

Cavaliere, Giuseppe; Nielsen, Morten Ørregaard; Taylor, A. M. Robert

Working Paper

Quasi-maximum likelihood estimation of heteroskedastic fractional time series models

Queen's Economics Department Working Paper, No. 1324

Provided in Cooperation with:

Queen's University, Department of Economics (QED)

Suggested Citation: Cavaliere, Giuseppe; Nielsen, Morten Ørregaard; Taylor, A. M. Robert (2014) : Quasi-maximum likelihood estimation of heteroskedastic fractional time series models, Queen's Economics Department Working Paper, No. 1324, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:

<https://hdl.handle.net/10419/122029>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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



Queen's Economics Department Working Paper No. 1324

Quasi-Maximum Likelihood Estimation of Heteroskedastic Fractional Time Series Models

Giuseppe Cavaliere
University of Bologna

Morten
Orregaard Nielsen
Queen's University and CREATES

A.M. Robert Taylor
University of Essex

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

6-2014

Quasi-Maximum Likelihood Estimation of Heteroskedastic Fractional Time Series Models*

Giuseppe Cavaliere^a, Morten Ørregaard Nielsen^b and A.M. Robert Taylor^c

^a University of Bologna

^b Queen's University and CREATES

^c University of Essex

June 10, 2014

Abstract

In a recent paper Hualde and Robinson (2011) establish consistency and asymptotic normality for conditional sum-of-squares estimators, which are equivalent to conditional quasi-maximum likelihood estimators, in parametric fractional time series models driven by conditionally homoskedastic shocks. In contrast to earlier results in the literature, their results apply over an arbitrarily large set of admissible parameter values for the (unknown) memory parameter covering both stationary and non-stationary processes and invertible and non-invertible processes. In this paper we extend their results to the case where the shocks can display conditional and unconditional heteroskedasticity of a quite general and unknown form. We establish that the consistency result presented in Hualde and Robinson (2011) continues to hold under heteroskedasticity as does asymptotic normality. However, we demonstrate that the covariance matrix of the limiting distribution of the estimator now depends on nuisance parameters which derive both from the weak dependence in the process (as is also the case for the corresponding result in Hualde and Robinson, 2011) but additionally from the heteroskedasticity present in the shocks. Asymptotically pivotal inference can be performed on the parameters of the heteroskedastic model, provided a conventional “sandwich” estimator of the variance is employed.

Keywords: (un)conditional heteroskedasticity; conditional sum-of-squares; fractional integration; quasi-maximum likelihood estimation.

J.E.L. Classifications: C13, C22.

*Parts of this paper were written while Cavaliere and Taylor visited the Economics Department at Queen's University, Canada, whose hospitality is gratefully acknowledged, as John Weatherall Distinguished Fellows. We are grateful to David Harris and Søren Johansen for comments. Nielsen thanks 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) for financial support. Cavaliere and Taylor thank the Danish Council for Independent Research, Sapere Aude | DFF Advanced Grant (Grant nr: 12-124980) for financial support. Correspondence to: Robert Taylor, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: rtaylor@essex.ac.uk

1 Introduction

This paper is concerned with establishing consistency and asymptotic normality results for the conditional quasi-maximum likelihood (QML) estimator, or equivalently the conditional sum-of-squares (CSS) estimator, in univariate fractionally integrated (or fractional) time series models which display time-variation in the variance of the driving shocks. Although the QML estimator has found widespread use in the literature, the conditions under which it is consistent in fractional time series models are only recently beginning to be well understood and those results which are available all pertain to the case where the shocks are (conditionally) homoskedastic. We extend these results to allow for both unconditional heteroskedasticity (often referred to as non-stationary volatility in the literature) and conditional heteroskedasticity of a very general form in our analysis.

Fractional models provide a parsimonious means to model a very wide range of dependence in time series data. In particular, they are flexible enough to allow for both weak and strong dependence, stationary and non-stationary behaviour, and invertibility and non-invertibility through the value of the long memory parameter in the model. However, it is this strength of the fractional model that has meant that, until only very recently, proofs of the consistency of standard parametric estimators, such as QML, have been eschewed in the literature. Consistency results are, of course, important in their own right but are also necessary prerequisites in any proof of asymptotic normality for implicitly defined estimators such as the QML estimator. The problem is due to the non-uniform convergence of the objective function when the range of values that the long memory parameter may take is large; see Hualde and Robinson (2011) and, for the multivariate extension thereof, Nielsen (2014) for detailed discussion on this point. In essence, the problem lies in the different rates of convergence of the estimator's objective function in different regions of values the long memory parameter can take. It should be noted at this stage that there exists a recent parallel literature on globally consistent semi-parametric estimators in the frequency domain. For an early example, see Shimotsu and Phillips (2005). However, these estimators achieve only semi-parametric rates of convergence for the estimator of the long memory parameter, treating weak dependence in the process non-parametrically, whereas the objective in this paper is to obtain globally consistent estimators with the usual parametric, i.e. \sqrt{T} , rate of convergence.

The aforementioned difficulties in establishing consistency with parametric estimators have previously been avoided by, for example, either restricting the range of values the long memory parameter is allowed to take to an interval of length less than one-half (see, for example, Fox and Taquq, 1986, Dahlhaus, 1989, Giraitis and Surgailis, 1990, Hosoya, 1996, and Robinson, 2006), or, as in, for example, Tanaka (1999) and Nielsen (2004) to give local consistency proofs. Alternatively, with some prior knowledge of the approximate magnitude of the long memory parameter one can (fractionally) difference the data, estimate the long memory parameter on the resulting data, and add back. However, in a recent paper, Hualde and Robinson (2011)

demonstrate the consistency and asymptotic normality of the QML estimator in parametric fractional time series models for an arbitrarily large set of admissible values of the long memory parameter which simultaneously includes the range of cases discussed above. They do so in the context of a fractional model driven by conditionally homoskedastic errors.

Our contribution in this paper is to extend the results in Hualde and Robinson (2011) to allow for both conditional and unconditional heteroskedasticity in the driving shocks. We do so using a new framework which includes the general form of non-stationary volatility considered in Cavaliere and Taylor (2005, 2008) as a special case and also includes a set of conditional heteroskedasticity conditions which are similar to those employed in Robinson (1991), Demetrescu, Kuzin and Hassler (2008) and Hassler, Rodrigues and Rubia (2009), among others. It is worth noting that neither of these conditions involve specifying a parametric model for the volatility process, and that our conditional heteroskedasticity conditions are weaker than have been previously considered in the fractional integration literature. We show, in the context of the resulting heteroskedastic fractional time series model, consistency and asymptotic normality of the QML estimator.¹ We then show that the variance of the limiting distribution of the QML estimator depends on nuisance parameters which derive both from the weak dependence present in the process (as in the corresponding result in Hualde and Robinson, 2011), but also from the heteroskedasticity present in the shocks. Asymptotically pivotal inference on the parameters of the fractional model can, however, be obtained by using standard sandwich-type estimators of the variance.

The extensions to the work of Hualde and Robinson (2011) considered in this paper are of considerable practical importance given that the well-documented failure of the assumption of conditional homoskedasticity in both empirical finance and macroeconomics; see section 2 of Gonçalves and Kilian (2004) for detailed discussion and empirical evidence on this point. Moreover, a large body of recent applied work has grown suggesting that the assumption of constant *unconditional* volatility is also at odds with the data. Sensier and van Dijk (2004) report that over 80% of the real and price variables in the Stock and Watson (1999) data-set reject the null of constant innovation variance, while Loretan and Phillips (1994) report evidence against the constancy of unconditional variances in stock market returns and exchange-rate data. Recent studies have reported that a general decline in the unconditional volatility of the shocks driving macroeconomic series in the twenty years or so leading up to the recent financial crisis is a relatively common phenomenon. This feature is known as the “great moderation”; see, *inter alia*, Kim and Nelson (1999), McConnell and Perez-Quiros (2000), Sensier and van Dijk (2004), and references therein. Indeed, the aforementioned financial crisis could be argued to

¹Notice that the results given in this paper necessarily apply to the special case of short memory processes, where $d = 0$, driven by conditionally and/or unconditionally heteroskedastic innovations. For earlier contributions relevant to the $d = 0$ case see, for example, Hannan and Heyde (1972), Hannan and Deistler (1988) and Gonçalves and Kilian (2004, 2007) who allow for conditional (but not unconditional) heteroskedasticity, and Phillips and Xu (2007) who allow for unconditional (but not conditional) heteroskedasticity, the latter in the context of a finite-order AR.

constitute a further break in the unconditional volatility of these series with volatility increasing sharply again.

The remainder of the paper is structured as follows. In the next section we present the heteroskedastic fractional time series model and the assumptions. Section 3 presents the estimator and the main results, which are the consistency theorem and the asymptotic distribution theory. Section 4 concludes, and the paper ends with appendices of proofs and auxiliary results used in the main proofs.

In the following, for a sequence of stochastic processes $X_T(s) \in \mathbb{R}$, $s = (s_1, \dots, s_m) \in S$, where S is a compact subset of m -dimensional Euclidean space, the notation $X_T \Rightarrow X$ or $X_T(s) \Rightarrow X(s)$ is used to indicate convergence in distribution of the sequence, either as continuous processes in $\mathcal{C}(S)$ or as cadlag processes in $\mathcal{D}(S)$, whereas $X_T(s) \xrightarrow{D} X(s)$ means convergence in distribution in \mathbb{R} for a fixed s and \xrightarrow{P} denotes convergence in probability. For any space Q , $\text{int}(Q)$ denotes the interior of Q ; $\mathbb{I}(\cdot)$ denotes the indicator function; $x := y$ indicates that x is defined by y . For any matrix, M , $\|M\|$ is used to denote the norm $\|M\|^2 := \text{tr}\{M'M\}$ and $(M)_{m,n}$ denotes its (m, n) 'th element; for any vector, v , $\|v\|$ denotes the usual Euclidean norm, $\|v\| := (v'v)^{1/2}$ and $(v)_m$ denotes its m 'th element; for any real number, x , $[x]$ denotes the integer part of x . A function $f(x) : \mathbb{R}^q \rightarrow \mathbb{R}$ satisfies a Lipschitz condition of order α , or is in $\text{Lip}(\alpha)$, 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 $x_1, x_2 \in \mathbb{R}^q$. Throughout, we use the notation c or K for a generic, finite constant, and, finally, as a convention, it is assumed that $j^{-1} = 0$ for $j = 0$ in summations over j .

2 The Heteroskedastic Fractional Model

We consider the fractional time series model

$$X_t = \Delta_+^{-d} u_t \text{ and } u_t = a(L, \psi) \varepsilon_t, \quad (1)$$

where the operator Δ_+^{-d} is given, for a generic variable x_t , by $\Delta_+^{-d} x_t := \Delta^{-d} x_t \mathbb{I}(t \geq 1) = \sum_{i=0}^{t-1} \pi_i(d) x_{t-i}$ with

$$\pi_i(d) := \frac{\Gamma(d+i)}{\Gamma(d)\Gamma(1+i)} = \frac{d(d+1)\dots(d+i-1)}{i!}$$

denoting the coefficients in the usual binomial expansion of $(1-z)^{-d}$, and where ψ is a p -dimensional parameter vector and $a(z, \psi) := \sum_{n=0}^{\infty} a_n(\psi) z^n$.

Remark 1. The parametric form (but not the parameters characterising it) of the function $a(z, \psi)$ will be assumed known, so that, specifically, u_t is assumed to be a linear process governed by an underlying p -dimensional parameter vector. For example, any process that can be written as a finite order ARMA model is permitted as is the exponential spectrum model of Bloomfield (1973). The latter is popular in the fractional literature owing to the relatively simple covariance matrix formula it offers in this setting; see, for example, Robinson (1994). Further discussion

on the function $a(z, \psi)$ can be found in Hualde and Robinson (2011) and also in Nielsen (2014) for the multivariate case.

Remark 2. The definition of fractional integration applied in (1) is that of 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, non-stationary, and over-differenced time series are permitted and that the range of admissible values of the fractional parameter can be arbitrarily large; see the discussion in Hualde and Robinson (2011).

Model (1) is analyzed under the following assumption on the innovations ε_t .

Assumption A. *The innovations $\{\varepsilon_t\}$ are such that $\varepsilon_t = \sigma_t z_t$, where $\{\sigma_t\}$ and $\{z_t\}$ satisfy the conditions in parts (a) and (b), respectively, below:*

- (a) $\{\sigma_t\}_{t \in \mathbb{Z}}$ is non-stochastic and uniformly bounded, and, for all $t = 1, \dots, T$ satisfies $\sigma_t := \sigma(t/T) > 0$, where $\sigma(\cdot) \in \mathcal{D}([0, 1])$, the space of càdlàg functions on $[0, 1]$.
- (b) $\{z_t\}_{t \in \mathbb{Z}}$ is a martingale difference sequence with respect to the natural filtration F_t , the sigma-field generated by $\{z_s\}_{s \leq t}$, such that $F_{t-1} \subseteq F_t$ for $t = \dots, -1, 0, 1, 2, \dots$, and satisfies
 - (i) $E(z_t^2) = 1$,
 - (ii) $\tau_{r,s} := E(z_t^2 z_{t-r} z_{t-s})$ is uniformly bounded for all $t \geq 1, r \geq 0, s \geq 0$, where also $\tau_{r,r} > 0$ for all $r \geq 0$,
 - (iii) For all integers $r_1, r_2 \geq 1$, the 4th order cumulants $\kappa_4(t, t, t-r_1, t-r_2)$ of $(z_t, z_t, z_{t-r_1}, z_{t-r_2})$ satisfy $\sup_t \sum_{r_1, r_2=1}^{\infty} |\kappa_4(t, t, t-r_1, t-r_2)| < \infty$.

Importantly, through the condition in Assumption A(b)(iii), only four moments are assumed finite in Assumption A. Moreover, Assumption A does not impose Gaussianity. Assumption A allows for both conditional and unconditional heteroskedasticity of very general forms. The conditions in part (a) of Assumption A, see Cavaliere and Taylor (2008), imply that the unconditional innovation variance σ_t^2 is only required to be bounded and to display at most a countable number of jumps, therefore allowing for an extremely wide class of potential models for the unconditional variance of ε_t . Models of single or multiple variance shifts satisfy part (a) of Assumption A with $\sigma(\cdot)$ piecewise constant. (Piecewise) affine functions are also permitted, thereby allowing for variances which follow a (broken) linear trend.

Part (b) of Assumption A allows for conditional heteroskedasticity in $\{z_t\}$. A parametric model, such as a member of the GARCH family, is not assumed. Rather, the conditions in Assumption A(b) allow for conditional heteroskedasticity of unknown and very general form and are typical of those used in this literature; see, for example, Robinson (1991), Demetrescu, Kuzin and Hassler (2008), Hassler, Rodrigues and Rubia (2009) and Kew and Harris (2009).

However, we note that the conditions given in part (b) of Assumption A are somewhat weaker than required by these authors. First of all, they impose the assumption that, for any integer q , $2 \leq q \leq 8$, and for q non-negative integers s_i , $E(\prod_{i=1}^q z_{t_i}^{s_i}) = 0$ when at least one s_i is exactly one and $\sum_{i=1}^q s_i \leq 8$, see, e.g., Assumption E(e) of Kew and Harris (2009). This implies, in particular, that $\tau_{r,s} = 0$ for $r \neq s$, which rules out a large class of asymmetric conditionally heteroskedastic processes. Secondly, these authors assume strict stationarity of z_t , which we do not.

A special case of Assumption A, which obtains for $\sigma(\cdot)$ constant and $\{z_t\}$ conditionally homoskedastic, is the following

Assumption H. *The innovations $\{\varepsilon_t\}$ form a martingale difference sequence with respect to the filtration F_t , where $E(\varepsilon_t^2 | F_{t-1}) = \sigma^2$ almost surely and $\sup_t E(\varepsilon_t^4) \leq K < \infty$.*

Remark 3. Assumption H is a fairly standard conditional homoskedasticity assumption used in the time series literature. It is, however, weaker than Assumption A2 in Hualde and Robinson (2011), which additionally imposes the conditions that ε_t is strictly stationary and ergodic and that the conditional third and fourth moments of ε_t are equal to the corresponding unconditional moments.

Remark 4. Observe that the assumption that z_t is a martingale difference sequence in Assumption A implies that for any $\kappa_q(\cdot)$, $q \geq 2$, if the highest argument in the cumulant appears only once, then the cumulant is zero. This result is stated and formally proved in Lemma A.2 in the appendix. For this reason, our stated assumptions deal only with cumulants where the first two (the highest) arguments coincide. Moreover, notice that the boundedness assumption in A(b)(ii) does in fact follow from the conditions imposed in A(b)(iii).

Remark 5. Since σ_t depends on (t/T) , a time series generated according to Assumption A formally constitutes a triangular array of the type $\{\varepsilon_{T,t} : 0 \leq t \leq T, T \geq 1\}$, where $\varepsilon_{T,t} = \sigma_{T,t} z_t$ and $\sigma_{T,t} = \sigma(t/T)$. Since the triangular array notation is not essential, for simplicity the subscript T is suppressed in the sequel.

We next impose the following assumption on the true parameter values, which are denoted by subscript zero.

Assumption B. *The true parameter values satisfy $(d_0, \psi_0) \in D \times \Psi$, where $D := [d_1, d_2]$ with $-\infty < d_1 \leq d_2 < \infty$ and the set $\Psi \subseteq \mathbb{R}^p$ is convex and compact.*

Assumption B permits the length of the interval D of admissible values of the parameter d to be arbitrarily large. Specifically, the length of D is not limited to less than 1/2 as in most previous studies of fractional time series models that include proofs of consistency. Thus, under Assumption B, the model can simultaneously accommodate both non-stationary, (asymptotically) stationary, and over-differenced processes.

