Accepted Manuscript

Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets

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

PII: S0304-4076(15)00064-0
DOI: http://dx.doi.org/10.1016/j.jeconom.2015.02.039
Reference: ECONOM 4061

To appear in: *Journal of Econometrics*

Please cite this article as: Cavaliere, G., Nielsen, M.Ø., Taylor, A.M.R., 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* (2015), http://dx.doi.org/10.1016/j.jeconom.2015.02.039

Bootstrap Score Tests for Fractional Integration in Heteroskedastic ARFIMA Models, with an Application to Price Dynamics in Commodity Spot and Futures Markets*

Giuseppe Cavaliere[a], Morten Ørregaard Nielsen[b] and A.M. Robert Taylor[c]
[a] University of Bologna
[b] Queen's University and CREATES
[c] University of Essex

Abstract

Empirical evidence from time series methods which assume the usual I(0)/I(1) paradigm suggests that the efficient market hypothesis, stating that spot and futures prices of a commodity should cointegrate with a unit slope on futures prices, does not hold. However, these statistical methods are known to be unreliable if the data are fractionally integrated. Moreover, spot and futures price data tend to display clear patterns of time-varying volatility which also has the potential to invalidate the use of these methods. Using new tests constructed within a more general heteroskedastic fractionally integrated model we are able to find a body of evidence in support of the efficient market hypothesis for a number of commodities. Our new tests are wild bootstrap implementations of score-based tests for the order of integration of a fractionally integrated time series. These tests are designed to be robust to both conditional and unconditional heteroskedasticity of a quite general and unknown form in the shocks. We show that the asymptotic tests do not admit pivotal asymptotic null distributions in the presence of heteroskedasticity, but that the corresponding tests based on the wild bootstrap principle do. A Monte Carlo simulation study demonstrates that very significant improvements in finite sample behaviour can be obtained by the bootstrap *vis-à-vis* the corresponding asymptotic tests in both heteroskedastic and homoskedastic environments.

Keywords: Bootstrap; efficient market hypothesis; fractional integration; score tests; spot and futures commodity prices; time-varying volatility

J.E.L. Classifications: C12, C22, C58, G13, G14.

1 Introduction

A large body of empirical literature has developed aimed at assessing to what extent futures commodity markets are efficient. Suppose we let s_t denote the (log) spot price of a particular commodity at time t, and let $f_t^{(k)}$ denote the (log) price of the corresponding k-period futures contract at time t, with k a positive constant. Then, in its simplest form, the Efficient Market Hypothesis (EMH, hereafter) states that in a frictionless market $f_t^{(k)}$ is an unbiased predictor of s_{t+k}; that is,

$$f_t^{(k)} = E\left(s_{t+k}|\mathcal{I}_t\right), \tag{1.1}$$

*Parts of this paper were written while Cavaliere and Taylor visited the Economics Department at Queen's University, Canada, whose hospitality is gratefully acknowledged, as John Weatherall Distinguished Fellows. We are grateful to the Guest Editor, Michael McAleer, three anonymous referees, David Harris and Søren Johansen for their helpful comments on earlier versions of this paper. We also thank Paresh Narayan for providing us with his data and David Harris for sharing his computer code with us. Nielsen thanks the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78) for financial support. Cavaliere and Taylor thank the Danish Council for Independent Research, Sapere Aude | DFF Advanced Grant (Grant nr: 12-124980) for financial support. Correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: `rtaylor@essex.ac.uk`

where $\mathcal{I}_t$ denotes the available information set; that is, the sigma-algebra generated by current and past values of $x_t := (s_t, f_t)'$. Equivalently, letting $u_t^{(k)} := f_{t-k}^{(k)} - s_t$ denote the so-called forward premium, the relation (1.1) can be reformulated as

$$E(u_{t+k}^{(k)}|\mathcal{I}_t) = 0, \tag{1.2}$$

which asserts that the expected forward risk premium is zero. Under the standard assumption of (log) spot prices being well approximated by (possibly heteroskedastic) I(1) processes, the relations in (1.1) and (1.2) imply that: (i) $f_t^{(k)}$ is I(1); (ii) s_{t+k} and $f_t^{(k)}$ are co-integrated; (iii) the co-integrating vector has the form $\beta = (1, -1)'$; (iv) the co-integrating residuals (or spread), $s_t - f_{t-k}^{(k)}$, form a (possibly heteroskedastic) martingale difference sequence. Weaker forms of the EMH require that, due to time varying risk premia, interest rates and storage costs, in equilibrium, the right-hand side of (1.2) is equal to some arbitrary (possibly nonzero) constant (see, e.g., Luo, 1998), and that in place of (iv) we have the weaker condition (iv') $u_t^{(k)}$ can be described as a mean reverting, stationary (aside from possible heteroskedasticity) process. Observe that this need not therefore be an I(0) process, as, for example, any fractionally integrated I(d) process with $d < 1/2$ satisfies condition (iv').

Despite widespread acceptance of the EMH in theory, the long-run one-for-one relationship between spot and futures prices it postulates has proven difficult to verify empirically; see e.g. Baillie and Bollerslev (1994) and Figuerola-Ferretti and Gonzalo (2010) for detailed discussions of early and more recent empirical evidence, respectively. Although the presence of unit roots in both spot and futures prices tends to be supported for most commodities when data are analyzed by means of standard stationarity and unit root tests, most of the early empirical evidence based on the usual I(0)/I(1) paradigm rejected the hypothesis of co-integration between spot and future prices; see the discussion in Westerlund and Narayan (2013) and the references therein. While more recent approaches, although still in the standard I(0)/I(1) paradigm, do often find some form of co-integration for most commodities, they still, however, tend to reject the $(1, -1)'$ co-integrating vector in (iii); see *inter alia* Figuerola-Ferretti and Gonzalo (2010), Westerlund and Narayan (2013), and the references therein.

Most of the empirical evidence is based on the assumptions that the data are well described by I(d) processes with $d = 0$ or $d = 1$ and that the degree of possible (conditional and unconditional) heteroskedasticity in the series is small enough to guarantee that standard statistical procedures for (co-)integrated conditionally i.i.d. data apply. Both assumptions, however, would appear to be at odds with the empirical features of price series in both spot and futures markets, and indeed in financial data more generally.[1] Regarding the first assumption, researchers have reasonably claimed that data seem to be better characterised by fractional integration, i.e. by a general I(d) process, in particular where the forward premium $u_t^{(k)}$ is concerned; see, e.g., Baillie and Bollerslev (1994, 2000). Consequently, inference methods which do not allow for the possibility of fractional integration in the data will be biased where it is present, in the sense that they will tend to reject $(1, -1)'$ co-integration between spot and forward prices; see Maynard and Phillips (2001). Regarding the second assumption, it is now a well established fact that the existence of time-varying conditional and unconditional volatility can seriously affect standard inference procedures for unit root and co-integration tests (Cavaliere and Taylor, 2007, 2008*a*, 2009, and Cavaliere, Rahbek and Taylor, 2014). Hence, existing evidence against co-integration and/or a $(1, -1)'$ co-integration relation between spot and futures prices is likely to be affected by time-varying conditional and/or unconditional volatility in the data. Moreover, as we show in this paper, inference on the fractional integration order is very likely to be affected by time-varying behaviour in the volatility process; that is, existing evidence of fractional integration in futures markets may also be driven by non-stationarity in the second-order moments.

In response to these issues we test fractional co-integration with co-integrating vector $(1, -1)$ imposed, by testing the degree of fractional integration the the spread or forward premium. Thus, we focus on the problem of conducting inference on the fractional integration (long memory) parameter, based around the score or Lagrange multiplier [LM] principle, in univariate autoregressive fractionally integrated moving average [ARFIMA] time series which display time-variation in the volatility process

[1] For example, Sensier and van Dijk (2004) report that over 80% of the real and price variables in the Stock and Watson (1999) data-set reject the null of constant innovation variance, while Loretan and Phillips (1994) report evidence against the constancy of unconditional variances in stock market returns and exchange-rate data.

of the driving shocks. We allow for both unconditional heteroskedasticity (often referred to as non-stationary volatility in the literature) and conditional heteroskedasticity in our analysis. The score test for fractional integration was pioneered by Robinson (1991, 1994) and has been applied in early empirical work by, e.g., Gil-Alana and Robinson (1997), among numerous other studies. The classical likelihood-based tests, and in particular the score-based tests, for inference on the long memory parameter have been derived under the assumption of conditionally (and, hence, unconditionally) homoskedastic shocks; see, among others, Robinson (1994), Agiakloglou and Newbold (1994), Tanaka (1999), Nielsen (2004), Lobato and Velasco (2007), and Johansen and Nielsen (2010). Very few contributions in the fractional integration literature investigate the impact of time-varying volatility on inference in long memory series. A small number of papers have considered the case where the shocks can display certain forms of conditional heteroskedasticity (but maintaining the assumption of unconditional homoskedasticity); see, for example, Robinson (1991), Baillie, Chung, Tieslau (1996), Ling and Li (1997), Ling (2003), Demetrescu, Kuzin and Hassler (2008) and Hassler, Rodrigues and Rubia (2009). To the best of our knowledge, the only paper in this literature which explicitly allows for non-stationary volatility is Kew and Harris (2009) who extend the idea of Demetrescu, Kuzin and Hassler (2008) to use heteroskedasticity-robust White (1980)-type standard errors when computing regression-based tests for fractional integration. They apply this approach to the tests proposed in Dolado, Gonzalo and Mayoral (2002) and Lobato and Velasco (2006, 2007), Agiakloglou and Newbold (1994) and Breitung and Hassler (2002).

