Queen's Economics Department Working Paper No. 1390

Adaptive inference in heteroskedastic fractional time series models

Giuseppe Cavaliere
University of Bologna

Morten ÃŸrregaard 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

5-2018

Adaptive Inference in Heteroskedastic Fractional Time Series Models*

Giuseppe Cavaliere[a], Morten Ørregaard Nielsen[b] and A.M. Robert Taylor[c,†]

[a] Department of Statistics, University of Bologna

[b] Department of Economics, Queen's University and CREATES

[c] Essex Business School, University of Essex

Abstract

We consider estimation and inference in fractionally integrated time series models driven by shocks which can display conditional and unconditional heteroskedasticity of unknown form. Although the standard conditional sum-of-squares (CSS) estimator remains consistent and asymptotically normal in such cases, unconditional heteroskedasticity inflates its variance matrix by a scalar quantity, $\lambda > 1$, thereby inducing a loss in efficiency relative to the unconditionally homoskedastic case, $\lambda = 1$. We propose an adaptive version of the CSS estimator, based on non-parametric kernel-based estimation of the unconditional variance process. This eliminates the factor λ from the variance matrix, thereby delivering the same asymptotic efficiency as that attained by the standard CSS estimator in the unconditionally homoskedastic case and, hence, asymptotic efficiency under Gaussianity. The asymptotic variance matrices of both the standard and adaptive CSS estimators depend on any conditional heteroskedasticity and/or weak parametric autocorrelation present in the shocks. Consequently, asymptotically pivotal inference can be achieved through the development of confidence regions or hypothesis tests using either heteroskedasticity robust standard errors and/or a wild bootstrap. Monte Carlo simulations and empirical applications are included to illustrate the practical usefulness of the methods proposed.

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

JEL classification: C12, C13, C22.

1 Introduction

Long memory models have proved highly effective in a wide range of fields of application including finance, economics, internet modeling, hydrology, climate studies, linguistics, opinion polling and DNA sequencing to name but a few; see, for example, the survey article by Samorodnitsky (2007) and the references therein. We contribute to the long memory literature in this paper by developing efficient methods of inference for the long and/or short memory parameters of univariate fractionally integrated time series models driven by shocks which display permanent changes in their unconditional volatility over time, often referred to as non-stationary volatility. We also allow for the presence of weak dependence (conditional heteroskedasticity and/or weak parametric autocorrelation) in the shocks.

Non-stationary volatility appears to be a relevant data phenomenon in a range of applied subject areas. For example, Sensier and van Dijk (2004) report that a large variety of both real and nominal economic and financial variables reject the null of constant unconditional variance. 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

*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.

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

observed after 2007; see, *inter alia*, McConnell and Perez-Quiros (2000), Stock and Watson (2012), and the references therein. In a recent paper in the climate change literature, Chang *et al.* (2016) analyse time series data on global temperature anomaly distributions and find that not only is the global mean temperature increasing, but that the variability of temperature anomalies around the mean anomaly is decreasing over time. In a number of fields, including climate data (for an example see the land and ocean climate change data set from 1850 to the present day available from the Berkeley Earth website `http://berkeleyearth.org/land-and-ocean-data/#section-0-7`) and opinion poll data (see, for example, Pickup and Johnston, 2008), improvements in data measurement over time have also effected reductions in the degree of volatility seen in the data.

Hualde and Robinson (2011) demonstrate the global consistency and asymptotic normality of the conditional quasi-maximum likelihood (QML) estimator — equivalently the conditional sum-of-squares (CSS) estimator — in parametric univariate fractional time series models for an arbitrarily large set of admissible values of the long memory parameter (see also Nielsen, 2015). They do so in the context of a fractional model driven by conditionally homoskedastic innovations. The estimator is asymptotically efficient when the innovations are Gaussian. Their results solve a long-standing problem arising from the non-uniform convergence of the objective function when the range of values the long memory parameter may take is large. In a recent paper, Cavaliere, Nielsen and Taylor (2017, henceforth CNT) demonstrate that the QML estimator retains its global consistency and asymptotic normality properties in cases where the innovations display non-stationary volatility and/or conditional heteroskedasticity. CNT show, however, that (other things equal) non-stationary volatility inflates the limiting covariance matrix of the QML estimator by a scalar quantity $\lambda > 1$ relative to the unconditionally homoskedastic case, thereby implying a loss in asymptotic efficiency. CNT show that, as a result, standard hypothesis tests based on the QML estimate lose asymptotic efficiency relative to the unconditionally homoskedastic case when non-stationary volatility is present.

In light of the findings of CNT, the contribution of this paper is to develop two-step estimation and inference methods for fractional time series models which non-parametrically adapt to unconditional heteroskedasticity of unknown form, thereby recovering the efficiency losses that occur with standard methods. In the first step the unconditional variance process is estimated using a kernel-based non-parametric regression on the squares of the residuals which obtain on fitting the model using the standard QML estimator. In the second step, the sum of squares criterion which is minimised to deliver the standard QML estimator is scaled by the estimated volatility process, and is then subsequently minimised. We will refer to this estimator as an adaptive CSS [ACSS] estimator. Adaptive inference based on an estimate of the unconditional variance was proposed in the context of inference on the parameters of finite-order unconditionally heteroskedastic but conditionally homoskedastic autoregressive models by Xu and Phillips (2008). Harris and Kew (2017) extend the score-based tests developed in Robinson (1994) for homoskedastic errors, to develop tests on the value of the long memory parameter in the context of an unconditionally heteroskedastic but conditionally homoskedastic ARFIMA model, using an adaptive estimate of the unconditional volatility process. More generally, the use adaptive estimators and tests designed to account for non-parametric heteroskedasticity in time series models have been widely used in the literature; see among others, Carroll (1982), Robinson (1987), Harvey and Robinson (1988), Hansen (1995), Xu and Phillips (2011) and Xu and Yang (2015).

Under suitable conditions, we demonstrate that our proposed ACSS estimator is asymptotically equivalent to an infeasible estimator obtained by minimising the sum of squares criterion when divided by the true (unknown) volatility process. As a consequence, the estimator is asymptotically efficient under Gaussianity. We further demonstrate the global consistency of the adaptive estimator and show that it attains the same limiting distribution as that which would be attained by the standard QML estimator under unconditional homoskedasticity, other things being equal. A key consequence of this result is that while there is no loss of asymptotic efficiency from using the ACSS estimator rather than the standard QML estimator in the conditionally homoskedastic case, efficiency gains relative to the QML estimator will be obtained where the innovations display non-stationary volatility (with a limiting relative efficiency given by λ). Although the limiting distribution of the ACSS estimator does not depend on any non-stationary volatility present in the shocks, our results show that it does, like the standard QML estimator, depend on any conditional heteroskedasticity and/or weak dependence present. Interval estimation for (functions of) the long run and/or short run parameters of the fractional time series model based on the ACSS estimator must therefore either be based on the use of heteroskedasticity robust standard errors or by using an asymptotically valid bootstrap method. We investigate both approaches, and for the latter we propose a wild bootstrap implementation of the ACSS estimator and show that this is asymptotically valid. We also develop associated heteroskedasticity-robust t and Wald statistics for testing hypotheses on these parameters. We demonstrate that these have standard limiting null distributions under our assumptions and may also be validly bootstrapped. The finite sample performance of our proposed methods are explored using Monte Carlo simulation.

In an important related literature, Baillie, Chung, and Tieslau (1996), Ling and Li (1997), Li, Ling, and McAleer (2002) and Ling (2003), among others, consider efficient maximum likelihood estimation of an ARFIMA model in the presence of parametric GARCH models under Gaussianity, in each case assuming unconditional homoskedasticity. The aim of these authors is different from ours here which is to develop adaptive methods of inference valid under general forms of both conditional and unconditional heteroskedasticity without requiring the practitioner to specify a parametric model for either form of heteroskedasticity. As our results show, adaptive methods can deliver asymptotically efficient inference under unconditional heteroskedasticity. To develop efficient inference in the presence of conditional heteroskedasticity, a parametric model would have to be fitted, similarly to the approach adopted in the papers cited above. This will only be asymptotically efficient if the correct parametric model for the conditional heteroskedasticity is chosen and has the potential to behave very poorly where a misspecified model is chosen.

The remainder of the paper is structured as follows. Section 2 outlines our reference heteroskedastic fractional time series model and our main assumptions. Section 3 outlines the properties of the QML estimator under our set-up, details our ACSS estimator and establishes its large sample properties. Section 4 develops methods of inference based on confidence intervals, heteroskedasticity-robust t and Wald statistics, and wild bootstrap implementations thereof. Monte Carlo simulation results are given in Section 5. A variety of data examples are reported in Section 6. Section 7 concludes. Proofs of Theorems 1, 2 and 3 are provided in Appendices A, B and C, respectively. Additional supporting proof material together with some additional data analysis are reported in a supplementary appendix.

Notation. As a convention, it is assumed that $j^{-1} = 0$ for $j = 0$ in summations over j. We use c or K to denote a generic, finite constant, $||\cdot||$ to denote the Euclidean norm and $||\cdot||_r$ to denote the L_r-norm. 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 $x_1, x_2 \in \mathbb{R}^q$. 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 $T \to \infty$, T denoting the sample size. The probability and expectation conditional on the realisation of the original sample are denoted P^* and E^*, respectively. For a given sequence X^*_T computed on the bootstrap data, $X^*_T \xrightarrow{p^*}_p 0$ or $X^*_T = o^*_p(1)$, in probability, denote that $P^*(|X^*_T| > \epsilon) \to 0$ in probability for any $\epsilon > 0$, $X^*_T = O^*_p(1)$, in probability, 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, in each case as $T \to \infty$.

2 The Heteroskedastic Fractional Model and Assumptions

We consider the fractional time series model[1]

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

[1]The model in (1) uses the so-called "type II" fractional integration. This has 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 d. Consequently, stationary, non-stationary, and over-differenced time series are all permitted because the range of admissible values of the fractional parameter can be arbitrarily large; see Hualde and Robinson (2011).

where L is the usual lag operator, and 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 $\pi_n(d) := \frac{\Gamma(d+n)}{\Gamma(d)\Gamma(1+n)} = \frac{d(d+1)...(d+n-1)}{n!}$ 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. The parametric form of the function $a(z,\psi)$ will be assumed known, so that, specifically, u_t is assumed to be a linear process governed by an underlying unknown p-dimensional parameter vector, ψ. For example, any process that can be written as a finite order ARMA model is permitted, as is the exponential spectrum model of Bloomfield (1973). Further discussion on the function $a(z,\psi)$ can be found in Hualde and Robinson (2011). Thus, our focus is model-based inference (which might be on the long memory parameter, d, or 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.

We now outline the assumptions that we will place on the model in (1). It is important to note that none of the assumptions which follow impose Gaussianity on (1).

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 with respect to the natural filtration $\mathcal{F}_t$, the sigma-field generated by $\{z_s\}_{s\leq t}$, 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 $r \geq 0, s \geq 0$,*
- *(iii) 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 $\sup_t \sum_{r_1,...,r_{q-2}=1}^{\infty} |\kappa_q(t, t, t-r_1, \ldots, t-r_{q-2})| < \infty$.*

(b) $\{\sigma_t\}_{t\in\mathbb{Z}}$ is a non-stochastic sequence satisfying

- *(i) $\sup_{t\in\mathbb{Z}} \sigma_t < \infty$,*
- *(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 $[0,1]$, satisfies $\inf_{0\leq u\leq 1} \sigma(u) > 0$.*

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 3. *It holds that:*

- *(i) For all z in the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$ and for all $\psi \in \Psi$, $a(z,\psi)$ is bounded and bounded away from zero and $a_0(\psi) = 1$.*
- *(ii) For all $\psi \in \Psi$, $a(e^{i\omega}, \psi)$ is twice differentiable in ω with second derivative in $\mathrm{Lip}(\zeta)$ for $\zeta > 0$.*
- *(iii) For all λ, $a(e^{i\omega}, \psi)$ is differentiable in ψ on $\psi \in \Psi$, and for all $\psi \in \Psi$, $\dot{a}(e^{i\omega}, \psi) := \frac{\partial a(e^{i\omega},\psi)}{\partial \psi}$ is twice differentiable in ω with derivative in $\mathrm{Lip}(\zeta)$ for $\zeta > 0$.*
- *(iv) For all ω, $a(e^{i\omega}, \psi)$ 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 for all $\psi \in \mathcal{N}_\delta(\psi_0)$, the second and third derivatives of $a(e^{i\omega}, \psi)$ with respect to ψ are themselves twice differentiable in ω with derivative in $\mathrm{Lip}(\zeta)$ for $\zeta > 0$.*

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.*

Remark 2.1. Under Assumption 1(a) volatility clustering, such as generalised autoregressive conditional heteroskedasticity (GARCH), is permitted by the fact that the quantity $\tau_{r,r}$ is not necessarily equal to $E(z_t^2)E(z_{t-r}^2) = 1$. Asymmetric volatility clustering is allowed for by non-zero $\tau_{r,s}$ for $r \neq s$. Statistical leverage is also permitted, which occurs when the quantity $E(z_t^2 z_{t-i})$ is non-zero for some $i \geq 1$, 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. The stated conditions, including the summability condition on the eighth-order cumulants of ε_t, are typical, but rather weaker than, those used in the fractional literature; see, for example, Robinson (1991), Hassler, Rodrigues and Rubia (2009), and Harris and Kew (2017). Notably, these authors impose further conditions which rule out, among other things, statistical leverage and asymmetric volatility clustering. Harris and Kew (2017) additionally impose conditional homoskedasticity on z_t. $\diamondsuit$

Remark 2.2. Assumption 1 implies that the time-varying scale factor σ_t^2 corresponds to the unconditional variance of ε_t. Thus, both the conditional and the unconditional variance of ε_t are allowed to display time-varying behaviour under Assumption 1. $\diamondsuit$

Remark 2.3. Assumption 1(b) imposes relatively mild conditions on the sequence $\{\sigma_t\}$. In particular, the càdlàg assumption on $\sigma(\cdot)$ appears much weaker than those usually applied in the literature. For example, Assumption S of Harris and Kew (2017) and Assumption (i) of Xu and Phillips (2008, p. 267) require $\sigma(\cdot)$ to satisfy a uniform first-order Lipschitz condition with at most a finite number of discontinuities. In contrast, our assumption allows a countable number of jumps, which admits an extremely wide class of potential models for the unconditional variance of ε_t.[2] Models of single or multiple variance shifts satisfy part (b) of Assumption 1 with $\sigma(\cdot)$ piecewise constant; for example, a one-time break in variance from σ_0^2 to σ_1^2 at time $\lfloor \tau T \rfloor$, $0 < \tau < 1$, corresponds to $\sigma(u) := \sigma_0 + (\sigma_1 - \sigma_0)\mathbb{I}(u > \tau)$. (Piecewise) affine functions are also permitted. $\diamondsuit$

Remark 2.4. Zhang and Wu (2012) consider shocks which are 'locally stationary'; that is, of the form $u_t = G(t/T; F_t)$, where $F_t = (..., \varepsilon_{t-1}, \varepsilon_t)$ is shift process (see, for example, Rosenblatt, 1959) of i.i.d. random variables $\{\varepsilon_t\}$ and G is a sufficiently smooth (Lipschitz) measurable function. It is not difficult to see that our u_t process in (1) satisfies local stationarity under Assumption 1, up to an $O_p(T^{-1})$ term, provided $\sigma(\cdot)$ is Lipschitz measurable. The main difference between the 'locally stationary' set-up and ours is that the former places smoothness (and moment) restrictions on the function G, whereas we impose a linear dependence structure on u_t through the function $a(z, \psi)$, leaving the scaling function $\sigma(\cdot)$ essentially unrestricted. $\diamondsuit$

Remark 2.5. Assumption 2 permits the length of the interval of admissible values of the parameter d to be arbitrarily large such that the model in (1) is sufficiently general to simultaneously accommodate both non-stationary, (asymptotically) stationary, and over-differenced processes. $\diamondsuit$

Remark 2.6. Assumption 3 relates to the coefficients of the linear filter $a(z, \psi)$ and is easily satisfied, for example, by stationary and invertible finite order ARMA processes. In particular, 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 and $b_0(\psi) = 1$. Under Assumption 3 the coefficients $a_n(\psi)$, $b_n(\psi)$, $\dot{a}_n(\psi) := \partial a_n(\psi)/\partial\psi$, and $\dot{b}_n(\psi) := \partial b_n(\psi)/\partial\psi$ satisfy

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

uniformly in $\psi \in \Psi$; see Zygmund (2003, pp. 46 and 71). The second and third derivatives with respect to ψ satisfy the same bounds uniformly over the neighborhood $\mathcal{N}_\delta(\psi_0)$. $\diamondsuit$

Remark 2.7. Assumption 3(i) coincides with Assumption A1(iv) of Hualde and Robinson (2011), while Assumption 3(ii) strengthens their Assumption A1(ii) from once differentiable in ω with derivative in Lip(ζ) for $\zeta > 1/2$, and Assumption 3(iii) strengthens their Assumption A1(iii) from

[2] For $t \leq 0$, σ_t is assumed only to be uniformly bounded, see Assumption 1(b)(i), whereas, for $t = 1, \ldots, T$, Assumption 1(b)(ii) entails 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 $T \geq 1$. The array notation is not essential and so for simplicity the subscript T is suppressed in what follows.

continuity in ψ to differentiability. Assumption 3(iv) requires $a(z,\psi)$ to be thrice differentiable in ψ rather than the corresponding twice differentiable condition in Assumption A3(ii) of Hualde and Robinson (2011) with associated Lipschitz conditions in ω. The latter are used to obtain the bounds in (2), and also appear to be needed to obtain the corresponding bounds in Hualde and Robinson (2011, p. 3169). $\diamondsuit$

Remark 2.8. 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 then be chosen as any compact and convex subset of the (open) set for which the roots of the AR and MA polynomials are strictly outside 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 by the user as any compact and convex subset of $(-1,1)$. $\diamondsuit$

Remark 2.9. The identification condition in Assumption 4 is identical to Assumption A1(i) in Hualde and Robinson (2011) and is satisfied, for example, by all stationary and invertible finite order ARMA processes whose AR and MA polynomials do not admit any common factors. $\diamondsuit$

To conclude this section we need to set up some additional notation and an associated final assumption that will be required in the next section when stating the large sample properties of the QML estimator and of our proposed adaptive estimator. To that end, 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 $\gamma_n(\psi) := \sum_{m=0}^{n-1} a_m(\psi)\dot{b}_{n-m}(\psi)$. Observe that A_0 (and hence B_0) is finite because $\sum_{n=0}^{\infty} ||\gamma_n(\psi)|| < \infty$ under Assumption 3 and $\sum_{n,m=1}^{\infty} |\tau_{n,m}| < \infty$ by Assumption 1(a)(iii). The matrix B_0 coincides with the matrix A in Hualde and Robinson (2011) and derives from the autocorrelation present in the process through $a(z,\psi)$. The matrix A_0 also includes the effects of any conditional heteroskedasticity present in ε_t. 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 are affected by any unconditional heteroskedasticity arising from Assumption 1(b). As in Hualde and Robinson (2011), in order to state the limiting distribution of the QML and adaptive estimators we will require B_0 to be invertible, as formally stated in Assumption 5.

Assumption 5. *The matrix B_0 is non-singular.*

This condition is satisfied by, for example, stationary and invertible ARMA processes.

3 Adaptive Estimation

Adaptive estimation requires a preliminary consistent estimator. For this purpose, in Section 3.1 we review the standard QML estimator analyzed in CNT. Then, in Section 3.2, we detail our adaptive estimator and its large sample properties.

3.1 Standard QML Estimation

Define the residuals

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

Then the conditional[3] Gaussian QML estimator of θ is identical to the classical least squares or CSS estimator, which is found by minimizing the sum of squared residuals; that is,

$$\tilde{\theta} := \arg\min_{\theta\in\Theta} Q_T(\theta), \quad Q_T(\theta) := T^{-1}\sum_{t=1}^{T} \varepsilon_t(\theta)^2. \tag{4}$$

[3] We use the term ‘conditional’ here in its usual sense to indicate that we have conditioned on the initial values of u_t. This has been done implicitly through the assumption that (1) generates a type II fractional process.

CNT show that if X_t is generated according to (1) under Assumptions 1–5, then it holds that

$$\sqrt{T}(\tilde{\theta}-\theta_0) \xrightarrow{w} N(0,\lambda C_0), \tag{5}$$

where $C_0 := B_0^{-1}A_0B_0^{-1}$ and $\lambda := \int_0^1 \sigma^4(s)\mathsf{d}s/(\int_0^1 \sigma^2(s)\mathsf{d}s)^2$.

Remark 3.1. A fairly standard conditionally homoskedastic alternative to Assumption 1 (see, for example, Hannan, 1973, and Hualde and Robinson, 2011) is one where the innovations $\{\varepsilon_t\}$ are assumed to form a conditionally homoskedastic martingale difference sequence with respect to the filtration $\mathcal{F}_t$, i.e., where $E\left(\varepsilon_t^2|\mathcal{F}_{t-1}\right)=\sigma^2$ almost surely and $\sup_t E(|\varepsilon_t|^q)\leq K<\infty$ for some $q\geq 4$. Under these conditions, $A_0=B_0$ and $\lambda=1$ and, hence, the result in (5) reduces to the result in Theorem 2.2 of Hualde and Robinson (2011). In the case where $\lambda=1$ and z_t in Assumption 1(a) is Gaussian (and hence i.i.d.), the QML estimator $\tilde{\theta}$ of (4) is asymptotically efficient. $\diamondsuit$

Remark 3.2. Where heteroskedasticity arises only through part (a) of Assumption 1 then so the $\lambda=1$ and the variance matrix in (5) reduces to C_0. On the other hand, where heteroskedasticity arises only through part (b) of Assumption 1 then so C_0 reduces to B_0^{-1}. $\diamondsuit$

Remark 3.3. As (5) shows, the variance of the asymptotic distribution of the standard QML estimator, $\tilde{\theta}$, depends on the scalar parameter λ. This parameter is a measure of the degree of unconditional heteroskedasticity (non-stationary volatility) present in $\{\varepsilon_t\}$. For an unconditionally homoskedastic process, where $\sigma(\cdot)$ is constant, $\lambda=1$, whereas when $\sigma(\cdot)$ is non-constant, $\lambda>1$ by the Cauchy-Schwarz inequality. Consequently, other things being equal, the variance of the asymptotic distribution of the QML estimator is seen to be inflated when unconditional heteroskedasticity is present in $\{\varepsilon_t\}$, *via-à-vis* the unconditionally homoskedastic case. $\diamondsuit$

Suppose for the present that $\{\sigma_t^2\}$ was known. In such circumstances, a (infeasible) weighted CSS estimate of θ could be formed as

$$\bar{\theta} := \arg\min_{\theta\in\Theta} \bar{Q}_T(\theta), \text{ where } \bar{Q}_T(\theta) := T^{-1}\sum_{t=1}^{T}\left(\frac{\varepsilon_t(\theta)}{\sigma_t}\right)^2. \tag{6}$$

In Theorem 1 below it is shown that if X_t is generated by (1) under Assumptions 1–5, then $\sqrt{T}(\bar{\theta}-\theta_0) \xrightarrow{w} N(0,C_0)$. The asymptotic variance matrix of the QML estimator, $\tilde{\theta}$, therefore differs from that of the infeasible weighted CSS estimator, $\bar{\theta}$, by the factor λ and since C_0 is invariant to the function $\sigma(\cdot)$ the inefficiency of the standard QML estimator relative to the weighted CSS estimator is solely determined by this factor in large samples. Where this factor is large, QML will be highly inefficient relative to the weighted CSS estimator, whereas if it is close to unity QML will lose little in efficiency relative to the weighted CSS estimator and would be close to being asymptotically optimal under Gaussianity.

3.2 Adaptive QML Estimation

The weighted CSS estimator $\bar{\theta}$ in (6) is infeasible, because the true values of σ_t^2 are unknown in practice. However, it is possible to implement a feasible version of $\bar{\theta}$ that has the same asymptotic distribution as $\bar{\theta}$ following the approach used in, e.g., Xu and Phillips (2008, p. 271) using a kernel-based nonparametric estimate of σ_t^2. To that end, we first need a preliminary estimator of θ which, although based on an assumption of homoskedasticity, is nonetheless (root-T) consistent under heteroskedasticity. The standard QML estimator, $\tilde{\theta}$ of (4), satisfies this requirement, see (5).

Defining $\tilde{\varepsilon}_t := \varepsilon_t(\tilde{\theta})$ as the standard QML residuals, our proposed feasible CSS estimator of θ is then defined as

$$\hat{\theta} := \arg\min_{\theta\in\Theta} \hat{Q}_T(\theta), \;\; \hat{Q}_T(\theta) := T^{-1}\sum_{t=1}^{T}\frac{\varepsilon_t(\theta)^2}{\hat{\sigma}_t^2}. \tag{7}$$

where

$$\hat{\sigma}_t^2 := \sum_{i=1}^{T} k_{ti}\tilde{\varepsilon}_i^2, \tag{8}$$

and where

$$k_{ti} := \frac{K(\frac{t-i}{Tb})}{\sum_{i=1}^{T} K(\frac{t-i}{Tb})}. \tag{9}$$

In (9), $K(\cdot)$ is a bounded, nonnegative, continuous kernel function and $b := b(T)$ is a bandwidth parameter which depends on the sample size, T. In what follows we will refer to $\hat{\theta}$ as the adaptive CSS [ACSS] estimator of θ; this in the sense that it is a feasible version of $\bar{\theta}$ based on adaptive estimation of σ_t^2.

To establish the large sample properties of the adaptive ACSS estimator $\hat{\theta}$ in (7), we impose some conditions on both the kernel function, $K(\cdot)$, and on the bandwidth, b. We formally state these in Assumptions 6 and 7, respectively.

