappa_3})$ is increasing in T from 0 (for $T=2$) to $(2\kappa_3)^{-1}$ and decreasing in κ_3 from $\alpha\log(T-1)$ (for $\kappa_3=0$) to 0, such that $(2\kappa_3)^{-1}(1-(T-1)^{-2\alpha\kappa_3})\to\infty$ as $(\kappa_3,T)\to(0,\infty)$. Because $\det\{\Sigma_0\}>0$ by Assumption B and $\det\{A(1,\psi)\}>0$, $\det\{B(1,\psi)\}>0$ uniformly in $\psi\in\Psi$ by Assumption C, it follows that for any $K>0,\eta>0$, there exists $\bar\kappa_3>0$ and $T_2\geq1$ such that

$$P(\inf_{\theta\in\Theta_2(\kappa_2,\bar\kappa_3)}\det\{R(\theta)\}>K)\geq1-\eta\text{ for all }T\geq T_2.\tag{25}$$

Note that (25) holds for any choice of $\kappa_2\in(0,1/2)$.

3.3.2 Convergence on $\Theta_1(\kappa_1)$

First of all, if $d\in D_1(\kappa_1)^p$ then $d_{0k}-d_k\geq1/2+\kappa_1$ such that $\varepsilon_t(\theta)=\sum_{k=1}^{p}\sum_{n=0}^{t-1}B_n^{(\cdot k)}(\psi)\sum_{j=0}^{t-n-1}\pi_j(d_{0k}-d_k)u_{t-j-n}^{(k)}$ is a linear combination of nonstationary processes. To normalize $R(\theta)$ correctly for convergence, some rotation is convenient. To that end, apply the Beveridge-Nelson decomposition (12) iteratively to $B^{(\cdot k)}(\psi, z)$,

$$B^{(\cdot k)}(z,\psi) = \sum_{n=0}^{\infty} B_n^{(\cdot k)}(\psi) z^n = \sum_{h=0}^{s} \tilde{B}_{(h)}^{(\cdot k)}(\psi)(1-z)^h + \sum_{n=0}^{\infty} \tilde{B}_{(s+1),n}^{(\cdot k)}(\psi)(1-z)^{s+1} z^n, \tag{26}$$

where $\tilde{B}_{(h)}^{(\cdot k)}(\psi) = \sum_{n=0}^{\infty} \tilde{B}_{(h),n}^{(\cdot k)}(\psi)$, $\tilde{B}_{(h),n}^{(\cdot k)}(\psi) = -\sum_{m=n+1}^{\infty} \tilde{B}_{(h-1),m}^{(\cdot k)}(\psi), h \geq 1$, $\tilde{B}_{(0),n}^{(\cdot k)}(\psi) = B_n^{(\cdot k)}(\psi)$, and s is the integer part of $\min(p-1, \nabla_2 - \nabla_1 - 3/2)$ as defined in Assumption C(ii). Suppose, without loss of generality, that $d_1 - d_{01} \leq \cdots \leq d_p - d_{0p}$. In view of the iterated Beveridge-Nelson decomposition (26), if $\tilde{B}_{(0)}^{(\cdot 1)}(\psi) \neq 0$, introduce the direction vector $\beta_1(\theta) = \tilde{B}_{(0)}^{(\cdot 1)}(\psi)/|\tilde{B}_{(0)}^{(\cdot 1)}(\psi)|$ such that $\beta_1(\theta)'\varepsilon_t(\theta)$ is fractionally differenced of order $d_1 - d_{01}$ (i.e., fractionally integrated of order $d_{01} - d_1$). Since it is clearly possible that the next term in the decomposition (26) for $k = 1$ can have both higher and lower fractional order than the first term for $k = 2$, the next direction depends on the difference between $d_2 - d_{02}$ and $d_1 - d_{01} + 1$. Hence, if $d_1 - d_{01} + 1 < d_2 - d_{02}$ and $\beta_1(\theta)'\tilde{B}_{(1)}^{(\cdot 1)}(\psi) \neq 0$ then define $\beta_2(\theta) = \tilde{B}_{(1)}^{(\cdot 1)}(\psi)/|\tilde{B}_{(1)}^{(\cdot 1)}(\psi)|$. If $d_2 - d_{02} \leq d_1 - d_{01} + 1$ and $\tilde{B}_{(0)}^{(\cdot 2)}(\psi) \neq 0$ define $\beta_2(\theta) = \tilde{B}_{(0)}^{(\cdot 2)}(\psi)/|\tilde{B}_{(0)}^{(\cdot 2)}(\psi)|$. This procedure defines $\beta_a(\theta)$ for $a = 1, \ldots, p$, and stacking these vectors next to each other defines the rotation matrix $\beta(\theta) = [\beta_1(\theta), \ldots, \beta_p(\theta)]$, which is orthonormal by contruction.

Let the fractional differencing order of $\beta_a(\theta)'\varepsilon_t(\theta)$ be denoted δ_a. Then the following version of Lemma 1 holds for the nonstationary case.

Lemma 2 *With the notation of this section and under the assumptions of Theorem 1, for* $a, b = 1, \ldots, p$,

$$T^{\delta_a+\delta_b+1}\beta_a(\theta)'R(\theta)\beta_b(\theta) = T^{\delta_a+\delta_b}\sum_{t=1}^{T}(\gamma_a(\theta)'\Delta_+^{\delta_a}\varepsilon_t)(\gamma_b(\theta)'\Delta_+^{\delta_b}\varepsilon_t) + r_T(\theta), \tag{27}$$

where $r_T(\theta) = o_P(1)$ *uniformly in* $\theta \in \Theta_1(\kappa_1)$ *and* $\gamma_a(\theta) = |\tilde{B}_{(q)}^{(\cdot l)}(\psi)|(\sum_{m=0}^{\infty} A_m^{(l\cdot)}(\psi_0))'$ *when* $\beta_a(\theta)$ *is obtained from the above procedure as the q'th term in the iterated Beveridge-Nelson decomposition (26) for* $k = l$.

Proof. If $\beta_a(\theta)$ is obtained from the above procedure as the q'th term in the iterated Beveridge-Nelson decomposition (26) for $k = l$ then $\beta_a(\theta) = \tilde{B}_{(q)}^{(\cdot l)}(\psi)/|\tilde{B}_{(q)}^{(\cdot l)}(\psi)|$ and $\delta_a = d_l - d_{0l} + q$. Here, $0 \leq q \leq s$ because there can be at most s nonstationary terms since the $(s+1)$'th term will either be stationary or $\beta_p(\theta)$ will have been reached already.

Applying a finite-summation version of the iterated Beveridge-Nelson decomposition (26), it holds that

$$B_+^{(\cdot k)}(z,\psi) = \sum_{n=0}^{t-1} B_n^{(\cdot k)}(\psi) z^n = \sum_{h=0}^{q}\sum_{n=0}^{t-h-1} B_{(h),n}^{(\cdot k)}(\psi)(1-z)^h + \sum_{n=0}^{t-q-2} B_{(q+1),n}^{(\cdot k)}(\psi)(1-z)^{q+1} z^n, \tag{28}$$

with $B_{(h),n}^{(\cdot k)}(\psi) = -\sum_{m=n+1}^{t-1} B_{(h-1),m}^{(\cdot k)}(\psi), h \geq 1$, and $B_{(0),n}^{(\cdot k)}(\psi) = B_n^{(\cdot k)}(\psi)$. Thus,

$$\beta_a(\theta)'\varepsilon_t(\theta) = \gamma_a(\theta)'\Delta_+^{\delta_a}\varepsilon_t + r_{1t}(\theta) + r_{2t}(\theta) + r_{3t}(\theta) + r_{4t}(\theta),$$

where

$$
\begin{aligned}
r_{1t}(\theta) &= \beta_a(\theta)' \sum_{h=0}^{q-1} (\sum_{n=0}^{t-h-1} B_{(h),n}^{(\cdot l)}(\psi)) \sum_{j=0}^{t-1} \pi_j(d_{0l} - d_l - h) u_{t-j}^{(l)}, q \geq 1, \\
r_{2t}(\theta) &= \beta_a(\theta)' \sum_{n=0}^{t-q-2} B_{(q+1),n}^{(\cdot l)}(\psi) \sum_{j=0}^{t-n-1} \pi_j(d_{0l} - d_l - q - 1) u_{t-j-n}^{(l)}, \\
r_{3t}(\theta) &= \beta_a(\theta)' (\sum_{n=0}^{t-q-1} B_{(q),n}^{(\cdot l)}(\psi)) \sum_{j=0}^{t-1} \pi_j(-\delta_a) u_{t-j}^{(l)} - \gamma_a(\theta)' \Delta_+^{\delta_a} \varepsilon_t, \\
r_{4t}(\theta) &= \beta_a(\theta)' \sum_{k=1, k \neq l}^{p} \sum_{n=0}^{t-1} B_n^{(\cdot k)}(\psi) \sum_{j=0}^{t-n-1} \pi_j(d_{0k} - d_k) u_{t-j-n}^{(k)},
\end{aligned}
$$

and the terms $r_{it}(\theta), i = 1, \ldots, 4$, are shown next to be asymptotically negligible uniformly in $\theta \in \Theta_1$.

First, for $r_{1t}(\theta)$, note that by construction of $\beta_a(\theta)$ it holds that $\beta_a(\theta)' \tilde{B}_{(h)}^{(\cdot l)}(\psi) = 0$ for $0 \leq h \leq q-1$, so that

$$
\begin{aligned}
r_{1t}(\theta) &= -\beta_a(\theta)' \sum_{h=0}^{q-1} (\sum_{n=t-h}^{\infty} B_{(h),n}^{(\cdot l)}(\psi)) \Delta_+^{\delta_a + h - q} u_t^{(l)} \\
&\quad + \beta_a(\theta)' \sum_{h=0}^{q-1} (\sum_{n=0}^{\infty} B_{(h),n}^{(\cdot l)}(\psi) - \tilde{B}_{(h)}^{(\cdot l)}(\psi)) \Delta_+^{\delta_a + h - q} u_t^{(l)}.
\end{aligned}
$$

By (4), $|\sum_{n=t-h}^{\infty} B_{(h),n}^{(\cdot l)}(\psi)| \leq c t^{-1-s+h-\xi}$ and $\sum_{n=0}^{\infty} (B_{(h),n}^{(\cdot l)}(\psi) - \tilde{B}_{(h),n}^{(\cdot l)}(\psi)) = 1_{\{h \geq 1\}} \sum_{n=t}^{\infty} (\frac{1}{h!} \prod_{j=0}^{h-1} (n - j)) B_n^{(\cdot l)}(\psi) \leq 1_{\{h \geq 1\}} c t^{-1-s+h-\xi}$, both uniformly in $\psi \in \Psi$. Using this together with Lemma A.1 and $u_{t-j} = O_P(1)$ shows

$$
\begin{aligned}
\sup_{\theta \in \Theta_1} T^{2\delta_a} \sum_{t=1}^{T} r_{1t}(\theta)^2 &= O_P(\sup_{\delta_a \leq -1/2 - \kappa_1} T^{2\delta_a} \sum_{t=1}^{T} (\sum_{h=0}^{q-1} t^{-1-s+h-\xi} \sum_{j=1}^{t-1} j^{q-h-\delta_a-1})^2) \\
&= O_P(\sup_{\delta_a \leq -1/2 - \kappa_1} T^{2\delta_a} \sum_{t=1}^{T} t^{-2-2s-2\xi+2q-2\delta_a}) \\
&= O_P(\sup_{\delta_a \leq -1/2 - \kappa_1} (\log T) T^{2\delta_a} T^{\max(0, -1-2s-2\xi+2q-2\delta_a)}) \\
&= O_P((\log T) T^{\max(-1-2\kappa_1, -1-2\xi)}) \xrightarrow{P} 0,
\end{aligned}
$$

where the last equality used $q \leq s$.

For $r_{2t}(\theta)$, note first that the coefficients $B_{(q+1),n}^{(\cdot l)}(\psi)$ are absolutely summable by Assumption C(ii), see (4). If $d_l - d_{0l} + q + 1 = \delta_a + 1 \leq -1/2 - \kappa_1/2$, it follows from Lemma B.5 that $\sup_\theta T^{2\delta_a} \sum_{t=1}^{T} r_{2t}(\theta)^2 = O_P(T^{-2})$, and if $d_l - d_{0l} + q + 1 \geq -1/2 - \kappa_1/2$, Lemma B.3 shows that $\sup_\theta T^{2\delta_a} \sum_{t=1}^{T} r_{2t}(\theta)^2 = T^{2\delta_a+1} O_P(T^{\kappa_1}) = O_P(T^{-\kappa_1})$ by definition of $D_1(\kappa_1)$.