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

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) $a_0(\psi) = 1$ and $a(z, \psi) = \sum_{n=0}^{\infty} a_n(\psi)z^n$ is bounded and bounded away from zero.

(ii) $a(e^{i\lambda}, \psi)$ is twice differentiable in λ with second derivative in $\text{Lip}(\zeta)$ for $\zeta > 0$.

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

Remark 6. Assumption C is the univariate version of Assumption C in Nielsen (2014). Assumption C(i) coincides with Assumption A1(iv) of Hualde and Robinson (2011), while Assumption C(ii) strengthens their Assumption A1(ii) from once differentiable in λ with derivative in $\text{Lip}(\zeta)$ for $\zeta > 1/2$, while Assumption C(iii) strengthens their Assumption A1(iii) from continuity in ψ to differentiability.

Remark 7. Assumptions C(i)-(ii) ensure that u_t in (1) is an invertible short-memory process (with spectrum that is bounded and bounded away from zero at all frequencies); under Assumption C(i) the function $b(z, \psi) := \sum_{n=0}^{\infty} b_n(\psi)z^n = a(z, \psi)^{-1}$ is well-defined by its power series expansion for $|z| \leq 1 + \delta$ for some $\delta > 0$, and is also bounded and bounded away from zero on the complex unit disk. The coefficients $a_n(\psi)$ and $b_n(\psi)$ then satisfy

$$|a_n(\psi)| = O(n^{-2-\zeta}) \text{ and } |b_n(\psi)| = O(n^{-2-\zeta}) \text{ uniformly in } \psi \in \Psi, \quad (2)$$

see Zygmund (2003, pp. 46 and 71). In contrast, under Hualde and Robinson's (2011) Assumption A1(ii) the rate is $O(n^{-1-\zeta})$ for $\zeta > 1/2$, which is slightly weaker. Assumption C is easily seen to be satisfied, for example, by stationary and invertible finite order ARMA processes due to the exponential decay of their linear representation coefficients.

Finally, the following identification condition will also be needed:

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

Assumption D is identical to Assumption D in Nielsen (2014) and Assumption A1(i) in Hualde and Robinson (2011). It is satisfied, for example, by all stationary and invertible finite order ARMA processes whose AR and MA polynomials are not both overspecified; that is, those which do not admit any common factors in their AR and MA polynomials. More generally, Assumption D ensures identification and is related to the standard condition for identification in extremum (or ML) estimation; see, for example, Hayashi (2000, equation (7.2.13)) for a textbook treatment. In a time series context, see also Hannan (1973, equation (4)).

3 Asymptotic Theory

Let $\theta := (d, \psi) \in D \times \Psi = \Theta$ and define the residuals

$$\varepsilon_t(\theta) := \sum_{n=0}^{t-1} b_n(\psi) \Delta_+^d X_{t-n}. \quad (3)$$

Then the conditional² Gaussian quasi-log-likelihood function, based on the assumption that the shocks u_t are Gaussian with constant variance equal to σ^2 , for the model (1) is, apart from a constant,

$$L(\theta, \sigma^2) = -\frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T \varepsilon_t(\theta)^2.$$

It follows in the usual way that the (conditional) QML estimator is identical to the classical least squares or CSS estimator, which is found by minimizing the sum of squared residuals; that is,

$$\hat{\theta} := \arg \min_{\theta \in \Theta} R(\theta), \quad (4)$$

$$R(\theta) := T^{-1} \sum_{t=1}^T \varepsilon_t(\theta)^2. \quad (5)$$

It is important to notice that Gaussianity is not needed for the asymptotic theory in this paper, and so (4) can be viewed as a (conditional) QML estimator.

In our first result we now establish the consistency of the QML estimator from (4) when the shocks, u_t , driving (1) are heteroskedastic, satisfying the conditions in Assumption A.

Theorem 1. *Let X_t be generated by model (1) satisfying Assumptions A–D, and let $(\hat{d}, \hat{\psi})$ be defined by (4). Then $(\hat{d}, \hat{\psi}) \xrightarrow{P} (d_0, \psi_0)$ as $T \rightarrow \infty$.*

Remark 8. The result in Theorem 1 establishes that the consistency result derived in Hualde and Robinson (2011) (see also Nielsen, 2014) under the assumption of conditionally homoskedastic errors remains valid under the conditions of Assumption A thereby allowing for conditional and/or unconditional heteroskedasticity in the driving shocks, u_t , in (1). This result does, however, require the stronger smoothness conditions in Assumption C relative to the corresponding conditions in Assumption A1 of Hualde and Robinson (2011); see again the discussion in Remarks 6 and 7. Notice that the result in Theorem 1 implies that this result also holds under Assumption H. Although this imposes somewhat weaker conditions than the corresponding Assumption A2 of Hualde and Robinson (2011), this must be traded off against the stronger conditions imposed by Assumption C relative to their Assumption A1.

²We are using the term ‘conditional’ here in its usual sense to indicate that we have conditioned on the initial values of u_t . This has, of course, been done implicitly through the assumption that (1) generates a type II fractional process; see again the discussion in Remark 2.

We now turn to establishing asymptotic distribution theory for the QML estimator from (4) when we allow for heteroskedasticity of the form given in Assumption A. In order to do so we will need to strengthen Assumptions A and C with the following additional assumptions.

Assumption E. For all integers q such that $3 \leq q \leq 8$ and for all integers $r_1, \dots, r_{q-2} \geq 1$, the q 'th order cumulants $\kappa_q(t, t, t-r_1, \dots, t-r_{q-2})$ of $(z_t, z_t, z_{t-r_1}, \dots, z_{t-r_{q-2}})$ satisfy the condition that $\sup_t \sum_{r_1, \dots, r_{q-2}=1}^{\infty} |\kappa_q(t, t, t-r_1, \dots, t-r_{q-2})| < \infty$.

Assumption F. For all z such that $|z| \leq 1$, $a(z, \psi) = \sum_{n=0}^{\infty} a_n(\psi)z^n$ 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 $\frac{\partial^k a_n(\psi)}{\partial \psi^{(k)}}$ satisfy $\sum_{n=0}^{\infty} |\frac{\partial^k a_n(\psi)}{\partial \psi^{(k)}}| < \infty$ for all $\psi \in \mathcal{N}_\delta(\psi_0)$ and $k = 2, 3$.

Remark 9. Notice that the moment condition $\sup_t E|z_t|^8 < \infty$, imposed by a number of other authors, is necessary for Assumption E with $q = 8$ to hold and therefore is not stated explicitly.

Remark 10. The strengthening of Assumption A(b)(iii) to a summability condition on the eighth cumulants of ε_t in Assumption E would appear to be standard, whether stated directly or indirectly, in the fractional literature where asymptotic distribution theory is derived under (conditional) heteroskedasticity, the leading example being the literature on hypothesis testing on the long memory parameter, d ; see, *inter alia*, Demetrescu, Kuzin and Hassler (2008), Hassler, Rodrigues and Rubia (2009) and Kew and Harris (2009). The additional moment conditions this strengthening implies are required in the proof of Theorem 2 to verify that, under heteroskedastic innovations of the form given in Assumption A, a Lindeberg-type condition holds for the score and for proving convergence in L_2 of the Hessian; see Appendices E.1 and E.2 respectively.

Remark 11. Assumption F requires $a(z, \psi)$ to be three times differentiable in ψ rather than the corresponding twice differentiable condition in Assumption A3(ii) of Hualde and Robinson (2011). This strengthening is required to prove tightness of the Hessian; see Appendix E.2. This condition is satisfied by standard stationary and invertible ARMA processes, and, indeed, is not restrictive in practice for the reasons outlined on page 3156 of Hualde and Robinson (2011).

Finally, to state the asymptotic variance of the limiting distribution of the QML estimator, we need to define the following matrices,

$$A_0 := \sum_{n,m=1}^{\infty} \tau_{n,m} \begin{bmatrix} n^{-1}m^{-1} & -\gamma_n(\psi_0)'/m \\ -\gamma_n(\psi_0)/m & \gamma_n(\psi_0)\gamma_m(\psi_0)' \end{bmatrix}$$

and

$$B_0 := \sum_{n=1}^{\infty} \begin{bmatrix} n^{-2} & -\gamma_n(\psi_0)'/n \\ -\gamma_n(\psi_0)/n & \gamma_n(\psi_0)\gamma_n(\psi_0)' \end{bmatrix},$$

where $\tau_{n,m}$ is defined in Assumption A(b)(ii), $\gamma_n(\psi) := \sum_{m=0}^{n-1} a_m(\psi) \dot{b}_{n-m}(\psi)$ and $\dot{b}_n(\psi) := \frac{\partial b_n(\psi)}{\partial \psi}$.

Remark 12. The matrix B_0 coincides with the matrix A in Hualde and Robinson (2011) and derives from the weak dependence present in the process through $a(z, \psi)$. On the other hand, the matrix A_0 also includes the effects of any conditional heteroskedasticity present in ε_t via the $\tau_{n,m}$ coefficients. If there is no conditional heteroskedasticity present, then $A_0 = B_0$ because here the only the $\tau_{n,n} = 1$ coefficients are non-zero. Notice that both A_0 and B_0 are not affected by any unconditional heteroskedasticity arising from part (a) of Assumption A.

Remark 13. Observe that A_0 is finite because, under Assumption C, $\gamma_n(\psi) \leq Kn^{-1-\zeta}$ such that $\|A_0\| \leq \sum_{n,m=1}^{\infty} |\tau_{n,m}| n^{-1} m^1 < \infty$ by Assumption A(b)(iii).

As in Hualde and Robinson (2011) we will require the matrix B_0 to be invertible. This is formally stated in Assumption G. Again this is easily satisfied by, for example, stationary and invertible ARMA process.

Assumption G. *The matrix B_0 is non-singular.*

In our second result we now establish asymptotic distribution theory for the QML estimator from (4) when the shocks, u_t , driving (1) are heteroskedastic, satisfying Assumption A.

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

$$\sqrt{T}((\hat{d}, \hat{\psi}) - (d_0, \psi_0)) \xrightarrow{D} N(0, C) \text{ as } T \rightarrow \infty, \quad (6)$$

where $C := \lambda B_0^{-1} A_0 B_0^{-1}$ with $\lambda := \int_0^1 \sigma^4(s) ds / (\int_0^1 \sigma^2(s) ds)^2$.

Remark 14. Theorem 2 generalises the corresponding result in Hualde and Robinson (2011) to the case where the shocks can display conditional and/or unconditional heteroskedasticity. Observe that, under Assumption H, $A_0 = B_0$ and $\lambda = 1$ and, hence, the result in (6) reduces to the result in Theorem 2.2 of Hualde and Robinson (2011). Where heteroskedasticity arises only through part (a) of Assumption A then so the variance matrix C in the right member of (6) reduces to λB_0^{-1} . Moreover, where heteroskedasticity arises only through part (b) of Assumption A then so C reduces to $B_0^{-1} A_0 B_0^{-1}$.

Remark 15. The result in (6) shows that the variance of the asymptotic distribution of the (scaled and centered) QML estimator depends on the scalar parameter λ . This parameter is a measure of the degree of unconditional heteroskedasticity (non-stationary volatility) present in $\{\varepsilon_t\}$. For a homoskedastic process, where $\sigma(\cdot)$ is constant, $\lambda = 1$, whereas when $\sigma(\cdot)$ is non-constant $\lambda > 1$, by the Cauchy-Schwarz inequality. Consequently the variance of the asymptotic distribution of the QML estimator is seen to be inflated when unconditional heteroskedasticity is present in $\{\varepsilon_t\}$.

Remark 16. The dependence of the asymptotic variance of the QML estimator on nuisance parameters arising from both the weak dependence and heteroskedasticity in $\{\varepsilon_t\}$ implies that asymptotically pivotal inference on the parameter vector $\theta = (d, \psi)$ will need to be based around a consistent estimator of C . An obvious candidate is the usual sandwich estimator

$$\hat{C} := \left(\left(\frac{\partial^2 L(\theta, \sigma^2)}{\partial \theta \partial \theta'} \right)^{-1} \left(T^{-1} \sum_{t=1}^T \frac{\partial L_t(\theta, \sigma^2)}{\partial \theta} \frac{\partial L_t(\theta, \sigma^2)}{\partial \theta'} \right) \left(\frac{\partial^2 L(\theta, \sigma^2)}{\partial \theta \partial \theta'} \right)^{-1} \right) \Bigg|_{\sigma^2 = \hat{\sigma}^2, \theta = \hat{\theta}}$$

with $L_t(\theta, \sigma^2) := -\frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \varepsilon_t(\theta)^2$ and $\hat{\sigma}^2 := \frac{1}{T} \sum_{t=1}^T \varepsilon_t(\theta)^2$. It is then straightforward to show that

$$\sqrt{T}((\hat{d}, \hat{\psi}) - (d_0, \psi_0))'(\hat{C})^{-1} \sqrt{T}((\hat{d}, \hat{\psi}) - (d_0, \psi_0)) \xrightarrow{D} \chi_{p+1}^2 \text{ as } T \rightarrow \infty,$$

where χ_k^2 denotes a Chi-squared distribution with k degrees of freedom. As an illustration of this general result, a test for the null hypothesis $H_0 : d = \bar{d}$ could then be based on the Wald-type statistic $W_T := T(\hat{d} - \bar{d})^2(\hat{C}^{-1})_{11}$, which under H_0 will have a χ_1^2 limiting distribution. An alternative testing approach, based on the application of the wild bootstrap method to the Lagrange Multiplier test for H_0 , is considered in Cavaliere, Nielsen and Taylor (2013).

4 Concluding remarks

In this paper we have shown that the consistency of QML estimators from parametric fractional time series models driven by conditionally homoskedastic shocks, obtained in Hualde and Robinson (2011), continue to hold under a wide class of conditionally and/or unconditionally heteroskedastic shocks. We have also shown that the QML estimator is asymptotically normal, the covariance matrix of which is dependent on nuisance parameters deriving from both the weak dependence (as in the corresponding result in Hualde and Robinson, 2011) and any heteroskedasticity present in the shocks. However, asymptotically pivotal inference on the parameters of the model can be obtained, provided a standard sandwich estimator of the variance is employed. Like the results in Hualde and Robinson (2011), a fundamental aspect of our results is that they apply over an arbitrarily large set of admissible parameter values for the (unknown) memory parameter covering both stationary and non-stationary processes and invertible and non-invertible processes.

We conclude by mentioning an interesting topic for further research. As noted in the Introduction, an alternative to parametric QML estimation is provided by semi-parametric estimation of the long memory parameter; see, in particular, Shimotsu and Phillips (2005) and Abadir *et al.* (2007). The former establishes consistency under the assumption of conditionally homoskedastic shocks, while the latter allows for some quite mild heterogeneity, such that a weak law of large numbers condition holds; see Equation (2.13) on page 1357 of Abadir *et al.* (2007) and the discussion of this they subsequently provide on pages 1358-1359. It would be useful to investigate if, as we have shown here to hold for the QML estimator, the consistency

results presented by these authors continue to hold under the general heteroskedastic conditions given in Assumption A of this paper.

Appendix A Auxiliary Lemmas

We first give a result of McLeish (1974) that will be used repeatedly.

Lemma A.1. *Let U_{Tt} be a martingale difference array with respect to some filtration F_t such that $F_{t-1} \subseteq F_t$ for $t = \dots, -1, 0, 1, 2, \dots$, and suppose, as $T \rightarrow \infty$, that*

$$(i) \sum_{t=1}^T E(U_{Tt}^2 1\{|U_{Tt}| > \delta\}) \rightarrow 0 \text{ for all } \delta > 0,$$

$$(ii) \sum_{t=1}^T U_{Tt}^2 \xrightarrow{P} V.$$

Then $\sum_{t=1}^T U_{Tt} \xrightarrow{D} N(0, V)$ as $T \rightarrow \infty$.

Proof. See Theorem 2.3 of McLeish (1974) and the comments in the two paragraphs following it. □

The next lemma derives an important consequence of the martingale difference property of z_t on the higher-order moments and cumulants of z_t . For the special case with $q = 2$ we obtain the well-known result that a martingale difference sequence is uncorrelated.

Lemma A.2. *Let z_t be a martingale difference sequence with respect to the natural filtration \mathcal{F}_t , the sigma-field generated by $\{z_s\}_{s \leq t}$, and suppose $E|z_t|^q < \infty$ for some integer $q \geq 2$. Then the q 'th order moments and cumulants satisfy*

$$E(z_t z_{t-r_1} \cdots z_{t-r_{q-1}}) = 0 \text{ and } \kappa_q(t, t-r_1, \dots, t-r_{q-1}) = 0$$

for all integers $r_k \geq 1, k = 1, \dots, q-1$.