Assumption 6. *The kernel function* $K(\cdot) : [-\infty, \infty] \to \mathbb{R}^+ \cup \{0\}$ *is continuous and satisfies* $\sup_{-\infty \le x \le \infty} K(x) < \infty$ *and* $\int_{-\infty}^{\infty} K(u)\mathrm{d}u \in (0, \infty)$.

Assumption 7. *The bandwidth* $b := b(T)$ *satisfies* $b + T^{-1}b^{-2} \to 0$.

We now detail the asymptotic distribution of the ACSS estimator, $\hat{\theta}$ of (7).

Theorem 1. *Let* X_t *be generated according to* (1) *and let Assumptions 1–7 be satisfied. Then,*

$$\sqrt{T}(\bar{\theta} - \hat{\theta}) \xrightarrow{p} 0, \tag{10}$$

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{w} N(0, C_0). \tag{11}$$

Remark 3.4. Notice that Theorem 1 holds without the need to strengthen the càdlàg condition in part (b) of Assumption 1. In contrast, additional smoothness conditions on $\sigma(\cdot)$ are routinely needed in the adaptive inference literature; see Remark 2.3. This generalization with respect to the extant literature is made possible because our method of proof for Theorem 1 is based on showing that $T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \sigma_t^2)^2 \xrightarrow{p} 0$ (see Lemma A.1), rather than showing $\sup_t |\hat{\sigma}_t^2 - \sigma_t^2| \xrightarrow{p} 0$ as is usually done. Of course, under additional smoothness conditions on the kernel function and on $\sigma(\cdot)$, the former convergence implies the latter. ♢

Remark 3.5. Implementation of $\hat{\sigma}_t^2$ depends on the choice of kernel function, $K(\cdot)$, and the bandwidth, b. Commonly used kernels which satisfy Assumption 6 include the uniform, Epanechnikov, biweight and Gaussian functions. The bandwidth condition in Assumption 7 implies that $b \to 0$ but at a slower rate than $T^{-1/2}$. In practice bandwidth selection is crucial to performance and here the data-driven method of Wong (1983), which uses cross-validation on the average squared error, could be employed. This cross-validatory choice of b is the value b^* which minimises $\widehat{CV}(b) := T^{-1}\sum_{t=1}^{T}(\tilde{\varepsilon}_t^2 - \hat{\sigma}_t^2)^2$, but where a leave-one-out procedure is used in (8) such that the observation $\tilde{\varepsilon}_t^2$ is omitted which is done by defining $K(\frac{t-i}{Tb}) := 0$ for $t = i$ in (9). Using a leave-one-out procedure does not impact on any of the large-sample results we provide. ♢

Remark 3.6. A comparison of the result in Theorem 1 with that given at the end of Section 3.1 for $\bar{\theta}$ of (6) shows that the asymptotic distribution of the ACSS estimator coincides with that of the infeasible weighted CSS estimator. The limiting distribution of the ACSS estimator is therefore seen not to depend on the non-stationary volatility process, $\sigma(\cdot)$. ♢

Remark 3.7. Noting that $\lambda \ge 1$ by the Cauchy-Schwarz inequality, a comparison of (5) and (11) shows that there can never be a loss of asymptotic efficiency from using the ACSS estimator rather than the QML estimator (the relative efficiency being given by λ) even in the case where the shocks are unconditionally homoskedastic. Thus, the ACSS estimator is asymptotically more efficient than the QML estimator. ♢

Remark 3.8. Where z_t in Assumption 1(a) is conditionally homoskedastic but ε_t is unconditionally heteroskedastic the ACSS estimator recovers the asymptotic distribution attained by the QML estimator in the purely homoskedastic case. ♢

Remark 3.9. Consider the case where z_t in Assumption 1(a) is Gaussian. Then the adaptive estimator $\hat{\theta}$ is asymptotically efficient in the sense that it has the same asymptotic variance, $C_0 = B_0^{-1}$, as obtains under homoskedasticity for the standard QML estimator $\tilde{\theta}$ in (4). As a consequence, the standard QML estimator is not efficient, even when z_t is Gaussian, except in the special case where $\lambda = 1$, due to the factor $\lambda \geq 1$ appearing in its asymptotic variance. $\diamondsuit$

4 Adaptive Inference

For inference purposes, a consistent estimator of C_0 is required. In Section 4.1 we discuss such an estimator and show that it can be used to obtain asymptotically valid confidence regions for (functions of) θ. Section 4.2 discusses asymptotically pivotal hypothesis tests based on this estimator, and details the asymptotic power functions of these tests under Pitman drift. Bootstrap implementations of these methods are explored in Section 4.3.

4.1 Confidence Regions

To construct adaptive confidence regions for (functions of) the elements of θ based on the limiting result in (11) we will require a consistent estimate of C_0. Following Eicker (1967), Huber (1967), and White (1982), we consider the familiar sandwich-type estimator of C_0,

$$\hat{C} := \left. \left(\left(\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} \frac{\partial q_t(\theta)}{\partial\theta} \frac{\partial q_t(\theta)}{\partial\theta'} \right) \left(\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \right) \right|_{\theta=\hat{\theta}} \tag{12}$$

with $q_t(\theta) := \varepsilon_t(\theta)^2/\hat{\sigma}_t^2$. The consistency of $\hat{C}$ for C_0 is formalised in Theorem 2.

Theorem 2. *Let the conditions of Theorem 1 hold. It then follows that* $\hat{C} \xrightarrow{p} C_0$.

Theorem 2, taken together with the result in (11), implies that we can construct asymptotically pivotal feasible adaptive confidence regions for the elements of θ, using $\hat{C}$, in the usual way. Let $f : R^{p+1} \to R^q$ be a (possibly non-linear) function which is continuously differentiable at θ_0 and let $\xi := f(\theta)$ denote the parameter of interest. The adaptive estimator of ξ is given by $\hat{\xi} := f(\hat{\theta})$. Given previous results, the asymptotic distribution of $\hat{\xi}$ can straightforwardly be shown, using the delta method, to be $\sqrt{T}(\hat{\xi} - \xi_0) \xrightarrow{w} N(0, F(\theta_0)C_0F(\theta_0)')$, where $F(\theta) := \frac{\partial}{\partial\theta}f(\theta)$ is the Jacobian of the function $f(\theta)$. Confidence regions for ξ can then be formed using a consistent estimator of $F(\theta_0)C_0F(\theta_0)'$, an obvious candidate for which is $F(\hat{\theta})\hat{C}F(\hat{\theta})'$.

4.2 Hypothesis Testing and Local Power Considerations

To complement the material on feasible adaptive confidence regions, we now discuss adaptive tests of general hypotheses on the elements of θ. To that end, suppose we wish to test the null hypothesis,

$$H_0 : \xi := f(\theta) = 0, \tag{13}$$

where $f(\theta)$ is as defined below Theorem 2, and consider the sequence of local (Pitman) alternatives,

$$H_{1,T} : \xi_T = \delta/\sqrt{T}, \tag{14}$$

where δ is a fixed q-vector. In the case where we wish to test for linear restrictions on θ, we would set $f(\theta) = M'\theta - m$ where 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-vector of constants. An obvious example involves testing hypotheses on the long memory parameter d, important cases thereof are $d = 0$ (short memory), $d = 0.5$ (the series being weakly stationary, in the absence of unconditional heteroskedasticity, if $d < 0.5$) 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 (possibly heteroskedastic) random walk hypothesis, while $d = 0 \cap \psi = 0$ yields a martingale difference sequence.

The null hypothesis in (13) can be tested using the familiar Wald statistic,

$$W_T := Tf(\hat{\theta})'(F(\hat{\theta})\hat{C}F(\hat{\theta})')^{-1}f(\hat{\theta}), \tag{15}$$

rejecting H_0 in favour of $H_1 : \xi \neq 0$ for large values of W_T. Where only a single restriction is being tested, so that $q = 1$, one can also use the t-type statistic

$$t_T := \frac{\sqrt{T}f(\hat{\theta})}{\sqrt{F(\hat{\theta})\hat{C}F(\hat{\theta})'}} \tag{16}$$

with H_0 rejected in favour of H_1 for large absolute values of this statistic. The statistic in (16) can also be used to test H_0 against one-sided alternatives of the form $H_{1,L} : \xi < 0$ and $H_{1,U} : \xi > 0$ by rejecting for large negative and large positive values of t_T, respectively. A familiar special case of t_T in (16) which obtains for testing the simple null hypothesis that the i'th element of θ is equal to some hypothesised value m_i, $H_{0,i} : \theta_i - m_i = 0$ say, is given by $t_{i,T} := T^{1/2}(\hat{\theta}_i - m_i)/(\hat{C}_{ii})^{1/2}$, which again can be performed as either a one-sided or two-sided test.

The asymptotic distributions of W_T and t_T under H_0 follow immediately from Theorems 1 and 2, which we state as a corollary.

Corollary 1. *Let the conditions of Theorem 1 hold. Then, under H_0 of (13) and provided $F(\theta_0)$ is of full row rank, $W_T \xrightarrow{w} \chi^2(q)$ and $t_T \xrightarrow{w} N(0, 1)$.*

As an obvious consequence of Corollary 1, critical regions for the tests are found from standard tables, and hence the tests are easily implemented in practice.

We proceed to discuss asymptotic local power and optimality of the tests under the assumption that z_t is Gaussian; c.f. Remark 3.9 relating to efficiency of the estimator. The following corollary is implied by Theorems 1 and 2 and Le Cam's Third Lemma.

Corollary 2. *Let the conditions of Theorem 1 hold and assume also that z_t is Gaussian. Then, under $H_{1,T}$ of (14) and provided $F(\theta_0)$ is of full row rank, it holds that a one-sided test based on the t_T statistic is asymptotically uniformly most powerful (UMP) while a two-sided test based on t_T will be an asymptotically UMP unbiased. Specifically, $W_T \xrightarrow{w} \chi^2_q\left(\delta'(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1}\delta\right)$ and $t_T \xrightarrow{w} N(\delta(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1/2}, 1)$, where $\chi^2_q(g)$ indicates a noncentral χ^2_q distribution with non-centrality parameter g.*

Remark 4.1. As with Remark 3.9, Corollary 2 imposes Gaussianity, and hence conditional homoskedasticity, on z_t, so that $A_0 = B_0$ and $C_0 = B_0^{-1}$. Consequently, the limiting distributions given in Corollary 2 for the W_T and t_T statistics, and as a result the asymptotic local power functions of the tests based on these statistics, coincide with those which would obtain for the corresponding statistics in the homoskedastic Gaussian case, and the optimality statements follow. That is, even in the presence of heteroskedasticity of the form in Assumption 1(b), and regardless of the value of λ, the tests based on W_T and t_T achieve the same asymptotic local power as in the homoskedastic Gaussian case. Examples of these asymptotic local power functions, showing the impact of λ, are graphed in Cavaliere *et al.* (2015). ◇

Remark 4.2. Asymptotically pivotal test statistics can also be constructed based on the QML estimator, $\tilde{\theta}$, using the sandwich estimator $\tilde{C}$ defined as in (12), but with $Q_T(\theta)$ given by (4) and $q_t(\theta) = \varepsilon_t(\theta)^2$ and evaluated at $\theta = \tilde{\theta}$. It is shown in Theorem 2 of CNT that if X_t is generated according to (1) under Assumptions 1–5 then $\tilde{C} - \lambda C_0 \xrightarrow{p} 0$. Consequently, defining $\tilde{W}_T$ and $\tilde{t}_T$ as in (15) and (16) but now based on $\tilde{\theta}$ and $\tilde{C}$, it follows that, under the additional assumption that z_t is Gaussian, $\tilde{W}_T \xrightarrow{w} \chi^2_q\left(\lambda^{-1}\delta'(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1}\delta\right)$ and $\tilde{t}_T \xrightarrow{w} N(\lambda^{-1/2}\delta(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1/2}, 1)$. It is seen from a comparison with the results in Corollary 2 that the noncentrality parameters of tests based on our ACSS estimator are larger than those based on the QML estimator by a factor of $\lambda \geq 1$ (for W_T) or $\lambda^{1/2} \geq 1$ (for t_T) and so asymptotic local power is correspondingly higher. ◇

Remark 4.3. Although possessing standard limiting null distributions, cf Corollary 1, it is seen from the results in Corollary 2 that tests based on W_T and t_T will have asymptotic local power functions that depend on any weak dependence present in u_t. $\diamondsuit$

4.3 Bootstrap Methods

As an alternative to the asymptotic approach to forming confidence regions for θ outlined in Section 4.1, we now consider bootstrap-based methods of constructing confidence regions for θ. We will subsequently also explore bootstrap implementations of the robust Wald and t tests from Section 4.2. To that end, we first outline our proposed bootstrap algorithm. Because we allow for the presence of conditional heteroskedasticity under Assumption 1, we use a wild bootstrap-based approach (Wu, 1986). Specifically, with $\hat{\varepsilon}_t := \varepsilon_t(\hat{\theta})$ denoting the residuals based on the ACSS estimate, $\hat{\theta}$, we construct the bootstrap innovations $\varepsilon_t^* := \hat{\varepsilon}_t w_t$, where w_t, $t = 1, ..., T$, is an i.i.d. sequence with $E(w_t) = 0$, $E(w_t^2) = 1$ and $E(w_t^{16}) < \infty$, setting $\varepsilon_t^* = 0$ for $t \leq 0$. Then the bootstrap sample $\{X_t^*\}$ is generated from the recursion

$$X_t^* := \Delta_+^{-\hat{d}} u_t^* \text{ with } u_t^* := a(L, \hat{\psi})\varepsilon_t^*, \quad t = 1, ..., T, \tag{17}$$

and the bootstrap ACSS estimator is given by

$$\hat{\theta}^* := \arg\min_{\theta \in \Theta} \hat{Q}_T^*(\theta), \quad \hat{Q}_T^*(\theta) := T^{-1} \sum_{t=1}^{T} \frac{\varepsilon_t^*(\theta)^2}{\hat{\sigma}_t^{*2}}, \tag{18}$$

where

$$\varepsilon_t^*(\theta) := \sum_{n=0}^{t-1} b_n(\psi) \Delta_+^d X_{t-n}^* \tag{19}$$

and $\hat{\sigma}_t^{*2}$ is defined as the estimator (8) computed from the residuals $\tilde{\varepsilon}_t^* := \varepsilon_t^*(\tilde{\theta}^*)$ with $\tilde{\theta}^*$ denoting the preliminary standard QML estimator computed on the bootstrap sample.

Remark 4.4. The assumption that $E(w_t^{16}) < \infty$ is not restrictive in practice as it is satisfied by all common choices of the distribution of w_t, e.g., Gaussian, Rademacher, and other two-point distributions (Mammen, 1993, or Liu, 1988). $\diamondsuit$

Remark 4.5. Notice that $\hat{\theta}^*$ employs an unrestricted estimate of θ in constructing the bootstrap data. Because the bootstrap data generating process is then based on $\hat{\theta}$, it is the distribution of $\sqrt{T}(\hat{\theta}^* - \hat{\theta})$, conditional on the original data, that will be used to approximate that of $\sqrt{T}(\hat{\theta} - \theta_0)$. As is standard, the former can be approximated numerically to any desired degree of accuracy. $\diamondsuit$

Remark 4.6. An alternative to $\hat{\theta}^*$ is to calculate the following bootstrap estimator,

$$\check{\theta}^* := \arg\min_{\theta \in \Theta} \check{Q}_T^*(\theta), \quad \check{Q}_T^*(\theta) := T^{-1} \sum_{t=1}^{T} \frac{\varepsilon_t^*(\theta)^2}{\hat{\sigma}_t^2}, \tag{20}$$

where $\hat{\sigma}_t^2$ is the kernel-based estimator of σ_t^2 computed on the original data. In contrast, $\hat{\theta}^*$ is based on the bootstrap analogue of the adaptive estimator, $\hat{\sigma}_t^{*2}$. The alternative estimator (20) has the advantage of being computationally significantly less intensive than $\hat{\theta}^*$ because it eliminates the need to calculate the preliminary QML estimator on each bootstrap sample. All of the large sample results given below for $\hat{\theta}^*$ also hold for $\check{\theta}^*$. $\diamondsuit$

We are now in a position to establish the large-sample distribution theory for $\hat{\theta}^*$ and its computationally simpler analogue, $\check{\theta}^*$. By analogy to the use of $\hat{C}$ to estimate C_0 in the case of the original data, we also consider the bootstrap analogue of $\hat{C}$, defined as

$$\hat{C}^* := \left. \left(\left(\frac{\partial^2 \hat{Q}_T^*(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} \frac{\partial q_t^*(\theta)}{\partial\theta} \frac{\partial q_t^*(\theta)}{\partial\theta'} \right) \left(\frac{\partial^2 \hat{Q}_T^*(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \right) \right|_{\theta=\hat{\theta}^*} \tag{21}$$

with $q_t^*(\theta) := \frac{\varepsilon_t^*(\theta)^2}{\hat{\sigma}_t^{*2}}$. The variance estimator corresponding to $\check{\theta}^*$, denoted $\check{C}^*$, is defined accordingly.

Theorem 3. *Let Assumptions 1–7 be satisfied and assume that $\theta_0 \in \mathrm{int}(\Theta)$. Then,*

$$\sqrt{T}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{w^*}_p N(0, C_0^\dagger) \text{ and } \hat{C}^* \xrightarrow{p^*}_p C_0^\dagger, \tag{22}$$

where $C_0^\dagger := B_0^{-1} A_0^\dagger B_0^{-1}$ with $A_0^\dagger := \sum_{n=1}^\infty \tau_{n,n} \left[\begin{array}{cc} n^{-2} & -\gamma_n(\psi_0)'/n \\ -\gamma_n(\psi_0)/n & \gamma_n(\psi_0)\gamma_n(\psi_0)' \end{array} \right]$. Furthermore,

$$\sqrt{T}(\check{\theta}^* - \hat{\theta}^*) \xrightarrow{p^*}_p 0 \text{ and } \check{C}^* - \hat{C}^* \xrightarrow{p^*}_p 0. \tag{23}$$

The large sample result in (22) can be used as a basis for developing asymptotically valid bootstrap confidence regions and hypothesis tests for θ. We describe these for $\hat{\theta}^*$ in the following remarks; corresponding results for $\check{\theta}^*$ follow entirely analogously.

Remark 4.7. It is immediately seen from a comparison of the limiting covariance matrices which appear in (11) and (22) that bootstrap confidence regions for θ based on non-studentized quantities will be (asymptotically) valid provided $C_0^\dagger = C_0$; that is, when $\tau_{r,s} = 0$ for $r \neq s$, so that $A_0^\dagger = A_0$. This additional condition rules out certain asymmetries in the fourth-order moments of z_t, but importantly does not place any restrictions on the third-order moments of z_t and hence does not restrict leverage, see also Remark 2.1. As an example, in the case where θ is a scalar parameter, letting $\hat{\theta}^*_{(\alpha)}$ denote the α percent quantile of the bootstrap distribution of $\hat{\theta}^*$, the asymptotic $(1-\alpha)\%$-level naïve (or basic) and percentile bootstrap confidence intervals for θ 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. ♢

Remark 4.8. The additional condition in Remark 4.7 can be avoided by bootstrapping pivotal statistics such as the studentized quantities W_T of (15) and t_T of (16), as will be considered in Remark 4.9 to follow, or studentized bootstrap confidence intervals. Because $\hat{C}^*$ converges to the correct limiting variance as shown in (22), the fact that $C_0 \neq C_0^\dagger$ is inconsequential for the validity of bootstrap procedures as long as these are properly studentized. For example, letting $t^*_{(\alpha)}$ and $|t^*|_{(\alpha)}$ denote the α percent quantiles of the bootstrap distributions of $t^*_{i,T} = T^{1/2}(\hat{\theta}^*_i - \hat{\theta}_i)/(\hat{C}^*_{ii})^{1/2}$ and $|t^*_{i,T}|$, respectively, the asymptotic $(1-\alpha)\%$-level equal-tailed and symmetric studentized (or percentile-t) bootstrap confidence intervals for θ_i are $[\hat{\theta}_i - t^*_{(1-\alpha/2)} T^{-1/2}(\hat{C}_{ii})^{1/2}; \hat{\theta}_i - t^*_{(\alpha/2)} T^{-1/2}(\hat{C}_{ii})^{1/2}]$ and $[\hat{\theta}_i - |t^*|_{(1-\alpha)} T^{-1/2}(\hat{C}_{ii})^{1/2}; \hat{\theta}_i + |t^*|_{(1-\alpha)} T^{-1/2}(\hat{C}_{ii})^{1/2}]$, respectively. As these intervals are based on studentized quantities, they do not require any additional conditions, and their (asymptotic) validity follows immmediately from Theorem 3 under the conditions stated there. ♢

Remark 4.9. Wild bootstrap analogues of the robust W_T and t_T statistics of (15) and (16), respectively, are given by

$$W_T^* := T(f(\hat{\theta}^*) - f(\hat{\theta}))'(F(\hat{\theta}^*)\hat{C}^* F(\hat{\theta}^*)')^{-1}(f(\hat{\theta}^*) - f(\hat{\theta})) \tag{24}$$

and

$$t_T^* := \frac{\sqrt{T}(f(\hat{\theta}^*) - f(\hat{\theta}))}{\sqrt{F(\hat{\theta}^*)\hat{C}^* F(\hat{\theta}^*)'}}. \tag{25}$$

It is immediate from Theorem 3 that these statistics attain the same first order limiting distributions as those attained under the null hypothesis by their non-bootstrap counterparts. This implies that the wild bootstrap tests based on W_T^* and t_T^* will have correct asymptotic size regardless of any conditional or unconditional heteroskedasticity (satisfying Assumption 1) present in ε_t and hence establishes their (asymptotic) validity. Again this result does not require the additional condition in Remark 4.7. ♢

5 Monte Carlo Simulations

We report results from a simulation study comparing the finite sample properties of confidence intervals based on the asymptotic and bootstrap theory described above, in the context of a fractionally integrated process allowing for autocorrelation 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 (26) or (27):

$$(1 - aL)u_t = \varepsilon_t, \tag{26}$$

$$u_t = (1 + bL)\varepsilon_t, \tag{27}$$

where in each case the structure of the innovations $\varepsilon_t = \sigma_t z_t$ will be defined below. Our focus will be on investigating the finite sample behaviour of confidence intervals for the long memory parameter, d. We set $d_0 = 0$ in what follows with no loss of generality.

We report results for asymptotic confidence intervals based on the QML estimator (reported under $\tilde{d}$) and the corresponding ACSS estimator (reported under $\hat{d}$). For the latter, $\hat{\sigma}_t$ was estimated using the Gaussian kernel and with the bandwidth parameter chosen by cross-validation as outlined in Remark 3.5. In each case the confidence intervals were based on robust standard errors, using $\tilde{C}$ and $\hat{C}$ (as defined in Remark 4.2 and (12)) for the QML and ACSS estimators, respectively. We also report results for symmetric studentized (or percentile-t) wild bootstrap intervals (see Remark 4.8 or Gonçalves and Kilian, 2004, p. 100) based on the QML estimator (reported under $\tilde{d}^*$) and based on the ACSS estimator. In the latter case, we report results based on (20) (reported under $\check{d}^*$) as well as results based on (18). For the results based on (18), the bandwidth is either re-determined by cross-validation for each bootstrap replication (reported under $\hat{d}_1^*$) or the bandwidth is simply chosen to be the same as that used for the original sample (reported under $\hat{d}_2^*$). For each method, we report the coverage percentage and the median length (across the Monte Carlo replications) of the confidence interval for d based on $10,000$ Monte Carlo replications. All of these methods of interval estimation are asymptotically pivotal for each of the models we will consider here.

Results are reported for samples of size $T = 100$, 250 and 500. All confidence intervals are nominal 90% intervals. The variance estimators required in, for example, (12) were implemented using numerical derivatives. For the bootstrap implementations, we used 999 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.

5.2 Results With Heteroskedastic, Uncorrelated Errors

We consider first the case where the shocks are not autocorrelated (i.e., $a = b = 0$, such that $u_t = \varepsilon_t$) and analyse the impact of heteroskedasticity on the confidence intervals, uncontaminated by the influence of autocorrelation in u_t.

The unconditional volatility process is generated according to the deterministic one-shift volatility process, $\sigma_t = \upsilon_1 + (\upsilon_2 - \upsilon_1)\mathbb{I}(t \geq \tau T)$; i.e., there is an abrupt single shift in the variance from υ_1^2 to υ_2^2 at time τT, for some $\tau \in (0,1)$. In this example, $\lambda = (\tau + (1-\tau)(\upsilon_2/\upsilon_1)^2)^{-2}(\tau + (1-\tau)(\upsilon_2/\upsilon_1)^4)$. This function is graphed in Xu and Phillips (2008, p. 270), whereby it is seen that the variance of the QML estimator will be least inflated by either early positive ($\upsilon_2/\upsilon_1 > 1$) or late negative ($\upsilon_2/\upsilon_1 < 1$) breaks, but most inflated by either early negative or late positive breaks. Without loss of generality we normalise $\upsilon_1^2 = 1$. We let the break date vary among $\tau \in \{1/4, 3/4\}$ and the ratio $\upsilon := \upsilon_2/\upsilon_1$ 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 volatility early in the sample and an increase in volatility late in the sample, respectively.