Next, using the Beveridge-Nelson decomposition (12) applied to $A(z,\psi_0)$,

$$\begin{aligned} r_{3t}(\theta) &= \beta_a(\theta)'(\sum_{n=0}^{t-q-1} B_{(q),n}^{(\cdot l)}(\psi))\Delta_+^{\delta_a}u_t^{(l)} - \gamma_a(\theta)'\Delta_+^{\delta_a}\varepsilon_t \\ &= \beta_a(\theta)'(\sum_{n=0}^{\infty} B_{(q),n}^{(\cdot l)}(\psi))\Delta_+^{\delta_a}u_t^{(l)} - \gamma_a(\theta)'\Delta_+^{\delta_a}\varepsilon_t - \beta_a(\theta)'(\sum_{n=t-q}^{\infty} B_{(q),n}^{(\cdot l)}(\psi))\Delta_+^{\delta_a}u_t^{(l)} \\ &= \beta_a(\theta)'(\sum_{n=0}^{\infty} B_{(q),n}^{(\cdot l)}(\psi))\sum_{m=0}^{\infty} A_{(1),m}^{(l\cdot)}(\psi_0)\Delta_+^{\delta_a+1}\varepsilon_{t-m} - \beta_a(\theta)'(\sum_{n=t-q}^{\infty} B_{(q),n}^{(\cdot l)}(\psi))\Delta_+^{\delta_a}u_t^{(l)} \\ &\quad + \beta_a(\theta)'(\sum_{n=0}^{\infty} B_{(q),n}^{(\cdot l)}(\psi))(\sum_{m=0}^{\infty} A_m^{(l\cdot)}(\psi_0))\Delta_+^{\delta_a}\varepsilon_t - \gamma_a(\theta)'\Delta_+^{\delta_a}\varepsilon_t, \end{aligned}$$

where $A_{(1),m}^{(l\cdot)}(\psi_0) = -\sum_{n=m+1}^{\infty} A_n^{(l\cdot)}(\psi_0)$. The first term is handled in the same way as $r_{2t}(\theta)$ and the second term in the same way as $r_{1t}(\theta)$ except with $h = q$. The last term is

$$\beta_a(\theta)'(\sum_{n=0}^{\infty}(B_{(q),n}^{(\cdot l)}(\psi) - \tilde{B}_{(q),n}^{(\cdot l)}(\psi)))(\sum_{m=0}^{\infty} A_m^{(l\cdot)}(\psi_0))\Delta_+^{\delta_a}\varepsilon_t,$$

where $\sum_{n=0}^{\infty}(B_{(q),n}^{(\cdot l)}(\psi) - \tilde{B}_{(q),n}^{(\cdot l)}(\psi)) = 1_{\{q\geq 1\}}\sum_{n=t}^{\infty}(\frac{1}{q!}\prod_{j=0}^{q-1}(n-j))B_n^{(\cdot l)}(\psi) \leq 1_{\{q\geq 1\}}ct^{-1-s+q-\xi}$ uniformly in $\psi \in \Psi$, see (4), such that this term can be handled in the same way as $r_{1t}(\theta)$. Note that $\gamma(\theta)' = \beta_a(\theta)'(\sum_{n=0}^{\infty}\tilde{B}_{(q),n}^{(\cdot l)}(\psi))(\sum_{m=0}^{\infty} A_m^{(l\cdot)}(\psi_0)) = \beta_a(\theta)'\tilde{B}_{(q)}^{(\cdot l)}(\psi)(\sum_{m=0}^{\infty} A_m^{(l\cdot)}(\psi_0)) = |\tilde{B}_{(q)}^{(\cdot l)}(\psi)|(\sum_{m=0}^{\infty} A_m^{(l\cdot)}(\psi_0))$.

Finally, applying (28) with $q = s$ to $r_{4t}(\theta)$ shows that $r_{4t}(\theta)$ contains terms with $\Delta_+^{d_k-d_{0k}+h}u_t^{(k)}, 0 \leq h \leq s$, and a term with $\sum_{k=1,k\neq l}^{p}\sum_{n=0}^{t-s-2} B_{(s+1),n}^{(\cdot k)}(\psi)\sum_{j=0}^{t-n-1}\pi_j(d_{0k}-d_k-s-1)u_{t-j-n}^{(k)}$. The last term is handled in the same way as $r_{2t}(\theta)$. For the other terms, if $d_k - d_{0k} + h \geq \delta_a + \kappa_1/2$, then, as in the proof for $r_{2t}(\theta)$, Lemma B.5 shows that $\sup_\theta T^{2\delta_a}\sum_{t=1}^{T} r_{4t}(\theta)^2 = O_P(T^{-\kappa_1})$ when $d_k - d_{0k} + h \leq -1/2 - \kappa_1/2$, and Lemma B.3 shows that $\sup_\theta T^{2\delta_a}\sum_{t=1}^{T} r_{4t}(\theta)^2 = T^{2\delta_a+1}O_P(T^{\kappa_1}) = O_P(T^{-\kappa_1})$ when $d_k - d_{0k} + h \geq -1/2 - \kappa_1/2$. If $d_k - d_{0k} + h \leq \delta_a + \kappa_1/2$, then by construction of $\beta_a(\theta)$ there will either be another β-vector obtained from those terms (certainly when $d_k - d_{0k} + h \leq \delta_a$ since those β-vectors are obtained prior to $\beta_a(\theta)$) or they will correspond to terms not chosen because $\beta_p(\theta)$ has already been determined. In either case, by orthonormality of $\beta(\theta)$, the factor $\sum_{n=t-h}^{\infty} B_{(h),n}^{(\cdot k)}(\psi)$ appears multiplicatively and the terms are handled in the same way as $r_{1t}(\psi)$. ∎

It follows from Lemma 2 that it is sufficient to consider (the determinant of) the product moment $R^*(\theta)$ with (i,j)'th element $T^{\delta_i+\delta_j}\sum_{t=1}^{T}(\Delta_+^{\delta_i}\gamma_i(\theta)'\varepsilon_t)(\Delta_+^{\delta_j}\gamma_j(\theta)'\varepsilon_t)$ since the determinant of $\beta(\theta)$ is one by orthonormality. Define the $T\times p$ matrices A_T and B_T, where the (t,i)'th element of A_T is $A_T^{(t,i)} = T^{\delta_i}\Delta_+^{\delta_i}\gamma_i(\theta)'\varepsilon_t$ and the j'th column of B_T has ones in the first $\lfloor T/(p-j+1)\rfloor$ elements and zeros otherwise, with $\lfloor x\rfloor$ denoting the largest integer not greater than x. Then $A_T'A_T = R^*(\theta)$, $\det\{B_T'B_T\} = c_pT^p$ for some constant $c_p \in (0,\infty)$, and the matrix $A_T'B_T$ has (i,j)'th element $T^{\delta_i}\sum_{t=1}^{\lfloor T/(p-j+1)\rfloor}\Delta_+^{\delta_i}\gamma_i(\theta)'\varepsilon_t = T^{\delta_i}\Delta_+^{\delta_i-1}\gamma_i(\theta)'\varepsilon_{\lfloor T/(p-j+1)\rfloor}$. From Magnus & Neudecker (1999, p. 201, eqn. (3)) the following generalization of the Cauchy-Schwarz inequality to matrix determinants is obtained: for any two real matrices A_T and B_T of the same dimensions it holds that

$$\det\{A_T'A_T\} \geq (\det\{B_T'B_T\})^{-1}(\det\{A_T'B_T\})^2 = c_p^{-1}(\det\{T^{-1/2}A_T'B_T\})^2, \qquad (29)$$

where the equality uses elementary properties of the determinant function. It follows that the determinant of the matrix $R^*(\theta)$ can be bounded below, uniformly in $\theta \in \Theta_1$, by c_p^{-1} times the square of the determinant of the matrix $Q_T(\theta) = T^{-1/2} A_T' B_T$ with (i,j)'th element $Q_T^{(i,j)}(\theta) = T^{\delta_i - 1/2} \Delta_+^{\delta_i - 1} \gamma_i(\theta)' \varepsilon_{\lfloor T/(p-j+1) \rfloor}$, whose limit is derived next.

Using the notation $\varepsilon_{it}^*(\theta) = \gamma_i(\theta)'\varepsilon_t$, $\varepsilon_t^*(\theta) = [\varepsilon_{1t}^*(\theta)', \ldots, \varepsilon_{pt}^*(\theta)']' = [\gamma_1(\theta), \ldots, \gamma_p(\theta)]'\varepsilon_t = G(\theta)'\varepsilon_t$, and $\delta = (\delta_1, \ldots, \delta_p)'$, the functional central limit theorem of Hosoya (2005, Theorem 2) applies to $\Lambda_+(\delta - 1)\varepsilon_t^*(\theta)$ upon verification of Hosoya's (2005) Assumptions A(i)–(iv) for ε_t satisfying Assumption A and $0 < |G(\theta)| < \infty$ uniformly in $\theta \in \Theta_1$. Because $E(\varepsilon_l^*(\theta)|\mathcal{F}_t) = 0$ for $l > t$ and $E(\varepsilon_l^*(\theta)\varepsilon_m^*(\theta)'|\mathcal{F}_t) - E(\varepsilon_l^*(\theta)\varepsilon_m^*(\theta)') = 0$ for $\min(l,m) > t$, Hosoya's (2005) Assumptions A(i) and A(ii) are trivially satisfied. Furthermore, Assumption A implies that the fourth-order cumulant spectral density function of $\varepsilon_{it}^*(\theta)$ is bounded such that by Lemma 1 and Theorem 3, respectively, of Hosoya (2005), his Assumptions A(iii) and A(iv) are satisfied. In particular, the moment condition in Assumption A implies that Hosoya's (2005) Assumption A(iv) is satisfied with $\gamma_5 = 2$ such that the fractional order, $-\delta_i + 1$, must satisfy $2 > (2(-\delta_i + 1) - 1)^{-1}$ or $\delta_i < 1/4$ which is clearly satisfied for all $d \in D_1(\kappa_1)^p$ for any $\kappa_1 > 0$ since the latter implies $\delta_i \leq -1/2 - \kappa_1$. Introduce the normalization matrix $S_T(\delta) = \mathrm{diag}(T^{\delta_1 - 1/2}, \ldots, T^{\delta_p - 1/2})$. Then, as $T \to \infty$, Theorem 2 of Hosoya (2005) implies that

$$S_T(\delta)\Delta_+(\delta - 1)\varepsilon_{\lfloor Tr \rfloor}^*(\theta) \Rightarrow W_{-\delta}(r) \text{ in } \mathcal{D}^p[0,1] \tag{30}$$

for fixed $d \in D_1(\kappa_1)^p$, where $W_{-\delta}(r)$ is the p-vector fractional Brownian motion of type II with $W_{-\delta}^{(i)}(r) = \Gamma(-\delta_i + 1)^{-1} \int_0^r (r-s)^{-\delta_i} \mathsf{d}W^{(i)}(s)$ and W denotes p-vector Brownian motion generated by $\varepsilon_t^*(\psi)$, i.e., with variance matrix $E(W(s)W(s)') = sG(\theta)'\Sigma_0 G(\theta)$. It then clearly holds that, as $T \to \infty$,

$$Q_T(\theta) \overset{D}{\to} Q(\theta) \text{ with } Q^{(i,j)}(\theta) = W_{-\delta}^{(i)}((p-j+1)^{-1}) \tag{31}$$

for fixed $\theta \in \Theta_1(\kappa_1) = D_1(\kappa_1)^p \times \Psi$, which shows the pointwise limit.

The random matrix $Q(\theta)$ is non-singular almost surely, uniformly in $\theta \in \Theta_1$. To see this, consider a linear combination of the columns of $Q(\theta)$,

$$\sum_{j=1}^p \mu_j Q^{(\cdot j)}(\theta) = \sum_{j=1}^p \mu_j [W_{-\delta}^{(1)}((p-j+1)^{-1}), \ldots, W_{-\delta}^{(p)}((p-j+1)^{-1})]',$$

and a linear combination of the rows of $Q(\theta)$,

$$\sum_{i=1}^p \mu_i Q^{(i \cdot)}(\theta) = \sum_{i=1}^p \mu_i [W_{-\delta}^{(i)}(1/p), \ldots, W_{-\delta}^{(i)}(1)].$$

Both linear combinations are non-zero almost surely, uniformly in $\theta \in \Theta_1$, for all $\mu \neq 0$, which implies that $Q(\theta)$ is non-singular almost surely, uniformly in $\theta \in \Theta_1$. This holds even though $\gamma_a(\theta)$ and $\gamma_b(\theta)$ may be proportional for some a and b in which case $G(\theta)$ has reduced rank. The reason is that, if $\gamma_a(\theta)$ and $\gamma_b(\theta)$ are proportional, their associated fractional parameters δ_a and δ_b are different such that $W_{-\delta}^{(a)}(r)$ and $W_{-\delta}^{(b)}(r)$ are linearly independent (noting that $W_{-\delta}^{(a)}(r)$ only depends on δ_a, etc.), and of course $\gamma_a(\theta)'\Sigma_0\gamma_a(\theta) > 0$ uniformly in θ for all $a = 1, \ldots, p$.

