Queen's Economics Department Working Paper No. 1324

Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form

Giuseppe Cavaliere
University of Bologna

Morten Ørregaard 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

11-2016

Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form

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

November 2016

Abstract

We consider the problem of conducting estimation and inference on the parameters of univariate heteroskedastic fractionally integrated time series models. We first extend existing results in the literature, developed for conditional sum-of-squares estimators in the context of parametric fractional time series models driven by conditionally homoskedastic shocks, to allow for conditional and unconditional heteroskedasticity both of a quite general and unknown form. Global consistency and asymptotic normality are shown to still obtain; however, the covariance matrix of the limiting distribution of the estimator now depends on nuisance parameters derived both from the weak dependence and heteroskedasticity present in the shocks. We then investigate classical methods of inference based on the Wald, likelihood ratio and Lagrange multiplier tests for linear hypotheses on either or both of the long and short memory parameters of the model. The limiting null distributions of these test statistics are shown to be non-pivotal under heteroskedasticity, while that of a robust Wald statistic (based around a sandwich estimator of the variance) is pivotal. We show that wild bootstrap implementations of the tests deliver asymptotically pivotal inference under the null. We demonstrate the consistency and asymptotic normality of the bootstrap estimators, and further establish the global consistency of the asymptotic and bootstrap tests under fixed alternatives. Monte Carlo simulations highlight significant improvements in finite sample behaviour using the bootstrap in both heteroskedastic and homoskedastic environments. Our theoretical developments and Monte Carlo simulations include two bootstrap algorithms which are based on model estimates obtained either under the null hypothesis or unrestrictedly. Our simulation results suggest that the former is preferable to the latter, displaying superior size control yet largely comparable power.

Keywords: conditional/unconditional heteroskedasticity; conditional sum-of-squares; fractional integration; quasi-maximum likelihood estimation; wild bootstrap.

J.E.L. Classifications: C12, C13, C22.

*Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: `robert.taylor@essex.ac.uk`

1 Introduction

We consider model-based inference on both the long and short memory parameters in univariate fractionally integrated models driven by innovations which can display both unconditional heteroskedasticity (also referred to as non-stationary volatility) and conditional heteroskedasticity, both of a very general form unknown to the practitioner. Doing so is of considerable practical importance given the well-documented failure of the conditional homoskedasticity assumption 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 suggests that the assumption of constant *unconditional* volatility is also at odds with the data. For example, Sensier and van Dijk (2004) report that over 80% of the real and price variables in the well-known Stock and Watson (1999) data-set reject the null of constant innovation variance, while Amado and Teräsvirta (2014) report evidence against the constancy of unconditional variances in stock market returns. Many empirical studies also report a substantial decline, often referred to as the Great Moderation, in the unconditional volatility of the shocks driving macroeconomic series in the twenty years or so leading up to the Great Recession that started in late 2007, with a subsequent sharp increase again in volatility observed after 2007; see, *inter alia*, McConnell and Perez-Quiros (2000), Clark (2009), Stock and Watson (2012), and the references therein.

Fractional time series models allow for parsimonious and flexible modelling of a very wide range of dependence in time series data. However, this flexibility has meant that, until only recently, proofs of global consistency of standard parametric estimators, such as the conditional quasi-maximum likelihood (QML) estimator, or equivalently the conditional sum-of-squares (CSS) estimator, have been eschewed in the literature. Consistency results, while important in their own right, are also necessary prerequisites in any proof of asymptotic normality for implicitly defined estimators such as the QML estimator. The problem lies with 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) [HR11 hereafter] or Nielsen (2015) for detailed discussion on this. 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.[1] While earlier consistency results for the QML estimator have avoided these difficulties by, for example, restricting the range of values the long memory parameter can take to an interval of length less than one-half, or giving only local consistency proofs, HR11 demonstrate the consistency and asymptotic normality of the QML estimator for an arbitrarily large set of admissible values of the long memory parameter including stationary and non-stationary and invertible and non-invertible processes. They do so in the context of a fractional model driven by conditionally homoskedastic errors.

Our first contribution is to extend the results in HR11 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 (2005), Cavaliere and Taylor (2005, 2008), and Phillips and Xu (2006) as a special case and also includes a set of conditional heteroskedasticity conditions which are similar, but

[1] It should be noted 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 root-T parametric rate of convergence.

somewhat weaker, than those employed previously in the fractional integration literature by Robinson (1991), Demetrescu, Kuzin and Hassler (2008) and Hassler, Rodrigues and Rubia (2009), among others. Neither of these conditions involve specifying a parametric model for the volatility process. In the context of the resulting heteroskedastic fractional time series model, we demonstrate the consistency and asymptotic normality of the QML estimator.[2] We then show that the variance of the limiting distribution of the QML estimator depends on nuisance parameters which derive from the weak dependence present (as in the corresponding result in HR11), but also from the heteroskedasticity.

Our second contribution builds on our results for the QML estimator by investigating the impact of time-varying volatility on statistical inference in long memory series. The classical likelihood-based Wald, likelihood ratio (LR) and Lagrange multiplier (LM) tests for inference on the long memory parameter have been derived under the assumption of conditionally (and, hence, unconditionally) homoskedastic shocks; see, among others, Robinson (1994), Agiakloglou and Newbold (1994), Tanaka (1999), Nielsen (2004), Lobato and Velasco (2007), and Johansen and Nielsen (2010, 2012). A small number of papers have considered the case where the shocks can display certain forms of conditional heteroskedasticity (but maintaining the assumption of unconditional homoskedasticity); see, for example, Robinson (1991), Baillie, Chung, Tieslau (1996), Ling and Li (1997), Ling (2003), Demetrescu, Kuzin and Hassler (2008) and Hassler, Rodrigues and Rubia (2009). Allowing for non-stationary volatility of a similar form to that considered in this paper, Kew and Harris (2009) extend the idea of Demetrescu, Kuzin and Hassler (2008) to use heteroskedasticity-robust White (1980)-type standard errors when computing regression-based tests for fractional integration. They apply this approach to the tests proposed in Agiakloglou and Newbold (1994), Breitung and Hassler (2002), Dolado, Gonzalo and Mayoral (2002) and Lobato and Velasco (2006, 2007). In earlier work, reported in Cavaliere, Nielsen and Taylor (2015) [CNT15 henceforth], we discuss one-sided and two-sided LM tests, like the foregoing authors, for hypotheses on just the (scalar) long memory parameter of the model and develop wild bootstrap implementations of these tests which we showed to be robust to heteroskedasticity of the form considered there.

We extend the analysis of CNT15 to analyse the impact of time-varying volatility on the classical trinity of tests for linear restrictions on the long and/or short memory parameters. This allows us to test hypotheses on, for example: the order of integration of the series; the autoregressive and moving average (ARMA) lag lengths, in cases where the short memory component is specialised to an ARMA structure, thereby developing specification tests for the form of the short memory component; joint tests on both allowing, for example, the null hypothesis that the series is white noise to be tested. In the case where one restriction is being tested, we also suggest additional tests for one-sided alternatives, an example of which is testing the null hypothesis of stationarity. The extension of the results for the LM test in CNT15 to the Wald and LR tests requires the global consistency result on the parameter estimates from the fitted fractional model, not required for the LM test which estimates the model only under the null. The limiting distributions of the classical test statistics under both the null and local alternatives are shown to be non-pivotal, their functional form depending on nuisance parameters

[2]Our results necessarily also apply to the special case of short memory processes, where the long memory parameter is $d = 0$, driven by conditionally and/or unconditionally heteroskedastic innovations. For earlier contributions relevant to the $d = 0$ case see, e.g., Gonçalves and Kilian (2004) who allow for conditional (but not unconditional) heteroskedasticity, and Phillips and Xu (2006) who allow for unconditional (but not conditional) heteroskedasticity in the context of stationary finite-order autoregressions.

deriving from the heteroskedasticity. We demonstrate that, when using the standard χ^2 asymptotic critical values (as are appropriate under homoskedasticity), this leads to tests which are not in general asymptotically correctly sized. The size-corrected power of the trinity of the tests is also shown to depend on nuisance parameters deriving from the volatility process. We further demonstrate that a robust version of the Wald test, implemented with a heteroskedasticity-robust standard error, is asymptotically corrected sized when using standard critical values, and has an asymptotic local power function which coincides with the size-corrected asymptotic local power of the standard tests. We also demonstrate the consistency of the classical tests against fixed alternatives.

As a result of the dependence of their limiting null distributions on nuisance parameters derived from the heteroskedasticity, we develop wild bootstrap implementations of the likelihood-based tests. These are shown to correctly replicate the (first-order) asymptotic null distributions of the standard test statistics. As a consequence, asymptotically valid bootstrap inference can be performed in the presence of time-varying volatility using either the wild bootstrap versions of these tests or the robust Wald test. Again, establishing these results for the bootstrap Wald, LR and robust Wald tests necessitates global consistency proofs for the estimates from the bootstrap data, not required for LM tests, which we therefore provide. The latter presents several new theoretical challenges, not required in the recent work of HR11 or Nielsen (2015). We also establish global consistency results for the bootstrap tests.

We analyse two possible bootstrap algorithms. The first is based on model estimates obtained the null hypothesis, and for the case of testing hypotheses solely on the long memory parameter is similar to the bootstrap algorithm proposed in CNT15, except that the wild bootstrap scheme considered in CNT15 is applied to restricted residuals (obtained estimating the null model), while here we apply the wild bootstrap to the corresponding unrestricted residuals. The second is based on unrestricted estimates of the model parameters. In much of the bootstrap hypothesis testing literature the advice is to always use the former. We show that both deliver asymptotically valid tests in the present setting. Our simulation results suggest that tests based on the restricted bootstrap display superior size control to those based on the unrestricted bootstrap (especially where short memory ARMA effects are present), yet the two deliver largely comparable finite sample power properties. As such, we therefore recommend the use of the restricted bootstrap. More generally, our simulation results highlight the superior finite sample properties of our proposed bootstrap tests over their asymptotic counterparts, in both heteroskedastic and homoskedastic environments, and even more so where weak dependence is also present.

The remainder of the paper is structured as follows. In Section 2 we present the heteroskedastic fractional time series model and the main assumptions. Section 3 presents the QML estimator together with the associated classical likelihood-based (asymptotic) tests and our wild bootstrap implementations of these tests. The large sample properties of these estimators and tests are provided in Section 4. Results relating to tests on the long memory parameter from a Monte Carlo simulation study are given in Section 5. Section 6 concludes and the paper ends with four mathematical appendices containing preliminary lemmas, variation bounds, and the proofs of Theorems 5 and 6. Additional Monte Carlo results relating to further tests within the fractional model, proofs of all lemmas, and proofs of Theorems 1–4, 7, and 8 appear in a Supplement.

Notation. In the following, for any set S, $\text{int}(S)$ denotes the interior of S. For any matrix, M, $\|M\|^2 := \text{tr}\{M'M\}$ and $(M)_{m,n}$ denotes its (m,n)'th element; for any vector, v, $\|v\| := (v'v)^{1/2}$, the Euclidean norm, and $(v)_m$ or v_m denote its m'th element. A function $f(x) : \mathbb{R}^q \to \mathbb{R}$ satisfies a Lipschitz condition of order α, or is in Lip(α), if there exists a finite constant $K > 0$ such that $|f(x_1) - f(x_2)| \leq K||x_1 - x_2||^\alpha$ for all $x_1, x_2 \in \mathbb{R}^q$. Throughout the paper we use the notation c or K for a generic, finite constant, and, as a convention, it is assumed that $j^{-1} = 0$ for $j = 0$ in summations over j. We use $\xrightarrow{w}$, $\xrightarrow{p}$ and $\xrightarrow{L_r}$ to denote convergence in distribution, in probability, and in L_r-norm, respectively, in each case as $T \to \infty$, where the L_r-norm is defined for a scalar random variable X as $||X||_r := (E|X|^r)^{1/r}$ and T is the sample size. The probability and expectation conditional on the realisation of the original sample is denoted P^* and E^*, respectively. Moreover, for a given sequence X_T^* computed on the bootstrap data, $X_T^* \xrightarrow{p^*}_p 0$ or $X_T^* = o_p^*(1)$ denote that $P^*(|X_T^*| > \epsilon) \to 0$ in probability for any $\epsilon > 0$, $X_T^* = O_p^*(1)$ denotes that there exists a $K > 0$ such that $P^*(|X_T^*| > K) \to 0$ in probability, and $\xrightarrow{w^*}_p$ denotes weak convergence in probability, again in each case as $T \to \infty$.

2 The Heteroskedastic Fractional Model

We consider the fractional time series model

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

where the operator Δ_+^{-d} is given, for a generic variable x_t, by $\Delta_+^{-d} x_t := \Delta^{-d} x_t \mathbb{I}(t \geq 1) = \sum_{n=0}^{t-1} \pi_n(d) x_{t-n}$, where $\mathbb{I}(\cdot)$ denotes the indicator function, and with $\pi_n(d) := \frac{\Gamma(d+n)}{\Gamma(d)\Gamma(1+n)} = \frac{d(d+1)\ldots(d+n-1)}{n!}$, denoting the coefficients in the usual binomial expansion of $(1-z)^{-d}$, and where ψ is a p-dimensional parameter vector and $a(z, \psi) := \sum_{n=0}^{\infty} a_n(\psi) z^n$. We let $\theta := (d, \psi')'$ denote the full parameter vector.

Remark 2.1. The parametric form (but not the parameters characterising it) of the function $a(z, \psi)$ will be assumed known, so that u_t is a linear process governed by an underlying p-dimensional parameter vector. For example, finite order ARMA models are permitted, as is the Bloomfield (1973) exponential spectrum model applied by Robinson (1994) in a fractional setting. Further discussion on the function $a(z, \psi)$ can be found in HR11 and Nielsen (2015). Thus, our focus is model-based inference (which might be on the long memory parameter, d, the short memory parameter, ψ, or jointly on both). As such we assume a statistical model characterised by a finite-dimensional vector of parameters and the objective is one of estimation and inference on those parameters. $\diamondsuit$

Remark 2.2. The model in (1) is that of 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 HR11. $\diamondsuit$

The main focus in this paper is to test the null hypothesis

$$H_0 : M'\theta = m \tag{2}$$

against the two-sided alternative

$$H_1 : M'\theta \neq m, \tag{3}$$

in (1). Here, M is a $(p+1) \times q$ full-rank matrix of constants defining q (linearly independent) restrictions on the parameter vector θ, and m is a $q \times 1$ vector of constants. An obvious example involves testing hypotheses on the long memory parameter d, important cases thereof are $d = 0$ (short memory) and $d = 1$ (unit root). As a second example, testing hypotheses on the elements of ψ could be used for order determination for the short memory dynamics, such as establishing an autoregressive order. Finally, joint hypotheses involving both d and ψ can be tested; for example, $d = 1 \cap \psi = 0$ corresponds to the pure random walk hypothesis, while $d = 0 \cap \psi = 0$ yields white noise.

In order to demonstrate the consistency of the QML estimates from model (1), outlined in Section 3.1 below, the following set of conditions, labelled collectively as Assumption 1, are required to hold on ε_t. Some strengthening of these conditions, detailed later, will subsequently be needed when we derive asymptotic distribution theory for the QML estimators and associated tests and of their bootstrap analogues.

Assumption 1. *The innovations* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ *are such that* $\varepsilon_t = \sigma_t z_t$*, where* $\{z_t\}_{t\in\mathbb{Z}}$ *and* $\{\sigma_t\}_{t\in\mathbb{Z}}$ *satisfy the conditions in parts (a) and (b), respectively, below:*

(a) $\{z_t\}_{t\in\mathbb{Z}}$ *is a conditionally heteroskedastic martingale difference sequence [MDS] with respect to the natural filtration* $\mathcal{F}_t$*, the sigma-field generated by* $\{z_s\}_{s\le t}$*, such that* $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ *for* $t = ..., -1, 0, 1, 2, ...,$ *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* $r \ge 0, s \ge 0$,
  - *(iii) for all integers* $r_1, r_2 \ge 1$*, the 4'th 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* $\sup_t \sum_{r_1,r_2=1}^{\infty} |\kappa_q(t, t, t - r_1, t - r_2)| < \infty$.

(b) $\{\sigma_t\}_{t\in\mathbb{Z}}$ *is a random sequence which is stochastically independent of* $\{z_t\}$ *and satisfies*
  - *(i)* $\sup_{t\in\mathbb{Z}} \sigma_t < \infty$ *a.s.*,
  - *(ii) for all* $t = 1, ..., T$*,* $\sigma_t = \sigma(t/T)$*, where* $\sigma(\cdot) \in \mathcal{D}([0, 1])$*, the space of càdlàg functions on* $[0, 1]$*, and the random function* $\sigma(\cdot)$ *is strictly positive a.s.*

Assumption 1 allows for stochastic conditional and unconditional heteroskedasticity; both of unknown and very general forms. Note that, through Assumption 1(a)(iii), only four moments of $\{z_t\}$ are assumed finite in Assumption 1 and Gaussianity is not assumed. The scaling factor, $\{\sigma_t\}$, in Assumption 1(b) is possibly random and, hence, since we are not placing any restrictions on the moments of σ_t, the *unconditional* moments of ε_t are not required to be finite under Assumption 1. In the special case where σ_t is deterministic, $E(|\varepsilon_t|^r) < \infty$ if and only if $E(|z_t|^r) < \infty$. Where σ_t is deterministic, σ_t^2 denotes the unconditional variance of ε_t because $E(\varepsilon_t^2) = \sigma_t^2 E(z_t^2) = \sigma_t^2$.

Remark 2.3. Assumption 1(a) is closely related to Assumption A of Gonçalves and Kilian (2004), and allows for conditional heteroskedasticity in $\{z_t\}$ as well as (possibly asymmetric) volatility clustering and statistical leverage effects. Volatility clustering, such as GARCH, is allowed for by the fact that the quantity $\tau_{r,r}$ is not necessarily equal to $E(z_t^2)E(z_{t-r}^2) = 1$. In particular, Deo (2000) in his Lemmas 1 and 2, respectively, provides examples of stochastic volatility and GARCH processes satisfying Assumption 1(a). It should be stressed, however, that a parametric model, such as a member of the GARCH family, is not assumed. Statistical leverage is permitted under Assumption 1(a), which occurs when the quantity $E(z_t^2 z_{t-r})$ is non-zero for some $r \ge 1$, noting that $E(z_t^2 z_{t-i}) = E(h_t z_{t-i})$, where $h_t := E(z_t^2|\mathcal{F}_{t-1})$ is the conditional variance function. Likewise, asymmetric volatility clustering is allowed by non-zero $\tau_{r,s}$ for $r \ne s$.[3] ♢

[3]The asymmetry may arise from either the conditional volatility function or the innovation distri-

Remark 2.4. The conditions in Assumption 1(a) are typical, but rather weaker than, those used in this literature; e.g., Robinson (1991), Demetrescu, Kuzin and Hassler (2008), Hassler, Rodrigues and Rubia (2009) and Kew and Harris (2009). First, these authors 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 $E(z_t^2 z_{t-r}) = 0$ for $r \geq 1$ and $\tau_{r,s} = 0$ for $r \neq s$, thus ruling out both statistical leverage and asymmetric volatility clustering. Second, these authors assume strict stationarity of z_t, which we do not. Further discussion of these conditions can be found in, e.g., Demetrescu, Kuzin and Hassler (2008, pp. 179–180), Kew and Harris (2009, p. 1739) and CNT15. $\diamondsuit$

Remark 2.5. Part (b) of Assumption 1 implies that the scale factor σ_t 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 scale of ε_t. In particular, σ_t need not be deterministic (as in, e.g., CNT15, Cavaliere and Taylor, 2008), but is allowed to be stochastic, albeit independent of z_t. For example, models of single or multiple random scale shifts satisfy part (b) of Assumption 1 with $\sigma(u) = \upsilon_0^2 + \sum_{i=1}^{m}(\upsilon_i^2 - \upsilon_0^2)\mathbb{I}(u \geq \tau_i)$, where $(\upsilon_0^2, \ldots, \upsilon_m^2, \tau_1, \ldots, \tau_m)$ is a random vector such that there are m random scale shifts occurring at random times τ_i. (Piecewise) affine (random) functions are also permitted, thereby allowing for a scale factor which follows a (broken) linear trend. Further examples and discussion of part (b) of Assumption 1, relevant to the case where σ_t is non-stochastic, can be found in Cavaliere and Taylor (2008, 2009) and CNT15. $\diamondsuit$

Remark 2.6. Assumption 1 differs from Assumptions 1–3 in Cavaliere and Taylor (2009) who consider the problem of autoregressive unit root testing under heteroskedasticity. In particular, they assume that ε_t is a MDS whose conditional variance, say $v_t := E(\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, ...)$, is slowly changing over time, in the sense that it can be approximated by a càdlàg process as $T \to \infty$. That is, $v_{\lfloor T\cdot \rfloor} \stackrel{w}{\to} v(\cdot) \in D([0,1])$ as $T \to \infty$. While their assumption allows, e.g., for persistent changes and long swings in volatility, such as non-stationary stochastic volatility or near integrated GARCH processes (see their discussion on pp. 1233–1235), it has the drawback that it does not allow for short-run volatility changes, such as those generated by stationary GARCH processes or stationary stochastic volatility processes. The latter are allowed in our setting through Assumption 1(a) as discussed in Remark 2.3. Moreover, in addition to allowing for very general short-run volatility dynamics through z_t, our Assumption 1 also allows for persistent (albeit exogenous) changes of the volatility through the scale factor σ_t, which is only required to be described by a random process in D (Assumption 1(b)). $\diamondsuit$

Remark 2.7. An important special case of Assumption 1, which obtains for $\sigma(\cdot)$ non-random and constant and $\{z_t\}$ conditionally homoskedastic, is that the innovations $\{\varepsilon_t\}$ themselves form a conditionally homoskedastic MDS with respect to the filtration $\mathcal{F}_t$, where $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ almost surely and $\sup_t E(|\varepsilon_t|^q) \leq K < \infty$ for some $q \geq 4$. This is a standard conditional homoskedasticity assumption in the time series literature; see, for example, Hannan (1973) for an early reference. Even this special case of our assumption is, however, weaker than Assumption A2 in HR11, which additionally imposes

bution. A simple example of the latter, where analytical calculations appear feasible, is the following special case of the asymmetric power GARCH model of Ding, Granger, and Engle (1993), $z_t = h_t^{1/2} e_t$ with $h_t^{3/2} = \omega + \alpha h_{t-1}^{3/2}|e_{t-1}|^3$ and e_t being i.i.d.(0,1) with $Ee_t^3 = \mu_3 \neq 0$, and $E(|e_t|^4 sign(e_t)) = \mu_4$, which has $\tau_{0,1} = E(h_t^{3/2} e_t^3 h_{t-1}^{1/2} e_{t-1}) = \mu_3 \alpha E(h_{t-1}^2 |e_t|^4 sign(e_t)) = \mu_3 \alpha \mu_4 E h_{t-1}^2 \neq 0$.

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. $\diamondsuit$

Remark 2.8. The MDS assumption on z_t implies that for any $q \geq 2$, if the highest argument in the cumulant $\kappa_q(\cdot)$ appears only once, then the cumulant is zero. This result is formally stated in Lemma A.2. Hence, our assumptions deal only with cumulants where the two highest arguments coincide. Notice also that the moment condition $\sup_t E(|z_t|^4) < \infty$ is necessary for part (a)(iii) to hold and so is not stated explicitly. $\diamondsuit$

Remark 2.9. For $t \leq 0$, we assume only that σ_t is uniformly bounded, see Assumption 1(b)(i), whereas, for $t = 1, \ldots, T$, we assume in Assumption 1(b)(ii) that σ_t depends on (t/T). Therefore, a time series generated according to Assumption 1 formally constitutes an array of the type $\{\varepsilon_{T,t} : t \leq T, T \geq 1\}$, where $\varepsilon_{T,t} = \sigma_{T,t} z_t$ and $\sigma_{T,t}$ satisfies Assumption 1(b) for all $T \geq 1$. Because the array notation is not essential, for simplicity the subscript T is suppressed in the sequel. $\diamondsuit$

In anticipation of the asymptotic local power analysis of the likelihood-based tests, we prove consistency and asymptotic normality of the QML estimator under Pitman drift, where model (1) holds with true parameter value given by the sequence

$$\theta_{0,T} = \theta_0 + \delta_\theta / \sqrt{T}. \tag{4}$$

Corresponding results under a fixed true value, θ_0, as commonly stated in estimation settings, are trivially obtained by setting $\delta_\theta = 0$ in our results that follow. We note that, under the true parameter value $\theta_{0,T}$ in (4), u_t and X_t in (1) are formally in fact triangular arrays. However, this is not essential in what follows, so we suppress the additional subscript T (cf. Remark 2.9). We impose the following assumption on θ_0.

Assumption 2. *It holds that* $\theta_0 = (d_0, \psi_0')' \in D \times \Psi =: \Theta$, *where* $D := [d_1, d_2]$ *with* $-\infty < d_1 \leq d_2 < \infty$ *and the set* $\Psi \subset \mathbb{R}^p$ *is convex and compact.*

Assumption 2 permits the length of the interval D of admissible values of the parameter d to be arbitrarily large, allowing the model in (1) to 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 3. *For all* $\psi \in \Psi$ *and all* z *in the complex unit disk* $\{z \in \mathbb{C} : |z| \leq 1\}$ *it holds that:*

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

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

(iii) The function $\dot{a}(z, \psi) := \frac{\partial a(z,\psi)}{\partial \psi} = \sum_{n=0}^{\infty} \dot{a}_n(\psi) z^n$ *exists and* $\dot{a}(e^{i\lambda}, \psi)$ *is differentiable in* λ *with derivative in* $\mathrm{Lip}(\zeta)$ *for* $\zeta > 0$.

Remark 2.10. Assumption 3(i) coincides with Assumption A1(iv) of HR11, while Assumption 3(ii) strengthens their Assumption A1(ii) from once differentiable in λ with derivative in $\mathrm{Lip}(\zeta)$ for $\zeta > 1/2$, and Assumption 3(iii) strengthens their Assumption A1(iii) from continuity in ψ to differentiability. $\diamondsuit$

Remark 2.11. Assumptions 3(i)-(ii) ensure that u_t in (1) is an invertible short-memory process (with power transfer function (scale-free spectral density) that is bounded and bounded away from zero at all frequencies); under Assumption 3(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 $|z| \leq 1+\epsilon$ for some $\epsilon > 0$, and is also bounded and bounded away from zero on the complex unit disk. Under Assumption 3 the coefficients $a_n(\psi)$, $b_n(\psi)$, $\dot{a}_n(\psi)$, and $\dot{b}_n(\psi) := \frac{\partial b_n(\psi)}{\partial \psi}$ satisfy

$$|a_n(\psi)| = O(n^{-2-\zeta}), |b_n(\psi)| = O(n^{-2-\zeta}), ||\dot{a}_n(\psi)|| = O(n^{-1-\zeta}), ||\dot{b}_n(\psi)|| = O(n^{-1-\zeta}) \quad (5)$$

uniformly in $\psi \in \Psi$; see Zygmund (2003, pp. 46 and 71). In contrast, under Hualde and Robinson's (2011) Assumption A1(ii) the rate for $a_n(\psi)$ and $b_n(\psi)$ is $O(n^{-1-\zeta})$ for $\zeta > 1/2$, which is slightly weaker. Assumption 3 is easily seen to be satisfied, for example, by stationary and invertible finite order ARMA processes. ♢

Remark 2.12. Assumption 3 is assumed to apply for all ψ in the user-chosen optimizing set Ψ. For example, in the case where u_t is an ARMA model, the set Ψ can be chosen as any compact and convex subset of the (open) set for which the inverse roots of the AR and MA polynomials are strictly inside the unit circle. Specifically, if u_t is modeled as a first-order AR model then Assumption 3 is clearly satisfied for all $\psi \in (-1, 1)$, and the optimizing set Ψ can be chosen as any compact and convex subset of $(-1, 1)$. ♢

Finally, the following identification condition will also be needed.

Assumption 4. *For all $\psi \in \Psi \backslash \{\psi_0\}$ it holds that $a(z,\psi) \neq a(z,\psi_0)$ on a subset of $\{z \in \mathbb{C} : |z| = 1\}$ of positive Lebesgue measure.*

Assumption 4 is identical to Assumption D in Nielsen (2015) and Assumption A1(i) in HR11. It is satisfied, for example, by all stationary and invertible finite order ARMA processes whose AR and MA polynomials do not admit any common factors. More generally, Assumption 4 ensures identification and is related to the standard condition for identification in extremum (or ML) estimation; see, e.g., Hayashi (2000, eqn. (7.2.13)) for a textbook treatment. In a time series context, see also Hannan (1973, eqn. (4)).

3 Estimation and Inference

In Section 3.1 we first outline QML estimation of the parameters of model (1). Section 3.2 outlines the classical trinity of QML-based tests for inference on the parameters of model (1). Wild bootstrap implementations of these tests are discussed in Section 3.3.

3.1 QML Estimation

Define the residuals

$$\varepsilon_t(\theta) := \sum_{n=0}^{t-1} b_n(\psi)\Delta_+^d X_{t-n}. \quad (6)$$

Then the conditional[4] Gaussian quasi-log-likelihood function, based on the assumption that the shocks ε_t are Gaussian with constant variance equal to σ^2, for model (1) is, up to a constant, $L_T(\theta,\sigma^2) = -\frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t(\theta)^2$. 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,

$$\hat{\theta} := \arg\min_{\theta\in\Theta} Q_T(\theta), \quad (7)$$

$$Q_T(\theta) := T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta)^2. \quad (8)$$

[4]We 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 2.2.

It is important to notice that Gaussianity is not needed for the asymptotic theory in this paper, and so (7) can be viewed as a (conditional) QML estimator.

Next, calculation of both the LR and LM test statistics, requires estimation carried out under the null hypothesis. To that end, let a tilde (˜) denote an estimator obtained under the restriction(s) imposed by the null hypothesis, i.e. while fixing $M'\theta = m$. Specifically, the restricted estimator $\tilde{\theta} := (\tilde{d}, \tilde{\psi}')'$ is defined by $\tilde{\theta} := \arg\min_{\theta \in \Theta} Q_T(\theta)$ subject to $M'\theta = m$, and $\tilde{\sigma}^2 := \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t(\tilde{\theta})^2$, see (8). Equivalently, define the subspace $\tilde{\Theta} := \{\theta \in \Theta : M'\theta = m\}$ such that

$$\tilde{\theta} = \arg\min_{\theta \in \tilde{\Theta}} Q_T(\theta). \tag{9}$$

3.2 Asymptotic Tests

We next present the classical trinity of likelihood-based test statistics, i.e. the (quasi) LR, LM and Wald test statistics under the assumption of homoskedastic Gaussian innovations, as well as an heteroskedasticity-robust version of the Wald test implemented using a White (1982) sandwich-type robust standard error. First, we consider a Wald statistic given by

$$W_T := T(M'\hat{\theta} - m)'(M'\hat{B}^{-1}M)^{-1}(M'\hat{\theta} - m), \tag{10}$$

where $\hat{B} := -\frac{\partial^2 L_T(\hat{\theta}, \hat{\sigma}^2)}{\partial\theta\partial\theta'}$ and $\hat{\sigma}^2 := T^{-1}\sum_{t=1}^{T} \varepsilon_t(\hat{\theta})^2$. Next, the (quasi) LR test statistic is given by

$$LR_T := T \log\left(Q_T(\tilde{\theta})/Q_T(\hat{\theta})\right), \tag{11}$$

as follows immediately from (8). Finally, the LM test statistic is

$$LM_T := \frac{\partial L_T(\tilde{\theta}, \tilde{\sigma}^2)}{\partial\theta'} \left(\frac{\partial^2 L_T(\tilde{\theta}, \tilde{\sigma}^2)}{\partial\theta\partial\theta'}\right)^{-1} \frac{\partial L_T(\tilde{\theta}, \tilde{\sigma}^2)}{\partial\theta}. \tag{12}$$

Remark 3.1. The LM statistic in (12) reduces to that considered in CNT15 in the case where $M = (1, 0, \ldots, 0)'$; that is, where the null hypothesis only involves d. Moreover, CNT15 considered a more restrictive formulation for the short memory component $a(z, \psi)$ in (1), and did not consider estimation of θ. ♢

In the remainder we will often use S_T as a generic notation for any of the three statistics W_T, LR_T and LM_T in (10), (11) and (12), respectively. As we show below, these three statistics are (conditional on $\sigma(\cdot)$) asymptotically equivalent under our assumptions.

The dependence of the asymptotic variance of the QML estimator, see Theorem 2 below, on nuisance parameters arising from both the weak dependence and heteroskedasticity in $\{\varepsilon_t\}$ implies that asymptotically pivotal inference on the parameter vector θ will need to be based around a heteroskedasticity-robust version of the Wald test statistic using, e.g., the usual White (1982) sandwich estimator $\hat{C} := \hat{B}^{-1}\hat{A}\hat{B}^{-1}$ with $\hat{A} := T^{-1}\sum_{t=1}^{T} \frac{\partial \ell_t(\hat{\theta}, \hat{\sigma}^2)}{\partial\theta} \frac{\partial \ell_t(\hat{\theta}, \hat{\sigma}^2)}{\partial\theta'}$ and $\ell_t(\theta, \sigma^2) := -\frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\varepsilon_t(\theta)^2$, which leads to the robust Wald statistic

$$RW_T := T(M'\hat{\theta} - m)'(M'\hat{C}M)^{-1}(M'\hat{\theta} - m). \tag{13}$$

Under conditional homoskedasticity, as in the conditions in Remark 2.7, it follows immediately from (16) that all the tests are asymptotically χ^2_q distributed under the null hypothesis. Where $M = (1, 0, \ldots, 0)'$, these results can be found in, for example, Robinson (1994), Tanaka (1999), Nielsen (2004) or HR11.

3.3 Bootstrap Tests

We now describe our two proposed wild bootstrap-based analogues of the LR, LM, Wald and robust Wald tests from Section 3.2.

Algorithm 1 (Restricted wild bootstrap)**.**

(i) Estimate model (1) using Gaussian QML (unrestricted), yielding the estimate $\hat{\theta}$ together with the corresponding residuals, $\hat{\varepsilon}_t := \varepsilon_t(\hat{\theta})$, see (6).

(ii) Compute the centered residuals $\hat{\varepsilon}_{c,t} := \hat{\varepsilon}_t - T^{-1}\sum_{i=1}^{T}\hat{\varepsilon}_i$ and construct the bootstrap errors $\varepsilon_t^* := \hat{\varepsilon}_{c,t}w_t$, where $w_t, t = 1, ..., T$, is an i.i.d. sequence with $E(w_t) = 0$, $E(w_t^2) = 1$ and $E(w_t^4) < \infty$.

(iii) Construct the bootstrap sample $\{X_t^*\}$ from

$$X_t^* := \Delta_+^{-\tilde{d}} u_t^* \text{ with } u_t^* := a(L, \tilde{\psi})\varepsilon_t^*, \quad t = 1, ..., T, \tag{14}$$

with the T bootstrap errors ε_t^* generated in step (ii) and setting $\varepsilon_t^* = 0$ for $t \leq 0$.

(iv) Using the bootstrap sample, $\{X_t^*\}$, compute the bootstrap test statistic S_T^*, denoting either the LM, LR or Wald statistic, as detailed in Section 3.2, for testing $M'\theta = m$. Define the corresponding p-value as $P_T^* := 1 - G_T^*(S_T)$ with $G_T^*(\cdot)$ denoting the conditional (on the original data) cdf[5] of S_T^*. The wild bootstrap robust Wald statistic, denoted RW_T^*, and associated p-value is defined analogously.

(v) The wild bootstrap test of H_0 against H_1 at level α rejects if $P_T^* \leq \alpha$.

Algorithm 2 (Unrestricted wild bootstrap)**.**

(i) & (ii) As in Algorithm 1.

(iii) Construct the bootstrap sample $\{X_t^*\}$ from

$$X_t^* := \Delta_+^{-\hat{d}} u_t^* \text{ with } u_t^* := a(L, \hat{\psi})\varepsilon_t^*, \quad t = 1, ..., T, \tag{15}$$

with the T bootstrap errors ε_t^* generated in step (ii) and with $\varepsilon_t^* = 0$ for $t \leq 0$.

(iv) Using the bootstrap sample, $\{X_t^*\}$, compute the bootstrap test statistic S_T^*, denoting either the LM, LR or Wald statistic, as detailed in Section 3.2, for testing $M'\theta = M'\hat{\theta}$. Define the corresponding p-value as $P_T^* := 1 - G_T^*(S_T)$ with $G_T^*(\cdot)$ denoting the conditional (on the original data) cdf of S_T^*. The wild bootstrap robust Wald statistic, denoted RW_T^*, and associated p-value is defined analogously.

(v) As in Algorithm 1.

Remark 3.2. In step (iii) of Algorithm 1 the parameters characterizing (1), which are used in constructing the bootstrap sample data, are estimated under the restriction of the null hypothesis, H_0 of (2). In contrast, in step (iii) of Algorithm 2 the corresponding unrestricted parameter estimates are used in constructing the bootstrap sample data. Since the bootstrap DGP in Algorithm 2 is then based on $\hat{\theta}$, the bootstrap test statistic computed in step (iv) is for the hypothesis that $M'\theta = M'\hat{\theta}$. ♢

Remark 3.3. In contrast to the bootstrap LM tests considered in CNT15, the wild bootstrap implementations in both Algorithm 1 and Algorithm 2 are based on the use of unrestricted residuals in step (i). Hence, the restricted wild bootstrap in Algorithm 1

[5] In practice, $G_T^*(\cdot)$ is unknown, but can be approximated in the usual way through numerical simulation by generating B (conditionally) independent bootstrap statistics, $S_{T:b}^*$, $b = 1, ..., B$. The simulated bootstrap p-value is then $\tilde{P}_T^* := B^{-1}\sum_{b=1}^{B}\mathbb{I}(S_{T:b}^* > S_T)$ and satisfies $\tilde{P}_T^* \to P_T^*$ almost surely as $B \to \infty$.

might possibly be better described as a type of hybrid bootstrap (because it is based on unrestricted residuals in step (i) and restricted parameter estimates in step (iii)). A similar bootstrap algorithm is mentioned in MacKinnon (2009), who argues that using unrestricted residuals together with restricted parameter estimates may be a sensible thing to do, see the discussion below his equation (6.10) on page 195. In our case, when testing null hypotheses involving the long memory parameter, d, the restricted residuals have the potential to be non-stationary under the alternative. This complication is safely avoided by using the unrestricted residuals. In what follows we will refer to Algorithm 1 as the restricted wild bootstrap and Algorithm 2 as the unrestricted wild bootstrap. $\diamondsuit$

4 Asymptotic Theory

We now derive the large sample behaviour of the standard QML-based estimator and test statistics under unconditional and/or conditional heteroskedasticity. We then derive sufficient conditions which ensure that the wild bootstrap implementations of these tests outlined in Algorithms 1 and 2 are asymptotically valid, both under the null hypothesis and, crucially, under the alternative. All results provided in this section hold conditionally on a given realisation of the random scale process, $\sigma(\cdot)$. Clearly, where the stated limiting distributions do not depend on $\sigma(\cdot)$, they also hold unconditionally. Results for non-stochastic $\sigma(\cdot)$ processes follow trivially as a special case of the stated results.

Our first result in this section is to establish the (global) consistency of the unrestricted QML estimator from (7) when the shocks, u_t, driving (1) satisfy Assumption 1. A corresponding result for the restricted QML estimator is given at the end of Section 4.1.

Theorem 1. *Let X_t be generated by model (1) and (4) with $\theta_0 \in \text{int}(\Theta)$, let $(\hat{d}, \hat{\psi})$ be defined by (7), and suppose Assumptions 1–4 are satisfied. Then, conditional on $\sigma(\cdot)$, $(\hat{d}, \hat{\psi}) \xrightarrow{p} (d_0, \psi_0)$.*

The consistency established in Theorem 1 applies under the true values given by the Pitman drift $\theta_{0,T}$ in (4). Of course, setting $\delta_\theta = 0$, the same result holds trivially when the data generating process is characterised by a fixed true value, θ_0. Also note that, because $\theta_{0,T} \to \theta_0$, it obviously follows from Theorem 1 that $\hat{\theta} - \theta_{0,T} \xrightarrow{p} 0$.

Remark 4.1. While somewhat unusual in typical consistency proofs for implicitly defined extremum estimators, the condition $\theta_0 \in \text{int}(\Theta)$ in Theorem 1 is needed under the Pitman drift (4) to ensure that $\theta_{0,T} \in \Theta$ for T sufficiently large, such that (5) applies for $a_n(\psi_{0,T})$. Inspection of the proof in Section S.5.1 in the Supplement shows that under fixed true value, i.e. with $\delta_\theta = 0$, it is sufficient that $\theta_0 \in \Theta$, see (S.17) and (S.18). $\diamondsuit$

Remark 4.2. The result in Theorem 1 establishes that the consistency result derived in HR11 (see also Nielsen, 2015) under the assumption of conditionally homoskedastic errors remains valid under the conditions of Assumption 1 thereby allowing for conditional and/or unconditional heteroskedasticity in the driving shocks, ε_t, in (1). This result does, however, require the stronger smoothness conditions in Assumption 3 relative to the corresponding conditions in Assumption A1 of HR11; see again the discussion in Remarks 2.10 and 2.11. Notice the result in Theorem 1 implies that this result also holds under the conditions in Remark 2.7. Although this imposes somewhat weaker conditions than the corresponding Assumption A2 of HR11, this must be traded off against the stronger conditions imposed by Assumption 3 relative to their Assumption A1. $\diamondsuit$

We next establish asymptotic distribution theory for the QML estimator from (7) under heteroskedasticity of the form given in Assumption 1. In order to do so we need to strengthen Assumptions 1 and 3 with the following additional assumptions.