The results with uncorrelated errors are presented in Table 1. We consider the following three models for $\{z_t\}$, in each case with $\{e_t\}$ forming an i.i.d. standard normal sequence:

$$\text{Panel A} : z_t = e_t,$$
$$\text{Panel B} : z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.2 z_{t-1}^2 + 0.79 h_{t-1},$$
$$\text{Panel C} : z_t = e_t \exp(h_t), h_t = 0.936 h_{t-1} + 0.424 v_t, (v_t, e_t) \sim N(0, I_2).$$

Thus, Panel A relates to the case where z_t is conditionally homoskedastic, while Panels B and C contain results pertaining to conditionally heteroskedastic GARCH(1,1) and first-order autoregressive stochastic volatility [ARSV] specifications for z_t, respectively. These generating mechanisms and parameter values are taken from Gonçalves and Kilian (2004), where empirical evidence documenting their practical relevance is also presented. When z_t follows either a GARCH or an ARSV process, we simulate $T + 100$ values and discard the first 100 as initialization.

Consider first the results in the first three rows of Panels A, B and C where ε_t is unconditionally homoskedastic ($\upsilon = 1$). In this case, all of the reported confidence intervals have coverage rates which lie reasonably close to the nominal 90% level, albeit the standard QML estimator with asymptotic standard errors ($\tilde{d}$) has a coverage rate somewhat below the nominal level for $T = 100$, but the wild bootstrap-based confidence interval ($\tilde{d}^*$) rectifies this. Here there is also little to choose between the median confidence interval lengths, each of which decreases as the sample size increases, as would be expected given the consistency of both the QML and ACSS estimates. Where conditional heteroskedasticity is present the confidence intervals based on asymptotic standard errors do not perform as well, most notably where ε_t displays ARSV, with coverage rates consistently below the nominal level, increasingly so the smaller the sample size and for QML *vis-à-vis* ACSS. With one exception, the corresponding bootstrap confidence intervals do a good job in correcting the coverage rates. The exception is the wild bootstrap confidence interval for the ACSS estimate which uses the same bandwidth in the bootstrap samples as was estimated on the original data ($\hat{d}_2^*$) which has a coverage rate considerably in excess of the nominal level even for $T = 500$. For all methods, the median length of the confidence intervals is larger under conditional heteroskedasticity than under homoskedasticity.[4] Comparing across methods, it is clear that the ACSS-based $\hat{d}^*$ and $\hat{d}_1^*$ intervals perform similarly to one another and are clearly superior to both the wild bootstrap interval based on the QML estimator ($\tilde{d}^*$) and also to $\hat{d}_2^*$, in that they deliver approximately correct coverage rates and the smallest width intervals.

Consider next the results where unconditional heteroskedasticity is present in ε_t. Relative to the unconditionally homoskedastic results, we see a clear deterioration in finite sample coverage rates for the standard QML estimator with asymptotic standard errors. Other things equal, it performs worst where λ is largest. Its performance is particularly poor in the case where z_t is also an ARSV process (Panel C); here, even for $T = 500$, the coverage rate is still only around 84%. The ACSS estimator with asymptotic standard errors displays significantly better finite sample coverage, albeit coverage rates under ARSV z_t are also significantly below the nominal level. For both the QML and ACSS estimators, these effects are considerably ameliorated when implemented with a wild bootstrap. Amongst the wild bootstrap confidence intervals, the ACSS-based interval where the bandwidth is determined in each bootstrap sample using cross validation ($\hat{d}_1^*$) performs best with an empirical coverage rate very close to the nominal level throughout. The median length of the QML-based confidence intervals are, other things equal, inflated (often considerably so) the larger the value of λ, consistent with the impact of unconditional heteroskedasticity on the asymptotic variance matrix of the QML estimate; see (5). To illustrate, in the case where z_t is IID (Panel A) the median length of the wild bootstrap confidence interval based on $\tilde{d}$ ($T = 100$) increases from 0.274 under unconditional homoskedasticity to 0.430 when a late positive break in variance occurs. In contrast, the median length of the confidence intervals based on the ACSS estimator appear relatively unaffected by unconditional heteroskedasticity, as anticipated by Theorems 1 and 3. Notice also that, consistent with the large sample theory, the ratio of the median length of

[4]Recall from the results in (5) and Theorem 1 that the asymptotic variance of the limiting distributions of $\tilde{d}$ and $\hat{d}$ depend, in general, on any conditional heteroskedasticity present in z_t.

Table 1: Simulation results with uncorrelated errors, 90% nominal intervals

			Empirical coverage rate (%)						Median length					
τ	υ	T	$\tilde{\theta}$	$\tilde{\theta}^*$	$\hat{\theta}$	$\check{\theta}^*$	$\hat{\theta}_1^*$	$\hat{\theta}_2^*$	$\tilde{\theta}$	$\tilde{\theta}^*$	$\hat{\theta}$	$\check{\theta}^*$	$\hat{\theta}_1^*$	$\hat{\theta}_2^*$
Panel A: IID errors														
	1	100	88.7	89.8	89.1	90.2	89.8	90.5	0.265	0.274	0.269	0.278	0.276	0.281
	1	250	89.0	89.4	89.2	89.6	89.5	89.8	0.166	0.167	0.167	0.168	0.168	0.169
	1	500	90.2	90.5	90.3	90.4	90.4	90.5	0.116	0.117	0.117	0.117	0.117	0.118
1/4	1/3	100	85.9	89.4	88.6	90.3	89.8	92.5	0.358	0.405	0.282	0.295	0.293	0.319
1/4	1/3	250	88.0	89.9	89.6	90.0	90.0	91.8	0.239	0.253	0.172	0.174	0.174	0.183
1/4	1/3	500	88.8	89.8	89.5	89.6	89.6	91.0	0.173	0.178	0.119	0.120	0.120	0.124
1/4	3	100	88.0	89.3	88.6	89.9	89.5	91.2	0.295	0.310	0.292	0.304	0.301	0.317
1/4	3	250	88.7	89.2	88.7	89.1	89.0	90.4	0.185	0.188	0.176	0.179	0.178	0.184
1/4	3	500	90.2	90.6	90.0	90.2	90.1	90.8	0.130	0.131	0.121	0.122	0.122	0.125
3/4	1/3	100	88.3	89.6	88.8	90.4	89.8	91.3	0.288	0.302	0.280	0.293	0.289	0.305
3/4	1/3	250	89.4	90.0	89.9	90.2	90.1	91.1	0.183	0.186	0.173	0.175	0.175	0.180
3/4	1/3	500	89.4	89.7	89.5	89.6	89.5	90.3	0.129	0.129	0.120	0.120	0.120	0.123
3/4	3	100	85.5	89.0	88.7	90.4	90.0	92.8	0.378	0.430	0.295	0.307	0.306	0.333
3/4	3	250	87.8	89.5	89.7	90.1	89.9	91.9	0.246	0.261	0.176	0.178	0.178	0.188
3/4	3	500	88.4	89.3	89.3	89.5	89.4	90.8	0.175	0.181	0.121	0.121	0.121	0.126
Panel B: GARCH errors														
	1	100	87.3	89.8	88.8	90.2	90.3	92.8	0.310	0.333	0.290	0.303	0.302	0.328
	1	250	88.2	89.5	89.8	90.3	90.3	93.8	0.216	0.226	0.180	0.183	0.183	0.205
	1	500	88.7	89.6	89.4	89.7	89.7	94.0	0.168	0.174	0.125	0.126	0.126	0.144
1/4	1/3	100	85.2	89.0	87.6	89.7	89.6	92.8	0.366	0.416	0.284	0.304	0.302	0.342
1/4	1/3	250	87.4	89.6	89.5	90.3	90.1	93.9	0.259	0.281	0.179	0.182	0.182	0.206
1/4	1/3	500	88.3	90.0	89.6	89.9	89.9	94.0	0.201	0.211	0.124	0.125	0.125	0.142
1/4	3	100	87.2	89.6	88.0	89.7	89.6	92.7	0.336	0.365	0.302	0.318	0.317	0.354
1/4	3	250	87.9	89.5	89.3	89.8	89.9	93.3	0.230	0.242	0.183	0.186	0.186	0.210
1/4	3	500	88.4	89.6	89.4	89.5	89.3	94.1	0.180	0.186	0.127	0.127	0.128	0.145
3/4	1/3	100	86.3	89.0	87.3	89.6	89.4	92.4	0.324	0.352	0.284	0.302	0.300	0.333
3/4	1/3	250	88.2	89.8	89.5	90.4	90.3	93.8	0.227	0.239	0.177	0.180	0.180	0.203
3/4	1/3	500	88.8	89.9	89.2	89.5	89.6	93.9	0.177	0.183	0.124	0.124	0.125	0.142
3/4	3	100	85.2	89.1	87.8	89.7	89.5	93.1	0.391	0.448	0.302	0.319	0.319	0.363
3/4	3	250	87.3	89.7	88.8	89.3	89.5	93.4	0.268	0.290	0.184	0.188	0.188	0.214
3/4	3	500	87.9	89.5	89.6	89.8	90.1	94.2	0.206	0.217	0.127	0.128	0.128	0.147
Panel C: SV errors														
	1	100	82.8	88.2	84.0	88.7	88.9	92.3	0.410	0.489	0.306	0.345	0.348	0.406
	1	250	83.6	88.2	86.1	88.9	89.5	94.1	0.325	0.375	0.181	0.195	0.199	0.251
	1	500	84.9	89.1	87.2	88.7	89.7	95.2	0.270	0.305	0.117	0.122	0.125	0.167
1/4	1/3	100	82.2	88.0	83.8	88.9	89.6	92.5	0.421	0.517	0.280	0.324	0.328	0.383
1/4	1/3	250	83.3	88.5	85.7	88.6	89.5	94.1	0.339	0.400	0.168	0.182	0.189	0.239
1/4	1/3	500	84.1	88.8	87.1	88.7	90.0	95.6	0.289	0.334	0.111	0.116	0.121	0.168
1/4	3	100	83.3	88.2	84.3	88.7	88.8	92.5	0.422	0.507	0.317	0.356	0.360	0.424
1/4	3	250	84.4	88.7	86.0	88.5	89.3	94.1	0.334	0.387	0.188	0.200	0.205	0.258
1/4	3	500	84.7	89.1	87.0	88.6	89.7	95.5	0.282	0.322	0.122	0.127	0.131	0.175
3/4	1/3	100	82.5	87.9	83.6	88.6	89.0	92.4	0.408	0.493	0.279	0.323	0.327	0.381
3/4	1/3	250	84.0	89.0	85.5	88.7	89.7	94.2	0.327	0.382	0.167	0.181	0.185	0.236
3/4	1/3	500	84.3	88.6	86.6	88.4	89.4	95.1	0.279	0.319	0.108	0.114	0.118	0.161
3/4	3	100	82.0	88.3	84.6	88.8	89.1	92.9	0.440	0.534	0.325	0.367	0.373	0.435
3/4	3	250	84.0	88.8	86.6	89.1	90.0	94.4	0.348	0.412	0.191	0.203	0.209	0.264
3/4	3	500	84.0	88.7	87.1	88.9	89.9	95.3	0.294	0.340	0.124	0.129	0.133	0.180

Notes: The table reports empirical coverage percentage and median length of confidence intervals for d based on 10,000 replications. The reported intervals are based on the QML estimator with robust standard errors ($\tilde{\theta}$), the wild bootstrap equivalent ($\tilde{\theta}^*$), the ACSS estimator ($\hat{\theta}$), the wild bootstrap ACSS in (20) ($\check{\theta}^*$), the wild bootstrap ACSS estimator in (18), where the bandwidth is re-determined for each bootstrap sample ($\hat{\theta}_1^*$), and the wild bootstrap ACSS in (18), using the bandwidth from the original sample on each bootstrap sample ($\hat{\theta}_2^*$). The bootstrap intervals are symmetric studentized bootstrap confidence intervals as described in Remark 4.8, based on 999 bootstrap replications.

Table 2: Simulation results with AR or MA errors, 90% nominal intervals

			Empirical coverage rate (%)						Median length					
a	b	T	$\tilde{\theta}$	$\tilde{\theta}^*$	$\hat{\theta}$	$\check{\theta}^*$	$\hat{\theta}_1^*$	$\hat{\theta}_2^*$	$\tilde{\theta}$	$\tilde{\theta}^*$	$\hat{\theta}$	$\check{\theta}^*$	$\hat{\theta}_1^*$	$\hat{\theta}_2^*$
Panel A: No break, $\upsilon = 1$														
0.0		100	83.7	91.3	83.7	91.5	91.2	91.4	0.432	0.622	0.438	0.637	0.626	0.630
0.0		250	88.8	91.4	88.8	91.6	91.4	91.7	0.270	0.282	0.272	0.284	0.283	0.285
0.0		500	89.2	89.5	89.2	89.5	89.6	89.8	0.189	0.191	0.190	0.191	0.190	0.192
−0.8		100	88.3	89.8	88.8	90.2	89.8	90.4	0.284	0.300	0.289	0.304	0.301	0.306
−0.8		250	89.0	89.5	89.2	89.7	89.6	90.1	0.176	0.180	0.177	0.181	0.180	0.182
−0.8		500	89.7	89.9	89.8	89.9	90.0	90.0	0.124	0.125	0.124	0.125	0.125	0.126
0.8		100	76.4	96.8	76.8	96.9	96.9	97.1	0.461	0.921	0.466	0.941	0.922	0.941
0.8		250	79.3	96.4	79.5	96.3	96.4	96.5	0.349	0.619	0.350	0.623	0.617	0.625
0.8		500	82.8	95.6	82.8	95.7	95.7	95.8	0.282	0.414	0.283	0.416	0.415	0.416
	0.0	100	87.3	90.2	87.5	90.6	90.2	90.7	0.445	0.488	0.453	0.501	0.492	0.503
	0.0	250	89.8	90.4	90.0	90.5	90.5	90.6	0.272	0.274	0.274	0.275	0.275	0.278
	0.0	500	89.7	89.6	89.7	89.6	89.7	89.7	0.189	0.189	0.190	0.189	0.189	0.190
	−0.8	100	69.9	94.1	70.3	94.2	93.9	94.2	0.447	1.472	0.451	1.467	1.456	1.492
	−0.8	250	74.1	93.7	74.3	93.9	93.7	93.9	0.353	0.942	0.355	0.949	0.945	0.947
	−0.8	500	79.3	93.8	79.4	93.9	93.8	93.8	0.288	0.547	0.290	0.547	0.546	0.550
	0.8	100	88.9	90.0	89.2	90.2	89.8	90.4	0.290	0.300	0.295	0.305	0.300	0.307
	0.8	250	89.1	89.5	89.2	89.5	89.3	89.7	0.178	0.179	0.179	0.181	0.180	0.182
	0.8	500	90.2	90.0	90.3	90.3	90.3	90.3	0.124	0.125	0.125	0.125	0.125	0.125
Panel B: Early break, $\tau = 1/4, \upsilon = 1/3$														
0.0		100	77.7	89.5	82.5	92.4	91.9	93.3	0.512	0.791	0.427	0.629	0.610	0.645
0.0		250	84.7	92.6	88.8	91.6	91.3	92.9	0.370	0.457	0.276	0.293	0.291	0.306
0.0		500	87.9	91.2	89.4	89.7	89.4	90.8	0.272	0.294	0.192	0.194	0.193	0.201
−0.8		100	84.0	89.2	87.6	90.1	89.4	92.1	0.375	0.441	0.298	0.319	0.316	0.344
−0.8		250	87.5	89.8	89.3	90.1	89.9	91.7	0.253	0.273	0.183	0.187	0.186	0.197
−0.8		500	88.6	90.0	89.6	90.0	89.8	91.1	0.184	0.191	0.127	0.128	0.128	0.133
0.8		100	75.9	96.1	78.6	97.9	97.5	98.1	0.545	1.007	0.471	0.950	0.907	0.960
0.8		250	78.2	96.5	82.2	97.1	96.9	97.3	0.432	0.785	0.351	0.590	0.580	0.610
0.8		500	80.5	96.0	84.0	96.3	96.2	96.6	0.354	0.602	0.278	0.394	0.392	0.407
	0.0	100	84.2	89.2	87.2	90.5	90.0	91.9	0.577	0.704	0.448	0.511	0.499	0.541
	0.0	250	87.9	89.7	90.0	90.5	90.3	92.1	0.386	0.414	0.278	0.280	0.279	0.295
	0.0	500	89.1	89.8	90.1	90.2	89.9	91.2	0.277	0.282	0.192	0.192	0.190	0.199
	−0.8	100	70.8	93.2	69.8	93.4	93.2	93.8	0.518	1.812	0.440	1.353	1.333	1.507
	−0.8	250	71.4	93.3	74.3	93.1	92.8	93.2	0.410	1.269	0.351	0.873	0.866	0.915
	−0.8	500	73.4	92.4	79.7	93.5	93.5	93.8	0.344	0.869	0.285	0.506	0.502	0.525
	0.8	100	86.7	89.3	89.0	90.3	89.7	92.5	0.391	0.431	0.305	0.318	0.313	0.343
	0.8	250	87.8	89.2	89.4	89.5	89.4	91.4	0.257	0.269	0.185	0.186	0.185	0.196
	0.8	500	88.6	89.2	90.2	90.2	90.1	91.3	0.184	0.188	0.128	0.128	0.128	0.133
Panel C: Late break, $\tau = 3/4, \upsilon = 3$														
0.0		100	75.4	88.9	81.9	90.7	90.4	92.1	0.567	0.945	0.475	0.741	0.723	0.766
0.0		250	84.9	93.1	88.9	92.3	92.2	93.5	0.390	0.495	0.287	0.306	0.304	0.320
0.0		500	87.3	90.9	89.7	89.9	89.9	91.2	0.282	0.304	0.197	0.199	0.198	0.207
−0.8		100	84.1	88.7	87.9	89.7	89.8	92.3	0.401	0.472	0.316	0.337	0.333	0.365
−0.8		250	87.3	89.7	89.1	89.7	89.6	91.6	0.260	0.281	0.188	0.191	0.191	0.202
−0.8		500	88.2	89.5	89.2	89.7	89.6	91.1	0.186	0.194	0.129	0.130	0.130	0.135
0.8		100	73.8	97.0	75.2	97.1	97.0	97.5	0.572	1.188	0.491	1.053	1.029	1.119
0.8		250	76.2	96.4	78.5	96.1	96.0	96.5	0.437	0.854	0.363	0.677	0.669	0.707
0.8		500	78.1	95.7	82.0	95.6	95.7	96.0	0.362	0.648	0.290	0.446	0.442	0.463
	0.0	100	81.2	89.4	86.0	90.6	90.2	92.4	0.627	0.886	0.499	0.592	0.579	0.633
	0.0	250	87.9	90.3	90.0	91.0	90.7	92.6	0.407	0.445	0.291	0.295	0.292	0.310
	0.0	500	89.0	90.0	90.0	89.9	89.7	91.1	0.287	0.294	0.197	0.197	0.196	0.205
	−0.8	100	70.7	95.7	70.7	95.5	95.0	95.8	0.553	1.861	0.476	1.544	1.519	1.719
	−0.8	250	71.0	94.9	73.7	94.1	94.1	94.5	0.428	1.400	0.365	1.026	1.014	1.088
	−0.8	500	73.6	93.1	78.8	94.0	93.9	94.2	0.359	0.940	0.294	0.590	0.585	0.613
	0.8	100	87.1	89.7	88.9	90.2	90.0	92.7	0.417	0.468	0.322	0.335	0.333	0.366
	0.8	250	88.0	89.8	89.1	89.4	89.5	91.5	0.267	0.280	0.189	0.191	0.190	0.202
	0.8	500	89.5	90.2	90.6	90.4	90.6	91.8	0.189	0.193	0.129	0.130	0.130	0.135

Notes: See notes to Table 1.

the wild bootstrap confidence intervals based on $\tilde{d}^*$ (QML) and $\hat{d}_1^*$ (ACSS) is approximately unity, regardless of T, when $\upsilon = 1$ (such that $\sqrt{\lambda} = 1$), and is 1.41, 1.47 and 1.50 for $T = 100, 250$ and 500, respectively, in the late positive break case (where $\sqrt{\lambda} \approx 1.53$).

5.3 Results With Heteroskedastic, Autocorrelated Errors

Table 2 reports results for cases where u_t can display both weak parametric autocorrelation and heteroskedasticity. Specifically, u_t is generated according to either (26) or (27) with $a, b \in \{-0.8, 0, 0.8\}$. The reported cases where either $a = 0$ or $b = 0$ correspond to the situation where an AR(1) or MA(1) specification is estimated, respectively, even though it is not present in the data generating process. Results are given for where ε_t is either IID (Panel A) or displays an early negative (Panel B) or late positive (Panel C) break in is unconditional variance.

Relative to the results in Table 1, autocorrelation can be seen to have a significant impact on both the coverage rate and median length of the confidence intervals. For both the QML and ACSS estimators, coverage rates based on asymptotic standard errors are not as accurate (relative to the corresponding results in Table 1) when autocorrelation is either present and modelled or not present but allowed for in the estimated model. For example, where $a = 0$ (so that an AR(1) is modelled but not actually present in the data) the coverage rates for $\tilde{d}$ and $\hat{d}$ are both 83.7% when $T = 100$ in the homoskedastic case compared to 88.7% and 89.1%, respectively, in the corresponding case in Table 1. Where autocorrelation is present, the empirical coverage rates of $\tilde{d}$ and $\hat{d}$ can lie significantly below the nominal level. This is seen most obviously for the cases where u_t is either positively autocorrelated ($a = 0.8$) or follows a negative moving average ($b = -0.8$). To illustrate, when $b = -0.8$ the coverage rates for $\tilde{d}$ and $\hat{d}$ are around 70% when $T = 100$, regardless of whether ε_t is homoskedastic or contains a break in variance. As with the impact of conditional heteroskedasticity, the wild bootstrap considerably improves the coverage rates of the confidence intervals, albeit in the most problematic cases above ($a = 0.8$ and $b = -0.8$) the wild bootstrap tends to rather over correct such that the resulting confidence intervals are somewhat too liberal.

Turning to confidence interval width, as with the presence of conditional heteroskedasticity, the median lengths of the confidence intervals are seen to vary with a and b, again as expected given the results for the QML and ACSS estimates in (5) and Theorem 1, respectively. The confidence intervals are considerably wider when either $a = 0.8$ or $b = -0.8$ than when $a = -0.8$ or $b = 0.8$. Controlling for the impact of autocorrelation, however, the results in Table 2 reveal qualitatively similar conclusions to those drawn from the results in Table 1; that is, the efficiency gains from basing confidence intervals around the ACSS estimator, $\hat{d}$, rather than the QML estimator, $\tilde{d}$ are clearly visible when unconditional heteroskedasticity is present in ε_t. As with the conclusions drawn from Table 1, the ACSS-based interval with the bandwidth determined in each bootstrap sample using cross validation ($\hat{d}_1^*$) appears to deliver the best overall performance.

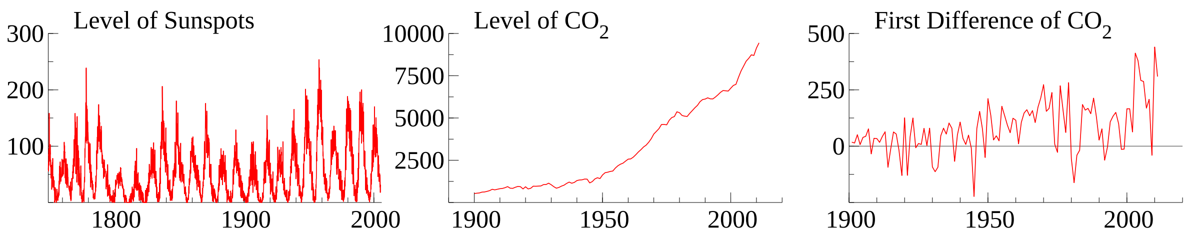
6 Data Examples

We now apply the methods discussed in this paper to a variety of data sets. All bootstrap confidence intervals were based on 999 replications using the Rademacher distribution for w_t. For each data set we consider we report the QML and ACSS estimates of d along with their (robust or sandwich) standard errors and 95% confidence intervals based on (5) and (11), respectively, together with 95% wild bootstrap percentile-t confidence intervals. For each data set an ARFIMA(p,d,0) model was fitted to the data with p chosen by a forward search algorithm starting from $p = 0$ ($p = 12$ for the sunspots data which includes a clear seasonal pattern) sequentially increasing p by one until the additional lag was deemed statistically insignificant (at the nominal asymptotic 10% level) using the appropriate form of the statistic t_T in (16). Additional graphical analysis of the residuals from the chosen models for each data set along with formal statistical tests for heteroskedasticity are reported in Section S.3 in the Supplementary Appendix.

6.1 Physical Data

The first data set we consider is the monthly mean sunspot number observed for 1749:1 to 2005:2. The data are plotted in the left panel of Figure 1 and were obtained from the NOAA Earth System

Figure 1: Physical Data



Notes: Sunspots data is monthly 1749:1 to 2005:2 and CO_2 data is annual 1900 to 2011.

Table 3: Physical Data

	Est. d	se(d)	Asympt. CI_L	Asympt. CI_U	WB_1 CI_L	WB_1 CI_U	WB_2 CI_L	WB_2 CI_U	WB_3 CI_L	WB_3 CI_U
Panel A: Sunspots										
QML	0.482	0.053	0.378	0.586	0.352	0.612				
ACSS	0.448	0.034	0.381	0.514	0.378	0.518	0.377	0.518	0.371	0.524
Panel B: CO_2										
QML	0.326	0.078	0.173	0.480	0.141	0.512				
ACSS	0.304	0.072	0.164	0.444	0.147	0.461	0.155	0.453	0.147	0.461

Notes: The sample sizes are $T = 3074$ (sunspots) and $T = 111$ (CO_2). An ARFIMA(12,d,0) model was estimated on the de-meaned sunspot data and an ARFIMA(0,d,0) model was estimated on the de-meaned first differenced CO_2 data. se(d) is the (robust) standard error. WB_1 does not re-estimate σ_t for each bootstrap replication; WB_2 re-estimates σ_t for each bootstrap replication, estimating the bandwidth parameter in each bootstrap replication using cross-validation; WB_3 re-estimates σ_t for each bootstrap replication using same bandwidth parameter as selected for the original data.