To strengthen the pointwise convergence in (31) to weak convergence in $\mathcal{C}^{p \times p}(\Theta_1)$, it is sufficient to show that $Q_T(\theta)$ is tight as a function of θ on $\theta \in \Theta_1$. As in Section 3.1, it is sufficient to prove tightness of $Q_T^{(i,j)}(\theta)$ for each $i, j = 1, \ldots, p$ by Billingsley (1968, Problem 6.6). Note that $Q_T^{(i,j)}(\theta) = \gamma_i(\theta)'\tilde{Q}_T(\delta_i)$, with $\tilde{Q}_T(\delta_i) = T^{\delta_i - 1/2}\Delta_+^{\delta_i - 1}\varepsilon_{\lfloor T/(p-j+1) \rfloor}$, is continuously differentiable in θ and the parameter ψ appears only through the coefficient $\gamma_i(\theta)$, which implies that $Q_T^{(i,j)}(\theta)$ is tight in θ if the p-vector $\tilde{Q}_T(\delta_i)$ is tight in δ_i; see Lemma A.2 of Johansen & Nielsen (2010). Tightness of the k'th element of $\tilde{Q}_T(\delta_i)$, i.e.,

$\tilde{Q}_T^{(k)}(d_i) = T^{\delta_i-1/2}\Delta_+^{\delta_i-1}\varepsilon_{k,\lfloor T/(p-j+1)\rfloor}$, is shown using the moment condition in Billingsley (1968, Theorem 12.3), which requires showing that $\tilde{Q}_T^{(k)}(\delta_i)$ is tight for fixed δ_i and that

$$||\tilde{Q}_T^{(k)}(z_1) - \tilde{Q}_T^{(k)}(z_2)||_2 \leq c|z_1 - z_2| \tag{32}$$

for some constant $c > 0$ that does not depend on T, z_1, or z_2. Tightness for any fixed δ_i is implied by the pointwise convergence in (30) and condition (32) is satisfied by (44) of Lemma B.1. Hence the convergence in (31) is strengthened to

$$Q_T(\theta) \Rightarrow Q(\theta) \text{ in } \mathcal{C}^{p\times p}(\Theta_1).$$

By the continuous mapping theorem applied to the $\inf_{\theta\in\Theta_1}(\det\{\cdot\})^2$ mapping, which is continuous because Θ_1 is compact, it then holds that

$$\inf_{\theta\in\Theta_1}(\det\{Q_T(\theta)\})^2 \xrightarrow{D} \inf_{\theta\in\Theta_1}(\det\{Q(\theta)\})^2. \tag{33}$$

It follows that

$$\inf_{\theta\in\Theta_1}\det\{R(\theta)\} \geq c_p^{-1}T^{2\kappa_1}\inf_{\theta\in\Theta_1}(\det\{Q_T(\theta)\})^2 + o_P(1)$$

and, because $2\kappa_1 > 0$ and $Q(\theta)$ is non-singular almost surely, it therefore holds that for any $K > 0$,

$$P(\inf_{\theta\in\Theta_1}\det\{R(\theta)\} > K) \to 1 \text{ as } T \to \infty. \tag{34}$$

Finally, note that (34) holds for any choice of $\kappa_1 > 0$ via the use of the generalized Cauchy-Schwarz inequality (29). Because the moment condition implied by Assumption A(iv) in Hosoya (2005) is in fact necessary, at least for general fractional processes, see Johansen & Nielsen (2012*b*), the application of the generalized Cauchy-Schwarz inequality appears to be necessary, as well, to avoid a stronger moment condition in Assumption A.

3.3.3 Notation for the proof of (13) in the general case

For the proof of the general case, first define the direction vectors $\beta_a(\theta)$ that generate all the processes with fractional differencing order $\delta_a \leq -1/2 + \kappa_3$ by the procedure in Section 3.3.2 above. Since these δ_a need to be classified into different subsets similar to D_1, D_2, and D_3, define $D_1^* = D_1^*(\kappa_1) = \{z \in \mathbb{R} : z \leq -1/2 - \kappa_1\}$, $D_2^* = D_2^*(\kappa_2, \kappa_3) = \{z \in \mathbb{R} : -1/2 - \kappa_2 \leq z \leq -1/2 + \kappa_3\}$, and $D_3^* = D_3^*(\kappa_3) = \{z \in \mathbb{R} : z \geq -1/2 + \kappa_3\}$. Then define the index sets (corresponding to D_1^* and D_2^*) $I = \{i : \delta_i \in D_1^*\}$ and $J = \{j : \delta_j \in D_2^*\}$ with number of elements p_I and p_J, respectively, and let $\beta_I(\theta) = [\beta_1(\theta), \ldots, \beta_{p_I}(\theta)]$ and $\beta_J(\theta) = [\beta_{p_I+1}(\theta), \ldots, \beta_{p_I+p_J}(\theta)]$ denote the β-vectors that generate processes whose fractional differencing parameters are in D_1^* and D_2^*, respectively. Note that if $p_I = p$ then J is empty and $p_J = 0$. If $p_I + p_J < p$ define also the index set (corresponding to D_3^*) $K = \{k : \delta_k \in D_3^*\}$ with $p_K = p - p_I - p_J$ elements, such that $p_I + p_J + p_K = p$, and define the $p \times p_K$ matrix $\beta_K(\theta)$ to be a basis for the null space of $[\beta_J(\theta), \beta_I(\theta)]$. These matrices depend on θ, but this dependence is very simple and suppressed in the following. Moreover, because the vectors $\tilde{B}_{(0)}^{(\cdot k)}(\psi) = \sum_{n=0}^{\infty} B_n^{(\cdot k)}(\psi)$ are orthogonal, the $p \times p$ matrix $\beta = [\beta_I, \beta_J, \beta_K]$ is orthonormal such that, in particular, its determinant is one. The fractional differencing order of $\beta_a(\theta)'\varepsilon_t(\theta)$ is δ_a, and conforming with the notation β_I, β_J, and β_K, these are stacked in $\delta_I = (\delta_1, \ldots, \delta_{p_I})'$, $\delta_J = (\delta_{p_I+1}, \ldots, \delta_{p_I+p_J})'$, $\delta_K = (\delta_{p_I+p_J+1}, \ldots, \delta_p)'$, and $\delta = (\delta_I', \delta_J', \delta_K')'$.

The residual product moment $R(\theta)$ will be analyzed in the directions given by β, and it is convenient to define the notation $R_{IK}(\theta) = \beta_I' R(\theta)\beta_K$, $R_{II}(\theta) = \beta_I' R(\theta)\beta_I$, etc., and let asterisks denote that nonstationary processes have been normalized, e.g., $R_{II}^{**}(\theta)$ has (i,j)'th element $T^{\delta_i+\delta_j+1}\beta_i(\theta)'R(\theta)\beta_j(\theta) = T^{\delta_i+\delta_j}\sum_{t=1}^{T}(\gamma_i(\theta)'\Delta_+^{\delta_i}\varepsilon_t)(\gamma_j(\theta)'\Delta_+^{\delta_j}\varepsilon_t) + o_P(1)$, where the $o_P(1)$ term is uniform in θ by Lemma 2 and will be suppressed in the following. Likewise, the processes $\beta_J'\varepsilon_t(\theta)$ and $\beta_K'\varepsilon_t(\theta)$ include $o_P(1)$ contributions as in Lemma 1, but also from terms like $r_{1t}(\theta)$ in the proof of Lemma 2, and these are suppressed in the following as well.

3.3.4 Proof of (13) when all processes are in I or all processes are in J

For the proof of (13) either I or J (or both) must be non-empty. The proof for the case where all processes are in I was given in Section 3.3.2 and placed no restrictions on κ_1. The proof for the case where all processes are in J was given in Section 3.3.1, where it was shown that there exists $\bar{\kappa}_3 > 0$ such that $\kappa_3 \leq \bar{\kappa}_3$ would obtain (25), so such a value is used subsequently.

3.3.5 Proof of (13) when J is empty

Suppose $\theta \in \Theta_I$, which is such that J is empty but I and K are non-empty. In this case,

$$\det\{R(\theta)\} = \det\{R_{II}(\theta)\} \det\{R_{KK|I}(\theta)\},$$

where $R_{KK|I}(\theta) = R_{KK}(\theta) - R^*_{KI}(\theta)R^{**}_{II}(\theta)^{-1}R^*_{IK}(\theta)$ denotes the conditional residual product moment. The analysis in Section 3.3.2 shows that $P(\inf_{\theta\in\Theta_I} \det\{R_{II}(\theta)\} > K) \to 1$ as $T \to \infty$, see (34), for any choice of $\kappa_1 > 0$.

The analysis in Section 3.1 shows that $R_{KK}(\theta)$ converges uniformly in probability to a positive definite matrix for any choice of $\kappa_3 > 0$. Next, $R^*_{KI}(\theta)$ converges to zero uniformly in probability by Lemma B.4 with $u_1 = \delta_k, u_2 = \delta_i$ such that $a = \kappa_3 > 0$ and $b = \kappa_1 > 0$. Finally, for any matrix A recall that $A^{-1} = \text{adj}\{A\}/\det\{A\}$, where $\text{adj}\{A\}$ denotes the adjoint matrix of A. In the case of $R^{**}_{II}(\theta)^{-1}$, the generalized Cauchy-Schwarz inquality (29) shows that $\det\{R^{**}_{II}(\theta)\} \geq c_p^{-1}(\det\{Q_T(\theta)\})^2$ and using (33) it holds that $\sup_{\theta\in\Theta_I}(\det\{R^{**}_{II}(\theta)\})^{-1} \leq c_p/\inf_{\theta\in\Theta_I}(\det\{Q_T(\theta)\})^2 = O_P(1)$. Each element of the adjoint matrix $\text{adj}\{R^{**}_{II}(\theta)\}$ is a simple function (addition and multiplication) of elements of $R^{**}_{II}(\theta)$, and therefore $\sup_{\theta\in\Theta_I} \text{adj}\{R^{**}_{II}(\theta)\} = O_P(1)$ because $\sup_{\theta\in\Theta_I} R^{**}_{II}(\theta) = O_P(1)$ by Lemma B.5. Thus, $\sup_{\theta\in\Theta_I} R^{**}_{II}(\theta)^{-1} = O_P(1)$ and $R^*_{KI}(\theta)R^{**}_{II}(\theta)^{-1}R^*_{IK}(\theta)$ converges to zero uniformly in probability such that $\det\{R_{KK|I}(\theta)\}$ has the same limit as $\det\{R_{KK}(\theta)\}$.

Hence, (13) follows for any $\kappa_1 > 0$ and any $\kappa_3 > 0$ when J is empty.

3.3.6 Proof of (13) when I is empty

Now suppose $\theta \in \Theta_J$, which is such that I is empty but J and K are non-empty. In this case,

$$\det\{R(\theta)\} = \det\{R_{KK}(\theta)\} \det\{R_{JJ|K}(\theta)\},$$

where the analysis in Section 3.1 shows that $R_{KK}(\theta)$ converges uniformly in probability to a positive definite matrix for any choice of $\kappa_3 > 0$. This also implies that $\sup_{\theta\in\Theta_J} R_{KK}(\theta)^{-1} = O_P(1)$. Since $\sup_{\theta\in\Theta_J} R_{KJ}(\theta) = O_P(1)$ by Lemma B.3 with $u_1 = \delta_k, u_2 = \delta_j$ such that $a = \min(1/2+\kappa_3, 1/2-\kappa_2, \kappa_3 - \kappa_2) = \kappa_3 - \kappa_2 > 0$ when choosing $\kappa_2 < \kappa_3$, it holds that $R_{JJ|K}(\theta) = R_{JJ}(\theta) - R_{JK}(\theta)R_{KK}(\theta)^{-1}R_{JK}(\theta) = R_{JJ}(\theta) + O_P(1)$, where the $O_P(1)$ term is uniform in $\theta \in \Theta_J$. Section 3.3.1 shows that there exists $\bar{\kappa}_3 > 0$ and $T_0 \geq 1$ such that $P(\inf_{\theta\in\Theta_J} \det\{R_{JJ}(\theta)\} > K) \geq 1 - \eta$ for all $T \geq T_0$ and all $\kappa_3 \leq \bar{\kappa}_3$, and the same result thus holds for $\inf_{\theta\in\Theta_J} \det\{R_{JJ|K}(\theta)\}$ by choosing $\bar{\kappa}_2 < \bar{\kappa}_3$ to define D^*_2 and D^*_3 (and hence J and K). Again, (13) follows for any $\kappa_1 > 0$.

3.3.7 Overlap-wedge argument

Because no restrictions have been placed on κ_1 so far, it can be chosen in the previous subsections as $0 < \kappa_1 < \bar{\kappa}_2$, such that there is an overlap between the intervals $D^*_1(\kappa_1)$ and $D^*_2(\bar{\kappa}_2, \bar{\kappa}_3)$. This implies that, for the case analyzed in the next subsection, where both I and J (and possibly K) are non-empty, the analysis can be based on $D^*_1(\bar{\kappa}_1)$ and $D^*_2(\bar{\kappa}_2, \bar{\kappa}_3)$ with $0 < \bar{\kappa}_2 < \min(\bar{\kappa}_1, \bar{\kappa}_3)$ such that $D^*_1(\bar{\kappa}_1)$ and $D^*_2(\bar{\kappa}_2, \bar{\kappa}_3)$ are disjoint and there is a wedge between them.

