Adaptive Inference in Heteroskedastic Fractional Time Series Models

Giuseppe Cavaliere, Morten Ørregaard Nielsen and Robert Taylor

Adaptive Inference in Heteroskedastic Fractional Time Series Models

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

[a] Department of Economics, University of Bologna and Department of Economics, Exeter Business School

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

[c] Essex Business School, University of Essex

August 23, 2019

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 volatility process. We show that adaptive estimation 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. Importantly, the asymptotic analysis is based on a novel proof strategy, which does not require consistent estimation (in the sup norm) of the volatility process. Consequently, we are able to work under a weaker set of assumptions than those employed in the extant literature. The asymptotic variance matrices of both the standard and adaptive CSS estimators depend on any weak parametric autocorrelation present in the fractional model and any conditional heteroskedasticity 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 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.

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

1 Introduction

Fractionally integrated time series models have proved highly effective in a wide range of fields of application including economics, finance, internet modeling, hydrology, climate studies, linguistics, opinion polling and DNA sequencing to name but a few; see, for example, the survey by Samorodnitsky (2007), the books by Giraitis *et al.* (2012) and Beran *et al.* (2013), and references therein.

As a result, a great many papers have appeared in the literature addressing estimation of and inference on the parameters of fractional time series models. The earliest work developed semi-parametric estimation methods; e.g., Geweke and Porter-Hudak (1983) and Robinson (1995). More relevant to this paper, parametric maximum likelihood and quasi-maximum likelihood (QML) based estimation and inference methods have been developed in, among others, Fox and Taqqu (1986), Robinson (1994), Tanaka (1999), Johansen and Nielsen (2010), Hualde and Robinson (2011), and Nielsen (2015). Of these estimation methods, the (conditional) QML estimator - equivalently the conditional sum-of-squares (CSS) estimator - was shown by Hualde and Robinson (2011) and Nielsen (2015) to be consistent over an arbitrarily large set of admissible values of the long memory parameter, thereby solving 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.

All of the papers cited above assume that the shocks driving the model are unconditionally homoskedastic (and in many cases the even stronger assumption that the shocks are conditionally homoskedastic or even independent and identically distributed, IID). That is, they do not allow for shocks which display permanent changes in their unconditional variance over time, what is often referred to as non-stationary volatility. 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 observed after 2007; see, *inter alia*, McConnell and Perez-Quiros (2000), Stock and Watson (2012), and the references therein. Fractional integration models have been very successfully applied in financial volatility modeling, particularly for the realised variance [RV]; see, e.g., Andersen *et al.* (2007) and references therein.

Moreover, the variance of RV is known to be time-varying; see Barndorff-Nielsen and Shephard (2002) for theoretical arguments and Corsi *et al.* (2008) for empirical evidence on this point.

In spite of this there are still relatively few papers which consider estimation and inference methods for fractionally integrated models driven by unconditionally heteroskedastic shocks. Harris and Kew (2017) extend the score-based tests of Robinson (1994) to develop LM-type tests on the value of the long memory parameter in the context of an unconditionally heteroskedastic, but conditionally homoskedastic, fractionally integrated model. Cavaliere, Nielsen and Taylor (2017, henceforth CNT) demonstrate that the QML estimator of Hualde and Robinson (2011), originally developed under the assumption of conditional homoskedasticity, retains its global consistency and asymptotic normality properties in cases where the innovations display non-stationary volatility and/or conditional heteroskedasticity. However, CNT also demonstrate that 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. 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.

The contribution of this paper to the literature is to develop asymptotically efficient QML-based methods of estimation and inference for the long and/or short memory parameters of univariate fractionally integrated time series models driven by shocks which display non-stationary volatility. To the best of our knowledge the methods developed in this paper are the first to provide a full theory of adaptive estimation and inference, including bootstrap-based inferential theory, in a heteroskedastic fractional time series model. This also allows us to develop methods of model selection and model-based inference. Our focus is on the QML estimation principle of Hualde and Robinson (2011). We will develop two-step QML-based estimation and inference methods for fractional time series models which non-parametrically adapt to unconditional heteroskedasticity of unknown form and we demonstrate that these recover the (asymptotic) efficiency losses experienced by the standard QML-based methods estimator under non-stationary volatility identified by CNT. 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 the adaptive CSS [ACSS] estimator.

2

A very large and well-established literature exists relating to the use of kernel-based methods for estimating local variance functions; see, among many others, Fan and Yao (1998), Hall and Tajividi (2000), Yu and Jones (2004) and Xu and Phillips (2011). Adaptive inference based on using such 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). More generally, adaptive estimators and tests designed to account for non-parametric heteroskedasticity have been widely used in the time series literature; see among others, Carroll (1982), Robinson (1987), Harvey and Robinson (1988), Hansen (1995), Xu and Phillips (2011) and Xu and Yang (2015). Adaptive methods have also been used in the context of testing for autoregressive conditional heteroskedasticity [ARCH] effects in unconditionally heteroskedastic autoregressive models by Patilea and Raïssi (2014), and for the adaptive estimation of VAR models in Patilea and Raïssi (2012, 2013). Gou and Koul (2008) consider adaptive estimation of the unconditional variance in the context of heteroskedastic regression models which display long memory in the regressors and regression disturbances. Of most relevance to the present paper is Harris and Kew (2017) whose LM-type tests on the value of the long memory parameter are based around an adaptive estimate of the unconditional volatility process. Importantly, however, Harris and Kew (2017) develop only LM tests and do not consider estimation. Moreover, they do not allow for the presence of conditional heteroskedasticity and impose considerably stronger conditions on the unconditional volatility process than we can allow for here; see Remark 3.4 for further details.

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 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. 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 present in the shocks and on

3

any weak parametric autocorrelation present in the model. 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 generalised ARCH [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 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. Data examples relating to realised variance and government bond yields are reported in Section 6. Section 7 concludes. All mathematical proofs and additional data analysis are reported in the Supplementary Appendix; Cavaliere, Nielsen and Taylor (2019).

NOTATION. We use c to denote a generic, finite constant, and $||\cdot||, ||\cdot||_r$ to denote the Euclidean and L_r-norms, respectively. 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 $c > 0$ such that $|f(x_1) - f(x_2)| \le c||x_1 - x_2||^\alpha$ 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, as the sample size $T \to \infty$. The probability and expectation conditional on the realisation of the original sample are denoted P^* and E^*, respectively. For a 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 $c > 0$ such that $P^*(|X_T^*| > c) \to 0$ 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 (type II) fractional time series model

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

where L is the usual lag operator, and the operator Δ_+^{-d} is given, for a generic variable x_t, by $\Delta_+^{-d} x_t := \Delta^{-d} x_t \mathbb{I}(t \ge 1) = \sum_{n=0}^{t-1} \pi_n(d) x_{t-n}$, 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 $a(z, \psi)$ can be found in Hualde and Robinson (2011). Thus, our focus is model-based inference (on either 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 $\{z_s\}_{s\le t}$, such that $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ for $t =$

..., $-1, 0, 1, 2, ...$, and satisfies

(i) $E(z_t^2) = 1$,

(ii) $\tau_{r,s} := E(z_t^2 z_{t-r} z_{t-s})$ is uniformly bounded for all $r \geq 0, s \geq 0$,

(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 $\sup_t \sum_{r_1, \ldots, r_{q-2}=1}^{\infty} |\kappa_q(t, t, t - r_1, \ldots, t - r_{q-2})| < \infty$.

(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 $\text{Lip}(\zeta)$ 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 $\text{Lip}(\zeta)$ 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 twice differentiable in ω with derivative in $\text{Lip}(\zeta)$ 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. Assumption 1(a), in particular part (iii), rules out long memory in z_t^2 because, for $q = 4$, it implies that $\sum_{r=1}^{\infty} |Cov(z_t^2, z_{t-r}^2)| < \infty$. However, volatility clustering, such as 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 *et al.* (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. $\diamond$

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. $\diamond$

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 (formally, $\varepsilon_t = \varepsilon_{T,t} = \sigma_{T,t} z_t$ is a triangular array, but for simplicity the subscript T is suppressed). 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. $\diamond$

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 IID 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. $\diamond$

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. $\diamond$

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 $\mathrm{Lip}(\zeta)$ for $\zeta > 1/2$, and Assumption 3(iii) strengthens their Assumption A1(iii) from 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, if 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 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 none is present $A_0 = B_0$ because here $\tau_{n,m} = \mathbb{I}(n = m)$. Neither A_0 nor B_0 are affected by any unconditional heteroskedasticity arising from Assumption 1(b). As in Hualde and Robinson (2011), to state the limiting distribution of the QML and adaptive estimators we will require B_0 to be invertible:

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. To this end, we first we review the standard QML estimator in Section 3.1. In Section 3.2, we then detail our adaptive estimator.

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 (on initial values) 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,

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

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

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 $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ almost surely and $\sup_t E(|\varepsilon_t|^q) \leq c < \infty$ 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. If heteroskedasticity arises only through part (a) of Assumption 1 then $\lambda = 1$ and the variance matrix in (5) reduces to C_0 whose form is determined only by the weak dependence (conditional heteroskedasticity and/or weak parametric autocorrelation) present in $\{u_t\}$. Alternatively, if heteroskedasticity arises only through part (b) of Assumption 1, then C_0 reduces to B_0^{-1} which is determined solely by the form of weak parametric autocorrelation present in $\{u_t\}$. $\diamondsuit$

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, which is a measure of the degree of unconditional heteroskedasticity (non-stationary volatility) present in $\{\varepsilon_t\}$ (and, hence, in $\{u_t\}$). 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. Other things being equal, the variance of the asymptotic distribution of the QML estimator is therefore inflated by unconditional heteroskedasticity in $\{\varepsilon_t\}$. $\diamondsuit$

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 $\sqrt{T}(\bar{\theta} - \theta_0) \overset{w}{\to} N(0, C_0)$. 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. As 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 λ 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 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

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

where

$$\hat{\sigma}_t^2 := \sum_{i=1}^{T} k_{ti} \tilde{\varepsilon}_i^2 \qquad \text{with} \qquad k_{ti} := \frac{K(\frac{t-i}{Tb})}{\sum_{i=1}^{T} K(\frac{t-i}{Tb})}. \tag{8}$$

Here $K(\cdot)$ is a bounded, nonnegative, continuous kernel function and $b := b(T)$ is a bandwidth parameter which depends on the sample size, T. 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 need to impose some additional conditions on the kernel function, $K(\cdot)$, and on the bandwidth, b.

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

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

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,*

$$\sqrt{T}(\bar{\theta} - \hat{\theta}) \xrightarrow{p} 0, \tag{9}$$

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{w} N(0, C_0). \tag{10}$$

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 previous works in the field is made possible because our method of proof for Theorem 1, which is novel in the literature, rests on showing that $T^{-1} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \sigma_t^2)^2 \xrightarrow{p} 0$ (see Lemma S.1), rather than establishing the stronger result that $\sup_t |\hat{\sigma}_t^2 - \sigma_t^2| \xrightarrow{p} 0$. The latter necessitates imposing additional smoothness conditions on the kernel function and, in particular, on $\sigma(\cdot)$. For instance, in addition to our Assumption 1(b), Xu and Phillips (2008) follow Cavaliere (2004) and impose that $\sigma(\cdot)$ also satisfies a Lipschitz condition except at a finite number of points of discontinuity; the same condition is also employed in Harris and Kew (2017) and in a series of papers by Patilea and Raïssi (2012, 2013, 2014). In the case of non-stationary time series, Chang and Park (2010) and Beare (2018) require twice differentiability of $\sigma(\cdot)$, while Boswijk and Zu (2018) assume continuity of $\sigma(\cdot)$. In addition, all of these papers (with the exception of Beare, 2018) also need to assume that z_t is conditionally homoskedastic or even that z_t is i.i.d. Conversely, our proof of Theorem 1 allows for conditional heteroskedasticity of a very general form. $\diamond$

Remark 3.5. Adaptive estimation of a function, such as $\sigma(\cdot)$ in our framework, has been pursued in various contexts in the statistics literature on long memory processes, although it is usually assumed to be smooth whereas our $\sigma(\cdot)$ is only assumed càdlàg. For example, Beran and Weiershäuser (2011) estimate a smooth (spline) time-varying location model where the errors are described by a subordinated Gaussian long memory process (for textbook discussions of Gaussian subordination methods, including key references, see, *inter alia*, Giraitis *et al.*, 2012, and Beran *et al.*, 2013). Although these authors focus on a (time-varying) mean, their approach could be implemented in order to estimate, under appropriate conditions, a time-varying volatility process. It is, however, important to stress that our approach differs from theirs, as well as from other approaches based on subordination of a Gaussian long memory process, in that we filter out the time dependence of the series, and in particular the long memory component, prior to adaptive estimation, rather than performing adaptive estimation on the subordinated series itself. $\diamond$

Remark 3.6. 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 $T^{-1/2}$. Noticeably, the condition does not depend on the long memory parameter, d, because adaptive estimation is applied to squared residuals, $\tilde{\varepsilon}_t^2$, in (8). As shown in Lemma S.1 and the proof of Theorem 1, these are close to the squared errors, ε_t^2, which do not display long memory; see also Remark 2.1. 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 (8). We note that, although Wong (1983) assumes independent data, there is some evidence in, e.g., Xu and Phillips (2008) and Harris and Kew (2017) that the method works well under assumptions similar to ours. $\diamond$

Remark 3.7. 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, $\sigma(\cdot)$. $\diamond$

Remark 3.8. Recalling that $\lambda \geq 1$, a comparison of (5) and (10) 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 where the shocks are unconditionally homoskedastic. Thus, the ACSS estimator is asymptotically more efficient than the QML estimator. $\diamond$

Remark 3.9. 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. $\diamond$

Remark 3.10. 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. $\diamond$

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 on the elements of θ 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 (10) 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,

$$\hat{C} := \left(\left(\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} \frac{\partial q_t(\theta)}{\partial\theta} \frac{\partial q_t(\theta)}{\partial\theta'} \right) \left(\frac{\partial^2 Q_T(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \right) \Bigg|_{\theta=\hat{\theta}} \tag{11}$$

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 (10), 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,

$$H_0 : \xi := f(\theta) = 0, \tag{12}$$

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

$$H_{1,T} : \xi_T = \delta/\sqrt{T}, \tag{13}$$

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 (12) can be tested using the familiar Wald statistic,

$$W_T := T f(\hat{\theta})'(F(\hat{\theta})\hat{C}F(\hat{\theta})')^{-1} f(\hat{\theta}), \tag{14}$$

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

$$t_T := \frac{\sqrt{T} f(\hat{\theta})}{\sqrt{F(\hat{\theta})\hat{C}F(\hat{\theta})'}} \tag{15}$$

with H_0 rejected in favour of H_1 for large absolute values of this statistic. The statistic in (15) 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 (15) 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 (12) and provided $F(\theta_0)$ is of full row rank, $W_T \overset{w}{\to} \chi^2(q)$ and $t_T \overset{w}{\to} N(0,1)$.*

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.

15

We proceed to discuss asymptotic local power and optimality of the tests under the assumption that z_t is Gaussian; c.f. Remark 3.10 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 (13) 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, $W_T \overset{w}{\to} \chi_q^2\left(\delta'(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1}\delta\right)$ and $t_T \overset{w}{\to} N(\delta(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1/2}, 1)$, where $\chi_q^2(g)$ indicates a noncentral χ_q^2 distribution with noncentrality parameter g.*

Remark 4.1. As with Remark 3.10, 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). $\diamond$

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 (11), 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 $\tilde{C} - \lambda C_0 \overset{p}{\to} 0$. Consequently, defining $\tilde{W}_T$ and $\tilde{t}_T$ as in (14) and (15) but now based on $\tilde{\theta}$ and $\tilde{C}$, it follows that, under the additional assumption that z_t is Gaussian, $\tilde{W}_T \overset{w}{\to} \chi_q^2\left(\lambda^{-1}\delta'(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1}\delta\right)$ and $\tilde{t}_T \overset{w}{\to} N(\lambda^{-1/2}\delta(F(\theta_0)B_0^{-1}F(\theta_0)')^{-1/2}, 1)$. 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. $\diamond$

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. $\diamond$

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

$$X_t^* := \Delta_+^{-\hat{d}} u_t^* \text{ with } u_t^* := a(L, \hat{\psi})\varepsilon_t^*, \quad t = 1, ..., T, \tag{16}$$

and the bootstrap ACSS estimator is given by

$$\hat{\theta}^* := \arg \min_{\theta \in \Theta} \hat{Q}_T^*(\theta), \quad \hat{Q}_T^*(\theta) := T^{-1} \sum_{t=1}^{T} \frac{\varepsilon_t^*(\theta)^2}{\hat{\sigma}_t^{*2}}, \tag{17}$$

where

$$\varepsilon_t^*(\theta) := \sum_{n=0}^{t-1} b_n(\psi) \Delta_+^d X_{t-n}^* \tag{18}$$

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). ◇

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. ◇

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

$$\check{\theta}^* := \arg\min_{\theta \in \Theta} \check{Q}_T^*(\theta), \quad \check{Q}_T^*(\theta) := T^{-1} \sum_{t=1}^{T} \frac{\varepsilon_t^*(\theta)^2}{\hat{\sigma}_t^2}, \tag{19}$$

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 estimator in (19) 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, but would be expected to display inferior finite sample performance. All of the large sample results given below for $\hat{\theta}^*$ also hold for $\check{\theta}^*$. ◇

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

$$\hat{C}^* := \left. \left(\left(\frac{\partial^2 \hat{Q}_T^*(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \left(T^{-1} \sum_{t=1}^{T} \frac{\partial q_t^*(\theta)}{\partial\theta} \frac{\partial q_t^*(\theta)}{\partial\theta'} \right) \left(\frac{\partial^2 \hat{Q}_T^*(\theta)}{\partial\theta\partial\theta'} \right)^{-1} \right) \right|_{\theta=\hat{\theta}^*} \tag{20}$$

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 $\theta_0 \in \text{int}(\Theta)$. Then,*

$$\sqrt{T}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{w^*}_p N(0, C_0^\dagger) \text{ and } \hat{C}^* \xrightarrow{p^*}_p C_0^\dagger, \tag{21}$$

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

$$\sqrt{T}(\check{\theta}^* - \hat{\theta}^*) \xrightarrow{p^*}_p 0 \text{ and } \check{C}^* - \hat{C}^* \xrightarrow{p^*}_p 0. \tag{22}$$

The large sample result in (21) 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 (10) and (21) 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. $\diamond$

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 (14) and t_T of (15), 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 (21), 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. $\diamond$

Remark 4.9. Wild bootstrap analogues of the robust W_T and t_T statistics of (14) and (15), respectively, are given by

$$W_T^* := T(f(\hat{\theta}^*) - f(\hat{\theta}))'(F(\hat{\theta}^*)\hat{C}^* F(\hat{\theta}^*)')^{-1}(f(\hat{\theta}^*) - f(\hat{\theta})), \tag{23}$$

$$t_T^* := \frac{\sqrt{T}(f(\hat{\theta}^*) - f(\hat{\theta}))}{\sqrt{F(\hat{\theta}^*)\hat{C}^* F(\hat{\theta}^*)'}}. \tag{24}$$

It is immediate from Theorem 3 that these statistics attain the same first order limiting null distributions as 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. $\diamond$

19

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, for a fractionally integrated process allowing for both weak parametric autocorrelation in the model 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, the errors, u_t, will satisfy

$$(1 - a_1 L)u_t = \varepsilon_t \qquad \text{or} \qquad u_t = (1 + a_2 L)\varepsilon_t, \tag{25}$$

where in each case the structure of the innovations $\varepsilon_t = \sigma_t z_t$ will be defined below. Our focus is on 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. Additional simulation results for estimators of d are reported in Section S.4 of the Supplementary Appendix (Cavaliere, Nielsen and Taylor, 2019).