Proof. The result for moments follows from the law of iterated expectations because

$$E(z_t z_{t-r_1} \cdots z_{t-r_{q-1}}) = E(E(z_t | \mathcal{F}_{t-1}) z_{t-r_1} \cdots z_{t-r_{q-1}}) = 0$$

by the martingale difference property of z_t . To show the result for cumulants, we start with $q = 2$. Then $\kappa_2(t, t-r) = E(z_t z_{t-r}) = 0$ because $r \geq 1$. When $q = 3$, $\kappa_3(t, t-r_1, t-r_2) = E(z_t z_{t-r_1} z_{t-r_2}) = 0$ by the result for moments. For $q = 4$ we find $\kappa_4(t, t-r_1, t-r_2, t-r_3) = E(z_t z_{t-r_1} z_{t-r_2} z_{t-r_3}) - E(z_t z_{t-r_1})E(z_{t-r_2} z_{t-r_3}) - E(z_t z_{t-r_2})E(z_{t-r_1} z_{t-r_3}) - E(z_t z_{t-r_3})E(z_{t-r_2} z_{t-r_1})$. Again, because $r_k \geq 1$ for $k = 1, 2, 3$, the cumulant is zero by the result for the second and fourth moments. For $q = 5$ we have $\kappa_5(t, t-r_1, \dots, t-r_4)$ for $r_k \geq 1$ and find that it contains the fifth moment, which is zero by the result for moments, and it contains ten products of pairs and corresponding triplets. In each of these there will be either a pair with $E(z_t z_{t-r_k}) = 0$ or there will be a triplet with $E(z_t z_{t-r_k} z_{t-r_j}) = 0$ as above. The same argument also applies to the higher-order cumulants and moments. □

Lemma A.3. Let σ_t and z_t satisfy Assumption A. Let $\xi_{0,n}$, $n \geq 1$, be vector-valued coefficients that satisfy $\|\xi_{0,n}\| \leq Kn^{-1}$, $K < \infty$, uniformly in $n \geq 1$, and let $g_{t,n,m}$, $t, n, m \geq 1$, be real coefficients satisfying $\sum_{n,m=1}^{\infty} n^{-1}m^{-1} \sup_t |g_{t,n,m}| < \infty$. Then

$$T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} \sigma_{t-n} \sigma_{t-m} g_{t,n,m} = T^{-1} \sum_{t=1}^T \sigma_t^4 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} g_{t,n,m} + o(1).$$

Proof. First notice that

$$\begin{aligned} & \left\| T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} \sigma_{t-n} \sigma_{t-m} g_{t,n,m} - T^{-1} \sum_{t=1}^T \sigma_t^4 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} g_{t,n,m} \right\| \\ &= \left\| T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} (\sigma_{t-n} \sigma_{t-m} - \sigma_t^2) g_{t,n,m} \right\| \\ &\leq KT^{-1} \sum_{t=1}^T \sum_{n=1}^{t-1} \sum_{m=n}^{t-1} n^{-1}m^{-1} |\sigma_{t-n} \sigma_{t-m} - \sigma_t^2| |g_{t,n,m}| = K(Q_{1T} + Q_{2T}), \end{aligned}$$

where the inequality follows because $\|\xi_{0,n}\| \leq Kn^{-1}$, by Assumption A(a) and by symmetry in n and m , and where we defined

$$\begin{aligned} Q_{1T} &:= \sum_{n=1}^{q_T} \sum_{m=n}^{q_T} n^{-1}m^{-1} \sup_t |g_{t,n,m}| T^{-1} \sum_{t=m+1}^T |\sigma_{t-n} \sigma_{t-m} - \sigma_t^2|, \\ Q_{2T} &:= \sum_{n=1}^{T-1} \sum_{m=\max(n, q_T+1)}^{T-1} n^{-1}m^{-1} \sup_t |g_{t,n,m}| T^{-1} \sum_{t=m+1}^T |\sigma_{t-n} \sigma_{t-m} - \sigma_t^2|. \end{aligned}$$

Let $q_T := \lfloor T^\varkappa \rfloor$ for $\varkappa \in (0, 1)$ and $M := \sup_{u \in [0,1]} \sigma(u)$, which is finite because $\sigma(u) \in \mathcal{D}([0, 1])$.

Then

$$|\sigma_{t-n} \sigma_{t-m} - \sigma_t^2| \leq \sigma_t |\sigma_{t-n} - \sigma_t| + \sigma_{t-n} |\sigma_{t-m} - \sigma_t| \leq M (|\sigma_{t-n} - \sigma_t| + |\sigma_{t-m} - \sigma_t|)$$

such that, for $m \geq n \geq 1$,

$$\sum_{t=m+1}^T |\sigma_{t-n} \sigma_{t-m} - \sigma_t^2| \leq M \sum_{t=m+1}^T (|\sigma_{t-n} - \sigma_t| + |\sigma_{t-m} - \sigma_t|) \leq 2M \sum_{t=m+1}^T |\sigma_{t-m} - \sigma_t|.$$

Hence, using the fact that $\sigma_t = \sigma(t/T) \in \mathcal{D}([0, 1])$,

$$\sup_{n,m=1,\dots,q_T} T^{-1} \sum_{t=m+1}^T |\sigma_{t-n} \sigma_{t-m} - \sigma_t^2| \leq 2M \sup_{m=1,\dots,q_T} T^{-1} \sum_{t=m+1}^T |\sigma_{t-m} - \sigma_t| \rightarrow 0 \text{ as } T \rightarrow \infty \quad (7)$$

by Lemma A.1 in Cavaliere and Taylor (2009).

The convergence in (7) allows us to show that Q_{1T} converges to zero as T diverges. Note that

$$Q_{1T} \leq \left(\sup_{n,m=1,\dots,q_T} T^{-1} \sum_{t=m+1}^T |\sigma_{t-n} \sigma_{t-m} - \sigma_t^2| \right) Q_{11T}$$

with $Q_{11T} := \sup_t \sum_{n=1}^{q_T} \sum_{m=n}^{q_T} n^{-1} m^{-1} |g_{t,n,m}| < \infty$. Because the first factor in Q_{1T} converges to zero as $T \rightarrow \infty$ by (7), it follows that $Q_{1T} \rightarrow 0$ as $T \rightarrow \infty$.

The term Q_{2T} is bounded as, by another application of Assumption A(a),

$$\begin{aligned} Q_{2T} &\leq 4M^2 \sum_{n=1}^{T-1} \sum_{m=\max(n, q_T+1)}^{T-1} n^{-1} m^{-1} \sup_t |g_{t,n,m}| \\ &\leq 4M^2 \sum_{m=q_T+1}^{\infty} \sum_{n=1}^{\infty} n^{-1} m^{-1} \sup_t |g_{t,n,m}| \rightarrow 0 \end{aligned}$$

as $T \rightarrow \infty$ because it is a tail sum ($q_T \rightarrow \infty$) of the convergent sum $\sum_{n,m=1}^{\infty} n^{-1} m^{-1} \sup_t |g_{t,n,m}|$. This completes the proof. \square

Appendix B Inequalities

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

Lemma B.1. *Uniformly in $-u_0 \leq v \leq u \leq u_0$ and for $j \geq 1, m \geq 0$ it holds that*

$$\left| \frac{\partial^m}{\partial u^m} \pi_j(u) \right| \leq c(1 + \log j)^m j^{u-1}, \quad (8)$$

$$\left| \frac{\partial^m}{\partial u^m} \pi_j(u) - \frac{\partial^m}{\partial v^m} \pi_j(v) \right| \leq c(u-v)(1 + \log j)^{m+1} j^{u-1}, \quad (9)$$

$$\left| \frac{\partial^m}{\partial u^m} \pi_{j+1}(u) - \frac{\partial^m}{\partial u^m} \pi_j(u) \right| \leq c(1 + \log j)^m j^{u-2}, \quad (10)$$

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

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

$$\pi_j(-v) \geq c j^{-v-1}, \quad (11)$$

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

Proof. See Lemma A.1 of Nielsen (2014). \square

Lemma B.2. *Let u and v be such that $\max(|u|, |v|) \leq a$ for some $a < \infty$. Then it holds that*

$$\sum_{j=1}^{t-1} j^{u-1} (t-j)^{v-1} \leq c(1 + \log t) t^{\max(u+v-1, u-1, v-1)},$$

where the constant $c > 0$ does not depend on u, v , or t .

Proof. See Lemma B.4 of Johansen and Nielsen (2010). \square

Lemma B.3. *Let $\xi_T(u, v, k) := \max_{1 \leq n, m \leq T} \sum_{t=\max(n,m)}^T |\zeta_{t-n}(-u, k) \zeta_{t-m}(-v, k)|$ for coefficients $\zeta_j(u, k)$ satisfying $\zeta_0(u, k) = 1$ and $\zeta_j(u, k) \leq c(\log j)^k j^{u-1}$ for $j \geq 1$, where $c > 0$ does not depend on u, k , or j . Then:*

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

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

where the constant $c > 0$ does not depend on u, v , or T .

(ii) For any $u > 0, v > 0$ it holds that

$$\sum_{t=0}^{\infty} |\zeta_{|t-n|}(-u, k)\zeta_t(-v, k)| \leq c(\log |n|)^k |n|^{\max(-u-1, -v-1)},$$

where the constant $c > 0$ does not depend on u, v , or n .

Proof. Part (i): See Lemma A.7 of Johansen and Nielsen (2012).

Part (ii): If $n \geq 0$ we split the summation and find the bound

$$\begin{aligned} & \sum_{t=0}^{\lfloor n/2 \rfloor} |\zeta_{|t-n|}(-u, k)\zeta_t(-v, k)| + \sum_{t=\lfloor n/2 \rfloor+1}^n |\zeta_{|t-n|}(-u, k)\zeta_t(-v, k)| + \sum_{t=n+1}^{\infty} |\zeta_{t-n}(-u, k)\zeta_t(-v, k)| \\ & \leq c \sum_{t=0}^{\lfloor n/2 \rfloor} (n-t)^{-u-1}(\log(n-t))^k t^{-v-1}(\log t)^k + c \sum_{t=\lfloor n/2 \rfloor+1}^n (n-t)^{-u-1}(\log(n-t))^k t^{-v-1}(\log t)^k \\ & \quad + \sum_{t=n+1}^{\infty} (t-n)^{-u-1}(\log(t-n))^k t^{-v-1}(\log t)^k \\ & \leq c(n/2)^{-u-1}(\log(n/2))^k \sum_{t=0}^{\lfloor n/2 \rfloor} t^{-v-1}(\log t)^k + c(n/2)^{-v-1}(\log(n/2))^k \sum_{t=\lfloor n/2 \rfloor+1}^n (n-t)^{-u-1}(\log(n-t))^k \\ & \quad + c(n+1)^{-v-1}(\log(n+1))^k \sum_{t=n+1}^{\infty} (t-n)^{-u-1}(\log(t-n))^k \\ & \leq c(\log n)^k n^{\max(-u-1, -v-1)}. \end{aligned}$$

If $n < 0$ we find the bound

$$\begin{aligned} & c \sum_{t=0}^{\infty} (t-n)^{-u-1}(\log(t-n))^k t^{-v-1}(\log t)^k \\ & \leq c(-n)^{-u-1}(\log(-n))^k \sum_{t=0}^{\infty} t^{-v-1}(\log t)^k \leq c|n|^{-u-1}(\log |n|)^k. \end{aligned}$$

□

Lemma B.4. Let $F_N(u) := \sum_{n=0}^{N-1} \pi_n(-u)^2$. For $N \geq 2, u \leq -1/2 + a$, and $a > 0$,

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

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

Proof. See Lemma A.3 of Nielsen (2014).

□

Appendix C Variation bounds

This section contains three lemmas that are used to verify tightness and stochastic equicontinuity conditions for the processes in the previous sections. The first lemma deals with non-stationary processes and the next lemma with product moments of processes that are nearly stationary. Lemma C.2 contains the truncation argument used to deal with the non-uniform convergence in Θ_2 ; see Appendix D.2. The final lemma 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 critical processes – and it is applied for the Hessian in the proof of asymptotic normality.

Lemma C.1. *Let ε_t satisfy Assumption A. Then, uniformly in $v_0 \leq v \leq u \leq u_0 < -1/2$,*

$$\|T^{u+1/2}\Delta_+^u\varepsilon_t\|_2 \leq c \text{ and } \|T^{u+1/2}\Delta_+^u\varepsilon_t - T^{v+1/2}\Delta_+^v\varepsilon_t\|_2 \leq c|u - v| \quad (12)$$

where the constant $c > 0$ does not depend on u, v , or T .

Proof. See Lemma C.3 in Johansen and Nielsen (2010), which also applies under Assumption A on ε_t in place of their i.i.d. assumption. \square

Lemma C.2. *Let $w_{1t} = w_{1t}(u) := \sum_{n=0}^{N-1} \pi_n(-u)\varepsilon_{t-n}$ and $w_{2t} = w_{2t}(u) := \sum_{n=N}^{t-1} \pi_n(-u)\varepsilon_{t-n}$, where ε_t satisfies Assumption A, and define the product moments $Q_{11NT}(u) := T^{-1} \sum_{t=N+1}^T w_{1t}^2 - E(T^{-1} \sum_{t=N+1}^T w_{1t}^2)$ and $Q_{12NT}(u) := T^{-1} \sum_{t=N+1}^T w_{1t}w_{2t}$. Then, for any $\kappa \in (0, 1/2)$ it holds that, uniformly in $-1/2 - \kappa \leq v \leq u \leq -1/2 + \kappa$,*

$$\|Q_{12NT}(u)\|_2 \leq c(\log T)T^{-1/2+\kappa}N^{1/2+\kappa}, \quad (13)$$

$$\|Q_{12NT}(u) - Q_{12NT}(v)\|_2 \leq c|u - v|(\log T)^2T^{-1/2+\kappa}N^{1/2+\kappa}, \quad (14)$$

$$\|Q_{11NT}(u)\|_2 \leq c(\log T)T^{-1/2}N^{1/2+2\kappa}, \quad (15)$$

$$\|Q_{11NT}(u) - Q_{11NT}(v)\|_2 \leq c|u - v|(\log T)^2T^{-1/2}N^{1/2+2\kappa}, \quad (16)$$

where the constant $c > 0$ does not depend on u, v , or T .

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

$$\sup_{|u+1/2| \leq \kappa} |Q_{11NT}(u)| \xrightarrow{P} 0 \text{ and } \sup_{|u+1/2| \leq \kappa} |Q_{12NT}(u)| \xrightarrow{P} 0. \quad (17)$$

Proof. Eqn. (13): First evaluate

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

The term $E(\prod_{k=1}^2 \varepsilon_{t_k-n_k}\varepsilon_{t_k-m_k})$ is non-zero only if the two highest subscripts are equal, see Lemma A.2. However, $n_k < N \leq m_k$ such that $t_k - n_k > t_k - m_k$ for $k = 1, 2$. This leaves only

one possibility, i.e., $t_1 - n_1 = t_2 - n_2$, in which case we eliminate $n_2 = t_2 - t_1 + n_1$ and note that $|t_1 - t_2| = |n_1 - n_2| \leq N$. In this case $EQ_{12NT}(u)^2$ is

$$T^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \sum_{n_1=0}^{N-1} \sum_{m_1=N}^{t_1-1} \sum_{m_2=\max(N, t_2-t_1+n_1)}^{t_2-1} \pi_{n_1}(-u) \pi_{t_2-t_1+n_1}(-u) \pi_{m_1}(-u) \pi_{m_2}(-u) \\ \times \sigma_{t_1-n_1}^2 \sigma_{t_1-m_1} \sigma_{t_2-m_2} E(z_{t_1-n_1}^2 z_{t_1-m_1} z_{t_2-m_2}). \quad (18)$$

If, in this expression, $t_1 - m_1 = t_2 - m_2$ we eliminate $m_2 = t_2 - t_1 + m_1$ and the expectation is $\tau_{m_1-n_1, m_1-n_1}$. Then, with $\xi_T(u, v, k)$ defined in Lemma B.3, $\sum_{n_1=0}^{N-1} \pi_{n_1}(-u) \pi_{t_2-t_1+n_1}(-u) \leq \xi_N(u, u, 0)$ and $\sum_{m_1=N}^{t_1-1} \pi_{m_1}(-u) \pi_{t_2-t_1+m_1}(-u) \leq \xi_T(u, u, 0)$ by (8) of Lemma B.1, so the contribution to $EQ_{12NT}(u)^2$ is bounded by

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

The result when $t_1 - m_1 = t_2 - m_2$ now follows from Lemma B.3(i). If, on the other hand, $t_1 - m_1 \neq t_2 - m_2$ in (18), the expectation in (18) is $\kappa_4(t_1 - n_1, t_1 - n_1, t_1 - m_1, t_2 - m_2)$ and the contribution to $EQ_{12NT}(u)^2$ is bounded by

$$cT^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \sum_{n_1=0}^{N-1} \pi_{n_1}(-u) \pi_{t_2-t_1+n_1}(-u) \pi_N(-u)^2 \\ \times \sum_{m_1=N}^{t_1-1} \sum_{m_2=\max(N, t_2-t_1+n_1)}^{t_2-1} |\kappa_4(t_1 - n_1, t_1 - n_1, t_1 - m_1, t_2 - m_2)| \\ \leq cT^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \xi_N(u, u, 0) N^{-2u-2}$$

using Assumption A(a),(b)(iii), and this proves the result.

Eqn. (14): Next consider $\|Q_{12NT}(u) - Q_{12NT}(v)\|_2$ which is bounded by

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

For the first term write $w_{1t}(u) - w_{1t}(v) = \sum_{n=0}^{N-1} (\pi_n(-u) - \pi_n(-v)) \varepsilon_{t-n} = (u-v) \sum_{n=0}^{N-1} \zeta_n(-u, 1) \varepsilon_{t-n}$, see (9) of Lemma B.1 and Lemma B.3. Now apply the same proof as for (13), noting that only a log-factor is added. The same proof can be used for the second term.

Eqn. (15): Note that

$$E(T^{-1} \sum_{t=N+1}^T w_{1t}^2) = T^{-1} \sum_{t=N+1}^T \sum_{n_1, n_2=0}^{N-1} \pi_{n_1}(-u) \pi_{n_2}(-u) E(\varepsilon_{t-n_1} \varepsilon_{t-n_2}) \\ = T^{-1} \sum_{t=N+1}^T \sum_{n=0}^{N-1} \pi_n(-u)^2 \sigma_{t-n}^2$$

such that the second moment of $Q_{11NT}(u)$ is

$$EQ_{11NT}(u)^2 = E(T^{-1} \sum_{t=N+1}^T w_{1t}^2)^2 - T^{-2} \sum_{t,s=N+1}^T \sum_{n,m=0}^{N-1} \pi_n(-u)^2 \pi_m(-u)^2 \sigma_{t-n}^2 \sigma_{s-m}^2. \quad (19)$$

Now,

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

where again the two highest subscripts in $\prod_{k=1}^2 \varepsilon_{t_k-n_k} \varepsilon_{t_k-m_k}$ have to be equal by Lemma A.2. By symmetry, there are three cases, which we now enumerate.