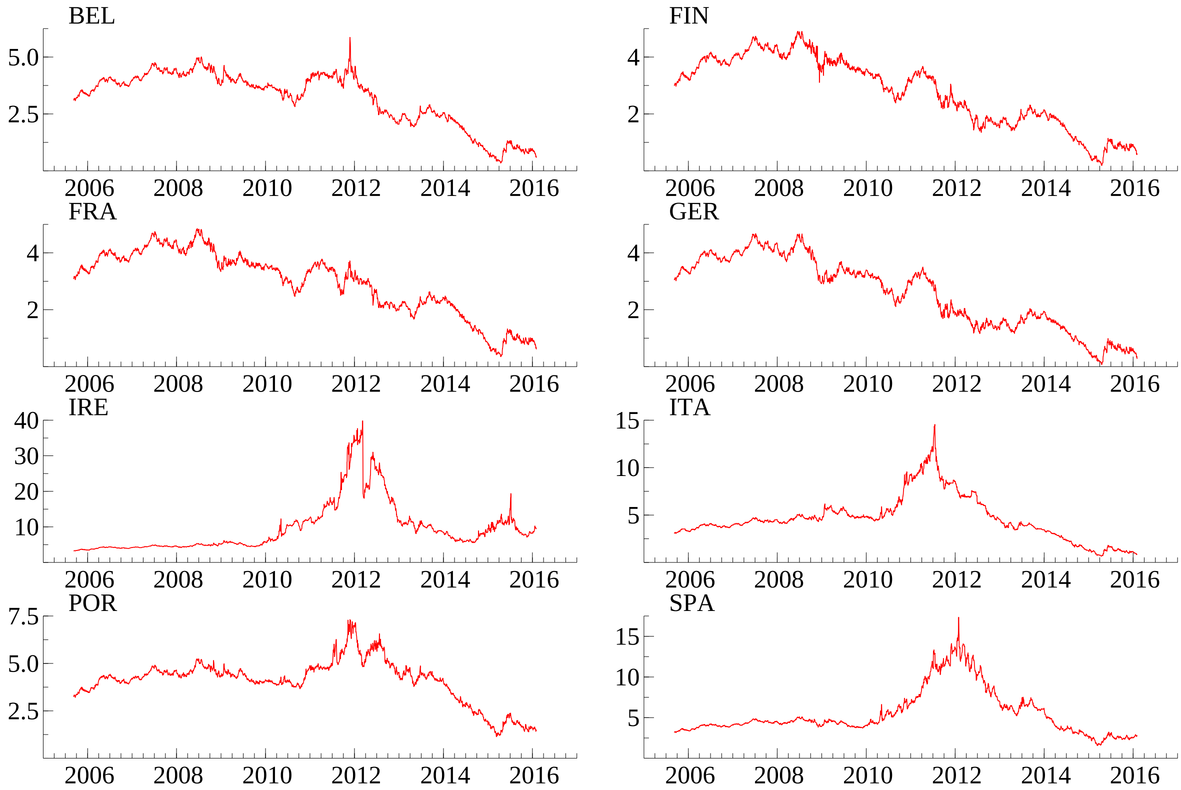
Research Laboratory time-series database (`www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/SUNSPOT/`). The results of the heteroskedasticity tests in Section S.3 highlight strong evidence of conditionally heteroskedastic behaviour in these data as well as evidence of unconditional heteroskedasticity. Panel A of Table 3 reports estimation results for the sunspot data. The ACSS estimate of d reported in Panel A of Table 3 is slightly lower than the QML estimate, and the ACSS-based confidence intervals are smaller than the QML intervals, consistent with the theoretical findings of the paper. For example, in the case of the wild bootstrap method which does not re-estimate σ_t for each bootstrap replication, the ACSS-based confidence interval is almost half the width of the corresponding QML-based confidence interval. The null hypothesis that $d = 0.5$ cannot, however, be rejected based on any of the reported intervals.

The middle panel of Figure 1 plots annual data from 1900–2011 on global CO_2 emissions from fossil-fuel burning, cement manufacture, and gas flaring. The data were obtained from the Carbon Dioxide Information Analysis Center. The data display a clear upward trend and, hence, we will analyse the first differences of the data shown in the right panel of Figure 1. The results in Section S.3 again highlight strong evidence of conditionally heteroskedastic behaviour in these data, but no significant evidence of any unconditional heteroskedasticity. Both the ACSS and QML estimates of d, reported in Panel B of Table 3 are smaller than 0.5 with the former lying further from 0.5 than the latter. The ACSS-based confidence intervals are also smaller. Interestingly, the null hypothesis that $d = 0.5$ is rejected, suggesting that the annual changes in CO_2 emissions are weakly stationary, using the ACSS wild bootstrap confidence intervals, but cannot be rejected the based on the QML wild bootstrap interval.

6.2 Sovereign Debt Data

We next analyse the sovereign debt data series for the following eight Eurozone countries considered in Martins and Amado (2016): Belgium, Finland, France, Germany, Ireland, Italy, Portugal, and Spain. These are (demeaned) daily data on 10-year government bond yields from September 5,

Figure 2: Sovereign Debt Data



Note: Daily data on 10-year government bond yields from September 5, 2005, to February 2, 2016.

2005, to February 2, 2016, for a sample size of $T = 2715$. These series are graphed in Figure 2 and detailed data descriptions can be found in Martins and Amado (2016, Section 3.1). Both the conditional and unconditional heteroskedasticity tests reported in Section S.3 display highly significant rejections, providing strong evidence of heteroskedastic behaviour in these data.

The ACSS and QML estimates of d, reported in Table 4 are similar for each country considered with estimated values for all countries lying relatively close to 1. However, the estimated ACSS standard errors are smaller, in many cases substantially so, for each country than the corresponding QML estimates, which translates into commensurately smaller confidence interval widths for d when based on ACSS rather than QML estimation. To illustrate, comparing the widths of wild bootstrap confidence intervals which do not re-estimate σ_t for each bootstrap replication, we see that for Portugal the width of the ACSS-based interval is only about a third of that of the QML-based interval, while for Ireland the ACSS-based interval is about half the width of the QML-based interval. Interestingly, the differences in the widths of the ACSS- and QML-based confidence intervals do not lead to a different outcome when using these intervals to test the null hypothesis that $d = 1$ for any of the series. In particular, the unit root null hypothesis cannot be rejected at the 95% level for all series except Spain and Ireland. The former (latter) appears to display lower (higher) persistence than a unit root process.

7 Conclusions

In this paper we have discussed estimation and inference on the parameters of fractionally integrated time series models driven by shocks which can display conditional and/or unconditional heteroskedasticity. The asymptotic variance matrix of the limiting distribution of the standard QML estimator is inflated under unconditional heteroskedasticity relative to the unconditionally homoskedastic case. We have shown that an adaptive version of the QML estimator, based on a non-parametric kernel-based estimator of the unconditional variance process, attains the same asymptotic variance matrix when unconditional heteroskedasticity is present as the standard QML

Table 4: Sovereign Debt Data

	Est. d	se(d)	Asympt. CI_L	Asympt. CI_U	WB_1 CI_L	WB_1 CI_U	WB_2 CI_L	WB_2 CI_U	WB_3 CI_L	WB_3 CI_U
Belgium ($p = 1$)										
QML	0.929	0.050	0.831	1.026	0.829	1.028				
ACSS	0.961	0.032	0.899	1.023	0.897	1.026	0.899	1.024	0.899	1.024
Finland ($p = 1$)										
QML	0.961	0.026	0.911	1.011	0.911	1.012				
ACSS	0.998	0.022	0.956	1.041	0.955	1.041	0.955	1.042	0.953	1.044
France ($p = 0$)										
QML	1.005	0.019	0.968	1.042	0.969	1.041				
ACSS	1.008	0.017	0.975	1.041	0.975	1.041	0.974	1.042	0.974	1.042
Germany ($p = 1$)										
QML	0.971	0.025	0.922	1.020	0.920	1.022				
ACSS	0.992	0.022	0.948	1.036	0.949	1.036	0.950	1.034	0.947	1.038
Ireland ($p = 0$)										
QML	1.098	0.043	1.014	1.182	1.011	1.185				
ACSS	1.082	0.024	1.034	1.130	1.033	1.131	1.034	1.131	1.034	1.130
Italy ($p = 2$)										
QML	0.940	0.048	0.846	1.033	0.844	1.036				
ACSS	0.961	0.029	0.903	1.018	0.901	1.021	0.906	1.016	0.902	1.019
Portugal ($p = 1$)										
QML	0.965	0.062	0.843	1.087	0.834	1.096				
ACSS	0.970	0.022	0.928	1.013	0.928	1.013	0.927	1.013	0.926	1.015
Spain ($p = 1$)										
QML	0.872	0.038	0.797	0.947	0.796	0.948				
ACSS	0.935	0.025	0.886	0.984	0.884	0.986	0.885	0.985	0.885	0.985

Notes: Daily data on 10-year government bond yields from September 5, 2005, to February 2, 2016. Sample size is $T = 2715$. An ARFIMA(p,d,0) model was estimated on the de-meaned data. se(d), WB_1, WB_2 and WB_3 are as described in the Note to Table 3.

estimator would achieve under unconditional homoskedasticity and, hence, achieves (asymptotic) efficiency gains over the QML estimator. Under Gaussianity, the adaptive estimator is asymptotically efficient. We have shown that asymptotically pivotal inference based on the adaptive estimator can be achieved through the development of confidence regions or hypothesis tests using either heteroskedasticity robust standard errors and/or a wild bootstrap, and that these deliver gains in asymptotic efficiency/local power over their counterparts based on the standard QML estimator. Monte Carlo simulation results reported suggest that the large sample advantages of basing confidence intervals on adaptive methods carry over into finite samples. Empirical illustrations of the proposed methods were given for a variety of data sets.

Appendix A: Proof of Theorem 1

The following lemma provides the technical results needed to prove our main theorems. Importantly, we do not prove a (uniform) consistency result for the volatility function itself, such as $\sup_t |\hat{\sigma}_t - \sigma_t| \xrightarrow{p} 0$, but only for a smoothed average of the volatility function as in part (d) below. This is sufficient for the proofs of the main theorems, and allows us to assume only that $\sigma(\cdot)$ is càdlàg, where, e.g., Hansen (1995) and Xu and Phillips (2008) need much stronger assumptions. The proof is given in Section S.2.1.

Lemma A.1. *Define* $\tilde{\sigma}_t^2 := \sum_{i=1}^T k_{ti}\varepsilon_i^2$ *and* $\bar{\sigma}_t^2 := \sum_{i=1}^T k_{ti}\sigma_i^2$. *Then:*

(a) Under Assumptions 1(b), 6, and 7, $T^{-1}\sum_{t=1}^T(\bar{\sigma}_t^2 - \sigma_t^2)^2 = o(1)$,

(b) Under Assumptions 1 and 6, $T^{-1}\sum_{t=1}^T(\tilde{\sigma}_t^2 - \bar{\sigma}_t^2)^2 = O_p(T^{-1}b^{-1})$,

(c) Under Assumptions 1–3 and 6, $T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^2-\tilde{\sigma}_t^2)^2=O_p(T^{-3/2}b^{-1})$*,*

(d) Under Assumptions 1–3, 6, and 7, $T^{-1}\sum_{t=1}^{T}(\sigma_t^2-\hat{\sigma}_t^2)^2=o_p(1)$*,*

(e) Under Assumptions 1 and 6, $T^{-1}\sum_{t=1}^{T}(\tilde{\sigma}_t^2-\bar{\sigma}_t^2)^4=O_p(T^{-2}b^{-2})$*,*

(f) Under Assumptions 1–3, 6, and 7, $\max_{1\le t\le T}\hat{\sigma}_t^2=O_p(1)$*,*

(g) Under Assumptions 1–3, 6, and 7, $(\min_{1\le t\le T}\hat{\sigma}_t^2)^{-1}=O_p(1)$*,* $(\min_{1\le t\le T}\tilde{\sigma}_t^2)^{-1}=O_p(1)$*, and* $(\min_{1\le t\le T}\bar{\sigma}_t^2)^{-1}=O_p(1)$*.*

A.1 Proof of consistency

As in other proofs of consistency for fractional time series models, e.g. Theorem 1 of CNT, the parameter space Θ is partitioned into disjoint compact subsets; in this case $\Theta_1:=\Theta_1(\kappa)=D_1\times\Psi$ and $\Theta_2:=\Theta_2(\kappa)=D_2\times\Psi$, where $D_1:=D_1(\kappa)=D\cap\{d:d-d_0\le -1/2+\kappa\}$ and $D_2:=D_2(\kappa)=D\cap\{d:d-d_0\ge -1/2+\kappa\}$ for some constant $\kappa\in(0,1/2)$ to be determined later. Clearly, $\theta_0\in\Theta_2$ and if $d_1>d_0-1/2$ then the choice $\kappa=d_1-d_0+1/2>0$ implies that Θ_1 is empty in which case the proof is easily simplified accordingly.

First we show that for any $K>0$ there exists a (fixed) $\bar{\kappa}>0$ such that

$$P(\inf_{\theta\in\Theta_1(\bar{\kappa})}\hat{Q}_T(\theta)>K)\to 1 \text{ as } T\to\infty, \tag{A.1}$$

$$P(\inf_{\theta\in\Theta_1(\bar{\kappa})}\bar{Q}_T(\theta)>K)\to 1 \text{ as } T\to\infty. \tag{A.2}$$

However, this follows easily from the lower bound $\hat{Q}_T(\theta)\ge(\max_{1\le s\le T}\hat{\sigma}_s^2)^{-1}T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta)^2$, because (A.1) is proven for $T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta)^2$ in CNT and $\max_{1\le s\le T}\hat{\sigma}_s^2=O_p(1)$ by Lemma A.1(f). It follows that $P(\hat{\theta}\in\Theta_2(\bar{\kappa}))\to 1$ as $T\to\infty$, so that the relevant parameter space is reduced to $\Theta_2(\bar{\kappa})$. The same holds for (A.2).

We define also the objective function

$$\bar{Q}_T^0(\theta):=T^{-1}\sum_{t=1}^{T}\left(\sum_{n=0}^{t-1}b_n(\psi)\sum_{j=0}^{t-n-1}\pi_j(d_0-d)\sum_{m=0}^{\infty}a_m(\psi_0)\frac{\varepsilon_{t-n-j-m}}{\sigma_{t-n-j-m}}\right)^2,$$

where $\varepsilon_t/\sigma_t=z_t$, from which it follows easily that $\arg\min_{\theta\in\Theta}\bar{Q}_T^0(\theta)\xrightarrow{p}\theta_0$ by (5). In view of (A.1) and (A.2), the desired results follow if, for any $\kappa>0$,

$$\sup_{\theta\in\Theta_2}|\hat{Q}_T(\theta)-\bar{Q}_T(\theta)|\xrightarrow{p}0 \text{ and } \sup_{\theta\in\Theta_2}|\bar{Q}_T^0(\theta)-\bar{Q}_T(\theta)|\xrightarrow{p}0. \tag{A.3}$$

For the first statement in (A.3) we find the difference

$$\begin{aligned}|\hat{Q}_T(\theta)-\bar{Q}_T^0(\theta)|&=\left|T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta)^2\frac{(\sigma_t^2-\hat{\sigma}_t^2)}{\hat{\sigma}_t^2\sigma_t^2}\right|\\&\le(\min_{1\le t\le T}\hat{\sigma}_t^2\sigma_t^2)^{-1}\left(T^{-1}\sum_{t=1}^{T}\varepsilon_t(\theta)^4\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}(\sigma_t^2-\hat{\sigma}_t^2)^2\right)^{1/2}\end{aligned}$$

by the Cauchy-Schwarz inequality. The first term on the right-hand side is $O_p(1)$ by Lemma A.1(g) and uniform strict positivity of σ_t^2 (Assumption 1(b)(ii)). For the second and third terms we apply Lemmas S.3 and A.1(d), respectively.

To prove the second statement in (A.3) we decompose the residual as

$$\varepsilon_t(\theta)=\sum_{j=0}^{t-1}\phi_j(\theta)\varepsilon_{t-j}+r_t(\theta), \tag{A.4}$$

where the coefficients $\phi_j(\theta)$ and remainder term $r_t(\theta)$ are subject to the bounds in Lemma S.2. We then split the infinite summation in $\bar{Q}_T^0(\theta)$ and find

$$\bar{Q}_T(\theta) - \bar{Q}_T^0(\theta) = T^{-1} \sum_{t=1}^{T} \sum_{j=0}^{t-1} \phi_j(\theta)^2 \varepsilon_{t-j}^2 \frac{\sigma_{t-j}^2 - \sigma_t^2}{\sigma_{t-j}^2 \sigma_t^2} \tag{A.5}$$

$$+ 2T^{-1} \sum_{t=1}^{T} \sum_{k>j=0}^{t-1} \phi_j(\theta)\phi_k(\theta)\varepsilon_{t-j}\varepsilon_{t-k} \frac{\sigma_{t-j}\sigma_{t-k} - \sigma_t^2}{\sigma_{t-j}\sigma_{t-k}\sigma_t^2} \tag{A.6}$$

$$+ T^{-1} \sum_{t=1}^{T} \sigma_t^{-2} r_t(\theta)^2 - T^{-1} \sum_{t=1}^{T} \tilde{r}_t(\theta)^2 \tag{A.7}$$

$$+ 2T^{-1} \sum_{t=1}^{T} \sum_{j=0}^{t-1} \sigma_t^{-2} \phi_j(\theta)\varepsilon_{t-j} r_t(\theta) - 2T^{-1} \sum_{t=1}^{T} \sum_{i=0}^{t-1} \phi_i(\theta) z_{t-i} \tilde{r}_t(\theta), \tag{A.8}$$

where $\tilde{r}_t(\theta)$ is defined in the same way as $r_t(\theta)$ but with z_t replacing ε_t. For (A.5) we reverse the order of the summations and find the bound

$$E \sup_{\theta \in \Theta_2} |(A.5)| \leq c \sum_{j=1}^{T-1} j^{-1-2\kappa} T^{-1} \sum_{t=j+1}^{T} |\sigma_{t-j}^2 - \sigma_t^2| \to 0$$

using Lemma S.2 and Lemma S.1 with $a_j = j^{-1-2\kappa}$ and $b_{j,T} = T^{-1} \sum_{t=j+1}^{T} |\sigma_{t-j}^2 - \sigma_t^2|$, which, by Cavaliere and Taylor (2009, Lemma A.1), satisfies the assumptions of Lemma S.1.

For (A.6) we apply summation by parts, noting that $\pi_{n+1}(d) - \pi_n(d) = \pi_{n+1}(d-1)$ implies $\phi_{n+1}(d,\psi) - \phi_n(d,\psi) = \phi_{n+1}(d-1,\psi)$, see Lemma S.2, and then

$$\begin{aligned}
&\sum_{k=j+1}^{T-1} \phi_k(d_0 - d, \psi) \sum_{t=k+1}^{T-1} \varepsilon_{t-j}\varepsilon_{t-k} \frac{\sigma_{t-j}\sigma_{t-k} - \sigma_t^2}{\sigma_{t-j}\sigma_{t-k}\sigma_t^2} \\
&= \phi_T(d_0 - d, \psi) \sum_{k=j+1}^{T-1} \sum_{t=k+1}^{T-1} \varepsilon_{t-j}\varepsilon_{t-k} \frac{\sigma_{t-j}\sigma_{t-k} - \sigma_t^2}{\sigma_{t-j}\sigma_{t-k}\sigma_t^2} \\
&\quad - \sum_{l=j+1}^{T-2} \phi_{l+1}(d_0 - d - 1, \psi) \sum_{k=j+1}^{l} \sum_{t=k+1}^{T-1} \varepsilon_{t-j}\varepsilon_{t-k} \frac{\sigma_{t-j}\sigma_{t-k} - \sigma_t^2}{\sigma_{t-j}\sigma_{t-k}\sigma_t^2},
\end{aligned}$$

where $\sum_{k=j+1}^{l} \sum_{t=k+1}^{T-1} \varepsilon_{t-j}\varepsilon_{t-k} \frac{\sigma_{t-j}\sigma_{t-k} - \sigma_t^2}{\sigma_{t-j}\sigma_{t-k}\sigma_t^2} = \sum_{t=k+1}^{T-1} \varepsilon_{t-j} \sum_{k=j+1}^{l} \varepsilon_{t-k} \frac{\sigma_{t-j}\sigma_{t-k} - \sigma_t^2}{\sigma_{t-j}\sigma_{t-k}\sigma_t^2} = \sum_{t=k+1}^{T-1} v_{t,j,l}$ and $v_{t,j,l}$ is a martingale difference sequence with variance of order l. It follows that $E|\sum_{t=k+1}^{T-1} v_{t,j,l}| \leq cT^{1/2} l^{1/2}$, uniformly in k, j. Thus, using Lemma S.2 we obtain the bound

$$\begin{aligned}
E \sup_{\theta \in \Theta_2} |(A.6)| &= E \sup_{\theta \in \Theta_2} |T^{-1} \sum_{j=0}^{T-1} \phi_j(d_0 - d, \psi) \sum_{k=j+1}^{T-1} \phi_k(d_0 - d, \psi) \sum_{t=k+1}^{T-1} \varepsilon_{t-j}\varepsilon_{t-k} \frac{\sigma_{t-j}\sigma_{t-k} - \sigma_t^2}{\sigma_{t-j}\sigma_{t-k}\sigma_t^2}| \\
&\leq cT^{-1} \sum_{j=1}^{T-1} j^{-1/2-\kappa} T^{-1/2-\kappa} T + cT^{-1} \sum_{j=1}^{T-1} j^{-1/2-\kappa} \sum_{l=j+1}^{T-2} l^{-3/2-\kappa} T^{1/2} l^{1/2} \leq cT^{-2\kappa}.
\end{aligned}$$

For the proofs of (A.7) and (A.8), first note that $\sigma_t^{-2} r_t(\theta)$, $r_t(\theta)$, and $\tilde{r}_t(\theta)$ are clearly subject to the same bounds due to Assumption 1(b), see Lemma S.2, and thus the same proof applies to each term in (A.7) and to each term in (A.8). For (A.7) we find that $E \sup_{\theta \in \Theta_2} |(A.7)| \leq cT^{-1} \sum_{t=1}^{T} t^{-1-2\kappa} \leq cT^{-1}$ by Lemma S.2, while for (A.8) we find

$$E \sup_{\theta \in \Theta_2} |(A.8)| \leq cT^{-1} \sum_{t=1}^{T} \sum_{j=0}^{t-1} j^{-1/2-\kappa} t^{-1/2-\kappa} \leq cT^{-2\kappa}$$

by Lemma S.2. This proves the second statement in (A.3), and hence completes the proof.

A.2 Proof of asymptotic normality

We show that

$$\sup_{\theta\in\mathcal{N}(\theta_0)}\left|\frac{\partial^2\hat{Q}_T(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2\bar{Q}_T(\theta)}{\partial\theta\partial\theta'}\right|\xrightarrow{p}0 \text{ and } \sup_{\theta\in\mathcal{N}(\theta_0)}\left|\frac{\partial^2\bar{Q}_T(\theta)}{\partial\theta\partial\theta'}-\frac{\partial^2\bar{Q}^0_T(\theta)}{\partial\theta\partial\theta'}\right|\xrightarrow{p}0, \tag{A.9}$$

$$\sqrt{T}\frac{\partial\hat{Q}_T(\theta_0)}{\partial\theta}-\sqrt{T}\frac{\partial\bar{Q}_T(\theta_0)}{\partial\theta}\xrightarrow{p}0 \text{ and } \sqrt{T}\frac{\partial\bar{Q}_T(\theta_0)}{\partial\theta}-\sqrt{T}\frac{\partial\bar{Q}^0_T(\theta_0)}{\partial\theta}\xrightarrow{p}0, \tag{A.10}$$

which together imply both (10) and (11) by (5). The proof of (A.9) follows by the same argument as that of (A.3), noting that the derivatives add at most a logarithmic term, see (2) and Lemma S.2.

To prove (A.10) we first decompose $\varepsilon_t(\theta_0)$, similarly to (A.4), as

$$\varepsilon_t(\theta_0)=\varepsilon_t+r_t, \tag{A.11}$$

where r_t is subject to the bound in Lemma S.2. We then let $v_t:=\varepsilon_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}$ and find that

$$\begin{aligned}\sqrt{T}\frac{\partial\hat{Q}_T(\theta_0)}{\partial\theta}-\sqrt{T}\frac{\partial\bar{Q}_T(\theta_0)}{\partial\theta}=&2T^{-1/2}\sum_{t=1}^T r_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}(\hat{\sigma}_t^{-2}-\sigma_t^{-2}) &\text{(A.12)}\\ &+2T^{-1/2}\sum_{t=1}^T v_t(\hat{\sigma}_t^{-2}-\tilde{\sigma}_t^{-2}) &\text{(A.13)}\\ &+2T^{-1/2}\sum_{t=1}^T v_t(\tilde{\sigma}_t^{-2}-\bar{\sigma}_t^{-2}) &\text{(A.14)}\\ &+2T^{-1/2}\sum_{t=1}^T v_t(\bar{\sigma}_t^{-2}-\sigma_t^{-2}). &\text{(A.15)}\end{aligned}$$

For (A.12) we apply the Cauchy-Schwarz inequality and find that the i'th element satisfies

$$|(A.12)_i|\leq 2\left(T^{-1/2}\sum_{t=1}^T r_t^2\left(\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta_i}\right)^2\right)^{1/2}\left(T^{-1/2}\sum_{t=1}^T(\hat{\sigma}_t^{-2}-\sigma_t^{-2})^2\right)^{1/2},$$

where the first term is $O_p(T^{-1/4})$ by Lemma S.2 and the second term satisfies $T^{-1/2}\sum_{t=1}^T(\hat{\sigma}_t^{-2}-\sigma_t^{-2})^2\leq(\min_{1\leq s\leq T}\hat{\sigma}_s^2\sigma_s^2)^{-2}\sum_{t=1}^T(\sigma_t^2-\hat{\sigma}_t^2)^2=o_p(T^{1/2})$ by Lemma A.1(d),(g), so that $|(A.12)_i|=o_p(1)$.