This paper makes two distinct contributions. The first is to the theoretical econometrics literature. Here we examine the impact of time-varying conditional and/or unconditional volatility on standard score-based tests for the long memory parameter. Our analysis is based on a new framework which includes the general form of non-stationary volatility considered in Cavaliere and Taylor (2008*a*) as a special case and also includes a set of conditional heteroskedasticity conditions similar to those employed in Robinson (1991), Demetrescu, Kuzin and Hassler (2008) and Hassler, Rodrigues and Rubia (2009), among others. Neither of these conditions involve specifying a parametric model for the volatility process. We show that the limiting distributions of the score test statistics under both the null and local alternatives are non-pivotal with their functional form depending on nuisance parameters which derive from the heteroskedasticity present in the shocks. Consequently, inference based on conventional asymptotic critical values leads to tests which are not in general asymptotically correctly sized under the null when heteroskedasticity is present. In response to this we propose wild bootstrap implementations of the score statistics[2] which are shown to correctly replicate their (first order) limiting null distributions. As a result, asymptotically valid bootstrap inference can be performed in the presence of time-varying volatility using wild bootstrap implementations of these tests. Simulation evidence is reported which clearly demonstrates the superior finite sample properties of our proposed bootstrap tests over their asymptotic counterparts in both homoskedastic and heteroskedastic environments.

Our second contribution is to employ our new bootstrap tests to re-analyse the sample of daily data covering the period 2005–2011 for four commodities – gold, silver, platinum and crude oil – recently analysed in Westerlund and Narayan (2013). As Narayan, Huson and Narayan (2012) point out, these four commodities together constitute 76% of total commodities trading, with crude oil the most commonly traded. Westerlund and Narayan (2013) find strong evidence of conditional heteroskedasticity in both the spot and futures prices of each of these commodities and, as a result, recommend using weighted least squares, based on the assumption that volatility follows a finite-order ARCH process, to estimate the co-integrating relationship between the spot and futures prices, again within an I(0)/I(1) paradigm. In recognition of the financial crisis, and the associated increase in the *unconditional* volatility apparent in all of these series, they also consider splitting the sample at September 2008. The methods developed in this paper can control for a wide class of conditionally heteroskedastic processes without the need to specify a parametric model, unlike Westerlund and Narayan (2013), and simultaneously to allow for changes in the unconditional volatility of the process,

[2] An i.i.d. bootstrap implementation is also considered in the working paper version, see Cavaliere, Nielsen, and Taylor (2013). As would be expected, it correctly replicates the asymptotic null distribution of the statistics only under constant volatility so that inference with the i.i.d. bootstrap is not valid under heteroskedasticity.

including any associated with the recent financial crisis. Our methods also allow us to move beyond the strictures of the pure I(0)/I(1) paradigm, thereby permitting valid testing on condition (iv') in cases where the spread is stationary but not I(0). We find significant evidence of conditional heteroskedasticity in all of the series and of unconditional heteroskedasticity in all but the silver series. The results from our bootstrap tests suggest that (the weak form of) the EMH holds within a standard I(1) to I(0) co-integrated relationship for silver and platinum. For gold the EMH is accepted but within a fractionally co-integrated stationary relationship. For oil, our results suggest the spread is fractionally co-integrated but non-stationary. A rolling sub-sample analysis of the data is also reported and this does not appear to uncover any major within-sample departures from these conclusions.

The remainder of the paper is organised as follows. Section 2 outlines our heteroskedastic, fractionally integrated ARFIMA model. Section 3 analyses the effects of time-varying volatility on the large sample behaviour of the standard (asymptotic) score-based tests for hypotheses on the fractional integration parameter. The bootstrap algorithm and related bootstrap score-based tests are outlined in section 4, and the large sample properties of these are established. The results of a Monte Carlo study are given in section 5. Section 6 contains the empirical analysis of the EMH for futures markets, and section 7 concludes. Mathematical proofs are contained in the appendix.

In the following, $\xrightarrow{w}$ denotes weak convergence, $\xrightarrow{p}$ convergence in probability, $\xrightarrow{L_r}$ convergence in L_r-norm, and $\xrightarrow{w}_p$ weak convergence in probability, in each case as $T \to \infty$; for any space A, $\text{int}(A)$ denotes the interior of A; $\mathbb{I}(\cdot)$ denotes the indicator function; $x := y$ indicates that x is defined by y; for any square matrix, A, $\|A\|$ is used to denote the norm $\|A\|^2 := \mathsf{tr}\{A'A\}$; for any vector, x, $\|x\|$ denotes the usual Euclidean norm, $\|x\| := (x'x)^{1/2}$; for any matrix, A, $(A)_{m,n}$ denotes its (m,n)'th element and for any vector, x, $(x)_m$ denotes its m'th element; for any real number, x, $\lfloor x \rfloor$ denotes the integer part of x. Throughout, we use the notation K for a generic, finite constant.

2 The Heteroskedastic ARFIMA Model

Consider the real-valued, fractionally integrated stochastic process $\{y_t, t = 1, 2, ..., T\}$ generated by

$$\Delta_+^d y_t = u_t, \tag{2.1}$$

where the operator Δ_+^d is given by $\Delta_+^d z_t := \Delta^d z_t \mathbb{I}(t \geq 1) = \sum_{i=0}^{t-1} \pi_i(-d) z_{t-i}$ with $\pi_i(v) := \frac{\Gamma(v+i)}{\Gamma(v)\Gamma(1+i)} = (i!)^{-1}(v(v+1)\dots(v+i-1))$ denoting the coefficients in the usual binomial expansion of $(1-z)^{-v}$. The unobserved shock process $\{u_t\}$ is assumed to have the following ARMA(p,q) generating mechanism

$$c(L, \psi) u_t = \varepsilon_t, \tag{2.2}$$

where $c(z, \psi) := a(z, \psi) / b(z, \psi)$ and $a(z, \psi)$ and $b(z, \psi)$ are polynomial functions (of orders p and q, respectively) in the complex variate z, depending on the $k \times 1$ parameter vector ψ. The parameters of the model are collected in the vector $\gamma := (d, \psi')'$ with true value denoted by $\gamma_0 := (d_0, \psi_0')'$. The polynomials $a(z, \psi)$ and $b(z, \psi)$ are assumed to satisfy the following standard conditions.

Assumption $\mathcal{R}$ *The parameter space for ψ is Ψ, which is convex, compact, and such that, for all $\psi \in \Psi$, the polynomial functions $a(z, \psi)$ and $b(z, \psi)$ of the complex variate z have no common roots and all their roots lie strictly outside the unit circle.*

The innovation process $\{\varepsilon_t\}$ is taken to satisfy the following assumption, which embodies both unconditional heteroskedasticity (part (a)) and conditional heteroskedasticity (part (b)).

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

(a) *$\{\sigma_t\}_{t \in \mathbb{Z}}$ is non-stochastic and uniformly bounded, and satisfies $\sigma_t := \sigma(t/T) > 0$ for all $t = 1, ..., T$, where $\sigma(\cdot) \in D[0,1]$, the space of càdlàg functions on $[0,1]$.*

(b) *$\{z_t\}$ is a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t$, the sigma-field generated by $\{z_s\}_{s \leq t}$, such that $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ for $t = ..., -1, 0, 1, 2, ...,$ and satisfies*

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

(ii) $\tau_{r,s} := E(z_t^2 z_{t-r} z_{t-s})$ *is uniformly bounded for all* $t \geq 1, r \geq 0, s \geq 0$, *where also* $\tau_{r,r} > 0$ *for all* $r \geq 0$,

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

A special case of Assumption $\mathcal{V}$, where $\sigma(\cdot)$ is constant and $\{z_t\}$ is conditionally homoskedastic, is, in addition to a higher-order moment condition, the following classical assumption.

Assumption $\mathcal{H}$ *The innovations* $\{\varepsilon_t\}$ *form a martingale difference sequence with respect to the filtration* $\mathcal{F}_t$, *where, almost surely,* $E\left(\varepsilon_t^2 | \mathcal{F}_{t-1}\right) = \sigma^2$.

Assumption $\mathcal{H}$ is a conditional homoskedasticity requirement for martingale differences, dating back to, at least, Hannan (1973), and is rather standard in the time series literature. Conversely, Assumption $\mathcal{V}$ allows both conditional and unconditional heteroskedasticity of very general forms.

The conditions in part (a) of Assumption $\mathcal{V}$, see Cavaliere and Taylor (2008*a*), imply that the unconditional innovation variance σ_t^2 is only required to be bounded and to display at most a countable number of jumps, therefore allowing for an extremely wide class of potential models for the behaviour of the unconditional variance of ε_t. Models of single or multiple variance shifts satisfy part (a) of Assumption $\mathcal{V}$ with $\sigma(\cdot)$ piecewise constant; e.g., the case of a single break at time $\lfloor T\tau \rfloor$ obtains for $\sigma(u) := \sigma_0 + (\sigma_1 - \sigma_0)\mathbb{I}(u > \tau)$. If $\sigma(\cdot)$ is an affine function, then σ_t displays a linear trend. Piecewise affine functions are permitted, thereby allowing for variances which follow a broken trend, as are smooth transition variance shifts. The requirement within part (a) of Assumption $\mathcal{V}$ that $\sigma(\cdot)$ is non-stochastic is made to simplify the analysis, but can be generalised to allow for cases where $\sigma(\cdot)$ is stochastic and independent of z_t; see Cavaliere and Taylor (2009) for further details.