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 3.6. 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 (11)) 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 (19) (reported under $\check{d}^*$) as well as results based on (17). For the results based on (17), 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 $T = 100$, 250 and 500. All confidence intervals are nominal 90% intervals. The variance estimators required in, for example, (11) were implemented using numerical

derivatives. For the bootstrap implementations, 999 bootstrap replications were used and the w_t were generated according to the Rademacher distribution.

5.2 Results With Heteroskedastic, Uncorrelated Errors

We consider first the case where the errors, u_t, are not autocorrelated; i.e. it is known that $a_1 = a_2 = 0$ such that $u_t = \varepsilon_t$. This allows us to 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:

$$\text{Panel A} : z_t = e_t, \qquad \text{Panel B} : z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.2 z_{t-1}^2 + 0.79 h_{t-1},$$

$$\text{Panel C} : z_t = e_t \exp(h_t), h_t = 0.936 h_{t-1} + 0.424 v_t, (v_t, e_t) \sim N(0, I_2).$$

Panel A relates to the case where z_t is conditionally homoskedastic, while Panels B and C contain results for 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 presented; see also Gonçalves and Kilian (2007). 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 uncondition-

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

			Empirical coverage rate (%)						Median length					
τ	υ	T	$\tilde{d}$	$\tilde{d}^*$	$\hat{d}$	$\check{d}^*$	$\hat{d}_1^*$	$\hat{d}_2^*$	$\tilde{d}$	$\tilde{d}^*$	$\hat{d}$	$\check{d}^*$	$\hat{d}_1^*$	$\hat{d}_2^*$
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 ($\tilde{d}$), the wild bootstrap equivalent ($\tilde{d}^*$), the ACSS estimator ($\hat{d}$), the wild bootstrap ACSS in (19) ($\check{d}^*$), the wild bootstrap ACSS estimator in (17), where the bandwidth is re-determined for each bootstrap sample ($\hat{d}_1^*$), and the wild bootstrap ACSS in (17), using the bandwidth from the original sample on each bootstrap sample ($\hat{d}_2^*$). The bootstrap intervals are symmetric studentized bootstrap confidence intervals (see 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_1	a_2	T	$\tilde{d}$	$\tilde{d}^*$	$\hat{d}$	$\check{d}^*$	$\hat{d}_1^*$	$\hat{d}_2^*$	$\tilde{d}$	$\tilde{d}^*$	$\hat{d}$	$\check{d}^*$	$\hat{d}_1^*$	$\hat{d}_2^*$
Panel A: No break, $v = 1$														
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, $\tau = 1/4, v = 1/3$														
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	0.190	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, $\tau = 3/4, v = 3$														
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.

23

ally homoskedastic ($v = 1$). 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, as expected from the results in (5) and Theorem 1, the median length of the confidence intervals is larger under conditional heteroskedasticity than under homoskedasticity. Comparing across methods, it is clear that the ACSS-based $\check{d}^*$ 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 the wild bootstrap confidence intervals based on $\tilde{d}^*$ (QML) and $\hat{d}_1^*$ (ACSS) is approximately unity, regardless of T, when $v = 1$ (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 one of the models in (25) with $a_1, a_2 \in \{-0.8, 0, 0.8\}$. The reported cases where either $a_1 = 0$ or $a_2 = 0$ 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 $a_1 = 0$ (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 ($a_1 = 0.8$) or follows a negative moving average ($a_2 = -0.8$). To illustrate, when $a_2 = -0.8$ 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 ($a_1 = 0.8$ and $a_2 = -0.8$) the wild bootstrap

tends to rather over correct such that the resulting confidence intervals are somewhat too liberal.

As with the presence of conditional heteroskedasticity, the median lengths of the confidence intervals are seen to vary with a_1 and a_2, 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 $a_1 = 0.8$ or $a_2 = -0.8$ than when $a_1 = -0.8$ or $a_2 = 0.8$. 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 report the QML and ACSS estimates of d along with their (robust or sandwich) standard errors and 95% confidence intervals based on (5) and (10), respectively, together with 95% wild bootstrap percentile-t confidence intervals (for the ACSS these re-estimate σ_t and estimate the bandwidth parameter in each bootstrap replication using cross-validation). In each case an ARFIMA(p,d,0) model was fitted to the (de-meaned) data with p chosen by the BIC applied to the adaptive estimation model (choosing over $p = 0, \ldots, 10$). Plots of all data series and 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.5 in the Supplementary Appendix (Cavaliere, Nielsen and Taylor, 2019). Both the conditional and unconditional heteroskedasticity tests reported display highly significant rejections for most of the series, providing strong evidence of heteroskedastic behaviour in these data sets.

6.1 Realised Variance Data

We first revisit a classic data set in finance from Andersen *et al.* (2007). The data set consists of daily realised variance [RV], based on intra-daily five-minute return data for the S&P500 index, DM/$ exchange rate futures, and 30-year US Treasury bond futures. The sample periods are

Table 3: Andersen *et al.* (2007) Realised Variance Data

| | | QML | | | | | ACSS | | | | |
	p	Est. d	se(d)	ACI_L	ACI_U	WBCI_L	WBCI_U	Est. d	se(d)	ACI_L	ACI_U	WBCI_L	WBCI_U
S&P500	0	0.321	0.046	0.231	0.412	0.211	0.431	0.360	0.039	0.284	0.436	0.276	0.444
DM/$	0	0.382	0.032	0.320	0.444	0.319	0.445	0.390	0.017	0.357	0.423	0.358	0.422
T-bond	0	0.186	0.015	0.156	0.215	0.155	0.216	0.198	0.015	0.168	0.228	0.168	0.228

Notes: Daily data from January 2, 1990 to December 31, 2002 (SP500 and Tbond) and December 2, 1986 to June 30, 1999 (DM/$). The sample sizes are $T = 3262$ and $T = 3045$, respectively. An ARFIMA(p,d,0) model was estimated on the de-meaned data. se(d) is the (robust) standard error, ACI are asymptotic confidence intervals and WBCI are wild bootstrap confidence intervals (for ACSS re-estimating σ_t and estimating the bandwidth parameter in each bootstrap replication using cross-validation). All intervals are 95% nominal level.

January 2, 1990 to December 31, 2002 for the S&P500 and T-bond data ($T = 3262$), and December 2, 1986 to June 30, 1999 for the DM/$ data ($T = 3045$). Detailed data descriptions can be found in Andersen *et al.* (2007, Section III.A).

The results are reported in Table 3. Both the QML and ACSS estimates of d are in line with the typical 0.3–0.4 found in the literature, although the estimates for the T-bond RV data are slightly lower. For the S&P500 RV data and the DM/$ RV data, the standard errors for ACSS are much smaller than those for QML, which translates into commensurately smaller confidence interval widths for d when based on ACSS rather than QML estimation. However, for the T-bond RV data they are the same. This is exactly as expected from the theoretical results in (5) and Theorem 1, given that the unconditional heteroskedasticity tests reject the null of unconditional homoskedasticity for the S&P500 and DM/$ data, but do not reject for the T-bond data. Thus, for the S&P500 and DM/$ data, we can draw much more accurate inferences from the ACSS estimator compared with the QML estimator, in the sense of smaller standard errors and narrower confidence intervals. Finally, the wild bootstrap confidence intervals are very similar to the asymptotic confidence intervals, although the former are slightly wider for the S&P500 data.

We next analyse a similar, but more recent, data set consisting of the RV for 28 of the 30 stocks in the U.S. market which comprise the Dow Jones Industrial Average (tickers TRV and V have missing data issues and, hence, are omitted). The RVs are calculated from continuously compounded intra-daily five-minute returns based on prices obtained from the NYSE Trades and Quotes (TAQ) database. The sample period is January 2, 2003 to December 31, 2014 for a sample size of $T = 3021$.

The results for this data set are presented in Table 4. The overall conclusions are in line with those for the classic data set in Table 3. Specifically, the estimates of d are all about 0.3–0.4, the ACSS standard errors are smaller than the corresponding QML standard errors in every case,

Table 4: Dow Jones Realised Variance Data

| | | QML | | | | | | ACSS | | | | |
	p	Est. d	se(d)	ACI_L	ACI_U	$WBCI_L$	$WBCI_U$	Est. d	se(d)	ACI_L	ACI_U	$WBCI_L$	$WBCI_U$
AAPL	0	0.322	0.020	0.283	0.360	0.281	0.362	0.319	0.016	0.288	0.351	0.286	0.353
AXP	0	0.405	0.025	0.356	0.453	0.352	0.457	0.374	0.013	0.349	0.398	0.347	0.400
BA	0	0.345	0.020	0.305	0.385	0.307	0.383	0.316	0.014	0.287	0.344	0.285	0.346
CAT	0	0.359	0.023	0.315	0.404	0.316	0.403	0.336	0.015	0.306	0.365	0.305	0.366
CSCO	0	0.339	0.022	0.296	0.381	0.295	0.383	0.320	0.015	0.290	0.349	0.291	0.349
CVX	0	0.423	0.038	0.348	0.499	0.344	0.503	0.376	0.024	0.329	0.422	0.328	0.423
DD	0	0.356	0.022	0.313	0.399	0.309	0.403	0.321	0.016	0.290	0.351	0.290	0.351
DIS	0	0.382	0.024	0.335	0.430	0.336	0.428	0.350	0.018	0.315	0.385	0.315	0.385
GE	0	0.410	0.036	0.340	0.479	0.335	0.485	0.355	0.016	0.324	0.386	0.324	0.387
GS	1	0.477	0.074	0.332	0.622	0.324	0.631	0.433	0.027	0.380	0.486	0.380	0.486
HD	0	0.358	0.027	0.304	0.412	0.303	0.413	0.330	0.015	0.300	0.359	0.301	0.359
IBM	0	0.380	0.022	0.337	0.424	0.336	0.425	0.338	0.016	0.307	0.368	0.308	0.367
INTC	0	0.330	0.018	0.296	0.365	0.296	0.365	0.310	0.013	0.284	0.335	0.286	0.333
JNJ	0	0.376	0.033	0.313	0.442	0.313	0.441	0.351	0.026	0.301	0.401	0.302	0.401
JPM	0	0.459	0.039	0.382	0.536	0.383	0.535	0.396	0.017	0.363	0.428	0.364	0.428
KO	0	0.369	0.029	0.313	0.426	0.307	0.431	0.331	0.020	0.292	0.369	0.291	0.370
MCD	0	0.314	0.032	0.252	0.376	0.242	0.386	0.325	0.022	0.282	0.367	0.280	0.370
MMM	0	0.317	0.036	0.246	0.388	0.236	0.397	0.313	0.019	0.275	0.351	0.275	0.352
MRK	0	0.337	0.028	0.281	0.392	0.283	0.390	0.322	0.018	0.286	0.357	0.286	0.357
MSFT	0	0.312	0.020	0.272	0.352	0.270	0.353	0.280	0.016	0.249	0.311	0.249	0.312
NKE	0	0.349	0.024	0.303	0.396	0.302	0.397	0.319	0.016	0.288	0.350	0.288	0.350
PFE	0	0.295	0.024	0.249	0.342	0.247	0.344	0.291	0.020	0.253	0.329	0.249	0.333
PG	0	0.238	0.057	0.125	0.350	0.072	0.404	0.267	0.026	0.216	0.319	0.211	0.323
UNH	0	0.348	0.039	0.272	0.424	0.270	0.426	0.318	0.019	0.281	0.355	0.281	0.355
UTX	0	0.360	0.027	0.308	0.412	0.306	0.414	0.323	0.017	0.289	0.356	0.291	0.354
VZ	0	0.357	0.028	0.301	0.413	0.298	0.415	0.324	0.018	0.289	0.360	0.287	0.361
WMT	0	0.350	0.036	0.280	0.420	0.276	0.424	0.312	0.021	0.271	0.353	0.271	0.352
XOM	0	0.436	0.042	0.354	0.517	0.351	0.520	0.381	0.025	0.332	0.430	0.332	0.430

Notes: Daily data from January 2, 2003 to December 31, 2014 ($T = 31021$)y. An ARFIMA(p,d,0) model was estimated on the de-meaned data. se(d) is the (robust) standard error, ACI are asymptotic confidence intervals and WBCI are wild bootstrap confidence intervals (for ACSS re-estimating σ_t and estimating the bandwidth parameter in each bootstrap replication using cross-validation). All intervals are 95% nominal level.

and the wild bootstrap and asymptotic confidence intervals are very similar with few exceptions (most notably for ticker PG). Most interestingly, for tickers CVX, GS, JPM, and XOM, the QML confidence intervals include the value $d = 1/2$, which defines the boundary between stationarity and nonstationarity in standard fractional time series models, and hence $d \geq 1/2$ cannot be excluded based on QML inference. However, for each of those series, due to a combination of slightly lower estimates of d and much narrower confidence intervals, the ACSS confidence intervals do not include $d = 1/2$, so based on the ACSS we can reject that $d \geq 1/2$ at the 5% level.

6.2   Government Bond Yield 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 daily data on 10-year government bond yields from September 5, 2005, to February

Table 5: Government Bond Yield Data

		QML						ACSS					
	p	Est. d	se(d)	ACI$_L$	ACI$_U$	WBCI$_L$	WBCI$_U$	Est. d	se(d)	ACI$_L$	ACI$_U$	WBCI$_L$	WBCI$_U$
Belgium	0	1.126	0.040	1.047	1.206	1.049	1.204	1.093	0.026	1.043	1.144	1.043	1.144
Finland	1	0.961	0.026	0.911	1.011	0.909	1.014	0.998	0.022	0.956	1.041	0.957	1.039
France	0	1.005	0.019	0.968	1.042	0.968	1.042	1.008	0.017	0.975	1.041	0.974	1.042
Germany	1	0.971	0.025	0.922	1.020	0.923	1.019	0.992	0.022	0.948	1.036	0.948	1.036
Ireland	1	1.052	0.071	0.912	1.192	0.890	1.213	1.008	0.028	0.952	1.064	0.951	1.065
Italy	1	0.896	0.042	0.814	0.979	0.814	0.979	0.928	0.026	0.877	0.979	0.877	0.979
Portugal	1	0.965	0.062	0.843	1.087	0.834	1.096	0.970	0.022	0.928	1.013	0.926	1.014
Spain	1	0.872	0.038	0.797	0.947	0.794	0.950	0.935	0.025	0.886	0.984	0.884	0.986

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) is the (robust) standard error, ACI are asymptotic confidence intervals and WBCI are wild bootstrap confidence intervals (for ACSS re-estimating σ_t and estimating the bandwidth parameter in each bootstrap replication using cross-validation). All intervals are 95% nominal level.

2, 2016, with $T = 2715$; see Martins and Amado (2016, Section 3.1) for detailed data descriptions.

The ACSS and QML estimates of d, reported in Table 5 are similar for each country considered with estimated values for all countries lying relatively close to 1. As for the RV data sets, the ACSS standard errors are smaller, in many cases substantially so, for each country than the corresponding QML ones. 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

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 applications to realised variance and government bond yield data were provided to illustrate the improvements in inference that the adaptive methods discussed in this paper can deliver when applied to economic and financial data displaying non-constant variance.

Finally, although we have chosen to focus attention on developing inference methods for the elements of θ, our results also facilitate running standard diagnostic (misspecification) tests on the residuals from the estimated model, such as tests for unmodelled serial correlation, provided they are implemented with either heteroskedasticity-robust standard errors or the wild bootstrap.

Acknowledgements

References

Andersen, T.G., T. Bollerslev and F.X. Diebold (2007), Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility, *Review of Economics and Statistics* 89, 701–720.

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.

Barndorff-Nielsen, O.E. and N. Shephard (2002), Econometric analysis of realised volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society Series B* 64, 253–280.

Beare, B. (2018), Unit root testing with unstable volatility, *Journal of Time Series Analysis* 39, 816–835.

Beran, J., Y. Feng, S. Ghosh and R. Kulik (2013), *Long-Memory Processes: Probabalistic Properties and Statistical Methods*, Springer, Berlin.

Beran, J. and A. Weershäuser (2011), On spline regression under Gaussian subordination with long memory, *Journal of Multivariate Analysis* 102, 315–335.

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

Boswijk, H.P. and Y. Zu (2018), Adaptive wild bootstrap tests for a unit root with non-stationary volatility, *Econometrics Journal* 21, 87–113.

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

Cavaliere, G. (2004), Unit root tests under time-varying variances, *Econometric Reviews* 23, 259–292.

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

Cavaliere, G., 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.

Cavaliere, G., M.Ø. Nielsen and A.M.R. Taylor (2019), Supplementary Appendix: Adaptive inference in heteroskedastic fractional time series models, QED working paper 1390, Queen's University.