For (A.15) we note that $v_t(\bar{\sigma}_t^{-2}-\sigma_t^{-2})$ is a martingale difference sequence so that the i'th element satisfies

$$E(A.15)_i^2=4T^{-1}\sum_{t=1}^T(Ev_{it}^2)(\bar{\sigma}_t^{-2}-\sigma_t^{-2})^2\leq 4(\min_{1\leq s\leq T}\bar{\sigma}_s^2\sigma_s^2)^{-2}(\sup_t Ev_{it}^2)T^{-1}\sum_{t=1}^T(\sigma_t^2-\bar{\sigma}_t^{-2})^2,$$

where we can apply Lemma A.1(a),(g) to the first and last terms on the right-hand side. Using the decomposition (A.4) and the Cauchy-Schwarz inequality the middle term is

$$\sup_t E v_{it}^2 = \sup_t E\varepsilon_t^2 \left(\sum_{j=0}^{t-1} \frac{\partial \phi_j(\theta_0)}{\partial \theta_i} \varepsilon_{t-j} + \frac{\partial r_t(\theta_0)}{\partial \theta_i} \right)^2$$

$$\leq \sup_t c \sum_{j=1}^{t-1} (\log j)^2 j^{-2} \sigma_t^2 \sigma_{t-j}^2 \tau_{j,j} + \sup_t c \sum_{j>k=1}^{t-1} (\log j)(\log k) j^{-1} k^{-1} \sigma_t^2 \sigma_{t-j} \sigma_{t-k} |\kappa_4(t,t,t-j,t-k)|$$

$$+ \sup_t \left(E\varepsilon_t^4\right)^{1/2} \left(E \left(\frac{\partial r_t(\theta_0)}{\partial \theta_i} \right)^4 \right)^{1/2} + \sup_t 2E\varepsilon_t^2 \sum_{j=0}^{t-1} \frac{\partial \phi_j(\theta_0)}{\partial \theta_i} \varepsilon_{t-j} \frac{\partial r_t(\theta_0)}{\partial \theta_i} \leq c \qquad \text{(A.16)}$$

by Assumption 1 and Lemma S.2.

Next, for (A.13) we apply the Cauchy-Schwarz inequality, so the i'th element is

$$(A.13)_i \leq 2 \left(T^{-1} \sum_{t=1}^{T} v_{it}^2 \right)^{1/2} \left(\sum_{t=1}^{T} (\hat{\sigma}_t^{-2} - \tilde{\sigma}_t^{-2})^2 \right)^{1/2},$$

where $T^{-1}\sum_{t=1}^T v_{it}^2 = O_p(1)$ by (A.16) and $\sum_{t=1}^T (\hat{\sigma}_t^{-2} - \tilde{\sigma}_t^{-2})^2 \leq (\min_{1\leq s\leq T} \hat{\sigma}_s^2 \tilde{\sigma}_s^2)^{-2} \sum_{t=1}^T (\tilde{\sigma}_t^2 - \hat{\sigma}_t^2)^2 = O_p(T^{-1/2}b^{-1}) \xrightarrow{p} 0$ by Lemma A.1(c),(g) and Assumption 7.

Finally, we decompose the i'th element of (A.14) as

$$(A.14)_i = 2T^{-1/2} \sum_{t=1}^{T} v_{it} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2) \bar{\sigma}_t^{-4} + 2T^{-1/2} \sum_{t=1}^{T} v_{it} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 \tilde{\sigma}_t^{-2} \bar{\sigma}_t^{-4}. \qquad \text{(A.17)}$$

For the second term on the right-hand side we apply the Cauchy-Schwarz inequality,

$$2T^{-1/2} \sum_{t=1}^{T} v_{it} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 \tilde{\sigma}_t^{-2} \bar{\sigma}_t^{-4} \leq 2 (\min_{1\leq s\leq T} \bar{\sigma}_s^4 \tilde{\sigma}_s^2)^{-1} \left(T^{-1} \sum_{t=1}^{T} v_{it}^2 \right)^{1/2} \left(\sum_{t=1}^{T} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2)^4 \right)^{1/2},$$

where the first two terms are $O_p(1)$ by Lemma A.1(g) and (A.16), while the last term is $o_p(1)$ by Lemma A.1(e) and Assumption 7. Using the decomposition (A.4), the first term on the right-hand side of (A.17) is

$$2T^{-1/2} \sum_{t=1}^{T} v_{it} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2) \bar{\sigma}_t^{-4} = 2T^{-1/2} \sum_{t=1}^{T} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2) \bar{\sigma}_t^{-4} \sum_{n=1}^{t-1} \frac{\partial \phi_n(\theta_0)}{\partial \theta_i} \varepsilon_t \varepsilon_{t-n} \qquad \text{(A.18)}$$

$$+ 2T^{-1/2} \sum_{t=1}^{T} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2) \bar{\sigma}_t^{-4} \varepsilon_t \frac{\partial r_t(\theta_0)}{\partial \theta_i}. \qquad \text{(A.19)}$$

For (A.19) we apply Hölder's inequality and find

$$(A.19) \leq 2T^{-1/2} \left(\sum_{t=1}^{T} \bar{\sigma}_t^{-8} (\bar{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 \right)^{1/2} \left(\sum_{t=1}^{T} \varepsilon_t^4 \right)^{1/4} \left(\sum_{t=1}^{T} \left(\frac{\partial r_t(\theta_0)}{\partial \theta_i} \right)^4 \right)^{1/4}$$

$$= 2T^{-1/2} O_p(b^{-1/2}) O_p(T^{1/4}) O_p(1) = O_p(T^{-1/4} b^{-1/2}) = o_p(1)$$

by Lemmas A.1(b),(g) and S.2 and Assumptions 1,7. Next, (A.18) has second moment

$$E(A.18)^2 = 4T^{-1}\sum_{t,s=1}^{T}\sum_{j_1,j_2=1}^{T}\sum_{n=1}^{t-1}\sum_{m=1}^{s-1}\bar{\sigma}_t^{-4}\bar{\sigma}_s^{-4}\sigma_t\sigma_s\sigma_{t-n}\sigma_{s-m}\sigma_{j_1}^2\sigma_{j_2}^2 k_{tj_1}k_{sj_2}\frac{\partial\phi_n(\theta_0)}{\partial\theta_i}\frac{\partial\phi_m(\theta_0)}{\partial\theta_i}$$
$$\times E(z_t z_s z_{t-n} z_{s-m}(z_{j_1}^2-1)(z_{j_2}^2-1)).$$

By symmetry, we assume $t \geq s$ and $j_1 \geq j_2$ such that also $t > t-n$ and $t > s-m$, which by Lemma S.4 leaves two possibilities: (i) $t = s \geq j_1$ and (ii) $j_1 \geq t$. The proofs for these cases are nearly identical, so we prove only the first case. Here we find that $E(z_t^2 z_{t-n} z_{t-m}(z_{j_1}^2-1)(z_{j_2}^2-1))$ is a combination of cumulants. When the expectation is a $\kappa_8(\cdot)$ cumulant, we eliminate the summations over n, m, j_1, j_2 by Assumption 1(a)(iii) and the contribution to the second moment is $T^{-1}\sum_{t=1}^{T}\sup_{1\leq j_1,j_2\leq T} k_{tj_1}k_{tj_2} \leq c(Tb)^{-2}$ because

$$\sup_{1\leq j\leq T, 1\leq t\leq T}|k_{tj}| = \sup_{1\leq j\leq T, 1\leq t\leq T}\frac{|K(\frac{t-j}{Tb})|\frac{1}{Tb}}{|\sum_{n=1}^{T}K(\frac{t-n}{Tb})\frac{1}{Tb}|} \leq c\frac{1}{Tb} \tag{A.20}$$

by boundedness and integrability of $K(\cdot)$, see Assumption 6. When the expectation is a $\kappa_2(\cdot)\kappa_6(\cdot)$ product or a $\kappa_4(\cdot)\kappa_4(\cdot)$ product, 3 summations are eliminated and the contribution to the second moment is $O((Tb)^{-1})$. Thus, $(A.18) \xrightarrow{p} 0$ by Assumption 7, which shows that $(A.17) \xrightarrow{p} 0$ and hence proves the first statement of (A.10).

To prove the second statement of (A.10) we find, as in (A.5)–(A.8),

$$\begin{aligned}
\sqrt{T}\frac{\partial\bar{Q}_T(\theta_0)}{\partial\theta_i} - \sqrt{T}\frac{\partial\bar{Q}_T^0(\theta_0)}{\partial\theta_i} =& 2T^{-1/2}\sum_{t=1}^{T}\sigma_t^{-2}(\varepsilon_t + r_t)\left(\sum_{j=0}^{t-1}\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}\varepsilon_{t-j} + \frac{\partial r_t(\theta_0)}{\partial\theta_i}\right)\\
&- 2T^{-1/2}\sum_{t=1}^{T}(\sigma_t^{-1}\varepsilon_t + \tilde{r}_t)\left(\sum_{j=0}^{t-1}\sigma_{t-j}^{-1}\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}\varepsilon_{t-j} + \frac{\partial\tilde{r}_t(\theta_0)}{\partial\theta_i}\right)\\
=& 2T^{-1/2}\sum_{t=1}^{T}\sum_{j=0}^{t-1}\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}z_t z_{t-j}\left(\frac{\sigma_{t-j}}{\sigma_t} - 1\right) && \text{(A.21)}\\
&+ 2T^{-1/2}\sum_{t=1}^{T}z_t\left(\sigma_t^{-1}\frac{\partial r_t(\theta_0)}{\partial\theta_i} - \frac{\partial\tilde{r}_t(\theta_0)}{\partial\theta_i}\right) && \text{(A.22)}\\
&+ 2T^{-1/2}\sum_{t=1}^{T}\left(\sigma_t^{-2}r_t\sum_{j=0}^{t-1}\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}\varepsilon_{t-j} - \tilde{r}_t\sum_{j=0}^{t-1}\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}z_{t-j}\right) && \text{(A.23)}\\
&+ 2T^{-1/2}\sum_{t=1}^{T}\left(\sigma_t^{-2}r_t\frac{\partial r_t(\theta_0)}{\partial\theta_i} - \tilde{r}_t\frac{\partial\tilde{r}_t(\theta_0)}{\partial\theta_i}\right). && \text{(A.24)}
\end{aligned}$$

where $\tilde{r}_t$ is defined in the same way as r_t but with z_t replacing ε_t. The last three terms on the right-hand side are all easily shown to be $o_p(1)$ using either L_1- or L_2-convergence and applying the bounds in Lemma S.2. For example, for (A.22) we find

$$T^{-1/2}\sum_{t=1}^{T}E|z_t\frac{\partial r_t(\theta_0)}{\partial\theta_i}| \leq T^{-1/2}\sum_{t=1}^{T}(Ez_t^2)^{1/2}(E(\frac{\partial r_t(\theta_0)}{\partial\theta_i})^2)^{1/2} \leq cT^{-1/2}\sum_{t=1}^{T}(\log t)t^{-1} \leq cT^{-1/2}(\log T)^2 \to 0.$$

We are left with (A.21), which for $j = 0$ is zero and otherwise has second moment

$$
\begin{aligned}
E(A.21)^2 =&4T^{-1}\sum_{t=1}^{T}\sum_{j,k=1}^{t-1}\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}\frac{\partial\phi_k(\theta_0)}{\partial\theta_i}E(z_t^2 z_{t-j}z_{t-k})\left(\frac{\sigma_{t-j}}{\sigma_t}-1\right)\left(\frac{\sigma_{t-k}}{\sigma_t}-1\right)\\
=&4T^{-1}\sum_{t=1}^{T}\sum_{j,k=1}^{t-1}\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}\frac{\partial\phi_k(\theta_0)}{\partial\theta_i}\kappa_4(t,t,t-j,t-k)\left(\frac{\sigma_{t-j}}{\sigma_t}-1\right)\left(\frac{\sigma_{t-k}}{\sigma_t}-1\right)\\
&+4T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1}\left(\frac{\partial\phi_j(\theta_0)}{\partial\theta_i}\right)^2\left(\frac{\sigma_{t-j}}{\sigma_t}-1\right)^2.
\end{aligned}
$$

The last term converges to zero by Lemma S.1 after reversing summations and setting $a_j = (\partial\phi_j(\theta_0)/\partial\theta_i)^2 \leq c(\log j)^2 j^{-2}$ (by Lemma S.2) and $b_{j,T} = T^{-1}\sum_{t=j+1}^{T}(\sigma_{t-j}/\sigma_t - 1)^2 \leq cT^{-1}\sum_{t=j+1}^{T}(\sigma_{t-j}-\sigma_t)^2$ (using Assumption 1(b)), which satisfy the assumptions of Lemma S.1 by Cavaliere and Taylor (2009, Lemma A.1). For the first term we find the bound $c\sum_{j,k=1}^{T-1}(\sup_t|\kappa_4(t,t,t-j,t-k)|)T^{-1}\sum_{t=\max(j,k)+1}^{T}|\sigma_t-\sigma_{t-j}||\sigma_t-\sigma_{t-k}| \to 0$ again by Lemma S.1 in view of Assumption 1(a)(iii). This concludes the proof of the second statement of (A.10) and hence that of (11).

Appendix B: Proof of Theorem 2

We first consider

$$
\begin{aligned}
\hat{A} =&\frac{1}{4}T^{-1}\sum_{t=1}^{T}\frac{\partial q_t(\hat\theta)}{\partial\theta}\frac{\partial q_t(\hat\theta)}{\partial\theta'} = \frac{1}{4}T^{-1}\sum_{t=1}^{T}\hat\sigma_t^{-4}\frac{\partial\varepsilon_t(\hat\theta)^2}{\partial\theta}\frac{\partial\varepsilon_t(\hat\theta)^2}{\partial\theta'}\\
=&\frac{1}{4}T^{-1}\sum_{t=1}^{T}(\hat\sigma_t^{-4}-\sigma_t^{-4})\frac{\partial\varepsilon_t(\hat\theta)^2}{\partial\theta}\frac{\partial\varepsilon_t(\hat\theta)^2}{\partial\theta'} && \text{(B.1)}\\
&+T^{-1}\sum_{t=1}^{T}\sigma_t^{-4}\left(\varepsilon_t(\hat\theta)^2\frac{\partial\varepsilon_t(\hat\theta)}{\partial\theta}\frac{\partial\varepsilon_t(\hat\theta)}{\partial\theta'}-\varepsilon_t(\theta_0)^2\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right) && \text{(B.2)}\\
&+T^{-1}\sum_{t=1}^{T}\sigma_t^{-4}\varepsilon_t(\theta_0)^2\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}. && \text{(B.3)}
\end{aligned}
$$

By the Cauchy-Schwarz inequality, the (i,j)'th element of (B.1) satisfies

$$
|(B.1)_{i,j}| \leq \frac{1}{4}\left(T^{-1}\sum_{t=1}^{T}(\hat\sigma_t^{-4}-\sigma_t^{-4})^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left(\frac{\partial\varepsilon_t(\hat\theta)^2}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat\theta)^2}{\partial\theta_j}\right)^2\right)^{1/2},
$$

where the first term is $o_p(1)$ by Assumption 1(b) and Lemma A.1(d),(f),(g) because $(\hat\sigma_t^{-4}-\sigma_t^{-4}) = \hat\sigma_t^{-4}\sigma_t^{-4}(\sigma_t^2+\hat\sigma_t^2)(\sigma_t^2-\hat\sigma_t^2)$, and the last term is $O_p(1)$ by the uniform convergence in Lemma S.3 combined with consistency of $\hat\theta$; see, e.g., Johansen and Nielsen (2010, Lemma A.3).

Next, we decompose the (i,j)'th element of (B.2) and apply the Cauchy-Schwarz inequality,

$$T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}(\varepsilon_t(\hat{\theta})^2-\varepsilon_t(\theta_0)^2)\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_j}+T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}\varepsilon_t(\theta_0)^2\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_j}-\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta_i}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta_j}\right)$$

$$\leq\left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}(\varepsilon_t(\hat{\theta})^2-\varepsilon_t(\theta_0)^2)^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_j}\right)^2\right)^{1/2} \tag{B.4}$$

$$+\left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}\varepsilon_t(\theta_0)^4\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_j}-\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta_i}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta_j}\right)^2\right)^{1/2}. \tag{B.5}$$

The proofs for (B.4) and (B.5) are nearly identical, so we give only the former. The second large parenthesis in (B.4) is $O_p(1)$ by Lemma S.3 and Assumption 1(b)(ii). By the mean value theorem,

$$T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\hat{\theta})^2-\varepsilon_t(\theta_0)^2)^2=4\sum_{i=1}^{p+1}(\hat{\theta}_i-\theta_{0,i})T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\bar{\theta})^2-\varepsilon_t(\theta_0)^2)\varepsilon_t(\bar{\theta})\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta_i}$$

for an intermediate value, $\bar{\theta}$, between $\hat{\theta}$ and θ_0. By another application of the Cauchy-Schwarz inequality,

$$T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\bar{\theta})^2-\varepsilon_t(\theta_0)^2)\varepsilon_t(\bar{\theta})\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta_i}\leq(T^{-1}\sum_{t=1}^{T}(\varepsilon_t(\bar{\theta})^2-\varepsilon_t(\theta_0)^2)^2)^{1/2}(T^{-1}\sum_{t=1}^{T}\varepsilon_t(\bar{\theta})^2(\frac{\partial\varepsilon_t(\bar{\theta})}{\partial\theta_i})^2)^{1/2},$$

which is also $O_p(1)$ by Lemma S.3. Because $\hat{\theta}_i-\theta_{0,i}=O_p(T^{-1/2})$ by Theorem 1 and using Assumption 1(b)(ii), it follows that (B.4) is $o_p(1)$. Next, $(B.3)\xrightarrow{p}A_0$ by the same arguments as applied to the second term of (A.10), and it follows that $\hat{A}\xrightarrow{p}A_0$.

Finally, we find that

$$\hat{B}=\frac{1}{2}\frac{\partial^2\hat{Q}_T(\hat{\theta})}{\partial\theta\partial\theta'}\xrightarrow{p}B_0$$

by the uniform convergence in (A.9) combined with consistency of $\hat{\theta}$; see, e.g., Johansen and Nielsen (2010, Lemma A.3). It now follows straightforwardly, using Slutsky's Theorem and Assumption 5, that $\hat{C}=\hat{B}^{-1}\hat{A}\hat{B}^{-1}\xrightarrow{p}B_0^{-1}A_0B_0^{-1}=C_0$.

Appendix C: Proof of Theorem 3

We first give the bootstrap equivalent of Lemma A.1, the proof of which is given in Section S.2.2.

Lemma C.1. *Define* $\tilde{\sigma}_t^{*2}:=\sum_{i=1}^{T}k_{ti}\varepsilon_i^{*2}$. *Under Assumptions 1–7 it holds that, in probability:*
(b)* $T^{-1}\sum_{t=1}^{T}(\tilde{\sigma}_t^{*2}-\hat{\sigma}_t^2)^2=O_p^*((Tb)^{-1})$,
(c)* $T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^{*2}-\tilde{\sigma}_t^{*2})^2=O_p^*(T^{-3/2}b^{-1})$,
(e)* $T^{-1}\sum_{t=1}^{T}(\tilde{\sigma}_t^{*2}-\hat{\sigma}_t^2)^4=O_p^*(T^{-2}b^{-3})$,
(f)* $\max_{1\leq t\leq T}\hat{\sigma}_t^{*2}=O_p^*(1)$,
(g)* $(\min_{1\leq t\leq T}\hat{\sigma}_t^{*2})^{-1}=O_p^*(1)$ *and* $(\min_{1\leq t\leq T}\tilde{\sigma}_t^{*2})^{-1}=O_p^*(1)$.

We next give two results which are applied several times in the proof of Theorem 3.

Lemma C.2 (CNT, Lemma D.1)**.** *Under the assumptions of Theorem 3,*

$$T^{-1}\sum_{t=1}^{T}(\hat{\varepsilon}_t^2-\varepsilon_t^2)^2=O_p(T^{-1/2}).$$

Lemma C.3. *Suppose the conditions of Theorem 3 are satisfied. Suppose also that the coefficients $\lambda_j(\theta)$ satisfy $\sup_\theta |\lambda_j(\theta)| = O_p(j^g)$ and $\sup_\theta |\lambda_{j+1}(\theta) - \lambda_j(\theta)| = O_p(j^{g-1})$, 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 $1 \leq m \leq k \leq T$,*

$$
\begin{aligned}
&E^* \sup_\theta \left| \sum_{j=m}^{k} \lambda_j(\theta) \sum_{t=\max(j,h)+1}^{T} \frac{1}{\hat{\sigma}_t} \varepsilon^*_{t-j}\varepsilon^*_{t-h} \right| \\
&\quad = \mathbb{I}(g > -1/2) O_p(T^{1/2}k^{1/2+g}) + \mathbb{I}(g < -1/2) O_p(T^{1/2}m^{1/2+g}) + \mathbb{I}(g = -1/2) O_p(T^{1/2}(\log k)).
\end{aligned}
$$

Proof. The proof follows by Lemma A.1(g) and Lemma D.2 of CNT. □

C.1 Proof of consistency

As in the proof of consistency in Theorem 1, see Section A.1, we partition the parameter space into two disjoint sets, this time depending on the bootstrap true value, $\hat{d}$. That is, we define $\hat{D}_1 := D \cap \{d : d - \hat{d} \leq -1/2 + \kappa\}$ and $\hat{D}_2 := D \cap \{d : -1/2 + \kappa \leq d - \hat{d}\}$. Note that these sets are random and depend on T since $\hat{d}$ is random and depends on T. This presents an additional complication, so we will need also $D_1^0 := D \cap \{d : d - d_0 \leq -1/2 + 2\kappa\}$ and $D_2^0 := D \cap \{d : -1/2 + \kappa/2 \leq d - d_0\}$, which are non-random and do not depend on T. Analogously to Θ_i, we define $\hat{\Theta}_i := \hat{D}_i \times \Psi$ and $\Theta_i^0 := D_i^0 \times \Psi$ for $i = 1, 2$. Note that the D_i^0 are defined such that, by definition of d_0,

$$P(D_1^0 \supseteq \hat{D}_1) = P(\hat{d} - d_0 \leq \kappa) \to 1, \tag{C.1}$$

$$P(D_2^0 \supseteq \hat{D}_2) = P(d_0 - \hat{d} \leq \kappa/2) \to 1. \tag{C.2}$$

The general strategy of the proof relies on analyzing these parts of the parameter space separately, as was also the case in the proof of consistency in Theorem 1. First, it is shown that for any $K > 0$ there exists a (fixed) $\bar{\kappa} > 0$ such that

$$P^*\Big(\inf_{\theta \in \hat{\Theta}_1(\bar{\kappa})} \hat{Q}^*_T(\theta) > K\Big) \xrightarrow{p} 1 \text{ as } T \to \infty, \tag{C.3}$$

$$P^*\Big(\inf_{\theta \in \hat{\Theta}_1(\bar{\kappa})} \check{Q}^*_T(\theta) > K\Big) \xrightarrow{p} 1 \text{ as } T \to \infty. \tag{C.4}$$

This implies that $P^*(\hat{\theta}^* \in \hat{\Theta}_2(\bar{\kappa})) \xrightarrow{p} 1$ and $P^*(\check{\theta}^* \in \hat{\Theta}_2(\bar{\kappa})) \xrightarrow{p} 1$ as $T \to \infty$, so that the relevant parameter space is reduced to $\hat{\Theta}_2(\bar{\kappa})$. As in the proofs of (A.1) and (A.2), the results (C.3) and (C.4) follow from the bounds in (D.12) of CNT and Lemma C.1(g*).

In view of (C.3) and (C.4), it follows that $\hat{\theta}^* - \check{\theta}^* \xrightarrow{p^*}_p 0$ because

$$
\begin{aligned}
\sup_{\theta \in \hat{\Theta}_2} |\hat{Q}^*_T(\theta) - \check{Q}^*_T(\theta)| &\leq \sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{t=1}^{T} \varepsilon^*_t(\theta)^2 \frac{|\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2}|}{\hat{\sigma}_t^2 \hat{\sigma}_t^{*2}} \\
&\leq (\min_t \hat{\sigma}_t^2 \hat{\sigma}_t^{*2})^{-1} \sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{t=1}^{T} \varepsilon^*_t(\theta)^2 |\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2}| \\
&\leq (\min_t \hat{\sigma}_t^2 \hat{\sigma}_t^{*2})^{-1} \left(\sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{t=1}^{T} \varepsilon^*_t(\theta)^4 \right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2})^2 \right)^{1/2},
\end{aligned}
\tag{C.5}
$$

which is $O_p^*(1) O_p^*(1) o_p^*(1) = o_p^*(1)$, in probability, by Lemmas A.1(g), C.1(g*), S.8, and C.1(b*),(c*).

Thus, we proceed with $\check{\theta}^*$. We define also $\tilde{\theta}^\dagger := \arg\min_{\theta \in \hat{\Theta}_2} \hat{Q}_T(\theta)$, which satisfies

$$\hat{\theta} - \tilde{\theta}^\dagger \xrightarrow{p} 0 \tag{C.6}$$

because $P(|\hat{\theta} - \tilde{\theta}^\dagger| > \epsilon) = P(\hat{\theta} \notin \hat{\Theta}_2) = 0$ by definition of $\hat{\Theta}_2$.