Part (b) of Assumption $\mathcal{V}$ allows for conditional heteroskedasticity in $\{z_t\}$. We do not assume a parametric model of the generalized autoregressive conditional heteroskedasticity form as in, e.g., Baillie *et al.* (1996), Ling and Li (1997) and Ling (2003). Instead, the conditions in Assumption $\mathcal{V}$(b) allow for conditional heteroskedasticity of unknown and very general form and are typical of those used in this literature; see, for example, Robinson (1991), Demetrescu, Kuzin and Hassler (2008), Hassler, Rodrigues and Rubia (2009) and Kew and Harris (2009). In particular, part (b)(iii) controls the extent of higher-order dependence in $\{z_t\}$. However, we note that the conditions given in part (b) of Assumption $\mathcal{V}$ are somewhat weaker than required by these authors. Firstly, they impose the assumption that, for any integer q, $2 \leq q \leq 8$, and for q non-negative integers s_i, $E(\prod_{i=1}^{q} z_{t_i}^{s_i}) = 0$ when at least one s_i is exactly one and $\sum_{i=1}^{q} s_i \leq 8$, see, e.g., Assumption E(e) of Kew and Harris (2009). This implies, in particular, that $\tau_{r,s} = 0$ for $r \neq s$, which rules out asymmetric conditionally heteroskedastic processes. We are not aware of any other work in the fractional integration literature that allows for $\tau_{r,s} \neq 0$. Secondly, these authors assume strict stationarity of z_t, which we do not.

Remark 2.1 Observe that the moment condition $\sup_t E|z_t|^8 < \infty$, imposed by a number of other authors, is necessary for part (b)(iii) with $q = 8$ to hold and therefore is not stated explicitly. Moreover, notice that the boundedness assumption in (b)(ii) does in fact follow from the conditions imposed in (b)(iii). Finally, notice also that the assumption that z_t is a martingale difference sequence implies that for any $\kappa_q(\cdot)$, $q \geq 2$, if the highest argument in the cumulant appears only once, then the cumulant is zero; see Lemma A.2 in Cavaliere, Nielsen and Taylor (2014) [CNT]. Hence, our stated assumptions deal only with cumulants where the first two (the highest) arguments coincide.

Remark 2.2 A time series generated according to Assumption $\mathcal{V}$ formally constitutes a triangular array of the type $\{\varepsilon_{T,t} : 0 \leq t \leq T, T \geq 1\}$, where $\varepsilon_{T,t} = \sigma_{T,t} z_t$ and $\sigma_{T,t} = \sigma(t/T)$. Because the triangular array notation is not essential, for simplicity the subscript T is suppressed in the sequel.

Remark 2.3 Deterministic terms such as an unknown mean, trend, and/or seasonal means can also be added to the model by assuming that the observed process is $\beta' x_t + y_t$, where y_t is generated by (2.1), x_t is the deterministic term, and the coefficient β is estimated by maximum likelihood jointly with the other parameters. Under very weak conditions, not even requiring the usual Grenander-Rosenblatt assumptions, estimated deterministic terms would not alter the form of the asymptotic distributions given in this paper due to the block-diagonality of the Hessian matrix; see, e.g., Robinson (1994) and Nielsen (2004). However, we leave out deterministic terms to simplify the notation and discussion.

Remark 2.4 Our model (2.1) is fractionally integrated of type II, where the fractional differencing filter is truncated, i.e. the Δ_+ operator. Alternatively, a fractional model of type I would apply integer differencing until the fractional integration order of y_t is in the interval $(-1/2, 1/2)$, and then apply the untruncated fractional differencing operator. The type II model applied in this paper has the advantage that it is applicable for any value of d and without any prior knowledge of the integration order. Furthermore, the likelihood analysis could be made conditional on initial values, in which case the observed sample points would be split between initial values and observations used for estimation. This has consequences for second-order bias terms, but does not influence first-order asymptotics as considered here. For details, see Johansen and Nielsen (2014).

Remark 2.5 As is often done in this literature, the results in this paper are based on knowledge of the AR and MA lag orders, p and q, or at least may be viewed as conditional on the pre-determination of these from a consistent model selection procedure. The results given in Proposition 4.2(a) and (c) of Sin and White (1996) would appear to imply that the usual BIC procedure is consistent in the context of selecting p and q in the heteroskedastic ARFIMA model considered here. In Section 5 we report Monte Carlo results for both known lag orders and where these are chosen by the BIC. An alternative approach, which we do not consider here, is to develop bootstrap versions of the tests outlined in DKH, or indeed the LM and score tests of Agiakloglou and Newbold (1994) and Breitung and Hassler (2002), for the case where (2.2) is specified to be an $AR(\infty)$ process, driven by ε_t and satisfying standard summability and invertibility conditions, approximated by an autoregression whose lag length increases with the sample size at an appropriate rate; see DKH for further details. □

3 Score-based Tests on d

In this section we first derive one-sided and two-sided (quasi-) score tests under the assumption of homoskedastic Gaussian innovations. We then establish the large sample properties of the standard (asymptotic) test statistics based on comparing these statistics with standard (homoskedastic) critical values when the innovations in fact display unconditional and/or conditional heteroskedasticity of an unknown form as given in Assumption $\mathcal{V}$.

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

$$H_0 : d = \bar{d} \tag{3.1}$$

in the context of (2.1). We will test this hypothesis by using score tests in the time domain. The score tests may be performed against either a one-sided or a two-sided alternative. An example of the former is $H_1 : d < \bar{d}$, in which case $d > \bar{d}$ is implicitly part of the null, or vice versa. On the other hand, the more traditional two-sided score test is performed against the two-sided alternative, $H_1 : d \neq \bar{d}$. The one-sided score test is often referred to as Rao's score test; see Lehmann and Romano (2005, pp. 512, 566) for further details. In what follows we will refer to the one-sided score test simply as the score test, and the two-sided score test as the LM test.

To derive the test statistics, first define $\hat{\varepsilon}_t(\gamma) := \hat{\varepsilon}_t(d, \psi) := c(L, \psi)\, \Delta_+^d y_t$. Then the (Gaussian) log-likelihood function, conditionally on the initial values and under the assumption of constant variance, $\sigma(\cdot) = \sigma$, is given, up to a constant term, by $L(d, \psi, \sigma^2) := -\frac{T}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T} \hat{\varepsilon}_t(d, \psi)^2$. Concentrating out the nuisance parameter σ^2 yields, aside from a constant, the concentrated likelihood

$$\ell(d, \psi) := -\frac{T}{2}\log\left(\hat{\sigma}^2(d, \psi)\right), \tag{3.2}$$

where

$$\hat{\sigma}^2(d, \psi) := \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t(d, \psi)^2 . \tag{3.3}$$

The unrestricted conditional quasi-maximum likelihood [QML] estimator is then given as the maximizer of (3.2), which is equivalent to the conditional sum-of-squares estimator given as the minimizer of (3.3). To calculate the score and LM test statistics, estimation is carried out under the null hypothesis. To that end, let a tilde (˜) denote an estimator obtained under the restrictions of the null, i.e. while fixing $d = \bar{d}$. Specifically,

$$\tilde{\psi} := \arg\max_{\psi\in\Psi} \ell\left(\bar{d}, \psi\right) = \arg\min_{\psi\in\Psi} \hat{\sigma}^2\left(\bar{d}, \psi\right), \tag{3.4}$$

and the estimator of the full parameter vector γ under the null is then given by $\tilde{\gamma} = (\bar{d}, \tilde{\psi}')'$.

Let $D_T(\gamma) := \partial \ell(\gamma) / \partial \gamma$ and $H_T(\gamma) := \partial^2 \ell(\gamma) / \partial \gamma \partial \gamma'$ denote the score vector and Hessian matrix, respectively, of the likelihood. We will consider the following one-sided score statistic,[3]

$$S_{1T} := D_T(\tilde{\gamma})_1 \sqrt{-H_T^{-1}(\tilde{\gamma})_{11}}, \tag{3.5}$$

as well as its square, which is the more traditional LM test statistic,

$$S_{2T} := -D_T(\tilde{\gamma})' H_T^{-1}(\tilde{\gamma}) D_T(\tilde{\gamma}). \tag{3.6}$$

Under the null hypothesis (3.1) and homoskedasticity, as in Assumption $\mathcal{H}$, the tests statistics (3.5) and (3.6) are asymptotically $N(0,1)$ and χ_1^2 distributed, respectively; see, for example, Robinson (1994), Tanaka (1999), or Nielsen (2004). The former result motivates the use of (3.5) as a test of (3.1) against one-sided alternatives, where the null would be rejected in favor of the right-tailed alternative if $S_{1T} > \mathcal{Z}_{1-\alpha}$, where $\mathcal{Z}_{1-\alpha}$ is such that $P(Z > \mathcal{Z}_{1-\alpha}) = \alpha$ when $Z \sim N(0,1)$, and vice versa for the left-tailed test; see Robinson (1994, p. 1424) for details. This allows the testing of interesting one-sided hypotheses such as testing $d = 1/2$ against either $d < 1/2$ or against $d > 1/2$, or testing the unit root $d = 1$ against $d < 1$, or even $d = 2$ against $d < 2$ to check whether y_t is I(2). Of course one-sided tests will be more powerful than two-sided tests (in the correct tail).

We now turn to detailing the asymptotic behaviour of the statistics (3.5) and (3.6) under unconditional and/or conditional homoskedasticity of the form given in Assumption $\mathcal{V}$. To do so, we introduce the parameter ω^2 which derives from the weak dependence present in the shocks. In the simplest case, where $p = q = 0$, $\omega^2 = (\pi^2/6)^{-1}$. In order to obtain a general expression for cases where $(p,q) \neq (0,0)$, first define $\xi(z,\gamma) := \partial \log\left((1-z)^d c(z,\psi)\right) / \partial \gamma$ and $\xi(z) := \xi(z,\gamma_0) =: \sum_{j=1}^{\infty} \xi_j z^j$. Observe in this expression that $\xi_j = (-j^{-1}, c_j')'$, where c_j is defined as the coefficient on z^j in the expansion of $\partial \log c(z,\psi) / \partial \psi|_{\psi=\psi_0}$ in powers of z. Under Assumption $\mathcal{R}$, it holds that c_j decays exponentially. Further define

$$\Xi := \sum_{j=1}^{\infty} \xi_j \xi_j' = \begin{bmatrix} \pi^2/6 & \kappa' \\ \kappa & \Phi \end{bmatrix} \tag{3.7}$$

with $\kappa := -\sum_{j=1}^{\infty} j^{-1} c_j$ and $\Phi := \sum_{j=1}^{\infty} c_j c_j'$; notice that, under Assumption $\mathcal{R}$, the matrix Ξ is finite and positive definite. With these definitions, $\omega^2 := (\Xi^{-1})_{1,1} = (\pi^2/6 - \kappa' \Phi^{-1} \kappa)^{-1}$. It is easily shown that Φ is the Fisher information for ψ; for example, if $\{u_t\}$ is an AR(1) process with coefficient a then $c_j = -a^{j-1}$ and $\Phi = (1-a^2)^{-1}$.