To see why this is the case, suppose first that there are two processes in $D^*_1 \cup D^*_2$ (and possibly some in D^*_3, but these are irrelevant for this argument). If $\delta_i \in D^*_2(\kappa^*_2, \kappa_3)$ for $i = 1, 2$ and some $\kappa^*_2 > 0$ then I is empty, while if $\delta_i \in D^*_1(\kappa^*_1)$ for $i = 1, 2$ and some $\kappa^*_1 < \kappa^*_2$ (the overlap) then J is empty, and in both cases the above arguments (in Sections 3.3.6 and 3.3.5, respectively) apply. The remaining case is that

$\delta_i \in (D_1^* \cup D_2^*) \backslash D_1^*(\kappa_1^*)$ and $\delta_j \in (D_1^* \cup D_2^*) \backslash D_2^*(\kappa_2^*, \kappa_3)$ for $i \neq j$, in which case I and J are defined using $\bar{\kappa}_1 = \kappa_2^*$ and $\bar{\kappa}_2 = \kappa_1^*$, respectively, such that $\bar{\kappa}_2 < \bar{\kappa}_1$ (the wedge).

Suppose next that there are three processes in $D_1^* \cup D_2^*$ (and possibly some in D_3^*). If all three are in $D_1^*(\kappa_1^*)$ or all are in $D_2^*(\kappa_2^*, \kappa_3)$, the arguments above (Sections 3.3.5 and 3.3.6) again apply. Thus, suppose at least one $\delta_i \in (D_1^* \cup D_2^*) \backslash D_1^*(\kappa_1^*)$ and at least one $\delta_j \in (D_1^* \cup D_2^*) \backslash D_2^*(\kappa_2^*, \kappa_3)$ for $i \neq j$, while $\delta_k \in D_1^* \cup D_2^*$ for $k \neq i, j$. Then there are two subcases: (a) If $\delta_k \leq -1/2 - \kappa_1^*/2 - \kappa_2^*/2$ then the analysis can be based on I and J defined using $\bar{\kappa}_1 = \kappa_1^*/2 + \kappa_2^*/2 > \kappa_1^*$ and $\bar{\kappa}_2 = \kappa_1^*$, respectively. (b) If $\delta_k \geq -1/2 - \kappa_1^*/2 - \kappa_2^*/2$ then the analysis can be based on I and J defined using $\bar{\kappa}_1 = \kappa_2^*$ and $\bar{\kappa}_2 = \kappa_1^*/2 + \kappa_2^*/2 < \kappa_2^*$, respectively. In either subcase, $\bar{\kappa}_1 > \bar{\kappa}_2$ (the wedge). The argument is straightforwardly generalized to cases with more than three processes in $D_1^* \cup D_2^*$.

3.3.8 Proof of (13) for the general case

Consider finally the general case, where $\bar{\Theta}$ is defined using $D_1^*(\bar{\kappa}_1)$, $D_2^*(\bar{\kappa}_2, \bar{\kappa}_3)$, and $D_3^*(\bar{\kappa}_3)$ and is such that I, J, and K are all non-empty with $0 < \bar{\kappa}_2 < \min(\bar{\kappa}_1, \bar{\kappa}_3)$. If one or more of I, J, or K are empty, the proof simplies easily. Now

$$\det\{R(\theta)\} = \det\{R_{II}(\theta)\} \det\{R_{KK|I}(\theta)\} \det\{R_{JJ|I,K}(\theta)\},$$

where the first two terms have already been analyzed in Section 3.3.5 above (the case where J is empty), and it is shown there that, for any $\kappa_1 > 0$ and any $\kappa_3 > 0$, it holds that $P(\inf_{\theta \in \bar{\Theta}} \det\{R_{II}(\theta)\} > K) \to 1$ as $T \to \infty$ and $R_{KK|I}(\theta)$ converges uniformly in probability to a positive definite matrix.

It remains to be shown that there exists a $\bar{\kappa}_3 > 0$ and a $T_0 \geq 1$ such that $P(\inf_{\theta \in \bar{\Theta}} \det\{R_{JJ|I,K}(\theta)\} > K) \geq 1 - \eta$ for all $T \geq T_0$. The analysis of $R_{JJ|I,K}(\theta)$ is similar to that in Sections 3.3.1 and 3.3.6, except for the conditioning on both stationary and nonstationary variables. First eliminate the stationary variables,

$$R_{JJ|I,K}(\theta) = R_{JJ|I}(\theta) - R_{JK|I}(\theta) R_{KK|I}(\theta)^{-1} R_{KJ|I}(\theta). \tag{35}$$

To show that there exists a $\bar{\kappa}_3 > 0$ and a $T_0 \geq 1$ such that $P(\inf_{\theta \in \bar{\Theta}} \det\{R_{JJ|I}(\theta)\} > K) \geq 1 - \eta$ for all $T \geq T_0$, it needs to be further decomposed, and it is convenient to introduce the notation Z_{It} and Z_{Jt} for the variables in I and J, respectively, the more compact notation $P_{T,N}(Z_{1t}, Z_{2t}) = T^{-1} \sum_{t=N+1}^{T} Z_{1t} Z_{2t}'$ and $P_{T,N}(Z_{1t}, Z_{2t}|Z_{3t}) = P_{T,N}(Z_{1t}, Z_{2t}) - P_{T,N}(Z_{1t}, Z_{3t}) P_{T,N}(Z_{3t}, Z_{3t})^{-1} P_{T,N}(Z_{3t}, Z_{2t})$ for product moments, and again asterisks denote that nonstationary processes have been normalized. Then, as in Section 3.3.1, bound $R_{JJ|I}(\theta)$ by

$$\begin{aligned} R_{JJ|I}(\theta) &\geq P_{T,N}(Z_{Jt}, Z_{Jt}|Z_{It}) \\ &\geq P_{T,N}(w_t, w_t|Z_{It}) + P_{T,N}(w_t, v_t|Z_{It}) + P_{T,N}(v_t, w_t|Z_{It}), \end{aligned}$$

where

$$P_{T,N}(w_t, v_t|Z_{It}) = P_{T,N}(w_t, v_t) - P_{T,N}^*(w_t, Z_{It}) P_{T,N}^{**}(Z_{It}, Z_{It})^{-1} P_{T,N}(Z_{It}, v_t), \tag{36}$$

$$P_{T,N}(w_t, w_t|Z_{It}) - P_{T,N}(w_t, w_t) = P_{T,N}^*(w_t, Z_{It}) P_{T,N}^{**}(Z_{It}, Z_{It})^{-1} P_{T,N}^*(Z_{It}, w_t). \tag{37}$$

As in Section 3.3.1, the desired result holds for $P_{T,N}(w_t, w_t)$, which determines $\bar{\kappa}_3 > 0$, and hence $\bar{\kappa}_1, \bar{\kappa}_2$ are chosen according to $0 < \bar{\kappa}_2 < \min(\bar{\kappa}_3, \bar{\kappa}_1) < 1/2$ such that there is a wedge between the intervals D_1^* and D_2^* as argued in Section 3.3.7. In the analysis of $R_{JJ|I}(\theta)$, it thus only remains to be shown that the right-hand sides of (36) and (37) are both negligible uniformly in $\theta \in \bar{\Theta}$.

First the result is shown for (36). From Lemma B.2 it holds that $\sup_{\theta \in \bar{\Theta}} P_{T,N}(w_t, v_t) = o_P(1)$ when $N = T^\alpha$ with $\alpha < (1/2 - \bar{\kappa}_2)/(1/2 + \bar{\kappa}_2)$. Lemma B.4 shows that $\sup_{\theta \in \bar{\Theta}} P_{T,N}^*(w_t, Z_{It}) = O_P((\log T) T^{-\bar{\kappa}_1} N^{\bar{\kappa}_1 + \bar{\kappa}_2})$ and $\sup_{\theta \in \bar{\Theta}} P_{T,N}^*(Z_{It}, v_t) = O_P((\log T) T^{\bar{\kappa}_2})$. Since $\sup_{\theta \in \bar{\Theta}} P_{T,N}^{**}(Z_{It}, Z_{It})^{-1} = O_P(1)$ as in the case where J is empty in Section 3.3.5, it thus holds that $\sup_{\theta \in \bar{\Theta}} |(36)|$ when $N = T^\alpha$

with $\alpha < (\bar{\kappa}_1 - \bar{\kappa}_2)/(\bar{\kappa}_1 + \bar{\kappa}_2)$, in addition to the previous constraints on α, recalling that $\bar{\kappa}_2 < \bar{\kappa}_1$. Thus, the wedge is needed here to ensure that $P^*_{T,N}(w_t, Z_{It})$ converges to zero faster than $P^*_{T,N}(v_t, Z_{It})$ diverges.

Next, for (37), Lemma B.4 shows that $\sup_{\theta\in\bar{\Theta}} P^*_{T,N}(w_t, Z_{It}) = O_P((\log T)T^{-\bar{\kappa}_1}N^{\bar{\kappa}_1+\bar{\kappa}_2})$ and, as before, $\sup_{\theta\in\bar{\Theta}} P^{**}_{T,N}(Z_{It}, Z_{It})^{-1} = O_P(1)$, such that $\sup_{\theta\in\bar{\Theta}} |(37)| = o_P(1)$ when $N = T^\alpha$ with $\alpha < \bar{\kappa}_1/(\bar{\kappa}_1 + \bar{\kappa}_2)$, in addition to the previous constraints on α.

Finally, it only remains to be shown that the second term on the right-hand side of (35) is $O_P(1)$ uniformly in $\theta \in \bar{\Theta}$. To see this, note first that $R_{KK|I}(\theta)$, and hence $R_{KK|I}(\theta)^{-1}$, converges uniformly in probability to a positive definite matrix for any choice of $\kappa_3 > 0$ as in the case where J is empty in Section 3.3.5. The term $R_{JK|I}(\theta) = R_{JK}(\theta) - R^*_{JI}(\theta)R^{**}_{II}(\theta)^{-1}R^*_{KI}(\theta)$, where $\sup_{\theta\in\bar{\Theta}} R_{JK}(\theta) = O_P(1)$ by Lemma B.3 with $u_1 + 1/2 \geq -\bar{\kappa}_2, u_2 + 1/2 \geq \bar{\kappa}_3$, and hence $a = \bar{\kappa}_3 - \bar{\kappa}_2 > 0$. For the term $R^*_{JI}(\theta)R^{**}_{II}(\theta)^{-1}R^*_{KI}(\theta)$ it holds that $\sup_{\theta\in\bar{\Theta}} R^{**}_{II}(\theta)^{-1} = O_P(1)$ as in Section 3.3.5. In addition, $\sup_{\theta\in\bar{\Theta}} R^*_{JI}(\theta) = O_P((\log T)T^{\bar{\kappa}_2})$ by Lemma B.4 with $a = -\bar{\kappa}_2$ and $b = \bar{\kappa}_1$ while $\sup_{\theta\in\bar{\Theta}} R^*_{KI}(\theta) = O_P((\log T)T^{-\min(\bar{\kappa}_1,\bar{\kappa}_3)})$ by Lemma B.4 with $a = \bar{\kappa}_3$ and $b = \bar{\kappa}_1$, such that $\sup_{\theta\in\bar{\Theta}} R^*_{JI}(\theta)R^{**}_{II}(\theta)^{-1}R^*_{KI}(\theta) = O_P((\log T)^2T^{\bar{\kappa}_2-\min(\bar{\kappa}_1,\bar{\kappa}_3)}) = o_P(1)$ because $\bar{\kappa}_2 < \min(\bar{\kappa}_1, \bar{\kappa}_3)$. Here, the wedge is needed once more to ensure that $R^*_{KI}(\theta)$ converges to zero faster than $R^*_{JI}(\theta)$ diverges.

This completes the proof of Theorem 1.

4 Asymptotic distribution theory

To prove asymptotic normality for the CSS estimator for model (3), the smoothness conditions on the linear coefficients need to be strengthened and an additional condition is needed to ensure that the asymptotic variance matrix of the estimator is well-defined.