Case 1) Suppose first that $t_1 - n_1 = t_1 - m_1$, i.e. $n_1 = m_1$. If also $t_2 - n_2 = t_2 - m_2$ the contribution is $T^{-2} \prod_{k=1}^2 \sum_{t_k=N+1}^T \sum_{n_k=0}^{N-1} \pi_{n_k}(-u)^2 \sigma_{t_k-n_k}^2$, which cancels with the second term of (19). If $t_2 - n_2 \neq t_2 - m_2$, then both these terms have to be no greater than $t_1 - n_1$ by Lemma A.2, so that $t_2 \leq t_1 - n_1 + n_2$ and $m_2 \geq t_2 - t_1 + n_1$. In this case the contribution is

$$\begin{aligned} & T^{-2} \sum_{t_1=N+1}^T \sum_{n_1, n_2=0}^{N-1} \sum_{t_2=N+1}^{\max(T, t_1-n_1+n_2)} \sum_{m_2=\max(0, t_2-t_1+n_1)}^{N-1} \pi_{n_1}(-u)^2 \pi_{n_2}(-u) \pi_{m_2}(-u) \\ & \quad \times \sigma_{t_1-n_1}^2 \sigma_{t_2-n_2} \sigma_{t_2-m_2} \kappa_4(t_1 - n_1, t_1 - n_1, t_2 - n_2, t_2 - m_2) \\ & \leq cT^{-2} \sum_{t_1=N+1}^T \sum_{n_1, n_2=0}^{N-1} \pi_{n_1}(-u)^2 \pi_{n_2}(-u) \\ & \leq cT^{-1} \sum_{n_1, n_2=0}^{N-1} n_1^{-2u-2} n_2^{-u-1} \leq cT^{-1} N^{1/2+3\kappa}, \end{aligned}$$

where the first inequality is by Assumption A(a),(b)(iii).

Case 2) If $t_1 - n_1 = t_2 - n_2 \geq t_k - m_k$ the restriction $|t_1 - t_2| = |n_1 - n_2| \leq N$ is implied such that the contribution is

$$\begin{aligned} & T^{-2} \sum_{\substack{t_1, t_2=N+1 \\ |t_1-t_2| \leq N}}^T \sum_{n_1=\max(0, t_1-t_2)}^{N-1} \sum_{m_1=n_1}^{N-1} \sum_{m_2=\max(0, t_2-t_1+n_1)}^{N-1} \pi_{n_1}(-u) \pi_{t_2-t_1+n_1}(-u) \pi_{m_1}(-u) \pi_{m_2}(-u) \\ & \quad \times \sigma_{t_1-n_1}^2 \sigma_{t_1-m_1} \sigma_{t_2-m_2} E(z_{t_1-n_1}^2 z_{t_1-m_1} z_{t_2-m_2}). \end{aligned}$$

If also $t_1 - m_1 = t_2 - m_2$, the expectation is $\tau_{m_1-n_1, m_1-n_1}$ and contribution is bounded by

$$\begin{aligned} & cT^{-2} \sum_{\substack{t_1, t_2=N+1 \\ |t_1-t_2| \leq N}}^T \left(\sum_{n=0}^{N-1} \pi_n(-u) \pi_{t_2-t_1+n}(-u) \right)^2 \\ & \leq cT^{-2} \sum_{\substack{t_1, t_2=N+1 \\ |t_1-t_2| \leq N}}^T \xi_N(u, u, 0)^2 \leq cT^{-1} N \xi_N(u, u, 0)^2 \leq c(\log T)^2 T^{-1} N^{1+4\kappa} \end{aligned}$$

by Assumption A(a) and Lemma B.3(i). If instead $t_1 - m_1 \neq t_2 - m_2$, the expectation is $\kappa_4(t_1 - n_1, t_1 - n_1, t_1 - m_1, t_2 - m_2)$ and the bound is

$$\begin{aligned} & cT^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \sum_{\substack{n_1 = \max(0, t_1 - t_2) \\ N-1}}^{N-1} \pi_{n_1}(-u)^2 \pi_{t_2 - t_1 + n_1}(-u)^2 \\ & \times \sum_{m_1 = n_1}^{N-1} \sum_{m_2 = \max(0, t_2 - t_1 + n_1)}^{N-1} |\kappa_4(t_1 - n_1, t_1 - n_1, t_1 - m_1, t_2 - m_2)| \\ & \leq cT^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \sum_{\substack{n_1 = \max(0, t_1 - t_2) \\ N-1}}^{N-1} \pi_{n_1}(-u)^2 \pi_{t_2 - t_1 + n_1}(-u)^2 \leq cT^{-1}N. \end{aligned}$$

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

$$\begin{aligned} & T^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \sum_{n_1=0}^{N-1} \sum_{m_1=0}^{N-1} \pi_{n_1}(-u) \pi_{t_2 - t_1 + m_1}(-u) \pi_{m_1}(-u) \pi_{t_2 - t_1 + n_1}(-u) \sigma_{t_1 - n_1}^2 \sigma_{t_1 - m_1}^2 \tau_{m_1 - n_1, m_1 - n_1} \\ & \leq c(\log T)^2 T^{-1} N^{1+4\kappa} \end{aligned}$$

and if $t_1 - n_1 = t_2 - m_2$ and $t_1 - m_1 \neq t_2 - n_2$ (both no greater than $t_1 - n_1$ by Lemma A.2) the contribution is

$$\begin{aligned} & T^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \sum_{n_1=0}^{N-1} \sum_{m_1 = n_1}^{N-1} \sum_{n_2 = \max(0, t_2 - t_1 + n_1)}^{N-1} \pi_{n_1}(-u) \pi_{m_1}(-u) \pi_{n_2}(-u) \pi_{t_2 - t_1 + n_1}(-u) \\ & \times \sigma_{t_1 - n_1}^2 \sigma_{t_1 - m_1} \sigma_{t_2 - n_2} \kappa_4(t_1 - n_1, t_1 - n_1, t_1 - m_1, t_2 - n_2) \\ & \leq cT^{-2} \sum_{\substack{t_1, t_2 = N+1 \\ |t_1 - t_2| \leq N}}^T \sum_{n_1=0}^{N-1} \pi_{n_1}(-u)^2 \pi_{t_2 - t_1 + n_1}(-u)^2 \leq cT^{-1}N \end{aligned}$$

in the same way as in case 2).

Eqn. (16): Apply the same decomposition as in the proof of (14) and then use the same proof as for (15) with an extra log-factor.

Eqn. (17): Using the condition on α , the right-hand sides of (13)–(16) all converge to zero. Pointwise convergence in probability then follows from (13) and (15) and tightness on the interval $|u + 1/2| \leq \kappa$ follows from (14) and (16) using the criterion (51). Together this implies uniform convergence in probability. \square

Lemma C.3. Let $Z_{it} := \sum_{n=0}^{\infty} \zeta_{in}(\psi) \varepsilon_{t-n}$, $i = 1, 2$, where ε_t satisfies Assumption A and the 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 $\tilde{\Theta} := \{(u_1, u_2, \psi) \in D \times D \times \tilde{\Psi} : \min(u_1 + 1, u_2 + 1, u_1 + u_2 + 1) \geq a\}$ for $a > 0$. Then

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

Proof. The proof is given only for $k, l = 0$ since the derivatives just add a log-factor, see (8), which does not change the proof. Rearranging the summations the product moment $Q_T(u_1, u_2, \psi)$ is

$$\begin{aligned} & T^{-1} \sum_{j,k=0}^{T-1} \pi_j(-u_1) \pi_k(-u_2) \sum_{n,m=0}^{\infty} \zeta_{1n}(\psi) \zeta_{2m}(\psi) \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m} \\ &= T^{-1} \sum_{j=0}^{T-1} \pi_j(-u_1) \sum_{n=0}^{\infty} \sum_{m=\max(0, j+n-T+1)}^{j+n} \zeta_{1n}(\psi) \zeta_{2m}(\psi) \pi_{j+n-m}(-u_2) \sum_{t=\max(j, j+n-m)+1}^T \varepsilon_{t-j-n}^2 \end{aligned} \quad (20)$$

$$+ 2T^{-1} \sum_{j=0}^{T-1} \pi_j(-u_1) \sum_{n,m=0}^{\infty} \zeta_{1n}(\psi) \zeta_{2m}(\psi) \sum_{k=0}^{\min(T, j+n-m)-1} \pi_k(-u_2) \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m}. \quad (21)$$

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 $\sup_{(u_1, u_2, \psi) \in \tilde{\Theta}} |(20)|$ is

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

If $a > 0$ the summation over j is bounded and then $\sup_{(u_1, u_2, \psi) \in \tilde{\Theta}} |(20)| = O_P(1)$ 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 (20).

Next, summation by parts yields

$$\begin{aligned} & \sum_{k=0}^{\min(T, j+n-m)-1} \pi_k(-u_2) \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m} \\ &= \pi_{j+n-m-1}(-u_2) \sum_{k=0}^{\min(T, j+n-m)-1} \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m} \\ &\quad - \sum_{l=0}^{\min(T, j+n-m)-2} (\pi_{l+1}(-u_2) - \pi_l(-u_2)) \sum_{k=0}^l \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m}, \end{aligned} \quad (22)$$

where

$$\sum_{k=0}^l \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m} = \sum_{s=\max(1-m, 1+j-l-m)}^{T-m} w_s,$$

with $w_s := \varepsilon_s \sum_{k=j+n-m-l}^{j+n-m} \varepsilon_{s-k}$ being an uncorrelated sequence that satisfies

$$E(w_s^2) = \sigma_s^2 E \left(\sum_{k=j+n-m-l}^{j+n-m} \varepsilon_{s-k} \right)^2 = O(l),$$

such that

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

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

Now, rearranging the summations and applying the summation by parts result, $\sup_{(u_1, u_2, \psi) \in \tilde{\Theta}} |(21)|$ is

$$O_P \left(\sup_{(u_1, u_2, \psi) \in \tilde{\Theta}} \sum_{n, m=0}^{\infty} |\zeta_{1n}(\psi)| |\zeta_{2m}(\psi)| \sum_{j=\max(0, 1+m-n)}^{\min(T-1, T+m-n)} |\pi_j(-u_1)| |\pi_{j+n-m-1}(-u_2)| \right) \quad (23)$$

$$+ O_P \left(\sup_{(u_1, u_2, \psi) \in \tilde{\Theta}} T^{-1} \sum_{j=0}^{T-1} |\pi_j(-u_1)| \sum_{n, m=0}^{\infty} |\zeta_{1n}(\psi)| |\zeta_{2m}(\psi)| \sum_{l=0}^{T-2} |\pi_{l+1}(-u_2) - \pi_l(-u_2)| (T+l-j)^{1/2} l^{1/2} \right). \quad (24)$$

The result for (23) follows as in the analysis of (20). For term (24) it holds, using (10) 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 B.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 $c(\log T)^2 T^{-1/2} T^{\max(0, 1/2-a)} = c(\log T)^2 T^{\max(-1/2, -a)}$ if $a > 0$ and $c(\log T) T^{-1/2} T^{1/2-a} = c(\log T) T^{-a}$ if $a \leq 0$. \square

Appendix D Proof of Theorem 1

The residual in (3) is given by $\varepsilon_t(\theta) = \sum_{n=0}^{t-1} b_n(\psi) \Delta_+^{d-d_0} u_{t-n}$, and clearly the convergence properties of $R(\theta)$ in (5) depend on $d - d_0$. Define the untruncated processes

$$e_t(\psi) := c(L, \psi) \varepsilon_t = \sum_{n=0}^{\infty} c_n(\psi) \varepsilon_{t-n}, \quad (25)$$

$$\eta_t(\theta) := \Delta^{d-d_0} e_t(\psi) = \sum_{n=0}^{\infty} \varphi_n(\theta) \varepsilon_{t-n}, \quad (26)$$

where $\eta_t(\theta)$ is well-defined for $d - d_0 > -1/2$ and where we used

$$c(z, \psi) := b(z, \psi)a(z, \psi_0) = \frac{a(z, \psi_0)}{a(z, \psi)} = \sum_{n=0}^{\infty} c_n(\psi)z^n, \quad (27)$$

$$\varphi_n(\theta) := \sum_{m=0}^n \pi_m(d_0 - d)c_{n-m}(\psi). \quad (28)$$

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

$$|c_n(\psi)| = O(n^{-2-\zeta}) \text{ uniformly in } \psi \in \Psi \quad (29)$$

and from Lemma B.2 the coefficients $\varphi_n(\theta)$ then satisfy

$$|\varphi_n(\theta)| = O(n^{\max(d_0-d-1, -2-\zeta)}) \text{ uniformly in } \psi \in \Psi, \quad (30)$$

such that, in particular, when $d - d_0 > -1/2$, $\eta_t(\theta)$ is a linear process with square summable coefficients.

Let the deterministic function $r(\theta)$ denote the pointwise probability limit of $R(\theta)$, shown subsequently to be given by

$$r(\theta) := \begin{cases} \int_0^1 \sigma(s)^2 ds \sum_{n=0}^{\infty} \varphi_n(\theta)^2 & \text{if } d - d_0 > -1/2, \\ \infty & \text{if } d - d_0 \leq -1/2. \end{cases} \quad (31)$$

According to (31) the parameter space Θ is partitioned into three disjoint compact subsets, $\Theta_1 := \Theta_1(\kappa_1) = D_1 \times \Psi$, $\Theta_2 := \Theta_2(\kappa_1, \kappa_2) = D_2 \times \Psi$, and $\Theta_3 := \Theta_3(\kappa_2) = D_3 \times \Psi$, where $D_1 := D_1(\kappa_1) = D \cap \{d : d - d_0 \leq -1/2 - \kappa_1\}$, $D_2 := D_2(\kappa_1, \kappa_2) = D \cap \{-1/2 - \kappa_1 \leq d - d_0 \leq -1/2 + \kappa_2\}$, and $D_3 := D_3(\kappa_2) = D \cap \{d - d_0 \geq -1/2 + \kappa_2\}$, for some constants $0 < \kappa_2 < \kappa_1 < 1/2$ to be determined later. Here, special care is taken with respect to Θ_2 , where the convergence of the objective function is non-uniform, as evident in (31).

Clearly, $\theta_0 \in \Theta_3$ and if $d_1 > d_0 - 1/2$ then the choice $\kappa_2 = d_1 - d_0 + 1/2 > 0$ implies that Θ_1 and Θ_2 are empty in which case the proof is easily simplified accordingly.

The proof proceeds as follows. First, it is shown that for any $K > 0$ there exists a (fixed) $\bar{\kappa}_2 > 0$ such that

$$P\left(\inf_{\theta \in \Theta_1(\kappa_1) \cup \Theta_2(\kappa_1, \bar{\kappa}_2)} R(\theta) > K\right) \rightarrow 1 \text{ as } T \rightarrow \infty. \quad (32)$$

This implies that $P(\hat{\theta} \in \Theta_3(\bar{\kappa}_2)) \rightarrow 1$ as $T \rightarrow \infty$, so that the relevant parameter space is reduced to $\Theta_3(\bar{\kappa}_2)$. From Theorem 5.7 of van der Vaart (1998) the desired result then follows if, for any fixed $\kappa_2 \in (0, 1/2)$,

$$\sup_{\theta \in \Theta_3(\kappa_2)} |R(\theta) - r(\theta)| \xrightarrow{P} 0 \text{ as } T \rightarrow \infty, \quad (33)$$

$$\inf_{\theta \in \Theta_3(\kappa_2) \cap \{\theta : |\theta - \theta_0| \geq \delta\}} r(\theta) > r(\theta_0) \text{ for all } \delta > 0. \quad (34)$$

Condition (33) entails uniform convergence of the objective function on Θ_3 , and condition (34) ensures that the optimum of the limit function is uniquely attained at the true value.

We show (32) in Appendix D.3, after having established convergence on Θ_1 and Θ_2 in Appendices D.1 and D.2, respectively. Then we prove (33) and (34) in Appendices D.4 and D.5, respectively.

In the proofs we make repeated use of the following lemma, which shows that the problem can be simplified by considering the sum of squares of $\Delta_+^{d-d_0} e_t(\psi)$ rather than that of $\varepsilon_t(\theta)$ in the analysis of $R(\theta)$. That is, the truncation in the residual in the definition of $R(\theta)$ can be dispensed with in the asymptotic analysis.