Chang, S.K. and J.Y. Park (2010), Cointegrating regressions with time heterogeneity, *Econometric Reviews* 29, 397–438.

Corsi, F., S. Mittnik, C. Pigorsch and U. Pigorsch (2008), The volatility of realized volatility,

Econometric Reviews 27, 46–78.

Eicker, F. (1967), Limit theorems for regression with unequal and dependent errors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* 1, 59–82, University of California Press, Berkeley.

Fan, J. and Q. Yao (1998), Efficient estimation of conditional variance functions in stochastic regression, *Biometrika* 85, 645–660.

Fox, R. and M.S. Taqqu (1986), Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics* 14, 517–532.

Geweke, J. and S. Porter-Hudak (1983), The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4, 221–238.

Giraitis, L., H.L. Koul and D. Surgailis (2012), *Large Sample Inference for Long Memory Processes*, Imperial College Press, London.

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

Gonçalves, S. and L. Kilian (2007), Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity, *Econometric Reviews* 26, 609–641.

Guo, H. and H.L. Koul (2008), Asymptotic inference for some regression models under heteroscedasticity and long memory design and errors, *Annals of Statistics* 36, 458–487.

Hall, P. and N. Tajvidi (2000), Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data, *Statistical Science* 15, 153–167.

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* 1, 221–233, 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.

Patilea, V. and H. Raïssi (2012), Adaptive estimation of vector autoregressive models with time varying variance: application to testing linear causality in mean, *Journal of Statistical Planning and Inference* 142, 2891–2912.

Patilea, V. and H. Raïssi (2013), Corrected portmanteau tests for VAR models with time-varying variance, *Journal of Multivariate Analysis* 116, 190–207.

Patilea, V. and H. Raïssi (2014), Testing second-order dynamics for autoregressive processes in presence of time-varying variance, *Journal of the American Statistical Association* 109, 1099–1111.

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.

Robinson, P.M. (1995), Gaussian semiparametric estimation of long range dependence, *Annals of Statistics* 23, 1630–1661.

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.

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

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.

Yu, K. and M.C. Jones (2004), Likelihood-based local linear estimation of the conditional variance function, *Journal of the American Statistical Association* 99, 139–144.

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.

Supplementary Appendix:

Adaptive Inference in Heteroskedastic Fractional Time Series Models

by

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

August 23, 2019

S.1 Introduction

This supplement to Cavaliere, Nielsen and Taylor (2019) has four main sections. Section S.2 contains the proofs of the main theorems from the paper. Section S.3 contains Lemmas S.1 and S.2, which describe the key properties of the adaptive volatility estimators, as well as some technical lemmas, which are used to prove the main results. Section S.4 contains additional Monte Carlo results relating to the bias and root-mean-squared-error of the QML, ACSS and local Whittle estimators of the long memory parameter, d. Finally, Section S.5 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 Proofs of Main Theorems

S.2.1 Proof of Theorem 1

S.2.1.1 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 $D_1 := D_1(\kappa) = D \cap \{d : d - d_0 \leq -1/2 + \kappa\}$ and $D_2 := D_2(\kappa) = D \cap \{d : d - d_0 \geq -1/2 + \kappa\}$ 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 $c > 0$ there exists a (fixed) $\bar{\kappa} > 0$ such that

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

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

However, this follows easily from the lower bound $\hat{Q}_T(\theta) \geq (\max_{1 \leq s \leq T} \hat{\sigma}_s^2)^{-1} T^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta)^2$, because (S.1) is proven for $T^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta)^2$ in CNT and $\max_{1 \leq s \leq T} \hat{\sigma}_s^2 = O_p(1)$ by Lemma S.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 (S.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 (S.1) and (S.2), the desired results follow if, for any $\kappa > 0$,

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

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

$$|\hat{Q}_T(\theta) - \bar{Q}_T^0(\theta)| = \left| T^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta)^2 \frac{(\sigma_t^2 - \hat{\sigma}_t^2)}{\hat{\sigma}_t^2 \sigma_t^2} \right|$$

$$\leq (\min_{1 \leq t \leq T} \hat{\sigma}_t^2 \sigma_t^2)^{-1} \left(T^{-1} \sum_{t=1}^{T} \varepsilon_t(\theta)^4 \right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} (\sigma_t^2 - \hat{\sigma}_t^2)^2 \right)^{1/2}$$

by the Cauchy-Schwarz inequality. The first term on the right-hand side is $O_p(1)$ by Lemma S.1(g) and uniform strict positivity of σ_t^2 (Assumption 1(b)(ii)). For the second and third terms we apply Lemmas S.5 and S.1(d), respectively.

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

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

where the coefficients $\phi_j(\theta)$ and remainder term $r_t(\theta)$ are subject to the bounds in Lemma S.4. We then split the infinite summation in $\bar{Q}_T^0(\theta)$ and find

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

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

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

3

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

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

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

using Lemma S.4 and Lemma S.3 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.3.

For (S.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.4, and then

$$\sum_{k=j+1}^{T-1} \phi_k(d_0 - d, \psi) \sum_{t=k+1}^{T-1} \varepsilon_{t-j} \varepsilon_{t-k} \frac{\sigma_{t-j} \sigma_{t-k} - \sigma_t^2}{\sigma_{t-j} \sigma_{t-k} \sigma_t^2}$$

$$= \phi_T(d_0 - d, \psi) \sum_{k=j+1}^{T-1} \sum_{t=k+1}^{T-1} \varepsilon_{t-j} \varepsilon_{t-k} \frac{\sigma_{t-j} \sigma_{t-k} - \sigma_t^2}{\sigma_{t-j} \sigma_{t-k} \sigma_t^2}$$

$$- \sum_{l=j+1}^{T-2} \phi_{l+1}(d_0 - d - 1, \psi) \sum_{k=j+1}^{l} \sum_{t=k+1}^{T-1} \varepsilon_{t-j} \varepsilon_{t-k} \frac{\sigma_{t-j} \sigma_{t-k} - \sigma_t^2}{\sigma_{t-j} \sigma_{t-k} \sigma_t^2},$$

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 $E|\sum_{t=k+1}^{T-1} v_{t,j,l}| \le cT^{1/2} l^{1/2}$, uniformly in k, j. Thus, using Lemma S.4 we obtain the bound

$$E \sup_{\theta \in \Theta_2} |(S.6)| = E \sup_{\theta \in \Theta_2} |T^{-1} \sum_{j=0}^{T-1} \phi_j(d_0 - d, \psi) \sum_{k=j+1}^{T-1} \phi_k(d_0 - d, \psi) \sum_{t=k+1}^{T-1} \varepsilon_{t-j} \varepsilon_{t-k} \frac{\sigma_{t-j} \sigma_{t-k} - \sigma_t^2}{\sigma_{t-j} \sigma_{t-k} \sigma_t^2}|$$

$$\le cT^{-1} \sum_{j=1}^{T-1} j^{-1/2-\kappa} T^{-1/2-\kappa} T + cT^{-1} \sum_{j=1}^{T-1} j^{-1/2-\kappa} \sum_{l=j+1}^{T-2} l^{-3/2-\kappa} T^{1/2} l^{1/2} \le cT^{-2\kappa}.$$

For the proofs of (S.7) and (S.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.4, and thus the same proof applies to each term in (S.7) and to each term in (S.8). For (S.7) we find that $E \sup_{\theta \in \Theta_2} |(S.7)| \le$

$cT^{-1}\sum_{t=1}^{T}t^{-1-2\kappa}\leq cT^{-1}$ by Lemma S.4, while for (S.8) we find

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

by Lemma S.4. This proves the second statement in (S.3), and hence completes the proof.

S.2.1.2 Proof of asymptotic normality We show that

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

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

which together imply both (9) and (10) by (5). The proof of (S.9) follows by the same argument as that of (S.3), noting that the derivatives add at most a logarithmic term, see (2) and Lemma S.4.

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

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

where r_t is subject to the bound in Lemma S.4. We then let $v_t:=\varepsilon_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}$ and find that

$$\sqrt{T}\frac{\partial\hat{Q}_T(\theta_0)}{\partial\theta}-\sqrt{T}\frac{\partial\bar{Q}_T(\theta_0)}{\partial\theta}=2T^{-1/2}\sum_{t=1}^{T}r_t\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}(\hat{\sigma}_t^{-2}-\sigma_t^{-2})\tag{S.12}$$

$$+2T^{-1/2}\sum_{t=1}^{T}v_t(\hat{\sigma}_t^{-2}-\tilde{\sigma}_t^{-2})\tag{S.13}$$

$$+2T^{-1/2}\sum_{t=1}^{T}v_t(\tilde{\sigma}_t^{-2}-\bar{\sigma}_t^{-2})\tag{S.14}$$

$$+2T^{-1/2}\sum_{t=1}^{T}v_t(\bar{\sigma}_t^{-2}-\sigma_t^{-2}).\tag{S.15}$$

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

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

5

where the first term is $O_p(T^{-1/4})$ by Lemma S.4 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 S.1(d),(g), so that $|(S.12)_i| = o_p(1)$.

For (S.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

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

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

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

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

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

by Assumption 1 and Lemma S.4.

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

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

where $T^{-1}\sum_{t=1}^{T}v_{it}^2 = O_p(1)$ by (S.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 S.1(c),(g) and Assumption 7.

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

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

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

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

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

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

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

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

by Lemmas S.1(b),(g) and S.4 and Assumptions 1,7. Next, (S.18) has second moment

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

$$\times E(z_tz_sz_{t-n}z_{s-m}(z_{j_1}^2 - 1)(z_{j_2}^2 - 1)).$$

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.6 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

7

is $T^{-1} \sum_{t=1}^{T} \sup_{1 \le j_1, j_2 \le T} k_{tj_1} k_{tj_2} \le c(Tb)^{-2}$ because

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

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, $(S.18) \overset{p}{\to} 0$ by Assumption 7, which shows that $(S.17) \overset{p}{\to} 0$ and hence proves the first statement of (S.10).

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

$$\sqrt{T} \frac{\partial \bar{Q}_T(\theta_0)}{\partial \theta_i} - \sqrt{T} \frac{\partial \bar{Q}_T^0(\theta_0)}{\partial \theta_i} = 2T^{-1/2} \sum_{t=1}^{T} \sigma_t^{-2} (\varepsilon_t + r_t) \left(\sum_{j=0}^{t-1} \frac{\partial \phi_j(\theta_0)}{\partial \theta_i} \varepsilon_{t-j} + \frac{\partial r_t(\theta_0)}{\partial \theta_i} \right)$$

$$- 2T^{-1/2} \sum_{t=1}^{T} (\sigma_t^{-1} \varepsilon_t + \tilde{r}_t) \left(\sum_{j=0}^{t-1} \sigma_{t-j}^{-1} \frac{\partial \phi_j(\theta_0)}{\partial \theta_i} \varepsilon_{t-j} + \frac{\partial \tilde{r}_t(\theta_0)}{\partial \theta_i} \right)$$

$$= 2T^{-1/2} \sum_{t=1}^{T} \sum_{j=0}^{t-1} \frac{\partial \phi_j(\theta_0)}{\partial \theta_i} z_t z_{t-j} \left(\frac{\sigma_{t-j}}{\sigma_t} - 1 \right) \tag{S.21}$$

$$+ 2T^{-1/2} \sum_{t=1}^{T} z_t \left(\sigma_t^{-1} \frac{\partial r_t(\theta_0)}{\partial \theta_i} - \frac{\partial \tilde{r}_t(\theta_0)}{\partial \theta_i} \right) \tag{S.22}$$

$$+ 2T^{-1/2} \sum_{t=1}^{T} \left(\sigma_t^{-2} r_t \sum_{j=0}^{t-1} \frac{\partial \phi_j(\theta_0)}{\partial \theta_i} \varepsilon_{t-j} - \tilde{r}_t \sum_{j=0}^{t-1} \frac{\partial \phi_j(\theta_0)}{\partial \theta_i} z_{t-j} \right) \tag{S.23}$$

$$+ 2T^{-1/2} \sum_{t=1}^{T} \left(\sigma_t^{-2} r_t \frac{\partial r_t(\theta_0)}{\partial \theta_i} - \tilde{r}_t \frac{\partial \tilde{r}_t(\theta_0)}{\partial \theta_i} \right). \tag{S.24}$$

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.4. For example, for (S.22) we find

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

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

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

The last term converges to zero by Lemma S.3 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.4) 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.3 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.3 in view of Assumption 1(a)(iii). This concludes the proof of the second statement of (S.10) and hence that of (10).

S.2.2 Proof of Theorem 2

We first consider

$$
\begin{aligned}
\hat{A} =&\frac{1}{4}T^{-1}\sum_{t=1}^{T}\frac{\partial q_t(\hat{\theta})}{\partial\theta}\frac{\partial q_t(\hat{\theta})}{\partial\theta'} = \frac{1}{4}T^{-1}\sum_{t=1}^{T}\hat{\sigma}_t^{-4}\frac{\partial\varepsilon_t(\hat{\theta})^2}{\partial\theta}\frac{\partial\varepsilon_t(\hat{\theta})^2}{\partial\theta'}\\
=&\frac{1}{4}T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^{-4}-\sigma_t^{-4})\frac{\partial\varepsilon_t(\hat{\theta})^2}{\partial\theta}\frac{\partial\varepsilon_t(\hat{\theta})^2}{\partial\theta'} \quad\quad\quad\quad (S.25)\\
&+T^{-1}\sum_{t=1}^{T}\sigma_t^{-4}\left(\varepsilon_t(\hat{\theta})^2\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta'}-\varepsilon_t(\theta_0)^2\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}\right) \quad (S.26)\\
&+T^{-1}\sum_{t=1}^{T}\sigma_t^{-4}\varepsilon_t(\theta_0)^2\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta'}. \quad\quad\quad\quad\quad (S.27)
\end{aligned}
$$

By the Cauchy-Schwarz inequality, the (i,j)'th element of (S.25) satisfies

$$
|(S.25)_{i,j}| \leq \frac{1}{4}\left(T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^{-4}-\sigma_t^{-4})^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\left(\frac{\partial\varepsilon_t(\hat{\theta})^2}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})^2}{\partial\theta_j}\right)^2\right)^{1/2},
$$

where the first term is $o_p(1)$ by Assumption 1(b) and Lemma S.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.5 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 (S.26) and apply the Cauchy-Schwarz inequality,

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

$$\leq \left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}(\varepsilon_t(\hat{\theta})^2 - \varepsilon_t(\theta_0)^2)^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_j}\right)^2\right)^{1/2} \tag{S.28}$$

$$+ \left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}\varepsilon_t(\theta_0)^4\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^4}\left(\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_i}\frac{\partial\varepsilon_t(\hat{\theta})}{\partial\theta_j} - \frac{\partial\varepsilon_t(\theta_0)}{\partial\theta_i}\frac{\partial\varepsilon_t(\theta_0)}{\partial\theta_j}\right)^2\right)^{1/2}. \tag{S.29}$$

The proofs for (S.28) and (S.29) are nearly identical, so we give only the former. The second large parenthesis in (S.28) is $O_p(1)$ by Lemma S.5 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.5. Because $\hat{\theta}_i - \theta_{0,i} = O_p(T^{-1/2})$ by Theorem 1 and using Assumption 1(b)(ii), it follows that (S.28) is $o_p(1)$. Next, $(S.27) \xrightarrow{p} A_0$ by the same arguments as applied to the second term of (S.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 (S.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 $\hat{C} = \hat{B}^{-1}\hat{A}\hat{B}^{-1} \xrightarrow{p} B_0^{-1}A_0B_0^{-1} = C_0$.

S.2.3 Proof of Theorem 3

S.2.3.1 Proof of consistency As in the proof of consistency in Theorem 1, see Section S.2.1.1, 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,

$$P(D_1^0 \supseteq \hat{D}_1) = P(\hat{d} - d_0 \leq \kappa) \to 1, \tag{S.30}$$

$$P(D_2^0 \supseteq \hat{D}_2) = P(d_0 - \hat{d} \leq \kappa/2) \to 1. \tag{S.31}$$

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 $c > 0$ there exists a (fixed) $\bar{\kappa} > 0$ such that

$$P^*(\inf_{\theta \in \hat{\Theta}_1(\bar{\kappa})} \hat{Q}_T^*(\theta) > c) \xrightarrow{p} 1 \text{ as } T \to \infty, \tag{S.32}$$

$$P^*(\inf_{\theta \in \hat{\Theta}_1(\bar{\kappa})} \check{Q}_T^*(\theta) > c) \xrightarrow{p} 1 \text{ as } T \to \infty. \tag{S.33}$$

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 (S.1) and (S.2), the results (S.32) and (S.33) follow from the bounds in (D.12) of CNT and Lemma S.2(g^*).

In view of (S.32) and (S.33), it follows that $\hat{\theta}^* - \check{\theta}^* \xrightarrow{p^*}_p 0$ because

$$\sup_{\theta \in \hat{\Theta}_2} |\hat{Q}_T^*(\theta) - \check{Q}_T^*(\theta)| \leq \sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{t=1}^{T} \varepsilon_t^*(\theta)^2 \frac{|\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2}|}{\hat{\sigma}_t^2 \hat{\sigma}_t^{*2}} \tag{S.34}$$

$$\leq (\min_t \hat{\sigma}_t^2 \hat{\sigma}_t^{*2})^{-1} \sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{t=1}^{T} \varepsilon_t^*(\theta)^2 |\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2}|$$

$$\leq (\min_t \hat{\sigma}_t^2 \hat{\sigma}_t^{*2})^{-1} \left(\sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{t=1}^{T} \varepsilon_t^*(\theta)^4 \right)^{1/2} \left(T^{-1} \sum_{t=1}^{T} (\hat{\sigma}_t^2 - \hat{\sigma}_t^{*2})^2 \right)^{1/2},$$

11

which is $O_p^*(1)O_p^*(1)o_p^*(1) = o_p^*(1)$, in probability, by Lemmas S.1(g), S.2(g*), S.12, and S.2(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

$$\hat{\theta} - \tilde{\theta}^\dagger \xrightarrow{p} 0 \tag{S.35}$$

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

$$\arg\min_{\theta \in \hat{\Theta}_2} \check{Q}_T^*(\theta) - \tilde{\theta}^\dagger \xrightarrow{p^*}_p 0. \tag{S.36}$$

With P^*-probability converging to one in probability, the first term in (S.36) is $\check{\theta}^*$, see (S.33), so that the required result follows by combining (S.35) and (S.36). We therefore prove that (for any $\kappa > 0$)

$$\sup_{\theta \in \hat{\Theta}_2} |\check{Q}_T^*(\theta) - \hat{Q}_T(\theta)| \xrightarrow{p^*}_p 0, \tag{S.37}$$

which implies (S.36).