Next, we prove that

$$\arg\min_{\theta\in\hat\Theta_2} \check Q_T^*(\theta) - \tilde\theta^\dagger \xrightarrow{p^*}_p 0. \tag{C.7}$$

With P^*-probability converging to one in probability, the first term in (C.7) is $\check\theta^*$, see (C.4), so that the required result follows by combining (C.6) and (C.7). We therefore prove that (for any $\kappa>0$)

$$\sup_{\theta\in\hat\Theta_2} |\check Q_T^*(\theta) - \hat Q_T(\theta)| \xrightarrow{p^*}_p 0, \tag{C.8}$$

which implies (C.7).

To show (C.8) we decompose

$$\begin{aligned}\check Q_T^*(\theta) - \hat Q_T(\theta) =& \check Q_T^*(\theta) - E^*\check Q_T^*(\theta) &\text{(C.9)}\\ &+ E^*\check Q_T^*(\theta) - \hat Q_T(\theta) &\text{(C.10)}\end{aligned}$$

and write $\varepsilon_t^*(\theta) = \sum_{n=0}^{t-1}\hat\phi_n(\theta)\varepsilon_{t-n}^*$, where $\sup_{\theta\in\hat\Theta_2}|\hat\phi_n(\theta)| = O_p(n^{-1/2-\kappa})$, uniformly in n, by Lemma S.7. By uncorrelatedness of ε_t^* conditional on the original data,

$$\begin{aligned}(C.9) =& T^{-1}\sum_{t=1}^T \frac{1}{\hat\sigma_t^2}\sum_{n=0}^{t-1}\hat\phi_n(\theta)^2(\varepsilon_{t-n}^{*2} - \hat\varepsilon_{t-n}^2) &\text{(C.11)}\\ &+ 2T^{-1}\sum_{t=1}^T \frac{1}{\hat\sigma_t^2}\sum_{n=0}^{t-1}\sum_{m=n+1}^{t-1}\hat\phi_n(\theta)\hat\phi_m(\theta)\varepsilon_{t-n}^*\varepsilon_{t-m}^*. &\text{(C.12)}\end{aligned}$$

Noting that, conditionally on the original sample, $\varepsilon_t^{*2} - \hat\varepsilon_t^2 = \hat\varepsilon_t^2(w_t^2-1)$ is a martingale difference sequence, it follows that, defining $\eta_4 := E((w_t^2-1)^2)$,

$$\begin{aligned}\left(E^*\left|\sum_{t=n+1}^T \frac{1}{\hat\sigma_t^2}(\varepsilon_{t-n}^{*2} - \hat\varepsilon_{t-n}^2)\right|\right)^2 &\le \sum_{t,s=n+1}^T \frac{1}{\hat\sigma_t^2\hat\sigma_s^2}E^*(\varepsilon_{t-n}^{*2} - \hat\varepsilon_{t-n}^2)(\varepsilon_{s-n}^{*2} - \hat\varepsilon_{s-n}^2)\\ &= \sum_{t=n+1}^T \frac{1}{\hat\sigma_t^4}E^*(\varepsilon_{t-n}^{*2} - \hat\varepsilon_{t-n}^2)^2\\ &\le \frac{1}{\min_t\hat\sigma_t^4}\eta_4\sum_{t=n+1}^T \hat\varepsilon_{t-n}^4 =: C_T^2 = O_p(T) &\text{(C.13)}\end{aligned}$$

uniformly in $0\le n\le T-1$ by Lemmas A.1(g) and S.3. Thus, reversing the order of the summations in (C.11) and using (C.13), we find

$$\begin{aligned}E^*\sup_{\theta\in\hat\Theta_2}|(C.11)| &\le \sup_{\theta\in\hat\Theta_2} T^{-1}\sum_{n=0}^{T-1}\hat\phi_n(\theta)^2E^*\left|\sum_{t=n+1}^T\frac{1}{\hat\sigma_t^2}(\varepsilon_{t-n}^{*2} - \hat\varepsilon_{t-n}^2)\right|\\ &\le C_T\sup_{\theta\in\hat\Theta_2} T^{-1}\sum_{n=0}^{T-1}\hat\phi_n(\theta)^2 \le C_T O_p(T^{-1}\sum_{n=0}^{T-1}n^{-1-2\kappa}) = O_p(T^{-1/2}),\end{aligned}$$

which shows that $\sup_{\theta\in\hat\Theta_2}|(C.11)| = O_p^*(T^{-1/2})$, in probability.

To deal with (C.12), we apply Lemmas A.1(g) and C.3 with $g=-1/2-\kappa$,

$$E^*\sup_{\theta\in\hat\Theta_2}\left|\sum_{m=n+1}^{T-1}\hat\phi_m(\theta)\sum_{t=m+1}^T\frac{1}{\hat\sigma_t^2}\varepsilon_{t-n}^*\varepsilon_{t-m}^*\right| = O_p(T^{1/2}n^{-\kappa}).$$

It follows that

$$E^* \sup_{\theta\in\hat\Theta_2} |(C.12)| = \sup_{\theta\in\hat\Theta_2} T^{-1}\sum_{n=0}^{T-1} |\hat\phi_n(\theta)| O_p(T^{1/2}n^{-\kappa}) = O_p(T^{-1/2})\sum_{n=0}^{T-1} n^{-1/2-2\kappa}$$
$$= O_p((\log T)T^{\max(-1/2,-2\kappa)}),$$

such that $\sup_{\theta\in\hat\Theta_2}|(C.12)| = o_p^*(1)$, in probability.

It remains to analyze (C.10), for which we find

$$E^*Q_T^*(\theta) - \hat Q_T(\theta) = T^{-1}\sum_{t=1}^{T}\frac{1}{\hat\sigma_t^2}\sum_{n=0}^{t-1}(\hat\phi_n(\theta)^2\hat\varepsilon_{t-n}^2 - \phi_n(\theta)^2\varepsilon_{t-n}^2) \tag{C.14}$$
$$- 2T^{-1}\sum_{t=1}^{T}\frac{1}{\hat\sigma_t^2}\sum_{n=0}^{\infty}\sum_{m=n+1}^{\infty}\phi_n(\theta)\phi_m(\theta)\varepsilon_{t-n}\varepsilon_{t-m}. \tag{C.15}$$

By identical arguments to those in the proof of (A.6), the term (C.15) is $o_p(1)$, uniformly in $\theta \in \Theta_2^0$ and $P(\Theta_2^0 \supseteq \hat\Theta_2) \to 1$ by (C.2). We therefore proceed with (C.14), which is

$$(C.14) = T^{-1}\sum_{t=1}^{T}\frac{1}{\hat\sigma_t^2}\sum_{n=0}^{t-1}\phi_n(\theta)^2(\hat\varepsilon_{t-n}^2 - \varepsilon_{t-n}^2) \tag{C.16}$$
$$+ T^{-1}\sum_{t=1}^{T}\frac{1}{\hat\sigma_t^2}\sum_{n=0}^{t-1}(\hat\phi_n(\theta)^2 - \phi_n(\theta)^2)\hat\varepsilon_{t-n}^2. \tag{C.17}$$

For (C.16) we apply the Cauchy-Schwarz inequality and find

$$(C.16)^2 \le \left(T^{-1}\sum_{t=1}^{T}\left(\frac{1}{\hat\sigma_t^2}\sum_{n=0}^{t-1}\phi_n(\theta)^2\right)^2\right)\left(T^{-1}\sum_{t=1}^{T}(\hat\varepsilon_{t-n}^2 - \varepsilon_{t-n}^2)^2\right),$$

where the term in the second parenthesis is $o_p(1)$ by Lemma C.2. From Lemmas A.1(g) and S.2, the term in the first parenthesis is bounded, uniformly in $\theta \in \Theta_2^0$, by $O_p(1)T^{-1}\sum_{t=1}^{T}(\sum_{n=0}^{t-1} n^{-1-\kappa})^2 = O_p(1)$. Because $P(\Theta_2^0 \supseteq \hat\Theta_2) \to 1$, see (C.2), this bound applies also uniformly in $\theta \in \hat\Theta_2$. Finally, for the term (C.17) we note that, by the mean value theorem,

$$|\hat\phi_n(\theta)-\phi_n(\theta)| \le |\hat d-d_0|\sum_{j=0}^{n}\sum_{m=0}^{n-j} b_m(\psi)\dot\pi_j(\bar d-d)a_{n-j-m}(\hat\psi) + |\hat\psi-\psi_0|\sum_{j=0}^{n}\sum_{m=0}^{n-j} b_m(\psi)\pi_j(\hat d-d)\dot a_{n-j-m}(\bar\psi),$$

so that $\sup_{\theta\in\hat\Theta_2}|\hat\phi_n(\theta) - \phi_n(\theta)| = O_p(T^{-1/2}n^{-1/2-\kappa}(\log n))$ using (2), Lemmas S.5,S.6, and $\hat\theta - \theta_0 = O_p(T^{-1/2})$ by Theorem 1, and noting that $P(\Theta_2^0 \supseteq \hat\Theta_2) \to 1$, see (C.2). Thus, by reversing the order of the summations such that $|(C.17)| \le (\min_t \hat\sigma_t^2)^{-1}\sum_{n=0}^{T-1}|\hat\phi_n(\theta) + \phi_n(\theta)||\hat\phi_n(\theta) - \phi_n(\theta)|T^{-1}\sum_{t=n+1}^{T}\hat\varepsilon_{t-n}^2$, we find that

$$\sup_{\theta\in\hat\Theta_2}|(C.17)| \le (\min_t \hat\sigma_t^2)^{-1}\sum_{n=0}^{T-1} O_p(n^{-1/2-\kappa})O_p(T^{-1/2}n^{-1/2-\kappa}(\log n))T^{-1}\sum_{t=n+1}^{T}\hat\varepsilon_{t-n}^2 = O_p(T^{-1/2}(\log T))$$

because $T^{-1}\sum_{t=n+1}^{T}\hat\varepsilon_{t-n}^2 \le T^{-1}\sum_{t=1}^{T}\hat\varepsilon_t^2 = O_p(1)$.

C.2 Proof of asymptotic normality

Following roughly the same steps as in the proof of asymptotic normality in Theorem 1, see Section A.2, we first prove the asymptotic first-order equivalence of $\hat{\theta}^*$ and $\check{\theta}^*$ by showing

$$\sup_{\theta \in \mathcal{N}(\hat{\theta})} \left| \frac{\partial^2 \hat{Q}_T^*(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \check{Q}_T^*(\theta)}{\partial \theta \partial \theta'} \right| \xrightarrow{p^*}_p 0, \tag{C.18}$$

$$\sqrt{T} \frac{\partial \hat{Q}_T^*(\hat{\theta})}{\partial \theta} - \sqrt{T} \frac{\partial \check{Q}_T^*(\hat{\theta})}{\partial \theta} \xrightarrow{p^*}_p 0. \tag{C.19}$$

The proof of (C.18) is identical to that of (C.5) recognizing that the derivatives add at most a logarithmic factor, see Lemma S.6. The proof of (C.19) is identical to that of (A.13)–(A.14) with appropriate adjustments to take into account the bootstrap errors, i.e., replacing $\tilde{\sigma}_t^2$ and $\bar{\sigma}_t^2$ with $\tilde{\sigma}_t^{*2}$ and $\hat{\sigma}_t^2$, respectively, and using Lemma C.1 and independence of the w_t sequence.

Next, we define

$$\bar{\theta}^* := \arg\min_{\theta} \bar{Q}_T^*(\theta) \text{ with } \bar{Q}_T^*(\theta) := T^{-1} \sum_{t=1}^{T} \left(\sum_{n=0}^{t-1} b_n(\psi) \sum_{j=0}^{t-n-1} \pi_j(\hat{d} - d) \sum_{m=0}^{\infty} a_m(\hat{\psi}) \frac{\varepsilon^*_{t-n-j-m}}{\sigma_{t-n-j-m}} \right)^2$$

as well as the objective function $\bar{Q}_T^{*0}(\theta) := T^{-1} \sum_{t=1}^{T} \varepsilon_t^*(\theta)^2 / \sigma_t^2$. From Theorem 6 of CNT it holds that $\sqrt{T}(\bar{\theta}^* - \hat{\theta}) \xrightarrow{w}_p N(0, C_0)$, so it is sufficient to show that

$$\sup_{\theta \in \mathcal{N}(\hat{\theta})} \left| \frac{\partial^2 \check{Q}_T^*(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \bar{Q}_T^{*0}(\theta)}{\partial \theta \partial \theta'} \right| \xrightarrow{p^*}_p 0 \text{ and } \sup_{\theta \in \mathcal{N}(\hat{\theta})} \left| \frac{\partial^2 \bar{Q}_T^{*0}(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \bar{Q}_T^*(\theta)}{\partial \theta \partial \theta'} \right| \xrightarrow{p^*}_p 0 \tag{C.20}$$

$$\sqrt{T} \frac{\partial \check{Q}_T^*(\hat{\theta})}{\partial \theta} - \sqrt{T} \frac{\partial \bar{Q}_T^{*0}(\hat{\theta})}{\partial \theta} \xrightarrow{p^*}_p 0 \text{ and } \sqrt{T} \frac{\partial \bar{Q}_T^{*0}(\hat{\theta})}{\partial \theta} - \sqrt{T} \frac{\partial \bar{Q}_T^*(\hat{\theta})}{\partial \theta} \xrightarrow{p^*}_p 0, \tag{C.21}$$

from which the result follows by the triangle inequality. The proofs of (C.20) and (C.21) follow nearly identically to those of (A.3) and (A.10), respectively, but are simpler because ε_t^* is independent, conditionally on the data, and $\varepsilon_t^* = 0$ for $t \leq 0$. Specifically, in the bootstrap case we use Lemmas S.7 and S.8 instead of Lemmas S.2 and S.3 and note that there is no $r_t(\theta)$ or r_t remainders in the bootstrap case because $\varepsilon_t^* = 0$ for $t \leq 0$. Finally, it follows from Section D.2.1 of CNT that $\sqrt{T} \partial \bar{Q}_T^*(\hat{\theta}) / \partial \theta \xrightarrow{w}_p N(0, 4A_0^\dagger)$.

C.3 Proof of consistency of bootstrap variance estimator

We define

$$\hat{A}^* := \frac{1}{4} T^{-1} \sum_{t=1}^{T} \hat{\sigma}_t^{*-4} \frac{\partial \varepsilon_t^*(\hat{\theta})^2}{\partial \theta} \frac{\partial \varepsilon_t^*(\hat{\theta})^2}{\partial \theta'} \text{ and } \check{A}^* := \frac{1}{4} T^{-1} \sum_{t=1}^{T} \hat{\sigma}_t^{-4} \frac{\partial \varepsilon_t^*(\hat{\theta})^2}{\partial \theta} \frac{\partial \varepsilon_t^*(\hat{\theta})^2}{\partial \theta'}$$

and also

$$\hat{B}^* := \frac{1}{2} \frac{\partial^2 \hat{Q}_T^*(\hat{\theta}^*)}{\partial \theta \partial \theta'} \text{ and } \check{B}^* := \frac{1}{2} \frac{\partial^2 \check{Q}_T(\check{\theta}^*)}{\partial \theta \partial \theta'}.$$

It follows directly from (C.18) combined with $\hat{\theta}^* - \check{\theta}^* \xrightarrow{p^*}_p 0$ that $\hat{B}^* - \check{B}^* \xrightarrow{p^*}_p 0$; see, e.g., Johansen and Nielsen (2010, Lemma A.3). From (C.20) it then follows that $\check{B}^* - B_0 \xrightarrow{p^*}_p 0$.

By the Cauchy-Schwarz inequality we find that

$$||\hat{A}^* - \check{A}^*||^2 \leq \frac{1}{4} \left(T^{-1} \sum_{t=1}^{T} (\hat{\sigma}_t^{*-4} - \hat{\sigma}_t^{-4})^2 \right) \left(T^{-1} \sum_{t=1}^{T} \left\| \frac{\partial \varepsilon_t^*(\hat{\theta})^2}{\partial \theta} \right\|^4 \right),$$

where the last parenthesis is $O_p^*(1)$, in probability, by Lemma S.8. To bound the first parenthesis we note that

$$\hat{\sigma}_t^{*-4} - \hat{\sigma}_t^{-4} = \frac{\hat{\sigma}_t^4 - \hat{\sigma}_t^{*4}}{\hat{\sigma}_t^{*-4}\hat{\sigma}_t^{-4}} = \frac{(\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2})(\hat{\sigma}_t^2 + \hat{\sigma}_t^{*2})}{\hat{\sigma}_t^{*-4}\hat{\sigma}_t^{-4}}$$

and apply the Cauchy-Schwarz inequality once more,

$$T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^{*-4} - \hat{\sigma}_t^{-4})^2 \leq \left(T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2})^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\frac{(\hat{\sigma}_t^2 + \hat{\sigma}_t^{*2})^2}{\hat{\sigma}_t^{*-8}\hat{\sigma}_t^{-8}}\right)^{1/2},$$

where the term inside the first parenthesis on the right-hand side is $O_p^*(T^{-1}b^{-1}) = o_p^*(1)$, in probability, by Lemma C.1(b*),(c*) and Assumption 7, and the last parenthesis on the right-hand side is $O_p^*(1)$, in probability, by Lemmas A.1(f),(g) and C.1(f*),(g*). It follows that $||\hat{A}^* - \check{A}^*||^2 = o_p^*(1)$, in probability.

The proof that $\check{A}^* - A_0^\dagger \xrightarrow{p^*}_p 0$ is nearly identical to that of $\hat{A} - A_0 \xrightarrow{p} 0$ given in Section Appendix B: and is therefore omitted. The required result now follows by Slutsky's Theorem.

References

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.

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

Carroll, R.J. (1982), Adapting for heteroscedasticity in linear models, *Annals of Statistics* 10, 1224–1233.

Cavaliere, G. and A.M.R. Taylor (2009), Heteroskedastic time series with a unit root, *Econometric Theory* 25, 1228–1270.

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., M.Ø. Nielsen and A.M.R. Taylor (2017), Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form, *Journal of Econometrics* 198, 165–188.

Chang, Y., R.K. Kaufmann, C.S. Kim, J.I. Miller, J.Y. Park, and S. Park (2016), Evaluating trends in time series of distributions: a spatial fingerprint of human effects on climate, Available from `https://economics.missouri.edu/sites/default/files/wp-files/cc1_20161219.pdf`

Eicker, F. (1967), Limit theorems for regression with unequal and dependent errors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley.

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. (1995), Regression with nonstationary volatility, *Econometrica* 63, 1113–1132.

Harris, D. and H. Kew (2017), Adaptive long memory testing under heteroskedasticity, *Econometric Theory* 33, 755–778.

Harvey, A.C. and P.M. Robinson (1988), Efficient estimation of nonstationary time series regression, *Journal of Time Series Analysis* 9, 201–204.

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

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

Huber, P.J. (1967), The behavior of the maximum likelihood estimator under nonstandard conditions, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley.

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

Li, W.K., S. Ling, and M. McAleer (2002), Recent theoretical results for time series models with GARCH errors, *Journal of Economic Surveys* 16, 245–249.

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.

Mammen, E. (1993), Bootstrap and wild bootstrap for high dimensional linear models, *Annals of Statistics* 21, 255–285.

Martins, S. and C. Amado (2016), Modelling sovereign debt contagion: a smooth transition approach. University of Minho working paper.

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.

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.

Pickup, M. and R. Johnston (2008), Campaign trial heats as election forecasts: measurement error and bias in 2004 presidential campaign polls, *International Journal of Forecasting* 24, 272–284.

Robinson, P.M. (1987), Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form, *Econometrica* 55, 875–891.

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.

Rosenblatt, M. (1959), Stationary processes as shifts of functions of independent random variables, *Journal of Mathematics and Mechanics* 8, 665–681.

Samorodnitsky, G. (2007), Long range dependence, *Foundations and Trends in Stochastic Systems* 1, 163–257.

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 (2012), Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* Spring Volume, 81–156.

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

Wong, W.H. (1983), On the consistency of cross validation in kernel nonparametric regression, *Annals of Statistics* 11, 1136–1141.

Wu, C.F.J. (1986), Jackknife bootstrap and other resampling methods in regression analysis, *Annals of Statistics* 14, 1261–1295.

Xu, K.-L. and P.C.B. Phillips (2008), Adaptive estimation of autoregressive models with time-varying variances, *Journal of Econometrics* 142, 265–280.

Xu, K.-L. and P.C.B. Phillips (2011), Tilted nonparametric estimation of volatility functions with empirical applications, *Journal of Business and Economic Statistics* 29, 518–528.

Xu, K.-L. and J.-C. Yang (2015), Towards uniformly efficient trend estimation under weak/strong correlation and non-stationary volatility, *Scandinavian Journal of Statistics* 42, 63–86.

Zhang, T. and W.B. Wu (2012), Inference of time-varying regression models, *Annals of Statistics* 40, 1376–1402.

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

Adaptive Inference in Heteroskedastic Fractional Time Series Models

Supplementary Appendix

by

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

S.1 Introduction

This supplement to our paper "Adaptive Inference in Heteroskedastic Fractional Time Series Models" has two main sections. The first section contains the proofs of Lemmas A.1 and C.1 as well as some technical lemmas and their proofs, which are used to prove the main results.

The second section contains some additional data analysis related to the empirical examples in the paper. Equation references (S.n) for $n \geq 1$ refer to equations in this supplementary appendix and other equation references are to the main paper. Additional references are included at the end of the supplement.

S.2 Lemmas and Proofs of Lemmas

S.2.1 Proof of Lemma A.1

Proof of (a): We have

$$\begin{aligned}
T^{-1}\sum_{t=1}^{T}(\bar{\sigma}_t^2-\sigma_t^2)^2 =& T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T}k_{ti}k_{tj}(\sigma_i^2-\sigma_t^2)(\sigma_j^2-\sigma_t^2) \leq 2G\sum_{i=1}^{T}k_{ti}T^{-1}\sum_{t=1}^{T}|\sigma_i^2-\sigma_t^2| \\
=& 2GT^{-1}\sum_{t=1}^{T}\sum_{i=1}^{t-\lfloor MTb\rfloor}k_{ti}|\sigma_i^2-\sigma_t^2| + 2GT^{-1}\sum_{t=1}^{T}\sum_{i=t-\lfloor MTb\rfloor+1}^{t+\lfloor MTb\rfloor}k_{ti}|\sigma_i^2-\sigma_t^2| \\
&+ 2GT^{-1}\sum_{t=1}^{T}\sum_{i=t+\lfloor MTb\rfloor+1}^{T}k_{ti}|\sigma_i^2-\sigma_t^2| =: A_{1T}+A_{2T}+A_{3T}
\end{aligned}$$

for $G := \sup_{0\leq u\leq 1}\sigma(u)^2 < \infty$ by Assumption 1(b), for some M to be chosen, and using $\sum_{j=1}^{T}k_{tj}=1$. The proofs for A_{1T} and A_{3T} are identical, so we give only the former. In this case we first find

$$\begin{aligned}
\frac{1}{Tb}\sum_{i=1}^{\lfloor Tx\rfloor-\lfloor MTb\rfloor}K\left(\frac{\lfloor Tx\rfloor-i}{Tb}\right) &= \frac{1}{b}\int_{1/T}^{(\lfloor Tx\rfloor-\lfloor MTb\rfloor)/T}K\left(\frac{\lfloor Tx\rfloor-\lfloor Ts\rfloor}{Tb}\right)\mathsf{d}s \\
&\overset{u:=(s-x)/b}{=} \int_{\frac{1/T-x}{b}}^{\frac{\lfloor Tx\rfloor-\lfloor MTb\rfloor-Tx}{Tb}}K(u)\mathsf{d}u+o(1) \\
&\leq \int_{\frac{1/T-1}{b}}^{-\frac{\lfloor MTb\rfloor}{Tb}}K(u)\mathsf{d}u+o(1) \to \int_{-\infty}^{-M}K(u)\mathsf{d}u
\end{aligned}$$

by Assumptions 6 and 7, where the bound is uniform in $x \in [0,1]$ and can be made arbitrarily small by picking M sufficiently large. Thus, $A_{1T} \leq c\sup_{x\in[0,1]}\frac{1}{Tb}\sum_{i=1}^{\lfloor Tx\rfloor-\lfloor MTb\rfloor}K(\frac{\lfloor Tx\rfloor-i}{Tb}) \leq \epsilon$ for M sufficiently large. Next,

$$A_{2T} = 2GT^{-1}\sum_{t=1}^{T}\sum_{j=-\lfloor MTb\rfloor+1}^{\lfloor MTb\rfloor}k_{t,t-j}|\sigma_{t-j}^2-\sigma_t^2| \leq c\left(\sup_{-\lfloor MTb\rfloor\leq j\leq\lfloor MTb\rfloor}T^{-1}\sum_{t=1}^{T}|\sigma_{t-j}^2-\sigma_t^2|\right)\to 0$$

for any $M < \infty$ by Lemma A.1 of Cavaliere and Taylor (2009) because $b \to 0$ by Assumption 7, and where the inequality used $\sum_{j=-\lfloor MTb\rfloor+1}^{\lfloor MTb\rfloor}k_{t,t-j} < \infty$ by Assumption 6.

Proof of (b): The left-hand side is a non-negative random variable with expectation

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T}k_{ti}k_{tj}\sigma_i^2\sigma_j^2E(z_i^2-1)(z_j^2-1) \\
&= T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{T}k_{ti}^2\sigma_i^4\kappa_4(i,i,i,i)+T^{-1}\sum_{t=1}^{T}\sum_{i\neq j}k_{ti}k_{tj}\sigma_i^2\sigma_j^2\kappa_4(i,i,j,j) \leq c\frac{1}{Tb}
\end{aligned}$$

because $\sum_{i=1}^{T} k_{ti} = 1$ and using Assumption 1(a)(iii),(b) together with (A.20).