Assumption 5. *For all integers q such that $3 \leq q \leq 8$ and for all integers $r_1, ..., r_{q-2} \geq 1$, the q'th order cumulants $\kappa_q(t, t, t - r_1, \ldots, t - r_{q-2})$ of $(z_t, z_t, z_{t-r_1}, \ldots, z_{t-r_{q-2}})$ satisfy the condition that $\sup_t \sum_{r_1,\ldots,r_{q-2}=1}^{\infty} |\kappa_q(t, t, t - r_1, \ldots, t - r_{q-2})| < \infty$.*

Assumption 6. *For all z such that $|z| \leq 1$, $a(z, \psi) = \sum_{n=0}^{\infty} a_n(\psi) z^n$ is thrice 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 4.3. The strengthening of Assumption 1(a)(iii) to a summability condition on the eighth order cumulants of ε_t in Assumption 5 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). Also, the moment condition $\sup_t E|z_t|^8 < \infty$, imposed for example by these authors, is necessary for Assumption 5 with $q = 8$ to hold and therefore is not stated explicitly. The additional cumulant conditions are required in the proof of Theorem 2 to verify that, under heteroskedastic innovations of the form given in Assumption 1, a Lindeberg-type condition holds for the score and for proving convergence in L_2 of the Hessian; see Sections S.5.2.1 and S.5.2.2, respectively. ♢

Remark 4.4. Assumption 6 requires $a(z, \psi)$ to be thrice differentiable in ψ rather than the corresponding twice differentiable condition in Assumption A3(ii) of HR11. This strengthening is used to prove tightness of the Hessian; see Section S.5.2.2. 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 HR11. ♢

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

$$A_0 := \sum_{n,m=1}^{\infty} \tau_{n,m} \begin{bmatrix} n^{-1}m^{-1} & -\gamma_n(\psi_0)'/m \\ -\gamma_n(\psi_0)/m & \gamma_n(\psi_0)\gamma_m(\psi_0)' \end{bmatrix} \text{ and } B_0 := \sum_{n=1}^{\infty} \begin{bmatrix} n^{-2} & -\gamma_n(\psi_0)'/n \\ -\gamma_n(\psi_0)/n & \gamma_n(\psi_0)\gamma_n(\psi_0)' \end{bmatrix},$$

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

Remark 4.5. The matrix B_0 coincides with the matrix A in HR11 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 $\tau_{n,m} = \mathbb{I}(n = m)$. Notice that neither A_0 nor B_0 include the effects of the random scale (unconditional heteroskedasticity) arising from part (b) of Assumption 1. ♢

Remark 4.6. Observe that A_0 (and hence B_0) is finite because $|\gamma_n(\psi)| = O(n^{-1-\varsigma}(\log n))$ by (5) and Lemma A.4, while $\tau_{n,m} = \kappa_4(t, t, t - n, t - m) + \kappa_2(t, t)\kappa_2(t - n, t - n)\mathbb{I}(n = m)$ as z_t is a MDS, so that $\sum_{n,m=1}^{\infty} n^{-1}m^{-1}|\tau_{n,m}| \leq \sum_{n,m=1}^{\infty} |\kappa_4(t, t, t - n, t - m)| + \sum_{n=1}^{\infty} n^{-2}\kappa_2(t, t)\kappa_2(t - n, t - n) < \infty$ by Assumption 1(a)(iii). ♢

As in HR11 we require B_0 to be invertible. This is formally stated in Assumption 7. Again this is satisfied by, for example, stationary and invertible ARMA processes.

Assumption 7. *The matrix B_0 is non-singular.*

In our second result we now establish asymptotic distribution theory for the QML estimator from (7) when the shocks, u_t, driving (1) are heteroskedastic. Using standard notation, $Y|X$ denotes the random variable Y conditional on X, and if $Y \sim N(a, b)$ we also use the notation $N(a, b)|X$.

Theorem 2. *Under the conditions of Theorem 1, Assumptions 5–7, and conditional on $\sigma(\cdot)$,*

$$\sqrt{T}(\hat{\theta} - \theta_{0,T}) \xrightarrow{w} N(0, C_0)|\sigma(\cdot), \tag{16}$$

$$\hat{A} - \lambda A_0 \xrightarrow{p} 0, \quad \hat{B} - B_0 \xrightarrow{p} 0, \quad \hat{C} - C_0 \xrightarrow{p} 0, \tag{17}$$

where $C_0 := \lambda B_0^{-1} A_0 B_0^{-1}$ with $\lambda := \int_0^1 \sigma^4(s)\mathsf{d}s / (\int_0^1 \sigma^2(s)\mathsf{d}s)^2$.

Again, the result in Theorem 2 is given for the Pitman drift true value, $\theta_{0,T}$, in (4). Under a fixed true value, θ_0, the same result holds with $\theta_{0,T} = \theta_0$ in (16).

Remark 4.7. Theorem 2 generalises the corresponding result in HR11 to the case where the shocks can display stochastic conditional and/or unconditional heteroskedasticity of the form in Assumption 1. Note that the limiting distribution in (16) is that of the scaled estimator, $\sqrt{T}(\hat{\theta} - \theta_{0,T})$, conditional on $\sigma(\cdot)$, by virtue of Assumption 1(b) which allows $\sigma(\cdot)$ to be random. Where $\sigma(\cdot)$ is non-random, (16) gives the unconditional limiting distribution. Under the conditions in Remark 2.7, $A_0 = B_0$ and $\lambda = 1$ and, hence, the result in (16) reduces to the (unconditional) result in Theorem 2.2 of HR11. Where heteroskedasticity arises only through part (a) of Assumption 1 then so the variance matrix C_0 in the right member of (16) reduces to $B_0^{-1} A_0 B_0^{-1}$. Moreover, where heteroskedasticity arises only through part (b) of Assumption 1 then so C_0 reduces to λB_0^{-1}. ♢

Remark 4.8. The result in (16) shows that the variance of the (conditional) asymptotic distribution of the (scaled and centered) QML estimator depends on the scalar parameter λ. When $\sigma(\cdot)$ is non-random so that $\sigma^2(\cdot)$ can be interpreted as the unconditional variance process, λ is a measure of the degree of unconditional heteroskedasticity (non-stationary volatility) present in $\{\varepsilon_t\}$. More generally, because $\sigma(\cdot)$ is allowed to be random, λ is a normalised measure of the time-variation in $\sigma(\cdot)$. When $\sigma(\cdot)$ is constant (e.g., for a homoskedastic process) then $\lambda = 1$, whereas when $\sigma(\cdot)$ is non-constant then $\lambda > 1$ by the Cauchy-Schwarz inequality. Consequently the variance of the asymptotic distribution of the QML estimator is seen to be inflated when time-variation is present in $\{\sigma_t\}$ (or unconditional heteroskedasticity is present in $\{\varepsilon_t\}$ in the non-random $\sigma(\cdot)$ case). ♢

4.1 Asymptotic Inference

We now turn to detailing the asymptotic behaviour of the standard LM, LR, Wald and robust Wald statistics under stochastic unconditional and/or conditional heteroskedasticity of the form given in Assumption 1. In order to investigate the impact of heteroskedasticity on both the asymptotic size and local power of the tests we will derive asymptotic distributions under the relevant (local) Pitman drift alternative; that is,

$$H_{1,T} : M'\theta = m + \delta/\sqrt{T}, \tag{18}$$

where δ is a fixed $q \times 1$ vector. Notice that for $\delta = 0$, $H_{1,T}$ reduces to H_0 of (2). Thus, in view of (4), when $H_{1,T}$ is true we have that θ_0 and δ_θ satisfy $m = M'\theta_0$ and $\delta = M'\delta_\theta$. When H_0 is true we further have that δ_θ satisfies $M'\delta_\theta = 0$. On the other hand, under the fixed alternative H_1 in (3), $\delta_\theta = 0$ and hence θ_0 is such that $M'\theta_0 \neq m$, i.e. $\theta_0 \notin \tilde{\Theta}$.

Theorem 3. *Let Assumptions 1–7 be satisfied and assume that* $\theta_0 \in \text{int}(\Theta)$. *Then, under* $H_{1,T}$ *of (18), it holds that, conditional on* $\sigma(\cdot)$,

$$S_T \xrightarrow{w} Y'F_0Y|\sigma(\cdot), \tag{19}$$

$$RW_T \xrightarrow{w} \chi^2_q\left(\delta'(M'C_0M)^{-1}\delta\right)|\sigma(\cdot), \tag{20}$$

where $Y \sim N((M'C_0M)^{-1/2}\delta, I_q)$, $F_0 := ((M'C_0M)^{1/2})'(M'B_0^{-1}M)^{-1}(M'C_0M)^{1/2}$, *and where* $\chi^2_q(g)$ *denotes a noncentral* χ^2_q *distribution with non-centrality parameter* g.

Corollary 1. *Let the conditions of Theorem 3 be satisfied. Then, under* H_0 *of (2) and conditional on* $\sigma(\cdot)$,

$$S_T \xrightarrow{w} Z'F_0Z|\sigma(\cdot), \quad Z \sim N(0, I_q) \tag{21}$$

$$RW_T \xrightarrow{w} \chi^2_q. \tag{22}$$

Remark 4.9. The quadratic forms on the right-hand sides of (19) and (21) are not, in general, χ^2-distributed because the matrix F_0 is not idempotent, which is a necessary and sufficient condition for the quadratic forms to be χ^2-distributed. However, noting that $C_0 = B_0^{-1}$ and $\lambda = 1$ in the homoskedastic case, the asymptotic null distribution of S_T is seen to be χ^2_q under the conditions in Remark 2.7. ♢

The right member of (22) does not depend on nuisance parameters, so Corollary 1 establishes that, conditional on $\sigma(\cdot)$, the robust Wald test is asymptotically correctly sized under heteroskedasticity of the form in Assumption 1. This is a remarkably strong result, since it implies correct asymptotic size not only *unconditionally*, i.e. on average across all possible realisations of $\sigma(\cdot)$, but also *conditionally*, i.e. for any possible realisation of the random process $\sigma(\cdot)$, the latter implying the former. This strong result does not, however, hold for the standard LR, LM and Wald tests whose limiting (conditional) distributions depend, through F_0, on $\sigma(\cdot)$ and on the volatility process of $\{z_t\}$.

Theorem 3 also demonstrates that the standard LR, LM and Wald tests and the robust Wald test will all have asymptotic local power properties that depend on the heteroskedasticity present in the process, even when size-corrected in the case of the standard LR, LM and Wald tests. The finite sample effects of a variety of shock processes which display a one-time change in variance and/or a GARCH-type structure on the size and power properties of the tests will be quantified by Monte Carlo methods in Section 5.

Remark 4.10. For the case where $M = (1, 0, \ldots, 0)'$, the results in Theorem 3, again conditional on $\sigma(\cdot)$, reduce to

$$S_T \xrightarrow{w} \frac{(C_0)_{1,1}}{(B_0^{-1})_{1,1}}\chi^2_q(\delta^2(C_0)^{-1}_{1,1})|\sigma(\cdot), \tag{23}$$

$$RW_T \xrightarrow{w} \chi^2_q(\delta^2(C_0)^{-1}_{1,1})|\sigma(\cdot). \tag{24}$$

In the case where $\sigma(\cdot)$ is non-stochastic, the result in (23) coincides with the result given in Theorem 1 of CNT15 for the LM statistic on noting that $(B_0^{-1})_{1,1}$ and $(C_0)_{1,1}$ correspond to the parameters ω^2 and $\lambda\varpi^2$, respectively, in CNT15. The asymptotic size and size-corrected asymptotic local power function which obtain under (23) in the case where $\sigma(\cdot)$ is non-random are given in analytical form and graphed in CNT15. These quantify the dependence of the size and local power of the LM, LR, and Wald tests on the degree and form of heteroskedasticity and/or weak dependence present. ♢

Remark 4.11. When $q = 1$, one-sided tests could also be considered using the usual t-statistic or a one-sided score statistic as in CNT15 or Robinson (1994, p. 1424). This would allow testing, for example, $d = 1/2$ against the alternative $d < 1/2$ or testing the unit root $d = 1$ against $d < 1$, or even $d = 2$ against $d < 2$ to check whether X_t is I(2). Such tests will be more powerful than the trinity of two-sided tests against these one-sided alternatives (in the correct tail). Indeed, in the Gaussian homoskedastic single-parameter model, the t-test and one-sided score test mentioned above are asymptotically uniformly most powerful (UMP), and the trinity of two-sided tests are asymptotically UMP unbiased, see Tanaka (1999) and Nielsen (2004) for the fractional model or Lehmann and Romano (2005) for a general treatment. The large sample theory for these one-sided tests follows entirely straightforwardly from the results given in this paper. ♢

We next establish the (global) consistency of the tests discussed in this section against fixed (non-local) alternatives.

Theorem 4. *Let the conditions of Theorem 3 be satisfied. Under the fixed alternative H_1 in (3), the statistics W_T, LR_T, LM_T and RW_T given in (10), (11), (12) and (13), respectively, are all of $O_p(T)$, where the rate is sharp. That is, the tests are consistent.*

Remark 4.12. The results in Theorem 4 show that the rate of consistency of the tests under H_1 is unaffected by heteroskedasticity of the form given in Assumption 1. ♢

Remark 4.13. It is straightforward, given the results in the proof of Theorem 4, to show that the t-test and one-sided score test for a single linear restriction discussed in Remark 4.11 are both of $O_p(T^{1/2})$ (sharp) under H_1. These tests will therefore be consistent, provided they are performed in the correct tail. ♢

We conclude this section with the following result on some properties of the restricted estimator, which will be needed for the bootstrap theory in the next section.

Theorem 5. *Let the conditions of Theorem 2 hold. Then the restricted estimator $\tilde{\theta}$ in (9) exists, and, furthermore:*

(i) If the local alternative (18) is true, then $\tilde{\theta} \xrightarrow{p} \theta_0$ and, conditional on $\sigma(\cdot)$, $\sqrt{T}M_\perp'(\tilde{\theta} - \theta_{0,T}) \xrightarrow{w} N(0, M_\perp' C_0 M_\perp)|\sigma(\cdot)$, where $M_\perp$ satisfies $M_\perp' M = M' M_\perp = 0$.

(ii) If the fixed alternative (3) is true such that $\theta_0 \notin \tilde{\Theta}$, then there exists a fixed value $\theta^\dagger \in \tilde{\Theta}$ such that $\tilde{\theta} \xrightarrow{p} \theta^\dagger$.

4.2 Bootstrap Inference

In this section we derive the conditions under which the wild bootstrap implementations of the QML-based tests from Section 3.2 correctly replicate the first-order asymptotic null distributions of the test statistics given in Corollary 1, under heteroskedasticity of the form considered in this paper. Our aim is to establish validity of the bootstrap not only *unconditionally*, i.e. on average across all possible realizations of $\sigma(\cdot)$, but also *conditionally*, i.e. for any possible realization of the random process $\sigma(\cdot)$. Conditional bootstrap validity is a much stronger result and also implies unconditional validity. Conditional validity can be shown to hold by establishing that the limiting distribution of the asymptotic test, conditional on $\sigma(\cdot)$, coincides with that of the bootstrap implementation of the test. Because the bootstrap statistics considered in this paper depend on $\sigma(\cdot)$ only through the original data, the distribution of a bootstrap statistic, conditional on both $\sigma(\cdot)$ and the original data, is identical to that conditional on the data alone. For this reason, we need only condition on the original data. As with the results for asymptotic tests obtained in the previous section, the random scale process $\sigma(\cdot)$ may appear in the bootstrap limiting distributions.

Before we state our results for the bootstrap tests, we consider the bootstrap estimator of the parameter θ. With the bootstrap data $\{X_t^*\}$ generated from step (iii) of either Algorithm 1 or 2, we define the bootstrap residuals and bootstrap estimator,

$$\varepsilon_t^*(\theta) := \sum_{n=0}^{t-1} b_n(\psi)\Delta_+^d X_{t-n}^*, \tag{25}$$

$$\hat{\theta}^* := \arg\min_{\theta\in\Theta} Q_T^*(\theta), \quad Q_T^*(\theta) := T^{-1}\sum_{t=1}^{T}\varepsilon_t^*(\theta)^2, \tag{26}$$

c.f. (6)–(8). To unify the notation, particularly in the generation of the bootstrap data in step (iii) of Algorithms 1 and 2, we consider the generic estimator $\check{\theta} := (\check{d}, \check{\psi}')'$ used to denote either $\tilde{\theta}$ (restricted) or $\hat{\theta}$ (unrestricted), depending on whether Algorithm 1 or 2, respectively, is applied to generate the bootstrap data. We further define $\theta^\dagger$ to be the limit in probability of $\check{\theta}$, such that $\check{\theta} \xrightarrow{p} \theta^\dagger$ as $T \to \infty$. When $\check{\theta}$ is used to denote the unrestricted estimator then $\theta^\dagger = \theta_0$ as shown in Theorem 1. When $\check{\theta}$ is used to denote the restricted estimator then $\theta^\dagger$ is given in Theorem 5 as θ_0 if the null or local alternative is true and as a fixed value $\theta^\dagger \neq \theta_0$ under the fixed alternative (the precise definition of $\theta^\dagger$ in the latter case can be found in the proof of Theorem 5 in Section C).

We are now able to establish the following results which are a necessary pre-requisite for establishing the large sample properties of the bootstrap tests in Theorems 7 and 8. All of the results given in the remainder of this section apply to both the restricted and unrestricted bootstrap procedures. Let

$$A^\dagger := \sum_{n=1}^{\infty}\tau_{n,n}\begin{bmatrix} n^{-2} & -\gamma_n(\psi^\dagger)'/n \\ -\gamma_n(\psi^\dagger)/n & \gamma_n(\psi^\dagger)\gamma_n(\psi^\dagger)' \end{bmatrix} \text{ and } B^\dagger := \sum_{n=1}^{\infty}\begin{bmatrix} n^{-2} & -\gamma_n(\psi^\dagger)'/n \\ -\gamma_n(\psi^\dagger)/n & \gamma_n(\psi^\dagger)\gamma_n(\psi^\dagger)' \end{bmatrix}.$$

Theorem 6. *Let Assumptions 1–7 be satisfied and assume that $\theta_0 \in \text{int}(\Theta)$. Then, under either $H_{1,T}$ of (18) or H_1 of (3),*

$$\hat{\theta}^* - \check{\theta} \xrightarrow{p^*}_p 0 \tag{27}$$

and

$$\sqrt{T}(\hat{\theta}^* - \check{\theta}) = O_p^*(1), \tag{28}$$

in probability. Moreover, if $\theta^\dagger \in \text{int}(\Theta)$ then,

$$\sqrt{T}(\hat{\theta}^* - \check{\theta}) \xrightarrow{w^*}_p N(0, C^\dagger)|\sigma(\cdot), \quad \textit{where } C^\dagger := \lambda B^{\dagger-1}A^\dagger B^{\dagger-1}. \tag{29}$$

The estimator $\check{\theta}$ is the bootstrap true value used to generate the bootstrap sample in step (iii) of Algorithms 1 and 2, and hence the consistency result in (27) of Theorem 6 is as expected. The results in Theorem 6 show that, in general, we can prove that the bootstrap estimator satisfies the asymptotic rate condition (28). When it further holds that the probability limit of the bootstrap true value, namely $\theta^\dagger$ is in the interior of the parameter space, then we can also show that the bootstrap estimator is (conditionally) asymptotically normally distributed as in (29).

We note that for the unrestricted bootstrap in Algorithm 2, $\theta^\dagger = \theta_0$, which is always assumed to be in the interior of the parameter space, so for the unrestricted bootstrap we always obtain (29), even under the fixed alternative H_1 in (3). However, for the restricted bootstrap in Algorithm 1, $\theta^\dagger$ may be on the boundary under the fixed alternative H_1 when the hypothesised value is very far from the true value, θ_0, see Theorem 5. In the latter case, we therefore only obtain (28). The result (28) is sufficient to prove consistency of the bootstrap tests based on either bootstrap algorithm, but not sufficient to show the validity of bootstrap confidence intervals which would require (29).

We note from the difference between $A^\dagger$ and A_0 that the asymptotic distribution of the bootstrap estimator, conditional on $\sigma(\cdot)$, in (29) is unable to capture the asymmetry in the dependence structure of the higher-order moments of the innovations, as represented by $\tau_{r,s} = E(z_t^2 z_{t-r} z_{t-s})$ for $r \neq s$, because of the independence imposed on the bootstrap errors. That is, conditional on the original data, $E^*(\varepsilon_t^{*2}\varepsilon_{t-r}^*\varepsilon_{t-s}^*) = \hat{\varepsilon}_{c,t}^2 \hat{\varepsilon}_{c,t-r}\hat{\varepsilon}_{c,t-s} E(w_t^2 w_{t-r} w_{t-s}) = 0$ for all $r \neq s$. With the exception of the wild bootstrap implementation of the robust Wald test, for asymptotic validity of our proposed wild bootstrap tests we therefore need to strengthen part (ii) of Assumption 1(a) as follows:

Assumption 1'. *Assumption 1 holds with part (a)(ii) replaced by:*
(ii') $\tau_{r,s} := E(z_t^2 z_{t-r} z_{t-s})$ is uniformly bounded for all $r \geq 0, s \geq 0$, and $\tau_{r,s} = 0$ for $r \neq s$.

Assumption 1' imposes the additional (symmetry) condition that $\tau_{r,s} = 0$ for $r \neq s$. We note, in particular, that Assumption 1' does not rule out leverage effects, which are related to third moments and occur when $E(z_t^2 z_{t-r}) \neq 0$ for some $r \geq 1$. Instead, Assumption 1' rules out certain types of asymmetric volatility clustering, i.e. correlations between z_t^2 and $z_{t-r}z_{t-s}$ for $r \neq s$. Thus, by still allowing for leverage, Assumption 1' is still weaker than the corresponding conditions imposed in Robinson (1991), Demetrescu, Kuzin and Hassler (2008), Hassler, Rodrigues and Rubia (2009) and Kew and Harris (2009); see also Remarks 2.3 and 2.4.

Remark 4.14. As noted above Theorem 6, $\theta^\dagger = \theta_0$ under $H_{1,T}$ of (18), regardless of the bootstrap algorithm applied. Consequently, the matrices C_0 and $C^\dagger$ coincide under the additional Assumption 1'. This coincidence is required for establishing bootstrap validity in Theorem 7 and Corollary 2 below. However, because the asymptotic null distribution of RW_T in Corollary 1 does not depend on C_0, bootstrap implementations of RW_T will not require the additional restrictions of Assumption 1' for asymptotic vaildity. $\diamondsuit$

Remark 4.15. Under Assumption 1', the conditional limit distributions given in (16) for the QML estimator and in (29) for the QML estimator obtained in connection with the unrestricted wild bootstrap of Algorithm 2, coincide to first order. This result can therefore be used as the basis for constructing asymptotically valid bootstrap confidence intervals for θ under Assumption 1'. For example, when θ is a scalar parameter, the asymptotic bootstrap distribution of $\hat{\theta}^*$, using Algorithm 2, is centered around $\hat{\theta}$ and bootstrap confidence intervals can be based on the empirical quantiles from the bootstrap distribution of $\hat{\theta}^*$, conditional on the original data. Specifically, letting $\hat{\theta}_\alpha^*$ denote the α% quantile of the conditional bootstrap distribution of $\hat{\theta}^*$, the asymptotic $(1-\alpha)$%-level naïve (or basic) and percentile bootstrap confidence intervals are given by $[2\hat{\theta} - \hat{\theta}^*_{(1-\alpha/2)}; 2\hat{\theta} - \hat{\theta}^*_{(\alpha/2)}]$ and

$[\hat{\theta}^*_{(\alpha/2)}, \hat{\theta}^*_{(1-\alpha/2)}]$, respectively. Alternatively, the studentised bootstrap confidence interval can be constructed from the associated t-statistics of the bootstrap estimates. ♢

We now report Theorem 7 and Corollary 2 which establish the large sample validity of our proposed bootstrap tests. These results apply to both bootstraps, described in Algorithms 1 and 2. That is, they apply regardless of whether the bootstrap pseudo-data are generated using the restricted estimates, in which case the original null hypothesis $M'\theta = m$ is tested on the bootstrap data, or using the unrestricted estimates, in which case the hypothesis $M'\theta = M'\hat{\theta}$ is tested on the bootstrap data.

Theorem 7. *Let Assumptions 1–7 be satisfied and $\theta_0 \in \text{int}(\Theta)$. Then, under $H_{1,T}$,*

(i) $RW_T^* \xrightarrow{w^*}_p \chi^2_q$;

(ii) if Assumption 1 is replaced by 1', then $S_T^* \xrightarrow{w^*}_p Z'F_0Z|\sigma(\cdot)$, where $Z \sim N(0, I_q)$.

Theorem 7 has the following corollary, where P_T^* denotes the (wild bootstrap) p-value associated with any of the four test statistics considered. The proof of the corollary follows straightforwardly using, conditional on a given realisation of $\sigma(\cdot)$, the same arguments as are made in the proof of Theorem 5 of Hansen (2000).

Corollary 2. *Let the conditions of Theorem 7 be satisfied. Under the null hypothesis (2) and conditional on $\sigma(\cdot)$, $P_T^* \xrightarrow{w} U[0,1]$, i.e. a uniform distribution on $[0,1]$.*

An immediate implication of the result in Corollary 2 is that the wild bootstrap implementations of the LM, LR and Wald tests will all have correct asymptotic size in the presence of stochastic unconditional and conditional heteroskedasticity as given in Assumption 1'. This result holds conditional on $\sigma(\cdot)$, i.e. for any possible realisation of the random scale process $\sigma(\cdot)$. In the case of the robust Wald test the results holds without the necessity to strengthen Assumption 1 with the stronger moment condition in Assumption 1'. Notice that these results are trivially also seen to be true under homoskedasticity since the conditions in Remark 2.7 are contained within both Assumptions 1 and 1'. Moreover, the results in Theorem 7 also imply immediately that under either Assumption 1' or Assumption 1, as appropriate, the wild bootstrap tests will attain the same asymptotic local power function as the size-adjusted (recalling that the robust Wald test is asymptotically correctly sized) asymptotic tests; cf. Theorem 3.

Remark 4.16. In the working paper version of CNT15, we also analyzed i.i.d. bootstrap implementations of the LM test. As would be expected, the i.i.d. bootstrap is not able to account for heteroskedasticity, and therefore suffers similar size distortions to the asymptotic tests. It is straightforward to demonstrate that the same holds true for i.i.d. implementations of the LR and Wald tests discussed in this paper. However, the i.i.d. bootstrap does, like the wild bootstrap version of this test, correctly replicate the limiting null distribution of the robust Wald statistic under Assumption 1. ♢

We conclude this section by establishing consistency of our proposed wild bootstrap tests against fixed alternatives.

Theorem 8. *Let Assumptions 1–7 be satisfied and assume that $\theta_0 \in \text{int}(\Theta)$. Then, under the fixed alternative H_1 in (3), S_T^* and RW_T^* are $O_p^*(1)$, in probability. Furthermore, if $\theta^\dagger \in \text{int}(\Theta)$, then the conclusions of Theorem 7 also hold under the fixed alternative.*

Taken in tandem with the results established for the original (non-bootstrapped) statistics in Theorem 4, the results in Theorem 8 establish that the wild bootstrap implementations of these tests, regardless of whether Algorithm 1 or Algorithm 2 is used, are consistent against fixed alternatives.

Remark 4.17. Because $\theta^\dagger = \theta_0$ for the unrestricted wild bootstrap (Algorithm 2), Theorem 8 establishes that the unrestricted wild bootstrap statistics attain the same first order (conditional) limiting distribution under fixed alternatives as they do under the null. This result does not, however, hold for the restricted wild bootstrap statistics from Algorithm 1 because their bootstrap true value $\check{\theta}$ (and specifically its limit in probability, $\theta^\dagger$) can lie on the boundary of the parameter space. As a result the restricted wild bootstrap statistics can only be shown to be $O_p^*(1)$, in probability, when considering all possible fixed alternatives. While this does not entail a loss of consistency for the restricted wild bootstrap tests against fixed alternatives, it is suggestive that the finite sample power of the restricted and unrestricted wild bootstrap implementations of a given test could potentially differ. These issues are explored further in Section 5. ♢

5 Monte Carlo Simulations

We report results from a simulation study comparing the finite sample properties of the asymptotic and bootstrap tests described above, in the context of a fractionally integrated process allowing for weak dependence and both homoskedastic and heteroskedastic errors.

5.1 Monte Carlo Setup

The Monte Carlo data are simulated from the model in (1) with u_t generated according to either an AR(1) or an MA(1) process; that is u_t will satisfy either (30) or (31):

$$(1 - aL)u_t = \varepsilon_t, \tag{30}$$

$$u_t = (1 + bL)\varepsilon_t, \tag{31}$$

where in each case the innovations $\varepsilon_t = \sigma_t z_t$ and σ_t, z_t will be defined below.

We report results for the asymptotic LR_T, W_T, LM_T and RW_T tests from Section 3.2, together with their wild bootstrap counterparts from Section 3.3. We consider three specific hypotheses in the context of model (1) with u_t generated according to either (30) or (31). Principally, with the results reported in Sections 5.2–5.4, we consider tests on the long memory parameter in (1), focusing on testing the null hypothesis $H_{0,1} : d = 1$ against $H_{1,1} : d \neq 1$. Here we will reports results for both finite sample size and power by setting $d = 1$ and $d = 1 + \delta/\sqrt{T}$, respectively, with $\delta = 2$ when weak dependence is not present (i.e., where $a = b = 0$) and $\delta = 3$ otherwise. Further results are reported in the accompanying Supplement which relate, in the context of (1) and (30), to tests on: (i) the autoregressive parameter, for testing $H_{0,2} : a = 0$ against $H_{1,2} : a \neq 0$, i.e., a specification test for autoregressive order, with finite sample size and power results reported for $a = 0$ and $a = 5/\sqrt{T}$, respectively, in (30); (ii) joint tests for $H_{0,3} : d = 1 \cap a = 0$ against $H_{1,3} : d \neq 1 \cup a \neq 0$, with finite sample size and power results reported for $d = 1, a = 0$ and $d = 1 + 1/\sqrt{T}, a = (5/3)/\sqrt{T}$, in (1) and (30). In hypothesis $H_{0,1}$ the choice of $d = 1$ (which is without loss of generality) delivers tests of the $I(1)$ null. The second hypothesis, $H_{0,2}$, yields model specification tests for whether or not the autoregressive lag is required. The final hypothesis, $H_{0,3}$, gives tests of the random walk null. A brief summary of the results reported in the Supplement for the last two hypotheses is given in Section 5.5.

Results are reported for samples of size $T = 100$ and $T = 250$, and under $T = \infty$ we also report the asymptotic size or size-corrected asymptotic local power calculated as described in Remark 4.10. Note that, for the asymptotic tests, the simulated finite sample power of the asymptotic tests has been size-corrected, while the reported power values for their bootstrap implementations has not been size-corrected. All tests were computed at 5% nominal size. The test statistics W_T, LM_T and RW_T in (10), (12) and (13), respectively, were implemented using numerical derivatives, and specifically, the variance estimator, $\hat{B}^{-1}$, in the calculation of W_T was the inverse of the negative (numerical) Hessian. For the bootstrap implementations, we used 499 bootstrap replications and the i.i.d. sequence w_t for the wild bootstrap was chosen as the simple two-point distribution $P(w_t = -1) = P(w_t = 1) = 0.5$, which we found to perform slightly better than other standard choices of w_t made in the bootstrap literature. All simulations were performed in Ox version 7.1, see Doornik (2007), and based on $10,000$ Monte Carlo replications.

5.2 Results With Unconditionally Heteroskedastic, Uncorrelated Errors

We consider first the case where the shocks do not display weak dependence (i.e., $a = b = 0$, such that $u_t = \varepsilon_t$) and analyse the impact of unconditional heteroskedasticity on the tests of $H_{0,1} : d = 1$, uncontaminated by the influence of weak dependence. For the present we take $\{z_t\}$ to be conditionally homoskedastic, and specifically we simulate it as an i.i.d. $N(0, 1)$ sequence. Notice that the wild bootstrap tests from Algorithms 1 and 2 coincide in this case because θ is simply the scalar long memory parameter, d.

The scale process is generated according to the deterministic one-shift volatility process, $\sigma_t = \upsilon_0 + (\upsilon_1 - \upsilon_0)\mathbb{I}(t \geq \tau T)$; i.e., there is an abrupt single shift in the variance from υ_0^2 to υ_1^2 at time τT, for some $\tau \in (0, 1)$.[6] Recall that when $\sigma(\cdot)$ is non-random, $\sigma^2(\cdot)$ can be interpreted as the unconditional variance profile of ε_t. Without loss of generality we normalise $\upsilon_0^2 = 1$. We let the break date vary among $\tau \in \{1/4, 3/4\}$ and the ratio $\upsilon := \upsilon_1/\upsilon_0$ among $\upsilon \in \{1/3, 1, 3\}$. Note that $\upsilon = 1$ corresponds to homoskedastic errors. These values of τ and υ are motivated by the so-called Great Moderation and the recent Great Recession, as mentioned in the introduction, suggesting a decline in the volatility early in the sample and an increase in the volatility late in the sample, respectively.

The results for the case with conditionally homoskedastic $\{z_t\}$ are in Table 1 (the rows with $\upsilon = 1$). Even in this case, a comparison between the results for the asymptotic tests in Panel A and the corresponding wild bootstrap results in Panel B shows that the bootstrap can deliver significant improvements over the empirical size of the asymptotic tests. For example, for $T = 100$ the empirical rejection frequency of the RW_T test is 6.33% while that of the corresponding wild bootstrap test is 5.03%.

It is where heteroskedasticity is present in the shocks (the rows where $\upsilon \neq 1$) that the wild bootstrap based tests display their superiority over the other available tests. From the results in Table 1 we see that the asymptotic LM, LR and Wald tests can be severely over-sized with this phenomenon persisting as the sample size is increased, as predicted by the asymptotic distribution theory in Theorem 3. Again as predicted by Theorem 3, the degree of over-sizing seen in these tests worsens as λ increases. For example, in the two cases where $\lambda = 2.333$ the empirical rejection frequency of these tests approaches 20% regardless of the sample size. The robust Wald test, RW_T, displays much better size control, as would be expected, under heteroskedasticity but for the $\lambda = 2.333$ case it can still be significantly over-sized, especially so for $T = 100$ where it displays an empirical

[6] We considered other models for the scale process σ_t such as those in Cavaliere and Taylor (2005, 2008). For comparable values of the parameter λ defined in Theorem 3 and discussed in Remark 4.8, these results were qualitatively similar to those reported here, as predicted by the asymptotic theory, and are consequently omitted in the interests of brevity.

Table 1: Tests of $H_{0,1}$: simulated size and power with one-time shift in unconditional volatility

				size				power			
τ	υ	T	λ	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests											
	1	100	1.00	5.93	5.08	5.68	6.33	63.04	70.07	62.76	60.05
	1	250	1.00	5.54	5.32	5.47	5.64	67.02	71.47	67.13	65.26
	1	∞	1.00	5.00	5.00	5.00	5.00	72.74	72.74	72.74	72.74
1/4	1/3	100	2.33	18.33	17.19	17.93	9.02	33.41	45.07	32.44	24.25
1/4	1/3	250	2.33	19.14	18.45	18.96	7.03	35.85	44.00	35.70	29.56
1/4	1/3	∞	2.33	19.95	19.95	19.95	5.00	38.96	38.96	38.96	38.96
1/4	3	100	1.24	8.46	7.73	8.18	6.38	53.26	61.37	52.75	50.37
1/4	3	250	1.24	8.13	7.80	8.02	5.39	58.49	63.86	58.42	57.00
1/4	3	∞	1.24	7.90	7.90	7.90	5.00	63.27	63.27	63.27	63.27
3/4	1/3	100	1.24	8.38	7.46	8.06	6.70	56.71	64.89	56.07	52.39
3/4	1/3	250	1.24	8.15	7.77	8.02	5.77	58.70	63.74	58.60	56.35
3/4	1/3	∞	1.24	7.90	7.90	7.90	5.00	63.27	63.27	63.27	63.27
3/4	3	100	2.33	19.54	18.20	19.03	9.50	29.25	40.13	28.14	20.15
3/4	3	250	2.33	19.41	18.91	19.26	7.21	33.06	40.83	32.83	27.10
3/4	3	∞	2.33	19.95	19.95	19.95	5.00	38.96	38.96	38.96	38.96
Panel B: wild bootstrap tests											
	1	100	1.00	5.35	5.26	5.38	5.03	63.44	69.97	63.11	60.01
	1	250	1.00	5.18	5.29	5.24	5.06	66.91	71.76	66.78	65.18
	1	∞	1.00	5.00	5.00	5.00	5.00	72.74	72.74	72.74	72.74
1/4	1/3	100	2.33	7.00	7.00	6.92	5.79	37.27	48.02	36.09	26.33
1/4	1/3	250	2.33	5.61	5.47	5.61	5.16	36.67	44.47	36.21	29.36
1/4	1/3	∞	2.33	5.00	5.00	5.00	5.00	38.96	38.96	38.96	38.96
1/4	3	100	1.24	5.35	5.44	5.30	4.93	53.41	61.72	52.93	48.72
1/4	3	250	1.24	4.88	5.02	4.88	4.67	57.38	63.12	57.27	55.37
1/4	3	∞	1.24	5.00	5.00	5.00	5.00	63.27	63.27	63.27	63.27
3/4	1/3	100	1.24	5.27	5.50	5.29	4.97	56.91	65.35	56.12	51.42
3/4	1/3	250	1.24	5.35	5.40	5.33	5.12	58.36	63.61	58.21	56.22
3/4	1/3	∞	1.24	5.00	5.00	5.00	5.00	63.27	63.27	63.27	63.27
3/4	3	100	2.33	7.29	7.20	7.25	6.11	33.29	42.34	32.50	23.09
3/4	3	250	2.33	5.91	5.92	5.90	5.45	34.32	41.61	34.17	27.66
3/4	3	∞	2.33	5.00	5.00	5.00	5.00	38.96	38.96	38.96	38.96

Notes: Entries for finite T are simulated rejection frequencies of the tests. Entries for $T = \infty$ are calculated as described in Remark 4.10. Power is measured at $\delta = 2$ and is size corrected for the asymptotic tests, but not for the bootstrap tests. All entries are based on $10,000$ Monte Carlo replications.

rejection frequency of almost 10%. In contrast, the wild bootstrap tests in Panel B of Table 1 display very good size control throughout; the largest entry relating to size in Panel A of Table 1 is a rejection frequency of 7.29% for the bootstrap LM_T test which occurs for $T = 100$ with $\tau = 0.75$ and $\upsilon = 3$; although observe that the wild bootstrap implementation of the robust Wald test, RW_T, has empirical size of 6.11% in this case. Indeed, among the bootstrap tests, it is the wild bootstrap RW_T test that appears to deliver the best finite sample size control overall across the results in Table 1.

Turning to the power of the tests, we see again from the results in Panel A of Table

1 that the predictions from the asymptotic theory are strongly reflected in finite samples with the size-corrected empirical power of the asymptotic tests being lower the larger the value of λ, and that, as with the size results, these effects do not vanish as the sample size is increased. Indeed, the size-adjusted power of the tests can be significantly lower; for example, when $\lambda = 1$ all of the tests display an empirical rejection frequency of around 70% but for $\lambda = 2.333$ power can be roughly half this level. A notable feature of the power results in Panel B of Table 1 is how close these results are to the size-adjusted power results for the asymptotic tests in Panel A of the table. Although this is predicted by the large sample distribution theory in Section 4, it is nonetheless interesting to see how closely the finite sample results adhere to this asymptotic prediction. Amongst the tests, for both the size-corrected asymptotic tests and the wild bootstrap tests, the LR test, LR_T, displays the highest finite sample power. However, it should be recalled that the wild bootstrap LR_T test generally displays slightly higher empirical size than the wild bootstrap RW_T test, suggesting that at least some of this finite sample power advantage may simply be be artefact of the relative finite sample sizes of the two tests. Certainly in those cases where the empirical sizes of the bootstrap LR_T and RW_T tests are closest, the power differences between them are relatively small.

5.3 Results With Conditionally Heteroskedastic, Uncorrelated Errors

Next, we consider the following models where $\{z_t\}$ is conditionally heteroskedastic, in each case with $\{e_t\}$ forming an i.i.d. sequence.

Model A : $\varepsilon_t = z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.5 z_{t-1}^2, e_t \sim N(0,1)$.

Model B : $\varepsilon_t = z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.5 z_{t-1}^2, e_t \sim (3/5)^{1/2} t_5$.

Model C : $\varepsilon_t = z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.2 z_{t-1}^2 + 0.79 h_{t-1}, e_t \sim N(0,1)$.

Model D : $\varepsilon_t = z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.2 z_{t-1}^2 + 0.79 h_{t-1}, e_t \sim (3/5)^{1/2} t_5$.

Model E : $\varepsilon_t = z_t = h_t^{1/2} e_t, \log h_t = -0.23 + 0.9 \log h_{t-1} + 0.25 \left(|e_{t-1}^2| - 0.3 e_{t-1} \right), e_t \sim N(0,1)$.

Model F : $\varepsilon_t = z_t = h_t^{1/2} e_t, h_t = 0.0216 + 0.6896 h_{t-1} + 0.3174 \left(z_{t-1} - 0.1108 \right)^2, e_t \sim N(0,1)$.

Model G : $\varepsilon_t = z_t = h_t^{1/2} e_t, h_t = 0.005 + 0.7 h_{t-1} + 0.28 \left(|z_{t-1}| - 0.23 z_{t-1} \right)^2, e_t \sim N(0,1)$.