Where conditional heteroskedasticity is present in $\{z_t\}$ we also need to define the quantity[4] $\Upsilon := \sum_{j,k=1}^{\infty} \xi_j \xi_k' \tau_{j,k}$, where the $\tau_{j,k}$ coefficients derive from the higher-order dependence in the shocks induced by the conditional heteroskedasticity; see part (b) of Assumption $\mathcal{V}$. In such cases, the relevant quantity is now given by $\varpi^2 := (\Xi^{-1} \Upsilon \Xi^{-1})_{1,1}$. In an estimation setting, ω^2 and $\lambda \varpi^2$ would be the (asymptotic) variances of the QML of $\hat{d}$ under Assumptions $\mathcal{H}$ and $\mathcal{V}$, respectively. If $\{z_t\}$ is conditionally homoskedastic, then $\tau_{j,k} = \mathbb{I}(j=k)$ such that $\Upsilon = \sum_{j=1}^{\infty} \xi_j \xi_j' = \Xi$, and, hence, $\varpi^2 = \omega^2$.

To investigate the impact of heteroskedasticity on both the asymptotic size and local power of the tests we will derive asymptotic distributions under the relevant (local) Pitman drift alternative; i.e.,

$$H_{1,T} : d_0 = d_{0T} := \bar{d} + \delta/\sqrt{T}, \tag{3.8}$$

where δ is a fixed scalar. Notice that for $\delta = 0$, $H_{1,T}$ reduces to H_0 of (3.1).

Theorem 1 *Let Assumptions $\mathcal{R}$ and $\mathcal{V}$ be satisfied and assume that $\psi_0 \in \operatorname{int}(\Psi)$. Then, under $H_{1,T}$ of (3.8), it holds that*

$$S_{1T} \xrightarrow{w} (\lambda \frac{\varpi^2}{\omega^2})^{1/2} N(\delta \varpi^{-1} \lambda^{-1/2}, 1) \tag{3.9}$$

$$S_{2T} \xrightarrow{w} (\lambda \frac{\varpi^2}{\omega^2}) \chi_1^2\left(\delta^2 \varpi^{-2} \lambda^{-1}\right), \tag{3.10}$$

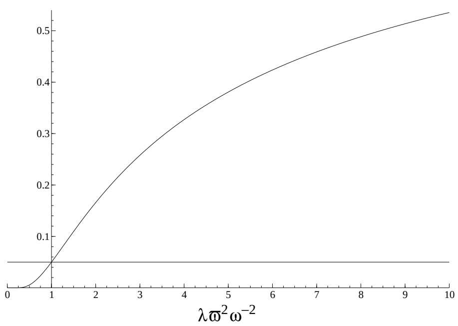
where $\lambda := (\int_0^1 \sigma^4(s) ds) / (\int_0^1 \sigma^2(s) ds)^2$.

[3] Note that $-H_T^{-1}(\tilde{\gamma})_{11}$ is not guaranteed to be positive in finite samples. To circumvent this issue, $-H_T(\tilde{\gamma})$ could be replaced by either an outer-product-of-gradients estimator or a positive definite estimator of its asymptotic limit Ξ_0 given in (3.7), although we prefer the computationally simpler version given here.

[4] Note that Assumptions $\mathcal{R}$ and $\mathcal{V}$ imply that Υ is finite. This follows because $||\xi_j|| \leq K j^{-1}$ under Assumption $\mathcal{R}$, and using condition (b)(iii) of Assumption $\mathcal{V}$ we thus find $||\Upsilon|| \leq \sum_{j,k=1}^{\infty} ||\xi_j|| ||\xi_k|| |\tau_{j,k}| \leq K \sum_{j,k=1}^{\infty} j^{-1} k^{-1} |\tau_{j,k}| < \infty$.

Figure 1. Size and power under heteroskedasticity

(a) Asymptotic size of S_{2T} at 5% nominal level (b) Size-corrected asymptotic local power of S_{2T}



Corollary 1 *Let Assumptions $\mathcal{R}$ and $\mathcal{V}$ be satisfied and assume that $\psi_0 \in \text{int}(\Psi)$. Then, under the null hypothesis H_0 of (3.1), $S_{1T} \xrightarrow{w} (\lambda\frac{\varpi^2}{\omega^2})^{1/2}N(0,1)$ and $S_{2T} \xrightarrow{w} (\lambda\frac{\varpi^2}{\omega^2})\chi_1^2$.*

Theorem 1 and Corollary 1 contain three key results. For concreteness, this discussion is based on the two-sided LM test, but similar remarks can be made about the one-sided score test.

First, if the shocks are (conditionally) homoskedastic then $\lambda = 1$ and $\frac{\varpi^2}{\omega^2} = 1$ and the standard results are special cases of the representations in (3.9) and (3.10). Specifically, under (3.8) and Assumption $\mathcal{H}$ it follows from Robinson (1994) that the S_{2T} test statistic is asymptotically non-central χ_1^2 distributed with non-centrality parameter $\delta^2\omega^{-2}$; that is, $S_{2T} \xrightarrow{w} \chi_1^2(\delta^2\omega^{-2})$.

Second, under the null, $\delta = 0$, the asymptotic distribution of S_{2T} is $(\lambda\frac{\varpi^2}{\omega^2})\chi_1^2$. When the factor $\lambda\frac{\varpi^2}{\omega^2} > 1$, for example if $\lambda > 1$ and $\frac{\varpi^2}{\omega^2} = 1$, it therefore follows that the LM test will over-reject asymptotically. Notice also therefore that a necessary (but not sufficient) condition for under-rejection to occur in the limit is for conditional heteroskedasticity to be present in $\{z_t\}$. Specifically, the LM test rejects if $S_{2T} > \chi^2_{1,1-\alpha}$, where $\chi^2_{1,1-\alpha}$ is such that $P(\chi_1^2 > \chi^2_{1,1-\alpha}) = \alpha$. Thus, rejection occurs with probability converging to

$$P((\lambda\frac{\varpi^2}{\omega^2})\chi_1^2 > \chi^2_{1,1-\alpha}) = P(\chi_1^2 > \chi^2_{1,1-\alpha}/(\lambda\frac{\varpi^2}{\omega^2})) = 1 - F_1(\chi^2_{1,1-\alpha}/(\lambda\frac{\varpi^2}{\omega^2})), \tag{3.11}$$

where $F_1(\cdot)$ denotes the cumulative density function [cdf] of the (central) χ_1^2 distribution. To illustrate this phenomenon, the asymptotic size of the LM test under heteroskedasticity is shown in Figure 1(a) as a function of the factor $\lambda\frac{\varpi^2}{\omega^2}$.

Third, under local alternatives the non-centrality parameter is scaled by $(\lambda\frac{\varpi^2}{\omega^2})^{-1}$, compared to the homoskedastic case, and the entire asymptotic distribution of S_{2T} is scaled by $\lambda\frac{\varpi^2}{\omega^2}$. The size-corrected LM test rejects when $S_{2T} > (\lambda\frac{\varpi^2}{\omega^2})\chi^2_{1,1-\alpha}$ such that size-corrected asymptotic local power is given by

$$P(\chi_1^2(\delta^2\varpi^{-2}\lambda^{-1}) > \chi^2_{1,1-\alpha}) = 1 - F_1(\chi^2_{1,1-\alpha}, \delta^2\varpi^{-2}\lambda^{-1}) = 1 - F_1(\chi^2_{1,1-\alpha}, \delta^2\omega^{-2}/(\lambda\frac{\varpi^2}{\omega^2})), \tag{3.12}$$

where $F_1(\cdot, c)$ is the cdf of the non-central χ_1^2 distribution with non-centrality parameter c. An implication of this is that the size-corrected asymptotic local power function of the S_{2T} test will be monotonically decreasing in $\lambda\frac{\varpi^2}{\omega^2}$ for a given value of δ. The size-corrected asymptotic local power of S_{2T} for various choices of δ is illustrated in Figure 1(b) (the figure is displayed with $\omega^2 = (\pi^2/6)^{-1}$).

The results in Theorem 1 therefore establish that the standard tests (obtained under the assumption of homoskedasticity) are not asymptotically correctly sized under heteroskedasticity of the form given in Assumption $\mathcal{V}$ and that these tests will also have asymptotic local power properties that depend on the degree of heteroskedasticity present in the process even when size-corrected. The finite sample effects of a variety of shock processes which display a one-time change in variance and/or a GARCH-type structure on the size and power properties of the LM test will be quantified by Monte Carlo simulation methods in section 5.