Lemma D.1. *With the notation of this section and under the assumptions of Theorem 1 and $0 < \kappa_1 < \min(1/2, \zeta/2 + 1/4)$ it holds that, as $T \rightarrow \infty$,*

$$\sup_{\theta \in \Theta_1} |T^{2(d-d_0)} \sum_{t=1}^T \varepsilon_t(\theta)^2 - T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2| \xrightarrow{P} 0, \quad (35)$$

$$\sup_{\theta \in \Theta_2 \cup \Theta_3} |T^{-1} \sum_{t=1}^T \varepsilon_t(\theta)^2 - T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2| \xrightarrow{P} 0. \quad (36)$$

Proof. First decompose

$$\sum_{t=1}^T \varepsilon_t(\theta)^2 - \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 = \sum_{t=1}^T \varepsilon_t(\theta) (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi)) \quad (37)$$

$$+ \sum_{t=1}^T \Delta_+^{d-d_0} e_t(\psi) (\varepsilon_t(\theta) - \Delta_+^{d-d_0} e_t(\psi)) \quad (38)$$

and note that

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

where $\phi_{tm} := - \sum_{j=0}^{t-1} \pi_j(d_0 - d) b_{m-j}(\psi)$ satisfies, see (2) and Lemmas B.1 and B.2,

$$\begin{aligned} \sup_{\psi \in \Psi} \sum_{m=t}^{\infty} |\phi_{tm}| &\leq c \sum_{m=t}^{\infty} \sum_{j=0}^{t-1} j^{d_0-d-1} (m-j)^{-2-\zeta} \\ &\leq c \sum_{j=0}^{t-1} j^{d_0-d-1} (t-j)^{-1-\zeta} \leq c(1 + \log t) t^{\max(d_0-d, -\zeta)-1}. \end{aligned} \quad (39)$$

Rewrite the term (38) as

$$(38) = \sum_{t=1}^T \sum_{j=0}^{t-1} \pi_j(d_0 - d) \sum_{n=0}^{\infty} b_n(\psi) \sum_{m=t}^{\infty} \phi_{tm} (u_{t-j-n} u_{t-m} - E(u_{t-j-n} u_{t-m})) \quad (40)$$

$$+ \sum_{t=1}^T \sum_{j=0}^{t-1} \pi_j(d_0 - d) \sum_{n=0}^{\infty} b_n(\psi) \sum_{m=t}^{\infty} \phi_{tm} E(u_{t-j-n} u_{t-m}). \quad (41)$$

The proof for (37) is identical to that for (38), except the summation over n in (37) is from t to ∞ . For (41) we note that $\sup_t |E(u_t u_{t-n})| = \sup_t |\sum_{m=n}^{\infty} a_m(\psi_0) a_{m-n}(\psi_0) \sigma_{t-m}^2| \leq$

$c \sum_{m=n}^{\infty} m^{-2-\zeta} (m-n)^{-2-\zeta} \leq cn^{-2-\zeta}$ with $\zeta > 0$ given in Assumption C(ii), such that $\sum_{m=t}^{\infty} |E(u_{t-j-n}u_{t-m})| \leq c|t-j-n|^{-1-\zeta}$. Using also (39) it holds that

$$\begin{aligned} \sup_{\psi \in \Psi} \sum_{m=t}^{\infty} \phi_{tm} E(u_{t-j-n}u_{t-m}) &\leq \left(\sup_{\psi \in \Psi} \sum_{m=t}^{\infty} |\phi_{tm}| \right) \left(\sum_{m=t}^{\infty} |E(u_{t-j-n}u_{t-m})| \right) \\ &\leq c(1 + \log t) t^{\max(d_0-d, -\zeta)-1} |t-j-n|^{-1-\zeta}. \end{aligned}$$

It also holds that

$$\sup_{\psi \in \Psi} \sum_{n=0}^{\infty} |b_n(\psi)| |t-j-n|^{-1-\zeta} \leq c \sum_{n=0}^{\infty} n^{-2-\zeta} |t-j-n|^{-1-\zeta} \leq c(t-j)^{-1-\zeta}$$

by (2) and Lemma B.3(ii). Consequently,

$$\begin{aligned} \sup_{\psi \in \Psi} |(41)| &\leq c \sum_{t=1}^T (1 + \log t) t^{\max(d_0-d, -\zeta)-1} \sum_{j=0}^{t-1} j^{d_0-d-1} (t-j)^{-1-\zeta} \\ &\leq c \sum_{t=1}^T (1 + \log t)^2 t^{2\max(d_0-d, -\zeta)-2} \end{aligned}$$

by Lemmas B.1 and B.2. Thus, $\sup_{\theta \in \Theta_1} T^{2(d-d_0)} |(41)| \leq c(\log T)^3 T^{-1} \rightarrow 0$ as $T \rightarrow \infty$ and $\sup_{\theta \in \Theta_2 \cup \Theta_3} T^{-1} |(41)| \leq c(\log T)^3 T^{-1+2\kappa_1} \rightarrow 0$ as $T \rightarrow \infty$.

Changing the order of the summations, (40) is

$$- \sum_{j=0}^{T-1} \pi_j (d_0 - d) \sum_{n=0}^{\infty} b_n(\psi) \sum_{m=j+1}^{\infty} \sum_{k=0}^{\min(m, T)-1} \pi_k (d_0 - d) b_{m-k}(\psi) \sum_{t=\max(j, k)+1}^{\min(m, T)} v_t, \quad (42)$$

where the summand $v_t := u_{t-j-n}u_{t-m} - E(u_{t-j-n}u_{t-m})$ is mean zero with autocovariances

$$\begin{aligned} E v_t v_s &= \sum_{k_1, k_2=0}^{\infty} \sum_{l_1, l_2=0}^{\infty} a_{k_1}(\psi_0) a_{k_2}(\psi_0) a_{l_1}(\psi_0) a_{l_2}(\psi_0) \sigma_{t-j-n-k_1} \sigma_{t-m-k_2} \sigma_{s-j-n-l_1} \sigma_{s-m-l_2} \\ &\quad \times [E(z_{t-j-n-k_1} z_{t-m-k_2} z_{s-j-n-l_1} z_{s-m-l_2}) - E(z_{t-j-n-k_1} z_{t-m-k_2}) E(z_{s-j-n-l_1} z_{s-m-l_2})]. \end{aligned}$$

The expectations are non-zero only if the two highest subscripts are equal (Lemma A.2). Routine calculations using (2), Assumption A, and Lemma B.3(ii) show that $|E v_t v_s| \leq c|s-t|^{-2-\zeta}$. Since the summation $\sum_{t=\max(j, k)+1}^{\min(m, T)}$ has at most m terms it follows that

$$E \left(\sum_{t=\max(j, k)+1}^{\min(m, T)} v_t \right)^2 = \sum_{t, s=\max(j, k)+1}^{\min(m, T)} E(v_t v_s) \leq c \sum_{t, s=\max(j, k)+1}^{\min(m, T)} |t-s|^{-2-\zeta} \leq cm$$

such that $\sum_{t=\max(j, k)+1}^{\min(m, T)} v_t = O_P(m^{1/2})$. From (42) it now follows, using Lemma B.1 and $\sup_{\psi \in \Psi} \sum_{n=0}^{\infty} |b_n(\psi)| < \infty$, that (40) satisfies

$$\sup_{\psi \in \Psi} |(40)| = O_P \left(\sum_{j=1}^{T-1} j^{d_0-d-1} \sup_{\psi \in \Psi} \sum_{m=j+1}^T \sum_{k=1}^{m-1} k^{d_0-d-1} |b_m(\psi)| m^{1/2} \right) \quad (43)$$

$$+ O_P \left(\sum_{j=1}^{T-1} j^{d_0-d-1} \sup_{\psi \in \Psi} \sum_{m=T+1}^{\infty} \sum_{k=1}^{T-1} k^{d_0-d-1} |b_m(\psi)| m^{1/2} \right). \quad (44)$$

For (43) change the order of the summations,

$$\begin{aligned} \sup_{\psi \in \Psi} \sum_{m=j+1}^T \sum_{k=1}^{m-1} k^{d_0-d-1} |b_m(\psi)| m^{1/2} &\leq \sum_{k=1}^{T-1} k^{d_0-d-1} \sum_{m=\max(j,k)+1}^T m^{-3/2-\zeta} \\ &\leq c(\log T) T^{\max(d_0-d,0)} (j+1)^{-1/2-\zeta}. \end{aligned}$$

Then

$$\begin{aligned} \sup_{d \in D_1} T^{2(d-d_0)} (\log T) T^{\max(d_0-d,0)} \sum_{j=1}^{T-1} j^{d_0-d-3/2-\zeta} &\leq c(\log T)^2 T^{-1/2+\max(-\kappa_1, -\zeta)}, \\ \sup_{d \in D_2 \cup D_3} T^{-1} (\log T) T^{\max(d_0-d,0)} \sum_{j=1}^{T-1} j^{d_0-d-3/2-\zeta} &\leq c(\log T)^2 T^{-1/2+\kappa_1+\max(0, \kappa_1-\zeta)}, \end{aligned}$$

which shows the result for (43). Similarly, for (44),

$$\begin{aligned} \sup_{\psi \in \Psi} \sum_{m=T+1}^{\infty} \sum_{k=1}^{T-1} k^{d_0-d-1} |b_m(\psi)| m^{1/2} &\leq \sum_{k=1}^{T-1} k^{d_0-d-1} \sum_{m=T+1}^{\infty} m^{-3/2-\zeta} \\ &\leq c(\log T) T^{\max(0, d_0-d)-1/2-\zeta}, \end{aligned}$$

which gives the bounds

$$\begin{aligned} \sup_{d \in D_1} T^{2(d-d_0)} (\log T) T^{\max(0, d_0-d)-1/2-\zeta} \sum_{j=1}^{T-1} j^{d_0-d-1} &\leq c(\log T)^2 T^{-1/2-\zeta}, \\ \sup_{d \in D_2 \cup D_3} T^{-1} (\log T) T^{\max(0, d_0-d)-1/2-\zeta} \sum_{j=1}^{T-1} j^{d_0-d-1} &\leq c(\log T)^2 T^{-1/2-\zeta+2\kappa_1}, \end{aligned}$$

showing the result for (44) and hence concluding the proof. \square

D.1 Convergence on $\Theta_1(\kappa_1)$

First, if $\theta \in \Theta_1(\kappa_1)$ then $\varepsilon_t(\theta)$ should be normalized by $T^{d-d_0+1/2}$, and by Lemma D.1 the difference between $T^{2(d-d_0)+1} R(\theta)$ and $T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2$ is negligible in probability uniformly in $\theta \in \Theta_1$, so it suffices to consider the latter product moment. We apply the Beveridge-Nelson decomposition,

$$e_t(\psi) = c(L, \psi) \varepsilon_t = \left(\sum_{n=0}^{\infty} c_n(\psi) \right) \varepsilon_t + \sum_{n=0}^{\infty} \tilde{c}_n(\psi) \Delta \varepsilon_{t-n}, \quad (45)$$

where $0 < |\sum_{n=0}^{\infty} c_n(\psi)| < \infty$ and $\tilde{c}_n(\psi) = -\sum_{k=n+1}^{\infty} c_k(\psi)$ satisfies $|\tilde{c}_n(\psi)| \leq cn^{-1-\zeta}$ uniformly in $\psi \in \Psi$, see (29) and also Phillips and Solo (1992, Lemma 2.1). This implies, in particular, that $\sum_{n=0}^{\infty} |\tilde{c}_n(\psi)| < \infty$ uniformly in $\psi \in \Psi$. The relevant product moment can then

be decomposed as

$$T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 \geq \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} \varepsilon_t)^2 \quad (46)$$

$$+ 2 \left(\sum_{n=0}^{\infty} c_n(\psi) \right) T^{2(d-d_0)} \sum_{t=1}^T \Delta_+^{d-d_0} \varepsilon_t \sum_{n=0}^{\infty} \tilde{c}_n(\psi) \Delta_+^{d-d_0+1} \varepsilon_{t-n}. \quad (47)$$

By the Cauchy-Schwarz inequality, (47) is bounded by

$$2 \left(\sum_{n=0}^{\infty} c_n(\psi) \right) \left(T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} \varepsilon_t)^2 \right)^{1/2} \left(T^{2(d-d_0)} \sum_{t=1}^T \left(\sum_{n=0}^{\infty} \tilde{c}_n(\psi) \Delta_+^{d-d_0+1} \varepsilon_{t-n} \right)^2 \right)^{1/2}.$$

The term in the first parenthesis satisfies $0 < |\sum_{n=0}^{\infty} c_n(\psi)| < \infty$ uniformly in $\psi \in \Psi$ by Assumption C, and the term in the second parenthesis will be shown to be $O_P(1)$ uniformly in $d \in D_1$ subsequently. The term inside the third parenthesis can be rewritten as

$$\begin{aligned} & T^{2(d-d_0)} \sum_{t=1}^T \sum_{n,m=0}^{\infty} \tilde{c}_n(\psi) \tilde{c}_m(\psi) \sum_{j,k=0}^{t-1} \pi_j(d_0-d-1) \pi_k(d_0-d-1) \varepsilon_{t-j-n} \varepsilon_{t-k-m} \\ &= T^{2(d-d_0)+1} \sum_{n,m=0}^{\infty} \tilde{c}_n(\psi) \tilde{c}_m(\psi) \sum_{j,k=0}^{T-1} \pi_j(d_0-d-1) \pi_k(d_0-d-1) T^{-1} \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m}, \end{aligned}$$

where $T^{-1} \sum_{t=\max(j,k)+1}^T \varepsilon_{t-j-n} \varepsilon_{t-k-m} = O_P(1)$ uniformly in $0 \leq j, k \leq T-1$ and $\sum_{n=0}^{\infty} |\tilde{c}_n(\psi)| < \infty$ uniformly in $\psi \in \Psi$. This leaves the bound

$$\begin{aligned} \sup_{\theta \in \Theta_1} T^{2(d-d_0)} \sum_{t=1}^T \left(\sum_{n=0}^{\infty} \tilde{c}_n(\psi) \Delta_+^{d-d_0+1} \varepsilon_{t-n} \right)^2 &\leq c \sup_{d \in D_1} T^{2(d-d_0)+1} \left(\sum_{j=0}^{T-1} |\pi_j(d_0-d-1)| \right)^2 O_P(1) \\ &= O_P\left(\sup_{d \in D_1} T^{2(d-d_0)+1} \left(\sum_{j=0}^{T-1} j^{d_0-d-2} \right)^2 \right) \\ &= O_P((\log T)^2 T^{-2\kappa_1}) \end{aligned}$$

by application of Lemma B.1, thus showing that (47) converges to zero in probability uniformly in $\theta \in \Theta_1$.

Next, the term (46) is analyzed. By the Cauchy-Schwarz inequality,

$$T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} \varepsilon_t)^2 \geq T^{2(d-d_0)-1} \left(\sum_{t=1}^T \Delta_+^{d-d_0} \varepsilon_t \right)^2 = (T^{d-d_0-1/2} \Delta_+^{d-d_0-1} \varepsilon_T)^2,$$

and we can write $T^{d-d_0-1/2} \Delta_+^{d-d_0-1} \varepsilon_T = T^{d-d_0-1/2} \sum_{j=0}^{T-1} \pi_j(d_0-d+1) \varepsilon_{T-j} = T^{d-d_0-1/2} \sum_{t=1}^T \pi_{T-t}(d_0-d+1) \varepsilon_t$ and apply Lemma A.1 with $U_{Tt} = T^{d-d_0-1/2} \pi_{T-t}(d_0-d+1) \varepsilon_t$, which is a martingale difference array by Assumption A. Firstly, the Lindeberg condition (i) of Lemma A.1 is satisfied by Lyapunov's sufficient condition because $\sum_{t=1}^T EU_{Tt}^4 = T^{4(d-d_0)-2} \sum_{t=1}^T \pi_{T-t}(d_0-d+1)^4 \varepsilon_t^4$

$1)^4 \sigma_t^4 E z_t^4 \leq cT^{-2} \sum_{t=1}^T \left(\frac{T-t}{T}\right)^{4(d_0-d)} \leq cT^{-1} \rightarrow 0$ as $T \rightarrow \infty$. Secondly, we verify condition (ii) of Lemma A.1 by showing L_2 -convergence. Thus,

$$\begin{aligned}
& E \left(\sum_{t=1}^T U_{Tt}^2 - E \sum_{t=1}^T U_{Tt}^2 \right)^2 \\
&= \sum_{t,s=1}^T E(U_{Tt}^2 U_{Ts}^2) - \sum_{t,s=1}^T E(U_{Tt}^2) E(U_{Ts}^2) \\
&= T^{4(d-d_0)-2} \sum_{t,s=1}^T \pi_{T-t}(d_0-d+1)^2 \pi_{T-s}(d_0-d+1)^2 \sigma_t^2 \sigma_s^2 [E(z_t^2 z_s^2) - E(z_t^2) E(z_s^2)] \\
&= T^{4(d-d_0)-2} \sum_{t=1}^T \pi_{T-t}(d_0-d+1)^4 \sigma_t^4 [E(z_t^4) - E(z_t^2)^2] \tag{48}
\end{aligned}$$

$$\begin{aligned}
& + 2T^{4(d-d_0)-2} \sum_{t=2}^T \sum_{s=1}^{t-1} \pi_{T-t}(d_0-d+1)^2 \pi_{T-s}(d_0-d+1)^2 \sigma_t^2 \sigma_s^2 [E(z_t^2 z_s^2) - E(z_t^2) E(z_s^2)]. \tag{49}
\end{aligned}$$

By Assumption A(a),(b)(ii) and Lemma B.1, the term (48) is bounded by

$$cT^{4(d-d_0)-2} \sum_{t=1}^T (T-t)^{4(d_0-d)} \leq cT^{-1} \rightarrow 0.$$

The term (49) is

$$\begin{aligned}
& 2T^{4(d-d_0)-2} \sum_{t=2}^T \sum_{r=1}^{t-1} \pi_{T-t}(d_0-d+1)^2 \pi_{T-t+r}(d_0-d+1)^2 \sigma_t^2 \sigma_{t-r}^2 \kappa_4(t, t, t-r, t-r) \\
& \leq cT^{-2} \sum_{t=2}^T \left(\frac{T-t}{T}\right)^{2(d_0-d)} \left(\frac{T-1}{T}\right)^{2(d_0-d)} \sum_{r=1}^{t-1} |\kappa_4(t, t, t-r, t-r)| \leq cT^{-1} \rightarrow 0
\end{aligned}$$

using Assumption A(a),(b)(iii) and Lemma B.1. Finally,

$$\begin{aligned}
E \sum_{t=1}^T U_{Tt}^2 &= T^{2(d-d_0)-1} \sum_{t=1}^T \pi_{T-t}(d_0-d+1)^2 \sigma_t^2 \\
&= \frac{1}{\Gamma(d_0-d+1)^2} T^{-1} \sum_{t=1}^T \left(\frac{T-t}{T}\right)^{2(d_0-d)} \sigma_t^2 (1+o(1)) \\
&\rightarrow \frac{1}{\Gamma(d_0-d+1)^2} \int_0^1 (1-s)^{2(d_0-d)} \sigma(s)^2 ds =: V(d),
\end{aligned}$$

and we conclude from Lemma A.1 and the above analysis that

$$Q_T(d) := T^{2(d-d_0)-1} \left(\sum_{t=1}^T \Delta_+^{d-d_0} \varepsilon_t \right)^2 = (T^{d-d_0-1/2} \Delta_+^{d-d_0-1} \varepsilon_T)^2 \xrightarrow{D} V(d) \chi_1^2 \text{ as } T \rightarrow \infty, \tag{50}$$

for any fixed $d \in D_1$, which shows the pointwise limit.