Model H : $\varepsilon_t = z_t = e_t \exp(h_t), h_t = 0.936 h_{t-1} + 0.5 v_t, (v_t, e_t) \sim N(0, \mathrm{diag}(\sigma_v^2, 1)), \sigma_v = 0.424$.

Model I : $\varepsilon_t = \sigma_t z_t, \sigma_t = 1 + 2\mathbb{I}(t \geq 0.75T), z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.5 z_{t-1}^2, e_t \sim N(0,1)$.

The conditionally heteroskedastic configurations for $\{z_t\}$ specified in Models A–H are a subset of those used in Section 4 of Gonçalves and Kilian (2004), to which the reader is referred for further discussion. Models A-D are standard stationary GARCH$(1,1)$ models driven by either Gaussian or t-distributed shocks with unit variance, while Model E is the is the exponential GARCH$(1,1)$ [EGARCH$(1,1)$] model of Nelson (1991). Model F is the asymmetric GARCH$(1,1)$ [AGARCH$(1,1)$] model of Engle (1990), Model G is the GJR-GARCH$(1,1)$ model of Glosten, Jaganathan and Runkle (1993), and Model H is a first-order autoregressive stochastic volatility model. Finally, Model I combines conditional heteroskedasticity in $\{z_t\}$, of the form specified by Model A, together with the one-time change model for the unconditional variance considered in the previous subsection (for the particular case of $\upsilon = 3$ and $\tau = 0.75$). The chosen parameter values in Models A–H are based on applied work see Section 4 of Gonçalves and Kilian (2004), where the relation between these models and the moment conditions in Assumptions 1 and 1' is also discussed. As noted by Gonçalves and Kilian (2004, p. 104), Models E, F, and G all fail the additional symmetry condition imposed by Assumption 1' required for the asymptotic validity of the wild bootstrap implementations of the LM, LR, and Wald tests. It is of nonetheless of interest to investigate the finite sample behaviour of the tests under models that may not in fact satisfy the assumptions needed for the asymptotic theory. The results relating to Models A-I are presented in Table 2.

Consider first the results in Panel A of Table 2 for the empirical sizes of the asymptotic tests. Here we see that for these commonly encountered models of conditional heteroskedasticity these tests can be very badly over-sized; indeed, the degree of over-sizing is, if anything, more pronounced than was observed in these tests for the models of unconditional heteroskedasticity in Table 1. While it was seen in Table 1 that the degree of size distortions under the single break model depends on both the change-point location and the magnitude of the break (with these distortions being relatively moderate for increases in variance early in the sample and decreases late in the sample), there are no entries for the asymptotic LM_T, LR_T and W_T tests in Table 2 that lie below 10%. Models H and I clearly effect the greatest degree of over-size, with the empirical sizes under Model H approaching 40%. Consistent with the results in Theorem 3, these size distortions do not disappear as the sample size is increased; indeed, the opposite occurs. Also in line with the asymptotic prediction from Theorem 3, the robust Wald test, RW_T, displays significantly better size control than the asymptotic LM_T, LR_T and W_T tests, although it remains uncomfortably over-sized even in moderately large sample sizes, again most notably under Models H and I. Turning to the results in Panel B of Table 2 we see, as with the case of unconditional heteroskedasticity in Table 1, that the wild bootstrap again does a good job overall in controlling size under all of Models A-I, although some finite sample over-size is still seen with the bootstrap LM_T, LR_T and W_T tests under Models B, H and I. Amongst the wild bootstrap tests, the best size control is again delivered by the wild bootstrap RW_T test.

As with the results in Table 1, the results in Panel B of Table 2 show that the size-corrected power of the asymptotic tests is very strongly affected by the presence of conditional heteroskedasticity in each of Models A-I, as expected from Theorem 3. In line with the empirical size results reported in Panel A this is seen to be most pronounced for Models H and I, and that these effects do not vanish (in fact, they tend to become more pronounced) as the sample size increases. As in Table 1, the size-adjusted power of the tests can be significantly lower than under homoskedasticity, particularly for Models H and I. The power results in Panel B of Table 2 for the wild bootstrap tests again lie relatively close to the size-adjusted power results for the asymptotic tests seen in Panel A of Table 2, although in a number of cases power can be somewhat higher. As with the results in Table 2, the LR tests display the highest finite sample power among the four tests, both for the size-corrected asymptotic implementations of these tests and their wild bootstrap analogues, but again for the bootstrap tests part of this appears to be attributable to differences in the empirical sizes of the four bootstrap tests.

5.4 Results With Weakly Dependent Errors

We now turn our attention to the results presented in Tables 3 and 4 which investigate the finite sample size and power properties, respectively, of the asymptotic and bootstrap tests of $H_{0,1} : d = 1$ for cases where the process is driven by shocks which can display both weak dependence and heteroskedasticity of the type considered in Table 1. Notice that because θ is now a two-dimensional vector — specifically, $\theta = (d, a)'$ in the case where u_t is generated according to (30), and $\theta = (d, b)'$ where u_t is generated according to (31)

Table 2: Tests of $H_{0,1}$: simulated size and power with conditionally heteroskedastic Models A–I

		size				power			
	T	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests									
Model A	100	16.14	14.77	15.84	8.68	36.41	47.54	35.13	30.81
	250	16.87	16.42	16.75	6.58	35.51	43.65	34.90	36.39
Model B	100	18.12	17.35	17.77	9.79	30.69	42.99	28.82	26.87
	250	23.54	22.96	23.38	7.80	24.00	32.59	23.59	27.68
Model C	100	12.20	10.95	11.89	7.73	43.71	54.29	43.07	40.79
	250	15.56	15.17	15.42	6.69	37.93	46.10	37.50	36.88
Model D	100	13.34	11.99	13.09	8.12	40.24	51.45	39.17	39.11
	250	18.38	17.72	18.26	7.06	31.57	40.92	31.13	33.55
Model E	100	16.38	15.38	15.99	8.74	34.56	46.09	33.37	28.83
	250	21.86	21.19	21.73	7.69	25.71	35.21	24.87	25.90
Model F	100	15.74	14.38	15.41	7.77	36.07	47.55	34.74	32.44
	250	23.84	22.71	23.74	7.88	24.21	33.08	23.79	23.86
Model G	100	14.83	13.42	14.53	8.14	36.46	48.15	35.33	30.97
	250	20.81	20.00	20.66	7.05	26.17	35.18	25.65	25.59
Model H	100	28.39	27.39	28.04	11.77	17.23	29.58	15.28	14.69
	250	38.86	38.28	38.72	10.67	10.57	18.50	9.82	9.32
Model I	100	27.21	25.93	26.77	10.75	19.26	30.15	17.65	13.04
	250	30.72	29.80	30.52	8.34	18.45	26.49	18.05	14.61
Panel B: wild bootstrap tests									
Model A	100	6.77	6.91	6.73	5.62	43.24	52.53	42.19	33.78
	250	5.84	5.89	5.84	4.81	42.06	48.31	41.80	35.51
Model B	100	7.50	7.34	7.45	6.17	42.51	51.65	41.67	32.24
	250	7.00	7.11	7.01	5.41	38.50	45.60	38.07	30.17
Model C	100	5.80	5.98	5.88	5.23	48.05	57.03	47.39	41.32
	250	5.59	5.59	5.55	5.21	43.04	49.25	42.73	37.68
Model D	100	6.14	6.17	6.19	5.60	47.75	56.30	47.15	40.42
	250	6.17	6.16	6.18	5.27	40.97	48.47	40.61	34.58
Model E	100	6.36	6.25	6.38	5.49	40.51	50.19	39.79	31.06
	250	6.25	6.10	6.32	5.14	35.34	42.36	35.01	27.87
Model F	100	5.94	5.76	5.98	5.07	41.11	50.19	40.29	32.09
	250	5.97	5.95	5.98	5.39	32.42	39.66	32.13	25.47
Model G	100	6.10	5.61	6.05	5.51	40.63	50.33	39.76	32.45
	250	5.63	5.53	5.65	4.96	32.67	39.57	32.47	25.63
Model H	100	8.20	8.49	8.33	6.52	30.58	40.61	29.48	20.94
	250	7.67	7.73	7.72	6.17	20.98	28.78	20.63	13.83
Model I	100	7.87	7.73	7.83	5.75	27.18	36.78	26.22	15.54
	250	6.38	6.44	6.39	5.01	25.15	31.69	24.79	16.14

Notes: Entries are simulated rejection frequencies of the tests. Power is measured at $\delta = 2$ and is size corrected for the asymptotic tests, but not for the bootstrap tests. All entries are based on $10,000$ Monte Carlo replications.

— the wild bootstrap tests which obtain under Algorithm 1 (restricted bootstrap) and Algorithm 2 (unrestricted bootstrap) now differ and, hence, results are reported for both. Results are reported for $a, b \in \{-0.8, 0.8\}$. In the cases where $a = 0.8$ and $b = -0.8$ the quantity $(C_0)_{1,1} = \lambda(B_0^{-1})_{1,1}$ discussed in Remark 4.10 is equal to 5.3221λ, while for $a = -0.8$ and $b = 0.8$, $(C_0)_{1,1} = 0.68937\lambda$. For the no weak dependence case ($a = b = 0$), $(C_0)_{1,1} = (\pi^2/6)^{-1}\lambda = 0.60793\lambda$, see also Figure 1 in Nielsen (2004).

Consider first the empirical size results for the homoskedastic case ($\lambda = 1$) in the first block of columns in Table 3. These highlight the poor finite sample size control of the asymptotic tests in the presence of weak dependence; most notably, severe over-sizing for the Wald and robust Wald tests when either a positive AR or negative MA component is present, and a degree of under-sizing in the LR test for a positive AR component. For example, for $a = 0.8$ and $T = 100$ the W_T, RW_T and LR_T tests have empirical rejection frequencies of around 10%, 16% and 3%, respectively, while for $b = -0.8$ and $T = 100$ the W_T and RW_T tests have empirical rejection frequencies of around 17% and 21%, respectively. In contrast, the wild bootstrap based analogues based on the restricted bootstrap of Algorithm 1 (Panel B) display very good size control throughout; in the first example above the corresponding restricted wild bootstrap W_T and RW_T reject only slightly under 5% of the time and for the restricted wild bootstrap LR_T test almost exactly 5% of the time. In the second example, the restricted wild bootstrap W_T and RW_T tests reject only slightly over 5% of the time. The unrestricted wild bootstrap of Algorithm 2 (Panel C) also controls size well in general but is not as effective in controlling size as the restricted wild bootstrap, most notably in the preceding example where the restricted wild bootstrap W_T and RW_T tests both reject the null around 9% of the time.

Turning to the two heteroskedastic cases reported in Table 3, the patterns of size distortions seen in the asymptotic LM, LR and Wald tests are very similar to those seen for these two cases (notice that the value of λ coincides for these two cases) in Table 1, with empirical sizes generally around 20%. This suggests that, even in relatively small samples, the impact of heteroskedasticity in the shocks on the empirical size of the tests largely dominates the impact of any weak dependence present, at least for the two heteroskedastic cases reported here; notice that in both of these cases $(C_0)_{1,1} = \lambda(B_0^{-1})_{1,1}$ (see Remark 4.10) so that the limiting null distributions of the asymptotic tests will not depend on any weak dependence present. In addition, the robust Wald test, RW_T, is seen to be quite unreliable in finite samples when both heteroskedasticity and weak dependence are present in the shocks. In contrast, the restricted wild bootstrap tests reported in Panel B of Table 3 again control size very well across the reported combinations of heteroskedasticity and weak dependence generally displaying empirical rejection frequencies close to the nominal level, albeit noting that a small degree of over-sizing is seen with the wild bootstrap LM_T, LR_T and W_T tests for $a - 0.8$ and $b = 0.8$ when $T = 100$. Somewhat larger distortions on average are seen with the unrestricted wild bootstrap tests in Panel C of Table 3, especially so in the case of the unrestricted wild bootstrap W_T and RW_T tests which can be significantly over-sized when $b = -0.8$. While we observed from the results in Tables 1 and 2 that the wild bootstrap RW_T test afforded the best size control amongst the four bootstrap tests when weak dependence was absent from the data, there is arguably somewhat less to choose between the four restricted wild bootstrap tests when weak dependence is present, although overall the RW_T test does still appear to offer the best size control amongst these four tests.

Consider next the power results under weak dependence in Table 4. The results in Panel A for the size-corrected asymptotic tests clearly demonstrate the dependence of

Table 3: Tests of $H_{0,1}$: simulated size with weakly dependent errors

			homoskedastic case				$\tau = 1/4$ and $\upsilon = 1/3$				$\tau = 3/4$ and $\upsilon = 3$			
a	b	T	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests														
	−0.8	100	4.74	5.51	17.08	21.10	7.80	12.36	22.05	20.68	8.57	14.16	24.31	18.25
	−0.8	250	6.48	7.03	16.56	19.01	12.09	17.31	27.20	20.98	12.47	20.23	30.55	22.25
	−0.8	∞	5.00	5.00	5.00	5.00	19.95	19.95	19.95	5.00	19.95	19.95	19.95	5.00
	0.8	100	5.82	5.27	5.51	5.87	17.20	15.99	16.62	8.24	18.74	17.45	18.23	8.48
	0.8	250	5.49	5.05	5.36	5.41	17.90	17.31	17.76	6.40	19.09	18.54	18.90	6.69
	0.8	∞	5.00	5.00	5.00	5.00	19.95	19.95	19.95	5.00	19.95	19.95	19.95	5.00
−0.8		100	6.10	5.14	5.75	6.49	18.59	16.77	18.06	10.18	20.79	18.32	20.10	10.63
−0.8		250	5.55	5.22	5.47	5.91	18.76	18.06	18.64	7.51	19.47	18.27	19.24	7.26
−0.8		∞	5.00	5.00	5.00	5.00	19.95	19.95	19.95	5.00	19.95	19.95	19.95	5.00
0.8		100	4.39	3.11	9.98	15.67	6.38	8.79	16.64	15.66	7.02	11.36	21.68	18.48
0.8		250	6.53	4.39	10.57	13.44	10.17	13.65	20.95	14.01	10.90	16.66	24.38	16.82
0.8		∞	5.00	5.00	5.00	5.00	19.95	19.95	19.95	5.00	19.95	19.95	19.95	5.00
Panel B: wild bootstrap tests (Algorithm 1)														
	−0.8	100	5.09	4.80	5.16	5.12	4.68	5.74	5.40	4.98	5.17	5.47	5.25	4.89
	−0.8	250	5.07	5.02	5.15	5.21	5.34	5.12	5.20	5.26	5.42	5.33	5.39	4.94
	−0.8	∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	0.8	100	5.00	5.09	4.97	4.77	6.65	6.61	6.60	5.39	6.63	6.61	6.70	5.43
	0.8	250	5.03	4.99	4.98	5.03	5.37	5.28	5.40	4.94	5.77	5.85	5.77	4.94
	0.8	∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
−0.8		100	4.94	5.13	4.92	4.94	6.56	6.75	6.59	5.58	6.71	6.99	6.70	5.54
−0.8		250	5.16	5.28	5.14	5.03	5.66	5.59	5.72	4.97	5.60	5.46	5.64	5.03
−0.8		∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
0.8		100	4.82	4.81	4.78	4.94	4.81	6.29	5.73	4.98	4.93	5.81	5.64	4.91
0.8		250	5.08	4.94	5.03	4.83	4.47	5.08	5.12	4.56	4.83	5.30	5.22	4.65
0.8		∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
Panel C: wild bootstrap tests (Algorithm 2)														
	−0.8	100	4.66	4.37	9.14	8.87	4.97	4.42	6.90	7.67	5.18	3.86	6.01	5.40
	−0.8	250	4.95	5.16	8.91	8.89	5.50	4.64	8.50	9.36	5.34	4.52	9.24	8.90
	−0.8	∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
	0.8	100	5.02	5.19	5.08	4.77	6.52	6.56	6.48	5.41	6.67	6.43	6.64	5.45
	0.8	250	5.05	5.02	5.02	5.01	5.47	5.25	5.42	4.95	5.83	5.84	5.83	4.92
	0.8	∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
−0.8		100	4.98	5.13	4.96	4.88	6.69	6.78	6.64	5.71	6.77	6.95	6.79	5.55
−0.8		250	5.13	5.17	5.13	5.07	5.69	5.62	5.69	4.97	5.58	5.46	5.69	5.01
−0.8		∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00
0.8		100	4.09	3.69	4.55	4.46	4.84	4.00	5.33	3.31	4.89	3.86	5.25	4.34
0.8		250	4.99	3.94	5.18	5.15	4.01	3.85	5.60	4.07	4.22	4.22	5.86	4.51
0.8		∞	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00

Notes: Entries for finite T are simulated rejection frequencies of the tests. Entries for $T = \infty$ are calculated as described in Remark 4.10. All entries are based on $10,000$ Monte Carlo replications.

Table 4: Tests of $H_{0,1}$: simulated power with weakly dependent errors

			homoskedastic case				$\tau = 1/4$ and $\upsilon = 1/3$				$\tau = 3/4$ and $\upsilon = 3$			
a	b	T	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests														
	−0.8	100	7.47	13.57	11.47	18.46	5.84	8.18	8.82	9.23	4.81	6.89	9.53	10.48
	−0.8	250	9.62	15.94	11.91	15.01	5.78	11.33	9.56	7.81	5.18	9.24	8.87	9.29
	−0.8	∞	25.52	25.52	25.52	25.52	13.62	13.62	13.62	13.62	13.62	13.62	13.62	13.62
	0.8	100	88.55	91.28	88.55	85.97	63.45	74.16	62.24	51.88	55.86	67.56	55.53	44.62
	0.8	250	91.06	93.11	91.05	89.88	63.91	71.48	63.79	56.53	58.87	67.29	58.85	53.16
	0.8	∞	95.09	95.09	95.09	95.09	65.74	65.74	65.74	65.74	65.74	65.74	65.74	65.74
−0.8		100	86.01	89.09	85.93	82.62	57.78	68.35	57.00	45.81	48.89	59.62	48.68	38.81
−0.8		250	90.42	92.25	90.40	89.04	59.15	66.65	58.58	52.35	56.89	64.35	56.36	49.96
−0.8		∞	95.09	95.09	95.09	95.09	65.74	65.74	65.74	65.74	65.74	65.74	65.74	65.74
0.8		100	32.93	49.90	15.20	24.26	30.67	38.12	10.29	20.26	25.24	31.06	8.84	16.45
0.8		250	15.52	39.39	12.62	21.31	17.78	28.79	7.39	17.35	14.03	23.47	6.26	14.00
0.8		∞	25.52	25.52	25.52	25.52	13.62	13.62	13.62	13.62	13.62	13.62	13.62	13.62
Panel B: wild bootstrap tests (Algorithm 1)														
	−0.8	100	6.46	14.12	10.89	15.49	6.38	11.20	8.01	10.92	5.11	7.41	7.00	7.16
	−0.8	250	9.83	17.57	12.72	15.86	7.49	13.53	9.71	9.95	6.17	11.07	9.13	8.93
	−0.8	∞	25.52	25.52	25.52	25.52	13.62	13.62	13.62	13.62	13.62	13.62	13.62	13.62
	0.8	100	88.23	91.07	88.09	85.05	66.26	74.92	64.87	52.57	58.78	68.59	57.91	45.92
	0.8	250	90.98	93.07	90.94	89.82	63.56	71.59	63.17	56.01	60.27	67.81	60.12	52.39
	0.8	∞	95.09	95.09	95.09	95.09	65.74	65.74	65.74	65.74	65.74	65.74	65.74	65.74
−0.8		100	85.33	88.77	85.27	82.13	60.57	70.01	59.97	45.02	51.81	62.63	52.46	39.19
−0.8		250	90.25	92.26	90.20	89.04	60.28	67.74	59.78	51.67	57.85	64.84	57.56	49.38
−0.8		∞	95.09	95.09	95.09	95.09	65.74	65.74	65.74	65.74	65.74	65.74	65.74	65.74
0.8		100	23.02	41.72	17.74	26.80	18.96	30.47	12.98	14.35	13.61	24.61	10.14	13.47
0.8		250	16.49	34.84	17.04	25.16	11.22	22.81	9.75	15.74	9.48	19.69	8.30	13.42
0.8		∞	25.52	25.52	25.52	25.52	13.62	13.62	13.62	13.62	13.62	13.62	13.62	13.62
Panel C: wild bootstrap tests (Algorithm 2)														
	−0.8	100	6.83	12.07	17.36	22.30	5.88	8.35	12.82	12.06	4.60	5.67	11.96	11.69
	−0.8	250	9.37	15.33	17.94	20.20	5.69	9.70	14.49	13.02	5.10	7.96	14.25	14.04
	−0.8	∞	25.52	25.52	25.52	25.52	13.62	13.62	13.62	13.62	13.62	13.62	13.62	13.62
	0.8	100	88.08	91.10	88.11	85.03	65.83	75.08	64.89	52.25	58.20	68.66	57.62	45.15
	0.8	250	90.90	93.09	90.90	89.81	63.47	71.58	62.95	55.70	60.23	67.92	59.97	52.29
	0.8	∞	95.09	95.09	95.09	95.09	65.74	65.74	65.74	65.74	65.74	65.74	65.74	65.74
−0.8		100	85.40	88.84	85.22	82.17	60.28	69.76	59.39	45.93	52.04	62.49	52.38	40.19
−0.8		250	90.25	92.19	90.24	89.04	60.22	67.69	59.60	52.01	57.88	64.87	57.54	49.74
−0.8		∞	95.09	95.09	95.09	95.09	65.74	65.74	65.74	65.74	65.74	65.74	65.74	65.74
0.8		100	30.39	48.76	12.52	21.60	31.53	40.99	11.82	20.09	27.49	33.33	9.07	15.43
0.8		250	15.43	38.36	10.07	18.23	16.93	28.85	6.38	15.21	15.00	24.53	5.93	11.97
0.8		∞	25.52	25.52	25.52	25.52	13.62	13.62	13.62	13.62	13.62	13.62	13.62	13.62

Notes: Entries for finite T are simulated rejection frequencies of the tests. Entries for $T = \infty$ are calculated as described in Remark 4.10. Power is measured at $\delta = 3$ and is size corrected for the asymptotic tests, but not for the bootstrap tests. All entries are based on 10,000 Monte Carlo replications.

the power of the asymptotic tests on both the degree of weak dependence present and on any heteroskedasticity present (even after controlling for the impact these have on the null distributions of the tests), again as predicted by Theorem 3; see also Remark 4.10. For example, when u_t is an MA(1) process with $b = -0.8$ the LM_T test has size-corrected power of under 8% for $T = 100$ in the homoskedastic case, yet for the AR(1) case with $a = -0.8$ this rises to about 86%. For all of the asymptotic tests, power is lowest, other things equal, for $b = -0.8$ and $a = 0.8$ and highest for $b = 0.8$ and $a = -0.8$, reflective of the value of the quantity $(C_0)_{1,1}$ in these cases, noted above, and its role in determining the value of the non-centrality parameter featuring in the local limiting distributions in (23) and (24) in Remark 4.10. When coupled with heteroskedastic effects, power can be diminished even further, most notably where $b = -0.8$ and $a = 0.8$. As with the results in Tables 1 and 2, the empirical power results in Panel B of Table 4 for the restricted wild bootstrap tests again lie close to the size-adjusted power results for the asymptotic tests in Panel A in most cases, although for $a = 0.8$ the power of the restricted bootstrap tests can in some cases lie somewhat below (but can also lie somewhat above) the size-adjusted powers of the asymptotic tests in finite samples. Consistent with the results in Tables 1 and 2, the LR tests again display the highest finite sample power among the four tests considered, for a given implementation, although again one should bear in mind the differences in the empirical sizes of the tests.

The results for the unrestricted bootstrap tests given in Panel C of Table 4 are generally very similar to those for the restricted bootstrap in Panel B, suggesting that overall the superior finite sample size control of the restricted bootstrap does not come at the cost of significantly reduced power relative to the unrestricted bootstrap. However, there are some cases, typically where $a = 0.8$, for which the unrestricted bootstrap can display somewhat higher finite sample power than the restricted bootstrap; for example, when $a = 0.8$ and $T = 100$ with $\tau = 1/4, \nu = 1/3$ the unrestricted bootstrap implementation of LM_T has power of about 32%, while the corresponding restricted bootstrap test has power of about 19%. This is not always the case, however; for example, where $a = 0.8$ and $T = 250$ the restricted bootstrap W_T test displays higher power than the unrestricted version. To investigate the relative power properties of the restricted and unrestricted bootstrap tests further, Figure S.1 in the Supplement graphs finite sample power functions of the tests for a subset of the models considered here. These results support the conclusions from the tables.

5.5 Summary of Results for Tests of $H_{0,2} : a = 0$ and $H_{0,3} : d = 1 \cap a = 0$

For the tests of $H_{0,2} : a = 0$, reported in Tables S.1 and S.2 in the accompanying Supplement, the conclusions we can draw from these results are qualitatively very similar to those drawn from the results in Tables 1 and 2. Again a significant deterioration is seen in the finite sample properties of the asymptotic tests under both unconditional and conditional heteroskedasticity in both Tables S.1 and S.2, relative to the corresponding results for conditionally homoskedastic errors in Table S.1. Of particular note is the poor performance of the robust Wald test, RW_T, which displays noticeably worse size distortions under heteroskedasticity than were seen in Tables 1 and 2. For the tests of $H_{0,3} : d = 1 \cap a = 0$, the conclusions drawn from the results in Tables S.3 and S.4 are again qualitatively similar to those drawn from the results in Tables 1 and 2, but with the observation that the size properties of the asymptotic tests can deteriorate considerably more here than was seen with the single parameter tests. Overall, the conclusions drawn from Tables S.1–S.4 regarding the wild bootstrap tests are much the same as those drawn for the tests on d discussed in Sections 5.2–5.4. The restricted wild bootstrap delivers very good size control throughout, again significantly better than is seen for the corresponding unrestricted wild bootstrap tests, with finite sample power of the tests based on the restricted wild bootstrap generally close to those of both the corresponding size-adjusted asymptotic tests and the unrestricted wild bootstrap tests (some exceptions are seen for the latter comparison, but it is important to observe that these occur where the unrestricted wild bootstrap tests suffer from significant over-sizing under the null while the corresponding restricted wild bootstrap tests do not).

Based on the simulation results reported in this section, coupled with the large sample properties of the tests detailed in Section 4, we recommend the use of the restricted wild bootstrap implementations of the LM_T, LR_T, W_T and RW_T tests in practice. Of these, the restricted wild bootstrap RW_T tests would appear to be preferred as it consistently delivers the best finite sample size control among the four bootstrap tests considered under both heteroskedasticity and weak dependence.

6 Concluding Remarks

We have made two contributions to the long memory literature. First, 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), continues 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 and any heteroskedasticity present in the shocks. 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.

Second, we have proposed classical asymptotic Wald, likelihood and Lagrange multiplier tests, and a robust Wald test formed using heteroskedastic-robust (sandwich-type) standard errors, for testing linear hypotheses on the long and/or short memory parameters of the heteroskedastic fractional time series model, together with wild bootstrap implementations of these tests. The latter were shown to yield tests which are asymptotically robust under the null to the heteroskedasticity in the shocks. Excepting the robust Wald test, this property was shown not to be shared by the asymptotic tests. The (global) consistency of our proposed bootstrap tests was established under fixed alternatives.

A simulation study highlighted both the potential for severe size distortions with the standard asymptotic tests in the presence of heteroskedastic shocks and the excellent job done by the bootstrap tests in controlling finite sample sizes here. The bootstrap tests were also shown to deliver considerably more reliable finite sample inference than the asymptotic tests in the homoskedastic case, particularly so for tests on the long memory parameter when weak dependence was present. The simulation study also compared the finite sample properties of using a bootstrap algorithm where the bootstrap sample data were generated using model estimates obtained under the null hypothesis (restricted) with one where they were estimated unrestrictedly. Based on these results we recommend the use of the wild bootstrap algorithm based on restricted estimates. Of the restricted wild bootstrap tests, that based on the robust Wald statistic appeared to deliver the best overall size control. This test also has the advantage that it is asymptotically valid under the same set of assumptions as are needed for establishing large sample theory for the asymptotic tests, while asymptotic validity for the wild bootstrap Wald, likelihood ratio and Lagrange multiplier tests requires an additional symmetry-type condition to hold on any conditional heteroskedasticity present.

A Preliminary Lemmas

For the purposes of Appendices A–D, all stated results and derivations shall be taken as conditional on $\sigma(\cdot)$. Due to the stochastic independence of $\{\sigma_t\}$ and $\{z_t\}$, see Assumption 1(b), and given the simple structure of conditional distributions on product spaces, this implies that $\{\sigma_t\}$ can be treated as fixed. In order to avoid repetition, this will not be repeated on every occasion. Where convergence obtains to a limit which does not depend on $\sigma(\cdot)$, it should be recalled that the stated convergence result also holds unconditionally.

This appendix presents a series of lemmas that will be used repeatedly in the proofs of our main results. The proofs of all lemmas are given in the Supplement.

Lemma A.1. *Let U_{Tt} be a martingale difference array with respect to some filtration $\mathcal{F}_t$ such that $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ for $t = ..., -1, 0, 1, 2, ...,$ and suppose, as $T \to \infty$, that*
(i) $\sum_{t=1}^{T} E(U_{Tt}^2 1\{|U_{Tt}| > \delta\}) \to 0$ for all $\delta > 0$,
(ii) Either (a) $\sum_{t=1}^{T} U_{Tt}^2 \xrightarrow{p} V$ or (b) $\sum_{t=1}^{T} E(U_{Tt}^2 | \mathcal{F}_{t-1}) \xrightarrow{p} V$.
Then $\sum_{t=1}^{T} U_{Tt} \xrightarrow{w} N(0, V)$ as $T \to \infty$.

The following 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 MDS is uncorrelated.

Lemma A.2. *Let z_t be a MDS with respect to the natural filtration $\mathcal{F}_t$, the sigma-field generated by $\{z_s\}_{s \le t}$, and suppose $E|z_t|^q < \infty$ for some integer $q \ge 2$. Then the q'th order moments and cumulants satisfy $E(z_t z_{t-r_1} \cdots z_{t-r_{q-1}}) = 0$ and $\kappa_q(t, t-r_1, \ldots, t-r_{q-1}) = 0$, for all integers $r_k \ge 1, k = 1, \ldots, q-1$.*

The next lemmas contains useful inequalities applied throughout the remaining proofs.

Lemma A.3. *Uniformly in $-u_0 \le v \le u \le u_0$ and for $j \ge 1, m \ge 0$ it holds that*

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

$$|\frac{\partial^m}{\partial u^m} T^{-u} \pi_j(u)| \le c(1 + \log |j/T|)^m T^{-u} j^{u-1}, \tag{A.2}$$

$$|\frac{\partial^m}{\partial u^m} \pi_j(u) - \frac{\partial^m}{\partial v^m} \pi_j(v)| \le c(u - v)(1 + \log j)^{m+1} j^{u-1}, \tag{A.3}$$

$$|\frac{\partial^m}{\partial u^m} T^{-u} \pi_j(u) - \frac{\partial^m}{\partial v^m} T^{-v} \pi_j(v)| \le c(u - v)(1 + \log |j/T|)^{m+1} T^{-v} j^{v-1}, \tag{A.4}$$

$$|\frac{\partial^m}{\partial u^m} \pi_{j+1}(u) - \frac{\partial^m}{\partial u^m} \pi_j(u)| \le c(1 + \log j)^m j^{u-2}, \tag{A.5}$$

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

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

$$\pi_j(-v) \ge c j^{-v-1}, \tag{A.6}$$

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

Lemma A.4. *Let u and v be such that* $\max(|u|, |v|) \le a$ *for some $a < \infty$. Then it holds that $\sum_{j=1}^{t-1} j^{u-1}(t-j)^{v-1} \le c(1 + \log t) t^{\max(u+v-1, u-1, v-1)}$, where the constant $c > 0$ does not depend on $u, v,$ or t.*

Lemma A.5. *Let σ_t satisfy Assumption 1(b). Let $\xi_n, n \geq 1$, be vector-valued coefficients and $g_{t,m,n}, t, n, m \geq 1$, be real coefficients. Suppose $\sum_{n,m=1}^{\infty} ||\xi_n|| ||\xi_m|| \sup_t |g_{t,n,m}| < \infty$. Then $T^{-1} \sum_{t=1}^{T} \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_n \xi_m' \sigma_{t-n} \sigma_{t-m} g_{t,n,m} = T^{-1} \sum_{t=1}^{T} \sigma_t^4 \sum_{n,m=1}^{t-1} \xi_n \xi_m' g_{t,n,m} + o(1)$.*

B Variation Bounds

This appendix contains three lemmas that are used to verify tightness and stochastic equicontinuity conditions for the processes in the proofs of the main theorems. The first deals with nonstationary processes and the next lemma with product moments of processes that are nearly stationary. Lemma B.2 contains the truncation argument used to deal with the non-uniform convergence in Θ_2. The third lemma covers product moments of stationary, nearly stationary, and nonstationary processes, and is applied in the consistency proof — both for stationary and nonstationary processes and to deal with certain cross-products of stationary and nearly stationary processes — and it is applied for the Hessian in the proof of asymptotic normality.

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

$$||T^{u+1/2} \Delta_+^u \varepsilon_t||_2 \leq c \text{ and } ||T^{u+1/2} \Delta_+^u \varepsilon_t - T^{v+1/2} \Delta_+^v \varepsilon_t||_2 \leq c|u - v|, \tag{B.1}$$

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

Lemma B.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 1, and define the product moments $M_{11NT}(u) := T^{-1} \sum_{t=N+1}^{T} w_{1t}^2 - E(T^{-1} \sum_{t=N+1}^{T} w_{1t}^2)$ and $M_{12NT}(u) := T^{-1} \sum_{t=N+1}^{T} w_{1t} w_{2t}$. Then, for any $\kappa \in (0, 1/2)$, if $N := \lfloor T^\alpha \rfloor$ with $0 < \alpha < \min(\frac{1/2-\kappa}{1/2+\kappa}, \frac{1/2}{1/2+2\kappa})$, where for any real number x, $\lfloor x \rfloor$ denotes the integer part of x, it holds that*

$$\sup_{|u+1/2| \leq \kappa} |M_{11NT}(u)| \xrightarrow{p} 0 \text{ and } \sup_{|u+1/2| \leq \kappa} |M_{12NT}(u)| \xrightarrow{p} 0. \tag{B.2}$$

Lemma B.3. *Let $Z_{it} := \sum_{n=0}^{\infty} \zeta_{in}(\psi) \varepsilon_{t-n}, i = 1, 2$, where ε_t satisfies Assumption 1 and the coefficients $\zeta_{in}(\psi)$ satisfy $\sum_{n=0}^{\infty} |\zeta_{in}(\psi)| < \infty, i = 1, 2$, uniformly in $\psi \in \tilde{\Psi} \subseteq \Psi$. The coefficients $\zeta_{in}(\psi)$ may depend on T as long as there exists a $T_0 \geq 1$ such that uniform (in ψ) absolute summability holds uniformly in $T \geq T_0$. Define the product moment $M_T(u_1, u_2, \psi) := T^{-1} \sum_{t=1}^{T} \frac{\partial^k}{\partial u_1^{(k)}} (\Delta_+^{u_1} Z_{1t}) \frac{\partial^l}{\partial u_2^{(l)}} (\Delta_+^{u_2} Z_{2t})$ for $k, l \geq 0$ and the sets $\tilde{\Theta} := \{(u_1, u_2, \psi) \in \mathbb{R} \times \mathbb{R} \times \tilde{\Psi} : \min(u_1+1, u_2+1, u_1+u_2+1) \geq a\}$ and $\bar{\Theta} := \{(u_1, u_2, \psi) \in \mathbb{R} \times \mathbb{R} \times \tilde{\Psi} : u_1 \leq -1/2 - \kappa_1, u_2 \leq -1/2 - \kappa_1\}$. Then*

$$\sup_{(u_1,u_2,\psi) \in \tilde{\Theta}} |M_T(u_1, u_2, \psi)| = O_p(1) \text{ for } a > 0, \tag{B.3}$$

$$\sup_{(u_1,u_2,\psi) \in \tilde{\Theta}} |M_T(u_1, u_2, \psi)| = O_p((\log T) T^{-a}) \text{ for } a \leq 0, \tag{B.4}$$

$$\sup_{(u_1,u_2,\psi) \in \bar{\Theta}} T^{1+u_1+u_2} |M_T(u_1, u_2, \psi)| = O_p(1) \text{ for } k, l = 0. \tag{B.5}$$

C Proof of Theorem 5

First note that $\tilde{\Theta}$ is convex and compact because Θ is convex and compact. The existence of $\tilde{\theta}$ follows because $\tilde{\Theta}$ is compact and $Q_T(\theta)$ is continuous.

C.1 Proof of Part (i)

As noted above, $\tilde{\Theta}$ satisfies all the assumptions on the parameter space from Theorem 1. By definition of $\tilde{\theta}$ we have $M'(\tilde{\theta} - \theta_{0,T}) = M'\tilde{\theta} - m - \delta/\sqrt{T} = -\delta/\sqrt{T} \to 0$, and the conclusion then follows from Theorems 1 and 2 because $M'_\perp\tilde{\theta}$ is unrestricted with true value in the interior of the parameter space.

C.2 Proof of Part (ii)

As in related work, e.g. the proof of Theorem 1, it is convenient to partition the parameter space due to the non-uniform convergence of the objective function. In this case we partition as $\tilde{\Theta}_1(\kappa) := \{\theta \in \tilde{\Theta} : d \le d_0 - 1/2 + \kappa\}$ and $\tilde{\Theta}_2(\kappa) := \{\theta \in \tilde{\Theta} : d \ge d_0 - 1/2 + \kappa\}$. Similarly define $\Theta_1(\kappa)$ and $\Theta_2(\kappa)$ in the obvious way. Because $\tilde{\Theta}$ is compact and convex, we can divide the proof into the following two cases: (a) $\tilde{\Theta}_2(\kappa) \ne \emptyset$ for some $\kappa > 0$; and (b) $\tilde{\Theta}_2(\kappa) = \emptyset$ for all $\kappa > 0$, and we treat these in turn. Intuitively, (a) is when $\tilde{\Theta}$ includes some $d > d_0 - 1/2$, and (b) is when $\tilde{\Theta}$ does not include any $d > d_0 - 1/2$, but may include the point $d = d_0 - 1/2$.

Case (a): In this case we will need two results from the proof of Theorem 1, given in the Supplement: It is shown in (S.20) and (S.21) that there exists a $\bar{\kappa} \in (0, \kappa)$ such that, for any $K > 0$,

$$P(\arg\min_{\theta \in \tilde{\Theta}_1(-\bar{\kappa})} Q_T(\theta) > K) \ge P(\arg\min_{\theta \in \Theta_1(-\bar{\kappa})} Q_T(\theta) > K) \to 1, \tag{C.1}$$

and that, for all $\kappa > 0$,

$$\sup_{\theta \in \Theta_2(\kappa)} |Q_T(\theta) - Q(\theta)| \xrightarrow{p} 0, \tag{C.2}$$

where $Q(\theta)$ is defined in (S.19). It follows from (C.1) and (C.2) that $P(\tilde{\theta} \in \tilde{\Theta}_2(\bar{\kappa})) \to 1$ as $T \to \infty$, and, hence, $\tilde{\theta} \xrightarrow{p} \theta^\dagger := \arg\min_{\theta \in \tilde{\Theta}_2(\bar{\kappa})} Q(\theta)$. We note that $\theta^\dagger$ exists because $\tilde{\Theta}_2(\bar{\kappa})$ is compact and $Q(\theta)$ is continuous, and, moreover, $\theta^\dagger \ne \theta_0$ because $\theta_0 \notin \tilde{\Theta}$.

Case (b): In this case, $\tilde{\Theta}$ contains only values of d such that $d \le d_0 - 1/2 - \kappa_1$ for some $\kappa_1 \ge 0$, and we need to normalize $Q_T(\theta)$ by an additional factor $(\log T)^{-1}$ when $d = d_0 - 1/2$ or $T^{2(d-d_0)+1}$ when $d < d_0 - 1/2$. We thus define the normalization $h(T, g) := (\log T)^{-1}\mathbb{I}(g = 0) + T^{-2g}\mathbb{I}(g > 0)$.