To show (S.37) we decompose

$$\check{Q}_T^*(\theta) - \hat{Q}_T(\theta) = \check{Q}_T^*(\theta) - E^*\check{Q}_T^*(\theta) \tag{S.38}$$

$$+ E^*\check{Q}_T^*(\theta) - \hat{Q}_T(\theta) \tag{S.39}$$

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.11. By uncorrelatedness of ε_t^* conditional on the original data,

$$(S.38) = T^{-1} \sum_{t=1}^{T} \frac{1}{\hat{\sigma}_t^2} \sum_{n=0}^{t-1} \hat{\phi}_n(\theta)^2 (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{t-n}^2) \tag{S.40}$$

$$+ 2T^{-1} \sum_{t=1}^{T} \frac{1}{\hat{\sigma}_t^2} \sum_{n=0}^{t-1} \sum_{m=n+1}^{t-1} \hat{\phi}_n(\theta)\hat{\phi}_m(\theta)\varepsilon_{t-n}^*\varepsilon_{t-m}^*. \tag{S.41}$$

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 $\eta_4 := E((w_t^2 - 1)^2)$,

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

$$= \sum_{t=n+1}^{T} \frac{1}{\hat{\sigma}_t^4} E^* (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{t-n}^2)^2$$

$$\leq \frac{1}{\min_t \hat{\sigma}_t^4} \eta_4 \sum_{t=n+1}^{T} \hat{\varepsilon}_{t-n}^4 =: C_T^2 = O_p(T) \tag{S.42}$$

uniformly in $0 \leq n \leq T-1$ by Lemmas S.1(g) and S.5. Thus, reversing the order of the summations in (S.40) and using (S.42), we find

$$E^* \sup_{\theta \in \hat{\Theta}_2} |(S.40)| \leq \sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{n=0}^{T-1} \hat{\phi}_n(\theta)^2 E^* \left| \sum_{t=n+1}^{T} \frac{1}{\hat{\sigma}_t^2} (\varepsilon_{t-n}^{*2} - \hat{\varepsilon}_{t-n}^2) \right|$$

$$\leq C_T \sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{n=0}^{T-1} \hat{\phi}_n(\theta)^2 \leq C_T O_p(T^{-1} \sum_{n=1}^{T-1} n^{-1-2\kappa}) = O_p(T^{-1/2}),$$

which shows that $\sup_{\theta \in \hat{\Theta}_2} |(S.40)| = O_p^*(T^{-1/2})$, in probability.

To deal with (S.41), we apply Lemmas S.1(g) and S.10 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

$$E^* \sup_{\theta \in \hat{\Theta}_2} |(S.41)| = \sup_{\theta \in \hat{\Theta}_2} T^{-1} \sum_{n=1}^{T-1} |\hat{\phi}_n(\theta)| O_p(T^{1/2} n^{-\kappa}) = O_p(T^{-1/2}) \sum_{n=1}^{T-1} n^{-1/2-2\kappa}$$

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

such that $\sup_{\theta \in \hat{\Theta}_2} |(S.41)| = o_p^*(1)$, in probability.

It remains to analyze (S.39), for which we find

$$E^* Q_T^*(\theta) - \hat{Q}_T(\theta) = T^{-1} \sum_{t=1}^{T} \frac{1}{\hat{\sigma}_t^2} \sum_{n=0}^{t-1} (\hat{\phi}_n(\theta)^2 \hat{\varepsilon}_{t-n}^2 - \phi_n(\theta)^2 \varepsilon_{t-n}^2) \tag{S.43}$$

$$- 2T^{-1} \sum_{t=1}^{T} \frac{1}{\hat{\sigma}_t^2} \sum_{n=0}^{\infty} \sum_{m=n+1}^{\infty} \phi_n(\theta) \phi_m(\theta) \varepsilon_{t-n} \varepsilon_{t-m}. \tag{S.44}$$

By identical arguments to those in the proof of (S.6), the term (S.44) is $o_p(1)$, uniformly in $\theta \in \Theta_2^0$

and $P(\Theta_2^0 \supseteq \hat{\Theta}_2) \to 1$ by (S.31). We therefore proceed with (S.43), which is

$$(S.43) = T^{-1} \sum_{t=1}^{T} \frac{1}{\hat{\sigma}_t^2} \sum_{n=0}^{t-1} \phi_n(\theta)^2 (\hat{\varepsilon}_{t-n}^2 - \varepsilon_{t-n}^2) \tag{S.45}$$

$$+ T^{-1} \sum_{t=1}^{T} \frac{1}{\hat{\sigma}_t^2} \sum_{n=0}^{t-1} (\hat{\phi}_n(\theta)^2 - \phi_n(\theta)^2) \hat{\varepsilon}_{t-n}^2. \tag{S.46}$$

For (S.45) we apply the Cauchy-Schwarz inequality and find

$$(S.45)^2 \leq \left(T^{-1} \sum_{t=1}^{T} \left(\frac{1}{\hat{\sigma}_t^2} \sum_{n=0}^{t-1} \phi_n(\theta)^2 \right)^2 \right) \left(T^{-1} \sum_{t=1}^{T} (\hat{\varepsilon}_{t-n}^2 - \varepsilon_{t-n}^2)^2 \right),$$

where the term in the second parenthesis is $o_p(1)$ by Lemma S.9. From Lemmas S.1(g) and S.4, the term in the first parenthesis is bounded, uniformly in $\theta \in \Theta_2^0$, by $O_p(1)T^{-1} \sum_{t=1}^{T} (\sum_{n=1}^{t-1} n^{-1-\kappa})^2 = O_p(1)$. Because $P(\Theta_2^0 \supseteq \hat{\Theta}_2) \to 1$, see (S.31), this bound applies also uniformly in $\theta \in \hat{\Theta}_2$. Finally, for the term (S.46) we note that, by the mean value theorem,

$$|\hat{\phi}_n(\theta) - \phi_n(\theta)| \leq |\hat{d} - d_0| \sum_{j=0}^{n} \sum_{m=0}^{n-j} b_m(\psi) \dot{\pi}_j(\bar{d} - d) a_{n-j-m}(\hat{\psi}) + |\hat{\psi} - \psi_0| \sum_{j=0}^{n} \sum_{m=0}^{n-j} b_m(\psi) \pi_j(\hat{d} - d) \dot{a}_{n-j-m}(\bar{\psi}),$$

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.7,S.8, 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 (S.31). Thus, by reversing the order of the summations such that $|(S.46)| \leq (\min_t \hat{\sigma}_t^2)^{-1} \sum_{n=0}^{T-1} |\hat{\phi}_n(\theta) + \phi_n(\theta)||\hat{\phi}_n(\theta) - \phi_n(\theta)|T^{-1} \sum_{t=n+1}^{T} \hat{\varepsilon}_{t-n}^2$, we find that

$$\sup_{\theta \in \hat{\Theta}_2} |(S.46)| \leq (\min_t \hat{\sigma}_t^2)^{-1} \sum_{n=1}^{T-1} O_p(n^{-1/2-\kappa}) O_p(T^{-1/2} n^{-1/2-\kappa}(\log n)) T^{-1} \sum_{t=n+1}^{T} \hat{\varepsilon}_{t-n}^2 = O_p(T^{-1/2}(\log T))$$

because $T^{-1} \sum_{t=n+1}^{T} \hat{\varepsilon}_{t-n}^2 \leq T^{-1} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 = O_p(1)$.

S.2.3.2   Proof of asymptotic normality Following roughly the same steps as in the proof of asymptotic normality in Theorem 1, see Section S.2.1.2, we first prove the asymptotic first-order

14

equivalence of $\hat{\theta}^*$ and $\check{\theta}^*$ by showing

$$\sup_{\theta \in \mathcal{N}(\hat{\theta})} \left| \frac{\partial^2 \hat{Q}_T^*(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \check{Q}_T^*(\theta)}{\partial \theta \partial \theta'} \right| \xrightarrow{p^*}_p 0, \tag{S.47}$$

$$\sqrt{T} \frac{\partial \hat{Q}_T^*(\hat{\theta})}{\partial \theta} - \sqrt{T} \frac{\partial \check{Q}_T^*(\hat{\theta})}{\partial \theta} \xrightarrow{p^*}_p 0. \tag{S.48}$$

The proof of (S.47) is identical to that of (S.34) recognizing that the derivatives add at most a logarithmic factor, see Lemma S.8. The proof of (S.48) is identical to that of (S.13)–(S.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 S.2 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

$$\sup_{\theta \in \mathcal{N}(\hat{\theta})} \left| \frac{\partial^2 \check{Q}_T^*(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \bar{Q}_T^{*0}(\theta)}{\partial \theta \partial \theta'} \right| \xrightarrow{p^*}_p 0 \text{ and } \sup_{\theta \in \mathcal{N}(\hat{\theta})} \left| \frac{\partial^2 \bar{Q}_T^{*0}(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \bar{Q}_T^*(\theta)}{\partial \theta \partial \theta'} \right| \xrightarrow{p^*}_p 0 \tag{S.49}$$

$$\sqrt{T} \frac{\partial \check{Q}_T^*(\hat{\theta})}{\partial \theta} - \sqrt{T} \frac{\partial \bar{Q}_T^{*0}(\hat{\theta})}{\partial \theta} \xrightarrow{p^*}_p 0 \text{ and } \sqrt{T} \frac{\partial \bar{Q}_T^{*0}(\hat{\theta})}{\partial \theta} - \sqrt{T} \frac{\partial \bar{Q}_T^*(\hat{\theta})}{\partial \theta} \xrightarrow{p^*}_p 0, \tag{S.50}$$

from which the result follows by the triangle inequality. The proofs of (S.49) and (S.50) follow nearly identically to those of (S.3) and (S.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.11 and S.12 instead of Lemmas S.4 and S.5 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 $\sqrt{T} \partial \bar{Q}_T^*(\hat{\theta}) / \partial \theta \xrightarrow{w}_p N(0, 4A_0^\dagger)$.

S.2.3.3 Proof of consistency of bootstrap variance estimator We define

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

15

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 (S.47) 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 (S.49) 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.12. 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 S.2(b*),(c*) and Assumption 7, and the last parenthesis on the right-hand side is $O_p^*(1)$, in probability, by Lemmas S.1(f),(g) and S.2(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 S.2.2 and is therefore omitted. The required result now follows by Slutsky's Theorem.

S.3 Lemmas

S.3.1 Lemma for Non-Bootstrap Asymptotics

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.

Lemma S.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 $(\min_{1\le t\le T}\bar{\sigma}_t^2)^{-1} = O_p(1)$.

Proof. Part (a): We have

$$T^{-1}\sum_{t=1}^T (\bar{\sigma}_t^2 - \sigma_t^2)^2 = T^{-1}\sum_{t=1}^T \sum_{i,j=1}^T k_{ti}k_{tj}(\sigma_i^2 - \sigma_t^2)(\sigma_j^2 - \sigma_t^2) \le 2G\sum_{i=1}^T k_{ti}T^{-1}\sum_{t=1}^T |\sigma_i^2 - \sigma_t^2|$$

$$= 2GT^{-1}\sum_{t=1}^T \sum_{i=1}^{t-\lfloor MTb\rfloor} k_{ti}|\sigma_i^2 - \sigma_t^2| + 2GT^{-1}\sum_{t=1}^T \sum_{i=t-\lfloor MTb\rfloor+1}^{t+\lfloor MTb\rfloor} k_{ti}|\sigma_i^2 - \sigma_t^2|$$

$$+ 2GT^{-1}\sum_{t=1}^T \sum_{i=t+\lfloor MTb\rfloor+1}^{T} k_{ti}|\sigma_i^2 - \sigma_t^2| =: A_{1T} + A_{2T} + A_{3T}$$

for $G := \sup_{0\le u\le 1}\sigma(u)^2 < \infty$ 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

$$\frac{1}{Tb}\sum_{i=1}^{\lfloor Tx\rfloor - \lfloor MTb\rfloor} K\left(\frac{\lfloor Tx\rfloor - i}{Tb}\right) = \frac{1}{b}\int_{1/T}^{(\lfloor Tx\rfloor-\lfloor MTb\rfloor)/T} K\left(\frac{\lfloor Tx\rfloor - \lfloor Ts\rfloor}{Tb}\right)\mathsf{d}s$$

$$\overset{u:=(s-x)/b}{=} \int_{\frac{1/T-x}{b}}^{\frac{\lfloor Tx\rfloor-\lfloor MTb\rfloor-Tx}{Tb}} K(u)\mathsf{d}u + o(1)$$

$$\le \int_{\frac{1/T-1}{b}}^{-\frac{\lfloor MTb\rfloor}{Tb}} K(u)\mathsf{d}u + o(1) \to \int_{-\infty}^{-M} K(u)\mathsf{d}u$$

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, $A_{1T} \le c\sup_{x\in[0,1]}\frac{1}{Tb}\sum_{i=1}^{\lfloor Tx\rfloor-\lfloor MTb\rfloor} K(\frac{\lfloor Tx\rfloor-i}{Tb}) \le \epsilon$ for M

17

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.

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

$$T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\sigma_i^2\sigma_j^2 E(z_i^2 - 1)(z_j^2 - 1)$$

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

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

Part (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 $R_T(\theta_0)$,

$$T^{-1}\sum_{t=1}^{T}(\hat{\sigma}_t^2 - \tilde{\sigma}_t^2)^2 = R_T(\tilde{\theta}) = R_T(\theta_0) + \frac{\partial R_T(\theta_0)}{\partial\theta}(\tilde{\theta} - \theta_0) + (\tilde{\theta} - \theta_0)'\frac{\partial^2 R_T(\theta_0)}{\partial\theta\partial\theta'}(\tilde{\theta} - \theta_0)$$

$$+ \sum_{k,m,n=1}^{p+1} (\tilde{\theta}_k - \theta_{0k})(\tilde{\theta}_m - \theta_{0m})(\tilde{\theta}_n - \theta_{0n})\frac{\partial^3 R_T(\dot{\theta})}{\partial\theta_k\partial\theta_m\partial\theta_n}, \qquad (S.51)$$

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.4. By repeated application of the Cauchy-Schwarz inequality and using that $E(\varepsilon_i r_i^3) = 0$,

$$\left(E|(r_i^2 + 2\varepsilon_i r_i)(r_j^2 + 2\varepsilon_j r_j)|\right)^2 \leq E(r_i^2 + 2\varepsilon_i r_i)^2 E(r_j^2 + 2\varepsilon_j r_j)^2$$

$$= E(r_i^4 + 4\varepsilon_i^2 r_i^2)E(r_j^4 + 4\varepsilon_j^2 r_j^2)$$

18

$$\leq \left(Er_i^4 + 4(E\varepsilon_i^4)^{1/2}(Er_i^4)^{1/2}\right)\left(Er_j^4 + 4(E\varepsilon_j^4)^{1/2}(Er_j^4)^{1/2}\right),$$

so that, by Lemma S.4 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 (S.20).

For the second term on the right-hand side of (S.51), we find in the same way that

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

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

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 (S.20), so that the contribution to the right-hand side of (S.51) is $O_p(T^{-3/2}b^{-1})$.

To prove the result for the third term on the right-hand side of (S.51) we first find

$$\frac{\partial^2 R_T(\theta_0)}{\partial \theta_m \partial \theta_n} = 8T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T}k_{ti}k_{tj}(\varepsilon_i + r_i)\frac{\partial \varepsilon_i(\theta_0)}{\partial \theta_m}(\varepsilon_j + r_j)\frac{\partial \varepsilon_j(\theta_0)}{\partial \theta_n}$$

$$+ 4T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T}k_{ti}k_{tj}(r_i^2 + 2\varepsilon_i r_i)\frac{\varepsilon_j(\theta_0)}{\partial \theta_m}\frac{\partial \varepsilon_j(\theta_0)}{\partial \theta_n}$$

$$+ 4T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T}k_{ti}k_{tj}(r_i^2 + 2\varepsilon_i r_i)\varepsilon_j(\theta_0)\frac{\partial^2 \varepsilon_j(\theta_0)}{\partial \theta_m \partial \theta_n}$$

$$=: 8B_{1T} + 4B_{2T} + 4B_{3T}.$$

19

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

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

so the contribution to (S.51) is $O_p(T^{-2}b^{-1})$. Next, we find that

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

where the term inside the first large square-root is

$$T^{-1} \sum_{t=1}^{T} \left(\sum_{i=1}^{T} k_{ti}\varepsilon_i \frac{\partial \varepsilon_i(\theta_0)}{\partial \theta_m} \right)^2 + T^{-1} \sum_{t=1}^{T} \left(\sum_{i=1}^{T} k_{ti} r_i \frac{\partial \varepsilon_i(\theta_0)}{\partial \theta_m} \right)^2$$

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

$$=: B_{11T} + B_{12T} + B_{13T},$$

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

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

using $\sum_{i=1}^{T} k_{ti} = 1$, (S.20), and Assumption 1, see also Lemma S.4. Similarly,

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

It follows that the contributions of B_{11T} and B_{12T}, and hence B_{13T} and B_{1T}, to (S.51) is $O_p(T^{-2}b^{-1})$.