To strengthen the pointwise convergence in (50) to weak convergence in $\mathcal{C}(D_1)$ it is sufficient to show that $Q_T(d)$ is tight as a function of the parameter. We prove tightness using the moment

condition in Billingsley (1968, Theorem 12.3), which requires showing that $Q_T(d)$ is tight for fixed $d \in D_1$ (which is implied by the pointwise convergence) and that

$$\|Q_T(u_1) - Q_T(u_2)\|_2 \leq c|u_1 - u_2| \quad (51)$$

for some constant $c > 0$ that does not depend on T , u_1 , or u_2 . The tightness condition (51) is satisfied by Lemma C.1, and hence the convergence in (50) is strengthened to

$$Q_T(d) \Rightarrow V(d)\chi_1^2 \text{ in } \mathcal{C}(D_1).$$

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

$$\inf_{d \in D_1} Q_T(d) \xrightarrow{D} \inf_{d \in D_1} V(d)\chi_1^2,$$

which is positive almost surely. It follows that

$$\inf_{\theta \in \Theta_1} R(\theta) \geq \inf_{\theta \in \Theta_1} \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 T^{2(d_0-d)+1} Q_T(d) + o_P(1)$$

and, for any $K > 0$,

$$P\left(\inf_{\theta \in \Theta_1} \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 T^{2(d_0-d)+1} Q_T(d) > K \right) \rightarrow 1 \text{ as } T \rightarrow \infty$$

because $\inf_{\psi \in \Psi} \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 > 0$ by Assumption C and $2(d_0 - d) + 1 \geq 2\kappa_1 > 0$ for $d \in D_1$.

D.2 Convergence on $\Theta_2(\kappa_1, \kappa_2)$

First note that by (36) of Lemma D.1 it suffices to prove the result for $T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2$. Letting $v := d - d_0 \in [-1/2 - \kappa_1, -1/2 + \kappa_2]$, $R_{1T}(v) := T^{-1} \sum_{t=1}^T (\Delta_+^v \varepsilon_t)^2$, and $R_{2T}(v, \psi) := T^{-1} \sum_{t=1}^T (\Delta_+^v \varepsilon_t) \left(\sum_{n=0}^{\infty} \tilde{c}_n(\psi) \Delta_+^{1+v} \varepsilon_{t-n} \right)$, and applying the decomposition (45), the relevant product moment is

$$T^{-1} \sum_{t=1}^T (\Delta_+^v e_t(\psi))^2 \geq \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 R_{1T}(v) + 2 \left(\sum_{n=0}^{\infty} c_n(\psi) \right) R_{2T}(v, \psi).$$

The second term, $R_{2T}(v, \psi)$, is $O_P(1)$ uniformly in $|v + 1/2| \leq \kappa_1$ and $\psi \in \Psi$ by Lemma C.3 with $\tilde{\Psi} = \Psi$, $\zeta_{1n}(\psi) = 1_{\{n=0\}}$, $\zeta_{2n}(\psi) = \tilde{c}_n(\psi)$, $u_1 = v \geq -1/2 - \kappa_1$, $u_2 = 1 + v \geq 1/2 - \kappa_1$ such that $a = \min(1/2 - \kappa_1, 1 - 2\kappa_2) > 0$.

To analyze $R_{1T}(v)$ decompose $\Delta_+^v \varepsilon_t$ as

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

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

$$R_{1T}(v) \geq T^{-1} \sum_{t=N+1}^T (\Delta_+^v \varepsilon_t)^2 \geq T^{-1} \sum_{t=N+1}^T w_{1t}^2 + 2T^{-1} \sum_{t=N+1}^T w_{1t} w_{2t}. \quad (52)$$

Setting $N = N_T := \lfloor T^\alpha \rfloor$ with $0 < \alpha < \min(\frac{1/2 - \kappa_1}{1/2 + \kappa_1}, \frac{1/2}{1/2 + 2\kappa_1})$, noting that such an α exists because $0 < \kappa_1 < 1/2$, it follows from (17) of Lemma C.2 that the second term on the right-hand side of (52) converges in probability to zero uniformly in $|v + 1/2| \leq \kappa_1$ and that

$$\sup_{|v+1/2| \leq \kappa_1} \left| T^{-1} \sum_{t=N_T+1}^T w_{1t}^2 - E \left(T^{-1} \sum_{t=N_T+1}^T w_{1t}^2 \right) \right| \xrightarrow{P} 0 \text{ as } T \rightarrow \infty.$$

Thus, the right-hand side of (52) minus $E(T^{-1} \sum_{t=N_T+1}^T w_{1t}^2)$ converges uniformly in probability to zero as $T \rightarrow \infty$. It follows, see Assumption A(a), that

$$\begin{aligned} \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 R_{1T}(v) &\geq \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 E \left(T^{-1} \sum_{t=N_T+1}^T w_{1t}^2 \right) + \mu_{1T}(\theta) \\ &= \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 T^{-1} \sum_{t=N_T+1}^T \sum_{n=0}^{N_T-1} \pi_n(-v)^2 \sigma_{t-n}^2 + \mu_{1T}(\theta) \\ &\geq \left(\inf_{0 \leq s \leq 1} \sigma(s)^2 \right) \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 T^{-1} (T - N_T) F_{N_T}(v) + \mu_{1T}(\theta), \end{aligned}$$

where $F_{N_T}(v) = \sum_{n=0}^{N_T-1} \pi_n(-v)^2$ and $\mu_{1T}(\theta) \xrightarrow{P} 0$ as $T \rightarrow \infty$ uniformly in $|v + 1/2| \leq \kappa_1$ and $\psi \in \Psi$.

D.3 Proof of Eqn. (32)

We need to show that, for any $K > 0, \eta > 0$, there exists a $\bar{\kappa}_2 > 0$ and a T_0 such that

$$P \left(\inf_{\theta \in \Theta_1(\kappa_1) \cup \Theta_2(\kappa_1, \bar{\kappa}_2)} R(\theta) < K \right) \leq \eta$$

for all $T \geq T_0$. Since $\inf_{\theta \in \Theta_1 \cup \Theta_2} R(\theta) \leq \sum_{j=1}^2 \inf_{\theta \in \Theta_j} R(\theta)$, the two sets Θ_1 and Θ_2 can be considered separately.

First consider the interval $\Theta_1(\kappa_1)$ with $\kappa_1 = \bar{\kappa}_1$ satisfying $0 < \bar{\kappa}_1 < \min(1/2, \zeta/2 + 1/4)$, and define $\bar{\Theta}_1 := \Theta_1(\bar{\kappa}_1)$. It holds from Section D.1 that $P(\inf_{\theta \in \bar{\Theta}_1} R(\theta) > K) \rightarrow 1$ as $T \rightarrow \infty$, i.e., for any $K > 0, \eta > 0$, there exists a T_1 such that $P(\inf_{\theta \in \bar{\Theta}_1} R(\theta) < K) \leq \eta/2$ for all $T \geq T_1$.

Second, having already fixed $\kappa_1 = \bar{\kappa}_1$, consider $\Theta_2(\bar{\kappa}_1, \kappa_2)$. From Section D.2 with $\kappa_1 = \bar{\kappa}_1$ and $\alpha = 1/6$,

$$R(\theta) \geq \left(\inf_{0 \leq s \leq 1} \sigma(s)^2 \right) \left(\sum_{n=0}^{\infty} c_n(\psi) \right)^2 T^{-1} (T - T^{1/6}) F_{T^{1/6}}(d - d_0) + \mu_T(\theta),$$

where $\mu_T(\theta) = O_P(1)$ as $T \rightarrow \infty$ uniformly in $d \in [d_0 - 1/2 - \bar{\kappa}_1, d_0 - 1/2 + \bar{\kappa}_1] \supset D_2$ and $\psi \in \Psi$. From Lemma B.4,

$$F_{T^{1/6}}(d - d_0) \geq 1 + c \frac{1 - (T - 1)^{-2\kappa_2/6}}{2\kappa_2}.$$

The factor $(2\kappa_2)^{-1}(1 - (T - 1)^{-2\kappa_2/6})$ is increasing in T from 0 (for $T = 2$) to $(2\kappa_2)^{-1}$ and decreasing in κ_2 from $\frac{\log(T-1)}{6}$ (for $\kappa_2 = 0$) to 0, such that $(2\kappa_2)^{-1}(1 - (T - 1)^{-2\kappa_2/6}) \rightarrow \infty$

as $(\kappa_2, T) \rightarrow (0, \infty)$. Because $(\sum_{n=0}^{\infty} c_n(\psi))^2 > 0$ uniformly in $\psi \in \Psi$ and $\inf_{0 \leq s \leq 1} \sigma(s)^2 > 0$, it follows that for any $K > 0, \eta > 0$, there exists $\bar{\kappa}_2 > 0$ (small) and T_2 such that, with $\bar{\Theta}_2 := \Theta_2(\bar{\kappa}_1, \bar{\kappa}_2)$, $P(\inf_{\theta \in \bar{\Theta}_2} R(\theta) < K) \leq \eta/2$ for all $T \geq T_2$.

Combining these results, for any $K > 0, \eta > 0$, there exists a $\bar{\kappa}_2 > 0$ such that

$$P\left(\inf_{\theta \in \bar{\Theta}_1 \cup \bar{\Theta}_2} R(\theta) < K\right) \leq \sum_{j=1}^2 P\left(\inf_{\theta \in \bar{\Theta}_j} R(\theta) < K\right) \leq \sum_{j=1}^2 \eta/2 = \eta$$

for all $T \geq \max(T_1, T_2) = T_0$, which proves (32).

D.4 Convergence on $\Theta_3(\kappa_2)$ and proof of (33)

Again, by Lemma D.1, it suffices to prove the result for $T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2$. In this case, define the untruncated process $\eta_t(\theta) := \Delta^{d-d_0} e_t(\psi)$ and note that $\eta_t(\theta) - \Delta_+^{d-d_0} e_t(\psi) = \sum_{n=t}^{\infty} \pi_n(d_0 - d) e_{t-n}(\psi) = \sum_{n=t}^{\infty} \varphi_n(\theta) \varepsilon_{t-n}$, see (28), with

$$E(\eta_t(\theta) - \Delta_+^{d-d_0} e_t(\psi))^2 = \sum_{n=t}^{\infty} \varphi_n(\theta)^2 \sigma_{t-n}^2 \leq c \sum_{n=t}^{\infty} n^{2 \max(d_0 - d - 1, -2 - \zeta)} \leq ct^{-2\kappa_2} \rightarrow 0$$

for all $\theta \in \Theta_3$ (pointwise), using (30) and Assumption A(a). It follows that

$$R(\theta) = T^{-1} \sum_{t=1}^T \eta_t(\theta)^2 + o_P(1),$$

and furthermore,

$$\begin{aligned} ET^{-1} \sum_{t=1}^T \eta_t(\theta)^2 &= T^{-1} \sum_{t=1}^T \sum_{n=0}^{\infty} \varphi_n(\theta)^2 \sigma_{t-n}^2 \\ &= T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n=0}^{\infty} \varphi_n(\theta)^2 + T^{-1} \sum_{t=1}^T \sum_{n=0}^{\infty} \varphi_n(\theta)^2 (\sigma_{t-n}^2 - \sigma_t^2). \end{aligned}$$

Let $q_T := \lfloor T^\chi \rfloor$ for some $\chi \in (0, 1)$. Then the last term is bounded as

$$T^{-1} \sum_{t=1}^T \sum_{n=0}^{\infty} \varphi_n(\theta)^2 (\sigma_{t-n}^2 - \sigma_t^2) \leq \sum_{n=0}^{q_T} \varphi_n(\theta)^2 T^{-1} \sum_{t=1}^T |\sigma_{t-n}^2 - \sigma_t^2| \quad (53)$$

$$+ \sum_{n=q_T+1}^{\infty} \varphi_n(\theta)^2 T^{-1} \sum_{t=1}^T |\sigma_{t-n}^2 - \sigma_t^2|. \quad (54)$$

Because $\sup_{n=1, \dots, q_T} T^{-1} \sum_{t=1}^T |\sigma_{t-n}^2 - \sigma_t^2| \rightarrow 0$ by Cavaliere and Taylor (2009, Lemma A.1) and $\sum_{n=0}^{q_T} \varphi_n(\theta)^2 \leq \sum_{n=0}^{\infty} \varphi_n(\theta)^2 < \infty$ uniformly in $\theta \in \Theta_3$, see (30), it holds that $|(53)| \rightarrow 0$. Next, by Assumption A(a) and by (30) we have, respectively, $\sup_t \sigma_t^2 \leq M < \infty$ such that $\sup_t T^{-1} \sum_{t=1}^T |\sigma_{t-n}^2 - \sigma_t^2| \leq 2M$ and $\sum_{n=q_T+1}^{\infty} \varphi_n(\theta)^2 \leq c \sum_{n=q_T+1}^{\infty} n^{2 \max(d_0 - d - 1, -2 - \zeta)} \leq cq_T^{-2\kappa_2} \rightarrow 0$ uniformly in $\theta \in \Theta_3$, and therefore $|(54)| \rightarrow 0$. This shows that $ET^{-1} \sum_{t=1}^T \eta_t(\theta)^2 = \int_0^1 \sigma(s)^2 ds \sum_{n=0}^{\infty} \varphi_n(\theta)^2 + o(1) = r(\theta) + o(1)$, see (31), because $T^{-1} \sum_{t=1}^T \sigma_t^2 \rightarrow \int_0^1 \sigma(s)^2 ds$ by Assumption A(a) and the continuous mapping theorem. To prove

$$R(\theta) = T^{-1} \sum_{t=1}^T \eta_t(\theta)^2 + o_P(1) \xrightarrow{P} r(\theta) \text{ as } T \rightarrow \infty, \quad (55)$$

i.e., the pointwise limit in probability, we show L_2 -convergence.

To that end we find, in a similar way as in (48) and (49), that

$$\begin{aligned}
& E \left(T^{-1} \sum_{t=1}^T \eta_t(\theta)^2 - ET^{-1} \sum_{t=1}^T \eta_t(\theta)^2 \right)^2 \\
&= T^{-2} \sum_{t,s=1}^T E(\eta_t(\theta)^2 \eta_s(\theta)^2) - T^{-2} \sum_{t,s=1}^T E(\eta_t(\theta)^2) E(\eta_s(\theta)^2) \\
&= T^{-2} \sum_{t,s=1}^T \sum_{n_1, n_2=0}^{\infty} \sum_{m_1, m_2=0}^{\infty} \left(\prod_{i=1}^2 \varphi_{n_i}(\theta) \varphi_{m_i}(\theta) \sigma_{t-n_i} \sigma_{s-m_i} \right) \\
&\quad \times [E(z_{t-n_1} z_{t-n_2} z_{s-m_1} z_{s-m_2}) - E(z_{t-n_1} z_{t-n_2}) E(z_{s-m_1} z_{s-m_2})],
\end{aligned}$$

where the expectations are zero unless the two highest subscripts are equal (Lemma A.2). By symmetry, we only need to consider three cases, which we now enumerate.

Case 1) $t - n_1 = t - n_2 = s - m_1 = s - m_2$, in which case the expectations and the σ_t 's are uniformly bounded by Assumption A and we find the contribution

$$cT^{-2} \sum_{t=1}^T \left(\sum_{n=0}^{\infty} \varphi_n(\theta)^2 \right)^2 \leq cT^{-1} \left(\sum_{n=0}^{\infty} n^{-1-2\kappa_2} \right)^2 \leq cT^{-1} \rightarrow 0$$

using (30).

Case 2) $t - n_1 = t - n_2 > s - m_1 \geq s - m_2$, where the contribution is

$$\begin{aligned}
& T^{-2} \sum_{t,s=1}^T \sum_{n=0}^{\infty} \sum_{m_1=\max(0, s-t+n+1)}^{\infty} \sum_{m_2=m_1}^{\infty} \varphi_n(\theta)^2 \varphi_{m_1}(\theta) \varphi_{m_2}(\theta) \\
&\quad \times \sigma_{t-n}^2 \sigma_{s-m_1} \sigma_{s-m_2} \kappa_4(t-n, t-n, s-m_1, s-m_2) \\
&\leq cT^{-2} \sum_{t,s=1}^T \sum_{n=0}^{\infty} n^{-1-2\kappa_2} \max(0, s-t+n+1)^{-1-2\kappa_2} \\
&\quad \times \sum_{m_1=\max(0, s-t+n+1)}^{\infty} \sum_{m_2=m_1}^{\infty} |\kappa_4(t-n, t-n, s-m_1, s-m_2)| \\
&\leq cT^{-2} \sum_{t,s=1}^T \sum_{n=0}^{\infty} n^{-1-2\kappa_2} \max(0, s-t+n+1)^{-1-2\kappa_2} \\
&\leq cT^{-2} \sum_{t,s=1}^T |t-s|^{-1-2\kappa_2} \leq cT^{-1} \rightarrow 0
\end{aligned}$$

using Assumption A(a),(b)(iii) together with (30).