Let $\epsilon > 0$ be an arbitrary, fixed constant. First, as in (C.1), see also Section S.5.1.2 in the Supplement for details, it holds that

$$P(\arg\min_{\theta \in \tilde{\Theta}_1(-\epsilon-\kappa_1)} h(T, \kappa_1) Q_T(\theta) > K) \to 1 \tag{C.3}$$

because $T^{2(\epsilon+\kappa_1)}h(T, \kappa_1) \to \infty$ when $\epsilon > 0, \kappa_1 \ge 0$. Next, we analyze $h(T, \kappa_1)Q_T(\theta)$ at the point $d = d_0 - 1/2 - \kappa_1$. Define the untruncated innovations $e_t(\psi) := c(L, \psi)\varepsilon_t = \sum_{n=0}^\infty c_n(\psi)\varepsilon_{t-n}$ with $c(z, \psi) := b(z, \psi)a(z, \psi_0) = \frac{a(z,\psi_0)}{a(z,\psi)} = \sum_{n=0}^\infty c_n(\psi)z^n$, which are such that $\sup_{\theta \in \{\theta \in \tilde{\Theta} : d = d_0 - 1/2 - \kappa_1\}} |h(T, \kappa_1)Q_T(\theta) - h(T, \kappa_1)T^{-1}\sum_{t=1}^T (\Delta_+^{-1/2-\kappa_1} e_t(\psi))^2| \xrightarrow{p} 0$, by slight modification of Lemma S.2 in the Supplement. Thus, we may consider $U_T(\psi) := h(T, \kappa_1)T^{-1}\sum_{t=1}^T (\Delta_+^{-1/2-\kappa_1} e_t(\psi))^2$ instead of $h(T, \kappa_1)Q_T(\theta)$. We apply the Beveridge-Nelson decomposition,

$$e_t(\psi) = c(L, \psi)\varepsilon_t = \left(\sum_{n=0}^\infty c_n(\psi)\right)\varepsilon_t + \sum_{n=0}^\infty \tilde{c}_n(\psi)\Delta\varepsilon_{t-n}, \tag{C.4}$$

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 $\psi \in \Psi$ by Assumption 3 and Lemma A.4, see also Phillips and Solo (1992, Lemma 2.1). This implies, in particular, that $\sum_{n=0}^{\infty} |\tilde{c}_n(\psi)| < \infty$ uniformly in $\psi \in \Psi$. With the additional normalization by $h(T, \kappa_1)$ we thus have

$$U_T(\psi) = A(\psi)^2 B_T + h(T, \kappa_1) R_T(\psi), \tag{C.5}$$

where $A(\psi) := \sum_{n=0}^{\infty} c_n(\psi)$, $B_T := h(T, \kappa_1) T^{-1} \sum_{t=1}^{T} (\Delta_+^{-1/2-\kappa_1} \varepsilon_t)^2$, and by Lemmas B.1 and B.3 we find $\sup_{\theta \in \{\theta \in \tilde{\Theta}: d = d_0 - 1/2 - \kappa_1\}} |R_T(\psi)| = O_p(1)$. Since $h(T, \kappa_1) = o(1)$ for all $\kappa_1 \geq 0$, it holds that $h(T, \kappa_1) R_T(\psi)$ is asymptotically negligible, uniformly in $\psi \in \Psi$. The term $A(\psi)$ clearly depends only on ψ and is non-random and bounded uniformly in $\psi \in \Psi$. The term B_T is non-negative, and we find from Assumption 1(b) and (A.1) of Lemma A.3 that$E(B_T) = h(T, \kappa_1) T^{-1} \sum_{t=1}^{T} \sum_{j=0}^{t-1} \pi_j (1/2 + \kappa_1)^2 \sigma_{t-j}^2 \leq K h(T, \kappa_1) T^{-1} \sum_{t=1}^{T} \sum_{j=0}^{t-1} j^{-1+2\kappa_1} \leq K$,, which implies $B_T = O_p(1)$.

From (C.3) it thus follows that $P(|\tilde{d} - d_0 + 1/2 - \kappa_1| > \epsilon) = P(\tilde{d} \leq d_0 - 1/2 - \kappa_1 - \epsilon) = P(\tilde{\theta} \in \tilde{\Theta}_1(-\epsilon - \kappa_1)) \to 0$. Because $\epsilon > 0$ was arbitrary, it follows that $\tilde{d} \xrightarrow{p} d_0 - 1/2 - \kappa_1 = d^\dagger$ as $T \to \infty$. Finally, from Assumption 1(b) and (A.6) of Lemma A.3 it follows that $E(B_T) \geq c$, so that the first term on the right-hand side of (C.5) asymptotically dominates the second term, and we then find $\psi^\dagger$ from the first term on the right-hand side of (C.5) and conclude for case (b) that $\psi^\dagger = \arg\min_{\psi \in \Psi} A(\psi)^2$.

D Proof of Theorem 6

We first give two results which are applied several times. Next, Lemma D.3 is a bootstrap version of Lemma B.2 designed to deal with the non-uniform convergence in $\check{\Theta}_2$.

Lemma D.1. *Under the conditions of Theorem 6,* $T^{-1} \sum_{t=1}^{T} (\hat{\varepsilon}_{c,t}^2 - \varepsilon_t^2)^2 = O_p(T^{-1/2})$.

Lemma D.2. *Let* ε_t^* *be defined as in Algorithms 1 or 2 and suppose the conditions of Theorem 6 are satisfied. Suppose also that the coefficients* $\lambda_j(\theta)$ *satisfy* $\sup_\theta |\lambda_j(\theta)| \leq cj^g$ *and* $\sup_\theta |\lambda_{j+1}(\theta) - \lambda_j(\theta)| \leq cj^{g-1}$*, where* g *is fixed and* $|g| < \infty$*. Introduce the notation* h *for a positive integer, which in the following can be either* $h = k + 1$ *or* $h \leq m - 1$*. Then, uniformly in* $1 \leq m \leq k \leq T$*,* $E^* \sup_\theta \left| \sum_{j=m}^{k} \lambda_j(\theta) \sum_{t=\max(j,h)+1}^{T} \varepsilon_{t-j}^* \varepsilon_{t-h}^* \right| = \mathbb{I}(g > -1/2) O_p(T^{1/2} k^{1/2+g}) + \mathbb{I}(g < -1/2) O_p(T^{1/2} m^{1/2+g}) + \mathbb{I}(g = -1/2) O_p(T^{1/2} (\log k))$.

Lemma D.3. *Let* ε_t^* *be defined as in Algorithms 1 or 2 and let the conditions of Theorem 6 be satisfied. Let* $w_{1t}^* := \sum_{n=0}^{N-1} \pi_n(-u) \varepsilon_{t-n}^*$ *and* $w_{2t}^* := \sum_{n=N}^{t-1} \pi_n(-u) \varepsilon_{t-n}^*$ *and define the product moments* $M_{11NT}^*(u) := T^{-1} \sum_{t=N+1}^{T} (w_{1t}^{*2} - \sum_{n=0}^{N-1} \pi_n(-u)^2 \sigma_{t-n}^2)$ *and* $M_{12NT}^*(u) := T^{-1} \sum_{t=N+1}^{T} w_{1t}^* w_{2t}^*$*. For any* $\kappa \in (0, 1/2)$*, 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* $\sup_{|u+1/2| \leq \kappa} |M_{11NT}^*(u)| \xrightarrow{p^*}_p 0$ *and* $\sup_{|u+1/2| \leq \kappa} |M_{12NT}^*(u)| \xrightarrow{p^*}_p 0$.

D.1 Proof of Consistency: Eqn. (27)

First we define the untruncated process

$$e_t^*(\psi) := \check{c}(L, \psi) \varepsilon_t^* = \sum_{n=0}^{\infty} \check{c}_n(\psi) \varepsilon_{t-n}^*, \tag{D.6}$$

$$\check{c}(z, \psi) := b(z, \psi) a(z, \check{\psi}) = \frac{a(z, \check{\psi})}{a(z, \psi)} = \sum_{n=0}^{\infty} \check{c}_n(\psi) z^n. \tag{D.7}$$

Conditional on the original data, $\check{\theta}$, and hence $\check{\psi}$, is fixed, and from Assumption 3 and Lemma A.4 the coefficients $\check{c}_n(\psi)$ then satisfy

$$|\check{c}_n(\psi)| = O(n^{-2-\zeta}) \text{ uniformly in } \psi \in \Psi. \tag{D.8}$$

In general, $\check{\theta}$ is random, in which case the bound (D.8) applies almost surely because $\check{\psi} \in \Psi$ almost surely. Also note that, although $\check{\theta}$ depends on T, we have suppressed the T subscript on the triangular array $e_t(\psi)$ and on the coefficients $c_n(\psi)$.

Next, we partition the parameter space into three disjoint sets depending on the bootstrap true value, $\check{d}$: $\check{D}_1 := D \cap \{d : d - \check{d} \leq -1/2 - \kappa_1\}$, $\check{D}_2 := D \cap \{d : -1/2 - \kappa_1 \leq d - \check{d} \leq -1/2 + \kappa_2\}$, and $\check{D}_3 := D \cap \{d : -1/2 + \kappa_2 \leq d - \check{d}\}$. Note that these sets are random and depend on T since $\check{d}$ is random and depends on T. This presents an additional complication compared with related work, e.g. the proof of consistency of the QML estimator in Theorem 1, so we will need also $D_1^\dagger := D \cap \{d : d - d^\dagger \leq -1/2 - \kappa_1/2\}$, $D_2^\dagger := D \cap \{d : -1/2 - 2\kappa_1 \leq d - d^\dagger \leq -1/2 + 2\kappa_2\}$, and $D_3^\dagger := D \cap \{d : -1/2 + \kappa_2/2 \leq d - d^\dagger\}$, which are non-random and do not depend on T. Note that the $D_i^\dagger$ are defined such that, by definition of $d^\dagger$,

$$P(D_1^\dagger \supseteq \check{D}_1) = P(|\check{d} - d^\dagger| \leq \kappa_1/2) \to 1, \tag{D.9}$$

$$P(D_2^\dagger \supseteq \check{D}_2) = P(|\check{d} - d^\dagger| \leq \kappa_1 \cap |\check{d} - d^\dagger| \leq \kappa_2) \to 1, \tag{D.10}$$

$$P(D_3^\dagger \supseteq \check{D}_3) = P(|\check{d} - d^\dagger| \leq \kappa_2/2) \to 1, \tag{D.11}$$

which is used below. We further define $\check{\Theta}_i := \check{D}_i \times \Psi$ and $\Theta_i^\dagger := D_i^\dagger \times \Psi$ for $i = 1, 2, 3$. Clearly, $\theta^\dagger \in \Theta_3^\dagger$ and if $d_1 > d^\dagger - 1/2$ then the choice $\kappa_2 = d_1 - d^\dagger + 1/2 > 0$ implies that $\Theta_1^\dagger$ and $\Theta_2^\dagger$ are empty in which case the proof is easily simplified accordingly.

The general strategy of the proof relies on analyzing these parts of the parameter space separately, as is also the case in related work, e.g. in the proof of Theorem 1, with appropriate modifications to account for the randomness of the sets and for the fact that the analysis is conducted under the bootstrap probability measure. The latter sometimes implies a simplification, e.g., because the ε_t^* are independent conditional on the original data, in which case we give some of the arguments in the Supplement. First, we prove that for any $K > 0$ there exists a (fixed) $\bar{\kappa}_2 > 0$ such that

$$P^*\Big(\inf_{\theta \in \check{\Theta}_1(\kappa_1) \cup \check{\Theta}_2(\kappa_1, \bar{\kappa}_2)} Q_T^*(\theta) > K\Big) \xrightarrow{p} 1 \text{ as } T \to \infty. \tag{D.12}$$

This implies that $P^*(\hat{\theta}^* \in \check{\Theta}_3(\bar{\kappa}_2)) \xrightarrow{p} 1$ as $T \to \infty$, so that the relevant parameter space is reduced to $\check{\Theta}_3(\bar{\kappa}_2)$, which is subsequently analyzed.

Before we analyze each of the sets $\check{\Theta}_i$, we first note the following important simplification in the bootstrap residual and the bootstrap objective function. The proof is a nearly trivial consequence of the definition of ε_t^* in step (iii) of Algorithms 1 and 2, and is given in Section S.6.4 in the Supplement.

Lemma D.4. *Under the conditions of Theorem 6, the bootstrap residual satisfies* $\varepsilon_t^*(\theta) = \Delta_+^{d-\check{d}} e_t^*(\psi)$, *and hence*

$$\sup_{\theta \in \check{\Theta}_1} |T^{2(d-\check{d})} \sum_{t=1}^T \varepsilon_t^*(\theta)^2 - T^{2(d-\check{d})} \sum_{t=1}^T (\Delta_+^{d-\check{d}} e_t^*(\psi))^2| = 0, \tag{D.13}$$

$$\sup_{\theta \in \check{\Theta}_2 \cup \check{\Theta}_3} |T^{-1} \sum_{t=1}^T \varepsilon_t^*(\theta)^2 - T^{-1} \sum_{t=1}^T (\Delta_+^{d-\check{d}} e_t^*(\psi))^2| = 0. \tag{D.14}$$

D.1.1 Analysis of $\check{\Theta}_1$ We first note that

$$\inf_{\theta \in \check{\Theta}_1} Q_T^*(\theta) \geq T^{2\kappa_1} \inf_{\theta \in \check{\Theta}_1} T^{2(d-\check{d})+1} Q_T^*(\theta) \tag{D.15}$$

because $2(\check{d}-d)-1 \geq 2\kappa_1 > 0$ on $d \in \check{D}_1$. By Lemma D.4 we analyze $T^{2(d-\check{d})} \sum_{t=1}^T (\Delta_+^{d-\check{d}} e_t^*(\psi))^2$ instead of $T^{2(d-\check{d})+1} Q_T^*(\theta)$. We apply the Beveridge-Nelson decomposition (C.4) to $e_t^*(\psi)$,

$$e_t^*(\psi) = \check{c}(L,\psi)\varepsilon_t^* = \left(\sum_{n=0}^{\infty} \check{c}_n(\psi)\right) \varepsilon_t^* + \sum_{n=0}^{\infty} \bar{c}_n(\psi) \Delta \varepsilon_{t-n}^*, \tag{D.16}$$

where $0 < |\sum_{n=0}^{\infty} \check{c}_n(\psi)| < \infty$ almost surely uniformly in $\psi \in \Psi$ and $\bar{c}_n(\psi) = -\sum_{k=n+1}^{\infty} \check{c}_k(\psi)$ satisfies $|\bar{c}_n(\psi)| \leq c n^{-1-\zeta}$ almost surely uniformly in $\psi \in \Psi$ because $\check{\psi} \in \Psi$, see (D.8). This implies, in particular, that $\sum_{n=0}^{\infty} |\bar{c}_n(\psi)| < \infty$ almost surely uniformly in $\psi \in \Psi$. The relevant product moment can then be decomposed as

$$T^{2(d-\check{d})} \sum_{t=1}^T (\Delta_+^{d-\check{d}} e_t^*(\psi))^2 \geq \left(\sum_{n=0}^{\infty} \check{c}_n(\psi)\right)^2 \check{M}_T^*(d) + q_{1,T}^*(\theta), \tag{D.17}$$

where we have defined $\check{M}_T^*(d) := T^{2(d-\check{d})} \sum_{t=1}^T (\Delta_+^{d-\check{d}} \varepsilon_t^*)^2$. The proof that $\sup_{\theta \in \check{\Theta}_1} |q_{1,T}(\theta)| = o_p^*(1)$, in probability, is relatively straightforward, due to the independence of ε_t^*, conditional on the original data, and is given in Section S.6.5 in the Supplement.

Next, by the Cauchy-Schwarz inequality,

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

and we now apply Lemma A.1 with $U_{Tt}^* = T^{d-\check{d}-1/2} \pi_{T-t}(\check{d}-d+1)\varepsilon_t^*$. We first verify the Lindeberg condition (i) by proving the sufficient Lyapunov condition that $\sum_{t=1}^T E^*(U_{Tt}^{*4}) \xrightarrow{p} 0$. Letting $\mu_4 := E(w_t^4)$, we find that

$$\begin{aligned} \sum_{t=1}^T E^*(U_{Tt}^{*4}) &= \mu_4 T^{-2} \sum_{t=1}^T T^{4(d-\check{d})} \pi_{T-t}(\check{d}-d+1)^4 \hat{\varepsilon}_{c,t}^4 \leq \mu_4 T^{-2} \sum_{t=1}^T c \left(\frac{T-t}{T}\right)^{4(\check{d}-d)} \hat{\varepsilon}_{c,t}^4 \\ &\leq c T^{-2} \sum_{t=1}^T \left(\frac{T-t}{T}\right)^{2+4\kappa_1} \hat{\varepsilon}_{c,t}^4 \leq c T^{-2} \sum_{t=1}^T \hat{\varepsilon}_{c,t}^4 = O_p(T^{-1}), \end{aligned}$$

where the first inequality is by Lemma A.3 and the second applies the definition of $\check{D}_1$.

Secondly, we verify the conditional variance condition (ii)(b) of Lemma A.1. By independence, conditional on the original data, of ε_t^* we find that

$$\sum_{t=1}^T E^*(U_{Tt}^{*2} | \mathcal{F}_{t-1}^*) = V^{\dagger}(d) + q_{2,T}(d), \tag{D.18}$$

where $V^{\dagger}(d) := \frac{1}{\Gamma(d^{\dagger}-d+1)^2} \int_0^1 (1-s)^{2(d^{\dagger}-d)} \sigma(s)^2 \mathsf{d}s$. Again, showing $\sup_{d \in \check{D}_1} |q_{2,T}(d)| = o_p^*(1)$, in probability, is straightforward and the proof is in Section S.6.6 in the Supplement. This verifies Lemma A.1 condition (ii)(b), and hence $\sum_{t=1}^T U_{Tt}^* \xrightarrow{w^*}_p N(0, V^{\dagger}(d))$. It follows that $G_T^*(d) := (T^{d-\check{d}-1/2} \sum_{t=1}^T \pi_{T-t}(\check{d}-d+1)\varepsilon_t^*)^2 \xrightarrow{w^*}_p V^{\dagger}(d)\chi_1^2$, for any $d \in \check{D}_1$ pointwise.

To strengthen the pointwise convergence of $G_T^*(d)$ to hold uniformly, it is sufficient to show that $G_T^*(d)$ is tight (in probability) as a stochastic process on the space of continuous functions indexed by d. However, that only works on fixed intervals; in this case $D_1^\dagger$. Thus, we show that on $D_1^\dagger$ the process $G_T^*(d)$ is tight as a function of d. Using the result (D.9), this would also be sufficient for tightness of $G_T^*(d)$ on the smaller set $\check{D}_1$. We prove tightness using the bootstrap equivalent of the moment condition in Billingsley (1968, Theorem 12.3), see also Swensen (2003, p. 114), which requires showing that $G_T^*(d)$ is tight for fixed $d \in D_1^\dagger$ and that

$$E^*(G_T^*(u_1) - G_T^*(u_2))^2 = (u_1 - u_2)^2 O_p(1), \tag{D.19}$$

for some $O_p(1)$ term that does not depend on T, u_1, or u_2. By the mean value theorem, $G_T^*(u_1) - G_T^*(u_2) = (u_1 - u_2)\frac{\partial G_T^*(u)}{\partial u}|_{u=u_3}$, where u_3 is an intermediate value between u_1 and u_2. We note that the derivative only adds a logarithmic factor, see (A.2), so the same proof as for the probability limit of the variance of U_{Tt}^* above shows that $E^*(G_T^*(u_1)-G_T^*(u_2))^2 = (u_1-u_2)^2 O_p(1)$, and hence that $G_T^*(d)$ is tight (in probability) on $D_1^\dagger$. We can then apply the continuous mapping theorem along with (D.15) to conclude that $\inf_{\theta\in\check{\Theta}_1} Q_T^*(\theta) \geq T^{2\kappa_1} \inf_{\theta\in\check{\Theta}_1}(\sum_{n=0}^\infty \check{c}_n(\psi))^2 G_T^*(d) + o_p^*(1)$, in probability. Because $\sum_{n=0}^\infty \check{c}_n(\psi) > 0$ almost surely uniformly in $\psi \in \Psi$ and $G_T^*(d) > 0$ almost surely, this shows that, for any $K > 0$ and any $\kappa_1 > 0$,

$$P^*(\inf_{\theta\in\check{\Theta}_1} Q_T^*(\theta) > K) \xrightarrow{p} 1. \tag{D.20}$$

D.1.2 Analysis of $\check{\Theta}_2$ This analysis is similar to that of $\check{\Theta}_1$ in the sense that we want to show that on $\check{\Theta}_2$ the objective function $Q_T^*(\theta)$ is arbitrarily large, uniformly in $\theta \in \check{\Theta}_2$. We define $\check{v} := d - \check{d}$, $R_{1T}^*(\check{v}) := T^{-1}\sum_{t=1}^T(\Delta_+^{\check{v}}\varepsilon_t^*)^2$, and $R_{2T}^*(\check{v},\psi) := T^{-1}\sum_{t=1}^T(\Delta_+^{\check{v}}\varepsilon_t^*)(\sum_{n=0}^\infty \bar{c}_n(\psi)\Delta_+^{1+\check{v}}\varepsilon_{t-n}^*)$, and apply (D.16) to obtain the lower bound

$$Q_T^*(\theta) \geq \left(\sum_{n=0}^\infty \check{c}_n(\psi)\right)^2 R_{1T}^*(\check{v}) + 2\left(\sum_{n=0}^\infty \check{c}_n(\psi)\right) R_{2T}^*(\check{v},\psi), \tag{D.21}$$

where $0 < \sum_{n=0}^\infty \check{c}_n(\psi) < \infty$ almost surely uniformly in $\psi \in \Psi$. Showing $\sup_{\theta\in\check{\Theta}_2}|R_{2T}^*(\check{v},\psi)| = o_p^*(1)$, in probability, is relatively straightforward, due to the independence of ε_t^* conditional on the original data, so this proof is given in Section S.6.7 in the Supplement.

For $R_{1T}^*(\check{v})$, we apply the truncation argument in Lemma D.3. Define $w_{1t}^* := \sum_{n=0}^{N-1}\pi_n(-\check{v})\varepsilon_{t-n}^*$ and $w_{2t}^* := \sum_{n=N}^{t-1}\pi_n(-\check{v})\varepsilon_{t-n}^*$ so that

$$\begin{aligned} R_{1T}^*(\check{v}) &\geq T^{-1}\sum_{t=N+1}^T(\Delta_+^{\check{v}}\varepsilon_t^*)^2 \geq T^{-1}\sum_{t=N+1}^T w_{1t}^{*2} + 2T^{-1}\sum_{t=N+1}^T w_{1t}^* w_{2t}^* \\ &= T^{-1}\sum_{t=N+1}^T\sum_{n=0}^{N-1}\pi_n(\check{d}-d)^2\sigma_{t-n}^2 + q_{3,T}(d) \\ &\geq \left(\inf_{0\leq s\leq 1}\sigma(s)^2\right) T^{-1}(T-N)F_N(\check{d}-d) + q_{3,T}(d) + q_{4,T}(d), \end{aligned} \tag{D.22}$$

where $F_N(u) = \sum_{n=0}^{N-1}\pi_n(-u)^2$, and where $\sup_{d\in\check{D}_2}|q_{3,T}(d)| = o_p^*(1)$, in probability, by Lemma D.3 and $\sup_{d\in\check{D}_2}|q_{4,T}(d)| = o_p^*(1)$, in probability, by Assumption 1(b).

Thus, we find $Q_T^*(\theta) \geq \left(\sum_{n=0}^{\infty} \check{c}_n(\psi)\right)^2 (\inf_{0\leq s\leq 1} \sigma(s)^2)\, T^{-1}(T-N)F_N(\check{d}-d) + q_{5,T}(\theta)$, where $\sup_{\theta\in\check{\Theta}_2} |q_{5,T}(\theta)| = o_p^*(1)$, in probability. From Lemma A.3 of Nielsen (2015) we have that $\inf_{u\leq -1/2+a} F_N(u) \geq 1 + c(2a)^{-1}\left(1-(N-1)^{-2a}\right)$ for $N\geq 2$ and $a>0$, where the constant $c>0$ does not depend on a or N. We apply this result with $u = \check{d} - d \leq -1/2 + \kappa_2$ and $N = \lfloor T^{1/6} \rfloor$, thus satisfying the assumptions of Lemma D.3, i.e., $F_{T^{1/6}}(d-d_0) \geq 1 + c(2\kappa_2)^{-1}(1-(T-1)^{-2\kappa_2/6})$. 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 $(2\kappa_2)^{-1}(1-(T-1)^{-2\kappa_2/6}) \to \infty$ as $(\kappa_2, T) \to (0,\infty)$. In view of (D.20), and because $\inf_{\psi\in\Psi} \sum_{n=0}^{\infty} \check{c}_n(\psi) > 0$ almost surely and $\inf_{0\leq s\leq 1} \sigma(s)^2 > 0$, the result (D.12) follows.

D.1.3 Analysis of $\check{\Theta}_3$ Analogously to X_t^*, $\varepsilon_t^*(\theta)$, and $Q_T^*(\theta)$, we define $X_t^\dagger := \Delta_+^{-d^\dagger} a(L,\psi^\dagger)\varepsilon_t$, $\varepsilon_t^\dagger(\theta) := \sum_{n=0}^{t-1} b_n(\psi)\Delta_+^d X_{t-n}^\dagger$ and $Q_T^\dagger(\theta) := T^{-1}\sum_{t=1}^T \varepsilon_t^\dagger(\theta)^2$. The objective function $Q_T^\dagger(\theta)$ has the two distinct advantages that it depends on the original errors (not the bootstrap errors as in $Q_T^*(\theta)$) and that the "true" value in $Q_T^\dagger(\theta)$ is $\theta^\dagger$, which is fixed (unlike $Q_T^*(\theta)$ where the "true" value is $\check{\theta}$, which is random in general).

Thus, defining also $\hat{\theta}^\dagger := \arg\min_{\theta\in\Theta_3^\dagger} Q_T^\dagger(\theta)$ and $\tilde{\theta}^\dagger := \arg\min_{\theta\in\check{\Theta}_3} Q_T^\dagger(\theta)$, we can apply Theorem 1 directly to conclude that

$$\hat{\theta}^\dagger \xrightarrow{p} \theta^\dagger. \tag{D.23}$$

Furthermore, it holds that

$$\hat{\theta}^\dagger - \tilde{\theta}^\dagger \xrightarrow{p} 0 \text{ and } \check{\theta} - \theta^\dagger \xrightarrow{p} 0; \tag{D.24}$$

in the latter case by definition of $\theta^\dagger$ and in the former because $P(|\hat{\theta}^\dagger - \tilde{\theta}^\dagger| > \epsilon) = P(\tilde{\theta}^\dagger \in \check{\Theta}_3 \backslash \Theta_3^\dagger) = P(\tilde{d}^\dagger \in \check{D}_3 \backslash D_3^\dagger) \leq P(\check{D}_3 \supsetneq D_3^\dagger) \to 0$, by (D.11).

Finally, suppose we can prove that

$$\arg\min_{\theta\in\check{\Theta}_3} Q_T^*(\theta) - \tilde{\theta}^\dagger \xrightarrow{p^*}_p 0. \tag{D.25}$$

With P^*-probability converging to one in probability, the first term in (D.25) is $\hat{\theta}^*$, see (D.12), so the required result follows by combining (D.23)–(D.25). We therefore prove

$$\sup_{\theta\in\check{\Theta}_3} |Q_T^*(\theta) - Q_T^\dagger(\theta)| \xrightarrow{p^*}_p 0, \tag{D.26}$$

which implies (D.25).

To show (D.26), first note that by (S.24) in the Supplement we can replace $Q_T^\dagger(\theta)$ with $T^{-1}\sum_{t=1}^T (\Delta_+^{d-d^\dagger} \sum_{n=0}^\infty c_n^\dagger(\psi)\varepsilon_{t-n})^2$, where $c_n^\dagger(\psi)$ is defined as in (D.7) with $\check{\psi}$ replaced by $\psi^\dagger$. Then decompose

$$Q_T^*(\theta) - \frac{1}{T}\sum_{t=1}^T (\Delta_+^{d-d^\dagger} \sum_{n=0}^\infty c_n^\dagger(\psi)\varepsilon_{t-n})^2 = Q_T^*(\theta) - E^* Q_T^*(\theta) \tag{D.27}$$

$$+ E^* Q_T^*(\theta) - T^{-1}\sum_{t=1}^T (\Delta_+^{d-d^\dagger} \sum_{n=0}^\infty c_n^\dagger(\psi)\varepsilon_{t-n})^2 \tag{D.28}$$

and write $\varepsilon_t^*(\theta) = \sum_{n=0}^{t-1} \check{\varphi}_n(\theta)\varepsilon_{t-n}^*$, where $\check{\varphi}_n(\theta) := \sum_{m=0}^n \pi_m(\check{d}-d)\check{c}_{n-m}(\psi)$ satisfies

$$|\check{\varphi}_n(\theta)| = O(n^{\max(\check{d}-d-1,-2-\zeta)}) \text{ uniformly in } \psi \in \Psi \tag{D.29}$$

by Lemmas A.3 and A.4 and (D.8). By uncorrelatedness of ε_t^* conditional on the original data, (D.27) is

$$Q_T^*(\theta) - E^* Q_T^*(\theta) = T^{-1} \sum_{t=1}^{T} \sum_{n=0}^{t-1} \breve{\varphi}_n(\theta)^2 (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{c,t-n}^2) \tag{D.30}$$

$$+ 2T^{-1} \sum_{t=1}^{T} \sum_{n=0}^{t-1} \sum_{m=n+1}^{t-1} \breve{\varphi}_n(\theta) \breve{\varphi}_m(\theta) \varepsilon_{t-n}^* \varepsilon_{t-m}^*. \tag{D.31}$$

Noting that, conditionally on the original sample, $\varepsilon_t^{*2} - \hat{\varepsilon}_{c,t}^2 = \hat{\varepsilon}_{c,t}^2 (w_t^2 - 1)$ is a MDS with respect to $\mathcal{F}_t^*$, it follows that, defining $\eta_4 := E((w_t^2 - 1)^2)$,

$$\left(E^* \left| \sum_{t=n+1}^{T} (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{c,t-n}^2) \right| \right)^2 \leq \sum_{t,s=n+1}^{T} E^* (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{c,t-n}^2)(\varepsilon_{s-n}^{*2} - \hat{\varepsilon}_{c,s-n}^2) = \sum_{t=n+1}^{T} E^* (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{c,t-n}^2)^2$$

$$\leq \eta_4 \sum_{t=n+1}^{T} \hat{\varepsilon}_{c,t-n}^4 \leq \eta_4 \sum_{t=1}^{T} \hat{\varepsilon}_{c,t}^4 = O_p(T) \tag{D.32}$$

uniformly in $0 \leq n \leq T-1$. Thus, reversing the order of the summations in (D.30) and using (D.29) and (D.32), we find

$$E^* \sup_{\theta \in \breve{\Theta}_3} |(D.30)| \leq \sup_{\theta \in \breve{\Theta}_3} T^{-1} \sum_{n=0}^{T-1} \breve{\varphi}_n(\theta)^2 E^* \left| \sum_{t=n+1}^{T} (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{c,t-n}^2) \right|$$

$$\leq c \sup_{\theta \in \breve{\Theta}_3} T^{-1/2} \sum_{n=0}^{T-1} \breve{\varphi}_n(\theta)^2 \leq c T^{-1/2} \sum_{n=0}^{T-1} n^{-1-2\kappa_2} \leq c T^{-1/2},$$

which shows that $\sup_{\theta \in \breve{\Theta}_3} |(D.30)| = O_p^*(T^{-1/2})$, in probability.

To deal with (D.31), we apply Lemma D.2 with $g = -1/2 - \kappa_2$ to conclude that $E^* \sup_{\theta \in \breve{\Theta}_3} \left| \sum_{m=n+1}^{T-1} \breve{\varphi}_m(\theta) \sum_{t=m+1}^{T} \varepsilon_{t-n}^* \varepsilon_{t-m}^* \right| = O_p(T^{1/2} n^{-\kappa_2})$. It follows that

$$E^* \sup_{\theta \in \breve{\Theta}_3} |(D.31)| = \sup_{\theta \in \breve{\Theta}_3} \sum_{n=0}^{T-1} |\breve{\varphi}_n(\theta)| n^{-\kappa_2} O_p(T^{-1/2}) = O_p(T^{-1/2}) \sum_{n=0}^{T-1} n^{-1/2-2\kappa_2}$$

$$= O_p((\log T) T^{\max(-1/2, -2\kappa_2)}),$$

such that $\sup_{\theta \in \breve{\Theta}_3} |(D.31)| = o_p^*(1)$, in probability.

It remains to analyze (D.28), for which we find

$$(D.28) = T^{-1} \sum_{t=1}^{T} \sum_{n=0}^{t-1} (\breve{\varphi}_n(\theta)^2 \hat{\varepsilon}_{c,t-n}^2 - \varphi_n^\dagger(\theta)^2 \varepsilon_{t-n}^2) \tag{D.33}$$

$$- T^{-1} \sum_{t=1}^{T} \sum_{n=t}^{\infty} \varphi_n^\dagger(\theta)^2 \varepsilon_{t-n}^2 \tag{D.34}$$

$$- 2T^{-1} \sum_{t=1}^{T} \sum_{n=0}^{\infty} \sum_{m=n+1}^{\infty} \varphi_n^\dagger(\theta) \varphi_m^\dagger(\theta) \varepsilon_{t-n} \varepsilon_{t-m} \tag{D.35}$$

with $\varphi_n^\dagger(\theta) := \sum_{m=0}^{\min(n,t-1)} \pi_m(d^\dagger - d)c_{n-m}^\dagger(\psi)$ satisfying $\sup_{\theta\in\Theta_3^\dagger} |\varphi_n^\dagger(\theta)| \leq cn^{-1/2-\kappa_2/2}$, see (D.29). By identical arguments to those in the proof for Θ_3 of Theorem 1 given in Section S.5.1.5 in the Supplement, the terms (D.34) and (D.35) are $o_p(1)$, uniformly in $\theta \in \Theta_3^\dagger$ and $P(\Theta_3^\dagger \supseteq \breve{\Theta}_3) \to 1$ by (D.11). We therefore proceed with (D.33), which is

$$(D.33) = T^{-1}\sum_{t=1}^{T}\sum_{n=0}^{t-1}\varphi_n^\dagger(\theta)^2(\hat{\varepsilon}_{c,t-n}^2 - \varepsilon_{t-n}^2) \tag{D.36}$$

$$+ T^{-1}\sum_{t=1}^{T}\sum_{n=0}^{t-1}(\breve{\varphi}_n(\theta)^2 - \varphi_n^\dagger(\theta)^2)\hat{\varepsilon}_{c,t-n}^2. \tag{D.37}$$

For (D.36) we apply the Cauchy-Schwarz inequality and find

$$|(D.36)| \leq \left(T^{-1}\sum_{t=1}^{T}\left(\sum_{n=0}^{t-1}\varphi_n^\dagger(\theta)^2\right)^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_{c,t-n}^2 - \varepsilon_{t-n}^2)^2\right)^{1/2},$$

where the term in the second parenthesis is $o_p(1)$ by Lemma D.1. From the bound on $\varphi_n^\dagger(\theta)$, the term in the first parenthesis is bounded by $cT^{-1}\sum_{t=1}^{T}(\sum_{n=0}^{t-1} n^{-1-\kappa_2})^2 \leq c$ uniformly in $\theta \in \Theta_3^\dagger$. Because $P(\Theta_3^\dagger \supseteq \breve{\Theta}_3) \to 1$, see (D.11), this bound applies also uniformly in $\theta \in \breve{\Theta}_3$.

Finally, for the term (D.37) we reverse the order of the summations such that $(D.37) = \sum_{n=0}^{T-1}(\breve{\varphi}_n(\theta)^2 - \varphi_n^\dagger(\theta)^2)T^{-1}\sum_{t=n+1}^{T}\hat{\varepsilon}_{c,t-n}^2$, and apply the mean value theorem,

$$(D.37) = 2(\breve{\theta} - \theta^\dagger)\sum_{n=0}^{T-1}\breve{\varphi}_n(\theta)\frac{\partial\breve{\varphi}_n(\theta)}{\partial\breve{\theta}}T^{-1}\sum_{t=n+1}^{T}\hat{\varepsilon}_{c,t-n}^2.$$

The product of $\breve{\varphi}_n(\theta)$ and its derivative (with respect to $\breve{\theta}$) is absolutely summable, uniformly in $\theta \in \Theta_3^\dagger$, and $T^{-1}\sum_{t=n+1}^{T}\hat{\varepsilon}_{c,t-n}^2 = O_p(1)$ uniformly in $n = 0, \ldots, T-1$, and therefore $\sup_{\theta\in\breve{\Theta}_3}|(D.37)| = o_p(1)$ because $|\breve{\theta} - \theta^\dagger| \xrightarrow{p} 0$ by definition of $\theta^\dagger$.

D.2 Proof of Asymptotic Normality: Eqn. (29)

We expand the score function around the bootstrap true value, $\breve{\theta}$, as

$$0 = T^{1/2}\frac{\partial Q_T^*(\hat{\theta}^*)}{\partial\theta} = T^{1/2}\frac{\partial Q_T^*(\breve{\theta})}{\partial\theta} + T^{1/2}\frac{\partial^2 Q_T^*(\bar{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}^* - \breve{\theta}),$$

where $\bar{\theta}$ is an intermediate value satisfying $|\bar{\theta}_i - \breve{\theta}_i| \leq |\hat{\theta}_i^* - \breve{\theta}_i|$ for $i = 1, \ldots, p+1$. For use in both this proof and subsequent proofs, we define the coefficients

$$\xi_n(\theta_1, \theta_2) := \frac{\partial}{\partial\theta_2}\sum_{m=0}^{n}\pi_m(d_1 - d_2)\sum_{k=0}^{n-m}a_k(\psi_1)b_{n-m-k}(\psi_2). \tag{D.38}$$

D.2.1 Convergence of the Score Function First note that $\varepsilon_t^*(\breve{\theta}) = \sum_{n=0}^{t-1} b_n(\breve{\psi})u_{t-n}^* = \varepsilon_t^*$, by step (iii) of either algorithms in Section 3.3 because $u_{t-n}^* = 0$ for $n \geq t$. Using this and the coefficients $\breve{\xi}_n := \xi_n(\breve{\theta}, \breve{\theta}) = [-n^{-1}, \gamma_n(\breve{\psi})]'$, see (D.38), which satisfy

$$\sum_{n=0}^{s}||\breve{\xi}_n|| = O_p(\log s) \text{ and } \sum_{n=0}^{s}(\breve{\xi}_n)_i^q = O_p(1) \text{ for any } q > 1, s \geq 2, i = 1, \ldots, p+1, \tag{D.39}$$

by Assumption 3(iii) and (5), the normalized score function evaluated at the bootstrap true value is $T^{1/2}\frac{\partial Q_T^*(\check{\theta})}{\partial\theta} = 2T^{-1/2}\sum_{t=1}^T \varepsilon_t^*(\check{\theta})\frac{\partial\varepsilon_t^*(\check{\theta})}{\partial\theta} = 2T^{-1/2}\sum_{t=1}^T \varepsilon_t^* \sum_{n=1}^{t-1}\check{\xi}_n\varepsilon_{t-n}^*$. Conditional on the original data, $U_{Tt}^* := 2T^{-1/2}\varepsilon_t^*\sum_{n=1}^{t-1}\check{\xi}_n\varepsilon_{t-n}^*$ is a MDS with respect to $\mathcal{F}_t^*$, so we apply Lemma A.1.

We first verify the Lyapunov sufficient condition for Lemma A.1(i) elementwise. By independence of ε_t^*, conditional on the original data,

$$\sum_{t=1}^T E^*(U_{Tt,i}^{*4}) = 2^4T^{-2}\sum_{t=1}^T E^*(\varepsilon_t^{*4})E^*((\sum_{n=1}^{t-1}(\check{\xi}_n)_i\varepsilon_{t-n}^*)^4)$$

$$= 16T^{-2}\sum_{t=1}^T \hat{\varepsilon}_{c,t}^4\sum_{n=1}^{t-1}(\check{\xi}_n)_i^4\hat{\varepsilon}_{c,t-n}^4 \tag{D.40}$$

$$+ 48T^{-2}\sum_{t=1}^T \hat{\varepsilon}_{c,t}^4\sum_{n=1}^{t-1}\sum_{m=n+1}^{t-1}(\check{\xi}_n)_i^2(\check{\xi}_m)_i^2\hat{\varepsilon}_{c,t-n}^2\hat{\varepsilon}_{c,t-m}^2. \tag{D.41}$$

Reverse the order of the summations in (D.40) to obtain

$$(D.40) = 16T^{-1}\sum_{n=1}^{T-1}(\check{\xi}_n)_i^4 T^{-1}\sum_{t=n+1}^T \hat{\varepsilon}_{c,t}^4\hat{\varepsilon}_{c,t-n}^4,$$

where $\sum_{n=1}^{T-1}(\check{\xi}_n)_i^4 = O_p(1)$ in view of (D.39) and, by the Cauchy-Schwarz inequality, $T^{-1}\sum_{t=n+1}^T \hat{\varepsilon}_{c,t}^4\hat{\varepsilon}_{c,t-n}^4 \le (T^{-1}\sum_{t=n+1}^T \hat{\varepsilon}_{c,t}^8)^{1/2}(T^{-1}\sum_{t=n+1}^T \hat{\varepsilon}_{c,t-n}^8)^{1/2} \le T^{-1}\sum_{t=1}^T \hat{\varepsilon}_{c,t}^8 = O_p(1)$ uniformly in $n = 1,\ldots,T-1$. Thus, $(D.40) = O_p(T^{-1})$. In the same way,

$$(D.41) = 48T^{-1}\sum_{n=1}^{T-1}\sum_{m=n+1}^{T-1}(\check{\xi}_n)_i^2(\check{\xi}_m)_i^2 T^{-1}\sum_{t=m+1}^T \hat{\varepsilon}_{c,t}^4\hat{\varepsilon}_{c,t-n}^2\hat{\varepsilon}_{c,t-m}^2,$$

where $\sum_{n=1}^{T-1}\sum_{m=n+1}^{T-1}(\check{\xi}_n)_i^2(\check{\xi}_m)_i^2 = O_p(1)$ in view of (D.39) and $T^{-1}\sum_{t=m+1}^T \hat{\varepsilon}_{c,t}^4\hat{\varepsilon}_{c,t-n}^2\hat{\varepsilon}_{c,t-m}^2 = O_p(1)$ uniformly in $1 \le n, m \le T-1$. Thus, $(D.41) = O_p(T^{-1})$.