Remark 3.1 To quantify and interpret the scalar parameter λ, suppose that there is no conditional heteroskedasticity present in $\{z_t\}$, such that $\varpi^2 = \omega^2$. Here the right members of (3.9) and (3.10) simplify to $N(\delta\varpi^{-1}, \lambda)$ and $\lambda\chi_1^2\left(\delta^2\omega^{-2}\lambda^{-1}\right)$, respectively. These limits thus depend on λ, which is then a measure of the degree of unconditional heteroskedasticity present in the process $\{\varepsilon_t\}$. For a homoskedastic process, where $\sigma(\cdot)$ is constant, $\lambda = 1$, whereas when $\sigma(\cdot)$ is non-constant $\lambda > 1$ by the Cauchy-Schwarz inequality. For the single break in volatility example discussed in Remark 2.1 with $\sigma_0 = 1$ and $\sigma_1 = 3\,(\sigma_1 = 1/3)$ then: for $\tau = 0.25$ we find $\lambda = 1.245\,(2.333)$; for $\tau = 0.75$, $\lambda = 2.333\,(1.245)$, and for $\tau = 0.5$, $\lambda = 1.640$ in both cases. On the other hand, under constant unconditional volatility, $\lambda = 1$, the right members of (3.9) and (3.10) simplify to $(\frac{\varpi^2}{\omega^2})^{1/2}N(\delta\varpi^{-1}, 1)$ and $(\frac{\varpi^2}{\omega^2})\chi_1^2\left(\delta^2\varpi^{-2}\right)$, respectively, so that both the asymptotic size and local power functions of S_{1T} and S_{2T} depend on both ω^2 and ϖ^2.

Remark 3.2 In the Gaussian homoskedastic single-parameter model the one-sided test based on (3.5) is asymptotically uniformly most powerful (UMP), and the two-sided test based on (3.6) is asymptotically UMP unbiased, see Tanaka (1999) and Nielsen (2004) for the fractional model or Lehmann and Romano (2005) for a general treatment. □

4 Wild Bootstrap Inference

In this section we outline bootstrap-based analogues of the score and LM tests from section 3. We will demonstrate that the wild bootstrap implementations of these tests are asymptotically valid under heteroskedasticity of unknown form since they correctly replicate the large sample distributions of the test statistics.

We first outline our proposed wild bootstrap algorithm.

Algorithm 1 (wild bootstrap):

(i) Estimate model (2.1)-(2.2) under the null hypothesis (3.1) using Gaussian QML yielding the estimates $(\bar{d}, \tilde{\psi})$, see (3.4), together with the corresponding residuals, $\tilde{\varepsilon}_t := \hat{\varepsilon}_t(\bar{d}, \tilde{\psi})$.

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

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

$$y_t^* = \Delta_+^{-\bar{d}} u_t^*, u_t^* = c(L, \tilde{\psi})^{-1}\varepsilon_t^*, \quad t = 1, ..., T, \tag{4.1}$$

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

(iv) Using the bootstrap sample, $\{y_t^*\}$, compute the bootstrap test statistic S_{iT}^*, denoting either the score statistic ($i = 1$) or the LM statistic ($i = 2$), as detailed in section 3. If S_{iT}^* is the score test statistic for a left-tailed test, define the corresponding p-value as $P_T^* := G_{iT}^*(S_{iT})$, and if S_{iT}^* is the score test statistic for a right-tailed test or the LM test statistic, define the corresponding p-value as $P_T^* := 1 - G_{iT}^*(S_{iT})$. In either case, $G_{iT}^*(\cdot)$ denotes the conditional (on the original data) cdf of S_{iT}^*.

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

Remark 4.1 For stationary data, it is often seen in the wild bootstrap literature (for a review, see Davidson and Flachaire, 2008) that improved bootstrap accuracy can be achieved by generating the pseudo-data according to an asymmetric distribution with $E(w_t) = 0$, $E(w_t^2) = 1$ and $E(w_t^3) = 1$. A well-known example of this is the Mammen (1993) distribution: $P(w_t = -0.5(\sqrt{5} - 1)) = 0.5(\sqrt{5} + 1)/\sqrt{5} =: \pi$, $P(w_t = 0.5(\sqrt{5} + 1)) = 1 - \pi$. Two other commonly used (symmetric) distributions are the simple two-point distribution $P(w_t = -1) = P(w_t = 1) = 0.5$ and an i.i.d. $N(0, 1)$ sequence. The large sample properties of the resulting bootstrap tests are the same for all three. In our simulations we found that, of these three, the simple two-point distribution gave slightly better small sample performance than the other two, and so we will present results only for this choice of w_t.

Remark 4.2 In step (i) of Algorithm 1 the parameters characterizing (2.1), which are then used in constructing the bootstrap sample data in steps (ii) and (iii), are estimated under the restriction of the null hypothesis, H_0 of (3.1). It is also possible to estimate these parameters unrestrictedly in step (i) of Algorithm 1, and use these unrestricted estimates in constructing the bootstrap sample data in steps (ii) and (iii). Let , $\hat{d}$ be the unrestricted estimate of d obtained from the original sample data. Since the bootstrap DGP is now based on $\hat{d}$, the bootstrap test statistic computed in step (iv) should be for the hypothesis that $d = \hat{d}$. A finite sample comparison of these two possible approaches is conducted in section 5, where it is shown that the bootstrap based on restricted estimates is preferred.

Remark 4.3 In practice, the cdf $G_T^*(\cdot)$ required in step (iv) of Algorithm 1 will be unknown, but can be approximated in the usual way through numerical simulation. This is achieved by generating B (conditionally) independent bootstrap statistics, $S_{iT:b}^*$, $i = 1, 2$, for $b = 1, ..., B$, computed as in Algorithm 1 above. The simulated bootstrap p-value for S_{2T}, for example, is then computed as $\tilde{P}_T^* := B^{-1} \sum_{b=1}^{B} \mathbb{I}(S_{2T:b}^* > S_{2T})$, and is such that $\tilde{P}_T^* \overset{a.s.}{\to} P_T^*$ as $B \to \infty$. The choice of B is discussed by, *inter alia*, Davidson and MacKinnon (2000). □

In Theorem 2 and Corollary 2, we now provide results which establish the asymptotic validity of our proposed wild bootstrap fractional integration tests. For these results to hold we need to strengthen part (ii) of Assumption $\mathcal{V}$(b) as follows:

Assumption $\mathcal{V}'$ *Assumption $\mathcal{V}$ holds with* (ii) *replaced by:*

(ii') $\tau_{r,s} := E(z_t^2 z_{t-r} z_{t-s})$ *is uniformly bounded for all* $t \geq 1, r \geq 0, s \geq 0$, *where* $\tau_{r,r} > 0$ *for all* $r \geq 0$ *and* $\tau_{r,s} = 0$ *for* $r \neq s$.

Remark 4.4 Assumption $\mathcal{V}'$ imposes the additional condition that $\tau_{r,s} = 0$ for $r \neq s$. However, Assumption $\mathcal{V}'$ is still slightly weaker than the corresponding conditions imposed in Robinson (1991), Demetrescu, Kuzin and Hassler (2008), Hassler, Rodrigues and Rubia (2009) and Kew and Harris (2009), see the remarks after Assumption $\mathcal{V}$. □

Theorem 2 *Let Assumptions $\mathcal{R}$ and $\mathcal{V}'$ hold. Then under (3.8) it holds that* $S_{1T}^* \overset{w}{\to}_p (\lambda \frac{\varpi^2}{\omega^2})^{1/2} N(0, 1)$ *and* $S_{2T}^* \overset{w}{\to}_p (\lambda \frac{\varpi^2}{\omega^2}) \chi_1^2$.

Theorem 2 has the following corollary, where P_T^* denotes the (wild bootstrap) p-value associated with any of the test statistics considered.

Corollary 2 *Let the conditions of Theorem 2 be satisfied. Under the null hypothesis (3.1),* $P_T^* \overset{w}{\to} U[0, 1]$, *i.e. a uniform distribution on* $[0, 1]$.

An immediate implication of the result in Corollary 2 is that the wild bootstrap implementations of the one-sided score and two-sided LM tests will have correct asymptotic size in the presence of both unconditional and conditional heteroskedasticity of the form given in Assumption $\mathcal{V}'$. Notice that these results are trivially also seen to be true under conditional homoskedasticity since that special case is contained within both Assumptions $\mathcal{V}$ and $\mathcal{V}'$. Moreover, the results in Theorem 2 also imply immediately that, under Assumption $\mathcal{V}'$, the wild bootstrap tests will attain the same asymptotic local power function as the size-adjusted asymptotic tests; cf. Theorem 1.

Remark 4.5 In the working paper version, Cavaliere, Nielsen, and Taylor (2013), we also analyzed i.i.d. bootstrap implementations of the tests. As would be expected, the i.i.d. bootstrap is not able to account for heteroskedasticity, and therefore suffers similar size distortions to the asymptotic tests. □

5 Monte Carlo Simulations

In this section we use Monte Carlo simulation methods to compare the finite sample size and power properties of the asymptotic tests and their bootstrap implementations described above (either imposing the null in step (i) of Algorithm 1, or not as in Remark 4.2), for both homoskedastic and heteroskedastic shocks. To conserve space in the tables, we present results only for the two-sided LM statistic, S_{2T} (results for the one-sided score test statistics are qualitatively similar). Comparison is also made with Kew and Harris' (2009) White-corrected analogue of the LM test of Agiakloglou and Newbold (1994) and Breitung and Hassler (2002). Of the tests considered in Kew and Harris (2009), this seems the most natural to compare with. This test will be referred to as the KH test in what follows. With the exception of the results reported in Table 4, the tests are calculated assuming knowledge of the autoregressive and moving average lag orders, p and q respectively, in (2.2).

5.1 Monte Carlo Setup

The Monte Carlo data are simulated from the model (2.1) with ARMA$(1,1)$ shocks $(1-aL)u_t = (1+bL)\varepsilon_t$, where $\varepsilon_t = \sigma_t z_t$ and σ_t, z_t are defined in the subsections below. The S_{2T} test is invariant to $\bar{d}$, but for the KH test we set $\bar{d} = 1$. Data is then generated with $d_0 = 1 + \delta/\sqrt{T}$, see (3.1) and (3.8).