Assumption E *For all z in the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$, $A(z,\psi) = \sum_{n=0}^\infty A_n(\psi)z^n$ is three times differentiable in ψ on the closed neighborhood $\mathcal{N}_\delta(\psi_0) = \{\psi \in \Psi : |\psi - \psi_0| \leq \delta\}$ for some $\delta > 0$, and the derivatives $\frac{\partial^k A_n^{(i,j)}(\psi)}{\partial\psi^{(k)}}$ satisfy $\sum_{n=0}^\infty |\frac{\partial^k A_n^{(i,j)}(\psi)}{\partial\psi^{(k)}}| < \infty$ for all $\psi \in \mathcal{N}_\delta(\psi_0)$ and $k = 2, 3$.*

Assumption F *The symmetric $(p + q) \times (p + q)$ matrix Ω_0 with (i, j)'th element*

$$\begin{aligned}
&\sum_{n,m=1}^\infty n^{-1}m^{-1}\operatorname{tr}\{\Sigma_0^{-1}\sum_{k=\max(0,m-n)}^\infty C_{ik}(\psi_0)\Sigma_0C_{j,k+n-m}(\psi_0)'\} && \text{if } i, j = 1, \ldots, p,\\
&-\sum_{n=1}^\infty n^{-1}\operatorname{tr}\{\Sigma_0^{-1}\sum_{k=0}^\infty C_{ik}(\psi_0)\Sigma_0\Gamma_{j,n+k}(\psi_0)'\} && \text{if } i = 1, \ldots, p \text{ and } j = p+1, \ldots, p+q,\\
&\sum_{n=1}^\infty \operatorname{tr}\{\Sigma_0^{-1}\Gamma_{in}(\psi_0)\Sigma_0\Gamma_{jn}(\psi_0)'\} && \text{if } i, j = p+1, \ldots, p+q,
\end{aligned}$$

is non-singular, where $C_{ik}(\psi)$ is defined in (10) and $\Gamma_{in}(\psi) = \sum_{m=0}^{n-1} A_m(\psi)\frac{\partial B_{n-m}(\psi)}{\partial\psi_i}$.

In the case with martingale difference errors, i.e. with $u_t = \varepsilon_t$, Assumption F reduces to $\Omega_0 = \frac{\pi^2}{6}\Sigma_0 \odot \Sigma_0^{-1}$ being non-singular, where $\odot$ denotes the elementwise (Hadamard) product. Thus, at least in this special case, Assumption F follows immediately from Assumption A and in particular from the assumption that $\Sigma_0 > 0$. Also in this special case the equations for Ω_0 given in Assumption F generalizes the well-known $\pi^2/6$ from the univariate case in a straightforward way.

Instead of Assumption E, Hualde & Robinson (2011) assume that $A(z, \psi)$ is twice continuously differentiable in ψ, which is slightly weaker. Assumption F is identical to Hualde & Robinson's (2011) Assumption A4(v) although it is stated differently. As above, both Assumptions E and F are easily satisfied by, e.g., the Bloomfield model or stationary and invertible ARMA processes.

Theorem 2 *Under the assumptions of Theorem 1, Assumptions E and F, and* $(d_0, \psi_0) \in \text{int}(D^p \times \Psi)$,

$$\sqrt{T}((\hat{d}', \hat{\psi}')' - (d_0', \psi_0')') \overset{D}{\to} N(0, \Omega_0^{-1}) \text{ as } T \to \infty.$$

Theorem 2 proves the anticipated result that the CSS estimator has the same asymptotic distribution as the (unconditional) Gaussian maximum likelihood estimator. For similar asymptotic distribution results for the CSS estimator, see also Beran (1995), Tanaka (1999), and Nielsen (2004) in the univariate case, and Hualde & Robinson (2011) for the multivariate model. An important consequence of these results is that the CSS estimator is efficient under Gaussianity, c.f. Fox & Taqqu (1986) and Dahlhaus (1989) for the univariate case. However, the asymptotic normality result in Theorem 2 is valid much more generally because Gaussianity is not assumed. Although Theorem 2 is proved in Hualde & Robinson (2011), the assumptions are stated in a different way here and allow for a rather brief proof, which is given next.

5 Proof of Theorem 2

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

$$0 = T^{1/2} \frac{\partial \det\{R(\hat{\theta})\}}{\partial \theta'} = T^{1/2} \frac{\partial \det\{R(\theta_0)\}}{\partial \theta'} + T^{1/2} \frac{\partial^2 \det\{R(\bar{\theta})\}}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0), \tag{38}$$

where $\bar{\theta}$ is an intermediate value satisfying $|\bar{\theta}_i - \theta_{0i}| \leq |\hat{\theta}_i - \theta_{0i}|, i = 1, \ldots, p+q$.

The normalized score function evaluated at the true value is

$$T^{1/2} \frac{\partial \det\{R(\theta_0)\}}{\partial \theta_i} = \det\{R(\theta_0)\} \operatorname{tr}\{R(\theta_0)^{-1} 2T^{-1/2} \sum_{t=1}^{T} y_{i,t-1} \varepsilon_t(\theta_0)'\} \text{ with } y_{i,t-1} = \frac{\partial}{\partial \theta_i} \varepsilon_t(\theta)|_{\theta=\theta_0}.$$

Define also the vector S_T by

$$S_T^{(i)} = 2T^{-1/2} \sum_{t=1}^{T} \varepsilon_t' \Sigma_0^{-1} z_{i,t-1} \text{ with } z_{i,t-1} = \frac{\partial}{\partial \theta_i} \eta_t(\theta)|_{\theta=\theta_0} = \frac{\partial}{\partial \theta_i} B(L,\psi) \Lambda(d - d_0) u_t|_{\theta=\theta_0}.$$

It is shown in Robinson (2006, pp. 135–136) that $2T^{-1/2} \sum_{t=1}^{T} \varepsilon_t(\theta_0)' R(\theta_0)^{-1} y_{i,t-1} - S_T^{(i)} = o_P(1)$ under the assumptions of Theorem 2. The (untruncated) stationary and ergodic process $z_{i,t-1} = \frac{\partial}{\partial \theta_i} B(L,\psi)\Lambda(d - d_0)u_t|_{\theta=\theta_0}$ is given by $-\sum_{n=1}^{\infty} n^{-1} \sum_{k=0}^{\infty} B_k^{(\cdot i)}(\psi_0) u_{t-n-k}^{(i)} = -\sum_{n=1}^{\infty} n^{-1} \sum_{k=0}^{\infty} C_{ik}(\psi_0) \varepsilon_{t-n-k}$ when $i = 1, \ldots, p$ and $\sum_{n=1}^{\infty} \frac{\partial B_n(\psi_0)}{\partial \psi_i} u_{t-n} = \sum_{n=1}^{\infty} \Gamma_{in}(\psi_0) \varepsilon_{t-n}$ when $i = p+1, \ldots, p+q$, where $C_{ik}(\psi)$ is defined in (10).

As usual, the Cramér-Wold device is used to obtain a central limit theorem for the score function, so, for any $(p+q)$-vector μ, it needs to be shown that $\mu' S_T = \sum_{i=1}^{p+q} \mu_i S_T^{(i)} \overset{D}{\to} N(0, 4\mu'\Omega_0 \mu)$. Because $z_{i,t-1}$ is measurable with respect to the sigma-algebra $\mathcal{F}_{t-1} = \sigma(\{\varepsilon_s, s \leq t-1\})$, $\nu_t = \sum_{i=1}^{p+q} \nu_{it} = \mu_i \varepsilon_t' \Sigma_0^{-1} z_{i,t-1}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_t$. By the law of large numbers for stationary and ergodic processes the sum of conditional variances is

$$\begin{aligned} T^{-1} \sum_{t=1}^{T} E(\nu_t^2 | \mathcal{F}_{t-1}) &= T^{-1} \sum_{t=1}^{T} \sum_{i,j=1}^{p+q} \mu_i \mu_j z_{i,t-1}' \Sigma_0^{-1} z_{j,t-1} \\ &= \sum_{i,j=1}^{p+q} \mu_i \mu_j \operatorname{tr}\{\Sigma_0^{-1} T^{-1} \sum_{t=1}^{T} z_{i,t-1} z_{j,t-1}'\} \\ &\overset{P}{\to} \sum_{i,j=1}^{p+q} \mu_i \mu_j \operatorname{tr}\{\Sigma_0^{-1} E(z_{i,t-1} z_{j,t-1}')\} = \sum_{i,j=1}^{p+q} \mu_i \mu_j \Omega_0^{(i,j)} \end{aligned}$$

and the Lindeberg condition is satisfied because ν_t is stationary with finite variance. It therefore follows from the central limit theorem for martingales, e.g. Hall & Heyde (1980, chp. 3), that $S_T \xrightarrow{D} N(0, 4\Omega_0)$ and hence also that $T^{1/2}\frac{\partial \det\{R(\theta_0)\}}{\partial\theta'} \xrightarrow{D} \det\{\Sigma_0\}N(0, 4\Omega_0)$. The components of the matrix Ω_0 given in Assumption F are easily found from $E(z_{i,t-1}z'_{j,t-1})$.

The second derivative in (38) is tight (stochastically equicontinuous) by Newey (1991, Corollary 2.2) if its derivative is dominated uniformly in $d \in D_3^p$ and $\psi \in \mathcal{N}_\delta(\psi_0)$ by a random variable $B_T = O_P(1)$. From Lemma B.3 with $u_1 = u_2 = d - d_0 \geq -1/2 + \kappa_3$ and $\tilde{\Psi} = \mathcal{N}_\delta(\psi_0)$ (noting that only summability of the linear coefficients is assumed in Lemma B.3 and this is satisfied uniformly on $\mathcal{N}_\delta(\psi_0)$ by the derivatives of $C_{kn}(\psi)$ by Assumption E) it holds that $B_T = \sup_{d\in D_3^p, \psi\in\mathcal{N}_\delta(\psi_0)} \frac{\partial^3 R(\theta)}{\partial\theta^{(3)}} = O_P(1)$, showing that the second derivative in (38) is tight. This result, together with consistency of $\hat{\theta}$ (Theorem 1), implies by Lemma A.3 of Johansen & Nielsen (2010) that the second derivative in (38) can be evaluated at the true value. Hence,

$$\begin{aligned}
\frac{\partial^2 \det\{R(\theta_0)\}}{\partial\theta_i\partial\theta_j} = {} & \det\{R(\theta_0)\}\operatorname{tr}\{R(\theta_0)^{-1}2T^{-1}\sum_{t=1}^{T} y_{i,j,t-1}\varepsilon_t(\theta_0)'\} + \det\{R(\theta_0)\}\operatorname{tr}\{R(\theta_0)^{-1}2T^{-1}\sum_{t=1}^{T} y_{i,t-1}y'_{j,t-1}\} \\
& + \det\{R(\theta_0)\}\operatorname{tr}\{R(\theta_0)^{-1}2T^{-1}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_t(\theta_0)'\}\operatorname{tr}\{R(\theta_0)^{-1}2T^{-1}\sum_{t=1}^{T} y_{j,t-1}\varepsilon_t(\theta_0)'\} \\
& - \det\{R(\theta_0)\}\operatorname{tr}\{R(\theta_0)^{-1}2T^{-1}\sum_{t=1}^{T} y_{i,t-1}\varepsilon_t(\theta_0)'R(\theta_0)^{-1}2T^{-1}\sum_{s=1}^{T} y_{j,s-1}\varepsilon_s(\theta_0)'\},
\end{aligned}$$

where $y_{i,j,t-1} = \frac{\partial}{\partial\theta_i\partial\theta_j}\varepsilon_t(\theta)|_{\theta=\theta_0}$. From the argument in Robinson (2006, pp. 135–136), $R(\theta_0)$, $\varepsilon_t(\theta_0)$, $y_{i,t-1}$, and $y_{i,j,t-1}$ can be replaced by Σ_0, ε_t, $z_{i,t-1}$, and $z_{i,j,t-1}$, respectively, where $z_{i,j,t-1} = \frac{\partial}{\partial\theta_i\partial\theta_j}\eta_t(\theta)|_{\theta=\theta_0}$. Because $z_{i,j,t-1}\varepsilon'_t$, $z_{i,t-1}\varepsilon'_t$, and $z_{j,t-1}\varepsilon'_t$ are martingale difference sequences with respect to $\mathcal{F}_t$ with finite second moments, the first, third, and fourth terms on the right-hand side are $o_P(1)$. Finally, from the law of large numbers it follows that $\operatorname{tr}\{\Sigma_0^{-1}T^{-1}\sum_{t=1}^{T} z_{i,t-1}z'_{j,t-1}\} \xrightarrow{P} \Omega_0^{(i,j)}$ such that $\frac{\partial^2 \det\{R(\theta_0)\}}{\partial\theta_i\partial\theta_j} \xrightarrow{P} 2\det\{\Sigma_0\}\Omega_0^{(i,j)}$, which proves the result.

6 Concluding remarks

This paper has proven consistency and asymptotic normality for the conditional-sum-of-squares estimator for multivariate fractional time series. The model considered is parametric and allows for a wide range of weak dependence in the linear process innovations. In particular, it encompasses the multivariate non-cointegrated fractional ARIMA model. In this fractional context, the consistency proof is the most challenging, since the objective function does not converge uniformly in probability on the (arbitrarily large) parameter set. The consistency result is of course important not only in its own right but also as a required prerequisite for a proof of asymptotic normality of the estimator. On the other hand, the proof of asymptotic normality, given the consistency result, is more standard.