Next we verify condition (ii)(b) of Lemma A.1 elementwise. We find

$$\sum_{t=1}^T E^*(U_{Tt,i}^*U_{Tt,j}^*|\mathcal{F}_{t-1}^*) = \sum_{t=1}^T(E^*(U_{Tt,i}^*U_{Tt,j}^*|\mathcal{F}_{t-1}^*) - E^*(U_{Tt,i}^*U_{Tt,j}^*) + E^*(U_{Tt,i}^*U_{Tt,j}^*))$$

$$= \sum_{t=1}^T(E^*(U_{Tt,i}^*U_{Tt,j}^*|\mathcal{F}_{t-1}^*) - E^*(U_{Tt,i}^*U_{Tt,j}^*)) \tag{D.42}$$

$$+ 4T^{-1}\sum_{t=1}^T \hat{\varepsilon}_{c,t}^2\sum_{n=1}^{t-1}(\check{\xi}_n)_i^2(\check{\xi}_n)_j^2\hat{\varepsilon}_{c,t-n}^2, \tag{D.43}$$

and want to show that (D.42) is negligible by L_2-convergence and that (D.43) converges in probability to the relevant limit. For the former, write $E^*(U_{Tt,i}^*U_{Tt,j}^*|\mathcal{F}_{t-1}^*) - E^*(U_{Tt,i}^*U_{Tt,j}^*) = 4T^{-1}\hat{\varepsilon}_{c,t}^2 y_{t-1}^*$ with $y_{t-1}^* := \sum_{n,m=1}^{t-1}(\check{\xi}_n)_i(\check{\xi}_m)_j\varepsilon_{t-n}^*\varepsilon_{t-m}^* - \sum_{n=1}^{t-1}(\check{\xi}_n)_i(\check{\xi}_n)_j\hat{\varepsilon}_{c,t-n}^2$ so that

$$E^*\left(\left(\sum_{t=1}^T(E^*(U_{Tt,i}^*U_{Tt,j}^*|\mathcal{F}_{t-1}^*) - E^*(U_{Tt,i}^*U_{Tt,j}^*))\right)^2\right) = 16T^{-2}\sum_{t,s=1}^T \hat{\varepsilon}_{c,t}^2\hat{\varepsilon}_{c,s}^2E^*(y_{t-1}^*y_{s-1}^*). \tag{D.44}$$

Here, decompose $y^*_{t-1} = \sum_{n=1}^{t-1} (\check{\xi}_n)_i (\check{\xi}_n)_j (\varepsilon^{*2}_{t-n} - \hat{\varepsilon}^2_{c,t-n}) + \sum_{n \neq m} (\check{\xi}_n)_i (\check{\xi}_m)_j \varepsilon^*_{t-n} \varepsilon^*_{t-m}$ and

$$y^*_{t-1} y^*_{s-1} = \sum_{n=1}^{t-1} (\check{\xi}_n)_i (\check{\xi}_n)_j (\varepsilon^{*2}_{t-n} - \hat{\varepsilon}^2_{c,t-n}) \sum_{m=1}^{t-1} (\check{\xi}_m)_i (\check{\xi}_m)_j (\varepsilon^{*2}_{t-m} - \hat{\varepsilon}^2_{c,t-m}) \tag{D.45}$$

$$+ \sum_{n \neq m} (\check{\xi}_n)_i (\check{\xi}_m)_j \varepsilon^*_{t-n} \varepsilon^*_{t-m} \sum_{k \neq l} (\check{\xi}_k)_i (\check{\xi}_l)_j \varepsilon^*_{t-k} \varepsilon^*_{t-l} \tag{D.46}$$

$$+ 2 \sum_{n=1}^{t-1} (\check{\xi}_n)_i (\check{\xi}_n)_j (\varepsilon^{*2}_{t-n} - \hat{\varepsilon}^2_{c,t-n}) \sum_{k \neq l} (\check{\xi}_k)_i (\check{\xi}_l)_j \varepsilon^*_{t-k} \varepsilon^*_{t-l}, \tag{D.47}$$

where we note immediately that $E^*(D.47) = 0$ by independence of ε^*_t, conditional on the original data, so that (D.47) does not contribute to (D.44). Next, $E^*(D.45) \neq 0$ only if $s = t-n+m$ and $E^*(D.46) \neq 0$ only if either $s = t-n+k$ or $s = t-n+l$. In either case, we can eliminate the summation over s in (D.44), which leaves the summation over t with T terms. The remaining summations in $E^*(D.45)$ and $E^*(D.46)$ at most contribute a term of order $O_p(\log T)$, in view of (D.39). This implies that $(D.44) = O_p((\log T)^4 T^{-1})$ and hence $(D.42) = O_p((\log T)^2 T^{-1/2})$.

It remains to find the limit in probability of (D.43), which we decompose as

$$(D.43) = 4 \sum_{n=1}^{T-1} (\check{\xi}_n)^2_i (\check{\xi}_n)^2_j T^{-1} \sum_{t=n+1}^{T} (\hat{\varepsilon}^2_{c,t} - \varepsilon^2_t) \hat{\varepsilon}^2_{c,t-n} \tag{D.48}$$

$$+ 4 \sum_{n=1}^{T-1} (\check{\xi}_n)^2_i (\check{\xi}_n)^2_j T^{-1} \sum_{t=n+1}^{T} \varepsilon^2_t (\hat{\varepsilon}^2_{c,t-n} - \varepsilon^2_{t-n}) \tag{D.49}$$

$$+ 4 \sum_{n=1}^{T-1} ((\check{\xi}_n)^2_i (\check{\xi}_n)^2_j - (\xi^\dagger_n)^2_i (\xi^\dagger_n)^2_j) T^{-1} \sum_{t=n+1}^{T} \varepsilon^2_t \varepsilon^2_{t-n} \tag{D.50}$$

$$+ 4 \sum_{n=1}^{T-1} (\xi^\dagger_n)^2_i (\xi^\dagger_n)^2_j T^{-1} \sum_{t=n+1}^{T} \varepsilon^2_t \varepsilon^2_{t-n}, \tag{D.51}$$

where $\xi^\dagger_n := \xi_n(\theta^\dagger, \theta^\dagger) = [-n^{-1}, \gamma_n(\psi^\dagger)]'$ similarly to $\check{\xi}_n$, see also (D.38). First we show that each of (D.48)–(D.50) are asymptotically negligible. The proofs for the terms (D.48) and (D.49) are identical, so we give only the former. By the Cauchy-Schwarz inequality,

$$|(D.48)| \leq 4 \sum_{n=1}^{T-1} (\check{\xi}_n)^2_i (\check{\xi}_n)^2_j \left(T^{-1} \sum_{t=n+1}^{T} (\hat{\varepsilon}^2_{c,t} - \varepsilon^2_t)^2 \right)^{1/2} \left(T^{-1} \sum_{t=n+1}^{T} \hat{\varepsilon}^4_{c,t-n} \right)^{1/2},$$

where the term in the middle parenthesis is $o_p(1)$ by Lemma D.1 and the last parenthesis is $O_p(1)$, both uniformly in $n = 0, \ldots, T-1$, while the summation over n is bounded using (D.39). Thus, $(D.48) = o_p(1)$. This leaves only (D.50), for which we first note that if $i = j = 1$ then $(D.50) = 0$ because $\check{\xi}_{n,1} = \xi^\dagger_{n,1} = -n^{-1}$. If $i \neq 1$ or $j \neq 1$, we reverse the order of the summations and apply the mean value theorem to find

$$(D.50) = 4(\check{\psi} - \psi^\dagger)' \sum_{n=1}^{T-1} \frac{\partial((\xi_n(\theta,\theta))^2_i (\xi_n(\theta,\theta))^2_j)}{\partial \psi} |_{\psi = \bar{\psi}} T^{-1} \sum_{t=n+1}^{T} \hat{\varepsilon}^2_{c,t} \hat{\varepsilon}^2_{c,t-n},$$

where $\bar{\psi}$ is an intermediate value between $\check{\psi}$ and $\psi^\dagger$. Now, $T^{-1} \sum_{t=n+1}^{T} \hat{\varepsilon}^2_{c,t} \hat{\varepsilon}^2_{c,t-n} = O_p(1)$ uniformly in $n = 1, \ldots, T-1$, $|\check{\psi} - \psi^\dagger| \overset{p}{\to} 0$ by definition of $\psi^\dagger$, and the summation over

n is bounded by the same argument that led to (D.39) (noting that when $d_1 = d_2$, the first element of $\xi_n(\theta_1, \theta_2)$ is $-n^{-1}$). It follows that $(D.50) = o_p(1)$.

Finally, reversing the order of the summations, (D.51) is decomposed as

$$(D.51) = 4T^{-1} \sum_{t=1}^{T} \sigma_t^2 \sum_{n=1}^{t-1} \sigma_{t-n}^2 (\xi_n^\dagger)_i^2 (\xi_n^\dagger)_j^2 (z_t^2 z_{t-n}^2 - \tau_{n,n}) \quad \text{(D.52)}$$

$$+ 4T^{-1} \sum_{t=1}^{T} \sigma_t^2 \sum_{n=1}^{t-1} \sigma_{t-n}^2 (\xi_n^\dagger)_i^2 (\xi_n^\dagger)_j^2 \tau_{n,n}. \quad \text{(D.53)}$$

The term (D.52) is mean zero with second moment bounded by

$$KT^{-2} \sum_{t,s=1}^{T} \sum_{n=1}^{t-1} \sum_{m=1}^{t-1} (\xi_n^\dagger)_i^2 (\xi_n^\dagger)_j^2 (\xi_m^\dagger)_i^2 (\xi_m^\dagger)_j^2 E(z_t^2 z_{t-n}^2 - \tau_{n,n})(z_s^2 z_{s-m}^2 - \tau_{m,m}) \leq KT^{-1} \to 0$$

using Assumption 5 and (D.39), so that $(D.52) = o_p(1)$. By Lemma A.5 with $g_{t,n,m} = \mathbb{I}(n = m)\tau_{n,m}$, (D.53) is, apart from a $o(1)$ term,

$$4T^{-1} \sum_{t=1}^{T} \sigma_t^4 \sum_{n=1}^{t-1} (\xi_n^\dagger)_i^2 (\xi_n^\dagger)_j^2 \tau_{n,n} = 4T^{-1} \sum_{t=1}^{T} \sigma_t^4 \sum_{n=1}^{\infty} (\xi_n^\dagger)_i^2 (\xi_n^\dagger)_j^2 \tau_{n,n} - 4T^{-1} \sum_{t=1}^{T} \sigma_t^4 \sum_{n=t}^{\infty} (\xi_n^\dagger)_i^2 (\xi_n^\dagger)_j^2 \tau_{n,n},$$

where the first term is $4A^\dagger T^{-1} \sum_{t=1}^{T} \sigma_t^4 \to 4A^\dagger \int_0^1 \sigma^4(s) \mathsf{d}s$ and the second term is bounded by $KT^{-1} \sum_{t=1}^{T} \sum_{n=t}^{\infty} ||\xi_n^\dagger||^4$, which converges to zero because it is the Cesàro mean of the sequence $\sum_{n=t}^{\infty} ||\xi_n^\dagger||^4$, which itself converges to zero as $t \to \infty$ since it is the tail of a convergent sum, see (D.39). This concludes the proof.

D.2.2 Convergence of the Hessian In anticipation of the proof of (28), our proof of convergence of the Hessian for the bootstrap estimator is somewhat different from our proof of convergence of the Hessian for the QML estimator in Section S.5.2.2 in the Supplement. Recalling that $\varepsilon_t^*(\theta) = \sum_{n=0}^{t-1} \check{\varphi}_n(\theta) \varepsilon_{t-n}^*$, see (D.29), the Hessian matrix is

$$\frac{\partial^2 Q_T^*(\theta)}{\partial\theta\partial\theta'} = 2T^{-1} \sum_{t=1}^{T} \varepsilon_t^*(\theta) \frac{\partial^2 \varepsilon_t^*(\theta)}{\partial\theta\partial\theta'} + 2T^{-1} \sum_{t=1}^{T} \frac{\partial \varepsilon_t^*(\theta)}{\partial\theta} \frac{\partial \varepsilon_t^*(\theta)}{\partial\theta'}$$

$$= 2T^{-1} \sum_{t=1}^{T} \sum_{m=0}^{t-1} \sum_{n=1}^{t-1} \frac{\partial \check{\xi}_n(\theta)}{\partial\theta'} \check{\varphi}_m(\theta) \varepsilon_{t-n}^* \varepsilon_{t-m}^* \quad \text{(D.54)}$$

$$+ 2T^{-1} \sum_{t=1}^{T} \sum_{n,m=1}^{t-1} \check{\xi}_n(\theta) \check{\xi}_m(\theta)' \varepsilon_{t-n}^* \varepsilon_{t-m}^*, \quad \text{(D.55)}$$

where we defined the coefficients $\check{\xi}_n(\theta) := \xi_n(\check{\theta}, \theta) = \frac{\partial \check{\varphi}_n(\theta)}{\partial\theta}$, which satisfy, for $\nu = 0, 1$,

$$\sum_{n=0}^{s} \sup_{||\theta - \check{\theta}|| \leq \epsilon} ||\frac{\partial^{(\nu)} \check{\xi}_n(\theta)_i}{\partial \theta_j^{(\nu)}}|| = O_p(s^\epsilon) \text{ and } \sum_{n=0}^{s} \sup_{||\theta - \check{\theta}|| \leq \epsilon} ||\frac{\partial \check{\xi}_n(\theta)_i \check{\xi}_n(\theta)_j}{\partial\theta}|| = O_p(1), \quad \text{(D.56)}$$

as easily proven by reversing summations in (D.38) and using Assumption 3(iii) and (5).

We show $\sup_{||\theta-\check{\theta}||\leq\epsilon}|(D.54)_{i,j}|\overset{p^*}{\to}_p 0$ and $\sup_{||\theta-\check{\theta}||\leq\epsilon}|(D.55)_{i,j}-2(B^\dagger)_{i,j}\int_0^1\sigma^2(s)\mathsf{d}s|\overset{p^*}{\to}_p 0$. First, since $\check{\varphi}_m(\check{\theta})=\mathbb{I}(m=0)$, by the mean value theorem,

$$(D.54)_{i,j}=2\sum_{n=1}^{T-1}\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\sum_{m=0}^{T-1}\check{\varphi}_m(\theta)T^{-1}\sum_{t=\max(n,m)+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_{t-m}$$
$$=2\sum_{n=1}^{T-1}\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}T^{-1}\sum_{t=n+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_t+(\theta-\check{\theta})'2\sum_{n=1}^{T-1}\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\sum_{m=0}^{T-1}\check{\xi}_m(\bar{\theta}_n)T^{-1}\sum_{t=\max(n,m)+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_{t-m}$$

for intermediate values, $\bar{\theta}_n$, and hence

$$E^*\sup_{||\theta-\check{\theta}||\leq\epsilon}|(D.54)_{i,j}|\leq 2T^{-1}\sum_{n=1}^{T-1}\sup_{||\theta-\check{\theta}||\leq\epsilon}\left|\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\right|E^*\left|\sum_{t=n+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_t\right|$$
$$+\epsilon T^{-1}\sum_{n=1}^{T-1}\sup_{||\theta-\check{\theta}||\leq\epsilon}\left\|\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\check{\xi}_n(\bar{\theta}_n)\right\|E^*\left|\sum_{t=n+1}^{T}\varepsilon^{*2}_{t-n}\right|$$
$$+\epsilon T^{-1}\sum_{n=1}^{T-1}\sum_{m=0,m\neq n}^{T-1}\sup_{||\theta-\check{\theta}||\leq\epsilon}\left\|\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\check{\xi}_m(\bar{\theta}_n)\right\|E^*\left|\sum_{t=\max(n,m)+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_{t-m}\right|.$$

Here, $E^*|\sum_{t=n+1}^T\varepsilon^{*2}_{t-n}|=O_p(T)$ uniformly in $n=1,\ldots,T-1$ and, for $n\neq m$,

$$\left(E^*\left|\sum_{t=\max(n,m)+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_{t-m}\right|\right)^2\leq\sum_{t,s=\max(n,m)+1}^{T}E^*(\varepsilon^*_{t-n}\varepsilon^*_{t-m}\varepsilon^*_{s-n}\varepsilon^*_{s-m})=2\sum_{t=\max(n,m)+1}^{T}E^*(\varepsilon^{*2}_{t-n}\varepsilon^{*2}_{t-m})$$
$$=2\sum_{t=\max(n,m)+1}^{T}\hat{\varepsilon}^2_{c,t-n}\hat{\varepsilon}^2_{c,t-m}\leq c\sum_{t=1}^{T}\hat{\varepsilon}^4_{c,t}=O_p(T)\qquad\text{(D.57)}$$

uniformly in $0\leq n,m\leq T-1$, where the first inequality is Jensen's and the second is Cauchy-Schwarz together with $\sum_{t=n+1}^T\hat{\varepsilon}^4_{c,t-n}\leq\sum_{t=1}^T\hat{\varepsilon}^4_{c,t}$. Thus,

$$E^*\sup_{||\theta-\check{\theta}||\leq\epsilon}|(D.54)_{i,j}|\leq O_p(T^{-1/2})\sum_{n=1}^{T-1}\sup_{||\theta-\check{\theta}||\leq\epsilon}\left|\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\right|+\epsilon O_p(1)\sum_{n=1}^{T-1}\sup_{||\theta-\check{\theta}||\leq\epsilon}\left\|\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\check{\xi}_n(\bar{\theta}_n)\right\|$$
$$+\epsilon O_p(T^{-1/2})\sum_{n=1}^{T-1}\sum_{m=0,m\neq n}^{T-1}\sup_{||\theta-\check{\theta}||\leq\epsilon}\left\|\frac{\partial\check{\xi}_n(\theta)_i}{\partial\theta_j}\check{\xi}_m(\bar{\theta}_n)\right\|$$
$$=O_p(T^{-1/2})O_p(T^{\epsilon})+\epsilon O_p(1)O_p(1)=O_p(T^{\epsilon-1/2})+o_p(1)$$

using (D.56). Hence, $\sup_{||\theta-\check{\theta}||\leq\epsilon}|(D.54)_{i,j}|=o^*_p(1)$, in probability, because ϵ can be arbitrarily small.

For (D.55) we reverse the order of the summations and expand as, noting that

$E^*(\varepsilon^*_{t-n}\varepsilon^*_{t-m}) = 0$ for $n \neq m$,

$$(D.55) = 2\sum_{n\neq m=1}^{T-1} \check{\xi}_n(\theta)\check{\xi}_m(\theta)'T^{-1}\sum_{t=\max(n,m)+1}^{T} \varepsilon^*_{t-n}\varepsilon^*_{t-m} \tag{D.58}$$

$$+ 2\sum_{n=1}^{T-1}(\check{\xi}_n(\theta)\check{\xi}_n(\theta)' - \check{\xi}_n\check{\xi}_n')T^{-1}\sum_{t=n+1}^{T} \varepsilon^{*2}_{t-n} \tag{D.59}$$

$$+ 2\sum_{n=1}^{T-1}\check{\xi}_n\check{\xi}_n'T^{-1}\sum_{t=n+1}^{T}(\varepsilon^{*2}_{t-n} - E^*(\varepsilon^{*2}_{t-n})) \tag{D.60}$$

$$+ 2\sum_{n=1}^{T-1}\check{\xi}_n\check{\xi}_n'T^{-1}\sum_{t=n+1}^{T}(E^*(\varepsilon^{*2}_{t-n}) - \varepsilon^2_{t-n}) \tag{D.61}$$

$$+ 2\sum_{n=1}^{T-1}(\check{\xi}_n\check{\xi}_n' - \xi^\dagger_n\xi^{\dagger\prime}_n)T^{-1}\sum_{t=n+1}^{T} \varepsilon^2_{t-n} \tag{D.62}$$

$$+ 2\sum_{n=1}^{T-1}\xi^\dagger_n\xi^{\dagger\prime}_n T^{-1}\sum_{t=n+1}^{T} \varepsilon^2_{t-n} - 2B^\dagger\int_0^1 \sigma^2(s)\mathsf{d}s. \tag{D.63}$$

The terms (D.61) and (D.62) are negligible by the same proofs as for (D.48) and (D.50). The term (D.60) is negligible as in the proof of (D.30) noting that the $\check{\xi}_n$ coefficients are square summable by (D.39). Next, we apply the mean value theorem to (D.59) and find

$$(D.59)_{i,j} = 2(\theta - \check{\theta})'\sum_{n=1}^{T-1} \left.\frac{\partial\check{\xi}_n(\theta)_i\check{\xi}_n(\theta)_j}{\partial\theta}\right|_{\theta=\bar{\theta}} T^{-1}\sum_{t=n+1}^{T} \varepsilon^{*2}_{t-n},$$

where $\bar{\theta}$ denotes an intermediate value. Then, by (D.56) and because $E^*|T^{-1}\sum_{t=n+1}^{T}\varepsilon^{*2}_{t-n}| = O_p(1)$ uniformly in $n = 0, \dots, T-1$, $\sup_{||\theta-\check{\theta}||\leq\epsilon}|(D.59)| \leq \epsilon O^*_p(1)$, in probability, which is $o^*_p(1)$, in probability since ϵ can be arbitrarily small. Thus, by (D.56) and (D.57), $\sup_{||\theta-\check{\theta}||\leq\epsilon}|(D.58)|$ is also $o^*_p(1)$, in probability, because its conditional expectation is bounded by $O_p(T^{-1/2})(\sum_{n=1}^{T-1}||\check{\xi}_n(\theta)||)^2 = O_p(T^{\epsilon-1/2})$.

Finally, the first term on the right-hand side of (D.63) has mean

$$2T^{-1}\sum_{t=1}^{T}\sum_{n=1}^{t-1}\sigma^2_{t-n}\xi^\dagger_n\xi^{\dagger\prime}_n = 2T^{-1}\sum_{t=1}^{T}\sigma^2_t\sum_{n=1}^{\infty}\xi^\dagger_n\xi^{\dagger\prime}_n - 2T^{-1}\sum_{t=1}^{T}\sigma^2_t\sum_{n=t}^{\infty}\xi^\dagger_n\xi^{\dagger\prime}_n + o(1) \tag{D.64}$$

by Lemma A.5 with $g_{t,n,m} = \mathbb{I}(n = m)\sigma^2_t$. Here, the first term is $2B^\dagger T^{-1}\sum_{t=1}^{T}\sigma^2_t \to 2B^\dagger\int_0^1\sigma^2(s)\mathsf{d}s$. For the second term we find the bound $KT^{-1}\sum_{t=1}^{T}\sum_{n=t}^{\infty}||\xi^\dagger_n||^2 \to 0$, which converges to zero because it is the Cesàro mean of the sequence $\sum_{n=t}^{\infty}||\xi^\dagger_n||^2$, which itself converges to zero as $t \to \infty$ since it is the tail of a convergent sum, see (D.39). The variance of (D.63) is shown in Section S.6.8 to converge to zero so that (D.63) converges to zero in L_2-norm, and hence in probability, which concludes the proof.

D.3 Proof of Rate: Eqn. (28)

We apply Theorem 1 of Andrews (1999). His Assumption 1 holds by (27) of our Theorem 6, his Assumptions 2^{2*}(a) and 2^{2*}(b) by our Assumption 2, and finally his Assumptions 2^{2*}(c) and 3 hold by the convergence in distribution of the score and the uniform convergence of the Hessian in the proof of (29).

Acknowledgements

This paper supercedes and includes all results from our earlier paper "Quasi-Maximum Likelihood Estimation of Heteroskedastic Fractional Time Series Models". 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 the Co-Editor, Oliver Linton, an Associate Editor, three referees, David Harris, Søren Johansen, and James MacKinnon for helpful comments on earlier versions of this paper. 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, DNRF78) 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.

References

Agiakloglou, C. and P. Newbold (1994), Lagrange multiplier tests for fractional difference, *Journal of Time Series Analysis* 15, 253–262.

Amado, C. and T. Teräsvirta (2014), Modelling changes in the unconditional variance of long stock return series, *Journal of Empirical Finance* 25, 15–35.

Andrews, D.W.K. (1999), Estimation when a parameter is on a boundary, *Econometrica* 67, 1341–1383.

Baillie, R.T., C.-F. Chung and M.A. Tieslau (1996), Analyzing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics* 11, 23–40.

Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

Bloomfield, P. (1973), An exponential model for the spectrum of a scalar time series, *Biometrika* 60, 217–226.

Breitung, J. and U. Hassler (2002), Inference on the cointegration rank in fractionally integrated processes, *Journal of Econometrics* 110, 167–185.

Cavaliere, G. (2005), Unit root tests under time-varying variances, *Econometric Reviews* 23, 259–292.

Cavaliere, G., M.Ø. Nielsen and A.M.R. Taylor (2015), Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets, *Journal of Econometrics* 187, 557–579.

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.

Clark, T. (2009), Is the Great Moderation over? An empirical analysis, *Federal Reserve Bank of Kansas City Economic Review* 94(4), 5–42.

Demetrescu, M., V. Kuzin and U. Hassler (2008), Long memory testing in the time domain, *Econometric Theory* 24, 176–215.

Deo, R.S. (2000), Spectral tests of the martingale hypothesis under conditional heteroskedasticity, *Journal of Econometrics* 99, 291–315.

Ding, Z., C.W.J. Granger and R.F. Engle (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance* 1, 83–106.

Dolado, J.J., J. Gonzalo and L. Mayoral (2002), A fractional Dickey-Fuller test for unit roots, *Econometrica* 70, 1963–2006.

Doornik, J.A. (2007), *Object-Oriented Matrix Programming Using Ox*, 3rd ed., Timberlake Consultants Press, London.

Gonçalves, S. and L. Kilian (2004), Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* 123, 89–120.

Hannan, E.J. (1973), The asymptotic theory of linear time-series models, *Journal of Applied Probability* 10, 130–145.

Hansen, B.E. (2000), Testing for structural change in conditional models, *Journal of Econometrics* 97, 93–115.

Hassler, U., P.M.M. Rodrigues and A. Rubia (2009), Testing for general fractional integration in the time domain, *Econometric Theory* 25, 1793–1828.

Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton.

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.

Lehmann, E.L. and J.P. Romano (2005), *Testing Statistical Hypotheses*, third edition, Springer, New York.

Ling, S. (2003), Adaptive estimators and tests of stationary and nonstationary short- and long-memory ARFIMA-GARCH models, *Journal of the American Statistical Association* 98, 955–967.

Ling, S. and W.K. Li (1997), On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity, *Journal of the American Statistical Association* 92, 1184–1194.

Liu, R.Y. (1988), Bootstrap procedures under some non-I.I.D. models, *Annals of Statistics* 16, 1696–1708.

Lobato, I.N. and C. Velasco (2006), Optimal fractional Dickey-Fuller tests, *Econometrics Journal* 9, 492–510.

Lobato, I.N. and C. Velasco (2007), Efficient Wald tests for fractional unit roots, *Econometrica* 75, 575–589.

MacKinnon, J.G. (2009), Bootstrap hypothesis testing, in D.A. Belsley and J. Kontoghiorghes (eds.), *Handbook of Computational Econometrics*, Wiley, Chichester, 183–213.

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.

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

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

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

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.

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 University Press, Oxford, 1–44.

Stock, J.H. and M.W. Watson (2012), Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* Spring Volume, 81–156.

Swensen, A.R. (2003), Bootstrapping unit root tests for integrated processes, *Journal of Time Series Analysis* 24, 99–126.

Tanaka, K. (1999), The nonstationary fractional unit root, *Econometric Theory* 15, 549–582.

van der Vaart, A.W. (1998), *Asymptotic Statistics*, Cambridge University Press, Cambridge.

White, H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica* 48, 817–838.

White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica* 50, 1–25.

Zygmund, A. (2003), *Trigonometric Series*, third edition, Cambridge University Press, Cambridge.

Supplementary Online Appendix

to

Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form

by

G. Cavaliere, M.Ø. Nielsen and A.M.R. Taylor

Date: November 25, 2016

S.1 Introduction

This supplement contains additional Monte Carlo results and proofs for our paper "Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form". Equation references (S.n) for $n \geq 1$ refer to equations in this supplement and other equation references are to the main paper.

The supplement is organised as follows. Section S.2 presents additional Monte Carlo results and in Sections S.3 and S.4, respectively, we give proofs of the preliminary lemmas in Appendix A and of the variation bounds lemmas in Appendix B. Additional proofs for the QML estimator and the asymptotic tests are provided in Section S.5 and additional proofs for bootstrap inference are in Section S.6. All additional references are included at the end of the supplement.

For this entire supplement, all stated results and derivations shall be taken as conditional on $\sigma(\cdot)$. Due to the stochastic independence of $\{\sigma_t\}$ and $\{z_t\}$, see Assumption 1(b), and given the simple structure of conditional distributions on product spaces, this implies that $\{\sigma_t\}$ can be treated as fixed. In order to avoid repetition, this will not be repeated on every occasion. Where convergence obtains to a limit which does not depend on $\sigma(\cdot)$, it should be recalled that the stated convergence result also holds unconditionally.

S.2 Additional Monte Carlo Results

Tables S.1 and S.2 report results relating to tests on the autoregressive parameter a in (30). In particular, results are reported for the asymptotic LM_T, LR_T, W_T and RW_T tests of $H_{0,2} : a = 0$ against $H_{1,2} : a \neq 0$, along with their restricted wild bootstrap (Algorithm 1) and unrestricted wild bootstrap (Algorithm 2) counterparts. Finite sample size and power results are reported for $a = 0$ and $a = 1 + 5/\sqrt{T}$, respectively, in (30). Table S.1 relates to the case of a one-time shift in volatility, while Table S.2 reports results for the conditionally heteroskedastic Models A-I outlined in Section 5.3. Tables S.3 and S.4 report corresponding results for the joint tests of $H_{0,3} : d = 1 \cap a = 0$ against $H_{1,3} : d \neq 1 \cup a \neq 0$, with finite sample size and power results reported for $d = 1, a = 0$ and $\delta = (1, 5/3)'$, in (1) and (30). Finally, Figure S.1 reports finite sample power functions for the bootstrap tests of $H_{0,1} : d = 1$ against $H_{1,1} : d \neq 1$ for a range of values of δ in $d = 1 + \delta/\sqrt{T}$. The Monte Carlo DGP and set-up of these experiments are exactly as detailed in Section 5.1.

S.3 Proofs of Preliminary Lemmas

S.3.1 Proof of Lemma A.1

The result for condition (ii)(a) follows from Theorem 2.3 of McLeish (1974) and the comments in the two paragraphs following it. For condition (ii)(b) the result is Theorem 2.2 of Dvoretzky (1972).

S.3.2 Proof of Lemma A.2

The result for moments follows 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 law of iterated expectations and the martingale difference property of z_t. To show the result for cumulants, we first have $\kappa_2(t, t-r) = E(z_t z_{t-r}) = 0$ because $r \geq 1$. When $q \geq 3$ we use the relation $E(z_t z_{t-r_1} \cdots z_{t-r_{q-1}}) = \sum_\pi \prod_{B\in\pi} \kappa(B)$, where π runs through the list of all partitions of $\{0, r_1, \ldots, r_{q-1}\}$ and B runs through the list of all blocks of the partition π. The required result then holds by induction on q because it has already been shown to hold for moments and for $q = 2$.

Table S.1: Tests of $H_{0,2}$: simulated size and power with one-time shift in unconditional volatility

				size				power			
τ	υ	T	λ	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests											
	1	100	1.00	6.32	7.93	8.93	10.34	44.73	59.89	45.36	53.25
	1	250	1.00	5.47	5.49	5.17	5.68	69.57	74.64	67.29	61.51
	1	∞	1.00	5.00	5.00	5.00	5.00	87.91	87.91	87.91	87.91
1/4	1/3	100	2.33	17.21	18.63	19.48	14.97	17.88	36.43	29.53	42.11
1/4	1/3	250	2.33	19.29	19.10	18.67	7.85	34.95	47.00	39.56	42.32
1/4	1/3	∞	2.33	19.95	19.95	19.95	5.00	53.57	53.57	53.57	53.57
1/4	3	100	1.24	9.18	11.25	12.45	12.00	36.67	52.71	29.50	41.33
1/4	3	250	1.24	8.27	8.27	7.82	5.72	59.46	65.89	58.03	54.23
1/4	3	∞	1.24	7.90	7.90	7.90	5.00	80.12	80.12	80.12	80.12
3/4	1/3	100	1.24	9.20	10.41	10.88	11.38	36.78	54.29	43.87	51.60
3/4	1/3	250	1.24	8.17	8.26	7.68	5.67	60.36	67.61	60.03	56.65
3/4	1/3	∞	1.24	7.90	7.90	7.90	5.00	80.12	80.12	80.12	80.12
3/4	3	100	2.33	19.31	21.29	22.73	16.61	14.07	32.55	20.12	33.64
3/4	3	250	2.33	19.56	19.66	19.15	8.31	31.65	44.31	35.54	38.14
3/4	3	∞	2.33	19.95	19.95	19.95	5.00	53.57	53.57	53.57	53.57
Panel B: wild bootstrap tests (Algorithm 1)											
	1	100	1.00	5.38	5.28	5.24	5.24	45.11	60.40	44.88	51.63
	1	250	1.00	5.06	5.03	5.14	5.26	69.58	74.63	67.68	62.30
	1	∞	1.00	5.00	5.00	5.00	5.00	87.91	87.91	87.91	87.91
1/4	1/3	100	2.33	6.31	6.43	6.08	5.39	20.86	39.55	31.03	42.21
1/4	1/3	250	2.33	5.88	5.97	5.87	5.12	36.70	48.28	40.22	42.24
1/4	1/3	∞	2.33	5.00	5.00	5.00	5.00	53.57	53.57	53.57	53.57
1/4	3	100	1.24	5.01	5.13	5.32	5.24	34.98	51.83	31.33	41.30
1/4	3	250	1.24	5.13	5.12	5.06	5.06	58.54	65.74	57.76	54.19
1/4	3	∞	1.24	5.00	5.00	5.00	5.00	80.12	80.12	80.12	80.12
3/4	1/3	100	1.24	5.85	5.59	5.06	5.03	38.86	55.98	42.84	50.54
3/4	1/3	250	1.24	5.00	5.09	5.17	4.93	59.96	67.61	59.65	55.69
3/4	1/3	∞	1.24	5.00	5.00	5.00	5.00	80.12	80.12	80.12	80.12
3/4	3	100	2.33	6.05	6.36	5.44	5.24	15.20	33.89	20.86	33.87
3/4	3	250	2.33	5.60	5.73	5.57	5.21	32.69	45.70	36.66	38.96
3/4	3	∞	2.33	5.00	5.00	5.00	5.00	53.57	53.57	53.57	53.57
Panel C: wild bootstrap tests (Algorithm 2)											
	1	100	1.00	6.07	8.72	7.05	6.70	47.83	72.84	31.10	39.56
	1	250	1.00	4.57	4.43	3.33	3.49	63.58	70.35	18.33	18.95
	1	∞	1.00	5.00	5.00	5.00	5.00	87.91	87.91	87.91	87.91
1/4	1/3	100	2.33	9.34	11.73	7.69	7.00	43.02	60.12	19.10	37.37
1/4	1/3	250	2.33	5.51	7.23	4.72	3.72	33.37	54.72	14.12	18.53
1/4	1/3	∞	2.33	5.00	5.00	5.00	5.00	53.57	53.57	53.57	53.57
1/4	3	100	1.24	6.50	9.64	8.21	7.96	45.47	68.69	29.03	40.55
1/4	3	250	1.24	4.42	4.86	3.47	3.52	50.12	62.99	14.78	15.56
1/4	3	∞	1.24	5.00	5.00	5.00	5.00	80.12	80.12	80.12	80.12
3/4	1/3	100	1.24	6.60	9.44	6.94	6.48	46.98	70.86	28.87	40.22
3/4	1/3	250	1.24	4.65	4.78	3.31	3.03	53.03	65.09	17.29	17.72
3/4	1/3	∞	1.24	5.00	5.00	5.00	5.00	80.12	80.12	80.12	80.12
3/4	3	100	2.33	10.42	10.84	8.70	8.78	41.91	52.38	17.80	37.52
3/4	3	250	2.33	5.19	7.18	4.88	4.19	29.05	52.17	13.66	18.27
3/4	3	∞	2.33	5.00	5.00	5.00	5.00	53.57	53.57	53.57	53.57

Notes: Entries for finite T are simulated rejection frequencies of the tests. Entries for $T = \infty$ are calculated as described in Remark 4.10. Power is measured at $\delta = 5$ and is size corrected for the asymptotic tests, but not for the bootstrap tests. All entries are based on $10,000$ Monte Carlo replications.

Table S.2: Tests of $H_{0,2}$: simulated size and power with conditionally heteroskedastic Models A–I

	T	size				power			
		LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests									
Model A	100	12.88	13.78	14.42	12.18	24.12	43.50	31.57	47.63
	250	13.81	13.85	13.09	6.23	42.18	51.95	42.65	50.61
Model B	100	14.44	15.23	15.83	13.05	22.76	42.59	30.39	46.32
	250	18.04	18.08	17.16	7.52	28.33	39.96	30.25	46.19
Model C	100	11.66	12.95	13.65	12.21	27.20	46.13	35.38	46.52
	250	15.37	15.27	14.65	6.90	39.26	49.22	40.53	45.16
Model D	100	12.29	13.78	14.39	12.61	27.32	44.11	32.20	44.39
	250	19.30	19.31	18.42	8.72	27.75	40.50	31.22	41.66
Model E	100	16.39	17.28	17.76	13.78	16.42	35.97	29.51	45.83
	250	20.08	20.02	19.22	7.87	27.08	40.91	32.27	42.34
Model F	100	15.68	17.03	18.05	13.71	18.16	38.68	26.60	41.67
	250	23.28	23.18	22.82	9.19	22.06	34.90	27.77	38.90
Model G	100	14.68	16.48	17.15	13.91	19.60	37.53	26.52	39.62
	250	21.08	21.35	20.49	8.23	25.56	39.47	30.03	42.24
Model H	100	24.55	26.77	27.43	18.58	8.10	20.02	17.44	35.72
	250	35.58	36.05	35.79	12.85	8.68	17.51	15.66	29.81
Model I	100	23.48	25.82	26.82	18.33	8.43	22.57	15.96	32.56
	250	27.38	27.73	26.74	9.20	15.91	30.41	23.29	33.97
Panel B: wild bootstrap tests (Algorithm 1)									
Model A	100	6.04	6.35	5.88	5.20	30.99	48.76	36.41	47.84
	250	5.88	5.78	5.83	4.93	47.24	56.80	47.52	50.01
Model B	100	6.57	6.10	6.05	5.62	30.51	47.90	36.29	47.70
	250	6.52	6.57	6.67	5.77	43.74	53.95	44.87	47.84
Model C	100	5.72	5.72	5.29	4.98	30.78	48.83	34.86	44.83
	250	5.27	5.76	5.50	5.01	42.46	52.97	44.50	45.20
Model D	100	5.69	5.58	5.23	5.02	29.66	47.78	34.84	44.79
	250	6.68	6.59	6.66	5.83	39.31	49.98	41.99	44.10
Model E	100	6.45	6.32	5.57	4.91	23.94	42.24	32.12	43.95
	250	5.95	5.96	5.95	5.22	35.34	46.48	38.67	42.38
Model F	100	6.01	6.08	5.70	5.18	24.10	42.24	31.58	41.83
	250	6.25	6.43	6.39	5.54	31.81	43.06	35.80	39.76
Model G	100	5.91	6.34	5.97	5.64	24.95	43.32	32.75	43.66
	250	5.70	5.48	5.74	5.15	36.02	47.22	39.34	42.78
Model H	100	6.76	7.27	6.32	5.75	15.53	31.11	25.21	39.10
	250	6.98	7.11	7.21	6.02	18.40	29.85	24.85	32.86
Model I	100	6.63	7.26	6.10	5.49	13.75	29.71	19.98	33.43
	250	6.27	6.52	6.52	5.64	23.50	36.85	28.83	35.00
Panel C: wild bootstrap tests (Algorithm 2)									
Model A	100	8.70	10.53	7.70	7.10	47.96	65.68	23.12	38.38
	250	5.99	6.46	4.60	3.49	47.14	61.42	16.13	17.53
Model B	100	9.49	11.08	8.19	7.54	49.29	63.18	21.79	38.10
	250	6.94	7.62	5.38	4.25	45.13	58.62	16.63	18.64
Model C	100	7.58	10.21	7.38	7.09	44.98	65.80	24.77	38.66
	250	4.93	6.43	4.09	3.61	39.65	56.41	14.98	17.82
Model D	100	7.90	10.10	7.41	7.13	45.32	64.33	23.88	38.18
	250	6.56	7.76	5.06	4.48	38.69	55.01	14.40	18.47
Model E	100	9.52	11.18	7.80	6.82	45.44	61.11	21.03	36.92
	250	6.47	7.50	4.78	3.66	37.42	54.36	13.93	18.25
Model F	100	8.82	10.49	7.83	7.43	44.70	61.64	22.01	37.45
	250	6.75	8.00	5.23	4.37	33.54	51.87	13.12	18.57
Model G	100	8.81	11.11	8.04	7.73	44.96	62.65	22.88	37.77
	250	6.03	7.31	4.55	3.92	37.10	54.98	13.82	19.30
Model H	100	13.54	12.14	8.53	8.64	45.57	51.34	15.17	35.51
	250	9.53	10.31	6.33	5.41	31.01	42.05	9.83	19.29
Model I	100	13.42	12.40	8.95	9.51	43.72	47.81	14.78	36.28
	250	7.51	9.33	6.09	4.95	29.67	47.84	11.56	17.97

Notes: Entries are simulated rejection frequencies of the tests. Power is measured at $\delta = 5$ and is size corrected for the asymptotic tests, but not for the bootstrap tests. All entries are based on 10,000 Monte Carlo replications.