20

Finally, we prove the result for the last term on the right-hand side of (S.51), where we find

$$\frac{\partial^3 R_T(\theta)}{\partial\theta_l\partial\theta_m\partial\theta_n} = 2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}(\varepsilon_i(\theta)^2 - \varepsilon_i^2)\frac{\partial^3(\varepsilon_j(\theta)^2)}{\partial\theta_l\partial\theta_m\partial\theta_n}$$

$$+ 2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\frac{\partial(\varepsilon_i(\theta)^2)}{\partial\theta_l}\frac{\partial^2(\varepsilon_j(\theta)^2)}{\partial\theta_m\partial\theta_n}$$

$$+ 2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\frac{\partial(\varepsilon_i(\theta)^2)}{\partial\theta_m}\frac{\partial^2(\varepsilon_j(\theta)^2)}{\partial\theta_l\partial\theta_n}$$

$$+ 2T^{-1}\sum_{t=1}^{T}\sum_{i,j=1}^{T} k_{ti}k_{tj}\frac{\partial(\varepsilon_i(\theta)^2)}{\partial\theta_n}\frac{\partial^2(\varepsilon_j(\theta)^2)}{\partial\theta_l\partial\theta_m}$$

$$=: 2C_{1T}(\theta) + 2C_{2T}(\theta) + 2C_{3T}(\theta) + 2C_{4T}(\theta).$$

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 (S.20) and Lemma S.5. 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.51) is $O_p(T^{-3/2}b^{-1})$.

Part (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.

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

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

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 (S.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 $O((Tb)^{-2})$.

Part (f): We apply the inequality $\hat{\sigma}_t^2 \leq \sigma_t^2 + |\hat{\sigma}_t^2 - \tilde{\sigma}_t^2| + |\tilde{\sigma}_t^2 - \bar{\sigma}_t^2| + |\bar{\sigma}_t^2 - \sigma_t^2|$, and note that $\max_{1 \leq t \leq T} \sigma_t^2$ 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 $M_T(\theta) := \max_{1 \leq t \leq T} |\sigma_t^2(\theta) - \tilde{\sigma}_t^2|$ and apply a mean-value expansion around $M_T(\theta_0)$,

$$\max_{1 \leq t \leq T} |\hat{\sigma}_t^2 - \tilde{\sigma}_t^2| = M_T(\tilde{\theta}) = M_T(\theta_0) + 2(\tilde{\theta} - \theta_0)\frac{\partial M_T(\dot{\theta})}{\partial \theta}$$

$$\leq M_T(\theta_0) + 2 \sum_{m=1}^{p+1} |\tilde{\theta}_m - \theta_{0,m}| \max_{1 \leq t \leq T} \left| \sum_{i=1}^{T} k_{ti}\varepsilon_i(\dot{\theta})\frac{\partial \varepsilon_i(\dot{\theta})}{\partial \theta_m} \right|, \qquad \text{(S.52)}$$

for an intermediate value, $\dot{\theta}$. The first term on the right-hand side of (S.52) is a non-negative random variable with expectation

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

$$\leq c(\sup_{1 \leq i \leq T, 1 \leq t \leq T} |k_{ti}|) \sum_{i=1}^{T} i^{-1-\zeta} \leq c\frac{1}{Tb},$$

where the first two inequalities are due to Lemma S.4 and Assumption 1 and the last inequality is due to (S.20). For second term on the right-hand side of (S.52) we apply the Cauchy-Schwarz inequality and find the bound

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

22

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

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

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

Finally,

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

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

Part (g): We apply the inequality

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

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 $t = 1, \ldots, T$, we find that

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

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

S.3.2 Lemma for Bootstrap Asymptotics

In this section we present the bootstrap equivalent of Lemma S.1.

Lemma S.2. *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^*) $\max_{1 \le t \le T} \hat\sigma_t^{*2} = O_p^*(1)$,

(g^*) $(\min_{1 \le t \le T} \hat\sigma_t^{*2})^{-1} = O_p^*(1)$ *and* $(\min_{1 \le t \le T} \tilde\sigma_t^{*2})^{-1} = O_p^*(1)$.

Proof. *Part (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 S.1(b) since $T^{-1} \sum_{t=1}^{T} \hat\varepsilon_t^4 = O_p(1)$.

Part (c^*): This follows as in the proof of Lemma S.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.

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

$$T^{-1} \sum_{t=1}^{T} \sum_{i,j,m,n=1}^{T} k_{ti} k_{tj} k_{tm} k_{tn} \hat\varepsilon_i^2 \hat\varepsilon_j^2 \hat\varepsilon_m^2 \hat\varepsilon_n^2 E^*(w_i^2 - 1)(w_j^2 - 1)(w_m^2 - 1)(w_n^2 - 1)$$

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

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

Part (f^*): We apply the inequality $\hat\sigma_t^{*2} \le \hat\sigma_t^2 + |\hat\sigma_t^{*2} - \tilde\sigma_t^{*2}| + |\tilde\sigma_t^{*2} - \hat\sigma_t^2|$, and note that $\max_{1 \le t \le T} \hat\sigma_t^2 = O_p(1)$ by Lemma S.1(f). The proof for the second term is exactly the same as for the corresponding term in the proof of Lemma S.1(f), but using that $(\tilde\theta^* - \hat\theta) = O_p^*(T^{-1/2})$, in probability (by Theorem 6 of CNT) and Lemma S.12.

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

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

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

which by the Cauchy-Schwarz inequality is bounded by

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

24

using (S.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 $\max_{1 \leq t \leq T} |\hat{\sigma}_t^{*2} - \hat{\sigma}_t^2|$, and hence $\max_{1 \leq t \leq T} \hat{\sigma}_t^{*2}$, is $O_p^*(1)$, in probability.

Part (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 S.1(g). $\qquad \square$

S.3.3 Additional Technical Lemmas

In Lemma S.4 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.3. *Let the sequences a_j and $b_{j,T}, j = 1, \ldots, T$, 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 $T \to \infty$, $\sup_{j \leq q_T, T} |b_{j,T}| \to 0$. Then $\sum_{j=1}^{T} a_j b_{j,T} \to 0$.*

Proof. By the triangle inequality,

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

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

Lemma S.4. *Let Assumptions 1–3 be satisfied. For $k = 0, 1$ define $\Psi_k := \Psi$ and for $k = 2, 3$ define $\Psi_k := \mathcal{N}_\delta(\psi_0) := \{\psi \in \Psi : ||\psi - \psi_0|| \leq \delta\}$ 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 $1 \leq h \leq 8$, for any finite constant g, and for $k = 0, 1, 2, 3$,

$$\sup_{d_0 - d \leq g, \psi \in \Psi_k} |\partial^k \phi_j(\theta) / \partial \theta^{(k)}| \leq c(\log j)^k j^{\max(g-1, -2-\zeta)},$$

$$E \sup_{d_0 - d \leq g, \psi \in \Psi_k} |\partial^k r_t(\theta) / \partial \theta^{(k)}|^h \leq c(\log t)^{hk} t^{h \max\{g-1, -1-\zeta\}},$$

$$E|r_t|^h \leq ct^{-h(1+\zeta)}.$$

Proof. The residual is given in (3) as

$$\varepsilon_t(\theta) = \sum_{n=0}^{t-1} \sum_{j=0}^{t-1-n} \sum_{m=0}^{\infty} b_n(\psi) \pi_j(d_0 - d) a_m(\psi_0) \varepsilon_{t-n-j-m}$$

$$= \sum_{j=0}^{t-1} \sum_{n=0}^{t-1-j} \sum_{m=0}^{t-n-j-1} b_n(\psi) \pi_j(d_0 - d) a_m(\psi_0) \varepsilon_{t-n-j-m}$$

$$+ \sum_{n=0}^{t-1} \sum_{j=0}^{t-1-n} \sum_{m=t-n-j}^{\infty} b_n(\psi) \pi_j(d_0 - d) a_m(\psi_0) \varepsilon_{t-n-j-m}$$

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

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

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

$$\leq c(\log j)^k \sum_{n=1}^{j-1} n^{g-1} (j - n)^{-2-\zeta} \leq c(\log j)^k j^{\max(g-1, -2-\zeta)}$$

by (2) and Lemmas S.7 and S.8. 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),

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

$$\leq c \left((\log t)^k \sum_{n=1}^{t-1} n^{-2-\zeta} \sum_{j=1}^{t-1-n} j^{g-1} (t - n - j)^{-1-\zeta} \right)^h$$

$$\leq c \left((\log t)^k \sum_{n=1}^{t-1} n^{-2-\zeta}(t-n)^{\max\{g-1,-1-\zeta\}} \right)^h \leq c(\log t)^{hk} t^{h\max\{g-1,-1-\zeta\}}.$$

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,

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

$\square$

Lemma S.5. *Let Assumptions 1–3 be satisfied. For $k = 0,1$ define $\Psi_k := \Psi$ and for $k = 2,3$ define $\Psi_k := \mathcal{N}_\delta(\psi_0) := \{\psi \in \Psi : ||\psi - \psi_0|| \leq \delta\}$ 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 $\kappa > 0$,*

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

Proof. 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.4 and Minkowski's inequality we find

$$\left(T^{-1}\sum_{t=1}^{T} \left| \frac{\partial^{k(i)}\varepsilon_t(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}} \right|^q \right)^{1/q} \leq \left(T^{-1}\sum_{t=1}^{T} \left| \sum_{j=0}^{t-1} \frac{\partial^{k(i)}\phi_j(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}}\varepsilon_{t-j} \right|^q \right)^{1/q} + \left(T^{-1}\sum_{t=1}^{T} \left| \frac{\partial^{k(i)}r_t(\theta)}{\partial\theta_{r_1}\ldots\partial\theta_{r_{k(i)}}} \right|^q \right)^{1/q}.$$
(S.53)

We note from Lemma S.4 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

27

from Lemma S.4 that the second term on the right-hand side of (S.53) satisfies

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

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

Next, we give the proof for the first term on the right-hand side of (S.53). 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

$$T^{-1} \sum_{t=1}^{T} \left| \sum_{j=0}^{t-1} \phi_j(\theta) \varepsilon_{t-j} \right|^q = T^{-1} \sum_{t=1}^{T} \left| \phi_{t-1}(\theta) \sum_{j=0}^{t-1} \varepsilon_{t-j} \right|^q \tag{S.54}$$

$$+ T^{-1} \sum_{t=1}^{T} \left| \sum_{j=0}^{t-2} (\phi_j(\theta) - \phi_{j+1}(\theta)) \sum_{l=0}^{j} \varepsilon_{t-l} \right|^q \tag{S.55}$$

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

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 $\sup_{d_0 - d \le 1/2 - \kappa, \psi \in \Psi} |\phi_{t-1}(\theta)| \le c t^{-1/2-\kappa}$ by Lemma S.4, we first find that $\sup_{d_0 - d \le 1/2 - \kappa, \psi \in \Psi} |(S.54)| = O_p(T^{\max\{-\kappa q, -1\}}) = o_p(1)$ because $\kappa > 0, q \ge 2$. To prove the result for (S.55), first note that, with obvious notation, $\phi_{j+1}(d, \psi) - \phi_j(d, \psi) = \phi_{j+1}(d-1, \psi)$, so that $\sup_{d_0 - d \le 1/2 - \kappa, \psi \in \Psi} |\phi_j(\theta) - \phi_{j+1}(\theta)| \le c j^{-3/2-\kappa}$ by Lemma S.4. It then follows that

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

uniformly in t, and therefore

$$\sup_{d_0 - d \le 1/2 - \kappa, \psi \in \Psi} |(S.55)| = O_p(1)$$

which proves the result. $\square$

Lemma S.6 (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 $\{z_s\}_{s \le t}$, and suppose $E|z_t|^q < \infty$ for some integer*

28

$q \geq 2$. 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.7 (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.8 (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*

$$|\frac{\partial^m}{\partial u^m} \pi_j(u)| \leq c(u_0)(1 + \log j)^m j^{u-1}, \tag{S.56}$$

$$|\frac{\partial^m}{\partial u^m} T^{-u} \pi_j(u)| \leq c(u_0)T^{-u}(1 + |\log(j/T)|)^m j^{u-1}. \tag{S.57}$$

Lemma S.9 (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 S.10. *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 $1 \leq m \leq k \leq T$,*

$$E^* \sup_\theta \left| \sum_{j=m}^{k} \lambda_j(\theta) \sum_{t=\max(j,h)+1}^{T} \frac{1}{\hat{\sigma}_t} \varepsilon_{t-j}^* \varepsilon_{t-h}^* \right|$$

$$= \mathbb{I}(g > -1/2)O_p(T^{1/2}k^{1/2+g}) + \mathbb{I}(g < -1/2)O_p(T^{1/2}m^{1/2+g}) + \mathbb{I}(g = -1/2)O_p(T^{1/2}(\log k)).$$

Proof. The proof follows by Lemma S.1(g) and Lemma D.2 of CNT. $\qquad \square$

Lemma S.11. *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.4, with the main difference being that $\varepsilon_t^* = 0$ for $t \leq 0$, and is omitted for brevity. □

Lemma S.12. *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.5 and is omitted for brevity. □

S.4   Additional Simulation Results

In this section we present additional simulation results in Tables S.1 and S.2. Specifically, we report bias and root-mean-squared-error [RMSE] for the QML and ACSS estimators of d using the same data generating mechanisms as in Tables 1 and 2. We compare with the local Whittle estimator of Künsch (1987) and Robinson (1995), noting that Shao and Wu (2007) have shown that this estimator remains valid under conditions that allow some (conditional) heteroskedasticity. For the local Whittle estimator we applied two different bandwidth parameters, $m = \lfloor T^{0.5} \rfloor$ and $m = \lfloor T^{0.8} \rfloor$, where $\lfloor \cdot \rfloor$ denotes the integer part of the argument.

The results in Tables S.1 and S.2 are very much in line with the earlier simulation results and the theoretical findings in the paper. Specifically, the bias of the QML and ACSS estimators are very similar, but the ACSS has much smaller RMSE in the presence of unconditional heteroskedasticity. Compared with the local Whittle estimator, both the QML and ACSS estimators have much

Table S.1: Simulation results for estimators of d with uncorrelated errors

			Bias				RMSE			
τ	υ	T	QML	ACSS	LW(0.5)	LW(0.8)	QML	ACSS	LW(0.5)	LW(0.8)
Panel A: IID errors										
	1	100	-0.01164	-0.01140	-0.03155	-0.00936	0.08406	0.08470	0.25457	0.09682
	1	250	-0.00573	-0.00573	-0.01895	-0.00465	0.05037	0.05062	0.18869	0.06148
	1	500	-0.00242	-0.00245	-0.01824	-0.00261	0.03586	0.03591	0.14354	0.04530
1/4	1/3	100	-0.01645	-0.01229	-0.03532	-0.01239	0.11704	0.09025	0.30609	0.13406
1/4	1/3	250	-0.00915	-0.00587	-0.02597	-0.00865	0.07458	0.05326	0.23314	0.08974
1/4	1/3	500	-0.00428	-0.00241	-0.01833	-0.00424	0.05309	0.03630	0.19056	0.06650
1/4	3	100	-0.01242	-0.01102	-0.02894	-0.00938	0.09484	0.09318	0.26651	0.10563
1/4	3	250	-0.00590	-0.00497	-0.02114	-0.00579	0.05820	0.05501	0.20030	0.06967
1/4	3	500	-0.00256	-0.00213	-0.01677	-0.00341	0.04016	0.03753	0.15171	0.05053
3/4	1/3	100	-0.01307	-0.01236	-0.03350	-0.01112	0.09298	0.09038	0.26470	0.10786
3/4	1/3	250	-0.00467	-0.00443	-0.02172	-0.00449	0.05612	0.05300	0.19864	0.06878
3/4	1/3	500	-0.00320	-0.00255	-0.01643	-0.00400	0.04004	0.03701	0.15261	0.05045
3/4	3	100	-0.01822	-0.01135	-0.03860	-0.01425	0.12510	0.09536	0.31257	0.13547
3/4	3	250	-0.01027	-0.00596	-0.02887	-0.00987	0.07748	0.05489	0.23916	0.09038
3/4	3	500	-0.00542	-0.00334	-0.01972	-0.00510	0.05358	0.03692	0.18848	0.06482
Panel B: GARCH errors										
	1	100	-0.01506	-0.01359	-0.03496	-0.01165	0.10288	0.09215	0.27456	0.11557
	1	250	-0.00739	-0.00587	-0.02289	-0.00686	0.07130	0.05639	0.21553	0.08375
	1	500	-0.00501	-0.00338	-0.02162	-0.00504	0.05610	0.03862	0.17590	0.06851
1/4	1/3	100	-0.01953	-0.01612	-0.03502	-0.01545	0.12695	0.09491	0.31759	0.14376
1/4	1/3	250	-0.00940	-0.00612	-0.02709	-0.00994	0.08757	0.05641	0.25048	0.10315
1/4	1/3	500	-0.00539	-0.00289	-0.02644	-0.00625	0.06787	0.03904	0.20610	0.08273
1/4	3	100	-0.01736	-0.01351	-0.03776	-0.01474	0.11313	0.09920	0.28199	0.12410
1/4	3	250	-0.00772	-0.00502	-0.02648	-0.00743	0.07629	0.05782	0.22296	0.08901
1/4	3	500	-0.00522	-0.00270	-0.01805	-0.00502	0.05998	0.03968	0.18531	0.07212
3/4	1/3	100	-0.01506	-0.01338	-0.03824	-0.01205	0.10681	0.09287	0.28677	0.12257
3/4	1/3	250	-0.00893	-0.00750	-0.02847	-0.00860	0.07474	0.05573	0.22674	0.08833
3/4	1/3	500	-0.00558	-0.00391	-0.02263	-0.00581	0.05896	0.03856	0.18569	0.07351
3/4	3	100	-0.01902	-0.01253	-0.03570	-0.01511	0.13877	0.10094	0.31344	0.14682
3/4	3	250	-0.01029	-0.00609	-0.02805	-0.01034	0.09131	0.05829	0.24858	0.10411
3/4	3	500	-0.00688	-0.00330	-0.02747	-0.00718	0.06923	0.03968	0.20729	0.08305
Panel C: SV errors										
	1	100	-0.02242	-0.01508	-0.05170	-0.02062	0.16380	0.12465	0.31455	0.17409
	1	250	-0.01433	-0.00886	-0.03387	-0.01507	0.12561	0.07046	0.24168	0.13876
	1	500	-0.00892	-0.00517	-0.03006	-0.01142	0.10556	0.04732	0.18746	0.11886
1/4	1/3	100	-0.02451	-0.01429	-0.04526	-0.02358	0.16689	0.12392	0.32216	0.18138
1/4	1/3	250	-0.01517	-0.00789	-0.03647	-0.01626	0.13348	0.07384	0.25710	0.14840
1/4	1/3	500	-0.00923	-0.00587	-0.02823	-0.01289	0.11518	0.04748	0.20106	0.13016
1/4	3	100	-0.02209	-0.01486	-0.04440	-0.02010	0.16712	0.12614	0.31287	0.17433
1/4	3	250	-0.01343	-0.00844	-0.03907	-0.01519	0.13049	0.07388	0.24301	0.14310
1/4	3	500	-0.00976	-0.00535	-0.02654	-0.01178	0.11092	0.04936	0.19557	0.12171
3/4	1/3	100	-0.02050	-0.01394	-0.04319	-0.02012	0.15918	0.11504	0.31292	0.17612
3/4	1/3	250	-0.01623	-0.00910	-0.03748	-0.01744	0.12666	0.06894	0.24795	0.14260
3/4	1/3	500	-0.01011	-0.00667	-0.02590	-0.01273	0.10828	0.04723	0.19282	0.12157
3/4	3	100	-0.02447	-0.01553	-0.04531	-0.02094	0.17972	0.13460	0.32244	0.18170
3/4	3	250	-0.01652	-0.00943	-0.03653	-0.02007	0.13952	0.07644	0.25366	0.14835
3/4	3	500	-0.01092	-0.00541	-0.03083	-0.01296	0.11547	0.05118	0.20484	0.12863

Notes: The table reports simulated bias and root-mean-squared error (RMSE) for d based on 10,000 replications. Results are presented for the QML and ACSS estimators along with the local Whittle estimator with bandwidths $m = \lfloor T^{0.5} \rfloor$ and $m = \lfloor T^{0.8} \rfloor$ denoted LW(0.5) and LW(0.8), respectively.