Empirical applications and finite sample simulations of the CSS estimation procedure for fractional time series models are well-known, see e.g. Nielsen & Frederiksen (2005), Hualde & Robinson (2011), and the references therein.

Appendix A Inequalities

This section presents some useful inequalities that are applied both in the proofs of the main theorems and in proofs of product moment bounds in the next section.

Lemma A.1 *Uniformly in $-u_0 \le v \le u \le u_0$ and for $j \ge 1, m \ge 0$ it holds that*

$$|\frac{\partial^m}{\partial u^m}\pi_j(u)| \le c(1+\log j)^m j^{u-1}, \tag{39}$$

$$|\frac{\partial^m}{\partial u^m}\pi_j(u) - \frac{\partial^m}{\partial v^m}\pi_j(v)| \le c(u-v)(1+\log j)^{m+1} j^{u-1}, \tag{40}$$

$$|\frac{\partial^m}{\partial u^m}\pi_{j+1}(u) - \frac{\partial^m}{\partial u^m}\pi_j(u)| \le c(1+\log j)^m j^{u-2}, \tag{41}$$

where the constant $c > 0$ does not depend on $u, \tilde{u}$, or j.

Uniformly in $-\delta_0 \le v + 1/2 \le \delta_0$ for $\delta_0 < 1/2$ and $j \ge 1$ it holds that

$$\pi_j(-v) \ge c j^{-v-1}, \tag{42}$$

where the constant $c > 0$ does not depend on v or j.

Proof. The results (39) and (40) are in Lemma B.3 of Johansen & Nielsen (2010). From (2) it holds that $\pi_{j+1}(u) - \pi_j(u) = \pi_j(u)(u-1)/(j+1) = \pi_{j+1}(u-1)$ such that (41) follows directly from (39).

To prove (42) 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 $\sup_{1/2-\delta_0 \le u \le 1/2+\delta_0} |\epsilon(u,j)| \to 0$ as $j \to \infty$. This proves the result and shows that the constant c can be chosen to depend only on δ_0. ■

Lemma A.2 *Uniformly for* $\max(|\alpha|, |\beta|) \le a_0$ *it holds that*

$$\sum_{j=1}^{t-1} j^{\alpha-1}(t-j)^{\beta-1} \le c(1+\log t) t^{\max(\alpha+\beta-1, \alpha-1, \beta-1)}, \tag{43}$$

where the constant $c > 0$ does not depend on α, β, or t.

Proof. See Lemma B.4 of Johansen & Nielsen (2010). ■

Lemma A.3 *Let $u = (u_k, u_l)$ and $F_N^{(k,l)}(u) = \sum_{n=0}^{N-1} \pi_n(-u_k)\pi_n(-u_l)$. For $N \ge 2, u_k, u_l \le -1/2 + a$, and $a > 0$,*

$$F_N^{(k,l)}(u) \ge 1 + c\frac{1-(N-1)^{-2a}}{2a},$$

where the constant $c > 0$ does not depend on a, u, or N.

Proof. Using (42) of Lemma A.1 it holds that $F_N^{(k,l)}(u) \ge 1 + c\sum_{n=1}^{N-1} n^{-u_k-u_l-2}$ and the result follows because $\sum_{n=1}^{N-1} n^{-u_k-u_l-2} \ge \int_1^{N-1} x^{-u_k-u_l-2}\mathrm{d}x = (u_k+u_l+1)^{-1}(1-(N-1)^{-u_k-u_l-1}) \ge (2a)^{-1}(1-(N-1)^{-2a})$. ■

Appendix B Product moment bounds

This section contains a series of lemmas that are used to verify tightness and stochastic equicontinuity conditions for the processes in the previous sections. The first lemma deals with nonstationary processes and the next lemma with product moments of processes that are nearly stationary. It is Lemma B.2 that contains the truncation argument used to deal with the non-uniform convergence in Θ_2, see Section 3.3.1. Lemma B.3 covers product moments of stationary and nearly stationary processes, and is applied in the consistency proof – both for the stationary processes and to deal with certain cross-products of stationary and nearly stationary processes – and it is applied for the Hessian in the proof of asymptotic normality. The final two lemmas derive bounds for cross-products of nonstationary and (nearly) stationary processes.

Lemma B.1 *Let ε_t satisfy Assumption A. Then, for $k = 1, \ldots, p$,*

$$||T^{u+1/2}\Delta_+^u \varepsilon_{k,t}||_2 \leq c \text{ and } ||T^{u+1/2}\Delta_+^u \varepsilon_{k,t} - T^{v+1/2}\Delta_+^v \varepsilon_{k,t}||_2 \leq c|u-v| \tag{44}$$

uniformly in $(u,v) \in \tilde{D} \times \tilde{D}$, where $\tilde{D} = \{u \in D : u + 1/2 \leq -b\}$ for $b > 0$ and where the constant $c > 0$ does not depend on u, v, or T.

Proof. See Lemma C.3 in Johansen & Nielsen (2010), which applies also under Assumption A on ε_t instead of Johansen & Nielsen's (2010) i.i.d. assumption. ■

Lemma B.2 *Let Assumptions A and B be satisfied. Define $w_{kt} = w_{kt}(u_k) = \sum_{n=0}^{N-1} \pi_n(-u_k) C_k^*(\psi)\varepsilon_{t-n}$ and $v_{kt} = v_{kt}(u_k) = \sum_{n=N}^{t-1} \pi_n(-u_k) C_k^*(\psi)\varepsilon_{t-n}$, where the $p \times p$ matrix $C_k^*(\psi)$ satisfies $0 < |C_k^*(\psi)| < \infty$ uniformly in $\psi \in \Psi$. Define the product moments $Q_{1NT}(u,\psi) = T^{-1}\sum_{t=N+1}^T w_{kt}w_{lt}' - C_k^*(\psi)\Sigma_0 C_l^*(\psi) T^{-1}(T-N)F_N^{(k,l)}(u)$ and $Q_{2NT}(u,\psi) = T^{-1}\sum_{t=N+1}^T w_{kt}v_{lt}$, with $F_N^{(k,l)}(u)$ given in Lemma A.3 and $u = (u_k, u_l)'$, and the set $\tilde{\Theta} = \{(u_k, u_l, \psi) \in D \times D \times \Psi : |u_k + 1/2| \leq a, |u_l + 1/2| \leq a\}$ for $a \in (0, 1/2)$. Then*

$$\sup_{(u,\psi)\in\tilde{\Theta}} |Q_{2NT}(u,\psi)| = O_P((\log T)^2 T^{-1/2+a} N^{1/2+a}), \tag{45}$$

$$\sup_{(u,\psi)\in\tilde{\Theta}} |Q_{1NT}(u,\psi)| = O_P((\log T)^2 T^{-1/2} N^{1/2+2a}). \tag{46}$$

In particular, if $N = T^\alpha$ with $0 < \alpha < \min(\frac{1/2-a}{1/2+a}, \frac{1/2}{1/2+2a})$, then it holds that

$$\sup_{(u,\psi)\in\tilde{\Theta}} |Q_{1NT}(u,\psi)| \xrightarrow{P} 0 \text{ and } \sup_{(u,\psi)\in\tilde{\Theta}} |Q_{2NT}(u,\psi)| \xrightarrow{P} 0. \tag{47}$$

Proof. *Proof of (45)*: Let $\varepsilon_{kt}^* = C_k^*(\psi)\varepsilon_t$. Rearranging the summations,

$$Q_{2NT}(u,\psi) = T^{-1}\sum_{n=0}^{N-1} \pi_n(-u_k) \sum_{m=N}^{T-1} \pi_m(-u_l) \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime},$$

where summation by parts yields

$$\begin{aligned}\sum_{m=N}^{T-1} \pi_m(-u_l) \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime} &= \pi_{T-1}(-u_l) \sum_{m=N}^{T-1} \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime} \\ &\quad - \sum_{r=N}^{T-2} (\pi_{r+1}(-u_l) - \pi_r(-u_l)) \sum_{m=N}^{r} \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime}.\end{aligned}$$

Now let

$$\sum_{m=N}^{K} \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime} = C_k^*(\psi) \sum_{s=1}^{T-n} z_s C_l^*(\psi)',$$

where $z_s = \varepsilon_s \sum_{k=N-n}^{K-n} \varepsilon_{s-k}'$ is mean zero with

$$E|z_s|^2 = E\,\mathrm{tr}\{z_s' z_s\} = O(K-N).$$

Using serial uncorrelatedness of z_s it follows that

$$E\left|\sum_{s=1}^{T-n} z_s\right|^2 \leq \sum_{s=1}^{T-n} E|z_s|^2 = O((T-n)(K-N)),$$

such that the L_2-norm is $||\sum_{s=1}^{T-n} z_s||_2 = (E|\sum_{s=1}^{T-n} z_s|^2)^{1/2} = O((T-n)^{1/2}(K-N)^{1/2})$ uniformly in n, K. Hence, $\sum_{s=1}^{T-n} z_s = O_P((T-n)^{1/2}(K-N)^{1/2})$, and because $0 < |C_k^*(\psi)| < \infty$ uniformly in $\psi \in \Psi$,

$$\begin{aligned}\sup_{\psi\in\Psi} |\sum_{m=N}^{K} \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime}| &\leq \sup_{\psi\in\Psi} |C_k^*(\psi)||C_l^*(\psi)'| \left|\sum_{s=1}^{T-n} z_s\right| \\ &= O_P((T-n)^{1/2}(K-N)^{1/2}).\end{aligned}$$

Now, rearranging the summations and applying the summation by parts result, $\sup_{(u,\psi)\in\tilde{\Theta}} |Q_{2NT}(u,\psi)|$ is

$$\begin{aligned}&O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}} T^{-1/2} \sum_{n=0}^{N-1} |\pi_n(-u_k)||\pi_{T-1}(-u_l)|(T-n)^{1/2}\right) \\ &+O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}} T^{-1} \sum_{n=0}^{N-1} |\pi_n(-u_k)| \sum_{m=N}^{T-2} (\pi_{m+1}(-u_l) - \pi_m(-u_l))(T-n)^{1/2}(m-N)^{1/2}\right).\end{aligned}$$

By Lemma A.1 the first term is

$$\sup_{(u,\psi)\in\tilde{\Theta}} T^{-1/2} \sum_{n=0}^{N-1} |\pi_n(-u_k)||\pi_{T-1}(-u_l)|(T-n)^{1/2} \leq \sup_{(u,\psi)\in\tilde{\Theta}} cT^{-u_l-1} \sum_{n=0}^{N-1} n^{-u_k-1} \leq c(\log T)T^{-1/2+a}N^{1/2+a}$$

and the second term is

$$\sup_{(u,\psi)\in\tilde{\Theta}} cT^{-1/2} \sum_{n=0}^{N-1} n^{-u_k-1} \sum_{m=N}^{T-2} m^{u_l-3/2} \leq cT^{-1/2} \sum_{n=0}^{N-1} n^{-1/2+a} \sum_{m=N}^{T-2} m^{a-1} \leq c(\log T)^2 T^{-1/2+a} N^{1/2+a}.$$

Proof of (46): Decompose $Q_{1NT}(u,\psi)$ as

$$\begin{aligned}Q_{1NT}(u,\psi) &= \sum_{n=0}^{N-1} \pi_n(-u_k)\pi_n(-u_l)C_k^*(\psi)T^{-1} \sum_{t=N+1}^{T} (\varepsilon_{t-n}\varepsilon_{t-n}' - \Sigma_0)C_l^*(\psi)' && (48) \\ &+2T^{-1} \sum_{m=1}^{N-1} \pi_m(-u_l) \sum_{n=0}^{m-1} \pi_n(-u_k) \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime}, && (49)\end{aligned}$$

where $T^{-1}\sum_{t=N+1}^{T}(\varepsilon_{t-n}\varepsilon_{t-n}' - \Sigma_0) = O_P(T^{-1/2})$ uniformly in N under Assumption A. Thus, because $0 < |C_k^*(\psi)| < \infty$ uniformly in $\psi \in \Psi$,

$$\begin{aligned}\sup_{(u,\psi)\in\tilde{\Theta}} |(48) &= O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}} T^{-1/2} \sum_{n=0}^{N-1} |\pi_n(-u_k)\pi_n(-u_l)|\right) \\ &= O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}} T^{-1/2} \sum_{n=0}^{N-1} n^{-u_k-u_l-2}\right) = O_P\left((\log T)T^{-1/2}N^{2a}\right).\end{aligned}$$