Table S.3: Tests of $H_{0,3}$: simulated size and power with one-time shift in unconditional volatility

				size				power			
τ	υ	T	λ	LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests											
	1	100	1	7.73	6.50	9.89	12.53	61.46	68.19	43.14	32.92
	1	250	1	5.91	5.39	5.77	6.28	65.53	69.32	64.89	63.80
	1	∞	1	5.00	5.00	5.00	5.00	70.33	70.33	70.33	70.33
1/4	1/3	100	2.33	25.46	23.32	27.03	20.06	29.92	42.28	21.10	12.96
1/4	1/3	250	2.33	27.27	26.06	26.56	11.08	32.89	39.43	31.17	21.50
1/4	1/3	∞	2.33	27.70	27.70	27.70	5.00	35.24	35.24	35.24	35.24
1/4	3	100	1.24	12.39	11.13	15.22	14.43	51.35	60.51	17.82	13.66
1/4	3	250	1.24	10.61	9.54	10.23	7.29	56.25	60.44	54.47	51.81
1/4	3	∞	1.24	9.01	9.01	9.01	5.00	60.13	60.13	60.13	60.13
3/4	1/3	100	1.24	12.30	10.40	13.94	14.25	49.18	59.45	33.59	23.29
3/4	1/3	250	1.24	9.87	9.16	9.39	7.23	56.98	61.84	55.31	51.39
3/4	1/3	∞	1.24	9.01	9.01	9.01	5.00	60.13	60.13	60.13	60.13
3/4	3	100	2.33	28.40	26.30	31.05	21.87	26.24	40.12	14.53	11.84
3/4	3	250	2.33	27.56	26.84	26.93	11.82	31.00	37.83	28.66	19.49
3/4	3	∞	2.33	27.70	27.70	27.70	5.00	35.24	35.24	35.24	35.24
Panel B: wild bootstrap tests (Algorithm 1)											
	1	100	1	5.45	5.06	5.20	5.12	62.06	68.10	40.31	33.03
	1	250	1	5.22	5.15	5.11	4.92	66.00	69.13	64.79	62.66
	1	∞	1	5.00	5.00	5.00	5.00	70.33	70.33	70.33	70.33
1/4	1/3	100	2.33	7.28	7.49	6.22	5.34	34.71	45.72	25.12	14.56
1/4	1/3	250	2.33	6.54	6.27	6.29	5.29	35.75	41.49	33.42	22.35
1/4	1/3	∞	2.33	5.00	5.00	5.00	5.00	35.24	35.24	35.24	35.24
1/4	3	100	1.24	5.51	5.47	5.54	5.32	52.29	60.57	22.38	17.57
1/4	3	250	1.24	5.26	5.28	5.17	5.06	56.80	60.56	54.72	51.22
1/4	3	∞	1.24	5.00	5.00	5.00	5.00	60.13	60.13	60.13	60.13
3/4	1/3	100	1.24	6.27	6.08	5.33	5.23	53.47	61.75	34.36	23.72
3/4	1/3	250	1.24	5.61	5.30	5.44	5.30	57.65	62.09	56.27	51.82
3/4	1/3	∞	1.24	5.00	5.00	5.00	5.00	60.13	60.13	60.13	60.13
3/4	3	100	2.33	7.16	7.79	5.68	5.19	30.04	43.15	17.38	13.04
3/4	3	250	2.33	6.24	6.52	6.48	5.70	34.38	40.31	32.02	21.84
3/4	3	∞	2.33	5.00	5.00	5.00	5.00	35.24	35.24	35.24	35.24
Panel C: wild bootstrap tests (Algorithm 2)											
	1	100	1	5.40	6.63	5.97	5.73	59.50	70.81	23.94	17.82
	1	250	1	4.82	4.94	3.86	3.76	64.73	67.84	53.59	48.69
	1	∞	1	5.00	5.00	5.00	5.00	70.33	70.33	70.33	70.33
1/4	1/3	100	2.33	8.38	8.66	7.63	6.63	37.11	48.74	21.61	14.43
1/4	1/3	250	2.33	6.20	6.81	5.05	3.76	34.06	42.98	21.54	9.97
1/4	1/3	∞	2.33	5.00	5.00	5.00	5.00	35.24	35.24	35.24	35.24
1/4	3	100	1.24	6.00	7.47	7.29	7.12	51.16	64.42	22.26	17.62
1/4	3	250	1.24	4.91	5.32	3.92	3.61	54.82	59.62	36.76	30.14
1/4	3	∞	1.24	5.00	5.00	5.00	5.00	60.13	60.13	60.13	60.13
3/4	1/3	100	1.24	6.47	7.67	6.02	5.56	51.91	64.94	23.15	15.84
3/4	1/3	250	1.24	5.25	5.28	4.05	3.70	56.55	61.22	43.37	34.56
3/4	1/3	∞	1.24	5.00	5.00	5.00	5.00	60.13	60.13	60.13	60.13
3/4	3	100	2.33	9.31	8.70	8.49	8.28	34.99	45.86	21.08	17.25
3/4	3	250	2.33	6.07	7.23	5.15	4.07	32.49	41.83	17.86	10.80
3/4	3	∞	2.33	5.00	5.00	5.00	5.00	35.24	35.24	35.24	35.24

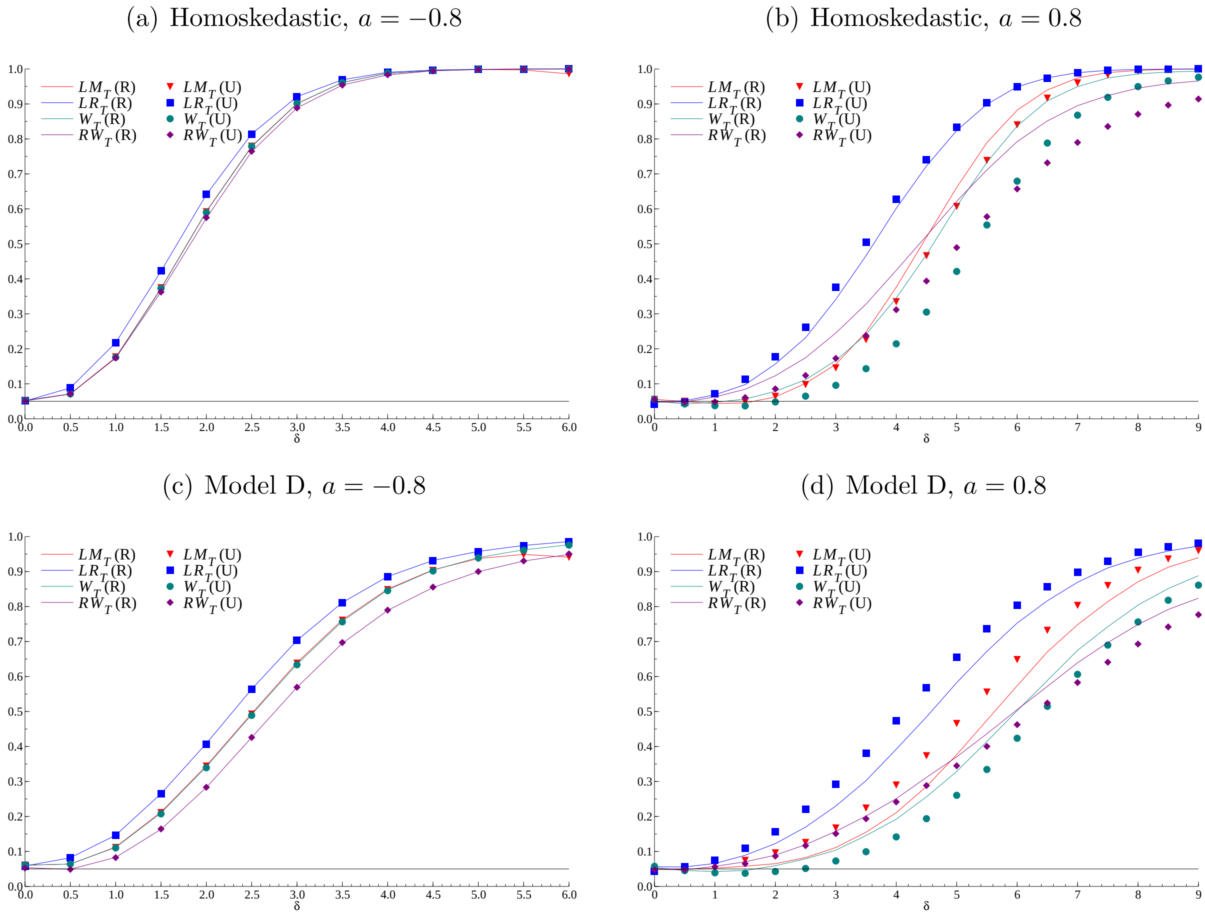
Notes: Entries for finite T are simulated rejection frequencies of the tests. Entries for $T = \infty$ are calculated as described in Remark 4.10. Power is measured at $\delta = [1, 5/3]'$ and is size corrected for the asymptotic tests, but not for the bootstrap tests. All entries are based on $10,000$ Monte Carlo replications.

Table S.4: Tests of $H_{0,3}$: simulated size and power with conditionally heteroskedastic Models A–I

	T	size				power			
		LM_T	LR_T	W_T	RW_T	LM_T	LR_T	W_T	RW_T
Panel A: asymptotic tests									
Model A	100	19.87	18.05	20.98	15.47	33.65	42.68	21.37	17.13
	250	21.71	20.42	20.84	9.12	33.38	37.22	31.97	27.24
Model B	100	23.15	20.87	24.11	17.24	29.41	38.24	19.39	15.65
	250	28.15	26.92	27.41	11.17	21.07	25.05	19.82	22.55
Model C	100	16.90	14.75	18.51	15.92	40.67	51.81	25.13	17.37
	250	21.49	20.42	21.12	9.74	34.78	40.13	32.04	28.54
Model D	100	17.84	15.57	19.64	16.24	37.35	47.97	22.00	16.58
	250	26.11	24.98	25.59	11.24	25.06	31.29	23.08	23.05
Model E	100	23.92	21.63	24.80	17.83	28.73	39.72	20.13	15.40
	250	29.48	28.24	28.91	11.07	25.09	30.88	23.13	20.32
Model F	100	22.60	20.24	24.08	17.74	31.49	41.58	17.85	12.76
	250	32.62	31.61	32.07	12.35	21.36	26.53	19.58	16.63
Model G	100	21.58	19.52	23.06	18.21	29.23	40.78	17.60	12.57
	250	29.97	28.57	29.38	11.36	22.40	27.79	19.63	16.84
Model H	100	38.16	37.24	40.48	26.25	15.18	25.09	12.01	10.60
	250	53.11	52.29	53.24	20.04	10.11	13.86	9.17	8.62
Model I	100	36.30	35.21	39.21	24.39	16.12	24.88	13.07	11.24
	250	41.17	39.82	40.48	14.15	16.34	20.71	14.68	11.49
Panel B: wild bootstrap tests (Algorithm 1)									
Model A	100	7.13	7.31	6.24	5.15	41.76	50.46	29.35	18.82
	250	6.89	6.79	6.91	5.65	41.30	45.70	40.48	30.64
Model B	100	8.21	8.92	7.19	5.68	41.80	50.30	31.42	20.63
	250	8.75	9.09	8.94	6.20	38.46	42.33	38.10	28.06
Model C	100	6.39	6.50	5.32	4.90	45.76	55.31	29.20	19.90
	250	6.12	6.20	6.15	5.38	42.07	47.41	40.52	31.65
Model D	100	6.42	6.49	5.38	5.19	44.25	54.09	29.11	20.55
	250	7.22	7.24	6.93	6.09	38.96	44.23	37.21	28.55
Model E	100	7.07	7.79	5.91	4.92	37.63	48.20	27.25	17.24
	250	6.62	6.46	6.48	5.33	34.73	39.94	33.71	22.83
Model F	100	6.55	7.06	6.08	5.50	38.78	49.18	26.03	16.67
	250	6.98	6.84	6.84	5.41	31.90	37.15	30.39	20.35
Model G	100	6.68	6.68	5.85	5.60	37.58	47.85	25.36	17.51
	250	6.34	6.35	6.49	5.72	32.44	37.84	31.18	22.10
Model H	100	8.66	10.12	7.85	6.15	26.63	39.18	22.18	14.54
	250	9.13	10.00	9.12	6.28	23.11	29.36	22.35	13.67
Model I	100	8.43	9.68	6.98	5.49	24.54	36.25	18.66	13.16
	250	8.07	8.44	8.18	5.80	26.27	31.19	24.76	15.06
Panel C: wild bootstrap tests (Algorithm 2)									
Model A	100	7.58	8.80	7.71	6.22	41.75	53.33	23.74	14.47
	250	6.78	7.21	5.73	4.06	40.36	46.25	32.16	17.36
Model B	100	9.09	9.99	8.58	6.92	41.89	52.52	25.69	15.25
	250	8.72	9.39	7.53	4.66	37.71	43.16	31.31	15.86
Model C	100	7.10	7.81	7.01	6.34	45.74	58.42	23.29	16.21
	250	5.99	6.58	4.81	3.85	40.88	48.11	29.44	18.10
Model D	100	6.99	7.74	6.88	6.43	45.10	56.94	23.19	16.47
	250	7.09	7.56	5.81	4.64	37.78	45.11	26.83	16.47
Model E	100	8.12	8.97	7.59	6.02	39.02	50.87	23.25	15.21
	250	6.60	6.95	5.50	3.77	34.29	41.21	24.88	12.41
Model F	100	7.61	8.14	7.51	7.06	40.09	51.94	22.51	15.68
	250	7.13	7.29	5.85	4.30	31.23	38.47	21.43	11.02
Model G	100	7.79	7.89	7.60	7.13	38.87	50.89	21.80	15.79
	250	6.34	6.94	5.47	4.29	31.66	39.23	22.09	11.96
Model H	100	10.49	10.93	9.40	8.99	30.29	41.67	20.53	15.20
	250	9.89	10.41	8.23	5.69	24.18	30.99	17.25	9.01
Model I	100	10.34	10.48	10.03	9.34	28.66	38.31	21.86	17.18
	250	8.42	8.97	7.40	4.91	25.75	32.83	16.86	8.22

Notes: Entries are simulated rejection frequencies of the tests. Power is measured at $\delta = [1, 5/3]'$ and is size corrected for the asymptotic tests, but not for the bootstrap tests. All entries are based on $10,000$ Monte Carlo replications.

Figure S.1: Finite sample power functions of bootstrap tests of $H_{0,1}$ with weakly dependent errors



Notes: Entries are simulated rejection frequencies of the tests measured at $\delta \in \{0.0, 0.5, 1.0, \ldots\}$. The notation (R) and (U) denotes the restricted and unrestricted bootstrap algorithms, respectively. All entries are based on 10,000 Monte Carlo replications.

S.3.3 Proofs of Lemmas A.3 and A.4

For the proof of Lemma A.3, see Lemma A.1 of Nielsen (2015) and Lemma B.3 of Johansen and Nielsen (2010), and for the proof of Lemma A.4, see Lemma B.4 of Johansen and Nielsen (2010).

S.3.4 Proof of Lemma A.5

First notice that

$$\begin{aligned}
&\left\| T^{-1} \sum_{t=1}^{T} \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_n \xi_m' \sigma_{t-n} \sigma_{t-m} g_{t,n,m} - T^{-1} \sum_{t=1}^{T} \sigma_t^4 \sum_{n,m=1}^{t-1} \xi_n \xi_m' g_{t,n,m} \right\| \\
&= \left\| T^{-1} \sum_{t=1}^{T} \sigma_t^2 \sum_{n,m=1}^{t-1} \xi_n \xi_m' (\sigma_{t-n} \sigma_{t-m} - \sigma_t^2) g_{t,n,m} \right\| \\
&\leq K T^{-1} \sum_{t=1}^{T} \sum_{n=1}^{t-1} \sum_{m=n}^{t-1} ||\xi_n|| ||\xi_m|| |\sigma_{t-n} \sigma_{t-m} - \sigma_t^2| |g_{t,n,m}| = K(r_{1T} + r_{2T}),
\end{aligned}$$

where the inequality follows by Assumption 1(b)(i) and by symmetry in n and m, and where we defined

$$r_{1T} := \sum_{n=1}^{q_T} \sum_{m=n}^{q_T} ||\xi_n||||\xi_m|| \sup_t |g_{t,n,m}| T^{-1} \sum_{t=m+1}^{T} |\sigma_{t-n}\sigma_{t-m} - \sigma_t^2|,$$
$$r_{2T} := \sum_{n=1}^{T-1} \sum_{m=\max(n,q_T+1)}^{T-1} ||\xi_n||||\xi_m|| \sup_t |g_{t,n,m}| T^{-1} \sum_{t=m+1}^{T} |\sigma_{t-n}\sigma_{t-m} - \sigma_t^2|.$$

Let $q_T := \lfloor T^{\varkappa} \rfloor$ for $\varkappa \in (0,1)$ and $G := \sup_{t\in\mathbb{Z}} \sigma_t$, which is finite by Assumption 1(b)(i). Then

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

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

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

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

$$\sup_{n,m=1,\ldots,q_T} T^{-1} \sum_{t=m+1}^{T} |\sigma_{t-n}\sigma_{t-m} - \sigma_t^2| \leq 2G \sup_{m=1,\ldots,q_T} T^{-1} \sum_{t=1}^{T-m} |\sigma_{t+m} - \sigma_t| \to 0 \text{ as } T \to \infty \tag{S.1}$$

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

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

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

The term r_{2T} is bounded as, by another application of Assumption 1(b),

$$\begin{aligned} r_{2T} &\leq 4G^2 \sum_{n=1}^{T-1} \sum_{m=\max(n,q_T+1)}^{T-1} ||\xi_n||||\xi_m|| \sup_t |g_{t,n,m}| \\ &\leq 4G^2 \sum_{m=q_T+1}^{\infty} \sum_{n=1}^{\infty} ||\xi_n||||\xi_m|| \sup_t |g_{t,n,m}| \to 0 \end{aligned}$$

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

S.4 Proofs of Variation Bounds Lemmas

We first present a lemma which contains uniform bounds on coefficient summations, which are used to prove the variation bounds in Lemmas B.1–B.3.

Lemma S.1. *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*.*

S.4.1 Proof of Lemma S.1

Part (i) is Lemma A.7 of Johansen and Nielsen (2012). To show part (ii) when $n \geq 0$ we split the summation and find the bound

$$\sum_{t=0}^{\lfloor n/2 \rfloor} |\zeta_{|t-n|}(-u,k)\zeta_t(-v,k)| + \sum_{t=\lfloor n/2 \rfloor+1}^{n} |\zeta_{|t-n|}(-u,k)\zeta_t(-v,k)| + \sum_{t=n+1}^{\infty} |\zeta_{t-n}(-u,k)\zeta_t(-v,k)|$$

$$\leq c\sum_{t=0}^{\lfloor n/2 \rfloor}(n-t)^{-u-1}(\log(n-t))^k t^{-v-1}(\log t)^k + c\sum_{t=\lfloor n/2 \rfloor+1}^{n}(n-t)^{-u-1}(\log(n-t))^k t^{-v-1}(\log t)^k$$

$$+ \sum_{t=n+1}^{\infty}(t-n)^{-u-1}(\log(t-n))^k t^{-v-1}(\log t)^k$$

$$\leq c(n/2)^{-u-1}(\log(n/2))^k \sum_{t=0}^{\lfloor n/2 \rfloor} t^{-v-1}(\log t)^k + c(n/2)^{-v-1}(\log(n/2))^k \sum_{t=\lfloor n/2 \rfloor+1}^{n}(n-t)^{-u-1}(\log(n-t))^k$$

$$+ c(n+1)^{-v-1}(\log(n+1))^k \sum_{t=n+1}^{\infty}(t-n)^{-u-1}(\log(t-n))^k$$

$$\leq c(\log n)^k n^{\max(-u-1,-v-1)}.$$

When $n < 0$ we find the bound

$$c\sum_{t=0}^{\infty}(t-n)^{-u-1}(\log(t-n))^k t^{-v-1}(\log t)^k$$

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

S.4.2 Proof of Lemma B.1

The proof is given in Lemma C.3 in Johansen and Nielsen (2010), which also applies under Assumption 1 on ε_t in place of their i.i.d. assumption.

S.4.3 Proof of Lemma B.2

We prove that, uniformly in $-1/2-\kappa \leq v \leq u \leq -1/2+\kappa$,

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

$$||M_{12NT}(u) - M_{12NT}(v)||_2 \leq c|u-v|(\log T)^2 T^{-1/2+\kappa}N^{1/2+\kappa}, \quad \text{(S.3)}$$

$$||M_{11NT}(u)||_2 \leq c(\log T)T^{-1/2}N^{1/2+2\kappa}, \quad \text{(S.4)}$$

$$||M_{11NT}(u) - M_{11NT}(v)||_2 \leq c|u-v|(\log T)^2 T^{-1/2}N^{1/2+2\kappa}, \quad \text{(S.5)}$$

where the constant $c>0$ does not depend on u,v, or T. Using the condition on α, the right-hand sides of (S.2)–(S.5) all converge to zero. Pointwise convergence in probability then follows from (S.2) and (S.4) and tightness on the interval $|u+1/2|\le\kappa$ follows from (S.3) and (S.5) using the criterion (S.38). Together this implies uniform convergence in probability.

Proof of (S.2): First evaluate

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

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, $n_k<N\le m_k$ 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 $|t_1-t_2|=|n_1-n_2|\le N$. In this case $EM_{12NT}(u)^2$ is

$$T^{-2}\sum_{\substack{t_1,t_2=N+1\\|t_1-t_2|\le N}}^{T}\sum_{n_1=0}^{N-1}\sum_{m_1=N}^{t_1-1}\sum_{m_2=\max(N,t_2-t_1+n_1)}^{t_2-1}\pi_{n_1}(-u)\pi_{t_2-t_1+n_1}(-u)\pi_{m_1}(-u)\pi_{m_2}(-u)$$
$$\times\,\sigma^2_{t_1-n_1}\sigma_{t_1-m_1}\sigma_{t_2-m_2}E(z^2_{t_1-n_1}z_{t_1-m_1}z_{t_2-m_2}). \tag{S.6}$$

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 S.1, $\sum_{n_1=0}^{N-1}\pi_{n_1}(-u)\pi_{t_2-t_1+n_1}(-u)\le\xi_N(u,u,0)$ and $\sum_{m_1=N}^{t_1-1}\pi_{m_1}(-u)\pi_{t_2-t_1+m_1}(-u)\le\xi_T(u,u,0)$ by (A.1) of Lemma A.3, so the contribution to $EM_{12NT}(u)^2$ is bounded by

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

The result when $t_1-m_1=t_2-m_2$ now follows from Lemma S.1(i). If, on the other hand, $t_1-m_1\ne t_2-m_2$ in (S.6), the expectation in (S.6) is $\kappa_4(t_1-n_1,t_1-n_1,t_1-m_1,t_2-m_2)$ and the contribution to $EM_{12NT}(u)^2$ is bounded by

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

using Assumption 1(a)(iii),(b), and this proves the result.

Proof of (S.3): Next consider $||M_{12NT}(u)-M_{12NT}(v)||_2$ 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 (A.3) of Lemma A.3 and Lemma S.1. Now apply the same proof as for (S.2), noting that only a log-factor is added. The same proof can be used for the second term.

Proof of (S.4): Note that

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

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

$$
EM_{11NT}(u)^2 = E(T^{-1}\sum_{t=N+1}^{T} w_{1t}^2)^2 - T^{-2}\sum_{t_1,t_2=N+1}^{T}\sum_{n,m=0}^{N-1}\pi_n(-u)^2\pi_m(-u)^2\sigma_{t_1-n}^2\sigma_{t_2-m}^2. \tag{S.7}
$$

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 (S.7). 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

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

where the first inequality is by Assumption 1(a)(iii),(b).

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

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

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

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

by Assumption 1(b) and Lemma S.1(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

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

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

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

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

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

in the same way as in case 2).

Proof of (S.5): Apply the same decomposition as in the proof of (S.3) and then use the same proof as for (S.4) with an extra log-factor.

S.4.4 Proof of Lemma B.3

The proof is given only for $k, l = 0$ since the derivatives just add a log-factor, see (A.1), which does not change the proof. Rearranging the summations the product moment

$M_T(u_1, u_2, \psi)$ is

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

$$= T^{-1} \sum_{j=0}^{T-1} \pi_j(-u_1) \sum_{n=0}^{\infty} \sum_{m=\max(0,j+n-T+1)}^{j+n} \zeta_{1n}(\psi)\zeta_{2m}(\psi)\pi_{j+n-m}(-u_2) \sum_{t=\max(j,j+n-m)+1}^{T} \varepsilon_{t-j-n}^2 \tag{S.8}$$

$$+ 2T^{-1} \sum_{j=0}^{T-1} \pi_j(-u_1) \sum_{n,m=0}^{\infty} \zeta_{1n}(\psi)\zeta_{2m}(\psi) \sum_{k=0}^{\min(T,j+n-m)-1} \pi_k(-u_2) \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m}. \tag{S.9}$$

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 $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |(S.8)|$ is

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

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

If $a > 0$ the summation over j is bounded and then $\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |(S.8)| = O_p(1)$ because $\sum_{n=0}^{\infty} |\zeta_{in}(\psi)| < \infty$ uniformly in $\psi \in \tilde{\Psi}, i = 1, 2$, showing (B.3) for (S.8). 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 (S.8), showing (B.4) for (S.8). In the case of (B.5), the summation over j is now

$$T^{-1} \sum_{j=1+\max(0,m-n)}^{\min(T-1,T-1+m-n)} (\frac{j}{T})^{-u_1-1}(\frac{j+n-m}{T})^{-u_2-1}$$
$$\leq cT^{-1} \sum_{j=1+\max(0,m-n)}^{\min(T-1,T-1+m-n)} (\frac{j}{T})^{-1/2+\kappa_1}(\frac{j+n-m}{T})^{-1/2+\kappa_1} < \infty$$

because $\kappa_1 > 0$, as seen easily by integral approximation, which shows (B.5) for (S.8).

Next, we analyze (S.9). Summation by parts yields

$$\sum_{k=0}^{\min(T,j+n-m)-1} \pi_k(-u_2) \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m}$$

$$= \pi_{\min(T,j+n-m)-1}(-u_2) \sum_{k=0}^{\min(T,j+n-m)-1} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m}$$

$$- \sum_{l=0}^{\min(T,j+n-m)-2} (\pi_{l+1}(-u_2) - \pi_l(-u_2)) \sum_{k=0}^{l} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m}, \tag{S.10}$$

where

$$E \sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |(S.10)| \leq \sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |\pi_{\min(T,j+n-m)-1}(-u_2)| E \left| \sum_{k=0}^{\min(T,j+n-m)-1} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} \right|$$
$$+ \sum_{l=0}^{\min(T,j+n-m)-2} \sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |\pi_{l+1}(-u_2) - \pi_l(-u_2)| E \left| \sum_{k=0}^{l} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} \right|.$$

Note that $\sum_{k=0}^{l}\sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} = \sum_{s=\max(1-m,1+j-l-m)}^{T-m} v_s$ with $v_s := \varepsilon_s \sum_{k=j+n-m-l}^{j+n-m} \varepsilon_{s-k}$ being an uncorrelated sequence that satisfies $E(v_s^2) \leq Kl$, such that

$$\left(E \left| \sum_{k=0}^{l} \sum_{t=\max(j,k)+1}^{T} \varepsilon_{t-j-n}\varepsilon_{t-k-m} \right| \right)^2 \leq E \left(\sum_{s=\max(1-m,1+j-l-m)}^{T-m} v_s \right)^2 \leq K(T+l-j)l.$$

It follows that $E\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} |(S.9)|$ is bounded by a constant times

$$\sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} \sum_{j=0}^{T-1} |\pi_j(-u_1)| \sum_{n,m=0}^{\infty} |\zeta_{1n}(\psi)\zeta_{2m}(\psi)||\pi_{\min(T,j+n-m)-1}(-u_2)| \tag{S.11}$$

$$+ \sup_{(u_1,u_2,\psi)\in\tilde{\Theta}} T^{-1} \sum_{j=0}^{T-1} |\pi_j(-u_1)| \sum_{n,m=0}^{\infty} |\zeta_{1n}(\psi)\zeta_{2m}(\psi)| \sum_{l=0}^{T-2} |\pi_{l+1}(-u_2) - \pi_l(-u_2)|(T+l-j)^{1/2} l^{1/2}. \tag{S.12}$$

The result for (S.11) follows as in the analysis of (S.8). To prove (B.3) and (B.4) for the term (S.12) we use (A.5) and that $\sum_{n=0}^{\infty} |\zeta_{in}(\psi)| < \infty$ uniformly in $\psi \in \tilde{\Psi}, i = 1, 2$, to obtain the bound

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

where the second inequality follows from Lemma A.4 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 $a \leq 0$. To prove (B.5) for the term (S.12) we note that $(T+l-j)^{1/2} \leq (2T)^{1/2}$ and find the simple bound

$$\sup_{(u_1,u_2,\psi)\in\bar{\Theta}} T^{-1} \sum_{j=1}^{T-1} (\frac{j}{T})^{-u_1-1} T^{-1} \sum_{l=1}^{T-2} (\frac{l}{T})^{-u_2-3/2} \leq cT^{-1} \sum_{j=1}^{T-1} (\frac{j}{T})^{-1/2+\kappa_1} T^{-1} \sum_{l=1}^{T-2} (\frac{l}{T})^{-1+\kappa_1} < \infty$$

because $\kappa_1 > 0$, as seen easily by integral approximation.

S.5 Proofs for QMLE and Asymptotic Tests

S.5.1 Proof of Theorem 1

The residual in (6) is given by $\varepsilon_t(\theta) = \sum_{n=0}^{t-1} b_n(\psi)\Delta_+^{d-d_{0,T}} u_{t-n}$, and clearly the convergence properties of $Q_T(\theta)$ in (8) depend on $\lim_{T\to\infty} d - d_{0,T} = d - d_0$. Define the untruncated processes

$$e_t(\psi) := c(L,\psi)\varepsilon_t = \sum_{n=0}^{\infty} c_n(\psi)\varepsilon_{t-n}, \tag{S.13}$$

$$\eta_t(\theta) := \Delta^{d-d_0} e_t(\psi) = \sum_{n=0}^{\infty} \varphi_n(\theta)\varepsilon_{t-n}, \tag{S.14}$$

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

$$c(z,\psi) := b(z,\psi)a(z,\psi_{0,T}) = \frac{a(z,\psi_{0,T})}{a(z,\psi)} = \sum_{n=0}^{\infty} c_n(\psi)z^n, \tag{S.15}$$

$$\varphi_n(\theta) := \sum_{m=0}^{n} \pi_m(d_0 - d)c_{n-m}(\psi). \tag{S.16}$$

Again, we have suppressed the T subscript on the triangular arrays $e_t(\psi)$ and $\eta_t(\theta)$ and on the coefficients $c_n(\psi)$ and $\varphi_n(\theta)$.

From Assumption 3 and Lemma A.4, there exists a $T_0 \geq 1$ such that the coefficients $c_n(\psi)$ satisfy

$$|c_n(\psi)| = O(n^{-2-\zeta}) \text{ uniformly in } \psi \in \Psi \text{ and } T \geq T_0. \tag{S.17}$$

From Lemmas A.3 and A.4 the coefficients $\varphi_n(\theta)$ then satisfy

$$|\varphi_n(\theta)| = O(n^{\max(d_0-d-1,-2-\zeta)}) \text{ uniformly in } \psi \in \Psi \text{ and } T \geq T_0, \tag{S.18}$$

such that, in particular, when $d - d_0 > -1/2$, $\eta_t(\theta)$ is a linear process with square summable coefficients. Note that the uniformity in T in (S.17) and (S.18) obtains from the uniform bound on $a_n(\psi)$ in (5), when T is sufficiently large that $\psi_{0,T} \in \Psi$.

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

$$Q(\theta) := \begin{cases} \int_0^1 \sigma(s)^2 \mathsf{d}s \sum_{n=0}^{\infty} \varphi_{0,n}(\theta)^2 & \text{if } d - d_0 > -1/2, \\ \infty & \text{if } d - d_0 \leq -1/2, \end{cases} \tag{S.19}$$

where $\varphi_{0,n}(\theta) := \sum_{m=0}^{n} \pi_m(d_0 - d) \sum_{k=0}^{n-m} b_k(\psi)a_{n-m-k}(\psi_0)$ is the same coefficient as in (S.16), but evaluated at ψ_0 instead of $\psi_{0,T}$. According to (S.19) 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 $D_1 := D_1(\kappa_1) = D \cap \{d : d - d_0 \leq -1/2 - \kappa_1\}$, $D_2 := D_2(\kappa_1,\kappa_2) = D \cap \{d : -1/2 - \kappa_1 \leq d - d_0 \leq -1/2 + \kappa_2\}$, and $D_3 := D_3(\kappa_2) = D \cap \{d : d - d_0 \geq -1/2 + \kappa_2\}$, for some constants $0 < \kappa_2 < \kappa_1 < 1/2$ 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 (S.19). 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

$$P(\inf_{\theta\in\Theta_1(\kappa_1)\cup\Theta_2(\kappa_1,\bar{\kappa}_2)} Q_T(\theta) > K) \to 1 \text{ as } T \to \infty. \tag{S.20}$$

This implies that $P(\hat{\theta} \in \Theta_3(\bar{\kappa}_2)) \to 1$ as $T \to \infty$, 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 $\kappa_2 \in (0, 1/2)$,

$$\sup_{\theta\in\Theta_3(\kappa_2)} |Q_T(\theta) - Q(\theta)| \xrightarrow{p} 0 \text{ as } T \to \infty, \tag{S.21}$$

$$\inf_{\theta\in\Theta_3(\kappa_2)\cap\{\theta:|\theta-\theta_0|\geq\epsilon\}} Q(\theta) > Q(\theta_0) \text{ for all } \epsilon > 0. \tag{S.22}$$

Condition (S.21) entails uniform convergence of the objective function on Θ_3, and condition (S.22) ensures that the optimum of the limit function is uniquely attained at the true value. For the proofs of (S.20) and (S.21) we make repeated use of the following lemma, which is the non-bootstrap version of Lemma D.4 and 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 $Q_T(\theta)$. This serves two purposes: First, the truncation in the residual in the definition of $Q_T(\theta)$ can be dispensed with in the asymptotic analysis. Secondly, the fractional order of $e_t(\psi)$ is $d_0 - d$, which is fixed and corresponds to the definitions of the parameter sets D_i, while the fractional order of $\varepsilon_t(\theta)$ is $d_{0,T} - d$, which depends on T.

Lemma S.2. *Under the assumptions of Theorem 1 and* $0 < \kappa_1 < \min(1/2, \zeta/2 + 1/4)$ *it holds that*

$$\sup_{\theta\in\Theta_1} |T^{2(d-d_0)} \sum_{t=1}^T \varepsilon_t(\theta)^2 - T^{2(d-d_0)} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2| \xrightarrow{p} 0, \tag{S.23}$$

$$\sup_{\theta\in\Theta_2\cup\Theta_3} |T^{-1} \sum_{t=1}^T \varepsilon_t(\theta)^2 - T^{-1} \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2| \xrightarrow{p} 0. \tag{S.24}$$

S.5.1.1 Proof of Lemma S.2

First decompose

$$\sum_{t=1}^T \varepsilon_t(\theta)^2 - \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2 = \sum_{t=1}^T \varepsilon_t(\theta)^2 - \sum_{t=1}^T (\Delta_+^{d-d_0} \sum_{n=0}^t b_n(\psi) u_{t-n})^2 \tag{S.25}$$

$$+ \sum_{t=1}^T (\Delta_+^{d-d_0} \sum_{n=0}^t b_n(\psi) u_{t-n})^2 - \sum_{t=1}^T (\Delta_+^{d-d_0} e_t(\psi))^2. \tag{S.26}$$

By the mean value theorem we find that

$$(S.25) = \frac{2\delta_\theta}{\sqrt{T}} \sum_{t=1}^T \left(\Delta_+^{\bar{v}} \sum_{n=0}^{t-1} b_n(\psi) u_{t-n} \right) \left(\frac{\partial}{\partial v} \Delta_+^{\bar{v}} \sum_{n=0}^{t-1} b_n(\psi) u_{t-n} \right)$$

for some intermediate value $\bar{v}$ between $d - d_{0,T}$ and $d - d_0$. We can apply Lemma B.3 directly to the right-hand side in both the non-stationary case (S.23) with normalization

$T^{2(d-d_0)}$ and in the nearly-stationary and stationary cases (S.24) with normalization T^{-1}. In either case, (S.25) is immediately shown to be uniformly negligible as required.

Next we write (S.26) as

$$(S.26) = \sum_{t=1}^{T} \Delta_+^{d-d_0} \sum_{n=0}^{t} b_n(\psi) u_{t-n} (\Delta_+^{d-d_0} \sum_{m=0}^{t} b_m(\psi) u_{t-m} - \Delta_+^{d-d_0} e_t(\psi)) \tag{S.27}$$

$$+ \sum_{t=1}^{T} \Delta_+^{d-d_0} e_t(\psi) (\Delta_+^{d-d_0} \sum_{n=0}^{t} b_n(\psi) u_{t-n} - \Delta_+^{d-d_0} e_t(\psi)) \tag{S.28}$$

and note that

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

where $\phi_{tm} := -\sum_{j=0}^{t-1} \pi_j(d_0 - d) b_{m-j}(\psi)$ satisfies, see (5) and Lemmas A.3 and A.4,

$$\begin{aligned}\sup_{\psi\in\Psi} \sum_{m=t}^{\infty} |\phi_{tm}| &\leq c \sum_{m=t}^{\infty} \sum_{j=0}^{t-1} j^{d_0-d-1} (m-j)^{-2-\zeta} \\ &\leq c \sum_{j=0}^{t-1} j^{d_0-d-1} (t-j)^{-1-\zeta} \leq c(1+\log t) t^{\max(d_0-d,-\zeta)-1}.\end{aligned} \tag{S.29}$$

Rewrite the term (S.28) as

$$(S.28) = \sum_{t=1}^{T} \sum_{j=0}^{t-1} \pi_j(d_0 - d) \sum_{n=0}^{\infty} b_n(\psi) \sum_{m=t}^{\infty} \phi_{tm} (u_{t-j-n} u_{t-m} - E(u_{t-j-n} u_{t-m})) \tag{S.30}$$

$$+ \sum_{t=1}^{T} \sum_{j=0}^{t-1} \pi_j(d_0 - d) \sum_{n=0}^{\infty} b_n(\psi) \sum_{m=t}^{\infty} \phi_{tm} E(u_{t-j-n} u_{t-m}). \tag{S.31}$$

The proof for (S.27) is identical to that for (S.28), except the summation over n in (S.27) is from t to ∞. For (S.31) we note that $\sup_t |E(u_t u_{t-n})| = \sup_t |\sum_{m=n}^{\infty} a_m(\psi_{0,T}) a_{m-n}(\psi_{0,T}) \sigma_{t-m}^2| \leq c \sum_{m=n}^{\infty} m^{-2-\zeta} (m-n)^{-2-\zeta} \leq c n^{-2-\zeta}$ where $\zeta > 0$ is given in Assumption 3(ii), such that

$$\sum_{m=t}^{\infty} |E(u_{t-j-n} u_{t-m})| \leq c |t-j-n|^{-1-\zeta}.$$

Using also (S.29) it holds that

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

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 (5) and Lemma S.1(ii). Consequently,

$$\sup_{\psi\in\Psi}|(S.31)| \leq c\sum_{t=1}^{T}(1+\log t)t^{\max(d_0-d,-\zeta)-1}\sum_{j=0}^{t-1}j^{d_0-d-1}(t-j)^{-1-\zeta}$$
$$\leq c\sum_{t=1}^{T}(1+\log t)^2 t^{2\max(d_0-d,-\zeta)-2}$$

by Lemmas A.3 and A.4. Thus, $\sup_{\theta\in\Theta_1}T^{2(d-d_0)}|(S.31)| \leq c(\log T)^3T^{-1} \to 0$ as $T\to\infty$ and $\sup_{\theta\in\Theta_2\cup\Theta_3}T^{-1}|(S.31)| \leq c(\log T)^3T^{-1+2\kappa_1} \to 0$ as $T\to\infty$.

Changing the order of the summations, (S.30) is

$$-\sum_{j=0}^{T-1}\pi_j(d_0-d)\sum_{n=0}^{\infty}b_n(\psi)\sum_{m=j+1}^{\infty}\sum_{k=0}^{\min(m,T)-1}\pi_k(d_0-d)b_{m-k}(\psi)\sum_{t=\max(j,k)+1}^{\min(m,T)}v_t, \qquad \text{(S.32)}$$

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

$$Ev_tv_s = \sum_{k_1,k_2=0}^{\infty}\sum_{l_1,l_2=0}^{\infty}a_{k_1}(\psi_{0,T})a_{k_2}(\psi_{0,T})a_{l_1}(\psi_{0,T})a_{l_2}(\psi_{0,T})\sigma_{t-j-n-k_1}\sigma_{t-m-k_2}\sigma_{s-j-n-l_1}\sigma_{s-m-l_2}$$
$$\times[E(z_{t-j-n-k_1}z_{t-m-k_2}z_{s-j-n-l_1}z_{s-m-l_2}) - E(z_{t-j-n-k_1}z_{t-m-k_2})E(z_{s-j-n-l_1}z_{s-m-l_2})].$$

The expectations are non-zero only if the two highest subscripts are equal (Lemma A.2). Routine calculations using (5), Assumption 1, and Lemma S.1(ii) show that $|Ev_tv_s| \leq c|s-t|^{-2-\zeta}$. Since the summation $\sum_{t=\max(j,k)+1}^{\min(m,T)}$ has at most m terms it follows that

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

such that $E|\sum_{t=\max(j,k)+1}^{\min(m,T)}v_t| \leq cm^{1/2}$. Using Lemma A.3 and $\sup_{\psi\in\Psi}\sum_{n=0}^{\infty}|b_n(\psi)| < \infty$, it now follows from (S.32) that (S.30) satisfies

$$E\sup_{\psi\in\Psi}|(S.30)| \leq c\sum_{j=1}^{T-1}j^{d_0-d-1}\sup_{\psi\in\Psi}\sum_{m=j+1}^{T}\sum_{k=1}^{m-1}k^{d_0-d-1}|b_m(\psi)|m^{1/2} \qquad \text{(S.33)}$$
$$+c\sum_{j=1}^{T-1}j^{d_0-d-1}\sup_{\psi\in\Psi}\sum_{m=T+1}^{\infty}\sum_{k=1}^{T-1}k^{d_0-d-1}|b_m(\psi)|m^{1/2}. \qquad \text{(S.34)}$$

For (S.33) change the order of the summations,

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

Then the bounds for (S.33) are

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

which shows the result for (S.33). Similarly, for (S.34),

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

which gives the bounds

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

showing the result for (S.34) and hence concluding the proof.