Table S.2: Simulation results for estimators of d with AR or MA errors

			Bias				RMSE			
a_1	a_2	T	QML	ACSS	LW(0.5)	LW(0.8)	QML	ACSS	LW(0.5)	LW(0.8)
Panel A: No break, $\upsilon = 1$										
0.0		100	-0.08056	-0.08298	-0.03176	-0.00941	0.25618	0.26304	0.25505	0.09696
0.0		250	-0.02503	-0.02466	-0.01916	-0.00467	0.12413	0.12361	0.18829	0.06147
0.0		500	-0.00966	-0.00915	-0.01823	-0.00263	0.06927	0.06477	0.14359	0.04533
-0.8		100	-0.01756	-0.01726	-0.04820	-0.28008	0.09196	0.09288	0.25930	0.29778
-0.8		250	-0.00766	-0.00767	-0.02678	-0.16231	0.05510	0.05529	0.18417	0.17411
-0.8		500	-0.00371	-0.00369	-0.01709	-0.11281	0.03812	0.03817	0.14295	0.12178
0.8		100	0.01447	0.01479	0.44368	0.70461	0.19112	0.19227	0.51343	0.71204
0.8		250	0.01291	0.01324	0.24663	0.67734	0.14538	0.14597	0.31070	0.68106
0.8		500	0.01176	0.01166	0.14795	0.65182	0.11182	0.11182	0.20515	0.65400
	0.0	100	-0.01162	-0.01058	-0.03315	-0.01024	0.15931	0.16283	0.25188	0.09846
	0.0	250	-0.00680	-0.00674	-0.02135	-0.00440	0.08650	0.08700	0.18625	0.06266
	0.0	500	-0.00447	-0.00443	-0.01408	-0.00327	0.05973	0.05925	0.14075	0.04540
	-0.8	100	-0.04122	-0.04312	-0.47124	-0.66670	0.23158	0.23430	0.53863	0.67525
	-0.8	250	-0.02750	-0.02722	-0.27214	-0.59559	0.17451	0.17413	0.33086	0.59924
	-0.8	500	-0.02181	-0.02220	-0.17512	-0.53891	0.13082	0.13148	0.22545	0.54196
	0.8	100	-0.01345	-0.01332	-0.01083	0.25670	0.09127	0.09196	0.25321	0.27636
	0.8	250	-0.00417	-0.00413	-0.01424	0.15680	0.05415	0.05434	0.18591	0.16933
	0.8	500	-0.00227	-0.00223	-0.01265	0.11126	0.03761	0.03769	0.14197	0.12032
Panel B: Early break, $\tau = 1/4, \upsilon = 1/3$										
0.0		100	-0.09116	-0.07896	-0.04071	-0.01521	0.27065	0.25123	0.30619	0.13767
0.0		250	-0.04236	-0.02597	-0.02450	-0.00898	0.16736	0.12488	0.23854	0.08994
0.0		500	-0.01773	-0.00980	-0.02286	-0.00489	0.10022	0.06650	0.18896	0.06691
-0.8		100	-0.02999	-0.01966	-0.05563	-0.28545	0.12907	0.09963	0.31038	0.31479
-0.8		250	-0.01456	-0.00856	-0.03293	-0.16702	0.08061	0.05741	0.23376	0.18907
-0.8		500	-0.00690	-0.00388	-0.02141	-0.11516	0.05700	0.03903	0.19077	0.13271
0.8		100	0.01385	0.00346	0.44363	0.70502	0.21957	0.18593	0.54490	0.71983
0.8		250	0.01488	0.00578	0.24630	0.67750	0.17590	0.13980	0.34508	0.68529
0.8		500	0.01423	0.00751	0.14318	0.65205	0.14466	0.10699	0.24191	0.65693
	0.0	100	-0.01889	-0.01576	-0.03885	-0.01427	0.19439	0.16016	0.30434	0.13520
	0.0	250	-0.00837	-0.00840	-0.02534	-0.00697	0.12525	0.08750	0.23657	0.08952
	0.0	500	-0.00783	-0.00499	-0.02061	-0.00604	0.08604	0.05962	0.18763	0.06668
	-0.8	100	-0.03874	-0.04997	-0.49925	-0.68709	0.24128	0.23586	0.58540	0.70007
	-0.8	250	-0.03029	-0.02894	-0.28771	-0.60734	0.19610	0.17322	0.37068	0.61459
	-0.8	500	-0.02404	-0.02254	-0.18390	-0.54781	0.16138	0.12895	0.26237	0.55275
	0.8	100	-0.01728	-0.01399	-0.01808	0.25388	0.12485	0.09751	0.30749	0.29358
	0.8	250	-0.00830	-0.00481	-0.02122	0.15401	0.07904	0.05582	0.23634	0.18049
	0.8	500	-0.00400	-0.00251	-0.01639	0.11069	0.05584	0.03855	0.18894	0.12995
Panel C: Late break, $\tau = 3/4, \upsilon = 3$										
0.0		100	-0.11051	-0.08745	-0.03487	-0.01207	0.31767	0.28075	0.30685	0.13594
0.0		250	-0.04473	-0.02615	-0.02314	-0.00750	0.18080	0.13116	0.23731	0.08933
0.0		500	-0.01987	-0.01032	-0.02235	-0.00504	0.10361	0.07201	0.18841	0.06624
-0.8		100	-0.03126	-0.01829	-0.05149	-0.28304	0.13882	0.10499	0.30972	0.31191
-0.8		250	-0.01473	-0.00803	-0.03366	-0.16637	0.08339	0.05876	0.23863	0.18817
-0.8		500	-0.00812	-0.00381	-0.02502	-0.11607	0.05808	0.03970	0.18856	0.13357
0.8		100	0.01145	0.01243	0.39940	0.68031	0.24352	0.20828	0.50684	0.69605
0.8		250	0.01890	0.01575	0.22896	0.66646	0.18678	0.15529	0.33546	0.67452
0.8		500	0.01727	0.01344	0.13810	0.64476	0.15061	0.11790	0.23744	0.64958
	0.0	100	-0.00191	-0.00250	-0.03823	-0.01450	0.24440	0.18876	0.30648	0.13432
	0.0	250	-0.00675	-0.00619	-0.02799	-0.00897	0.14052	0.09533	0.23817	0.09015
	0.0	500	-0.00695	-0.00424	-0.01686	-0.00532	0.09098	0.06163	0.18549	0.06640
	-0.8	100	-0.04956	-0.04057	-0.41284	-0.62999	0.26265	0.24038	0.51293	0.64516
	-0.8	250	-0.03266	-0.02889	-0.25177	-0.57948	0.20833	0.18268	0.34099	0.58730
	-0.8	500	-0.02516	-0.02274	-0.16815	-0.53259	0.16824	0.13633	0.24855	0.53761
	0.8	100	-0.01600	-0.01377	-0.01515	0.25050	0.13400	0.10298	0.30367	0.29070
	0.8	250	-0.00845	-0.00418	-0.02131	0.15275	0.08255	0.05810	0.24087	0.17979
	0.8	500	-0.00463	-0.00219	-0.01782	0.10912	0.05784	0.03916	0.18793	0.12916

Notes: See notes to Table S.1.

smaller RMSEs in the majority of the cases considered, reflecting the fact that QML and ACSS are fully parametric estimators whereas the local Whittle estimator is semiparametric and hence converges at a slower rate (Robinson, 1995; Shao and Wu, 2007). An exception occurs where the parametric model used in connection with the QML and ACSS estimates is over-fitted; see the rows in Table S.2 relating to either $a_1 = 0$ (such that the AR order is being over-fitted) or $a_2 = 0$ (such that the MA order is being over-fitted). Here we see that the local Whittle estimator with bandwidth $m = \lfloor T^{0.8} \rfloor$ has a lower RMSE than the QML and ACSS estimators, although it is important to notice that the bias displayed by the local Whittle estimator with bandwidth $m = \lfloor T^{0.8} \rfloor$ when either $a_1 \neq 0$ or $a_2 \neq 0$ is much higher than that of the local Whittle estimator with bandwidth $m = \lfloor T^{0.5} \rfloor$ and very much higher than that of the QML and ACSS estimators.

S.5 Data Plots and Heteroskedasticity Diagnostics

We show plots of the data series in Figures S.3–S.4. For detailed descriptions we refer to the main text as well as Andersen *et al.* (2007, Section III.A) for the classic realised variance data and Martins and Amado (2016, Section 3.1) for the government bond yield data.

To investigate the possible presence of heteroskedasticity in the residuals, we report in Table S.3 several tests for conditional and unconditional heteroskedasticity for all the 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.3 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 almost all series the null hypothesis is easily rejected at any conventional significance level.

In the last four columns of Table S.3 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). Most series 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.5–S.11. In the middle panels of Figures S.5–S.11 we plot the sample variance profiles of the residuals, say $\tilde{\varepsilon}_t$, of the fitted ARFIMA

Table S.3: 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: Andersen *et al.* Realized Variance Data Examples						
S&P500	114.093^a	153.603^a	2.004^a	1.989^a	1.724^a	10.155^a
DM/$	147.178^a	180.247^a	1.708^c	1.577^b	0.671^b	3.281^b
T-bond	149.859^a	204.588^a	1.797^c	1.007	0.201	0.997
Panel B: Dow Jones Realized Variance Data Examples						
AAPL	90.632^a	103.711^a	1.976^b	1.915^a	0.808^a	4.023^a
AXP	260.734^a	401.376^a	3.806^a	1.992^a	1.672^a	8.126^a
BA	228.501^a	290.754^a	3.183^a	1.768^a	0.923^a	4.664^a
CAT	265.167^a	388.255^a	3.477^a	1.771^a	1.211^a	6.126^a
CSCO	238.891^a	272.842^a	2.690^a	2.186^a	1.071^a	5.724^a
CVX	367.848^a	478.312^a	2.590^a	1.710^a	0.875^a	4.361^a
DD	247.257^a	295.765^a	3.041^a	1.616^b	0.993^a	5.129^a
DIS	331.309^a	413.546^a	2.869^a	1.760^a	0.878^a	4.505^a
GE	231.169^a	347.691^a	3.529^a	1.861^a	1.169^a	5.692^a
GS	434.517^a	787.135^a	2.739^a	1.476^b	0.892^a	4.445^a
HD	160.401^a	365.690^a	2.985^a	1.861^a	0.998^a	4.817^a
IBM	330.191^a	383.240^a	3.205^a	1.856^a	1.015^a	4.958^a
INTC	355.052^a	436.555^a	3.275^a	2.357^a	1.345^a	6.640^a
JNJ	370.667^a	396.315^a	1.785^b	1.444^b	0.521^b	2.767^b
JPM	647.448^a	802.441^a	3.884^a	2.129^a	1.586^a	7.700^a
KO	443.383^a	578.905^a	2.845^a	1.854^a	0.969^a	4.654^a
MCD	94.682^a	118.466^a	1.633^c	1.589^b	0.723^b	4.190^a
MMM	22.438^a	22.617	1.320	0.726	0.150	0.763
MRK	181.159^a	222.480^a	2.748^a	2.157^a	1.017^a	4.998^a
MSFT	90.319^a	115.216^a	2.158^a	1.435^b	0.570^b	2.741^b
NKE	292.171^a	421.712^a	2.875^a	1.582^b	0.883^a	4.503^a
PFE	68.115^a	73.463^a	1.622^c	1.590^b	0.821^a	4.227^a
PG	11.417^b	11.424	1.082	0.562	0.115	0.552
UNH	389.962^a	687.891^a	2.955^a	1.548^b	0.925^a	4.540^a
UTX	209.449^a	384.358^a	2.819^a	1.474^b	0.792^a	3.877^a
VZ	178.057^a	269.501^a	2.444^a	1.654^a	0.708^b	3.528^b
WMT	556.368^a	613.850^a	2.513^a	1.843^a	0.760^a	3.566^b
XOM	327.195^a	355.369^a	2.286^a	1.481^b	0.619^b	3.093^b
Panel C: Government Bold Yield Data Examples						
Belgium	780.609^a	850.677^a	2.846^a	1.759^a	1.166^a	5.852^a
Finland	618.360^a	873.567^a	2.598^a	1.424^b	0.842^a	4.245^a
France	585.122^a	650.438^a	3.583^a	1.920^a	0.817^a	4.955^a
Germany	688.557^a	773.014^a	3.876^a	2.627^a	2.129^a	11.559^a
Ireland	370.082^a	474.890^a	4.121^a	2.372^a	1.867^a	9.261^a
Italy	324.714^a	486.262^a	3.979^a	3.034^a	2.309^a	10.734^a
Portugal	134.993^a	340.272^a	3.232^a	2.221^a	1.471^a	6.988^a
Spain	167.000^a	245.053^a	3.814^a	2.696^a	2.284^a	10.845^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.

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[1], are reported in the right-hand panels of Figures S.5–S.11.

References

Andersen, T.G., T. Bollerslev and F.X. Diebold (2007), Roughing it up: including jump components in the measurement, modeling, and forecasting of return volatility, *Review of Economics and Statistics* 89, 701–720.

Cavaliere, G., M.Ø. Nielsen and A.M.R. Taylor (2019), Adaptive inference in heteroskedastic fractional time series models, QED working paper 1390, Queen's University.

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.

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

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

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

Künsch, H.R. (1987), Statistical aspects of self-similar processes, *Proceedings of the First World Congress of the Bernoulli Society* 1, 67–74, VNU Science Press, Utrecht.

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

Robinson, P.M. (1995), Gaussian semiparametric estimation of long range dependence, *Annals of Statistics* 23, 1630–1661.

Shao, X. and W.B. Wu (2007), Local Whittle estimation of fractional integration for nonlinear

---

[1]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.

processes, *Econometric Theory* 23, 899–929.

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

Figure S.1: Andersen *et al.* (2007) Realised Variance Data
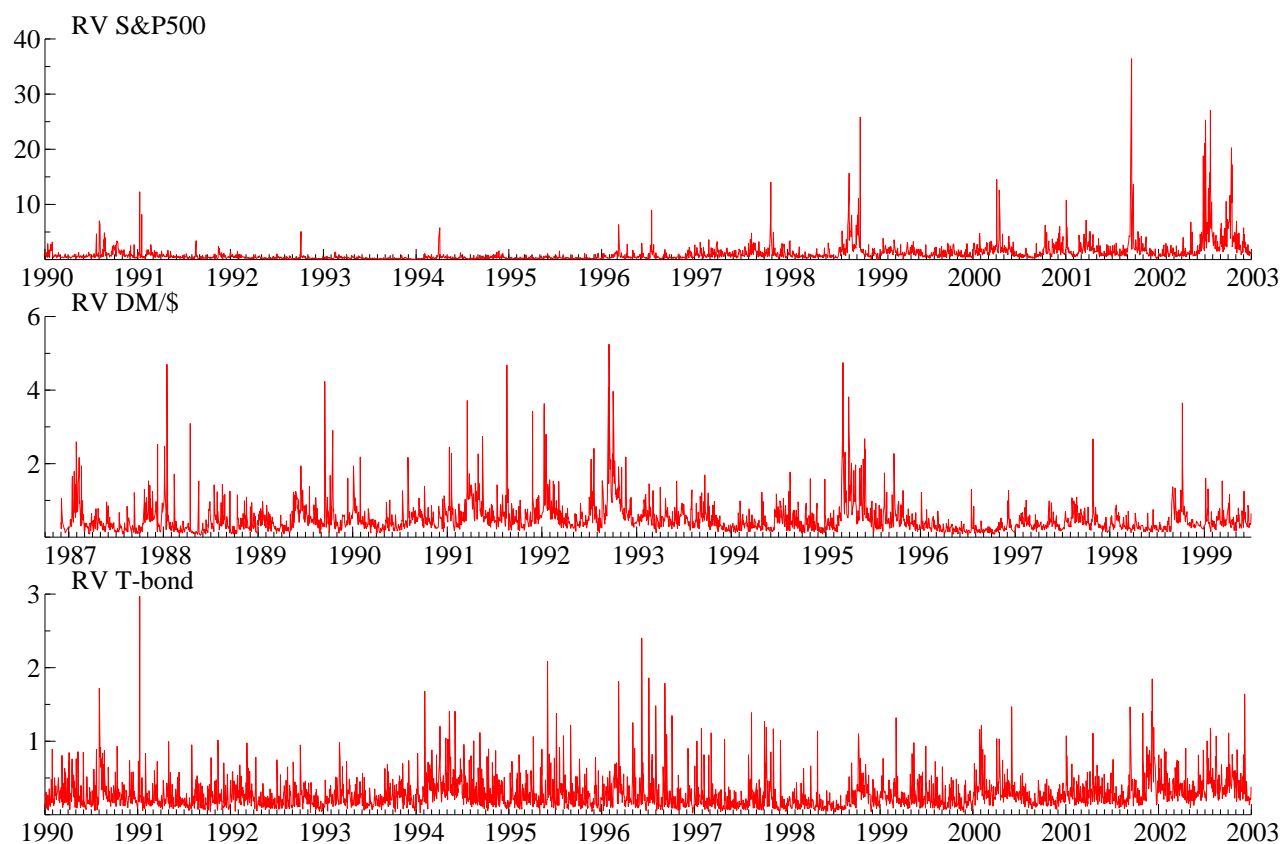
Note: Daily data from January 2, 1990 to December 31, 2002 (SP500 and Tbond) and December 2, 1986 to June 30, 1999 (DEM/USD).