Case 3) $t - n_1 = s - m_1 > t - n_2 \geq s - m_2$, where we distinguish between the two subcases:

Case 3a) $t - n_2 = s - m_2$ with the contribution

$$\begin{aligned}
& T^{-2} \sum_{t,s=1}^T \sum_{n_1=\max(0,t-s)}^{\infty} \sum_{n_2=n_1+1}^{\infty} \varphi_{n_1}(\theta) \varphi_{n_2}(\theta) \varphi_{s-t+n_1}(\theta) \varphi_{s-t+n_2}(\theta) \sigma_{t-n_1}^2 \sigma_{t-n_2}^2 \tau_{n_2-n_1, n_2-n_1} \\
& \leq cT^{-2} \sum_{t,s=1}^T \sum_{n_1=\max(0,t-s)}^{\infty} n_1^{-1/2-\kappa_2} (s-t+n_1)^{-1/2-\kappa_2} \sum_{n_2=n_1+1}^{\infty} n_2^{-1/2-\kappa_2} (s-t+n_2)^{-1/2-\kappa_2} \\
& = cT^{-2} \sum_{t=1}^T \sum_{n_1=0}^{\infty} n_1^{-1-2\kappa_2} \sum_{n_2=n_1+1}^{\infty} n_2^{-1-2\kappa_2} \\
& \quad + 2cT^{-2} \sum_{t>s=1}^T \sum_{n_1=t-s}^{\infty} n_1^{-1/2-\kappa_2} (s-t+n_1)^{-1/2-\kappa_2} \sum_{n_2=n_1+1}^{\infty} n_2^{-1/2-\kappa_2} (s-t+n_2)^{-1/2-\kappa_2} \\
& \leq cT^{-1} + cT^{-2} \sum_{t>s=1}^T \sum_{n_1=\max(0,t-s)}^{\infty} n_1^{-1/2-2\kappa_2} (s-t+n_1)^{-1/2-\kappa_2} \\
& \leq cT^{-1} + cT^{-2} \sum_{t>s=1}^T (t-s)^{-2\kappa_2} \leq cT^{-1} + cT^{-2\kappa_2} \rightarrow 0,
\end{aligned}$$

where we once again used (30) and Assumption A(a),(b)(iii).

Case 3b) $t - n_2 > s - m_2$ with the contribution

$$\begin{aligned}
& T^{-2} \sum_{t,s=1}^T \sum_{n_1=\max(0,t-s)}^{\infty} \sum_{n_2=n_1+1}^{\infty} \sum_{m=s-t+n_2+1}^{\infty} \varphi_{n_1}(\theta) \varphi_{n_2}(\theta) \varphi_{s-t+n_1}(\theta) \varphi_m(\theta) \\
& \quad \times \sigma_{t-n_1}^2 \sigma_{t-n_2} \sigma_{s-m} \kappa_4(t-n_1, t-n_1, t-n_2, s-m) \\
& \leq cT^{-1} + cT^{-2} \sum_{t>s=1}^T \sum_{n_1=\max(0,t-s)}^{\infty} n_1^{-1/2-2\kappa_2} (s-t+n_1)^{-1/2-\kappa_2} \\
& \leq cT^{-1} + cT^{-2\kappa_2} \rightarrow 0
\end{aligned}$$

as in Case 3a). This shows that (55) holds pointwise for all $\theta \in \Theta_3$.

The result (55) can be strengthened to uniform convergence in probability by showing that $R(\theta)$ is stochastically equicontinuous (or tight). From Newey (1991, Corollary 2.2) this holds if the derivative of $R(\theta)$ is dominated uniformly in $\theta \in \Theta_3$ by a random variable $B_T = O_P(1)$. From Lemma C.3 with $u_1 = u_2 = d - d_0 \geq -1/2 + \kappa_2$, $a = 2\kappa_2$, and $\tilde{\Psi} = \Psi$ (noting that only summability of the linear coefficients is assumed in Lemma C.3 and this is satisfied uniformly on Θ by the derivatives of $c_n(\psi)$ by Assumption C(iii)) it holds that $B_T = \sup_{\theta \in \Theta_3} |\frac{\partial R(\theta)}{\partial \theta}| = O_P(1)$, showing that $R(\theta)$ is stochastically equicontinuous on Θ_3 and hence that (55) holds uniformly in $\theta \in \Theta_3$. Since the result holds for any κ_2 it proves (33).

D.5 Proof of Eqn. (34)

Since $r(\theta_0) = \int_0^1 \sigma(s)^2 ds$ it is sufficient to prove that

$$\inf_{\theta \in \Theta_3 \cap \{\theta: |\theta - \theta_0| \geq \delta\}} \sum_{n=0}^{\infty} \varphi_n(\theta)^2 > 1 \text{ for all } \delta > 0 \text{ and all } \kappa_2 \in (0, 1/2).$$

Because $\varphi_0(\theta) = 1$ for all $\theta \in \Theta_3$ by Assumption C, it is clear that $\sum_{n=0}^{\infty} \varphi_n(\theta)^2 = 1 + \sum_{n=1}^{\infty} \varphi_n(\theta)^2 \geq 1$, and by Assumption D the inequality is strict for all $\theta \neq \theta_0$, which proves (34) by continuity of $\varphi_n(\cdot)$ and compactness of Θ_3 .

Appendix E Proof of Theorem 2

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

$$0 = T^{1/2} \frac{\partial R(\hat{\theta})}{\partial \theta} = T^{1/2} \frac{\partial R(\theta_0)}{\partial \theta} + T^{1/2} \frac{\partial^2 R(\bar{\theta})}{\partial \theta \partial \theta'} (\hat{\theta} - \theta_0), \quad (56)$$

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

E.1 Convergence of the score function

The normalized score function evaluated at the true value is

$$T^{1/2} \frac{\partial R(\theta_0)}{\partial \theta} = 2T^{-1/2} \sum_{t=1}^T \varepsilon_t(\theta_0) \hat{y}_{1,t-1} \text{ with } \hat{y}_{k,t-1} = \frac{\partial^k}{\partial \theta^{(k)}} \varepsilon_t(\theta) |_{\theta=\theta_0}.$$

Define also

$$S_T := 2T^{-1/2} \sum_{t=1}^T \varepsilon_t y_{1,t-1},$$

where the first element of $y_{1,t-1}$ is $-\sum_{n=1}^{t-1} n^{-1} \varepsilon_{t-n}$ and the remaining p elements are given by $\sum_{n=1}^{t-1} \gamma_n(\psi_0) \varepsilon_{t-n}$. Similarly, the first element of $\hat{y}_{1,t-1}$ is $-\sum_{n=1}^{t-1} n^{-1} \varepsilon_{t-n}(\theta_0)$ and the remaining elements are $\sum_{n=1}^{t-1} \dot{b}_n(\psi_0) u_{t-n}$.

We find that the difference is

$$T^{1/2} \frac{\partial R(\theta_0)}{\partial \theta} - S_T = 2T^{-1/2} \sum_{t=1}^T (\varepsilon_t(\theta_0) - \varepsilon_t) \hat{y}_{1,t-1} + 2T^{-1/2} \sum_{t=1}^T \varepsilon_t (\hat{y}_{1,t-1} - y_{1,t-1}), \quad (57)$$

where

$$\varepsilon_t(\theta_0) - \varepsilon_t = - \sum_{n=t}^{\infty} b_n(\psi_0) u_{t-n}$$

and

$$\hat{y}_{1,t-1} - y_{1,t-1} = \begin{bmatrix} - \sum_{n=1}^{t-1} n^{-1} \sum_{k=t-n}^{\infty} b_k(\psi_0) u_{t-n-k} \\ \sum_{n=1}^{t-1} \dot{b}_n(\psi_0) \sum_{k=t}^{\infty} a_k(\psi_0) \varepsilon_{t-k} \end{bmatrix}.$$

The first term of (57) is then

$$-2T^{-1/2} \sum_{t=1}^T \sum_{n=t}^{\infty} b_n(\psi_0) u_{t-n} \hat{y}_{1,t-1},$$

which has second moment

$$\begin{aligned}
& 4T^{-1} \sum_{t,s=1}^T \sum_{n=t}^{\infty} \sum_{m=s}^{\infty} b_n(\psi_0) b_m(\psi_0) E(u_{t-n} \hat{y}_{1,t-1} u_{s-m} \hat{y}_{1,s-1}) \\
& \leq KT^{-1} \sum_{t,s=1}^T \sum_{n=t}^{\infty} \sum_{m=s}^{\infty} |b_n(\psi_0)| |b_m(\psi_0)| \leq KT^{-1} \sum_{t,s=1}^T t^{-1-\zeta} s^{-1-\zeta} \leq KT^{-1-2\zeta} \rightarrow 0,
\end{aligned}$$

see (2). The second term of (57) is

$$\left[\begin{array}{c} -2T^{-1/2} \sum_{t=1}^T \varepsilon_t \sum_{n=1}^{t-1} n^{-1} \sum_{k=t-n}^{\infty} b_k(\psi_0) u_{t-n-k} \\ 2T^{-1/2} \sum_{t=1}^T \varepsilon_t \sum_{n=1}^{t-1} \dot{b}_n(\psi_0) \sum_{k=t}^{\infty} a_k(\psi_0) \varepsilon_{t-k} \end{array} \right]. \quad (58)$$

The first term in (58) has second moment

$$\begin{aligned}
& 4T^{-1} \sum_{t,s=1}^T \sum_{n=1}^{t-1} \sum_{m=1}^{s-1} n^{-1} m^{-1} \sum_{k=t-n}^{\infty} \sum_{l=s-m}^{\infty} b_k(\psi_0) b_l(\psi_0) E(\varepsilon_t \varepsilon_s u_{t-n-k} u_{s-m-l}) \\
& = 4T^{-1} \sum_{t=1}^T \sum_{n,m=1}^{t-1} n^{-1} m^{-1} \sum_{k=t-n}^{\infty} \sum_{l=t-m}^{\infty} b_k(\psi_0) b_l(\psi_0) \sum_{r=0}^{\infty} \sum_{q=0}^{\infty} a_r(\psi_0) a_q(\psi_0) \sigma_t^2 \sigma_{t-k-n-r} \sigma_{t-l-m-q} \\
& \quad \times (\kappa_4(t, t, t-k-n-r, t-l-m-q) + \kappa_2(t, t) \kappa_2(t-k-n-r, t-l-m-q)) \\
& \leq KT^{-1} \sum_{t=1}^T \sum_{n,m=1}^{t-1} n^{-1} m^{-1} \sum_{k=t-n}^{\infty} \sum_{l=t-m}^{\infty} |b_k(\psi_0)| |b_l(\psi_0)| \sum_{r=0}^{\infty} \sum_{q=0}^{\infty} |a_r(\psi_0)| |a_q(\psi_0)| \\
& \leq KT^{-1} \sum_{t=1}^T \sum_{n,m=1}^{t-1} n^{-1} m^{-1} \sum_{k=t-n}^{\infty} \sum_{l=t-m}^{\infty} k^{-2-\zeta} l^{-2-\zeta} \\
& \leq KT^{-1} \sum_{t=1}^T \sum_{n,m=1}^{t-1} n^{-1} m^{-1} (t-n)^{-1-\zeta} (t-m)^{-1-\zeta} \leq KT^{-1} \sum_{t=1}^T t^{-2} \leq KT^{-1} \rightarrow 0,
\end{aligned}$$

where the first two inequalities use Assumption A(a),(b)(iii) and (2), and the fourth inequality uses Lemma B.2. The second term in (58) has second moment

$$\begin{aligned}
& 4T^{-1} \sum_{t,s=1}^T \sum_{n=1}^{t-1} \sum_{m=1}^{s-1} \dot{b}_n(\psi_0) \dot{b}_m(\psi_0) \sum_{k=t}^{\infty} \sum_{l=s}^{\infty} a_k(\psi_0) a_l(\psi_0) E(\varepsilon_t \varepsilon_s \varepsilon_{t-k} \varepsilon_{s-l}) \\
& \leq KT^{-1} \sum_{t=1}^T \left(\sum_{k=t}^{\infty} |a_k(\psi_0)| \right)^2 \leq KT^{-1} \sum_{t=1}^T (t^{-1-\zeta})^2 \rightarrow 0
\end{aligned}$$

using Lemma A.2, Assumption C(iii), and (2). Thus, each of the terms in (58) and hence those in (57) converge to zero in L_2 -norm and therefore in probability.

Because $y_{1,t-1}$ is measurable with respect to the sigma-algebra $\mathcal{F}_{t-1} := \sigma(\{\varepsilon_s, s \leq t-1\})$, it holds that $v_{Tt} := 2T^{-1/2} \varepsilon_t \sum_{n=1}^{t-1} \xi_{0,n} \varepsilon_{t-n} = 2T^{-1/2} \sigma_t z_t \sum_{n=1}^{t-1} \xi_{0,n} \sigma_{t-n} z_{t-n}$ is a martingale difference sequence with respect to the filtration \mathcal{F}_t . Here we have defined $\xi_{0,n} := [-n^{-1}, \gamma_n(\psi_0)]'$, which satisfies $\|\xi_{0,n}\| \leq Kn^{-1}$ by Assumption C, see (2).

To apply the central limit theorem for martingales, see Lemma A.1, we first verify the Lindeberg condition via Lyapunov's sufficient condition that $\sum_{t=1}^T E\|v_{Tt}\|^{2+\epsilon} \rightarrow 0$ for some

$\epsilon > 0$. Thus,

$$\begin{aligned} E\|v_{Tt}\|^{2+\epsilon} &= E\left((2T^{-1/2})^{2+\epsilon} |\sigma_t z_t|^{2+\epsilon} \left\| \sum_{n=1}^{t-1} \xi_{0,n} \sigma_{t-n} z_{t-n} \right\|^{2+\epsilon} \right) \\ &\leq KT^{-1-\epsilon/2} E\left(|z_t|^{2+\epsilon} \left(\sum_{n=1}^{t-1} n^{-1} |z_{t-n}| \right)^{2+\epsilon} \right) \end{aligned}$$

using that the σ_t 's are bounded by Assumption A(a) and that $\|\xi_{0,n}\| \leq Kn^{-1}$. From Minkowski's inequality we find $E(\sum_{n=1}^{t-1} |z_t| n^{-1} |z_{t-n}|)^{2+\epsilon} \leq (\sum_{n=1}^{t-1} (E(|z_t| n^{-1} |z_{t-n}|)^{2+\epsilon})^{1/(2+\epsilon)})^{2+\epsilon}$ such that

$$\begin{aligned} E\|v_{Tt}\|^{2+\epsilon} &\leq KT^{-1-\epsilon/2} \left(\sum_{n=1}^{t-1} (E(|z_t| n^{-1} |z_{t-n}|)^{2+\epsilon})^{1/(2+\epsilon)} \right)^{2+\epsilon} \\ &\leq KT^{-1-\epsilon/2} \left(\sum_{n=1}^{t-1} n^{-1} \right)^{2+\epsilon} \leq KT^{-1-\epsilon/2} (\log T)^{2+\epsilon} \end{aligned}$$

where the second inequality is due to Assumption E provided ϵ is chosen such that $2\epsilon + 4 \leq 8$.

Therefore,

$$\sum_{t=1}^T E\|v_{Tt}\|^{2+\epsilon} \leq KT^{-\epsilon/2} (\log T)^{2+\epsilon} \rightarrow 0. \quad (59)$$

Next, the sum of squares of v_{Tt} is equal to

$$\begin{aligned} &4T^{-1} \sum_{t=1}^T \sigma_t^2 z_t^2 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} \sigma_{t-n} \sigma_{t-m} z_{t-n} z_{t-m} \\ &= 4T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} \sigma_{t-n} \sigma_{t-m} \tau_{nm} \end{aligned} \quad (60)$$

$$+ 4T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} \sigma_{t-n} \sigma_{t-m} (z_t^2 z_{t-n} z_{t-m} - \tau_{nm}). \quad (61)$$

By Lemma A.3 with $g_{t,n,m} = \tau_{nm}$, (60) is

$$\begin{aligned} &4T^{-1} \sum_{t=1}^T \sigma_t^4 \sum_{n,m=1}^{t-1} \xi_{0,n} \xi'_{0,m} \tau_{nm} + o(1) \\ &= 4T^{-1} \sum_{t=1}^T \sigma_t^4 \sum_{n,m=1}^{T-1} \xi_{0,n} \xi'_{0,m} \tau_{nm} - 4T^{-1} \sum_{t=1}^T \sigma_t^4 \sum_{n,m=t}^{T-1} \xi_{0,n} \xi'_{0,m} \tau_{nm} + o(1), \end{aligned}$$

where the first term is $4A_0 T^{-1} \sum_{t=1}^T \sigma_t^4 (1 + o(1)) \rightarrow 4A_0 \int_0^1 \sigma^4(s) ds$ and the second term is bounded by

$$KT^{-1} \sum_{t=1}^T \sum_{n,m=t}^{T-1} j^{-1} k^{-1} \tau_{nm} \leq KT^{-1} \sum_{t=1}^T t^{-2} \sum_{n,m=t}^{T-1} \tau_{nm},$$

which converges to zero by Assumption A(b)(iii).