We report results for $T = 100$ and $T = 250$, and under $T = \infty$ we also report the asymptotic size (for $\delta = 0$) or size-corrected local power (for $\delta \neq 0$) calculated from (3.11) and (3.12). Note that the simulated finite sample powers of the asymptotic S_{2T} test and of the KH test were both size-corrected, while those for the wild bootstrap test were not. All tests were computed at 5% nominal size. The S_{2T} test statistic in (3.6) was implemented using numerical derivatives. For the wild bootstrap, we used $B = 499$ bootstrap replications and the i.i.d. sequence $\{w_t\}$ was chosen such that $P(w_t = -1) = P(w_t = 1) = 0.5$. All simulation results were programmed in Ox version 6.3, see Doornik (2007), using $10,000$ Monte Carlo replications. Our programs are available on request.

5.2 Results With Unconditionally Heteroskedastic, Uncorrelated Shocks

We first consider the case where the shocks do not display either weak dependence or conditional heteroskedasticity so as to analyse the impact of unconditional heteroskedasticity alone on the tests. Accordingly, we set $a = b = 0$ and simulate $\{z_t\}$ as an i.i.d. $N(0,1)$ sequence.

The unconditional variance is generated according to the following one-shift volatility process, $\sigma_t^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)I(t \geq \tau T)$, that is, an abrupt single shift in the variance from σ_0^2 to σ_1^2 at time τT, for some $\tau \in (0,1)$. Without loss of generality we normalize $\sigma_0^2 = 1$. We consider $\tau \in \{1/4, 3/4\}$ and vary the ratio $\theta := \sigma_1/\sigma_0$ among $\theta \in \{1/3, 1, 3\}$. Note that $\theta = 1$ corresponds to homoskedastic shocks, in which case τ is irrelevant. These values of τ and θ are motivated by the so-called "great moderation" and the recent financial crisis, as mentioned in the Introduction, suggesting a decline in the volatility early in the sample and an increase in the volatility late in the sample, respectively.

The results for the case with conditionally homoskedastic $\{z_t\}$ are given in Table 1. Even in the homoskedastic case (the rows relating to $\theta = 1$ in Table 1), a comparison between the results for the asymptotic S_{2T} and KH tests and the corresponding results for the wild bootstrap implementations of the S_{2T} test shows that the bootstrap can deliver some improvements. For example, for $T = 100$ the empirical rejection frequencies of the asymptotic S_{2T} and KH tests are 5.87% and 5.89%, respectively, while that of the (null) wild bootstrap test is 4.95%.

It is where heteroskedasticity is present in the shocks (the rows where $\theta \neq 1$) that the wild bootstrap based tests clearly display their superiority over the asymptotic S_{2T} test. From the results in Table 1 we see that the asymptotic S_{2T} test can be severely over-sized with this phenomenon persisting as the sample size is increased, as predicted by the results in Theorem 1. Again as predicted by Theorem 1 the degree of over-sizing in the asymptotic S_{2T} test worsens as λ increases. For example, in the two cases where $\lambda = 2.333$ (see Remark 3.1 and column 4 in Table 1) the empirical rejection frequency of these tests is about 19% regardless of the sample size. The wild bootstrap test which imposes the null in estimating the parameters of (2.1) is clearly the best performing test in Table 1, and significantly outperforms the bootstrap test based on unrestricted estimates. The null bootstrap test displays excellent size control throughout; the largest entry relating to size for this test is a rejection frequency of 5.55% which occurs for $T = 100$ with $\tau = 0.75$ and $\theta = 3$. The asymptotic KH clearly also performs much better than the asymptotic S_{2T} test, but nonetheless it displays larger finite sample size distortions under heteroskedasticity throughout Table 1 than either of the two bootstrap tests.

Turning to the power of the tests, the results in Table 1 again show that the predictions from the asymptotic theory are strongly reflected in finite samples with the size-corrected empirical power of the asymptotic tests being lower the larger the value of λ, and that, as with the size results, these effects do not vanish as the sample size is increased. Indeed, the size-adjusted power of the tests

Table 1. Simulated size and power: one-time shift in unconditional volatility

				size				power			
τ	θ	T	λ	S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	KH	S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	KH
	1	100	1	5.87	4.95	5.09	5.89	40.58	39.62	41.41	45.64
	1	250	1	5.16	4.73	4.93	5.52	44.97	43.85	44.31	46.63
	1	∞	1	5.00	5.00	5.00	5.00	48.56	48.56	48.56	48.56
1/4	1/3	100	2.333	17.78	5.28	6.73	6.86	19.46	21.06	23.29	28.11
1/4	1/3	250	2.333	18.71	4.60	5.32	5.85	22.85	21.66	22.67	25.76
1/4	1/3	∞	2.333	19.95	5.00	5.00	5.00	24.24	24.24	24.24	24.24
1/4	3	100	1.245	9.37	5.38	5.74	6.18	30.63	31.69	32.68	38.12
1/4	3	250	1.245	8.17	5.03	5.17	5.56	36.36	36.34	36.92	39.69
1/4	3	∞	1.245	7.90	5.00	5.00	5.00	40.69	40.69	40.69	40.69
3/4	1/3	100	1.245	8.36	5.21	5.66	6.05	33.50	32.98	34.33	40.80
3/4	1/3	250	1.245	8.43	5.24	5.42	5.57	34.41	35.18	35.70	40.95
3/4	1/3	∞	1.245	7.90	5.00	5.00	5.00	40.69	40.69	40.69	40.69
3/4	3	100	2.333	19.78	5.55	6.91	7.42	16.92	18.14	20.45	21.60
3/4	3	250	2.333	19.02	5.24	5.86	5.88	19.41	19.60	20.97	23.01
3/4	3	∞	2.333	19.95	5.00	5.00	5.00	24.24	24.24	24.24	24.24

Notes: S_{2T} denotes the asymptotic LM test, $S_{2T}^{*,0}$ denotes the bootstrap LM test of Algorithm 1, and $S_{2T}^{*,a}$ denotes the alternative bootstrap test outlined in Remark 4.2, and KH denotes the asymptotic test of Kew and Harris (2009). Entries for finite T are simulated rejection frequencies of the tests. Entries for $T = \infty$ are calculated from (3.11) and (3.12). Power is measured at $\delta = 1.5$ and is size corrected for the asymptotic and KH tests, but not size corrected for the bootstrap tests. The bootstrap procedures are based on $B = 499$ bootstrap replications and all entries are based on $10,000$ Monte Carlo replications.

can be significantly lower; for example, when $\lambda = 1$ all of the tests display an empirical rejection frequency of 40-50%, but for $\lambda = 2.333$ (size-corrected) power is roughly half this level. A notable feature of the power results for the wild bootstrap test calculated under the null is how close these results are to the size-adjusted power results for the asymptotic S_{2T} test. This is of course predicted by the large sample distribution theory in sections 3 and 4, but it is interesting to see how closely the finite sample results adhere to this prediction. Interestingly, even though, as noted above, the unrestricted wild bootstrap yields a test with, in general, more liberal finite sample size properties than the corresponding restricted wild bootstrap test, it is seen from Table 1 that the power of the two bootstrap tests differ only very slightly, suggesting that the improved finite sample size control of the restricted bootstrap does not come at the cost of reduced power, and hence that the restricted version should be preferred. Interestingly, the asymptotic KH test displays higher finite sample size-adjusted power than any of the other tests throughout the experiments reported in Table 1, suggesting a useful complementarity between the asymptotic KH test and the restricted bootstrap S_{2T} test.

5.3 Results With Conditionally Heteroskedastic, Uncorrelated Shocks

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

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

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

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

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

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

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

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

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

Table 2. Simulated size and power: conditionally heteroskedastic Models A-I

	T	size				power			
		S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	KH	S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	KH
Model A	100	15.77	5.30	6.75	6.55	20.81	26.47	27.74	41.14
	250	17.33	5.35	6.38	5.70	21.09	25.20	25.96	40.85
Model B	100	17.87	5.17	7.40	6.28	17.50	24.20	26.74	43.10
	250	22.03	5.14	7.01	5.72	13.68	22.02	24.04	42.52
Model C	100	11.31	4.93	5.59	6.61	26.85	29.43	30.64	34.19
	250	15.68	4.87	5.53	5.93	21.27	25.38	26.31	29.62
Model D	100	13.51	5.18	6.35	6.11	24.22	28.33	30.09	39.84
	250	17.89	4.99	5.70	5.85	18.02	24.45	25.88	38.42
Model E	100	16.93	5.39	6.78	6.97	18.64	22.96	24.31	29.73
	250	21.55	5.35	6.57	6.11	15.71	20.75	22.07	25.23
Model F	100	15.85	4.96	6.06	6.86	20.41	23.69	24.72	29.71
	250	22.92	5.10	6.19	6.24	14.49	19.58	20.67	23.52
Model G	100	15.12	4.88	5.89	6.36	19.59	22.88	24.07	30.09
	250	21.33	5.05	5.96	5.71	13.30	18.99	20.09	22.92
Model H	100	28.41	5.14	7.71	7.42	9.54	15.27	19.16	31.79
	250	38.90	5.01	7.78	5.94	6.63	11.19	14.66	28.61
Model I	100	28.06	5.52	7.88	7.91	11.39	15.34	17.63	20.16
	250	31.91	5.36	7.22	6.11	10.44	14.06	15.93	21.47

Notes: See Notes for Table 1.

$$\text{Model I } : \varepsilon_t = \sigma_t z_t, \sigma_t = 1 + 2I(t \geq \frac{3}{4}T), z_t = h_t^{1/2} e_t, h_t = 0.1 + 0.5 z_{t-1}^2, e_t \sim N(0,1).$$

The conditionally heteroskedastic configurations for $\{z_t\}$ specified in Models A-H are a subset of those used in Section 4 of Gonçalves and Kilian (2004). Models A-D are standard stationary GARCH$(1,1)$ models driven by either Gaussian or t-distributed shocks with unit variance, while Model E is the is the exponential GARCH$(1,1)$ [EGARCH$(1,1)$] model of Nelson (1991). Model F is the asymmetric GARCH$(1,1)$ [AGARCH$(1,1)$] model of Engle (1990), Model G is the GJR-GARCH$(1,1)$ model of Glosten, Jaganathan and Runkle (1993), and Model H is a first-order autoregressive stochastic volatility model. The chosen parameter values in Models A-H are based on applied work, see Section 4 of Gonçalves and Kilian (2004), where the relation between these models and the moment conditions in Assumptions $\mathcal{V}$ and $\mathcal{V}'$ is also discussed. In any case, it is of interest to investigate the finite sample behavior of the tests under models that may not in fact satisfy the assumptions needed for the asymptotic theory. Finally, Model I combines conditional heteroskedasticity in $\{z_t\}$, of the form specified by Model A, together with the one-time change model for the unconditional variance considered in the previous subsection (for the particular case of $\theta = 3$ and $\tau = 0.75$). The results relating to Models A-I are presented in Table 2.