Proceeding as in the proof of (45), summation by parts yields

$$\begin{aligned}\sum_{n=0}^{m-1} \pi_n(-u_k) \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^* &= \pi_{m-1}(-u_k) \sum_{n=0} \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime} \\ &- \sum_{r=0}^{m-2} (\pi_{r+1}(-u_k) - \pi_r(-u_k)) \sum_{n=0}^{l} \sum_{t=\max(N,m)+1}^{T} \varepsilon_{k,t-n}^* \varepsilon_{l,t-m}^{*\prime},\end{aligned}$$

where $\sup_{\psi\in\Psi}|\sum_{n=0}^{K}\sum_{t=\max(N,m)+1}^{T}\varepsilon_{k,t-n}^{*}\varepsilon_{l,t-m}^{*\prime}|=O_P(T^{1/2}K^{1/2})$. Thus, $\sup_{(u,\psi)\in\tilde{\Theta}}|(49)$ is

$$O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}}T^{-1/2}\sum_{m=1}^{N-1}|\pi_m(-u_l)||\pi_{m-1}(-u_k)|m^{1/2}\right)$$
$$+O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}}T^{-1/2}\sum_{m=1}^{N-1}|\pi_m(-u_l)||\sum_{r=0}^{m-2}(\pi_{r+1}(-u_k)-\pi_r(-u_k))|r^{1/2}\right),$$

where the first term is

$$O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}}T^{-1/2}\sum_{m=1}^{N-1}m^{-u_k-u_l-3/2}\right)=O_P((\log T)T^{-1/2}N^{1/2+2a})$$

and the second term is

$$O_P\left(\sup_{(u,\psi)\in\tilde{\Theta}}T^{-1/2}\sum_{m=1}^{N-1}m^{-u_l-1}\sum_{r=1}^{m-2}r^{-u_k-3/2}\right)=O_P((\log T)^2T^{-1/2}N^{1/2+2a}).$$

Proof of (47): Using the condition on α, the right-hand sides of (45) and (46) converge to zero. ■

Lemma B.3 *Let Assumptions A and B be satisfied and let* $Z_{it}=\sum_{n=0}^{\infty}\zeta_{in}(\psi)\varepsilon_{t-n}, i=1,2,$ *where the* $1\times p$ *coefficients* $\zeta_{in}(\psi)$ *satisfy* $\sum_{n=0}^{\infty}|\zeta_{in}(\psi)|<\infty, i=1,2,$ *uniformly in* $\psi\in\tilde{\Psi}\subseteq\Psi$. *Define the product moment* $Q_T(u_1,u_2,\psi)=T^{-1}\sum_{t=1}^{T}\frac{\partial^k}{\partial u_1^{(k)}}(\Delta_+^{u_1}Z_{1t})\frac{\partial^l}{\partial u_2^{(l)}}(\Delta_+^{u_2}Z_{2t})$ *for* $k,l\geq 0$ *and the set* $\tilde{\Theta}=\{(u_1,u_2,\psi)\in D\times D\times\tilde{\Psi}:\min(u_1+1,u_2+1,u_1+u_2+1)\geq a\}$. *Then*

$$\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}|Q_T(u_1,u_2,\psi)| = O_P(1) \text{ for } a>0,$$
$$\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}|Q_T(u_1,u_2,\psi)| = O_P((\log T)^{1+k+l}T^{-a}) \text{ for } a\leq 0$$

Proof. The proof is given only for $k,l=0$ since the derivatives just add a log-factor, see (39), which does not change the proof. Also, $p=1$ is assumed in the proof to ease the notation.

Rearranging the summations and using symmetry, $Q_T(u_1,u_2,\psi)$ is

$$T^{-1}\sum_{j=0}^{T-1}\pi_j(-u_1)\sum_{n,m=0}^{\infty}\zeta_{1n}(\psi)\zeta_{2m}(\psi)\sum_{k=0}^{T-1}\pi_k(-u_2)\sum_{t=\max(j,k)+1}^{T}\varepsilon_{t-j-n}\varepsilon_{t-k-m}$$
$$= T^{-1}\sum_{j=0}^{T-1}\pi_j(-u_1)\sum_{n=0}^{\infty}\sum_{m=\max(0,j+n-T+1)}^{j+n}\zeta_{1n}(\psi)\zeta_{2m}(\psi)\pi_{j+n-m}(-u_2)\sum_{t=\max(j,j+n-m)+1}^{T}\varepsilon_{t-j-n}^2 \quad (50)$$
$$+2T^{-1}\sum_{j=0}^{T-1}\pi_j(-u_1)\sum_{n,m=0}^{\infty}\zeta_{1n}(\psi)\zeta_{2m}(\psi)\sum_{k=0}^{\min(T,j+n-m)-1}\pi_k(-u_2)\sum_{t=\max(j,k)+1}^{T}\varepsilon_{t-j-n}\varepsilon_{t-k-m}. \quad (51)$$

Since $T^{-1}\sum_{t=\max(j,j+n-m)+1}^{T}\varepsilon_{t-j-n}^2=O_P(1)$ uniformly in j,n,m it holds that $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}|(50)|$ is

$$O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}\sum_{n=0}^{\infty}\sum_{m=\max(0,n-T+1)}^{T-1+n}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=\max(0,m-n)}^{\min(T-1,T-1+m-n)}|\pi_j(-u_1)||\pi_{j+n-m}(-u_2)|\right)$$
$$=O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}\sum_{n=0}^{\infty}\sum_{m=\max(0,n-T+1)}^{T-1+n}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=1+\max(0,m-n)}^{\min(T-1,T-1+m-n)}j^{-u_1-1}(j+n-m)^{-u_2-1}\right).$$

If $a > 0$ the summation over j is bounded and then $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |(50)| = O_P(1)$ because $\sum_{n=0}^{\infty} |\zeta_{in}(\psi)| < \infty$ uniformly in $\psi \in \tilde{\Psi}, i = 1, 2$. If $a \leq 0$ the summation over j is $O_P((\log T)T^{-a})$ which is then also the bound for the supremum of (50).

Next, summation by parts yields

$$\begin{aligned}
&\sum_{k=0}^{\min(T,j+n-m)-1} \pi_k(-u_2) \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} \\
&= \pi_{j+n-m-1}(-u_2) \sum_{k=0}^{\min(T,j+n-m)-1} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} \\
&\quad - \sum_{l=0}^{\min(T,j+n-m)-2} (\pi_{l+1}(-u_2) - \pi_l(-u_2)) \sum_{k=0}^{l} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m}, \qquad (52)
\end{aligned}$$

where

$$\begin{aligned}
\sum_{k=0}^{\min(T,j+n-m)-1} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} &= \sum_{s=\max(j-T+2-m,2-n,1-m)}^{T-m} w_s, \\
\sum_{k=0}^{l} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} &= \sum_{s=\max(1-m,1+j-l-m)}^{T-m} v_s.
\end{aligned}$$

Here, $w_s = \varepsilon_s \sum_{k=\max(1,1+j+n-m-T)}^{j+n-m} \varepsilon_{s-k}$ and $v_s = \varepsilon_s \sum_{k=j+n-m-l}^{j+n-m} \varepsilon_{s-k}$ satisfy

$$\begin{aligned}
E(w_s^2) &= \sigma_0^2 E\left(\sum_{k=\max(1,1+j+n-m-T)}^{j+n-m} \varepsilon_{s-k}\right)^2 = O(T), \\
E(v_s^2) &= \sigma_0^2 E\left(\sum_{k=j+n-m-l}^{j+n-m} \varepsilon_{s-k}\right)^2 = O(l),
\end{aligned}$$

and, furthermore, w_s and v_s are both uncorrelated sequences such that

$$\begin{aligned}
E\left(\sum_{s=\max(j-T+2-m,2-n,1-m)}^{T-m} w_s\right)^2 &= \sum_{s=\max(j-T+2-m,2-n,1-m)}^{T-m} E(w_s^2) = O(T^2), \\
E\left(\sum_{s=\max(1-m,1+j-l-m)}^{T-m} v_s\right)^2 &= \sum_{s=\max(1-m,1+j-l-m)}^{T-m} E(v_s^2) = O((T+l-j)l).
\end{aligned}$$

It follows that $\sum_{k=0}^{\min(T,j+n-m)-1} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} = O_P(T)$ and $\sum_{k=0}^{l} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} = O_P((T+l-j)^{1/2}l^{1/2})$, in both cases uniformly in j, n, m.

Now, rearranging the summations and applying the summation by parts result, $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |(51)|$ is

$$\begin{aligned}
&O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} \sum_{n,m=0}^{\infty} |\zeta_{1n}(\psi)||\zeta_{2m}(\psi)| \sum_{j=\max(0,1+m-n)}^{\min(T-1,T+m-n)} |\pi_j(-u_1)||\pi_{j+n-m-1}(-u_2)|\right) \qquad (53) \\
&+O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-1} \sum_{j=0}^{T-1} |\pi_j(-u_1)| \sum_{n,m=0}^{\infty} |\zeta_{1n}(\psi)||\zeta_{2m}(\psi)| \sum_{l=0}^{T-2} |\pi_{l+1}(-u_2) - \pi_l(-u_2)|(T+l-j)^{1/2}l^{1/2}\right) \qquad (54)
\end{aligned}$$

The result for (53) follows as in the analysis of (50). For term (54) it holds, using (41) and that $\sum_{n=0}^{\infty}|\zeta_{in}(\psi)|<\infty$ uniformly in $\psi\in\tilde{\Psi}, i=1,2$, that the order is

$$\begin{aligned}
&\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-1}\sum_{j=1}^{T-1} j^{-u_1-1}\sum_{l=1}^{T-2} l^{-u_2-3/2}(T+l-j)^{1/2} \\
\leq\ &\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-1}\sum_{l=1}^{T-2} l^{-u_2-3/2}\sum_{j=1}^{T+l-1} j^{-u_1-1}(T+l-j)^{1/2} \\
\leq\ &\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} c(\log T)T^{-1}\sum_{l=1}^{T-2} l^{-u_2-3/2}(T+l)^{\max(1/2,1/2-u_1)} \\
\leq\ &\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} c(\log T)T^{-1/2}\sum_{l=1}^{T-2} l^{-u_2-3/2+\max(0,-u_1)},
\end{aligned}$$

where the second inequality follows from Lemma A.2 and the third because $(T+l)^{\max(1/2,1/2-u_1)}=(T+l)^{1/2}(T+l)^{\max(0,-u_1)}\leq(2T)^{1/2}l^{\max(0,-u_1)}$. Since $-u_2-3/2+\max(0,-u_1)=-\min(u_2+1,u_1+u_2+1)-1/2\leq -a-1/2$, the right-hand side is bounded by $c(\log T)^2T^{-1/2}T^{\max(0,1/2-a)}=c(\log T)^2T^{\max(-1/2,-a)}$ if $a>0$ and $c(\log T)T^{-1/2}T^{1/2-a}=c(\log T)T^{-a}$ if $a\leq 0$. ■

In the next lemma, note that when $N=0$, $Q_{2T}(u_1,u_2,\psi)=T^{u_2-1/2}\sum_{t=1}^{T}(\Delta_+^{u_1}Z_{1t})(\Delta_+^{u_2}Z_{2t})$.

Lemma B.4 *Let Assumptions A and B be satisfied and let* $Z_{it}=\sum_{n=0}^{\infty}\zeta_{in}(\psi)\varepsilon_{t-n}, i=1,2$, *where the* $1\times p$ *coefficients* $\zeta_{in}(\psi)$ *satisfy* $\sum_{n=0}^{\infty}|\zeta_{in}(\psi)|<\infty, i=1,2$, *uniformly in* $\psi\in\tilde{\Psi}\subseteq\Psi$. *Define the product moments* $Q_{1T}(u_1,u_2,\psi)=T^{u_2-1/2}\sum_{t=N+1}^{T}(\sum_{j=0}^{N-1}\pi_j(-u_1)Z_{1,t-j})(\Delta_+^{u_2}Z_{2t})$ *and* $Q_{2T}(u_1,u_2,\psi)=T^{u_2-1/2}\sum_{t=N+1}^{T}(\sum_{j=N}^{t-1}\pi_j(-u_1)Z_{1,t-j})(\Delta_+^{u_2}Z_{2t})$ *and the set* $\tilde{\Theta}=\{(u_1,u_2,\psi)\in D\times D\times\tilde{\Psi}:u_1+1/2\geq a, u_2+1/2\leq -b\}$ *for some* $b>0$. *Then*

$$\begin{aligned}
\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}|Q_{1T}(u_1,u_2,\psi)| &= O_P((\log T)T^{-b}N^{\max(b-a,0)})+O_P((\log T)T^{-1/2}N^{\max(1/2-a,0)}),\\
\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}|Q_{2T}(u_1,u_2,\psi)| &= O_P((\log T)T^{-\min(a,b,1/2)}).
\end{aligned}$$

Proof. The proof follows that of Lemma B.3 and is given for Q_{1T} only, since that for Q_{2T} follows in exactly the same way replacing N by T in the relevant summation limits. As in the proof of Lemma B.3, $p=1$ is assumed to ease the notation.