S.5.1.2 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 S.2 the difference between $T^{2(d-d_0)+1}Q_T(\theta)$ 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 (C.4) and decompose the relevant product moment as

$$T^{2(d-d_0)}\sum_{t=1}^{T}(\Delta_+^{d-d_0}e_t(\psi))^2 \geq \left(\sum_{n=0}^{\infty}c_n(\psi)\right)^2 T^{2(d-d_0)}\sum_{t=1}^{T}(\Delta_+^{d-d_0}\varepsilon_t)^2 \tag{S.35}$$
$$+2\left(\sum_{n=0}^{\infty}c_n(\psi)\right)T^{2(d-d_0)}\sum_{t=1}^{T}\Delta_+^{d-d_0}\varepsilon_t\sum_{n=0}^{\infty}\tilde{c}_n(\psi)\Delta_+^{d-d_0+1}\varepsilon_{t-n}. \tag{S.36}$$

By the Cauchy-Schwarz inequality, (S.36) is bounded by

$$2\left(\sum_{n=0}^{\infty}c_n(\psi)\right)\left(T^{2(d-d_0)}\sum_{t=1}^{T}(\Delta_+^{d-d_0}\varepsilon_t)^2\right)^{1/2}\left(T^{2(d-d_0)}\sum_{t=1}^{T}(\sum_{n=0}^{\infty}\tilde{c}_n(\psi)\Delta_+^{d-d_0+1}\varepsilon_{t-n})^2\right)^{1/2}. \tag{S.37}$$

The term in the first parenthesis satisfies $0<|\sum_{n=0}^{\infty}c_n(\psi)|<\infty$ uniformly in $\psi\in\Psi$ for T sufficiently large by Assumption 3. For the term in the second parenthesis we define $M_T(d):=T^{2(d-d_0)}\sum_{t=1}^{T}(\Delta_+^{d-d_0}\varepsilon_t)^2$, which is $O_p(1)$ by Lemma B.1. To strengthen this to hold uniformly in $d\in D_1$ it is sufficient to show that $M_T(d)$ 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 $M_T(d)$ is tight for fixed $d\in D_1$ and that

$$||M_T(u_1)-M_T(u_2)||_2 \leq c|u_1-u_2| \tag{S.38}$$

for some constant $c>0$ that does not depend on T, u_1, or u_2. The tightness condition in (S.38) is satisfied by Lemma B.1, and hence the second term in (S.37) is $O_p(1)$ uniformly in $d\in D_1$.

The term inside the third parenthesis in (S.37) can be rewritten as

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

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

where $E(T^{-1}\sum_{t=\max(j,k)+1}^{T}\varepsilon_{t-j-n}\varepsilon_{t-k-m})\leq c$ uniformly in $0\leq j,k\leq T-1$ and $\sum_{n=0}^{\infty}|\tilde{c}_n(\psi)|<\infty$ uniformly in $\psi\in\Psi$. Thus, the term inside the third parenthesis in (S.37) is a non-negative random variable with expectation

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

$$\leq c\sup_{d\in D_1}T^{2(d-d_0)+1}(\sum_{j=0}^{T-1}j^{d_0-d-2})^2\leq c(\log T)^2T^{-2\kappa_1}$$

by application of Lemma A.3, thus showing that (S.36) converges to zero in probability uniformly in $\theta\in\Theta_1$.

Next, the term (S.35) is analyzed. By the Cauchy-Schwarz inequality,

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

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 1. Firstly, the Lindeberg condition (i) of Lemma A.1 is satisfied by Lyapunov's sufficient condition because $\sum_{t=1}^{T}EU_{Tt}^4=T^{4(d-d_0)-2}\sum_{t=1}^{T}\pi_{T-t}(d_0-d+1)^4\sigma_t^4Ez_t^4\leq cT^{-2}\sum_{t=1}^{T}\left(\frac{T-t}{T}\right)^{4(d_0-d)}\leq cT^{-1}\to 0$ as $T\to\infty$. Secondly, we verify condition (ii)(a) of Lemma A.1 by showing L_2-convergence. Thus,

$$E\left(\sum_{t=1}^{T}U_{Tt}^2-E\sum_{t=1}^{T}U_{Tt}^2\right)^2=\sum_{t,s=1}^{T}E(U_{Tt}^2U_{Ts}^2)-\sum_{t,s=1}^{T}E(U_{Tt}^2)E(U_{Ts}^2)$$

$$=T^{4(d-d_0)-2}\sum_{t,s=1}^{T}\pi_{T-t}(d_0-d+1)^2\pi_{T-s}(d_0-d+1)^2\sigma_t^2\sigma_s^2[E(z_t^2z_s^2)-E(z_t^2)E(z_s^2)]$$

$$=T^{4(d-d_0)-2}\sum_{t=1}^{T}\pi_{T-t}(d_0-d+1)^4\sigma_t^4[E(z_t^4)-E(z_t^2)^2] \tag{S.39}$$

$$+2T^{4(d-d_0)-2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\pi_{T-t}(d_0-d+1)^2\pi_{T-s}(d_0-d+1)^2\sigma_t^2\sigma_s^2[E(z_t^2z_s^2)-E(z_t^2)E(z_s^2)]. \tag{S.40}$$

By Assumption 1(a)(ii),(b) and Lemma A.3, the term (S.39) is bounded by

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

The term (S.40) is

$$2T^{4(d-d_0)-2}\sum_{t=2}^{T}\sum_{r=1}^{t-1}\pi_{T-t}(d_0-d+1)^2\pi_{T-t+r}(d_0-d+1)^2\sigma_t^2\sigma_{t-r}^2\kappa_4(t,t,t-r,t-r)$$

$$\leq cT^{-2}\sum_{t=2}^{T}\left(\frac{T-t}{T}\right)^{2(d_0-d)}\left(\frac{T-1}{T}\right)^{2(d_0-d)}\sum_{r=1}^{t-1}|\kappa_4(t,t,t-r,t-r)|\leq cT^{-1}\to 0$$

using Assumption 1(a)(iii),(b) and Lemma A.3. Finally,

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

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

$$G_T(d) := T^{2(d-d_0)-1}(\sum_{t=1}^{T}\Delta_+^{d-d_0}\varepsilon_t)^2 = (T^{d-d_0-1/2}\Delta_+^{d-d_0-1}\varepsilon_T)^2 \overset{w}{\to} V(d)\chi_1^2, \tag{S.41}$$

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

To strengthen the pointwise convergence in (S.41) to weak convergence in $\mathcal{C}(D_1)$, denoted $\Rightarrow$, it is sufficient to show that $G_T(d)$ is tight (stochastically equicontinuous) as a function of the parameter, which follows by the tightness condition (S.38) and Lemma B.1. Hence the convergence in (S.41) is strengthened to $G_T(d)\Rightarrow V(d)\chi_1^2$ 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 $\inf_{d\in D_1}G_T(d)\overset{w}{\to}\inf_{d\in D_1}V(d)\chi_1^2$, which is positive almost surely. It follows that

$$\inf_{\theta\in\Theta_1}Q_T(\theta)\geq\inf_{\theta\in\Theta_1}(\textstyle\sum_{n=0}^{\infty}c_n(\psi))^2T^{2(d_0-d)-1}G_T(d)+o_p(1)$$

and, for any $K>0$,

$$P(\inf_{\theta\in\Theta_1}(\textstyle\sum_{n=0}^{\infty}c_n(\psi))^2T^{2(d_0-d)-1}G_T(d)>K)\to 1\text{ as }T\to\infty$$

because $\inf_{\psi\in\Psi}(\sum_{n=0}^{\infty}c_n(\psi))^2>0$ by Assumption 3 and $2(d_0-d)-1\geq 2\kappa_1>0$ for $d\in D_1$.

S.5.1.3 Convergence on $\Theta_2(\kappa_1,\kappa_2)$ First note that by (S.24) of Lemma S.2 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 $R_{2T}(v,\psi):=T^{-1}\sum_{t=1}^{T}(\Delta_+^{v}\varepsilon_t)(\sum_{n=0}^{\infty}\tilde{c}_n(\psi)\Delta_+^{1+v}\varepsilon_{t-n})$, and applying the decomposition (C.4), the relevant product moment is

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

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

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

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

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

$$R_{1T}(v) \ge T^{-1} \sum_{t=N+1}^{T} (\Delta_+^v \varepsilon_t)^2 \ge T^{-1} \sum_{t=N+1}^{T} w_{1t}^2 + 2T^{-1} \sum_{t=N+1}^{T} w_{1t} w_{2t}. \tag{S.42}$$

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 (B.2) of Lemma B.2 that the second term on the right-hand side of (S.42) converges in probability to zero uniformly in $|v+1/2| \le \kappa_1$ and that

$$\sup_{|v+1/2|\le\kappa_1} \left| T^{-1} \sum_{t=N+1}^{T} w_{1t}^2 - E\left(T^{-1} \sum_{t=N+1}^{T} w_{1t}^2\right)\right| \xrightarrow{p} 0 \text{ as } T \to \infty.$$

Thus, the right-hand side of (S.42) minus $E(T^{-1}\sum_{t=N+1}^{T} w_{1t}^2)$ converges uniformly in probability to zero as $T \to \infty$. It follows, see Assumption 1(b), that

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

where $F_N(v) = \sum_{n=0}^{N-1} \pi_n(-v)^2$ and $\mu_{1T}(\theta) \xrightarrow{p} 0$ as $T \to \infty$ uniformly in $|v+1/2| \le \kappa_1$ and $\psi \in \Psi$.

S.5.1.4 Proof of Eqn. (S.20) Because of restrictions imposed on the κ_i in the analysis of the sets Θ_i, we need to be careful in the proof of (S.20). We need to show that, for any $K > 0, \eta > 0$, there exists a $\bar{\kappa}_2 > 0$ and a T_0 such that

$$P(\inf_{\theta\in\Theta_1(\kappa_1)\cup\Theta_2(\kappa_1,\bar{\kappa}_2)} Q_T(\theta) < K) \le \eta$$

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

First consider the set $\Theta_1(\kappa_1)$ 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 S.5.1.2 that $P(\inf_{\theta\in\bar{\Theta}_1} Q_T(\theta) > K) \to 1$ as $T \to \infty$, i.e., for any $K > 0, \eta > 0$, there exists a T_1 such that $P(\inf_{\theta\in\bar{\Theta}_1} Q_T(\theta) < K) \le \eta/2$ for all $T \ge T_1$.

Second, having already fixed $\kappa_1 = \bar{\kappa}_1$, consider $\Theta_2(\bar{\kappa}_1, \kappa_2)$. From Section S.5.1.3 with $\kappa_1 = \bar{\kappa}_1$ and $\alpha = 1/6$,

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

where $\mu_T(\theta) = O_p(1)$ as $T \to \infty$ uniformly in $d \in [d_0 - 1/2 - \bar{\kappa}_1, d_0 - 1/2 + \bar{\kappa}_1] \supset D_2$ and $\psi \in \Psi$. As in Section D.1.2, $F_{T^{1/6}}(d - d_0) \geq 1 + c(2\kappa_2)^{-1}(1 - (T-1)^{-2\kappa_2/6})$ and $(2\kappa_2)^{-1}(1 - (T-1)^{-2\kappa_2/6}) \to \infty$ as $(\kappa_2, T) \to (0, \infty)$. 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 $\bar{\Theta}_2 := \Theta_2(\bar{\kappa}_1, \bar{\kappa}_2)$, $P(\inf_{\theta \in \bar{\Theta}_2} Q_T(\theta) < K) \leq \eta/2$ for all $T \geq T_2$.

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

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

for all $T \geq \max(T_1, T_2) = T_0$, which proves (S.20).

S.5.1.5 Convergence on $\Theta_3(\kappa_2)$ and Proof of Eqn. (S.21) First, by Lemma S.2, it suffices to demonstrate the result for $T^{-1}\sum_{t=1}^{T}(\Delta_+^{d-d_0} e_t(\psi))^2$. In this case, recall the untruncated process $\eta_t(\theta)$ defined in (S.14), 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 (S.16), with

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

for all $\theta \in \Theta_3$ (pointwise), using (S.18) and Assumption 1(b). It follows that

$$Q_T(\theta) = T^{-1} \sum_{t=1}^{T} \eta_t(\theta)^2 + o_p(1). \tag{S.43}$$

Next,

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

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

$$T^{-1} \sum_{t=1}^{T} \sum_{n=0}^{\infty} \varphi_n(\theta)^2 (\sigma_{t-n}^2 - \sigma_t^2) \leq \sum_{n=0}^{q_T} \varphi_n(\theta)^2 T^{-1} \sum_{t=1}^{T} |\sigma_{t-n}^2 - \sigma_t^2| \tag{S.44}$$

$$+ \sum_{n=q_T+1}^{\infty} \varphi_n(\theta)^2 T^{-1} \sum_{t=1}^{T} |\sigma_{t-n}^2 - \sigma_t^2|. \tag{S.45}$$

Notice that $T^{-1}\sum_{t=1}^{T} |\sigma_{t-n}^2 - \sigma_t^2| = T^{-1}\sum_{t=1}^{n} |\sigma_{t-n}^2 - \sigma_t^2| + T^{-1}\sum_{t=n+1}^{T} |\sigma_{t-n}^2 - \sigma_t^2| \leq T^{-1} 2n \sup_{t \in \mathbb{Z}} \sigma_t^2 + T^{-1}\sum_{t=n+1}^{T} |\sigma_{t-n}^2 - \sigma_t^2|$. Therefore, because $q_T = o(T)$ and $\sup_{t \in \mathbb{Z}} \sigma_t^2 \leq$

$M < \infty$ by Assumption 1(b)(i), we have

$$\sup_{n=1,\ldots,q_T} T^{-1} \sum_{t=1}^{T} |\sigma_{t-n}^2 - \sigma_t^2| \leq T^{-1} 2 q_T \sup_{t \in \mathbb{Z}} \sigma_t^2 + \sup_{n=1,\ldots,q_T} T^{-1} \sum_{t=1}^{T-n} |\sigma_{t+n}^2 - \sigma_t^2|$$
$$= \sup_{n=1,\ldots,q_T} T^{-1} \sum_{t=1}^{T-n} |\sigma_{t+n}^2 - \sigma_t^2| + o(1) \to 0,$$

where the last convergence follows from Cavaliere and Taylor (2009, Lemma A.1). Because $\sum_{n=0}^{q_T} \varphi_n(\theta)^2 \leq \sum_{n=0}^{\infty} \varphi_n(\theta)^2 < \infty$ uniformly in $\theta \in \Theta_3$, see (S.18), it thus holds that $|(S.44)| \to 0$. Next, by Assumption 1(b)(i) and by (S.18) we have, respectively, $\sup_{t \in \mathbb{Z}} \sigma_t^2 \leq M < \infty$ such that $\sup_{t \in \mathbb{Z}} T^{-1} \sum_{t=1}^{T} |\sigma_{t-n}^2 - \sigma_t^2| \leq 2M$ and $\sum_{n=q_T+1}^{\infty} \varphi_n(\theta)^2 \leq c \sum_{n=q_T+1}^{\infty} n^{2\max(d_0-d-1,-2-\zeta)} \leq c q_T^{-2\kappa_2} \to 0$ uniformly in $\theta \in \Theta_3$, and therefore $|(S.45)| \to 0$. Because $T^{-1} \sum_{t=1}^{T} \sigma_t^2 \to \int_0^1 \sigma(s)^2 \mathsf{d}s$ by Assumption 1(b)(ii) we thus have that $ET^{-1} \sum_{t=1}^{T} \eta_t(\theta)^2 = \int_0^1 \sigma(s)^2 \mathsf{d}s \sum_{n=0}^{\infty} \varphi_n(\theta)^2 + o(1)$. To prove

$$T^{-1} \sum_{t=1}^{T} \eta_t(\theta)^2 - \int_0^1 \sigma(s)^2 \mathsf{d}s \sum_{n=0}^{\infty} \varphi_n(\theta)^2 \xrightarrow{p} 0, \tag{S.46}$$

pointwise in $\theta \in \Theta_3$, it suffices to show L_2-convergence. In a similar way as in (S.39) and (S.40), we find that

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

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 1 and we find the contribution

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

using (S.18).

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

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

using Assumption 1(a)(iii),(b) together with (S.18).

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

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

where we once again used (S.18) and Assumption 1(a)(ii),(b).

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

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

as in Case 3a). This shows that (S.46) holds pointwise for all $\theta\in\Theta_3$.

Comparing the pointwise limit found in (S.46) with the definition of $Q(\theta)$ in (S.19), it remains only to show that

$$
\sup_{\theta\in\Theta_3}\left|\sum_{n=0}^{\infty}\varphi_n(\theta)^2-\sum_{n=0}^{\infty}\varphi_{0,n}(\theta)^2\right|\to 0. \tag{S.47}
$$

By the mean value theorem, the term inside the absolute value on the left-hand side is

$$\sum_{n=0}^{\infty}\left(\left(\sum_{m=0}^{n}\pi_m(d_0-d)c_{n-m}(\psi)\right)^2-\left(\sum_{l=0}^{n}\pi_l(d_0-d)\sum_{k=0}^{n-l}b_k(\psi)a_{n-l}(\psi_0)\right)^2\right)$$

$$=\frac{2\delta_\psi}{\sqrt{T}}\sum_{n=0}^{\infty}\sum_{m=0}^{n}\pi_m(d_0-d)\sum_{k=0}^{n-m}b_k(\psi)a_{n-m-k}(\bar{\psi})\sum_{l=0}^{n}\pi_l(d_0-d)\sum_{j=0}^{n-l}b_j(\psi)\dot{a}_{n-l-j}(\bar{\psi}),$$

where $\bar{\psi}$ is an intermediate value between ψ_0 and $\psi_{0,T}$. Taking the supremum of the absolute value we first find, using Assumption 3(iii) and Lemmas A.3 and A.4, that $\sup_{\psi\in\Psi}\sum_{j=0}^{n-l}|b_j(\psi)||\dot{a}_{n-l-j}(\bar{\psi})|\leq c\sum_{j=0}^{n-l}j^{-2-\zeta}(n-l-j)^{-1-\zeta}\leq c(n-l)^{-1-\zeta}$ and $\sup_{\psi\in\Psi}\sum_{k=0}^{n-m}|b_k(\psi)||a_{n-m-k}(\bar{\psi})|\leq c\sum_{k=0}^{n-m}k^{-2-\zeta}(n-m-k)^{-2-\zeta}\leq c(n-m)^{-2-\zeta}$. Thus, the left-hand side of (S.47) is bounded by

$$\frac{c}{\sqrt{T}}\sum_{n=0}^{\infty}\sum_{m=0}^{n}m^{-1/2-\kappa_2}(n-m)^{-2-\zeta}\sum_{l=0}^{n}l^{-1/2-\kappa_2}(n-l)^{-1-\zeta}\leq\frac{c}{\sqrt{T}}\sum_{n=0}^{\infty}n^{-1/2-\kappa_2}n^{-1/2-\kappa_2}\leq\frac{c}{\sqrt{T}}\to 0.$$

Combining (S.43), (S.46), and (S.47), we obtain the pointwise limit, i.e.

$$Q_T(\theta)\xrightarrow{p}Q(\theta). \tag{S.48}$$

The result (S.48) can be strengthened to uniform convergence in probability by showing that $T^{-1}\sum_{t=1}^{T}(\Delta_+^{d-d_0}e_t(\psi))^2$ is stochastically equicontinuous (or tight). From Newey (1991, Corollary 2.2) this holds if the derivative of $Q_T(\theta)$ is dominated uniformly in $\theta\in\Theta_3$ by a random variable $B_T=O_p(1)$. From Lemma B.3 with $u_1=u_2=d-d_0\geq -1/2+\kappa_2, a=2\kappa_2$, and $\tilde{\Psi}=\Psi$ it holds that $B_T=\sup_{\theta\in\Theta_3}|\frac{\partial}{\partial\theta}T^{-1}\sum_{t=1}^{T}(\Delta_+^{d-d_0}e_t(\psi))^2|=O_p(1)$ (noting that only summability of the linear coefficients is assumed in Lemma B.3 and this is satisfied uniformly on Θ by the derivatives of $c_n(\psi)$ by Assumption 3(iii)). This shows that $T^{-1}\sum_{t=1}^{T}(\Delta_+^{d-d_0}e_t(\psi))^2$ is stochastically equicontinuous on Θ_3 and hence that (S.48) holds uniformly in $\theta\in\Theta_3$ in view of Lemma S.2. Since the result holds for any κ_2 it proves (S.21).

S.5.1.6 Proof of Eqn. (S.22) Since $Q(\theta_0)=\int_0^1\sigma(s)^2\mathsf{d}s$ it is sufficient to prove that

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

Because $\varphi_0(\theta)=1$ for all $\theta\in\Theta_3$ by Assumption 3, 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 4 the inequality is strict for all $\theta\neq\theta_0$, which proves (S.22) by continuity of $\varphi_n(\cdot)$ and compactness of Θ_3.

S.5.2 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,

$$0=T^{1/2}\frac{\partial Q_T(\hat{\theta})}{\partial\theta}=T^{1/2}\frac{\partial Q_T(\theta_{0,T})}{\partial\theta}+T^{1/2}\frac{\partial^2 Q_T(\bar{\theta})}{\partial\theta\partial\theta'}(\hat{\theta}-\theta_{0,T}), \tag{S.49}$$

where $\bar{\theta}$ is an intermediate value satisfying $|\bar{\theta}_i - \theta_{0,T,i}| \leq |\hat{\theta}_i - \theta_{0,T,i}|$ for $i = 1, \ldots, p+1$. Recalling the definition of $\xi_n(\theta_1, \theta_2)$ in (D.38), we note for this proof that, for example, $\xi_n(\theta_0, \theta) = \frac{\partial \varphi_{0,n}(\theta)}{\partial \theta}$, see (S.14) and (S.16). We also define $\xi_{0,n} := \xi_n(\theta_0, \theta_0) = [-n^{-1}, \gamma_n(\psi_0)']'$, which satisfies

$$\sum_{n=0}^{s} ||\xi_{0,n}|| = O(\log s) \text{ and } \sum_{n=0}^{s} (\xi_{0,n})_i^q = O(1) \text{ for any } q > 1, s \geq 2, i = 1, \ldots, p+1, \quad \text{(S.50)}$$

by Assumption 3(iii) and (5).

S.5.2.1 Convergence of the Score Function The normalized score function evaluated at the true value is

$$T^{1/2} \frac{\partial Q_T(\theta_{0,T})}{\partial \theta} = 2T^{-1/2} \sum_{t=1}^{T} \varepsilon_t(\theta_{0,T}) \hat{y}_{1,t-1} \text{ with } \hat{y}_{k,t-1} := \frac{\partial^k}{\partial \theta^{(k)}} \varepsilon_t(\theta_{0,T}).$$

Define also $S_T := 2T^{-1/2} \sum_{t=1}^{T} \varepsilon_t y_{1,t-1}$, where $y_{1,t-1} := \sum_{n=1}^{t-1} \xi_{0,n} \varepsilon_{t-n}$. That is, 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 $-\sum_{n=1}^{t-1} n^{-1} \varepsilon_{t-n}(\theta_{0,T})$ and the remaining elements are $\sum_{n=1}^{t-1} \dot{b}_n(\psi_{0,T}) u_{t-n}$.

We next show that

$$T^{1/2} \frac{\partial Q_T(\theta_{0,T})}{\partial \theta} - S_T = o_p(1). \quad \text{(S.51)}$$

The left-hand side of (S.51) is

$$T^{1/2} \frac{\partial Q_T(\theta_{0,T})}{\partial \theta} - S_T = 2T^{-1/2} \sum_{t=1}^{T} (\varepsilon_t(\theta_{0,T}) - \varepsilon_t) \hat{y}_{1,t-1} + 2T^{-1/2} \sum_{t=1}^{T} \varepsilon_t (\hat{y}_{1,t-1} - y_{1,t-1}),$$

where

$$\varepsilon_t(\theta_{0,T}) - \varepsilon_t = -\sum_{n=t}^{\infty} b_n(\psi_{0,T}) u_{t-n}$$

and

$$\hat{y}_{1,t-1} - y_{1,t-1} = \begin{bmatrix} -\sum_{n=1}^{t-1} n^{-1} \sum_{k=t-n}^{\infty} b_k(\psi_{0,T}) u_{t-n-k} \\ \sum_{n=1}^{t-1} \dot{b}_n(\psi_{0,T}) \sum_{k=t}^{\infty} a_k(\psi_{0,T}) \varepsilon_{t-k} \end{bmatrix}.$$

The first term on the right-hand side of (S.51) is then

$$-2T^{-1/2} \sum_{t=1}^{T} \sum_{n=t}^{\infty} b_n(\psi_{0,T}) u_{t-n} \hat{y}_{1,t-1},$$

which has second moment

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

see (5). The second term on the right-hand side of (S.51) is

$$\begin{bmatrix} -2T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{n=1}^{t-1}n^{-1}\sum_{k=t-n}^{\infty}b_k(\psi_{0,T})u_{t-n-k} \\ 2T^{-1/2}\sum_{t=1}^{T}\varepsilon_t\sum_{n=1}^{t-1}\dot{b}_n(\psi_{0,T})\sum_{k=t}^{\infty}a_k(\psi_{0,T})\varepsilon_{t-k} \end{bmatrix}. \tag{S.52}$$

The first term in (S.52) has second moment

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

where the first two inequalities use Assumption 1(a)(iii),(b) and (5), and the fourth inequality uses Lemma A.4. The second term in (S.52) has second moment

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

using Lemma A.2, Assumption 3(iii), and (5). Thus, each of the terms in (S.52), and hence those in (S.51), converge to zero in L_2-norm and therefore in probability.

Because $y_{1,t-1}$ is measurable with respect to the sigma-algebra $\mathcal{F}_{t-1} := \sigma(\{\varepsilon_s, s \le t-1\})$, 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 MDS with respect to the filtration $\mathcal{F}_t$. To apply the central limit theorem for martingales, see Lemma A.1, we first verify the Lindeberg condition (i) via Lyapunov's sufficient condition that $\sum_{t=1}^{T}E||v_{Tt}||^{2+\epsilon}\to 0$ for some $\epsilon > 0$. Thus,

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

using Assumption 1(b), Minkowski's inequality, (S.50), and Assumption 5 with ϵ chosen such that $2\epsilon+4\le 8$. It follows that $\sum_{t=1}^{T}E||v_{Tt}||^{2+\epsilon}\le KT^{-\epsilon/2}(\log T)^{2+\epsilon}\to 0$.

Next, we verify condition (ii)(a) of Lemma A.1. The sum of squares of v_{Tt} is

$$4T^{-1}\sum_{t=1}^{T}\sigma_t^2 z_t^2 \sum_{n,m=1}^{t-1}\xi_{0,n}\xi_{0,m}'\sigma_{t-n}\sigma_{t-m}z_{t-n}z_{t-m}$$

$$= 4T^{-1}\sum_{t=1}^{T}\sigma_t^2 \sum_{n,m=1}^{t-1}\xi_{0,n}\xi_{0,m}'\sigma_{t-n}\sigma_{t-m}\tau_{n,m} \tag{S.53}$$

$$+ 4T^{-1}\sum_{t=1}^{T}\sigma_t^2 \sum_{n,m=1}^{t-1}\xi_{0,n}\xi_{0,m}'\sigma_{t-n}\sigma_{t-m}(z_t^2 z_{t-n}z_{t-m}-\tau_{n,m}). \tag{S.54}$$

The second moment of the (i,j)'th element of (S.54) is

$$16T^{-2}\sum_{t,s=1}^{T}\sigma_t^2\sigma_s^2\sum_{n,m=1}^{s-1}\sum_{k,l=1}^{t-1}(\xi_{0,m})_i(\xi_{0,n})_j(\xi_{0,k})_i(\xi_{0,l})_j\sigma_{s-n}\sigma_{s-m}\sigma_{t-k}\sigma_{t-l}Cov(z_t^2 z_{t-k}z_{t-l}, z_s^2 z_{s-n}z_{s-m})$$

$$\leq KT^{-2}\sum_{t,s=1}^{T}\sum_{n,m=1}^{s-1}\sum_{k,l=1}^{t-1}||\xi_{0,m}||||\xi_{0,n}||||\xi_{0,k}||||\xi_{0,l}|||Cov(z_t^2 z_{t-k}z_{t-l}, z_s^2 z_{s-n}z_{s-m})|$$

$$= KT^{-2}\sum_{t=1}^{T}\sum_{n,m=1}^{t-1}\sum_{k,l=1}^{t-1}||\xi_{0,m}||||\xi_{0,n}||||\xi_{0,k}||||\xi_{0,l}|||Cov(z_t^2 z_{t-n}z_{t-m}, z_t^2 z_{t-k}z_{t-l})| \tag{S.55}$$

$$+ KT^{-2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{n,m=1}^{s-1}\sum_{k,l=1}^{t-1}||\xi_{0,m}||||\xi_{0,n}||||\xi_{0,k}||||\xi_{0,l}|||Cov(z_t^2 z_{t-k}z_{t-l}, z_s^2 z_{s-n}z_{s-m})|. \tag{S.56}$$

For (S.55) we find the simple bound

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

using (S.50) and that z_t has finite eighth order moments by Assumption 5. The covariance in (S.56) is a combination of the cumulants of z_t up to order eight, where, apart from the eighth order term, each term is a product of two cumulants whose orders sum to eight. For the term with the eighth order cumulant we find the bound

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

by Assumption 5. There are no seventh order cumulants in (S.56) because they would be multiplied by a first order cumulant, which is zero. For the terms with products of sixth and second order cumulants we find, for example,

$$T^{-2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{n,m=1}^{s-1}\sum_{k,l=1}^{t-1}||\xi_{0,m}||||\xi_{0,n}||||\xi_{0,k}||||\xi_{0,l}|||\kappa_2(t-k,t-l)|\kappa_6(t,t,s,s,s-n,s-m)|$$

$$\leq KT^{-2}\sum_{t=2}^{T}\left(\sum_{s=1}^{t-1}\sum_{n,m=1}^{s-1}|\kappa_6(t,t,s,s,s-n,s-m)|\right)\left(\sum_{k=1}^{t-1}||\xi_{0,k}||^2\right) \leq KT^{-1}(\log T) \to 0$$

by (S.50) and Assumption 5. Another example is

$$\begin{aligned}
&T^{-2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{n,m=1}^{s-1}\sum_{k,l=1}^{t-1}||\xi_{0,m}||||\xi_{0,n}||||\xi_{0,k}||||\xi_{0,l}|||\kappa_2(t,t)|\kappa_6(t-k,t-l,s,s,s-n,s-m)| \\
&\leq KT^{-2}\sum_{t=2}^{T}\sum_{k=1}^{t-1}\sum_{n,m=1}^{t-1}\sum_{s=\max(n,m)+1}^{t-1-k}||\xi_{0,m}||||\xi_{0,n}||||\xi_{0,k}||^2|\kappa_6(t-k,t-k,s,s,s-n,s-m)| \\
&\quad+KT^{-2}\sum_{t=2}^{T}\sum_{k,l=1}^{t-1}\sum_{s=t-\min(k,l)}^{t-1}\sum_{n,m=1}^{s-1}||\xi_{0,m}||||\xi_{0,n}||||\xi_{0,k}||||\xi_{0,l}|||\kappa_6(t-k,t-l,s,s,s-n,s-m)| \\
&\leq KT^{-2}\sum_{t=2}^{T}\sum_{k=1}^{t-1}||\xi_{0,k}||^2\sum_{n,m=1}^{t-1}\sum_{s=\max(n,m)+1}^{t-1-k}|\kappa_6(t-k,t-k,s,s,s-n,s-m)| \\
&\quad+KT^{-2}\sum_{s=1}^{T-1}\sum_{k=1}^{t-1}||\xi_{0,k}||\sum_{t=\max(k,s)+1}^{s+k-1}\sum_{n,m=1}^{s-1}\sum_{l=k}^{t-1}|\kappa_6(t-k,t-l,s,s,s-n,s-m)|
\end{aligned}$$

using Lemma A.2 and symmetry. Here, both terms are clearly $O(T^{-1}(\log T))$ by (S.50) and Assumption 5. The remaining products of sixth and second order cumulants, as well as products of lower order cumulants, are treated similarly, thus proving that (S.56) and hence (S.54) is $o_p(1)$.

By Lemma A.5 with $g_{t,n,m}=\tau_{n,m}$, (S.53) is, apart from a $o(1)$ term,

$$4T^{-1}\sum_{t=1}^{T}\sigma_t^4\sum_{n,m=1}^{t-1}\xi_{0,n}\xi_{0,m}'\tau_{n,m}=4T^{-1}\sum_{t=1}^{T}\sigma_t^4\sum_{n,m=1}^{\infty}\xi_{0,n}\xi_{0,m}'\tau_{n,m}-4T^{-1}\sum_{t=1}^{T}\sigma_t^4\sum_{n,m=t}^{\infty}\xi_{0,n}\xi_{0,m}'\tau_{n,m},$$

where the first term on the right-hand side is $4A_0T^{-1}\sum_{t=1}^{T}\sigma_t^4\to 4A_0\int_0^1\sigma^4(s)\mathsf{d}s$ and the second term on the right-hand side is bounded by $KT^{-1}\sum_{t=1}^{T}\sum_{n,m=t}^{\infty}||\xi_{0,n}||||\xi_{0,m}|||\tau_{n,m}|$, which converges to zero because it is the Cesàro mean of the sequence $\sum_{n,m=t}^{\infty}||\xi_{0,n}||||\xi_{0,m}|||\tau_{n,m}|$, which itself converges to zero as $t\to\infty$ since it is the tail of a convergent sum, see Assumption 1(a)(iii) and Remark 4.6.

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

$$4T^{-1}\sum_{t=1}^{T}\sigma_t^2z_t^2\sum_{m,n=1}^{t-1}\xi_{0,m}\xi_{0,n}'\sigma_{t-m}\sigma_{t-n}z_{t-m}z_{t-n}\xrightarrow{p}4A_0\int_0^1\sigma^4(s)\mathsf{d}s. \tag{S.57}$$

Hence, by the central limit theorem for martingales, see Lemma A.1, we have $S_T\xrightarrow{w}N(0,4A_0\int_0^1\sigma^4(s)\mathsf{d}s)$ and therefore also $T^{1/2}\frac{\partial Q_T(\theta_{0,T})}{\partial\theta}\xrightarrow{w}N(0,4A_0\int_0^1\sigma^4(s)\mathsf{d}s)$ by (S.51).

S.5.2.2 Convergence of the Hessian The second derivative in (S.49) 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 B.3 with $u_1=u_2=d-d_{0,T}\geq-1/2+\kappa_2/2$ (for T sufficiently large) and $\tilde{\Psi}=\mathcal{N}_\delta(\psi_0)$ (noting that only summability of the linear coefficients is assumed in Lemma B.3 and this is satisfied uniformly on $\mathcal{N}_\delta(\psi_0)$ by the derivatives of $c_n(\psi)$ by Assumption 6) it holds that $B_T=\sup_{d\in D_3,\psi\in\mathcal{N}_\delta(\psi_0)}|\frac{\partial^3Q_T(\theta)}{\partial\theta^{(3)}}|=O_p(1)$, showing that the second derivative in (S.49) is tight. This result, together with $|\hat{\theta}-\theta_{0,T}|\xrightarrow{p}0$ (Theorem 1), implies by Lemma A.3 of

Johansen and Nielsen (2010) that the second derivative in (S.49) can be evaluated at the true value, $\theta_{0,T}$. Hence, we examine

$$\frac{\partial^2 Q_T(\theta_{0,T})}{\partial\theta\partial\theta'} = 2T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta_{0,T})\hat{y}_{2,t-1} + 2T^{-1}\sum_{t=1}^{T}\hat{y}_{1,t-1}\hat{y}'_{1,t-1},$$

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 MDS, and it has finite variance such that the first term of H_T is $o_p(1)$.

The second term of H_T is $2T^{-1}\sum_{t=1}^{T}\sum_{n,m=1}^{t-1}\xi_{0,n}\varepsilon_{t-n}\xi'_{0,m}\varepsilon_{t-m}$, which converges in L_2-norm, and hence in probability, to $2B_0\int_0^1\sigma^2(s)\mathsf{d}s$ exactly as in Section D.2.2 (just replacing $\xi_n^\dagger$ with $\xi_{0,n}$).

S.5.2.3 Proof of (17) To prove the result for $\hat{A}$ we write

$$\begin{aligned}\hat{A} &= T^{-1}\sum_{t=1}^{T}\frac{\partial\ell_t(\hat{\theta},\hat{\sigma}^2)}{\partial\theta}\frac{\partial\ell_t(\hat{\theta},\hat{\sigma}^2)}{\partial\theta'} = \frac{1}{\hat{\sigma}^4}T^{-1}\sum_{t=1}^{T}\varepsilon_t(\hat{\theta})^2\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta'} \\ &= \frac{1}{\hat{\sigma}^4}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_t(\hat{\theta})^2\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta'} - T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta_{0,T})^2\hat{y}_{1,t-1}\hat{y}'_{1,t-1}\right) \qquad \text{(S.58)} \\ &+ \frac{1}{\hat{\sigma}^4}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta_{0,T})^2\hat{y}_{1,t-1}\hat{y}'_{1,t-1} - T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 y_{1,t-1}y'_{1,t-1}\right) \qquad \text{(S.59)} \\ &+ \frac{1}{\hat{\sigma}^4}T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 y_{1,t-1}y'_{1,t-1}. \qquad \text{(S.60)}\end{aligned}$$

First of all, $\hat{\sigma}^2 = Q_T(\hat{\theta}) \xrightarrow{p} Q(\theta_0) = \int_0^1\sigma^2(s)\mathsf{d}s$ by the uniform convergence in (S.48), Theorem 1, and Johansen and Nielsen (2010, Lemma A.3).

Next, we decompose the (i,j)'th element of (S.58) and apply the Cauchy-Schwarz inequality,

$$\begin{aligned}&\frac{1}{\hat{\sigma}^4}T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2 - \varepsilon_t(\theta_{0,T})^2)\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta'_j} + \frac{1}{\hat{\sigma}^4}T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta_{0,T})^2\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta'_j} - \hat{y}_{1,t-1,i}\hat{y}'_{1,t-1,j}\right) \\ &\le \frac{1}{\hat{\sigma}^4}\left(T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2 - \varepsilon_t(\theta_{0,T})^2)^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta'_j}\right)^2\right)^{1/2} \qquad \text{(S.61)} \\ &+ \frac{1}{\hat{\sigma}^4}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta_{0,T})^4\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta'} - \hat{y}_{1,t-1}\hat{y}'_{1,t-1}\right)^2\right)^{1/2}. \qquad \text{(S.62)}\end{aligned}$$

The proofs for (S.61) and (S.62) are nearly identical, so we give only the former. The second large parenthesis in (S.61) is $O_p(1)$ by Lemma B.3. By the mean value theorem,

$$T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2 - \varepsilon_t(\theta_{0,T})^2)^2 = 4\sum_{i=1}^{p+1}(\hat{\theta}_i - \theta_{0,T,i})T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2 - \varepsilon_t(\theta_{0,T})^2)\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta_i}$$

for an intermediate value, $\bar{\theta}$, between $\hat{\theta}$ and $\theta_{0,T}$. By another application of the Cauchy-Schwarz inequality,

$$T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2-\varepsilon_t(\theta_{0,T})^2)\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta_i}\leq(T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2-\varepsilon_t(\theta_{0,T})^2)^2)^{1/2}(T^{-1}\sum_{t=1}^{T}(\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta_i})^2)^{1/2},$$

which is also $O_p(1)$ by Lemma B.3. Because $(\hat{\theta}_i-\theta_{0,T,i})=O_p(T^{-1/2})$ by Theorem 2 and $\hat{\sigma}^4\xrightarrow{p}(\int_0^1\sigma^2(s)\mathsf{d}s)^2$, it follows that (S.61) is $o_p(1)$. Next, (S.59) is negligible by the exact same argument as in the proof of (S.51), and finally (S.60) is $\hat{\sigma}^{-4}\sum_{t=1}^{T}v_{Tt}v_{Tt}'/4\xrightarrow{p}(\int_0^1\sigma^2(s)\mathsf{d}s)^{-2}A_0\int_0^1\sigma^4(s)\mathsf{d}s=\lambda A_0$ by (S.57) and using $\hat{\sigma}^2\xrightarrow{p}\int_0^1\sigma^2(s)\mathsf{d}s$. It follows that $\hat{A}\xrightarrow{p}\lambda A_0$.