Proof of (c): We let $\sigma_t^2(\theta) := \sum_{i=1}^{T} k_{ti}\varepsilon_i(\theta)^2$ and define $R_T(\theta) := T^{-1}\sum_{t=1}^{T}(\sigma_t^2(\theta) - \tilde{\sigma}_t^2)^2$. We then apply a third-order Taylor expansion of $R_T(\tilde{\theta})$ around $R_T(\theta_0)$,

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 = R_T(\tilde{\theta}) =& R_T(\theta_0) + \frac{\partial R_T(\theta_0)}{\partial \theta}(\tilde{\theta}-\theta_0) + (\tilde{\theta}-\theta_0)'\frac{\partial^2 R_T(\theta_0)}{\partial\theta\partial\theta'}(\tilde{\theta}-\theta_0) \\ &+ \sum_{k,m,n=1}^{p+1}(\tilde{\theta}_k - \theta_{0k})(\tilde{\theta}_m - \theta_{0m})(\tilde{\theta}_n - \theta_{0n})\frac{\partial^3 R_T(\dot{\theta})}{\partial\theta_k\partial\theta_m\partial\theta_n}, \end{aligned} \tag{S.1}$$

for an intermediate value, $\dot{\theta}$. The first term on the right-hand side, $R_T(\theta_0)$, is a non-negative random variable with expectation

$$ER_T(\theta_0) \leq T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}E|(r_i^2 + 2\varepsilon_i r_i)(r_j^2 + 2\varepsilon_j r_j)|,$$

where r_i is defined in Lemma S.2. By repeated application of the Cauchy-Schwarz inequality and using that $E(\varepsilon_i r_i^3) = 0$,

$$\begin{aligned} \left(E|(r_i^2 + 2\varepsilon_i r_i)(r_j^2 + 2\varepsilon_j r_j)|\right)^2 &\leq E(r_i^2 + 2\varepsilon_i r_i)^2 E(r_j^2 + 2\varepsilon_j r_j)^2 \\ &= E(r_i^4 + 4\varepsilon_i^2 r_i^2)E(r_j^4 + 4\varepsilon_j^2 r_j^2) \\ &\leq \left(Er_i^4 + 4(E\varepsilon_i^4)^{1/2}(Er_i^4)^{1/2}\right)(Er_j^4 + 4(E\varepsilon_j^4)^{1/2}(Er_j^4)^{1/2}), \end{aligned}$$

so that, by Lemma S.2 and Assumption 1, $E|(r_i^2 + 2\varepsilon_i r_i)(r_j^2 + 2\varepsilon_j r_j)| \leq ci^{-1-\zeta}j^{-1-\zeta}$, and thus

$$ER_T(\theta_0) \leq cT^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}i^{-1-\zeta}j^{-1-\zeta} \leq cT^{-2}b^{-2}$$

using (A.20).

For the second term on the right-hand side of (S.1), we find in the same way that

$$\begin{aligned} E\left|\frac{\partial R_T(\theta_0)}{\partial\theta_m}\right| &= E\left|4T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}(\varepsilon_i(\theta_0)^2 - \varepsilon_i^2)\varepsilon_j(\theta_0)\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_m}\right| \\ &\leq 4T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}E\left|(r_i^2 + 2\varepsilon_i r_i)\varepsilon_j(\theta_0)\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_m}\right|, \end{aligned}$$

where the square of the expectation is bounded by

$$\left(Er_i^4 + 4(E\varepsilon_i^4)^{1/2}(Er_i^4)^{1/2}\right)E\left(\varepsilon_j(\theta_0)^2\left(\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_m}\right)^2\right) \leq ci^{-2-2\zeta}.$$

It follows that

$$E\left|\frac{\partial R_T(\theta_0)}{\partial\theta_m}\right| \leq cT^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}i^{-1-\zeta} \leq c\frac{1}{Tb}$$

using $\sum_{j=1}^{T} k_{tj} = 1$ and (A.20), so that the contribution to the right-hand side of (S.1) is $O_p(T^{-3/2}b^{-1})$.

To prove the result for the third term on the right-hand side of (S.1) we first find

$$
\begin{aligned}
\frac{\partial^2 R_T(\theta_0)}{\partial\theta_m\partial\theta_n} =& 8T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}(\varepsilon_i + r_i)\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}(\varepsilon_j + r_j)\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_n} \\
&+ 4T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}(r_i^2 + 2\varepsilon_i r_i)\frac{\varepsilon_j(\theta_0)}{\partial\theta_m}\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_n} \\
&+ 4T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}(r_i^2 + 2\varepsilon_i r_i)\varepsilon_j(\theta_0)\frac{\partial^2\varepsilon_j(\theta_0)}{\partial\theta_m\partial\theta_n} \\
=:& 8B_{1T} + 4B_{2T} + 4B_{3T}.
\end{aligned}
$$

The proofs for B_{2T} and B_{3T} are nearly identical, so we give only the latter, for which we find, as above, that

$$
E|B_{3T}| \le T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}E\left|(r_i^2 + 2\varepsilon_i r_i)\varepsilon_j(\theta_0)\frac{\partial^2\varepsilon_j(\theta_0)}{\partial\theta_m\partial\theta_n}\right| \le cT^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}i^{-1-\zeta} \le c\frac{1}{Tb},
$$

so the contribution to (S.1) is $O_p(T^{-2}b^{-1})$. Next, we find that

$$
B_{1T} \le \left(T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{T} k_{ti}(\varepsilon_i + r_i)\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}\right)^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left(\sum_{j=1}^{T} k_{tj}(\varepsilon_j + r_j)\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_n}\right)^2\right)^{1/2},
$$

where the term inside the first large square-root is

$$
\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{T} k_{ti}\varepsilon_i\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}\right)^2 + T^{-1}\sum_{t=1}^{T}\left(\sum_{i=1}^{T} k_{ti}r_i\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}\right)^2 \\
&+ 2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\varepsilon_i r_j\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_m} \\
&=: B_{11T} + B_{12T} + B_{13T},
\end{aligned}
$$

and if the desired result can be shown for B_{11T} and B_{12T} it then follows for B_{13T} by application of the Cauchy-Schwarz inequality. We note that $k_{ti}\varepsilon_i\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}$ is a martingale difference sequence and hence that B_{11T} is a non-negative random variable with

$$
EB_{11T} = T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{T} k_{ti}^2 E\left(\varepsilon_i\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}\right)^2 \le c\frac{1}{Tb}
$$

using $\sum_{i=1}^{T} k_{ti} = 1$, (A.20), and Assumption 1, see also Lemma S.2. Similarly,

$$
EB_{12T} = T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}E\left(r_i r_j\frac{\partial\varepsilon_i(\theta_0)}{\partial\theta_m}\frac{\partial\varepsilon_j(\theta_0)}{\partial\theta_m}\right) \le c\frac{1}{Tb}.
$$

It follows that the contributions of B_{11T} and B_{12T}, and hence B_{13T} and B_{1T}, to (S.1) is $O_p(T^{-2}b^{-1})$.

Finally, we prove the result for the last term on the right-hand side of (S.1), where we find

$$\begin{aligned}\frac{\partial^3 R_T(\theta)}{\partial\theta_l\partial\theta_m\partial\theta_n} =& 2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}(\varepsilon_i(\theta)^2-\varepsilon_i^2)\frac{\partial^3(\varepsilon_j(\theta)^2)}{\partial\theta_l\partial\theta_m\partial\theta_n}\\ &+2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\frac{\partial(\varepsilon_i(\theta)^2)}{\partial\theta_l}\frac{\partial^2(\varepsilon_j(\theta)^2)}{\partial\theta_m\partial\theta_n}\\ &+2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\frac{\partial(\varepsilon_i(\theta)^2)}{\partial\theta_m}\frac{\partial^2(\varepsilon_j(\theta)^2)}{\partial\theta_l\partial\theta_n}\\ &+2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\frac{\partial(\varepsilon_i(\theta)^2)}{\partial\theta_n}\frac{\partial^2(\varepsilon_j(\theta)^2)}{\partial\theta_l\partial\theta_m}\\ =&:2C_{1T}(\theta)+2C_{2T}(\theta)+2C_{3T}(\theta)+2C_{4T}(\theta).\end{aligned}$$

The proofs for $C_{iT}(\theta)$, $i=1,\ldots,4$, are nearly identical, so we give only the proof for $i=1$. With the supremum taken over an arbitrarily small neighborhood of θ_0, we apply the Cauchy-Schwarz inequality such that

$$\sup_\theta\left|\sum_{i=1}^{T} k_{ti}(\varepsilon_i(\theta)^2-\varepsilon_i^2)\right|^2 \le \left(\sum_{i=1}^{T}k_{ti}^2\right)\left(\sup_\theta\sum_{i=1}^{T}(\varepsilon_i(\theta)^2-\varepsilon_i^2)^2\right)=O_p(b^{-1})$$

using (A.20) and Lemma S.3. In the same way,

$$\sup_\theta\left|\sum_{j=1}^{T}k_{tj}\frac{\partial^3(\varepsilon_j(\theta)^2)}{\partial\theta_l\partial\theta_m\partial\theta_n}\right|^2=O_p(b^{-1}),$$

and it follows that $\sup_\theta|C_{1T}(\theta)|=O_p(b^{-1})$ and hence the contribution to (S.1) is $O_p(T^{-3/2}b^{-1})$.

Proof of (d): We find the decomposition

$$T^{-1}\sum_{t=1}^{T}(\hat\sigma_t^2-\sigma_t^2)^2=T^{-1}\sum_{t=1}^{T}(\hat\sigma_t^2-\tilde\sigma_t^2)^2+T^{-1}\sum_{t=1}^{T}(\tilde\sigma_t^2-\bar\sigma_t^2)^2+T^{-1}\sum_{t=1}^{T}(\bar\sigma_t^2-\sigma_t^2)^2+\text{cross-terms},$$

which proves part (d) in light of parts (a)–(c), Assumption 7, and the Cauchy-Schwarz inequality applied to the cross-terms.

Proof of (e): The left-hand side is a non-negative random variable with expectation

$$T^{-1}\sum_{t=1}^{T}\sum_{i,j,m,n=1}^{T}k_{ti}k_{tj}k_{tm}k_{tn}\sigma_i^2\sigma_j^2\sigma_m^2\sigma_n^2E(z_i^2-1)(z_j^2-1)(z_m^2-1)(z_n^2-1),$$

which is a combination of cumulants. When the right-hand side is a $\kappa_8(\cdot)$ cumulant, 3 summations are eliminated by Assumption 1(a)(iii) and the contribution is $O((Tb)^{-3})$ using (A.20) and $\sum_{i=1}^{T}k_{ti}=1$, and when it is a $\kappa_2(\cdot)\kappa_6(\cdot)$ product or a $\kappa_4(\cdot)\kappa_4(\cdot)$ product, 2 summations are eliminated and the contribution is $O((Tb)^{-2})$.

Proof of (f): We apply the inequality $\hat\sigma_t^2\le\sigma_t^2+|\hat\sigma_t^2-\tilde\sigma_t^2|+|\tilde\sigma_t^2-\bar\sigma_t^2|+|\bar\sigma_t^2-\sigma_t^2|$, and note that $\max_{1\le t\le T}\sigma_t^2$ is $O(1)$ by Assumption 1(b), so we show that the max of each of the remaining terms are $O_p(1)$. First, following the proof of part (c) above, define $M_T(\theta):=\max_{1\le t\le T}|\sigma_t^2(\theta)-\tilde\sigma_t^2|$ and apply a mean-value expansion around $M_T(\theta_0)$,

$$\max_{1\le t\le T}|\hat{\sigma}_t^2-\tilde{\sigma}_t^2| = M_T(\tilde{\theta}) = M_T(\theta_0)+2(\tilde{\theta}-\theta_0)\frac{\partial M_T(\dot{\theta})}{\partial\theta}$$
$$\le M_T(\theta_0)+2\sum_{m=1}^{p+1}|\tilde{\theta}_m-\theta_{0,m}|\max_{1\le t\le T}\left|\sum_{i=1}^{T}k_{ti}\varepsilon_i(\dot{\theta})\frac{\partial\varepsilon_i(\dot{\theta})}{\partial\theta_m}\right|, \tag{S.2}$$

for an intermediate value, $\dot{\theta}$. The first term on the right-hand side of (S.2) is a non-negative random variable with expectation

$$EM_T(\theta_0) = E\max_{1\le t\le T}\left|\sum_{i=1}^{T}k_{ti}(\varepsilon_i(\theta_0)^2-\varepsilon_i^2)\right| \le E(\sup_{1\le i\le T,1\le t\le T}|k_{ti}|)\sum_{i=1}^{T}|2\varepsilon_i r_i - r_i^2|$$
$$\le c(\sup_{1\le i\le T,1\le t\le T}|k_{ti}|)\sum_{i=1}^{T}i^{-1-\zeta}\le c\frac{1}{Tb},$$

where the first two inequalities are due to Lemma S.2 and Assumption 1 and the last inequality is due to (A.20). For second term on the right-hand side of (S.2) we apply the Cauchy-Schwarz inequality and find the bound

$$2|\tilde{\theta}_m-\theta_{0,m}|\left(\max_{1\le t\le T}\sum_{i=1}^{T}k_{ti}^2\right)^{1/2}\left(\sum_{i=1}^{T}\varepsilon_i(\dot{\theta})^2\left(\frac{\partial\varepsilon_i(\dot{\theta})}{\partial\theta_m}\right)^2\right)^{1/2},$$

where $|\tilde{\theta}_m-\theta_{0,m}| = O_p(T^{-1/2})$ by (5), $\max_{1\le t\le T}\sum_{i=1}^{T}k_{ti}^2 = O(T^{-1}b^{-1})$ by (A.20) and $\sum_{i=1}^{T}k_{ti}=1$, and the term inside the last parenthesis is $O_p(T)$ by Lemma S.3 with $k_1=0, k_2=1$. Thus, $\max_{1\le t\le T}|\hat{\sigma}_t^2-\tilde{\sigma}_t^2| = O_p(T^{-1/2}b^{-1/2})\xrightarrow{p}0$ by Assumption 7.

Next, for $\tilde{\sigma}_t^2-\bar{\sigma}_t^2=\sum_{i=1}^{T}k_{ti}\sigma_i^2(z_i^2-1)$ we apply Bonferroni's and Markov's inequalities and find

$$P(\max_{1\le t\le T}|\tilde{\sigma}_t^2-\bar{\sigma}_t^2|>\epsilon)\le\sum_{t=1}^{T}P(|\tilde{\sigma}_t^2-\bar{\sigma}_t^2|>\epsilon)\le\frac{1}{\epsilon^4}\sum_{t=1}^{T}E(\tilde{\sigma}_t^2-\bar{\sigma}_t^2)^4,$$

which is $O(T^{-1}b^{-2})\to 0$ by Assumption 7 as in the proof of part (e).

Finally,

$$|\bar{\sigma}_t^2-\sigma_t^2| = \left|\sum_{i=1}^{T}k_{ti}(\sigma_i^2-\sigma_t^2)\right|\le 2G\sum_{i=1}^{T}k_{ti}=2G<\infty,$$

where $G:=\sup_{0\le u\le 1}\sigma(u)^2<\infty$ by Assumption 1(b) and using $\sum_{i=1}^{T}k_{ti}=1$. This implies that $\max_{1\le t\le T}|\bar{\sigma}_t^2-\sigma_t^2|$, and hence $\max_{1\le t\le T}\hat{\sigma}_t^2$, is $O_p(1)$.

Proof of (g): We apply the inequality

$$\min_{1\le t\le T}\hat{\sigma}_t^2\ge\min_{1\le t\le T}\bar{\sigma}_t^2-\max_{1\le t\le T}|\hat{\sigma}_t^2-\tilde{\sigma}_t^2|-\max_{1\le t\le T}|\tilde{\sigma}_t^2-\bar{\sigma}_t^2|,$$

and note that the last two terms are shown to be $o_p(1)$ in the proof of part (f). Thus, proving the result for the first term on the right-hand side is sufficient for proving all the results in part (g). Because $\sum_{i=1}^{T}k_{ti}=1$ for all $t=1,\dots,T$, we find that

$$\min_{1\le t\le T}\bar{\sigma}_t^2=\min_{1\le t\le T}\sum_{i=1}^{T}k_{ti}\sigma_i^2\ge(\min_{1\le i\le T}\sigma_i^2)\min_{1\le t\le T}\sum_{i=1}^{T}k_{ti}\ge\inf_{u\in[0,1]}\sigma^2(u)>0$$

by Assumption 1(b), so that $(\min_{1\le t\le T}\bar{\sigma}_t^2)^{-1}\le(\inf_{u\in[0,1]}\sigma^2(u))^{-1}<\infty$.

S.2.2 Proof of Lemma C.1

Proof of (b)*: The left-hand side is $T^{-1}\sum_{t=1}^{T}(\sum_{i=1}^{T}k_{ti}\hat{\varepsilon}_i^2(w_i^2-1))^2$, which is a non-negative random variable with expectation, conditional on the original sample, given by $T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{T}k_{ti}^2\hat{\varepsilon}_i^4\eta_4$, where $\eta_4 := E(w_i^2-1)^2$. The result now follows as in the proof of Lemma A.1(b) since $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^4 = O_p(1)$.

Proof of (c)*: This follows as in the proof of Lemma A.1(c) by defining $R_T^*(\theta) := T^{-1}\sum_{t=1}^{T}(\sigma_t^{*2}(\theta) - \sigma_t^{*2}(\hat{\theta}))^2$ and $\sigma_t^{*2}(\theta) = \sum_{i=1}^{T}k_{ti}\varepsilon_i^*(\theta)^2$, noting that $(\tilde{\theta}^* - \hat{\theta}) = O_p^*(T^{-1/2})$, in probability, by Theorem 6 of CNT.

Proof of (e)*: The left-hand side is a non-negative random variable with expectation, conditional on the original data, given by

$$T^{-1}\sum_{t=1}^{T}\sum_{i,j,m,n=1}^{T}k_{ti}k_{tj}k_{tm}k_{tn}\hat{\varepsilon}_i^2\hat{\varepsilon}_j^2\hat{\varepsilon}_m^2\hat{\varepsilon}_n^2E^*(w_i^2-1)(w_j^2-1)(w_m^2-1)(w_n^2-1)$$

$$\leq cT^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T}k_{ti}^2k_{tj}^2\hat{\varepsilon}_i^4\hat{\varepsilon}_j^4 \leq c\frac{1}{T^3b^2}\sum_{t=1}^{T}\left(\sum_{i=1}^{T}k_{ti}\hat{\varepsilon}_i^4\right)^2 \leq c\frac{1}{T^3b^2}\sum_{t=1}^{T}\left(\sum_{i=1}^{T}k_{ti}^2\right)\left(\sum_{i=1}^{T}\hat{\varepsilon}_i^8\right),$$

which is $O_p(T^{-2}b^{-3})$ by (A.20) since $T^{-1}\sum_{i=1}^{T}\hat{\varepsilon}_i^8 = O_p(1)$.

Proof of (f)*: We apply the inequality $\hat{\sigma}_t^{*2} \leq \hat{\sigma}_t^2 + |\hat{\sigma}_t^{*2} - \tilde{\sigma}_t^{*2}| + |\tilde{\sigma}_t^{*2} - \hat{\sigma}_t^2|$, and note that $\max_{1\leq t\leq T}\hat{\sigma}_t^2 = O_p(1)$ by Lemma A.1(f). The proof for the second term is exactly the same as for the corresponding term in the proof of Lemma A.1(f), but using that $(\tilde{\theta}^* - \hat{\theta}) = O_p^*(T^{-1/2})$, in probability (by Theorem 6 of CNT) and Lemma S.8.

Next, for $\tilde{\sigma}_t^{*2} - \hat{\sigma}_t^2 = \sum_{i=1}^{T}k_{ti}\hat{\varepsilon}_i^2(w_i^2-1)$ we apply Bonferroni's and Markov's inequalities and find

$$P^*(\max_{1\leq t\leq T}|\tilde{\sigma}_t^{*2} - \hat{\sigma}_t^2| > \epsilon) \leq \sum_{t=1}^{T}P^*(|\tilde{\sigma}_t^{*2} - \hat{\sigma}_t^2| > \epsilon) \leq \frac{1}{\epsilon^8}\sum_{t=1}^{T}E^*(\tilde{\sigma}_t^{*2} - \hat{\sigma}_t^2)^8$$

$$\leq c\sum_{t=1}^{T}E^*\prod_{i=1}^{8}\sum_{j_i=1}^{T}k_{tj_i}\hat{\varepsilon}_{j_i}^2(w_{j_i}^2-1) = c\eta_4^4\sum_{t=1}^{T}(\sum_{i=1}^{T}k_{ti}^2\hat{\varepsilon}_i^4)^4,$$

which by the Cauchy-Schwarz inequality is bounded by

$$c\eta_4^4\sum_{t=1}^{T}(\sum_{i=1}^{T}k_{ti}^4)^2(\sum_{i=1}^{T}\hat{\varepsilon}_i^8)^2 \leq cT(T^{-3}b^{-3})^2O_p(T^2) = O_p(T^{-3}b^{-6})$$

using (A.20), integrability of the kernel, and $T^{-1}\sum_{i=1}^{T}\hat{\varepsilon}_i^8 = O_p(1)$. The required result then follows by Assumption 7. This implies that $\max_{1\leq t\leq T}|\tilde{\sigma}_t^{*2} - \hat{\sigma}_t^2|$, and hence $\max_{1\leq t\leq T}\hat{\sigma}_t^{*2}$, is $O_p^*(1)$, in probability.

Proof of (g):* We apply the inequality

$$\min_{1\leq t\leq T}\hat{\sigma}_t^{*2} \geq \min_{1\leq t\leq T}\hat{\sigma}_t^2 - \max_{1\leq t\leq T}|\hat{\sigma}_t^{*2} - \tilde{\sigma}_t^{*2}| - \max_{1\leq t\leq T}|\tilde{\sigma}_t^{*2} - \hat{\sigma}_t^2|,$$

and note that the last two terms are shown to be $o_p^*(1)$, in probability, in the proof of part (f*) and the first term is $O_p(1)$ by Lemma A.1(g).

S.2.3 Technical Lemmas

The proofs of these lemmas are in Section S.2.4.

In Lemma S.2 we prove bounds for certain coefficients and remainder terms, which need to be uniform in the parameters, although for the second and third derivatives uniformity is only needed in a neighborhood of the true value. For any function $f(\theta) : \mathbb{R}^n \to \mathbb{R}$, we define $\partial^k f(\theta)/\partial\theta^{(k)}$ as a short-hand notation for a generic element of the k'th derivative with respect to the vector θ.

Lemma S.1. *Let the sequences a_j and $b_{j,T}, j = 1, \dots, T$, be such that $\sum_{j=1}^T |a_j| < \infty$, $\sup_{j,T} |b_{j,T}| < \infty$, and, for some $q_T \to \infty$ as $T \to \infty$, $\sup_{j \le q_T, T} |b_{j,T}| \to 0$. Then $\sum_{j=1}^T a_j b_{j,T} \to 0$.*

Lemma S.2. *Let Assumptions 1–3 be satisfied. For $k = 0, 1$ define $\Psi_k := \Psi$ and for $k = 2, 3$ define $\Psi_k := \mathcal{N}_\delta(\psi_0) := \{\psi \in \Psi : ||\psi - \psi_0|| \le \delta\}$ for some $\delta > 0$. Then it holds that*

$$\varepsilon_t(\theta) = \sum_{j=0}^{t-1} \phi_j(\theta)\varepsilon_{t-j} + r_t(\theta) \text{ and } \varepsilon_t(\theta_0) = \varepsilon_t + r_t,$$

where, for any integer h such that $1 \le h \le 8$, for any finite constant g, and for $k = 0, 1, 2, 3$,

$$\sup_{d_0 - d \le g, \psi \in \Psi_k} |\partial^k \phi_j(\theta)/\partial\theta^{(k)}| \le c(\log j)^k j^{\max(g-1, -2-\zeta)},$$
$$E \sup_{d_0 - d \le g, \psi \in \Psi_k} |\partial^k r_t(\theta)/\partial\theta^{(k)}|^h \le c(\log t)^{hk} t^{h\max\{g-1, -1-\zeta\}},$$
$$E|r_t|^h \le ct^{-h(1+\zeta)}.$$

Lemma S.3. *Let Assumptions 1–3 be satisfied. For $k = 0, 1$ define $\Psi_k := \Psi$ and for $k = 2, 3$ define $\Psi_k := \mathcal{N}_\delta(\psi_0) := \{\psi \in \Psi : ||\psi - \psi_0|| \le \delta\}$ for some $\delta > 0$. Also, for all integers q such that $2 \le q \le 8$, let $k(i) = 0, 1, 2, 3$ for $i = 1, \dots, q$, and define integers $r_1, \dots, r_{k(i)}$ such that $1 \le r_m \le p + 1$ for $m = 1, \dots, k(i)$. Then, for any $\kappa > 0$,*

$$\sup_{d_0 - d \le 1/2 - \kappa, \psi \in \Psi_{\max k(i)}} T^{-1} \sum_{t=1}^T \left(\prod_{i=1}^q \frac{\partial^{k(i)} \varepsilon_t(\theta)}{\partial\theta_{r_1} \dots \partial\theta_{r_{k(i)}}} \right) = O_p(1).$$

S.2.4 Proofs of Technical Lemmas

S.2.4.1 Proof of Lemma S.1 By the triangle inequality,

$$\left|\sum_{j=1}^T a_j b_{j,T}\right| \le \left|\sum_{j=1}^{q_T} a_j b_{j,T}\right| + \left|\sum_{j=q_T+1}^T a_j b_{j,T}\right| \le (\sup_{j \le q_T, T} |b_{j,T}|) \sum_{j=1}^{q_T} |a_j| + (\sup_{j,T} |b_{j,T}|) \sum_{j=q_T+1}^T |a_j|,$$

where the first term converges to zero by assumption and the last converges to zero because it is the tail of a convergent sum.