Consider first the results in Table 2 for the empirical size of the asymptotic S_{2T} test. Here we see that for these commonly encountered models of conditional heteroskedasticity the asymptotic S_{2T} test can be very badly over-sized; indeed, the degree of over-sizing is, if anything, more pronounced than was observed in this test for the models of unconditional heteroskedasticity in Table 1. While it was seen in Table 1 that the degree of size distortions under the single break model depends on both the change-point location and the magnitude of the break (with these distortions being relatively moderate for increases in variance early in the sample and decreases late in the sample), there are no entries for size of the asymptotic S_{2T} test in Table 2 that lie below 11%. Models H and I clearly effect the greatest degree of over-sizing, with the empirical size under Model H approaching 40% for $T = 250$. Consistent with the results in Theorem 1, it is observed that these size distortions do not disappear as the sample size is increased; indeed, the opposite phenomenon occurs. Turning to the results for the two wild bootstrap tests and the KH test in Table 2 we see, as with the case of unconditional heteroskedasticity in Table 1, that all of these again do a decent job of controlling finite sample size under all of Models A-I. As with the results in Table 1, the best performance among these is again clearly achieved by the restricted wild bootstrap (using step (i) of Algorithm 1); no empirical sizes are observed for the restricted wild bootstrap in Table 2 which are in excess of 5.39% or below 4.87%.

Table 3. Simulated size: weakly dependent shocks with known (p, q)

		homoskedastic case				$\tau = 1/4$ and $\theta = 1/3$				$\tau = 3/4$ and $\theta = 3$			
a/b	T	S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	KH	S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	KH	S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	KH
Panel A: moving average shocks ($a = 0$)													
-0.80	100	8.44	5.05	6.63	–	15.83	5.61	7.33	–	17.85	5.62	8.08	–
-0.80	250	6.83	5.00	5.26	–	18.13	5.51	6.35	–	19.88	5.28	6.51	–
-0.80	∞	5.00	5.00	5.00	–	19.95	5.00	5.00	–	19.95	5.00	5.00	–
0.80	100	8.71	5.38	6.63	–	15.05	5.44	6.99	–	17.88	5.72	8.13	–
0.80	250	6.89	5.24	5.49	–	18.34	5.13	6.07	–	20.67	5.02	6.35	–
0.80	∞	5.00	5.00	5.00	–	19.95	5.00	5.00	–	19.95	5.00	5.00	–
Panel B: autoregressive shocks ($b = 0$)													
-0.80	100	6.67	5.08	5.39	6.09	18.53	5.43	6.74	7.53	19.80	5.41	6.52	7.82
-0.80	250	6.03	5.37	5.50	5.87	20.07	5.36	5.84	6.19	19.53	5.08	5.75	6.50
-0.80	∞	5.00	5.00	5.00	5.00	19.95	5.00	5.00	5.00	19.95	5.00	5.00	5.00
0.80	100	4.28	4.76	4.15	6.01	6.73	4.91	5.03	8.53	6.84	4.50	4.69	7.89
0.80	250	6.28	4.95	4.91	5.15	11.07	4.94	4.67	6.73	10.82	4.78	4.06	6.07
0.80	∞	5.00	5.00	5.00	5.00	19.95	5.00	5.00	5.00	19.95	5.00	5.00	5.00

Notes: See Notes for Table 1.

The unrestricted wild bootstrap and KH tests are somewhat more liberal, both displaying empirical sizes approaching 8% in a number of cases.

Consider next the power results for the tests. As with the results in Table 1, we see from the results in Table 2 that the size-corrected empirical power of the asymptotic S_{2T} test is very strongly affected by the presence of conditional heteroskedasticity in each of Models A-I, which is expected from Theorem 1. In line with the empirical size results reported in the table we again see that this is most pronounced for Models H and I and that these effects do not vanish (indeed they again become more pronounced) as the sample size is increased. Again it is seen that the size-adjusted power of the asymptotic S_{2T} test can be significantly lower than in the homoskedastic case; for example, under Model H the size-corrected power is barely above the nominal level. The empirical power of the restricted wild bootstrap test now lies above the size-adjusted power results for the asymptotic S_{2T} test. Again there are only very slight differences between the power of the restricted and unrestricted wild bootstrap tests. The size-corrected power of the KH test is also clearly affected by conditional heteroskedasticity but to a lesser extent than the other tests reported. Again, as with the results in Table 1, the KH test displays the highest size-adjusted power of all the tests throughout.

5.4 Results With Weakly Dependent Shocks

We now consider the results in Table 3 which relate to the finite sample size properties of the asymptotic and bootstrap tests for processes driven by shocks which can display both weak dependence and unconditional heteroskedasticity of the type considered in Table 1. We do not report results for the KH test when an MA component is present because their test procedure assumes a finite-order AR for the shocks. Consider first the results for the homoskedastic case, $\lambda = 1$, presented in the first block of columns in Table 3. These results demonstrate that the asymptotic S_{2T} test has the potential for really quite poor finite sample size control under weak dependence; most notably, over-sizing when an MA component is present. For example, for $b = 0.8$ and $T = 100$ the asymptotic S_{2T} test has empirical rejection frequency of 8.71%. In contrast, the wild bootstrap tests display very good size control throughout, particularly so where the restricted bootstrap is used; in the above example the corresponding restricted wild bootstrap test displays a rejection frequency of 5.38%. For the case of AR shocks the two bootstrap methods again display superior finite sample size control to the KH test.

Turning to the results for the two heteroskedastic cases reported in Table 3, the patterns of size distortions seen in the asymptotic S_{2T} test are very similar to those seen for these two cases in Table 1, with empirical sizes generally around 15-20%. This suggests that even in relatively small samples the impact of any heteroskedasticity in the shocks largely dominates the impact of any weak dependence present, at least for the two heteroskedastic cases reported here. In contrast, the wild bootstrap tests reported in Table 3 do an excellent job for all the reported combinations of heteroskedasticity and

Table 4. Simulated size: weakly dependent shocks with (p, q) chosen by BIC

a/b	T	homoskedastic case S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	$\tau = 1/4$ and $\theta = 1/3$ S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$	$\tau = 3/4$ and $\theta = 3$ S_{2T}	$S_{2T}^{*,0}$	$S_{2T}^{*,a}$
Panel A: moving average shocks ($a = 0$)										
-0.80	100	34.55	28.60	27.44	37.62	19.52	20.44	45.83	34.32	32.35
-0.80	250	14.14	8.79	8.48	25.95	14.14	13.21	29.78	21.74	19.99
-0.80	500	8.12	4.82	3.20	19.15	5.21	4.87	19.84	6.73	6.36
0.80	100	11.25	3.54	2.56	24.10	4.59	3.83	25.51	4.38	3.94
0.80	250	7.59	4.28	3.54	23.53	4.19	3.53	27.44	4.79	3.94
0.80	500	6.67	4.61	4.02	24.53	3.77	2.80	26.30	4.19	3.15
Panel B: autoregressive shocks ($b = 0$)										
-0.80	100	10.04	4.11	3.21	24.36	4.57	4.39	27.21	4.56	4.49
-0.80	250	7.52	4.33	4.24	24.57	4.13	3.54	27.79	4.67	3.97
-0.80	500	6.40	4.37	4.07	25.37	3.88	2.91	26.53	4.08	3.11
0.80	100	40.61	36.11	35.49	36.87	26.59	25.18	48.75	39.52	37.44
0.80	250	16.66	12.35	12.26	26.63	18.61	17.81	33.33	26.23	24.86
0.80	500	7.93	4.66	2.27	18.02	7.09	6.81	19.65	8.21	8.11

Notes: See Notes for Table 1.