For the second result we find that

$$\hat{B}=-\frac{\partial^2L_T(\hat{\theta},\hat{\sigma}^2)}{\partial\theta\partial\theta'}=\frac{1}{2\hat{\sigma}^2}\frac{\partial^2Q_T(\hat{\theta},\hat{\sigma}^2)}{\partial\theta\partial\theta'}\xrightarrow{p}\frac{1}{2\int_0^1\sigma^2(s)\mathsf{d}s}2B_0\int_0^1\sigma^2(s)\mathsf{d}s=B_0$$

by the proof in Section S.5.2.2 and using $\hat{\sigma}^2\xrightarrow{p}\int_0^1\sigma^2(s)\mathsf{d}s$. Finally, it now follows straightforwardly, using Assumption 7 and Slutsky's Theorem, that $\hat{C}\xrightarrow{p}C_0$.

S.5.3 Proof of Theorem 3

Consider first the Wald statistic. From (16) of Theorem 2 we find, under $H_{1,T}$, that

$$\sqrt{T}(M'\hat{\theta}-m)\xrightarrow{w}N(\delta,M'C_0M).$$

It follows by (17) and the continuous mapping theorem that $W_T\xrightarrow{w}Y'F_0Y$. For the robust Wald statistic, the result follows in the same way by Theorem 2 and the continuous mapping theorem. Finally, the proofs for the LM and LR statistics apply standard mean-value or Taylor series expansions; for a textbook treatment, see for example Hayashi (2000, Section 7.4).

S.5.4 Proof of Theorem 4

Again, consider first the Wald statistic. Under the fixed alternative H_1 in (3) the true value is θ_0, i.e. $\delta_\theta=0$, and is such that $M'\theta_0=\bar{m}\neq m$. From Theorem 2 we then find

$$\sqrt{T}(M'\hat{\theta}-m)+\sqrt{T}(m-\bar{m})\xrightarrow{w}N(0,M'C_0M).$$

Since $\hat{B}\xrightarrow{p}B_0$ by (17), it follows that

$$\begin{aligned}W_T&=(\sqrt{T}(m-\bar{m})+O_p(1))'(M'B_0^{-1}M+o_p(1))^{-1}(\sqrt{T}(m-\bar{m})+O_p(1))\\&=T(m-\bar{m})'(M'B_0^{-1}M)^{-1}(m-\bar{m})+O_p(T^{1/2}).\end{aligned}$$

The proofs for the LM, LR, and robust Wald statistics again follow by standard expansions.

S.6 Additional Proofs for Bootstrap Estimator and Tests

S.6.1 Proof of Lemma D.1

Recall that $\hat{\varepsilon}_{c,t} = \varepsilon_t(\hat{\theta}) - T^{-1}\sum_{s=1}^{T}\varepsilon_s(\hat{\theta})$ and decompose as

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_{c,t}^2 - \varepsilon_t^2)^2 &= T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2 + (T^{-1}\sum_{s=1}^{T}\varepsilon_s(\hat{\theta}))^2 - 2\varepsilon_t(\hat{\theta})T^{-1}\sum_{s=1}^{T}\varepsilon_s(\hat{\theta}) - \varepsilon_t^2)^2 \\
&= T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2 - \varepsilon_t^2)^2 + (T^{-1}\sum_{s=1}^{T}\varepsilon_s(\hat{\theta}))^4 + 4T^{-1}\sum_{t=1}^{T}\varepsilon_t(\hat{\theta})^2(T^{-1}\sum_{s=1}^{T}\varepsilon_s(\hat{\theta}))^2 \qquad \text{(S.63)} \\
&\quad + \text{cross product terms.}
\end{aligned}
$$

The cross product terms are asymptotically of the required order by the Cauchy-Schwarz inequality, after dealing with the first three terms on the right-hand side.

First we write $\varepsilon_t(\hat{\theta}) = \varepsilon_t(\hat{\theta}) - \varepsilon_t(\theta_{0,T}) + \varepsilon_t(\theta_{0,T}) - \varepsilon_t + \varepsilon_t$ and find that

$$
T^{-1}\sum_{s=1}^{T}\varepsilon_s(\hat{\theta}) = T^{-1}\sum_{s=1}^{T}(\varepsilon_s(\hat{\theta}) - \varepsilon_s(\theta_{0,T})) + T^{-1}\sum_{s=1}^{T}(\varepsilon_s(\theta_{0,T}) - \varepsilon_s) + T^{-1}\sum_{s=1}^{T}\varepsilon_s, \quad \text{(S.64)}
$$

where the last term is clearly $O_p(T^{-1/2})$ under Assumption 1. Using (6), the second term of (S.64) is $T^{-1}\sum_{s=1}^{T}\sum_{m=s}^{\infty}b_m(\psi_{0,T})u_{s-m}$, which has zero mean and variance bounded by

$$
cT^{-2}\sum_{t,s=1}^{T}\sum_{m=s}^{\infty}\sum_{n=t}^{\infty}b_m(\psi_{0,T})b_n(\psi_{0,T}) \le cT^{-2}\sum_{t,s=1}^{T}\sum_{m=s}^{\infty}\sum_{n=t}^{\infty}m^{-2-\zeta}n^{-2-\zeta} \le cT^{-2}\sum_{t,s=1}^{T}s^{-1-\zeta}t^{-1-\zeta} \le cT^{-2},
$$

see (5), so that the second term of (S.64) is $O_p(T^{-1})$ by L_2-convergence. For the first term of (S.64) we apply the mean value theorem,

$$
T^{-1}\sum_{s=1}^{T}(\varepsilon_s(\hat{\theta}) - \varepsilon_s(\theta_{0,T})) = (\hat{\theta} - \theta_{0,T})'T^{-1}\sum_{t=1}^{T}\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta},
$$

where $\bar{\theta}$ is an intermediate value between $\hat{\theta}$ and $\theta_{0,T}$. By the Cauchy-Schwarz inequality and Lemma B.3, $T^{-1}\sum_{t=1}^{T}\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta} \le (T^{-1}\sum_{t=1}^{T}(\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta})^2)^{1/2} = O_p(1)$ when $\bar{d}$ is close to d_0. Since $(\hat{\theta} - \theta_{0,T}) = O_p(T^{-1/2})$ by Theorem 2, this shows that the first term of (S.64), and hence (S.64), is $O_p(T^{-1/2})$. Because $T^{-1}\sum_{t=1}^{T}\varepsilon_t(\hat{\theta})^2 = O_p(1)$ it follows that the second and third terms of (S.63) are both $O_p(T^{-1})$ such that we are left with the first term on the right-hand side of (S.63).

To deal with the first term of (S.63), we again write $\varepsilon_t(\hat{\theta}) = \varepsilon_t(\hat{\theta}) - \varepsilon_t(\theta_{0,T}) + \varepsilon_t(\theta_{0,T}) - \varepsilon_t + \varepsilon_t$ and find that

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2 - \varepsilon_t^2)^2 &= T^{-1}\sum_{t=1}^{T}((\varepsilon_t(\hat{\theta}) - \varepsilon_t(\theta_{0,T}))^4 + T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\theta_{0,T}) - \varepsilon_t)^4 \qquad \text{(S.65)} \\
&\quad + \text{cross product terms.}
\end{aligned}
$$

Again, the cross product terms are asymptotically of the required order by the Cauchy-Schwarz inequality, if the first two terms on the right-hand side are dealt with. Using

(6), the second term on the right-hand side of (S.65) is $T^{-1}\sum_{t=1}^{T}(\sum_{m=t}^{\infty} b_m(\psi_{0,T})u_{t-m})^4$, which is a non-negative random variable with mean

$$T^{-1}\sum_{t=1}^{T} E(\sum_{m=t}^{\infty} b_m(\psi_{0,T})u_{t-m})^4 \le cT^{-1}\sum_{t=1}^{T}(\sum_{m=t}^{\infty} b_m(\psi_{0,T}))^4 \le cT^{-1}\sum_{t=1}^{T}(\sum_{m=t}^{\infty} m^{-2-\zeta})^4$$
$$\le cT^{-1}\sum_{t=1}^{T} t^{-4-4\zeta} \le cT^{-1},$$

see (5), which shows that the second term of (S.65) is $O_p(T^{-1})$ by L_1-convergence. For the first term of (S.65), we apply the mean value theorem followed by the Cauchy-Schwarz inequality,

$$T^{-1}\sum_{t=1}^{T}((\varepsilon_t(\hat{\theta}) - \varepsilon_t(\theta_{0,T}))^4$$
$$= 4\sum_{i=1}^{p+1}(\hat{\theta}_i - \theta_{0,T,i})T^{-1}\sum_{t=1}^{T}((\varepsilon_t(\hat{\theta}) - \varepsilon_t(\theta_{0,T}))^3 \frac{\partial \varepsilon_t(\bar{\theta})}{\partial \theta_i}$$
$$\le 4\sum_{i=1}^{p+1}(\hat{\theta}_i - \theta_{0,T,i})\left(T^{-1}\sum_{t=1}^{T}((\varepsilon_t(\hat{\theta}) - \varepsilon_t(\theta_{0,T}))^6\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left(\frac{\partial \varepsilon_t(\bar{\theta})}{\partial \theta_i}\right)^2\right)^{1/2},$$

where $\bar{\theta}$ is an intermediate value between $\hat{\theta}$ and $\theta_{0,T}$ and $T^{-1}\sum_{t=1}^{T}((\varepsilon_t(\hat{\theta}) - \varepsilon_t(\theta_{0,T}))^6$ is at most $O_p(1)$. Since $T^{-1}\sum_{t=1}^{T}(\frac{\partial \varepsilon_t(\bar{\theta})}{\partial \theta_i})^2 = O_p(1)$ by Lemma B.3 and $(\hat{\theta} - \theta_{0,T}) = O_p(T^{-1/2})$ by Theorem 2, this shows that the first term of (S.65) is $O_p(T^{-1/2})$, and hence completes the proof.

S.6.2 Proof of Lemma D.2

The proofs for the two cases $h = k + 1$ and $h \le m - 1$ are identical, so we give only the former. First apply summation by parts,

$$\sum_{j=m}^{k} \lambda_j(\theta) \sum_{t=k+2}^{T} \varepsilon^*_{t-j}\varepsilon^*_{t-k-1} = \lambda_k(\theta)\sum_{j=m}^{k}\sum_{t=k+2}^{T} \varepsilon^*_{t-j}\varepsilon^*_{t-k-1}$$
$$- \sum_{q=m}^{k-1}(\lambda_{q+1}(\theta) - \lambda_q(\theta))\sum_{j=m}^{q}\sum_{t=k+2}^{T} \varepsilon^*_{t-j}\varepsilon^*_{t-k-1},$$

which implies that

$$E^*\sup_{\theta}\left|\sum_{j=m}^{k} \lambda_j(\theta) \sum_{t=k+2}^{T} \varepsilon^*_{t-j}\varepsilon^*_{t-k-1}\right| \le \sup_{\theta}|\lambda_k(\theta)|E^*\left|\sum_{j=m}^{k}\sum_{t=k+2}^{T} \varepsilon^*_{t-j}\varepsilon^*_{t-k-1}\right|$$
$$+ \sup_{\theta}\sum_{q=m}^{k-1}|\lambda_{q+1}(\theta) - \lambda_q(\theta)|E^*\left|\sum_{j=m}^{q}\sum_{t=k+2}^{T} \varepsilon^*_{t-j}\varepsilon^*_{t-k-1}\right|.$$

Next notice that, by Jensen's inequality,

$$\left(E^*\left|\sum_{j=m}^{k}\sum_{t=k+2}^{T}\varepsilon^*_{t-j}\varepsilon^*_{t-k-1}\right|\right)^2 \leq E^*\left|\sum_{j=m}^{k}\sum_{t=k+2}^{T}\varepsilon^*_{t-j}\varepsilon^*_{t-k-1}\right|^2$$
$$= \sum_{j=m}^{k}\sum_{j'=m}^{k}\sum_{t=k+2}^{T}\sum_{t'=k+2}^{T}E^*(\varepsilon^*_{t-j}\varepsilon^*_{t'-j'}\varepsilon^*_{t-k-1}\varepsilon^*_{t'-k-1})$$
$$= 2\sum_{j=m}^{k}\sum_{t=k+2}^{T}E^*(\varepsilon^{*2}_{t-j}\varepsilon^{*2}_{t-k-1}) = 2\sum_{j=m}^{k}\sum_{t=k+2}^{T}\hat{\varepsilon}^2_{c,t-j}\hat{\varepsilon}^2_{c,t-k-1},$$

where, by the Cauchy-Schwarz inequality,

$$\sum_{j=m}^{k}\sum_{t=k+2}^{T}\hat{\varepsilon}^2_{c,t-j}\hat{\varepsilon}^2_{c,t-k-1} \leq \sum_{j=m}^{k}(\sum_{t=k+2}^{T}\hat{\varepsilon}^4_{c,t-j})^{1/2}(\sum_{t=k+2}^{T}\hat{\varepsilon}^4_{c,t-k-1})^{1/2} \leq \sum_{j=m}^{k}\sum_{t=1}^{T}\hat{\varepsilon}^4_{c,t} = (k-m+1)O_p\left(T\right).$$

Therefore,

$$E^*\sup_\theta\left|\sum_{j=m}^{k}\lambda_j(\theta)\sum_{t=k+2}^{T}\varepsilon^*_{t-j}\varepsilon^*_{t-k-1}\right| \leq \sup_\theta|\lambda_k(\theta)|k^{1/2}O_p(T^{1/2}) + \sup_\theta\sum_{q=m}^{k-1}|\lambda_{q+1}(\theta)-\lambda_q(\theta)|q^{1/2}O_p(T^{1/2})$$
$$\leq O_p(T^{1/2}k^{g+1/2}) + O_p(T^{1/2})\sum_{q=m}^{k}q^{g-1/2},$$

which proves the result.

S.6.3 Proof of Lemma D.3

Reversing the summations, we find $M^*_{12NT}(u) = T^{-1}\sum_{n=0}^{N-1}\pi_n(-u)\sum_{m=N}^{T-1}\pi_m(-u)\sum_{t=m+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_{t-m}$, and we apply Lemma D.2 with $g = -1/2+\kappa$,

$$E^*\sup_{|u+1/2|\leq\kappa}\left|\sum_{m=N}^{T-1}\pi_m(-u)\sum_{t=m+1}^{T}\varepsilon^*_{t-n}\varepsilon^*_{t-m}\right| = O_p(T^{1/2+\kappa}),$$

which implies that

$$E^*\sup_{|u+1/2|\leq\kappa}|M^*_{12NT}(u)| \leq \sup_{|u+1/2|\leq\kappa}T^{-1}\sum_{n=0}^{N-1}|\pi_n(-u)|O_p(T^{1/2+\kappa}) = O_p(N^{\kappa+1/2}T^{\kappa-1/2}),$$

so that $\sup_{|u+1/2|\leq\kappa}|T^{-1}\sum_{t=N+1}^{T}w^*_{1t}w^*_{2t}| = o^*_p(1)$, in probability, by setting $N = T^\alpha$ with $\alpha < (1/2-\kappa)/(1/2+\kappa)$.

Next, we decompose $M^*_{11NT}(u)$ as

$$M^*_{11NT}(u) = T^{-1}\sum_{t=N+1}^{T}\sum_{n=0}^{N-1}\pi_n(-u)^2(\varepsilon^{*2}_{t-n}-\sigma^2_{t-n}) \tag{S.66}$$
$$+ T^{-1}\sum_{t=N+1}^{T}\sum_{n\neq m=0}^{N-1}\pi_n(-u)\pi_m(-u)\varepsilon^*_{t-n}\varepsilon^*_{t-m}, \tag{S.67}$$

where

$$E^* \sup_{|u+1/2|\le\kappa} |(S.67)| = \sup_{|u+1/2|\le\kappa} \sum_{n\neq m=0}^{N-1} |\pi_n(-u)||\pi_m(-u)| E^* \left| T^{-1} \sum_{t=N+1}^{T} \varepsilon^*_{t-n}\varepsilon^*_{t-m} \right|$$
$$\le c \sum_{n\neq m=0}^{N-1} n^{\kappa-1/2} m^{\kappa-1/2} O_p(T^{-1/2}) = O_p(N^{2\kappa+1}T^{-1/2}),$$

with the first inequality following from (D.57). Thus, $E^* \sup_{|u+1/2|\le\kappa} |(S.67)| = o_p(1)$ when $N = T^\alpha$ with $\alpha < 1/(4\kappa+2)$. We decompose (S.66) as

$$(S.66) = T^{-1} \sum_{t=N+1}^{T} \sum_{n=0}^{N-1} \pi_n(-u)^2 (\varepsilon^{*2}_{t-n} - \hat\varepsilon^2_{c,t-n}) \tag{S.68}$$
$$+ T^{-1} \sum_{t=N+1}^{T} \sum_{n=0}^{N-1} \pi_n(-u)^2 (\hat\varepsilon^2_{c,t-n} - \varepsilon^2_{t-n}) \tag{S.69}$$
$$+ T^{-1} \sum_{t=N+1}^{T} \sum_{n=0}^{N-1} \pi_n(-u)^2 (\varepsilon^2_{t-n} - \sigma^2_{t-n}), \tag{S.70}$$

and show that each of these terms are asymptotically negligible (in the sense of $\xrightarrow{p^*}_p 0$). First,

$$E^* \sup_{|u+1/2|\le\kappa} |(S.68)| \le \sup_{|u+1/2|\le\kappa} \sum_{n=0}^{N-1} \pi_n(-u)^2 E^* \left| T^{-1} \sum_{t=N+1}^{T} (\varepsilon^{*2}_{t-n} - \hat\varepsilon^2_{c,t-n}) \right|$$
$$\le c \sum_{n=0}^{N-1} n^{2\kappa-1} O_p(T^{-1/2}) = O_p(N^{2\kappa}T^{-1/2}),$$

where the second inequality follows by (D.32). Thus, $E^* \sup_{|u+1/2|\le\kappa} |(S.68)| = o_p(1)$ for $N = T^\alpha$ with $\alpha < 1/(4\kappa)$.

Next, using $\varepsilon_t = \sigma_t z_t$,

$$\left(E \left| T^{-1} \sum_{t=N+1}^{T} \sigma^2_{t-n}(z^2_{t-n} - 1) \right| \right)^2 \le E \left(T^{-1} \sum_{t=N+1}^{T} \sigma^2_{t-n}(z^2_{t-n} - 1) \right)^2$$
$$= T^{-2} \sum_{t,s=N+1}^{T} \sigma^2_{t-n}\sigma^2_{s-n}\kappa_4(t-n, t-n, s-n, s-n)$$
$$\le cT^{-2} \sum_{t,s=N+1}^{T} |\kappa_4(t-n, t-n, s-n, s-n)| \le cT^{-1}$$

by Assumption 1(a)(iii),(b). Thus,

$$E \sup_{|u+1/2|\le\kappa} |(S.70)| \le c \sup_{|u+1/2|\le\kappa} \sum_{n=0}^{N-1} \pi_n(-u)^2 T^{-1/2} = O(N^{2\kappa}T^{-1/2}),$$

such that $\sup_{|u+1/2|\le\kappa} |(S.70)| = o_p(1)$ for $N = T^\alpha$ with $\alpha < 1/(4\kappa)$.

For the term (S.69), we apply the Cauchy-Schwarz inequality,

$$|(S.69)| \leq \sum_{n=0}^{N-1} \pi_n(-u)^2 \left(T^{-1} \sum_{t=N+1}^{T} (\hat{\varepsilon}_{c,t-n}^2 - \varepsilon_{t-n}^2)^2 \right)^{1/2},$$

where the last term is $O_p(T^{-1/2})$ by Lemma D.1, uniformly in $n = 0, \ldots, N-1$ and $N = 1, \ldots, T-1$, and the first term satisfies $\sup_{|u+1/2| \leq \kappa} \sum_{n=0}^{N-1} \pi_n(-u)^2 \leq cN^{2\kappa}$. Thus, $\sup_{|u+1/2|\leq\kappa} |(S.69)| = O_p(N^{2\kappa} T^{-1/2})$ which is $o_p(1)$ when $N = T^\alpha$ with $\alpha < 1/(4\kappa)$.

S.6.4 Proof of Lemma D.4

The bootstrap residual is

$$\begin{aligned}
\varepsilon_t^*(\theta) &= \sum_{n=0}^{t-1} b_n(\psi) \Delta_+^{d-\check{d}} \sum_{m=0}^{t-n-1} a_m(\check{\psi}) \varepsilon_{t-n-m}^* \\
&= \sum_{n=0}^{\infty} b_n(\psi) \Delta_+^{d-\check{d}} \sum_{m=0}^{\infty} a_m(\check{\psi}) \varepsilon_{t-n-m}^* = \Delta_+^{d-\check{d}} e_t^*(\psi),
\end{aligned}$$

where the first equality is the definition in (25), the second is because $\varepsilon_t^* = 0$ for $t \leq 0$ in step (iii) of Algorithms 1 and 2, and the final equality is by definition of $e_t^*(\psi)$ and $\check{c}(L, \psi)$, see (D.6) and (D.7). The results (D.13) and (D.14) are trivial consequences of $\varepsilon_t^*(\theta) = \Delta_+^{d-\check{d}} e_t^*(\psi)$.

S.6.5 Proof for remainder in Eqn. (D.17)

With $\check{M}_T^*(d) := T^{2(d-\check{d})} \sum_{t=1}^{T} (\Delta_+^{d-\check{d}} \varepsilon_t^*)^2$ we find from (D.16) that

$$\begin{aligned}
T^{2(d-\check{d})} \sum_{t=1}^{T} (\Delta_+^{d-\check{d}} e_t^*(\psi))^2 &= \left(\sum_{n=0}^{\infty} \check{c}_n(\psi) \right)^2 \check{M}_T^*(d) + T^{2(d-\check{d})} \sum_{t=1}^{T} \left(\sum_{n=0}^{\infty} \bar{c}_n(\psi) \Delta_+^{d-\check{d}+1} \varepsilon_{t-n}^* \right)^2 \\
&\quad + 2 \left(\sum_{n=0}^{\infty} \check{c}_n(\psi) \right) T^{2(d-\check{d})} \sum_{t=1}^{T} \Delta_+^{d-\check{d}} \varepsilon_t^* \sum_{n=0}^{\infty} \bar{c}_n(\psi) \Delta_+^{d-\check{d}+1} \varepsilon_{t-n}^*
\end{aligned}$$

such that (because the second term on the right-hand side is non-negative)

$$\begin{aligned}
q_{1,T}^*(\theta) &= 2 \left(\sum_{n=0}^{\infty} \check{c}_n(\psi) \right) T^{2(d-\check{d})} \sum_{t=1}^{T} \Delta_+^{d-\check{d}} \varepsilon_t^* \sum_{n=0}^{\infty} \bar{c}_n(\psi) \Delta_+^{d-\check{d}+1} \varepsilon_{t-n}^* \\
&\leq 2 \left(\sum_{n=0}^{\infty} \check{c}_n(\psi) \right) \check{M}_T^*(d)^{1/2} \left(T^{2(d-\check{d})} \sum_{t=1}^{T} (\sum_{n=0}^{\infty} \bar{c}_n(\psi) \Delta_+^{d-\check{d}+1} \varepsilon_{t-n}^*)^2 \right)^{1/2} \qquad \text{(S.71)}
\end{aligned}$$

using the Cauchy-Schwarz inequality. The term in the first parenthesis satisfies $0 < |\sum_{n=0}^{\infty} \check{c}_n(\psi)| < \infty$ almost surely uniformly in $\psi \in \Psi$.

Next, we show that $\check{M}_T^*(d) = O_p^*(1)$, in probability, uniformly in $d \in \check{D}_1$. For the pointwise argument, first note that

$$\begin{aligned}
E^* \check{M}_T^*(d) &= T^{2(d-\check{d})} \sum_{t=1}^{T} \sum_{j,k=0}^{t-1} \pi_j(\check{d} - d) \pi_k(\check{d} - d) E^*(\varepsilon_{t-j}^* \varepsilon_{t-k}^*) \\
&= T^{2(d-\check{d})} \sum_{t=1}^{T} \sum_{j=0}^{t-1} \pi_j(\check{d} - d)^2 \hat{\varepsilon}_{c,t-j}^2 = T^{2(d-\check{d})+1} \sum_{j=0}^{T-1} \pi_j(\check{d} - d)^2 T^{-1} \sum_{t=j+1}^{T} \hat{\varepsilon}_{c,t-j}^2,
\end{aligned}$$

where the second equality follows by uncorrelatedness of ε_t^*, conditional on the original data, the third equality by reversing the order of the summations, and where $T^{-1}\sum_{t=j+1}^{T}\hat{\varepsilon}_{c,t-j}^2 = O_p(1)$ uniformly in $j = 0, \ldots, T-1$. Thus, by Lemma A.3,

$$E^*\check{M}_T^*(d) = O_p(1)T^{-1}\sum_{j=0}^{T-1}(j/T)^{2(\check{d}-d-1)} \leq O_p(1)T^{-1}\sum_{j=0}^{T-1}(j/T)^{-1+2\kappa_1},$$

where the inequality applies the definition of $\check{D}_1$ and $T^{-1}\sum_{j=0}^{T-1}(j/T)^{-1+2\kappa_1} \to \int_0^1 u^{-1+2\kappa_1}\mathsf{d}u < \infty$ because $-1+2\kappa_1 > -1$. Thus, $\check{M}_T^*(d) = O_p^*(1)$, in probability, pointwise for any $d \in \check{D}_1$. To strengthen this to hold uniformly in $d \in D_1^\dagger$ it is sufficient to show that $\check{M}_T^*(d)$ is tight (in probability) as a stochastic process on the space of continuous functions indexed by the parameter d. Using the mean value theorem, the tightness condition in (D.19) is satisfied by the same proof as the pointwise proof that $\check{M}_T^*(d) = O_p^*(1)$, in probability, except the derivative means we apply (A.2) from Lemma A.3 and find $T^{-1}\sum_{j=0}^{T-1}(j/T)^{-1+2\kappa_1}(1+\log|j/T|) \to \int_0^1 u^{-1+2\kappa_1}(1+\log|u|)\mathsf{d}u < \infty$ because $-1+2\kappa_1 > -1$. It follows that the second term on the right-hand side of (S.71), i.e. $\check{M}_T^*(d)$, is $O_p^*(1)$ in probability, uniformly in $d \in \check{D}_1$.

The term inside the second large parenthesis in (S.71) can be rewritten as

$$T^{2(d-\check{d})}\sum_{t=1}^{T}\sum_{n,m=0}^{\infty}\bar{c}_n(\psi)\bar{c}_m(\psi)\sum_{j,k=0}^{t-1}\pi_j(\check{d}-d-1)\pi_k(\check{d}-d-1)\varepsilon_{t-j-n}^*\varepsilon_{t-k-m}^*$$
$$= T^{2(d-\check{d})+1}\sum_{n,m=0}^{\infty}\bar{c}_n(\psi)\bar{c}_m(\psi)\sum_{j,k=0}^{T-1}\pi_j(\check{d}-d-1)\pi_k(\check{d}-d-1)T^{-1}\sum_{t=\max(j,k)+1}^{T}\varepsilon_{t-j-n}^*\varepsilon_{t-k-m}^*.$$

Taking the supremum we find the bound

$$\sup_{\theta\in\check{\Theta}_1} T^{2(d-\check{d})+1}\sum_{n,m=0}^{\infty}|\bar{c}_n(\psi)\bar{c}_m(\psi)|\sum_{j,k=0}^{T-1}|\pi_j(\check{d}-d-1)\pi_k(\check{d}-d-1)|\left|T^{-1}\sum_{t=\max(j,k)+1}^{T}\varepsilon_{t-j-n}^*\varepsilon_{t-k-m}^*\right|, \tag{S.72}$$

which is $o_p^*(1)$, in probability, thereby implying that $\sup_{\theta\in\check{\Theta}_1}|(S.71)| = o_p^*(1)$, in probability. To see that (S.72) is $o_p^*(1)$, in probability, note that

$$E^*\left|T^{-1}\sum_{t=\max(j,k)+1}^{T}\varepsilon_{t-j-n}^*\varepsilon_{t-k-m}^*\right| \leq T^{-1}\sum_{t=\max(j,k)+1}^{T}|\hat{\varepsilon}_{c,t-j-n}||\hat{\varepsilon}_{c,t-k-m}|$$
$$\leq \left(T^{-1}\sum_{t=\max(j,k)+1}^{T}\hat{\varepsilon}_{c,t-j-n}^2\right)^{1/2}\left(T^{-1}\sum_{t=\max(j,k)+1}^{T}\hat{\varepsilon}_{c,t-k-m}^2\right)^{1/2} \leq T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{c,t}^2.$$

Then (S.72) is a non-negative random variable with (conditional) expectation

$$\sup_{\theta\in\check{\Theta}_1} T^{2(d-\check{d})+1}\sum_{n,m=0}^{\infty}|\bar{c}_n(\psi)\bar{c}_m(\psi)|\sum_{j,k=0}^{T-1}|\pi_j(\check{d}-d-1)\pi_k(\check{d}-d-1)|T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_{c,t}^2,$$

where $\sum_{n=0}^{\infty} |\bar{c}_n(\psi)| < \infty$ almost surely uniformly in $\psi \in \Psi$. This leaves the bound

$$E^* \sup_{\theta\in\check{\Theta}_1} T^{2(d-\check{d})} \sum_{t=1}^{T} (\sum_{n=0}^{\infty} \bar{c}_n(\psi)\Delta_+^{d-\check{d}+1}\varepsilon_{t-n}^*)^2 \leq c \sup_{d\in\check{D}_1} T^{2(d-\check{d})+1} \left(\sum_{j=0}^{T-1} |\pi_j(\check{d}-d-1)|\right)^2 T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_{c,t}^2$$
$$= O_p(\sup_{d\in\check{D}_1} T^{2(d-\check{d})+1} (\sum_{j=0}^{T-1} j^{\check{d}-d-2})^2) = O_p((\log T)^2 T^{-2\kappa_1})$$

by application of Lemma A.3.

S.6.6 Proof for remainder in Eqn. (D.18)

By independence (conditional on the original data) of ε_t^* we find $\sum_{t=1}^{T} E^*(U_{Tt}^{*2}|\mathcal{F}_{t-1}^*) = T^{2(d-\check{d}-1/2)} \sum_{t=1}^{T} \pi_{T-t}(\check{d}-d+1)^2 \hat{\varepsilon}_{c,t}^2$, such that

$$q_{2,T}(d) = T^{2(d-\check{d}-1/2)} \sum_{t=1}^{T} \pi_{T-t}(\check{d}-d+1)^2 (\hat{\varepsilon}_{c,t}^2 - \varepsilon_t^2) \tag{S.73}$$
$$+ \sum_{t=1}^{T} (T^{2(d-\check{d}-1/2)} \pi_{T-t}(\check{d}-d+1)^2 - T^{2(d-d^\dagger-1/2)} \pi_{T-t}(d^\dagger - d+1)^2) \varepsilon_t^2 \tag{S.74}$$
$$+ T^{2(d-d^\dagger-1/2)} \sum_{t=1}^{T} \pi_{T-t}(d^\dagger - d+1)^2 \varepsilon_t^2 - V^\dagger(d). \tag{S.75}$$

Applying the Cauchy-Schwarz inequality, (S.73) is bounded as

$$|(S.73)| \leq \left(T^{4(d-\check{d})-1} \sum_{t=1}^{T} \pi_{T-t}(\check{d}-d+1)^4\right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} (\hat{\varepsilon}_{c,t}^2 - \varepsilon_t^2)^2\right)^{1/2},$$

where the term in the second parenthesis is $o_p(1)$ by Lemma D.1 and the term in the first parenthesis is bounded (uniformly for $d \in \check{D}_1$) using Lemma A.3 as $cT^{-1} \sum_{t=1}^{T} (\frac{T-t}{T})^{4(\check{d}-d)} \leq cT^{-1} \sum_{t=1}^{T} (\frac{T-t}{T})^{2+4\kappa_1} \leq c$.

To analyze (S.74), we apply the mean value theorem and note that the derivative of $f(\check{d}) := T^{2(d-\check{d}-1/2)} \pi_{T-t}(\check{d}-d+1)^2$ is bounded as

$$\left|\frac{\partial f(\check{d})}{\partial \check{d}}\right| \leq c(1 + \log|\tfrac{T-t}{T}|) T^{2(d-\check{d}-1/2)} (T-t)^{2(\check{d}-d)}$$

using (A.2) of Lemma A.3. Then, $(S.74) = (\check{d} - d^\dagger) \sum_{t=1}^{T} \frac{\partial f(\bar{d})}{\partial \check{d}} \varepsilon_t^2$, where $\bar{d}$ is an intermediate value between $\check{d}$ and $d^\dagger$. For any $\epsilon > 0$, $|\bar{d} - \check{d}| \leq \epsilon$ with probability converging to one, so that

$$\sup_{d\in\check{D}_1} |(S.74)| \leq |\check{d} - d^\dagger| T^{-1} \sum_{t=1}^{T} (1 + \log|\tfrac{T-t}{T}|)(\tfrac{T-t}{T})^{1-2\kappa_1-2\epsilon} \varepsilon_t^2$$
$$\leq |\check{d} - d^\dagger| T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 = O_p(|\check{d} - d^\dagger|) = o_p(1).$$

Finally, (S.75) is $o_p(1)$ by the same argument as the corresponding term in Section S.5.1.2.

S.6.7 Proof of Bound for $R^*_{2T}(\check{v},\psi)$ in Eqn. (D.21)

To bound $R^*_{2T}(\check{v},\psi)$ we note that the summation over n can be truncated at $n = t-1$ because $\varepsilon^*_t = 0$ for $t \le 0$, and we decompose as

$$R^*_{2T}(\check{v},\psi) = T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\pi_j(-\check{v})\varepsilon^*_{t-j}\sum_{n=0}^{t-1}\bar{c}_n(\psi)\sum_{k=0}^{t-n-1}\pi_k(-\check{v}-1)\varepsilon^*_{t-n-k}$$

$$= T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\pi_j(-\check{v})\sum_{n=0}^{j-1}\bar{c}_n(\psi)\pi_{j-n}(-\check{v}-1)\varepsilon^{*2}_{t-j} \tag{S.76}$$

$$+ T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\pi_j(-\check{v})\sum_{n=0}^{t-1}\bar{c}_n(\psi)\sum_{k=j-n+1}^{t-n-1}\pi_k(-\check{v}-1)\varepsilon^*_{t-j}\varepsilon^*_{t-n-k} \tag{S.77}$$

$$+ T^{-1}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\pi_j(-\check{v})\varepsilon^*_{t-j}\sum_{n=0}^{t-1}\bar{c}_n(\psi)\sum_{k=0}^{j-n-1}\pi_k(-\check{v}-1)\varepsilon^*_{t-n-k}. \tag{S.78}$$

We give the proofs for (S.76) and (S.77) only, since the proof for (S.78) is the same as that for (S.77). Reversing the summations,

$$(S.76) = \sum_{j=0}^{T-1}\pi_j(-\check{v})\sum_{n=0}^{j-1}\bar{c}_n(\psi)\pi_{j-n}(-\check{v}-1)T^{-1}\sum_{t=j+1}^{T}\varepsilon^{*2}_{t-j},$$

where $T^{-1}\sum_{t=j+1}^{T}\varepsilon^{*2}_{t-j} = O^*_p(1)$, in probability, uniformly in $j = 0,\ldots,T-1$, which leaves the bound

$$\sup_{\theta\in\check{\Theta}_2}|(S.76)| \le \sup_{\theta\in\check{\Theta}_2} c\sum_{j=0}^{T-1} j^{-\check{v}-1}\sum_{n=0}^{j-1}|\bar{c}_n(\psi)|(j-n)^{-\check{v}-2}O^*_p(1)$$
$$= \sup_{\theta\in\check{\Theta}_2} c\sum_{n=0}^{T-1}|\bar{c}_n(\psi)|\sum_{j=n+1}^{T-1} j^{-\check{v}-1}(j-n)^{-\check{v}-2}O^*_p(1) = O^*_p(1),$$

in probability, using Lemma A.3 and that $\sum_{n=0}^{T-1}|\bar{c}_n(\psi)| < \infty$ almost surely uniformly in $\psi\in\Psi$.

Next write $(S.77) = T^{-1}\sum_{n=0}^{T-1}\bar{c}_n(\psi)\sum_{k=n+1}^{T-1}\pi_k(-\check{v}-1)\sum_{j=0}^{k-1+n}\pi_j(-\check{v})\sum_{t=k+n+1}^{T}\varepsilon^*_{t-j}\varepsilon^*_{t-n-k}$ and apply Lemma D.2 with $g = -1/2+\kappa_1$,

$$E^*\sup_{d\in\check{D}_2}\left|\sum_{j=0}^{k-1+n}\pi_j(-\check{v})\sum_{t=k+n+1}^{T}\varepsilon^*_{t-j}\varepsilon^*_{t-n-k}\right| = O_p(T^{1/2}(k+n)^{\kappa_1}),$$

and hence

$$E^*\sup_{\theta\in\check{\Theta}_2}|(S.77)| \le \sup_{\theta\in\check{\Theta}_2} T^{-1/2}\left|\sum_{n=0}^{T-1}\bar{c}_n(\psi)\right|\sum_{k=n+1}^{T-1}k^{-\check{v}-2}(k+n)^{\kappa_1}O_p(1) \le O_p((\log T)T^{\max(2\kappa_1-1,-1/2)}),$$

because $\sum_{n=0}^{T-1}|\bar{c}_n(\psi)| < \infty$ almost surely uniformly in $\psi\in\Psi$. Thus, $\sup_{\theta\in\check{\Theta}_2}|(S.77)| = o^*_p(1)$, in probability.

S.6.8 Proof of variance of (D.63)

The variance of the (i,j)'th element of (D.63) is, apart from an asymptotically negligible term due to (D.64),

$$4T^{-2}\sum_{t,s=1}^{T}\sum_{m,n=1}^{s-1}\sum_{k,l=1}^{t-1}(\xi_m^\dagger)_i(\xi_n^\dagger)_j(\xi_k^\dagger)_i(\xi_l^\dagger)_j\sigma_{s-m}\sigma_{s-n}\sigma_{t-k}\sigma_{t-l}$$
$$\times[E(z_{s-m}z_{s-n}z_{t-k}z_{t-l})-E(z_{s-m}z_{s-n})E(z_{t-k}z_{t-l})]$$
$$\leq KT^{-2}\sum_{t,s=1}^{T}\sum_{m,n=1}^{s-1}\sum_{k,l=1}^{t-1}||\xi_m^\dagger||||\xi_n^\dagger||||\xi_k^\dagger||||\xi_l^\dagger|||E(z_{s-m}z_{s-n}z_{t-k}z_{t-l})-E(z_{s-m}z_{s-n})E(z_{t-k}z_{t-l})|,$$

using Assumption 1(b) to bound the σ_t's. Here, the expectations are zero unless the two highest subscripts are equal (Lemma A.2). By symmetry, we only need to consider three cases as follows.

Case 1) $s-m=s-n=t-k=t-l$, in which case the expectations are uniformly bounded by Assumption 1 and we find the contribution $cT^{-2}\sum_{t=1}^{T}(\sum_{n=0}^{\infty}||\xi_n^\dagger||^2)^2 \leq cT^{-1}\to 0$ using (D.39).

Case 2) $s-m=s-n>t-k\geq t-l$, where the contribution is

$$cT^{-2}\sum_{t,s=1}^{T}\sum_{n=1}^{s-1}\sum_{k,l=1}^{t-1}||\xi_n^\dagger||^2||\xi_k^\dagger||||\xi_l^\dagger|||\kappa_4(s-n,s-n,t-k,t-l)|$$
$$\leq cT^{-2}\sum_{s=1}^{T}\sum_{n=1}^{s-1}\sum_{k=1}^{T-1}||\xi_n^\dagger||^2||\xi_k^\dagger||\sum_{t=k+1}^{T}\sum_{l=1}^{t-1}|\kappa_4(s-n,s-n,t-k,t-l)|\leq cT^{-1}(\log T)\to 0$$

using Assumption 1(a)(iii) and (D.39).

Case 3) $s-m=t-k>s-n\geq t-l$, where we distinguish between the two subcases:

Case 3a) $s-n=t-l$ with the contribution

$$cT^{-2}\sum_{t,s=1}^{T}\sum_{m,n=\max(0,s-t)}^{s-1}||\xi_m^\dagger||||\xi_n^\dagger||||\xi_{t-s+m}^\dagger||||\xi_{t-s+n}^\dagger||\tau_{n-m,n-m}$$
$$\leq cT^{-2}\sum_{t,s=1}^{T}\sum_{m,n=\max(0,s-t)}^{s-1}||\xi_m^\dagger||||\xi_n^\dagger||||\xi_{t-s+m}^\dagger||||\xi_{t-s+n}^\dagger||\leq cT^{-1}(\log T)^3\to 0,$$

where we once again used Assumption 1(a)(ii) and (D.39).

Case 3b) $s-n>t-l$ with the contribution

$$cT^{-2}\sum_{t,s=1}^{T}\sum_{m,n=1}^{s-1}\sum_{l=1}^{t-1}||\xi_m^\dagger||||\xi_n^\dagger||||\xi_k^\dagger||||\xi_l^\dagger|||\kappa_4(s-m,s-m,s-n,t-l)|\leq cT^{-2}(\log T)^2\to 0$$

as in Case 2).

S.6.9 Proofs of Theorems 7 and 8

These results follow from Theorem 6 in the same way that Theorems 3 and 4 follow from Theorem 2.

Additional References

Dvoretzky, A. (1972), Asymptotic normality for sums of dependent random variables, in L.M. LeCam, J. Neyman, and E.L. Scott (eds.), *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability Vol. 2: Probability Theory*, University of California Press, Berkeley, 513–535.

McLeish, D.L. (1974), Dependent central limit theorems and invariance principles, *Annals of Statistics* 2, 620–628.