Figure S.2: Dow Jones Realised Variance Data, part 1
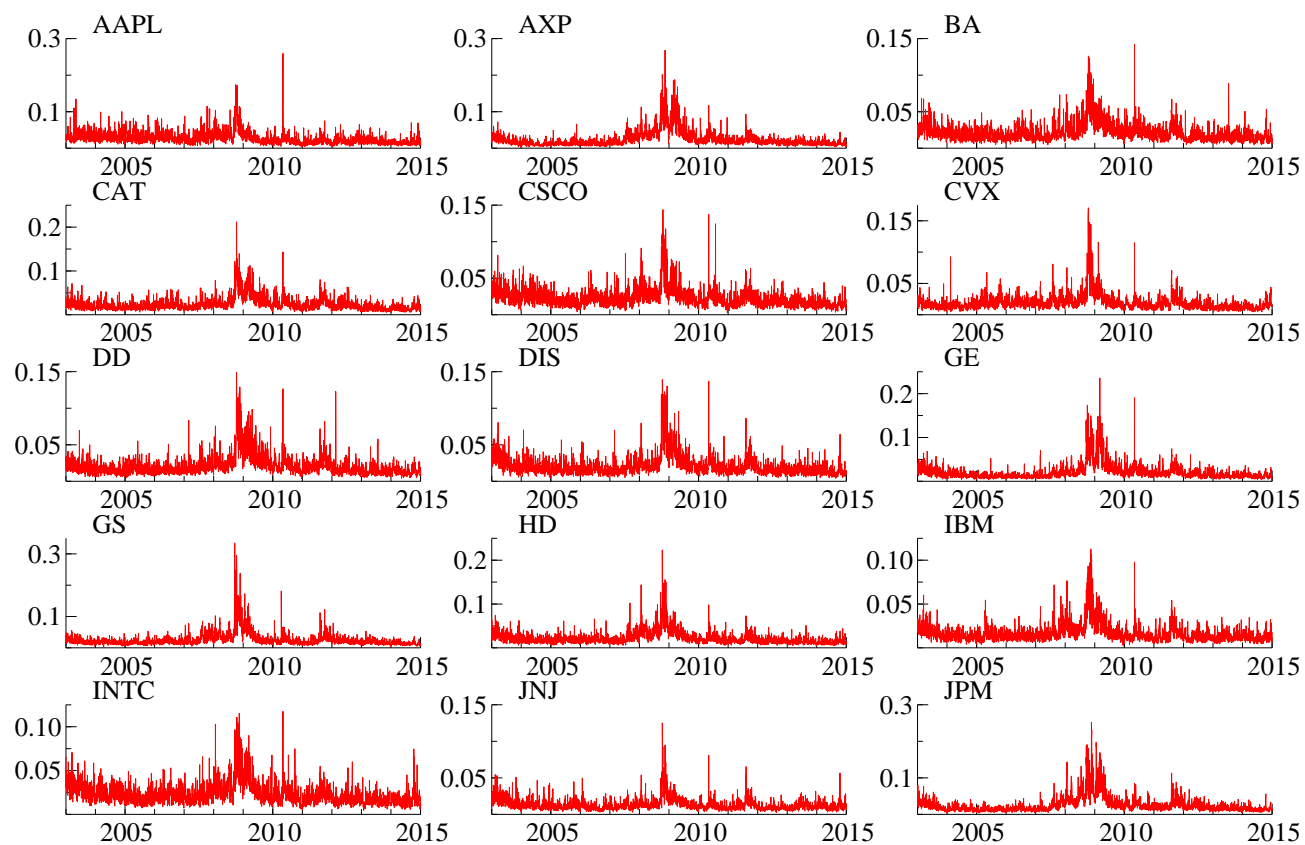
Note: Daily data from January 2, 2003 to December 31, 2014).

Figure S.3: Dow Jones Realised Variance Data, part 2
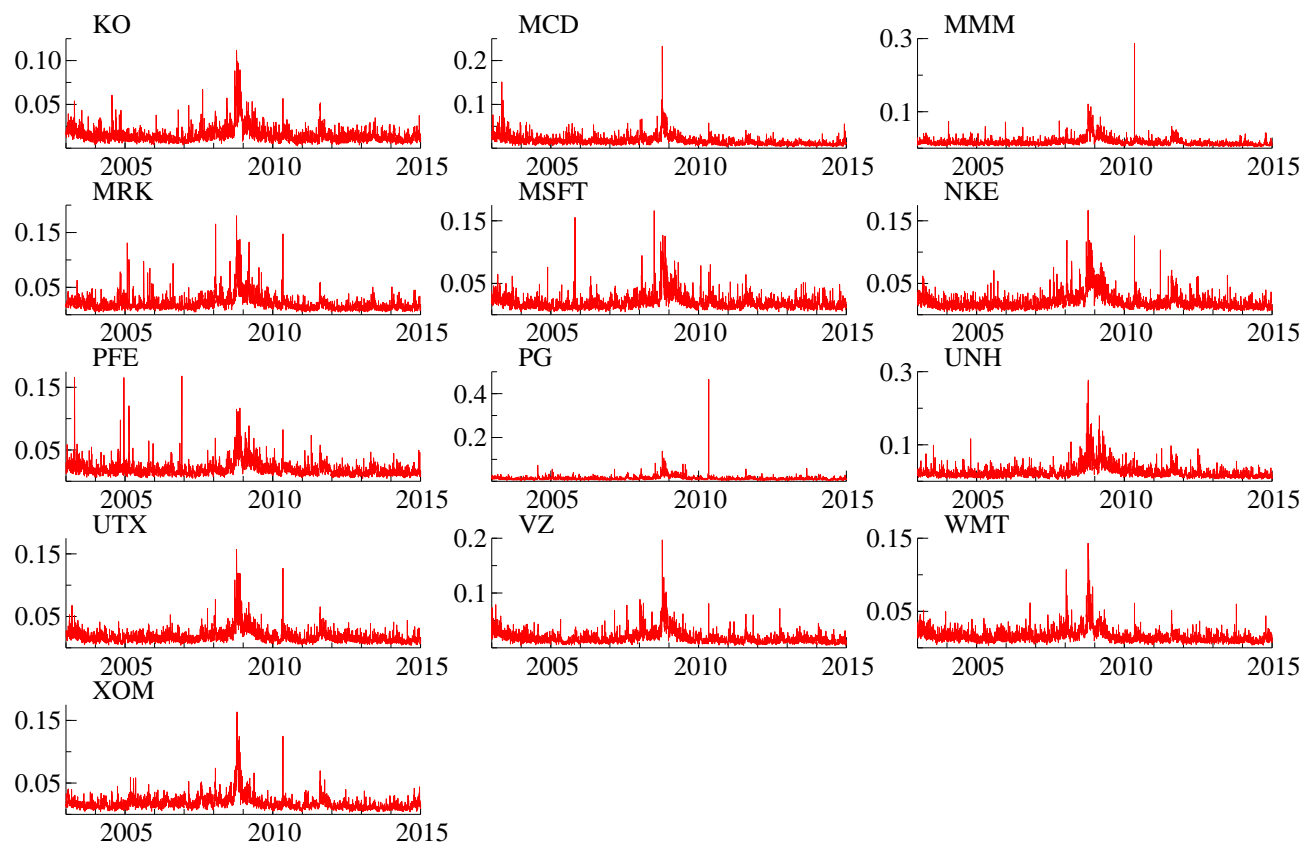
Note: Daily data from January 2, 2003 to December 31, 2014).

Figure S.4: Government Bond Yield Data
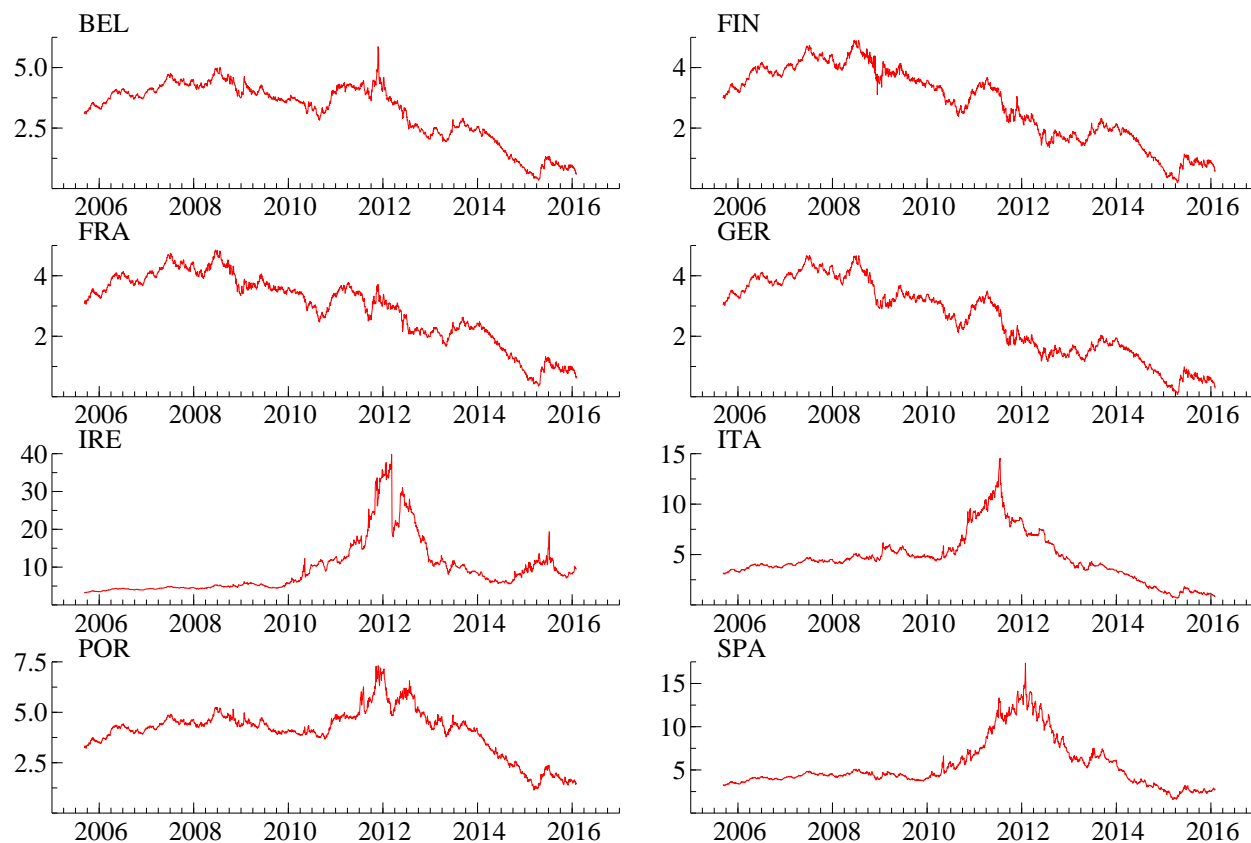
Note: Daily data on 10-year government bond yields from September 5, 2005, to February 2, 2016.