S.2.4.2 Proof of Lemma S.2 The residual is given in (3) as

$$\begin{aligned}
\varepsilon_t(\theta) =& \sum_{n=0}^{t-1} \sum_{j=0}^{t-1-n} \sum_{m=0}^{\infty} b_n(\psi)\pi_j(d_0 - d)a_m(\psi_0)\varepsilon_{t-n-j-m} \\
=& \sum_{j=0}^{t-1} \sum_{n=0}^{t-1-j} \sum_{m=0}^{t-n-j-1} b_n(\psi)\pi_j(d_0 - d)a_m(\psi_0)\varepsilon_{t-n-j-m} \\
&+ \sum_{n=0}^{t-1} \sum_{j=0}^{t-1-n} \sum_{m=t-n-j}^{\infty} b_n(\psi)\pi_j(d_0 - d)a_m(\psi_0)\varepsilon_{t-n-j-m} \\
=& \sum_{j=0}^{t-1} \phi_j(\theta)\varepsilon_{t-j} + r_t(\theta),
\end{aligned}$$

where $\phi_j(\theta) := \sum_{n=0}^{j}\sum_{m=0}^{j-n} b_m(\psi)\pi_n(d_0-d)a_{j-n-m}(\psi_0)$ satisfies

$$\sup_{d_0-d\leq g,\psi\in\Psi}\left|\frac{\partial^k\phi_j(\theta)}{\partial\theta^{(k)}}\right| \leq c\sum_{n=0}^{j}(\log n)^k n^{g-1}\sum_{m=0}^{j-n} m^{-2-\zeta}(j-n-m)^{-2-\zeta}$$
$$\leq c(\log j)^k\sum_{n=0}^{j} n^{g-1}(j-n)^{-2-\zeta} \leq c(\log j)^k j^{\max(g-1,-2-\zeta)}$$

by (2) and Lemmas S.5 and S.6. The remainder term, $r_t(\theta) := \sum_{n=0}^{t-1}\sum_{j=0}^{t-1-n}\sum_{m=t-n-j}^{\infty} b_n(\psi)\pi_j(d_0-d)a_m(\psi_0)\varepsilon_{t-n-j-m}$, satisfies, by the same arguments and using also Assumption 1(a)(iii),(b),

$$E\sup_{d_0-d\leq g,\psi\in\Psi}\left|\frac{\partial^k r_t(\theta)}{\partial\theta^{(k)}}\right|^h \leq cE\prod_{i=1}^{h}\sum_{n_i=0}^{t-1} n_i^{-2-\zeta}\sum_{j_i=0}^{t-1-n_i}(\log j_i)^k j_i^{g-1}\sum_{m_i=t-n_i-j_i}^{\infty} m_i^{-2-\zeta}|\varepsilon_{t-n_i-j_i-m_i}|$$
$$\leq c\left((\log t)^k\sum_{n=0}^{t-1} n^{-2-\zeta}\sum_{j=0}^{t-1-n} j^{g-1}(t-n-j)^{-1-\zeta}\right)^h$$
$$\leq c\left((\log t)^k\sum_{n=0}^{t-1} n^{-2-\zeta}(t-n)^{\max\{g-1,-1-\zeta\}}\right)^h \leq c(\log t)^{hk}t^{h\max\{g-1,-1-\zeta\}}.$$

At $\theta=\theta_0$ we find, using $\pi_j(0)=\mathbb{I}(j=0)$,

$$\varepsilon_t(\theta_0) = \sum_{n=0}^{t-1}\sum_{m=0}^{\infty} b_n(\psi_0)a_m(\psi_0)\varepsilon_{t-n-m} = \varepsilon_t + r_t,$$

where $r_t := -\sum_{n=t}^{\infty}\sum_{m=0}^{\infty} b_n(\psi_0)a_m(\psi_0)\varepsilon_{t-n-m}$ satisfies, by the same arguments as above,

$$E|r_t|^h \leq c\left(\sum_{n=t}^{\infty}\sum_{m=0}^{\infty}|b_n(\psi_0)||a_m(\psi_0)|\right)^h \leq c\left(\sum_{n=t}^{\infty} n^{-2-\zeta}\right)^h \leq c\left(t^{-1-\zeta}\right)^h \leq ct^{-h(1+\zeta)}.$$

S.2.4.3 Proof of Lemma S.3 First apply Hölder's inequality,

$$T^{-1}\sum_{t=1}^{T}\left(\prod_{i=1}^{q}\frac{\partial^{k(i)}\varepsilon_t(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}}\right) \leq \prod_{i=1}^{q}\left(T^{-1}\sum_{t=1}^{T}\left|\frac{\partial^{k(i)}\varepsilon_t(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}}\right|^q\right)^{1/q}.$$

Next, by Lemma S.2 and Minkowski's inequality we find

$$\left(T^{-1}\sum_{t=1}^{T}\left|\frac{\partial^{k(i)}\varepsilon_t(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}}\right|^q\right)^{1/q} \leq \left(T^{-1}\sum_{t=1}^{T}\left|\sum_{j=0}^{t-1}\frac{\partial^{k(i)}\phi_j(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}}\varepsilon_{t-j}\right|^q\right)^{1/q} + \left(T^{-1}\sum_{t=1}^{T}\left|\frac{\partial^{k(i)}r_t(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}}\right|^q\right)^{1/q}. \quad \text{(S.3)}$$

We note from Lemma S.2 that the derivatives add at most a logarithmic factor, which is inconsequential to the proof, so we give the proof only for $k(i)=0$ to lighten the notation. We first find from Lemma S.2 that the second term on the right-hand side of (S.3) satisfies

$$E\sup_{d_0-d\leq 1/2-\kappa,\psi\in\Psi} T^{-1}\sum_{t=1}^{T}|r_t(\theta)|^q \leq cT^{-1}\sum_{t=1}^{T} t^{q(-1/2-\kappa)} \leq cT^{-1}$$

for any $\kappa>0$ because $q\geq 2$.

Next, we give the proof for the first term on the right-hand side of (S.3). By summation by parts,

$$\sum_{j=0}^{t-1}\phi_j(\theta)\varepsilon_{t-j} = \phi_{t-1}(\theta)\sum_{j=0}^{t-1}\varepsilon_{t-j} + \sum_{j=0}^{t-2}(\phi_j(\theta)-\phi_{j+1}(\theta))\sum_{l=0}^{j}\varepsilon_{t-l},$$

so that

$$T^{-1}\sum_{t=1}^{T}\left|\sum_{j=0}^{t-1}\phi_j(\theta)\varepsilon_{t-j}\right|^q = T^{-1}\sum_{t=1}^{T}\left|\phi_{t-1}(\theta)\sum_{j=0}^{t-1}\varepsilon_{t-j}\right|^q \tag{S.4}$$

$$+ T^{-1}\sum_{t=1}^{T}\left|\sum_{j=0}^{t-2}(\phi_j(\theta)-\phi_{j+1}(\theta))\sum_{l=0}^{j}\varepsilon_{t-l}\right|^q \tag{S.5}$$

$$+ \text{cross-terms},$$

where the cross-terms will be $O_p(1)$, uniformly in $\theta \in \Theta_2$, by the Cauchy-Schwarz inequality after showing that the same is true for the two main terms. Because $\sum_{j=0}^{t-1}\varepsilon_{t-j} = O_p(t^{1/2})$, uniformly in t, and $\sup_{d_0-d\leq 1/2-\kappa,\psi\in\Psi}|\phi_{t-1}(\theta)| \leq ct^{-1/2-\kappa}$ by Lemma S.2, we first find that $\sup_{d_0-d\leq 1/2-\kappa,\psi\in\Psi}|(S.4)| = O_p(T^{\max\{-\kappa q,-1\}}) = o_p(1)$ because $\kappa > 0, q \geq 2$. To prove the result for (S.5), first note that, with obvious notation, $\phi_{j+1}(d,\psi) - \phi_j(d,\psi) = \phi_{j+1}(d-1,\psi)$, so that $\sup_{d_0-d\leq 1/2-\kappa,\psi\in\Psi}|\phi_j(\theta)-\phi_{j+1}(\theta)| \leq cj^{-3/2-\kappa}$ by Lemma S.2. It then follows that

$$\sup_{d_0-d\leq 1/2-\kappa,\psi\in\Psi}|\sum_{j=0}^{t-2}(\phi_j(\theta)-\phi_{j+1}(\theta))\sum_{l=0}^{j}\varepsilon_{t-l}| = O_p(1),$$

uniformly in t, and therefore

$$\sup_{d_0-d\leq 1/2-\kappa,\psi\in\Psi}|(S.5)| = O_p(1)$$

which proves the result.

S.2.5 Additional Lemmas Without Proofs

Lemma S.4 (CNT, Lemma A.2). *Let z_t be a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t$, the sigma-field generated by $\{z_s\}_{s\leq t}$, and suppose $E|z_t|^q < \infty$ for some integer $q \geq 2$. Then the q'th order moments and cumulants satisfy*

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

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

Lemma S.5 (Johansen and Nielsen, 2010, Lemma B.4). *Uniformly for* $\max\{|\alpha|,|\beta|\} \leq a_0$ *it holds that*

$$\sum_{j=1}^{t-1} j^{\alpha-1}(t-j)^{\beta-1} \leq c(a_0)(1+\log t)t^{\max\{\alpha+\beta-1,\alpha-1,\beta-1\}}.$$

Lemma S.6 (Johansen and Nielsen, 2010, Lemma B.3). *For $|u| \leq u_0$, $m \geq 0$, and all $j \geq 1$ it holds uniformly in u that*

$$|\frac{\partial^m}{\partial u^m}\pi_j(u)| \leq c(u_0)(1+\log j)^m j^{u-1}, \tag{S.6}$$

$$|\frac{\partial^m}{\partial u^m}T^{-u}\pi_j(u)| \leq c(u_0)T^{-u}(1+|\log(j/T)|)^m j^{u-1}. \tag{S.7}$$

Lemma S.7. *Let the assumptions of Theorem 3 be satisfied. For $k = 0, 1$ define $\Psi_k := \Psi$ and for*

$k = 2, 3$ *define* $\Psi_k := \mathcal{N}_\delta(\psi_0) := \{\psi \in \Psi : ||\psi - \psi_0|| \leq \delta\}$ *for some* $\delta > 0$. *Then it holds that*

$$\varepsilon_t^*(\theta) = \sum_{j=0}^{t-1} \hat{\phi}_j(\theta)\varepsilon_{t-j}^* \text{ and } \varepsilon_t^*(\hat{\theta}) = \varepsilon_t^*,$$

where, for any finite constant g *and for* $k = 0, 1, 2, 3$,

$$\sup_{\hat{d}-d \leq g, \psi \in \Psi_k} |\partial^k \hat{\phi}_j(\theta)/\partial\theta^{(k)}| = O_p((\log j)^k j^{\max(g-1,-2-\zeta)}).$$

Proof. The proof is almost identical to that of Lemma S.2, with the main difference being that $\varepsilon_t^* = 0$ for $t \leq 0$, and is omitted for brevity. □

Lemma S.8. *Let the assumptions of Theorem 3 be satisfied. For* $k = 0, 1$ *define* $\Psi_k := \Psi$ *and for* $k = 2, 3$ *define* $\Psi_k := \mathcal{N}_\delta(\psi_0) := \{\psi \in \Psi : ||\psi - \psi_0|| \leq \delta\}$ *for some* $\delta > 0$. *Also, for all integers* q *such that* $2 \leq q \leq 8$, *let* $k(i) = 0, 1, 2, 3$ *for* $i = 1, \ldots, q$, *and define integers* $r_1, \ldots, r_{k(i)}$ *such that* $1 \leq r_m \leq p + 1$ *for* $m = 1, \ldots, k(i)$. *Then, for any* $\kappa > 0$, *in probability,*

$$\sup_{\hat{d}-d \leq 1/2-\kappa, \psi \in \Psi_{\max k(i)}} T^{-1} \sum_{t=1}^{T} \left(\prod_{i=1}^{q} \frac{\partial^{k(i)} \varepsilon_t^*(\theta)}{\partial\theta_{r_1} \ldots \partial\theta_{r_{k(i)}}} \right) = O_p^*(1).$$

Proof. The proof is almost identical to that of Lemma S.3 and is omitted for brevity. □

S.3 Heteroskedasticity Diagnostics

To investigate the possible presence of heteroskedasticity in the residuals, we report in Table S.1 several tests for conditional and unconditional heteroskedasticity for both the physical and the sovereign debt data series. The superscripts a, b, and c denote significance at the 1%, 5%, and 10% nominal (asymptotic) levels, respectively.

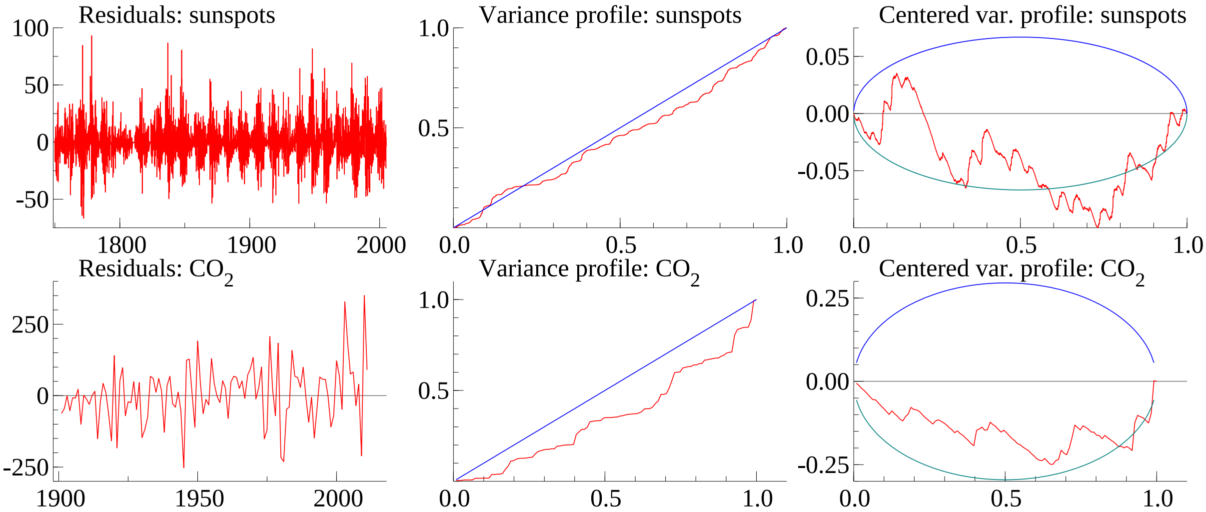
In the first two columns of results in Table S.1 we report LM tests of the null hypothesis of conditional homoskedasticity against the alternative of ARCH(k) dynamics. These tests are based on an AR(k) regression fitted to the squared residuals. For all series the null hypothesis is easily rejected at any conventional significance level.

In the last four columns of Table S.1 we report the $\mathcal{H}_{\mathrm{R}}$, $\mathcal{H}_{\mathrm{KS}}$, $\mathcal{H}_{\mathrm{CvM}}$, and $\mathcal{H}_{\mathrm{AD}}$ stationary volatility tests of Cavaliere and Taylor (2008, p. 312). These are tests of the null of stationary volatility, i.e. allowing in particular for conditional heteroskedasticity under the null, against the alternative of non-stationary volatility (unconditional heteroskedasticity). All series except CO_2 show strong evidence of unconditional heteroskedasticity.

To visualize the possible presence of unconditional heteroskedasticity in the residuals, we first plot the residual series in the left-hand panels in Figures S.1–S.3. In the middle panels of Figures S.1–S.3 we plot the sample variance profiles of the residuals, say $\tilde{\varepsilon}_t$, of the fitted ARFIMA models. The sample variance profiles, see Cavaliere and Taylor (2008), are plots of $\hat{\eta}(u) := (\sum_{t=1}^{T} \tilde{\varepsilon}_t^2)^{-1} \sum_{t=1}^{\lfloor Tu \rfloor} \tilde{\varepsilon}_t^2$ against $u \in [0, 1]$. In large samples, $\hat{\eta}(u) \approx (\int_0^1 \sigma^2(s)\, ds)^{-1} \int_0^u \sigma^2(s)\, ds$, which equals u when the unconditional volatility is constant; that is, when there is no unconditional heteroskedasticity. Consequently, under conditional homoskedasticity or, more generally, under stationary conditional heteroskedasticity, $\hat{\eta}(u)$ should be close to the 45 degree line, and significant deviations of this function from the 45 degree line point to the presence of persistent changes in volatility. These deviations, along with the corresponding 95% confidence bands[5], are reported in the right-hand panels of Figures S.1–S.3.

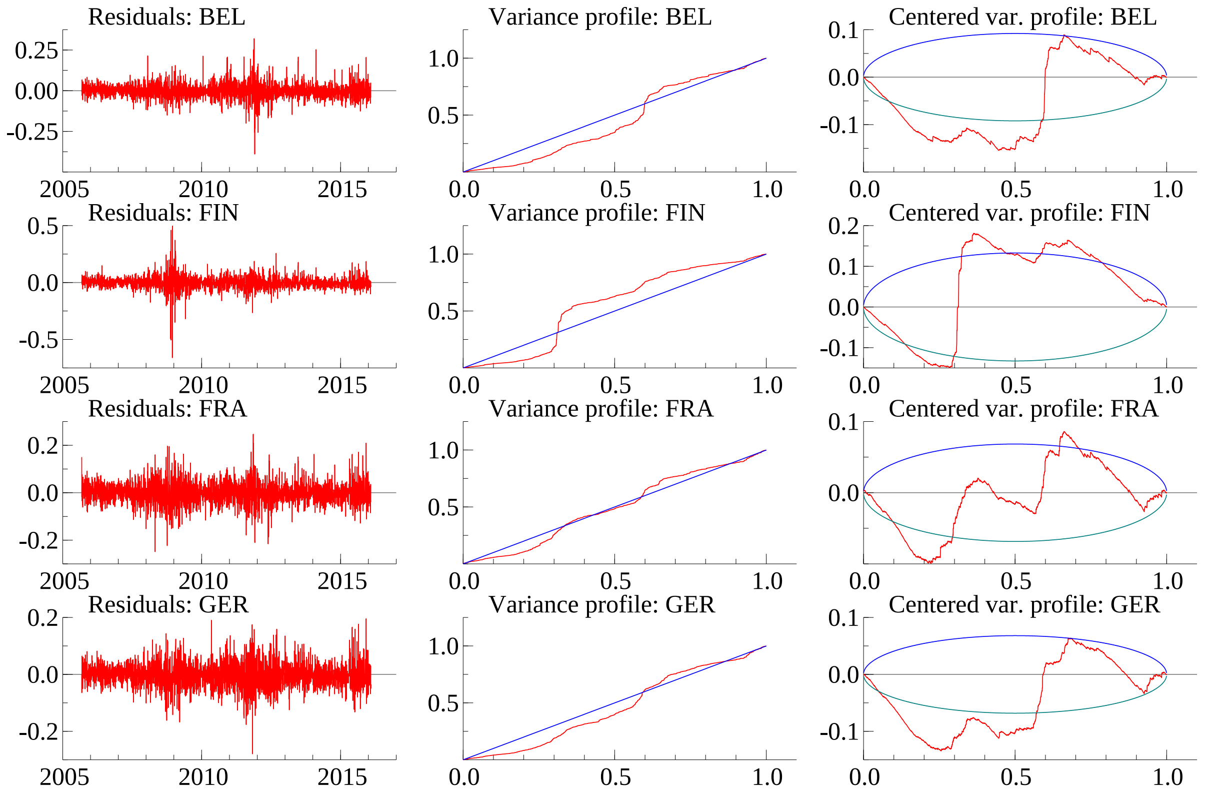
[5] The confidence bands are obtained as suggested by Cavaliere and Taylor (2008). This requires estimation of the long-run variance of $\tilde{\varepsilon}_t^2$ under the null hypothesis, which is done here using a sums-of-covariances estimator with the Bartlett kernel and a lag truncation of five.

Figure S.1: Residual graphics for physical data examples



Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Figure S.2: Residual graphics for sovereign debt data examples, part 1



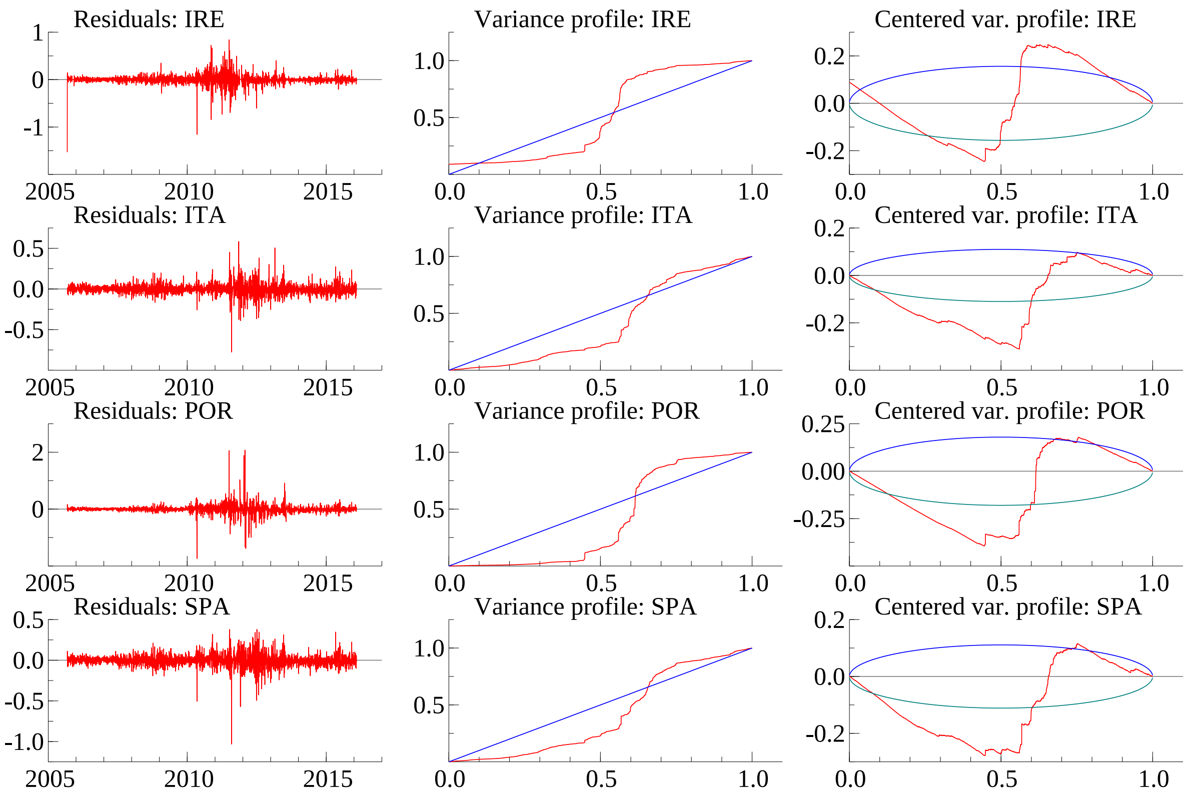
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Table S.1: Conditional and Unconditional Heteroskedasticity Tests for Data Examples

Series	ARCH(5)	ARCH(20)	$\mathcal{H}_{\mathrm{R}}$	$\mathcal{H}_{\mathrm{KS}}$	$\mathcal{H}_{\mathrm{CvM}}$	$\mathcal{H}_{\mathrm{AD}}$
Panel A: Physical Data Examples						
Sunspots	686.129^a	794.302^a	1.980^b	1.464^b	0.540^b	2.942^b
CO_2	32.616^a	48.050^a	0.833	0.826	0.260	1.549
Panel B: Sovereign Debt Data Examples						
Belgium	704.842^a	773.152^a	2.592^a	1.642^a	0.987^a	4.962^a
Finland	618.360^a	873.567^a	2.437^a	1.335^c	0.740^b	3.733^b
France	585.122^a	650.438^a	2.654^a	1.423^b	0.448^c	2.719^b
Germany	688.557^a	773.014^a	2.855^a	1.935^a	1.155^a	6.272^a
Ireland	146.756^a	196.892^a	3.092^a	1.554^b	0.924^a	5.901^a
Italy	319.061^a	482.451^a	3.636^a	2.769^a	1.920^a	8.943^a
Portugal	134.993^a	340.272^a	3.124^a	2.147^a	1.374^a	6.529^a
Spain	167.000^a	245.054^a	3.470^a	2.453^a	1.891^a	8.980^a

Notes: ARCH(k) denotes the LM test for ARCH(k) based on a AR(k) regression fitted to the squared residuals, and $\mathcal{H}_{\mathrm{R}}$, $\mathcal{H}_{\mathrm{KS}}$, $\mathcal{H}_{\mathrm{CvM}}$, and $\mathcal{H}_{\mathrm{AD}}$ denote the stationary volatility tests proposed in Cavaliere and Taylor (2008, p. 312). The superscripts a, b, and c denote significance at the 1%, 5%, and 10% nominal (asymptotic) levels, respectively.

Figure S.3: Residual graphics for sovereign debt data examples, part 2



Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Additional References

These are the references cited in this supplementary appendix but not listed in the main paper.

Cavaliere, G. and A.M.R. Taylor (2008), Time-transformed unit root tests for models with non-stationary volatility, *Journal of Time Series Analysis* 29, 300–330.