The second moment of the (i, j) 'th element of (61) is

$$\begin{aligned}
& 16T^{-2} \sum_{t,s=1}^T \sigma_t^2 \sigma_s^2 \sum_{n,m=1}^{s-1} \sum_{k,l=1}^{t-1} (\xi_{0,m})_i (\xi_{0,n})_j (\xi_{0,k})_i (\xi_{0,l})_j \sigma_{s-n} \sigma_{s-m} \sigma_{t-k} \sigma_{t-l} \text{Cov}(z_t^2 z_{t-k} z_{t-l}, z_s^2 z_{s-n} z_{s-m}) \\
& \leq KT^{-2} \sum_{t,s=1}^T \sum_{n,m=1}^{s-1} \sum_{k,l=1}^{t-1} m^{-1} n^{-1} k^{-1} l^{-1} |\text{Cov}(z_t^2 z_{t-k} z_{t-l}, z_s^2 z_{s-n} z_{s-m})| \\
& = KT^{-2} \sum_{t=1}^T \sum_{n,m=1}^{t-1} \sum_{k,l=1}^{t-1} m^{-1} n^{-1} k^{-1} l^{-1} |\text{Cov}(z_t^2 z_{t-n} z_{t-m}, z_t^2 z_{t-k} z_{t-l})| \tag{62}
\end{aligned}$$

$$+ KT^{-2} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{n,m=1}^{s-1} \sum_{k,l=1}^{t-1} m^{-1} n^{-1} k^{-1} l^{-1} |\text{Cov}(z_t^2 z_{t-k} z_{t-l}, z_s^2 z_{s-n} z_{s-m})|. \tag{63}$$

For (62) we find the simple bound

$$KT^{-2} \sum_{t=1}^T \left(\sum_{k=1}^{t-1} k^{-1} \right)^4 \leq KT^{-1} (\log T)^4 \rightarrow 0$$

because z_t has finite eighth order moments by Assumption E. The covariance in (63) is a combination of the cumulants of z_t up to order eight, where each term is a product of two cumulants whose orders sum to eight. For the eighth order cumulant we find

$$T^{-2} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{n,m=1}^{s-1} \sum_{k,l=1}^{t-1} n^{-1} m^{-1} k^{-1} l^{-1} |\kappa_8(t, t, t-k, t-l, s, s, s-n, s-m)| \leq KT^{-1} \rightarrow 0$$

by Assumption E. There are no seventh order cumulants in (63) because they would be multiplied by a first order cumulant, which is zero. For products of sixth and second order cumulants we find, for example,

$$\begin{aligned}
& T^{-2} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{n,m=1}^{s-1} \sum_{k,l=1}^{t-1} n^{-1} m^{-1} k^{-1} l^{-1} \kappa_2(t-k, t-l) |\kappa_6(t, t, s, s, s-n, s-m)| \\
& = T^{-2} \sum_{t=2}^T \left(\sum_{s=1}^{t-1} \sum_{n,m=1}^{s-1} n^{-1} m^{-1} |\kappa_6(t, t, s, s, s-n, s-m)| \right) \left(\sum_{k=1}^{t-1} k^{-2} \kappa_2(t-k, t-k) \right) \\
& \leq KT^{-1} \rightarrow 0
\end{aligned}$$

by Assumption E. Another example is

$$\begin{aligned}
& T^{-2} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{n,m=1}^{s-1} \sum_{k,l=1}^{t-1} n^{-1} m^{-1} k^{-1} l^{-1} \kappa_2(t, t) |\kappa_6(t-k, t-l, s, s, s-n, s-m)| \\
& \leq KT^{-2} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{1 \leq m \leq n \leq s-1} \sum_{1 \leq l \leq k \leq t-1} n^{-1} m^{-1} k^{-1} l^{-1} \kappa_2(t, t) |\kappa_6(t-k, t-l, s, s, s-n, s-m)| \\
& \leq KT^{-2} \sum_{t=2}^T \sum_{s=1}^{t-1} \sum_{1 \leq m \leq n \leq s-1} \sum_{t-s \leq l \leq k \leq t-1} n^{-1} m^{-1} k^{-1} l^{-1} \kappa_2(t, t) |\kappa_6(t-k, t-l, s, s, s-n, s-m)| \\
& \quad + KT^{-2} \sum_{t=2}^T \sum_{1 \leq l \leq k \leq t-1} \sum_{s=1}^{t-l-1} \sum_{1 \leq m \leq n \leq s-1} n^{-1} m^{-1} k^{-1} l^{-1} \kappa_2(t, t) |\kappa_6(t-k, t-l, s, s, s-n, s-m)| \\
& \leq KT^{-2} \sum_{s=1}^{T-1} \sum_{t=s+1}^T \sum_{1 \leq m \leq n \leq s-1} \sum_{t-s \leq l \leq k \leq t-1} n^{-1} m^{-1} k^{-1} l^{-1} |\kappa_6(s, s, t-k, t-l, s-n, s-m)| \\
& \quad + KT^{-2} \sum_{t=2}^T \sum_{k=1}^{t-1} \sum_{s=1}^{t-l-1} \sum_{1 \leq m \leq n \leq s-1} n^{-1} m^{-1} k^{-2} |\kappa_6(t-k, t-k, s, s, s-n, s-m)|
\end{aligned}$$

using Lemma A.2. Here, the second term is clearly $O(T^{-1})$ by Assumption E and the first term is

$$\begin{aligned}
& T^{-2} \sum_{s=1}^{T-1} \sum_{t=s+1}^T \sum_{1 \leq m \leq n \leq s-1} \sum_{t-s \leq l \leq k \leq t-1} n^{-1} m^{-1} k^{-1} l^{-1} |\kappa_6(s, s, t-k, t-l, s-n, s-m)| \\
& = T^{-2} \sum_{s=1}^{T-1} \sum_{t=s+1}^T \sum_{1 \leq m \leq n \leq s-1} \sum_{0 \leq u \leq v \leq s-1} n^{-1} m^{-1} (v-s+t)^{-1} (u-s+t)^{-1} \\
& \quad \times |\kappa_6(s, s, s-v, s-u, s-n, s-m)| \\
& \leq T^{-2} \sum_{s=1}^{T-1} \sum_{1 \leq m \leq n \leq s-1} \sum_{0 \leq u \leq v \leq s-1} n^{-1} m^{-1} |\kappa_6(s, s, s-v, s-u, s-n, s-m)| \left(\sum_{t=s+1}^T (t-s)^{-2} \right),
\end{aligned}$$

which is also $O(T^{-1})$ using Assumption E. The remaining products of sixth and second order cumulants, as well as products of lower order cumulants, are treated similarly.

It follows that the sum of squares of v_{Tt} satisfies

$$4T^{-1} \sum_{t=1}^T \sigma_t^2 z_t^2 \sum_{m,n=1}^{t-1} \xi_{0,m} \xi_{0,n}^t \sigma_{t-m} \sigma_{t-n} z_{t-m} z_{t-n} \xrightarrow{p} 4A_0 \int_0^1 \sigma^4(s) ds.$$

Hence, by the central limit theorem for martingales, see Lemma A.1, we have that $S_T \xrightarrow{D} N(0, 4A_0 \int_0^1 \sigma^4(s) ds)$ and therefore also that $T^{1/2} \frac{\partial R(\theta_0)}{\partial \theta} \xrightarrow{D} N(0, 4A_0 \int_0^1 \sigma^4(s) ds)$.

E.2 Convergence of the Hessian

The second derivative in (56) is tight (stochastically equicontinuous) by Newey (1991, Corollary 2.2) if its derivative is dominated uniformly in $d \in D_3, \psi \in \mathcal{N}_\delta(\psi_0)$ by a random variable $B_T = O_P(1)$. From Lemma C.3 with $u_1 = u_2 = d - d_0 \geq -1/2 + \kappa_2$ and $\tilde{\Psi} = \mathcal{N}_\delta(\psi_0)$

(noting that only summability of the linear coefficients is assumed in Lemma C.3 and this is satisfied uniformly on $\mathcal{N}_\delta(\psi_0)$ by the derivatives of $c_n(\psi)$ by Assumption F) it holds that $B_T = \sup_{d \in D_3, \psi \in \mathcal{N}_\delta(\psi_0)} |\frac{\partial^3 R(\theta)}{\partial \theta^{(3)}}| = O_P(1)$, showing that the second derivative in (56) is tight. This result, together with consistency of $\hat{\theta}$ (Theorem 1), implies by Lemma A.3 of Johansen and Nielsen (2010) that the second derivative in (56) can be evaluated at the true value.

Hence, we examine

$$\frac{\partial^2 R(\theta_0)}{\partial \theta \partial \theta'} = 2T^{-1} \sum_{t=1}^T \varepsilon_t(\theta_0) \hat{y}_{2,t-1} + 2T^{-1} \sum_{t=1}^T \hat{y}_{1,t-1} \hat{y}'_{1,t-1},$$

and by the same argument as for the score, it is enough to consider

$$H_T := 2T^{-1} \sum_{t=1}^T \varepsilon_t \hat{y}_{2,t-1} + 2T^{-1} \sum_{t=1}^T y_{1,t-1} y'_{1,t-1}.$$

Because $\hat{y}_{2,t-1}$ is measurable with respect to \mathcal{F}_t , $\varepsilon_t \hat{y}_{2,t-1}$ is a martingale difference sequence, and it has finite variance such that the first term of H_T is $o_P(1)$.

The second term of H_T has mean

$$\begin{aligned} 2T^{-1} \sum_{t=1}^T \sum_{n=1}^{t-1} \sigma_{t-n}^2 \xi_{0,n} \xi'_{0,n} &= 2T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n=1}^{t-1} \xi_{0,n} \xi'_{0,n} + o(1) \\ &= 2T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n=1}^{T-1} \xi_{0,n} \xi'_{0,n} - 2T^{-1} \sum_{t=1}^T \sigma_t^2 \sum_{n=t}^{T-1} \xi_{0,n} \xi'_{0,n} + o(1) \end{aligned}$$

by Lemma A.3 with $g_{t,n,m} = \mathbb{I}(n = m) \sigma_t^{-2}$. The first term converges to $2B_0 \int_0^1 \sigma^2(s) ds$ and the second term is bounded by $KT^{-1} \sum_{t=1}^T \sum_{n=t}^{T-1} \|\xi_{0,n}\|^2 \leq KT^{-1} \sum_{t=1}^T \sum_{n=t}^{T-1} n^{-2} \leq KT^{-1}(\log T) \rightarrow 0$. The variance of the (i, j) 'th element of the second term of H_T is

$$\begin{aligned} &4T^{-2} \sum_{t,s=1}^T \sum_{m,n=1}^{s-1} \sum_{k,l=1}^{t-1} (\xi_{0,m})_i (\xi_{0,n})_j (\xi_{0,k})_i (\xi_{0,l})_j \sigma_{s-m} \sigma_{s-n} \sigma_{t-k} \sigma_{t-l} \\ &\quad \times [E(z_{s-m} z_{s-n} z_{t-k} z_{t-l}) - E(z_{s-m} z_{s-n}) E(z_{t-k} z_{t-l})] \\ &\leq KT^{-2} \sum_{t,s=1}^T \sum_{m,n=1}^{s-1} \sum_{k,l=1}^{t-1} m^{-1} n^{-1} k^{-1} l^{-1} |E(z_{s-m} z_{s-n} z_{t-k} z_{t-l}) - E(z_{s-m} z_{s-n}) E(z_{t-k} z_{t-l})|. \end{aligned}$$

If either $m \neq n$ or $k \neq l$, the contribution to the variance of the second term of H_T is bounded

by

$$\begin{aligned}
& KT^{-2} \sum_{t,s=1}^T \sum_{m,n=1}^{s-1} \sum_{k,l=1}^{t-1} m^{-1} n^{-1} k^{-1} l^{-1} |E(z_{s-m} z_{s-n} z_{t-k} z_{t-l})| \\
& \leq KT^{-2} \sum_{t,s=1}^T \sum_{n=1}^{s-1} \sum_{m=s-t+1}^{s-1} \sum_{k=1}^{t-1} m^{-1} n^{-1} k^{-1} (t-s+m)^{-1} |\kappa_4(s-m, s-m, s-n, t-k)| \\
& \quad + KT^{-2} \sum_{t,s=1}^T \sum_{n=1}^{s-1} \sum_{k,l=1}^{t-1} n^{-2} k^{-1} l^{-1} |\kappa_4(s-n, s-n, t-k, t-l)| \\
& \leq KT^{-2} \sum_{t,s=1}^T \sum_{m=s-t+1}^{s-1} m^{-1} (t-s+m)^{-1} + KT^{-1} \sum_{s=1}^T \sum_{n=1}^{s-1} n^{-2} \\
& \leq KT^{-2} (\log T) \sum_{t,s=1}^T (|s-t|+1)^{-1} + KT^{-1} \sum_{s=1}^T s^{-1} \leq KT^{-1} (\log T)^2 \rightarrow 0,
\end{aligned}$$

whereas if $m = n$ and $k = l$, the contribution is bounded by

$$\begin{aligned}
& KT^{-2} \sum_{t,s=1}^T \sum_{n=1}^{s-1} \sum_{k=1}^{t-1} n^{-2} k^{-2} |\kappa_2(s-n, s-n)| |\kappa_2(t-k, t-k)| \\
& \leq KT^{-2} \sum_{t,s=1}^T \sum_{n=1}^{s-1} \sum_{k=1}^{t-1} n^{-2} k^{-2} \leq KT^{-2} \sum_{t,s=1}^T s^{-1} t^{-1} \leq KT^{-1} (\log T)^2 \rightarrow 0.
\end{aligned}$$

Thus, the second term of H_T converges in L_2 -norm, and hence in probability, to $2B_0 \int_0^1 \sigma^2(s) ds$.

References

- Abadir, K.M., W. Distaso and L. Giraitis (2007), Nonstationarity-extended local Whittle estimation, *Journal of Econometrics* 141, 1353–1384.
- Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.
- Cavaliere, G., M.Ø. Nielsen, and A.M.R. Taylor (2013), Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets, QED working paper 1309, Queen’s University.
- Cavaliere, G. and A.M.R. Taylor (2005), Stationarity tests under time-varying variances, *Econometric Theory* 21, 1112–1129.
- Cavaliere, G. and A.M.R. Taylor (2008), Bootstrap unit root tests for time series with non-stationary volatility, *Econometric Theory* 24, 43–71.
- Cavaliere, G. and A.M.R. Taylor (2009), Heteroskedastic time series with a unit root, *Econometric Theory* 25, 1228–1270.
- Dahlhaus, R. (1989), Efficient parameter estimation for self-similar processes, *Annals of Statistics* 17, 1749–1766.
- Demetrescu, M., V. Kuzin and U. Hassler (2008), Long memory testing in the time domain, *Econometric Theory* 24, 176–215.

- Fox, R. and M.S. Taqqu (1986), Large-sample properties of parameter estimates for strongly dependent stationary gaussian series, *Annals of Statistics* 14, 517–532.
- Giraitis, L. and D. Surgailis (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.
- Gonçalves, S. and L. Kilian (2004), Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* 123, 89–120.
- Gonçalves, S. and L. Kilian (2007), Asymptotic and bootstrap inference for $AR(\infty)$ processes with conditional heteroskedasticity, *Econometric Reviews* 26, 609–641.
- Hannan, E.J. (1973), The asymptotic theory of linear time-series models, *Journal of Applied Probability* 10, 130–145.
- Hannan, E.J. and C. C. Heyde (1972), On limit theorems for quadratic functions of discrete time series, *The Annals of Mathematical Statistics* 43, 2058–2066.
- Hannan, E. J. and M. Deistler (1988), *The Statistical Theory of Linear Systems*, Wiley, New York.
- Hassler, U., P.M.M. Rodrigues and A. Rubia (2009), Testing for general fractional integration in the time domain, *Econometric Theory* 25, 1793–1828.
- Hosoya, Y. (1996), The quasi-likelihood approach to statistical inference on multiple time-series with long-range dependence, *Journal of Econometrics* 73, 217–236.
- Hualde, J. and P.M. Robinson (2011), Gaussian pseudo-maximum likelihood estimation of fractional time series models, *Annals of Statistics* 39, 3152–3181.
- Johansen, S. and M.Ø. Nielsen (2010), Likelihood inference for a nonstationary fractional autoregressive model, *Journal of Econometrics* 158, 51–66.
- Johansen, S. and M.Ø. Nielsen (2012), Likelihood inference for a fractionally cointegrated vector autoregressive model, *Econometrica* 80, 2667–2732.
- Kew, H. and D. Harris (2009), Heteroskedasticity-robust testing for a fractional unit root, *Econometric Theory* 25, 1734–1753.
- Kim, C.-J. and C.R. Nelson (1999), Has the U.S. economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economics and Statistics* 81, 608–616.
- Loretan, M. and P.C.B. Phillips (1994), Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets, *Journal of Empirical Finance* 1, 211–248.
- McConnell, M.M. and G. Perez-Quiros (2000), Output fluctuations in the United States: what has changed since the early 1980s?, *American Economic Review* 90, 1464–1476.
- McLeish, D.L. (1974), Dependent central limit theorems and invariance principles, *Annals of Statistics* 2, 620–628.
- Newey, W.K. (1991), Uniform convergence in probability and stochastic equicontinuity, *Econo-*

- metrica* 59, 1161–1167.
- Nielsen, M.Ø. (2004), Efficient likelihood inference in nonstationary univariate models, *Econometric Theory* 20, 116–146.
- Nielsen, M.Ø. (2014), Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time series models, QED working paper 1259, Queen’s University.
- Phillips, P.C.B. and V. Solo (1992), Asymptotics for linear processes, *Annals of Statistics* 20, 971–1001.
- Phillips, P.C.B. and K.-L. Xu (2006), Inference in autoregression under heteroskedasticity, *Journal of Time Series Analysis* 27, 289–308.
- Robinson, P.M. (1991), Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, *Journal of Econometrics* 47, 67–84.
- Robinson, P.M. (1994), Efficient tests of nonstationary hypotheses, *Journal of the American Statistical Association* 89, 1420–1437.
- Robinson, P.M. (2006), Conditional-sum-of-squares estimation of models for stationary time series with long memory, in H.-C. Ho, C.-K. Ing & T.L. Lai, eds, *Time Series and Related Topics: In Memory of Ching-Zong Wei*, IMS Lecture Notes–Monograph Series Vol. 52, pp. 130–137.
- Shimotsu, K. and P.C.B. Phillips (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.
- Sensier, M. and D. van Dijk (2004), Testing for volatility changes in U.S. macroeconomic time series, *Review of Economics and Statistics* 86, 833–839.
- Stock, J.H. and M.W. Watson (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in R.F. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford: Oxford University Press, 1–44.
- van der Vaart, A.W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.
- Zygmund, A. (2003), *Trigonometric Series (3rd Edition)*, Cambridge University Press, Cambridge.