Figure S.5: Residual Graphics for Andersen *et al.* RV Data Examples
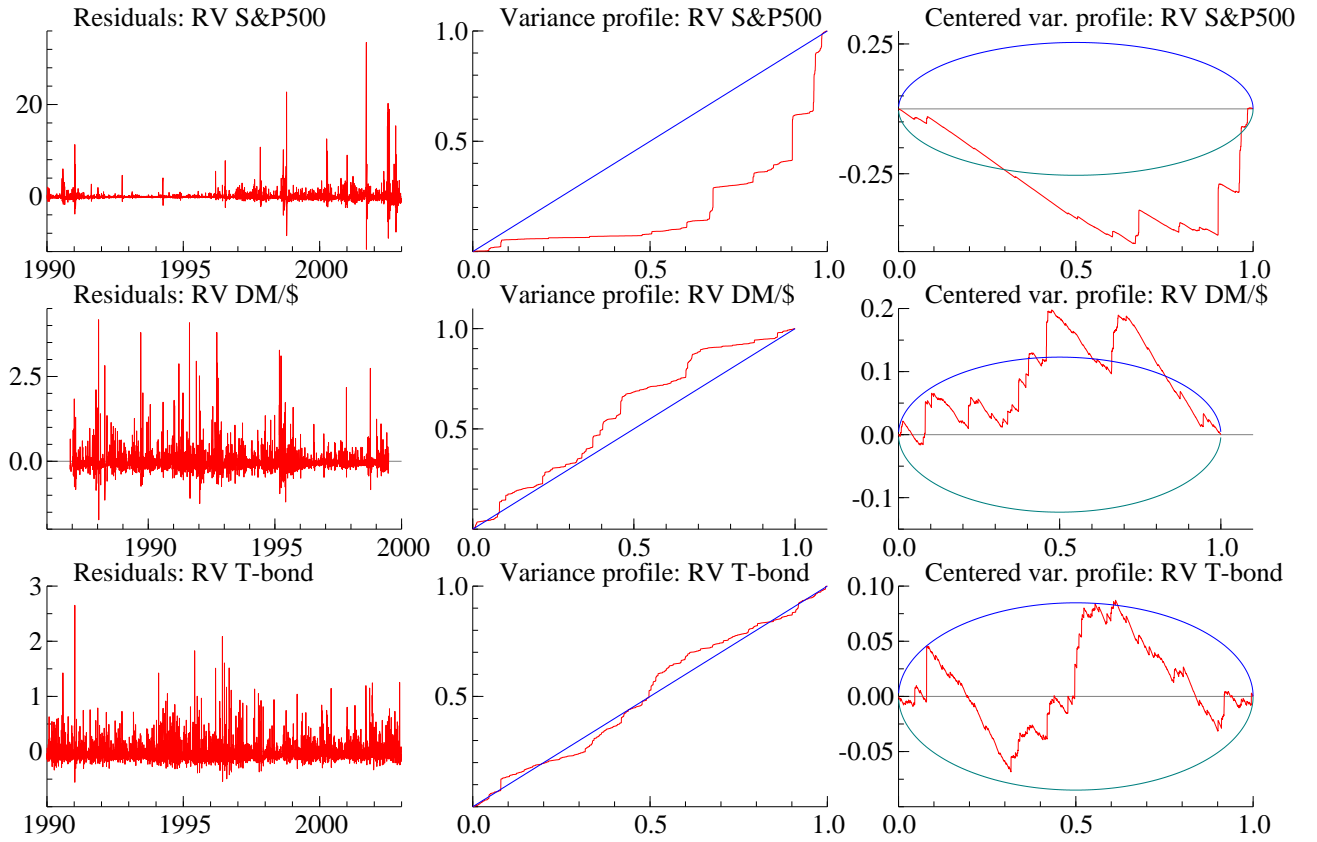
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Figure S.6: Residual Graphics for Dow Jones RV Data Examples, part 1
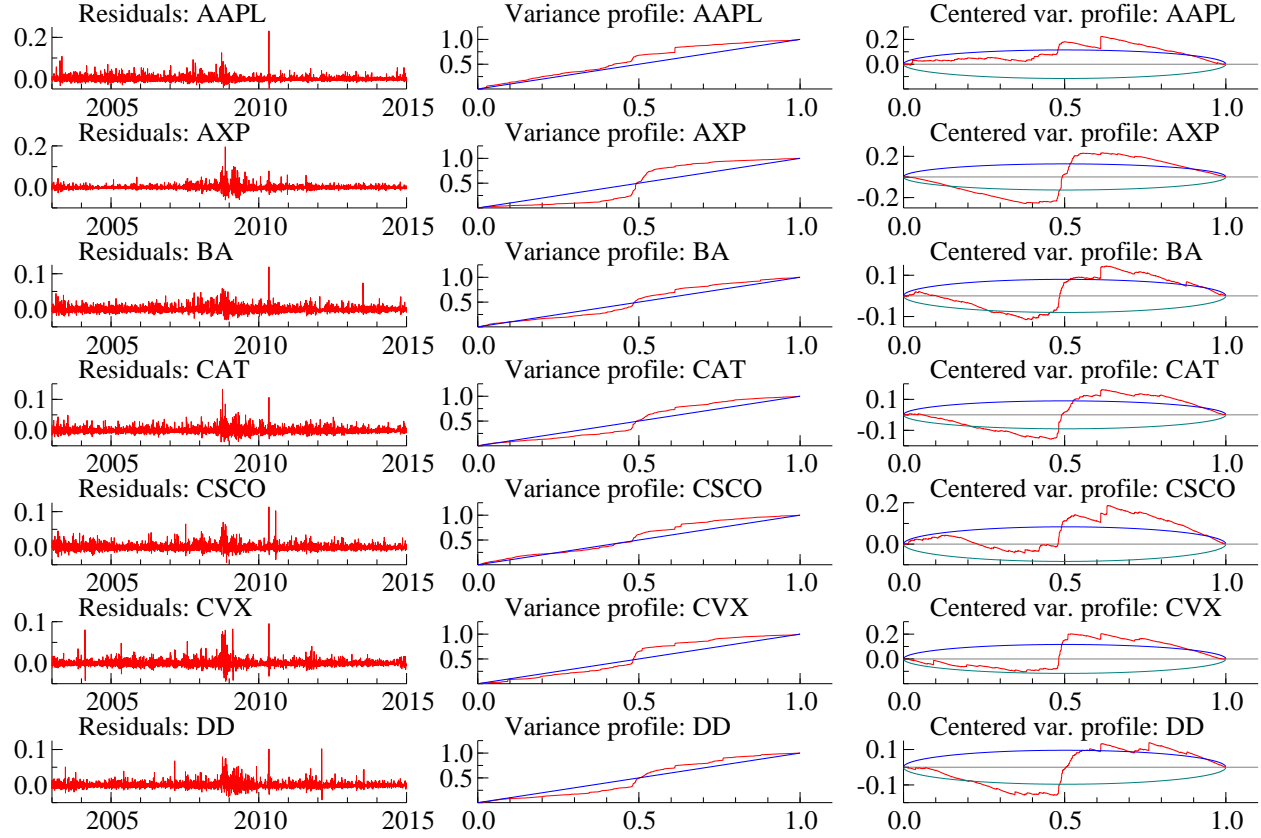
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Figure S.7: Residual Graphics for Dow Jones RV Data Examples, part 2
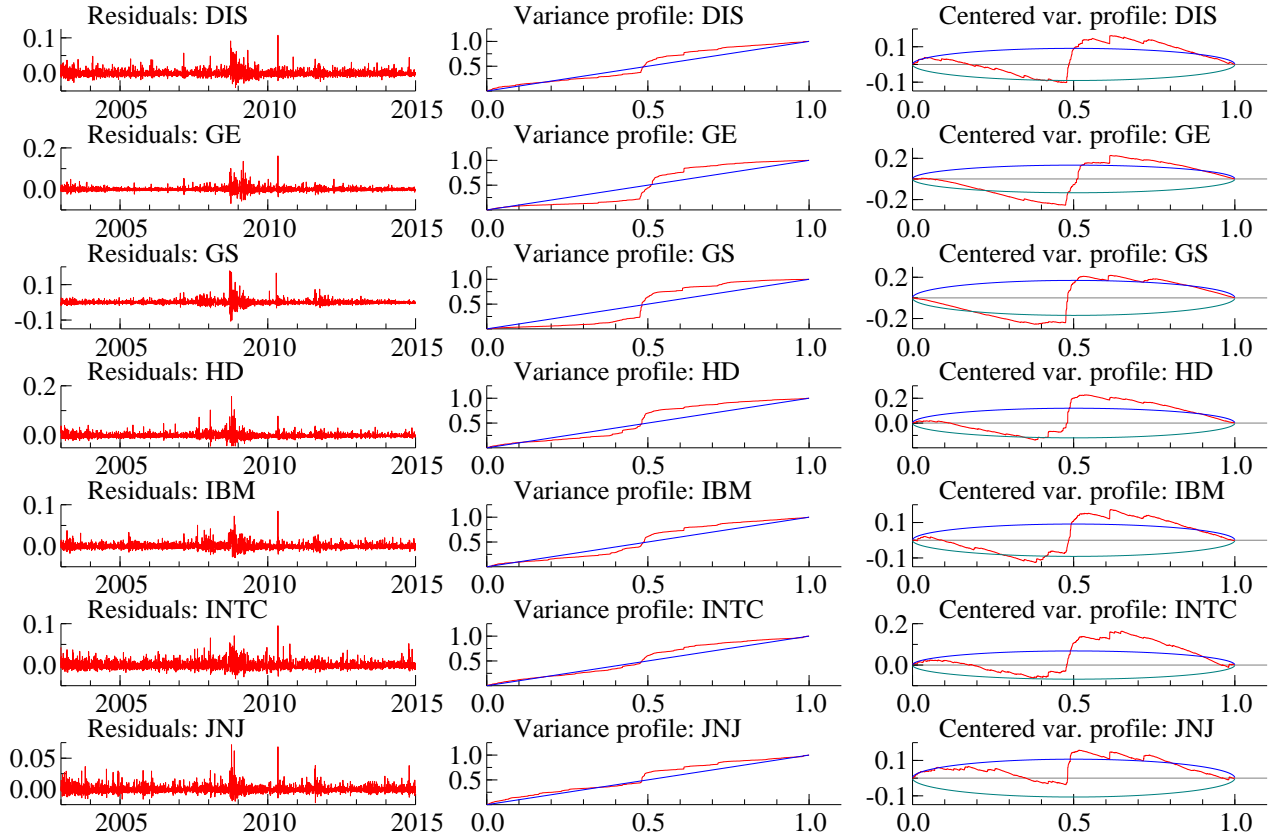
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Figure S.8: Residual Graphics for Dow Jones RV Data Examples, part 3
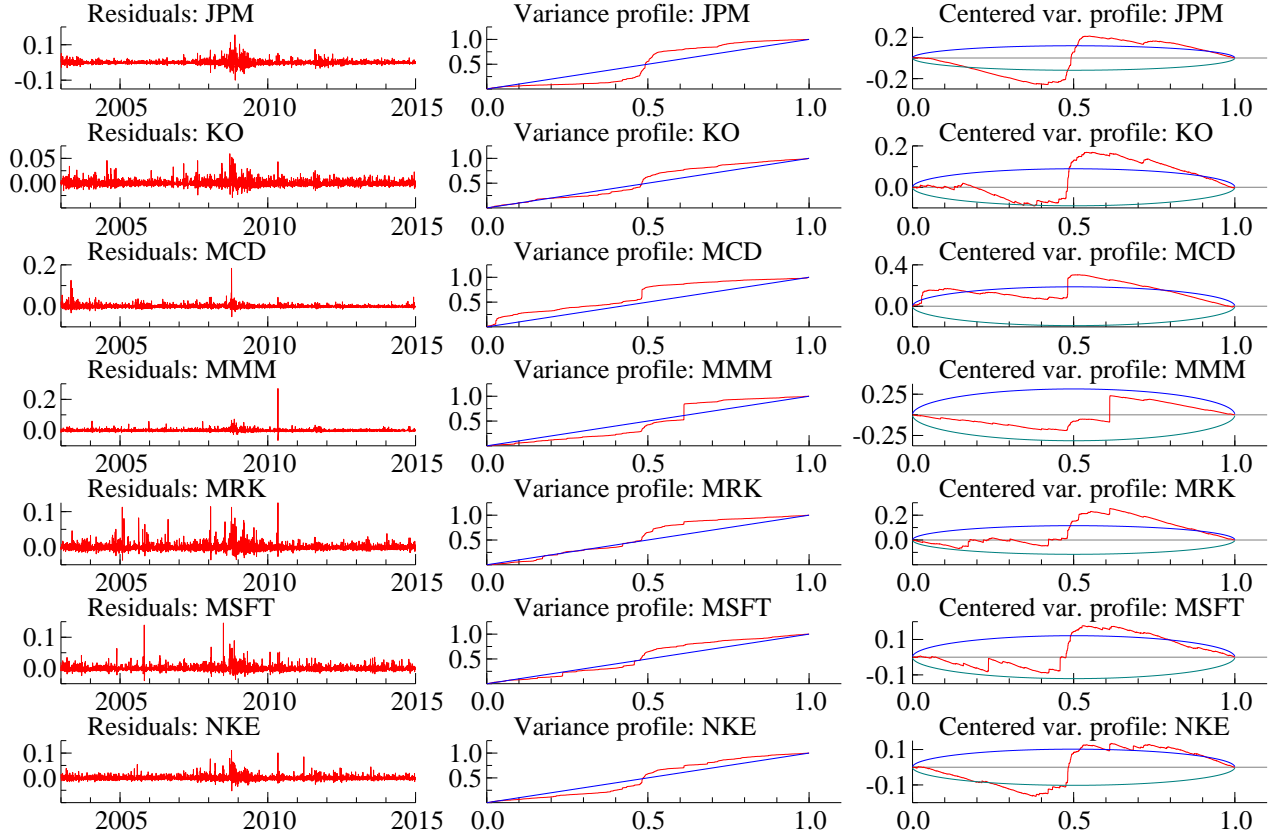
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Figure S.9: Residual Graphics for Dow Jones RV Data Examples, part 4
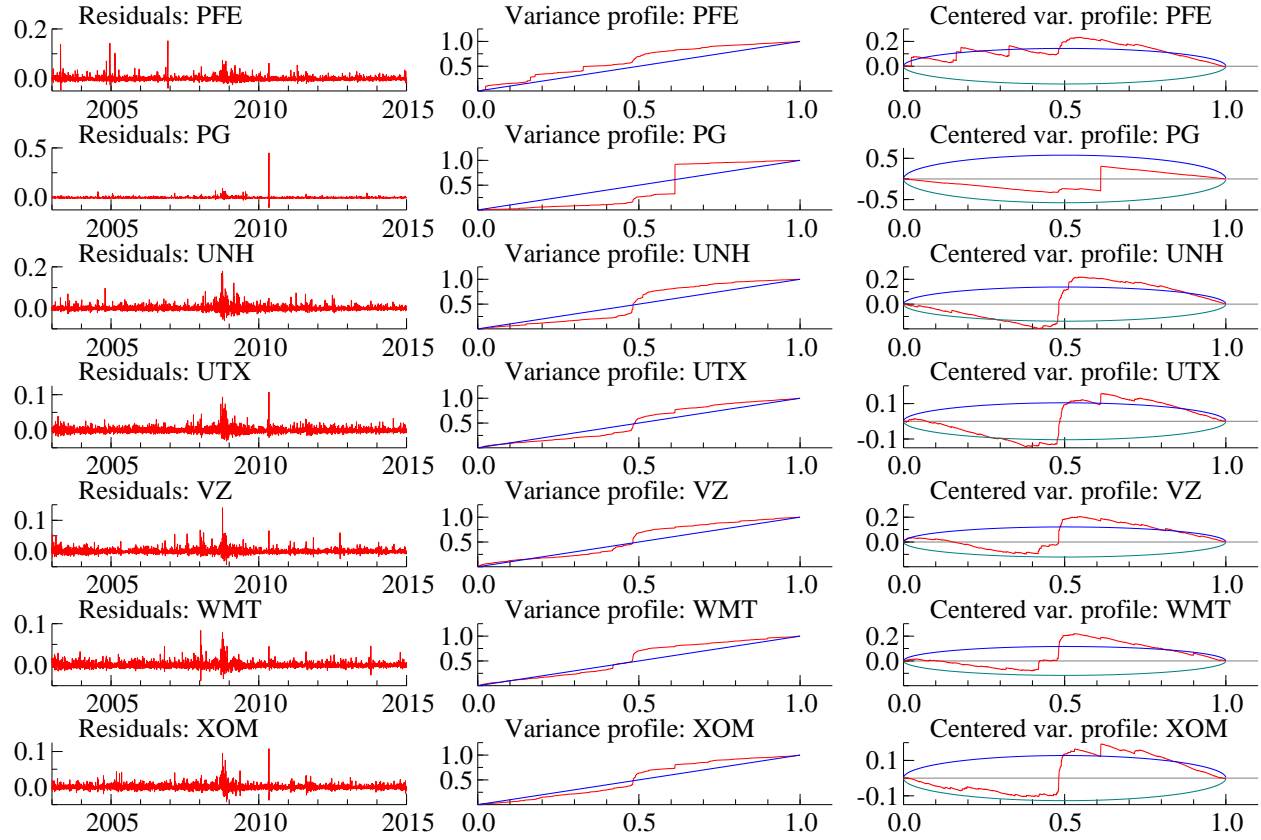
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Figure S.10: Residual Graphics for Government Bond Yield Data Examples, part 1
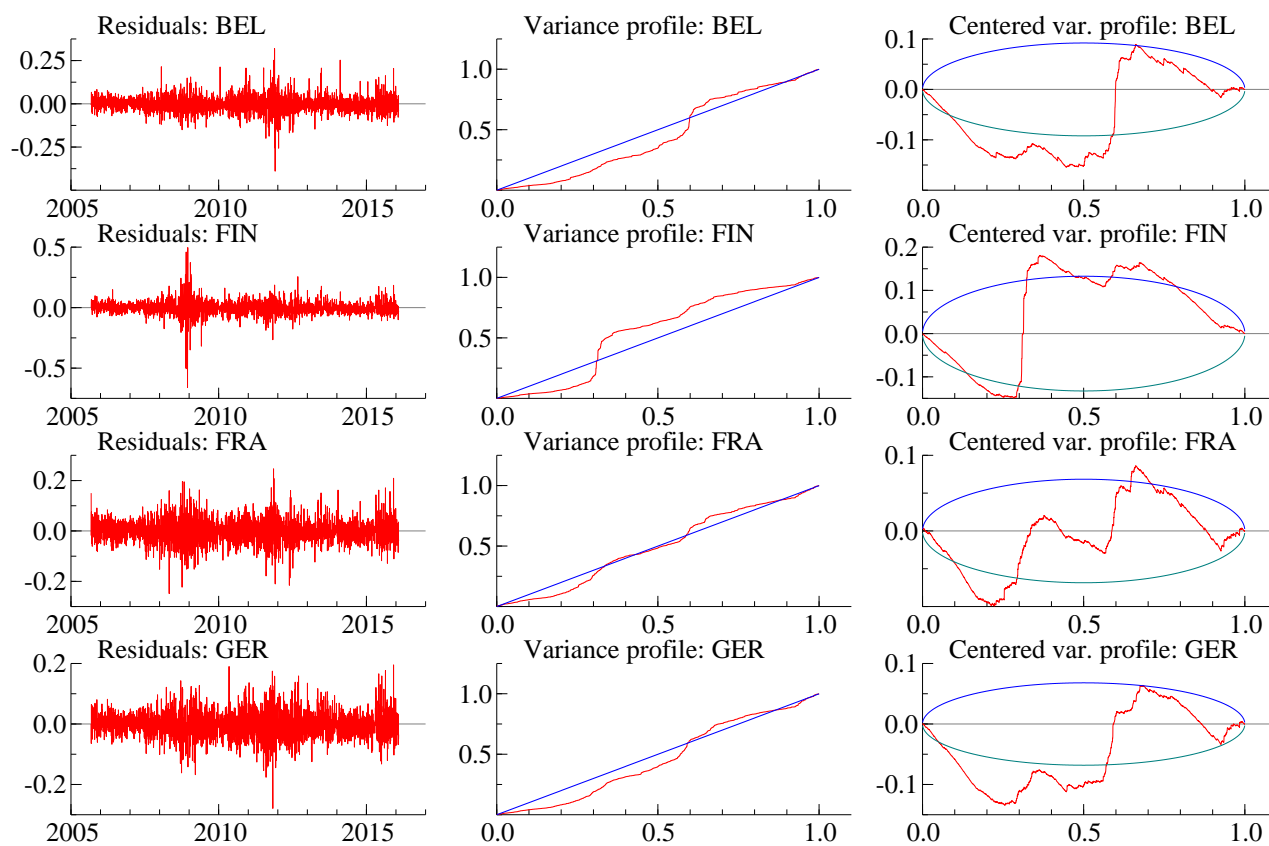
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Figure S.11: Residual Graphics for Government Bond Yield Data Examples, part 2
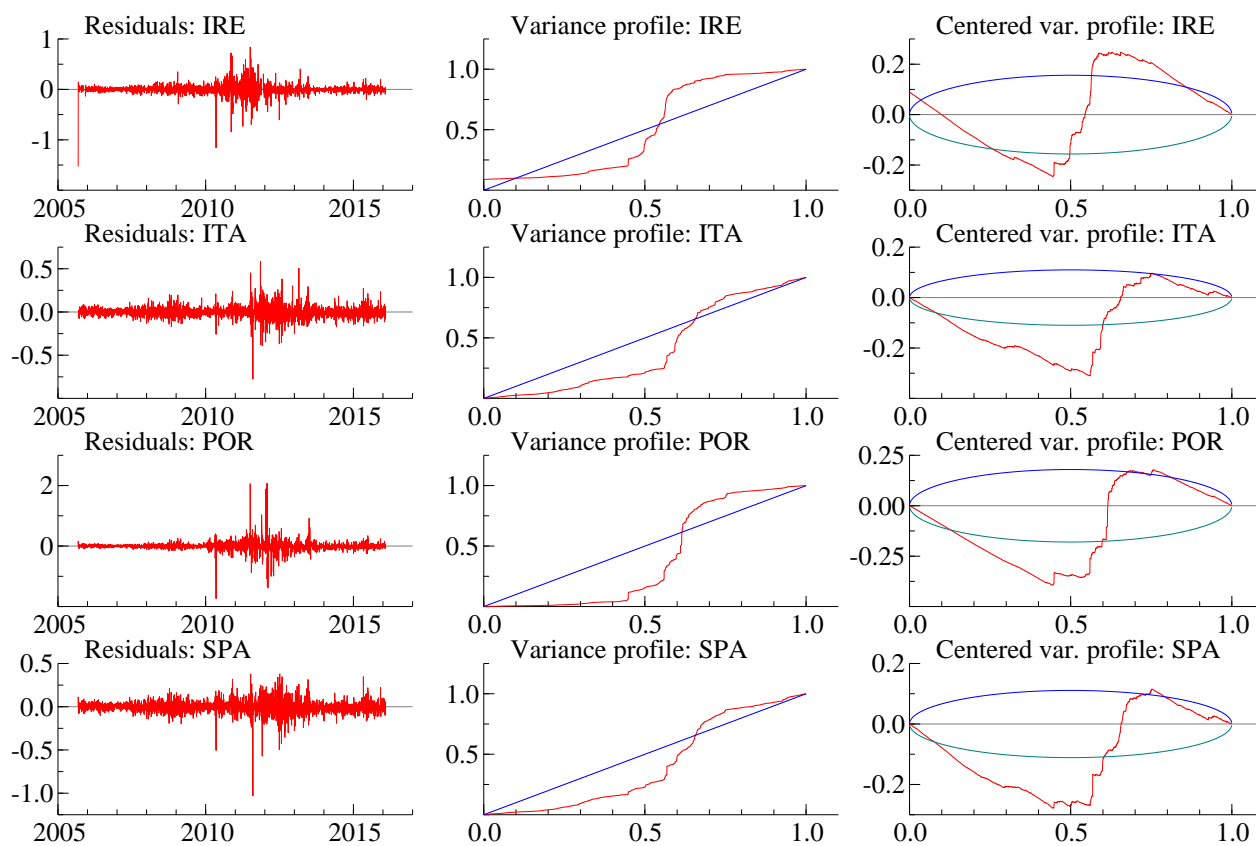
Note: Left panels show time series plots of residuals, middle panels show residual variance profiles, $\hat{\eta}(u)$, and right panels show centered variance profiles with 95% confidence bands.

Research Papers
2020

CREATES

Center for Research in Econometric
Analysis of Time Series