Rearranging the summations, $Q_{1T}(u_1,u_2,\psi)$ is

$$\begin{aligned}
&T^{u_2-1/2}\sum_{j=0}^{N-1}\pi_j(-u_1)\sum_{n,m=0}^{\infty}\zeta_{1n}(\psi)\zeta_{2m}(\psi)\sum_{k=0}^{T-1}\pi_k(-u_2)\sum_{t=\max(j,k,N)+1}^{T}\varepsilon_{t-j-n}\varepsilon_{t-k-m}\\
=\ &T^{u_2-1/2}\sum_{j=0}^{N-1}\pi_j(-u_1)\sum_{n=0}^{\infty}\sum_{m=\max(0,j+n-T+1)}^{j+n}\zeta_{1n}(\psi)\zeta_{2m}(\psi)\pi_{j+n-m}(-u_2)\sum_{t=\max(j,j+n-m,N)+1}^{T}\varepsilon_{t-j-n}^2 && (55)\\
&+T^{u_2-1/2}\sum_{j=0}^{N-1}\pi_j(-u_1)\sum_{n,m=0}^{\infty}\zeta_{1n}(\psi)\zeta_{2m}(\psi)\sum_{k=0}^{\min(T,j+n-m)-1}\pi_k(-u_2)\sum_{t=\max(j,k,N)+1}^{T}\varepsilon_{t-j-n}\varepsilon_{t-k-m} && (56)\\
&+T^{u_2-1/2}\sum_{j=0}^{N-1}\pi_j(-u_1)\sum_{n,m=0}^{\infty}\zeta_{1n}(\psi)\zeta_{2m}(\psi)\sum_{k=\max(0,j+n-m+1)}^{T-1}\pi_k(-u_2)\sum_{t=\max(j,k,N)+1}^{T}\varepsilon_{t-j-n}\varepsilon_{t-k-m} && (57)
\end{aligned}$$

Since $T^{-1}\sum_{t=\max(j,j+n-m,N)+1}^{T}\varepsilon_{t-j-n}^2 = O_P(1)$ uniformly in j,n,m,N it holds that $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}|(55)|$ is

$$O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2+1/2}\sum_{n=0}^{\infty}\sum_{m=\max(0,n-T+1)}^{N-1+n}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=\max(0,m-n)}^{\min(N-1,T-1+m-n)}|\pi_j(-u_1)||\pi_{j+n-m}(-u_2)|\right)$$
$$= O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-b}\sum_{n=0}^{\infty}\sum_{m=\max(0,n-T+1)}^{N-1+n}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=1+\max(0,m-n)}^{\min(N-1,T-1+m-n)} j^{-a-1/2}(j+n-m)^{b-1/2}\right).$$

If $b\geq 1/2$ then $(j+n-m)^{b-1/2}\leq T^{b-1/2}$ and the bound for (55) is

$$O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-1/2}\sum_{n,m=0}^{\infty}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=1}^{N-1} j^{-a-1/2}\right) = O_P((\log T)T^{-1/2}N^{\max(1/2-a,0)})$$

because $\sum_{n=0}^{\infty}|\zeta_{in}(\psi)|<\infty$ uniformly in $\psi\in\tilde{\Psi}, i=1,2$. If $b<1/2$ the bound is

$$O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-b}\sum_{n=0}^{\infty}\sum_{m=\max(0,n-T+1)}^{n-1}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=1}^{\min(N-1,T-1+m-n)} j^{-a-1/2}(j+n-m)^{b-1/2}\right) \quad (58)$$
$$+O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-b}\sum_{n=0}^{\infty}\sum_{m=n}^{N-1+n}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=1+m-n}^{N-1} j^{-a-1/2}(j+n-m)^{b-1/2}\right). \quad (59)$$

In (58), $(j+n-m)^{b-1/2}\leq j^{b-1/2}$ such that the bound is

$$O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-b}\sum_{n=0}^{\infty}\sum_{m=\max(0,n-T+1)}^{n-1}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{j=1}^{N-1} j^{b-a-1}\right) = O_P((\log T)T^{-b}N^{\max(b-a,0)})$$

because $\sum_{n=0}^{\infty}|\zeta_{in}(\psi)|<\infty$ uniformly in $\psi\in\tilde{\Psi}, i=1,2$. For (59), the summation over j is bounded if $b<a$ and $O_P((\log T)N^{b-a})$ if $b\geq a$. Because $\sum_{n=0}^{\infty}|\zeta_{in}(\psi)|<\infty$ uniformly in $\psi\in\tilde{\Psi}, i=1,2$, it then follows that (59) is $O_P((\log T)T^{-b}N^{\max(b-a,0)})$.

Next, summation by parts as in (52) shows that $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}}|(56)|$ is

$$O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2+1/2}\sum_{j=0}^{N-1}|\pi_j(-u_1)|\sum_{n,m=0}^{\infty}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)||\pi_{j+n-m-1}(-u_2)|\right) \quad (60)$$
$$+O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2}\sum_{j=0}^{N-1}|\pi_j(-u_1)|\sum_{n,m=0}^{\infty}|\zeta_{1n}(\psi)||\zeta_{2m}(\psi)|\sum_{l=0}^{T-2}|\pi_{l+1}(-u_2)-\pi_l(-u_2)|l^{1/2}\right) \quad (61)$$

The analysis of (60) is identical to that of (55) and the bound is the same. Since $\sum_{n=0}^{\infty}|\zeta_{in}(\psi)|<\infty$ uniformly in $\psi\in\tilde{\Psi}, i=1,2$, and $-u_2-3/2\geq b-1>-1$, (61) is of order

$$\begin{aligned}\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2}\sum_{j=1}^{N-1} j^{-u_1-1}\sum_{l=1}^{T-2} l^{-u_2-3/2} &\leq c\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2}(\log T)N^{\max(0,-u_1)}T^{-u_2-1/2}\\ &\leq c(\log T)N^{\max(0,1/2-a)}T^{-1/2}.\end{aligned}$$

Finally, for (57) apply summation by parts again,

$$\begin{aligned}
&\sum_{k=\max(0,j+n-m+1)}^{T-1} \pi_k(-u_2) \sum_{t=\max(j,k,N)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} \\
= \;& \pi_{T-1}(-u_2) \sum_{k=\max(0,j+n-m+1)}^{T-1} \sum_{t=\max(j,k,N)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} \\
& - \sum_{l=\max(0,j+n-m+1)}^{T-1} (\pi_{l+1}(-u_2)-\pi_l(-u_2)) \sum_{k=\max(0,j+n-m+1)}^{l} \sum_{t=\max(j,k,N)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m},
\end{aligned}$$

where, as before, it holds that, uniformly in j, n, m,

$$\begin{aligned}
\sum_{k=\max(0,j+n-m+1)}^{T-1} \sum_{t=\max(j,k,N)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} &= \sum_{s=\max(1-n,2-m,1+N-j-n)}^{T-j-n} \varepsilon_s \sum_{k=\max(1,m-n-j)}^{T-1+m-j-n} \varepsilon_{s-k} = O_P(T), \\
\sum_{k=\max(0,j+n-m+1)}^{l} \sum_{t=\max(j,k,N)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} &= \sum_{s=\max(1-n,2-m,1+N-j-n)}^{T-j-n} \varepsilon_s \sum_{k=\max(1,m-n-j)}^{l+m-j-n} \varepsilon_{s-k} = O_P(l^{1/2}T^{1/2}),
\end{aligned}$$

such that $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |(57)|$ is

$$O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2+1/2} \sum_{j=0}^{N-1} |\pi_j(-u_1)| \sum_{n,m=0}^{\infty} |\zeta_{1n}(\psi)||\zeta_{2m}(\psi)||\pi_{T-1}(-u_2)|\right) \tag{62}$$

$$+O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2} \sum_{j=0}^{N-1} |\pi_j(-u_1)| \sum_{n,m=0}^{\infty} |\zeta_{1n}(\psi)||\zeta_{2m}(\psi)| \sum_{l=\max(0,j+n-m+1)}^{T-1} |\pi_{l+1}(-u_2)-\pi_l(-u_2)| l^{1/2}\right) \tag{63}$$

Because $\sum_{n=0}^{\infty} |\zeta_{in}(\psi)| < \infty$ uniformly in $\psi \in \tilde{\Psi}, i = 1, 2$, (62) is bounded by

$$\begin{aligned}
O_P\left(\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_2+1/2}|\pi_{T-1}(-u_2)| \sum_{j=0}^{N-1} |\pi_j(-u_1)|\right) &= O_P\left(T^{-1/2}\sum_{j=1}^{N-1} j^{-a-1/2}\right) \\
&= O_P((\log T)T^{-1/2}N^{\max(1/2-a,0)}),
\end{aligned}$$

while (63) is bounded in an identical way to (61). ■

Lemma B.5 *Let Assumptions A and B be satisfied and let* $Z_{it} = \sum_{n=0}^{\infty} \zeta_{in}(\psi)\varepsilon_{t-n}, i = 1, 2$, *where the* $1 \times p$ *coefficients* $\zeta_{in}(\psi)$ *satisfy* $\sum_{n=0}^{\infty} |\zeta_{in}(\psi)| < \infty, i = 1, 2$, *uniformly in* $\psi \in \Psi$. *Define the product moment* $Q_T(u_1, u_2, \psi) = T^{u_1+u_2} \sum_{t=1}^{T} (\Delta_+^{u_1} Z_{1t})(\Delta_+^{u_2} Z_{2t})$ *and the set* $\tilde{\Theta} = \{(u_1, u_2, \psi) \in D \times D \times \Psi : u_i \leq -1/2 - b, i = 1, 2\}$ *for* $b > 0$. *Then*

$$\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |Q_T(u_1, u_2, \psi)| = O_P(1).$$

Proof. The proof is almost identical to that of Lemma B.3, so only the differences are outlined. The product moment $Q_T(u_1, u_2, \psi)$ is decomposed into (50) and (51), multiplied by $T^{u_1+u_2+1}$, where (51) is further decomposed into (53) and (54) using the summation by parts result (52).

For the term (50), with the normalization of this lemma, the relevant bound is

$$\begin{aligned}
&\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_1+u_2+1} \sum_{j=\max(0,m-n)}^{\min(T-1,T-1+m-n)} |\pi_j(-u_1)||\pi_{j+n-m}(-u_2)| \\
\leq\ & cT^{-1} \sum_{j=1+\max(0,m-n)}^{\min(T-1,T-1+m-n)} (j/T)^{b-1/2}((j+n-m)/T)^{b-1/2} \\
\leq\ & cT^{-1} \sum_{j=1}^{T-1} (j/T)^{2b-1} \to c\int_0^1 x^{2b-1}\mathsf{d}x < \infty,
\end{aligned}$$

where the second inequality is Cauchy-Schwarz and the last line is because $b > 0$. The bound for term (53) is identical to that just derived for (50). Finally, the bound for term (54), with the normalization of this lemma, is

$$\begin{aligned}
&\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{u_1+u_2} \sum_{l=1}^{T-2} l^{-u_2-3/2} \sum_{j=1}^{T+l-1} j^{-u_1-1}(T+l-j)^{1/2} \\
\leq\ & T^{-1}\sum_{l=1}^{T-2}(l/T)^{b-1}T^{-1}\sum_{j=1}^{T+l-1}(j/T)^{b-1/2}((T+l-j)/T)^{1/2} \\
\leq\ & T^{-1}\sum_{l=1}^{T-2}(l/T)^{b-1}\left(T^{-1}\sum_{j=1}^{T+l-1}(j/T)^{2b-1}\right)^{1/2}\left(T^{-1}\sum_{j=1}^{T+l-1}(j/T)\right)^{1/2} \\
\leq\ & T^{-1}\sum_{l=1}^{T-2}(l/T)^{b-1}\left(T^{-1}\sum_{j=1}^{2T}(j/T)^{2b-1}\right)^{1/2}\left(T^{-1}\sum_{j=1}^{2T}(j/T)\right)^{1/2} \\
\to\ & \int_0^1 x^{b-1}\mathsf{d}x\left(\int_0^2 x^{2b-1}\mathsf{d}x\right)^{1/2}\left(\int_0^2 x\mathsf{d}x\right)^{1/2} < \infty
\end{aligned}$$

because $b > 0$. ■

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.

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

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

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

Johansen, S. & Nielsen, M. Ø. (2012*a*), 'Likelihood inference for a fractionally cointegrated vector autoregressive model', *Econometrica* **80**, 2667–2732.

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

Li, W. K. & McLeod, A. I. (1986), 'Fractional time series modelling', *Biometrika* **73**, 217–221.

Lieberman, O., Rosemarin, R. & Rousseau, J. (2012), 'Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes', *Econometric Theory* **28**, 457–470.

Magnus, J. R. & Neudecker, H. (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, revised edn, John Wiley and Sons, New York.

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

Nielsen, M. Ø. (2004), 'Efficient likelihood inference in nonstationary univariate models', *Econometric Theory* **20**, 116–146.

Nielsen, M. Ø. & Frederiksen, P. H. (2005), 'Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration', *Econometric Reviews* **24**, 405–443.

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.

Shao, X. (2010), 'Nonstationarity-extended Whittle estimation', *Econometric Theory* **26**, 1060–1087.

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.