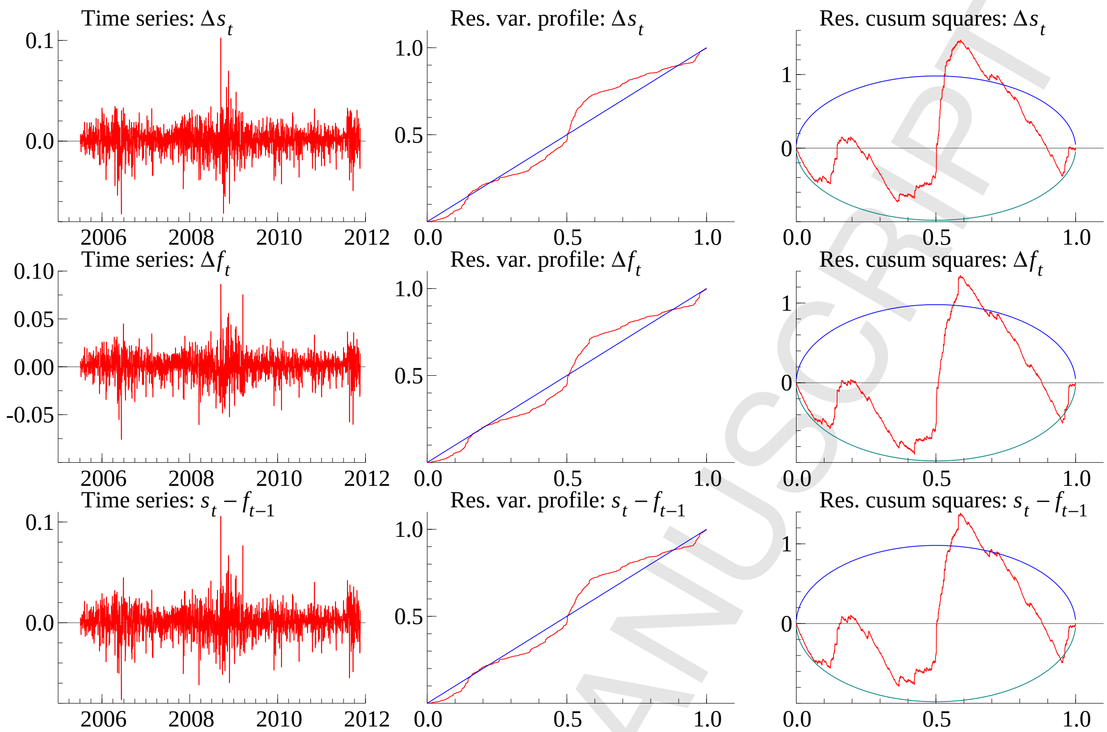
weak dependence; most of the empirical sizes reported for the restricted wild bootstrap test lie very close to the nominal level, with no entry in excess of 5.72% or below 4.50%. Slightly higher distortions on average are again seen with the unrestricted wild bootstrap test, confirming our previous recommendation to use the restricted version of the bootstrap. The KH test is again not as efficacious at controlling finite sample size as the two bootstrap tests. Moreover, it is seen that the KH test displays significantly poorer finite sample size control when both weak dependence and heteroskedasticity are present, *vis-à-vis* when only weak dependence is present.

Finally, in Table 4 we investigate the impact of data-based model selection on the finite sample size behaviour of the asymptotic S_{2T} test and our two bootstrap analogues in the case where p and q in (2.2) are not assumed known. Rather, as suggested in Remark 2.5, these are chosen (in both the original and bootstrap data) using the BIC and choosing across $p, q \in \{0, 1, 2\}$. Again we do not report results for the KH test because its finite-order AR set-up for the shocks means that we cannot make a like-for-like comparison of model selection with these tests. Regardless of whether heteroskedasticity is present or not, the most striking feature of these results is the large size distortions seen in all of the tests when either a large negative MA or positive AR coefficient is present. As might be anticipated this over-sizing is most pronounced for the asymptotic S_{2T} test, where it is also observed, albeit to a much lesser degree, for the positive MA and negative AR cases. Moreover, over-size is still seen for $T = 250$ even for the bootstrap tests and so we also report results for $T = 500$. By $T = 500$ the bootstrap tests are once again close to the nominal level in both the homoskedastic and heteroskedastic settings, while the asymptotic S_{2T} test is still oversized even in the homoskedastic case. Our results parallel those of Demetrescu, Kuzin and Hassler (2008) who also find considerable finite sample size distortions in the LM-type tests they propose and in the tests of Dolado, Gonzalo and Mayoral (2002) and Breitung and Hassler (2002), in each case when either AIC or BIC is used to select the lag length in an autoregressive approximating model for ARMA$(1, 1)$ shocks with $a = b = 0.5$.

6 Empirical Analysis

In this section we employ the asymptotic score-based tests and their bootstrap counterparts from sections 3 and 4 to re-assess the degree of support provided for (the weak form of) the EMH in a number of commodity markets. By adopting the heteroskedastic ARFIMA model of section 2, along with the novel (wild bootstrap) testing procedures outlined in section 4, we simultaneously allow for the possibility of both fractional integration and time-varying conditional and unconditional volatility in the data. This allows us to analyse the empirical validity or otherwise of the EMH in a more general and empirically well-grounded model framework than those which have previously been employed in the extant empirical literature.

Figure 2. Graphics for gold



Note: Left panels show time series plots of Δs_t, Δf_t, $s_t - f_{t-1}$, middle panels show the residual variance profiles, $\hat{\eta}(u)$, and right panels show the residual cusum of squares process with 95% confidence bands.

Our analysis is based on the data-set recently considered in Westerlund and Narayan (2013). This consists of (logged) spot prices (s_t) and corresponding one-period futures contract prices ($f_t := f_t^{(1)}$) of four commodities: gold, silver, platinum, and crude oil. Prices are recorded at the daily frequency (five observations per week) and cover the period July 5, 2005, to November 22, 2011. The number of available observations is $T = 1665$. All data were obtained from Bloomberg; see Westerlund and Narayan (2013) for full details and data definitions. Plots of s_t, f_t (in first differences) and of (minus) the forward premium (the spread) $s_t - f_{t-1}$ are reported in the left-hand panels of Figures 2-5.

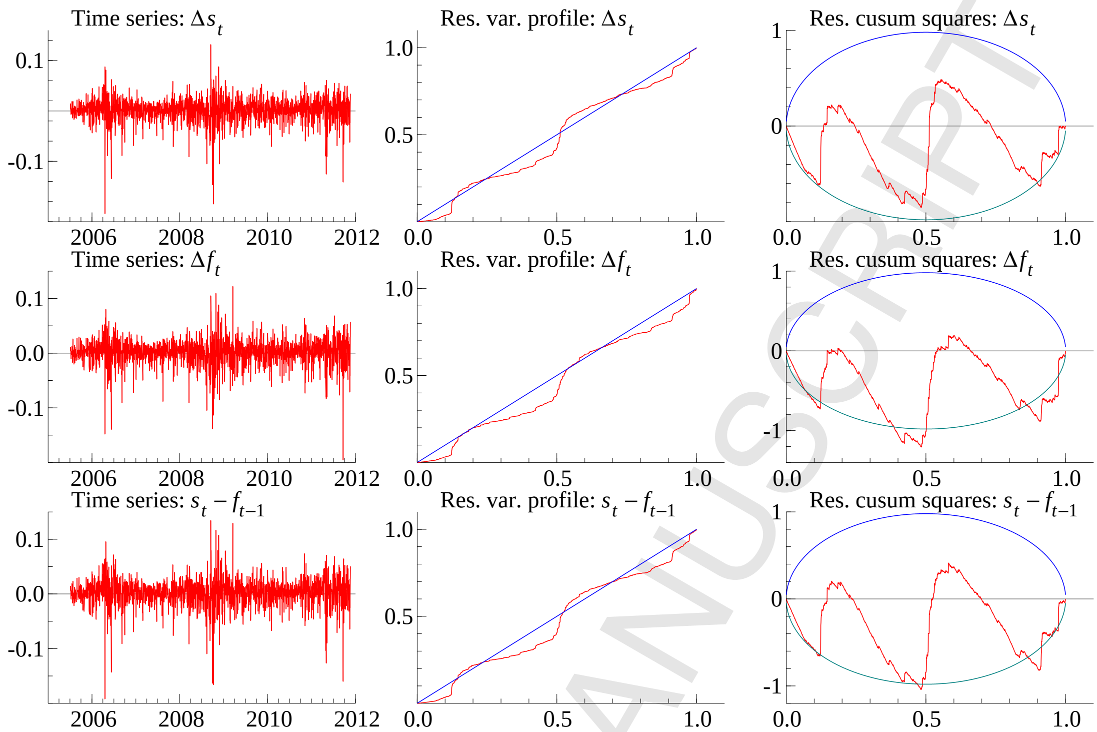
To investigate for the possible presence of heteroskedasticity in the series, we first report in the top panel of Table 5 results for the LM test of the null hypothesis of conditional homoskedasticity against the alternative of ARCH(k) dynamics. These tests are based on the squared residuals[5] of an ARFIMA(p, d, q) model fitted to each series (Δs_t, Δf_t and $s_t - f_{t-1}$) individually.[6] The AR and MA orders p and q for the ARFIMA model are selected using the BIC, while the number of ARCH lags k used for the LM test regression is set to either 5 (weekly frequency) or 21 (monthly frequency). For all commodities, conditional homoskedasticity is easily rejected at any conventional significance level for spot and futures prices and for the spread, $s_t - f_{t-1}$.

To visualise the possible presence of non-stationary volatility (unconditional heteroskedasticity) in the data, we plot in the central panels of Figures 2-5 the sample *variance profiles* of the residuals, say $\hat{\varepsilon}_t$, of the fitted ARFIMA models. The sample variance profiles, see Cavaliere and Taylor (2007), are plots of $\hat{\eta}(u) := \hat{V}_T(u)/\hat{V}_T(1)$ against $u \in [0, 1]$, where $\hat{V}_T(u) := T^{-1}\sum_{t=1}^{\lfloor Tu \rfloor} \hat{\varepsilon}_t^2$ denotes the cumulated

[5] Comparable results are obtained when the test statistics are computed on the original series rather than on the residuals.

[6] In all estimations and tests here and in the remainder of the empirical analysis, we allowed for a constant term in the model; see Remark 2.3, and in particular Robinson (1994) and Nielsen (2004). For all of the series considered, an additional linear trend term was found to be statistically insignificant at all conventional levels.

Figure 3. Graphics for silver



Note: Left panels show time series plots of Δs_t, Δf_t, $s_t - f_{t-1}$, middle panels show the residual variance profiles, $\hat{\eta}(u)$, and right panels show the residual cusum of squares process with 95% confidence bands.

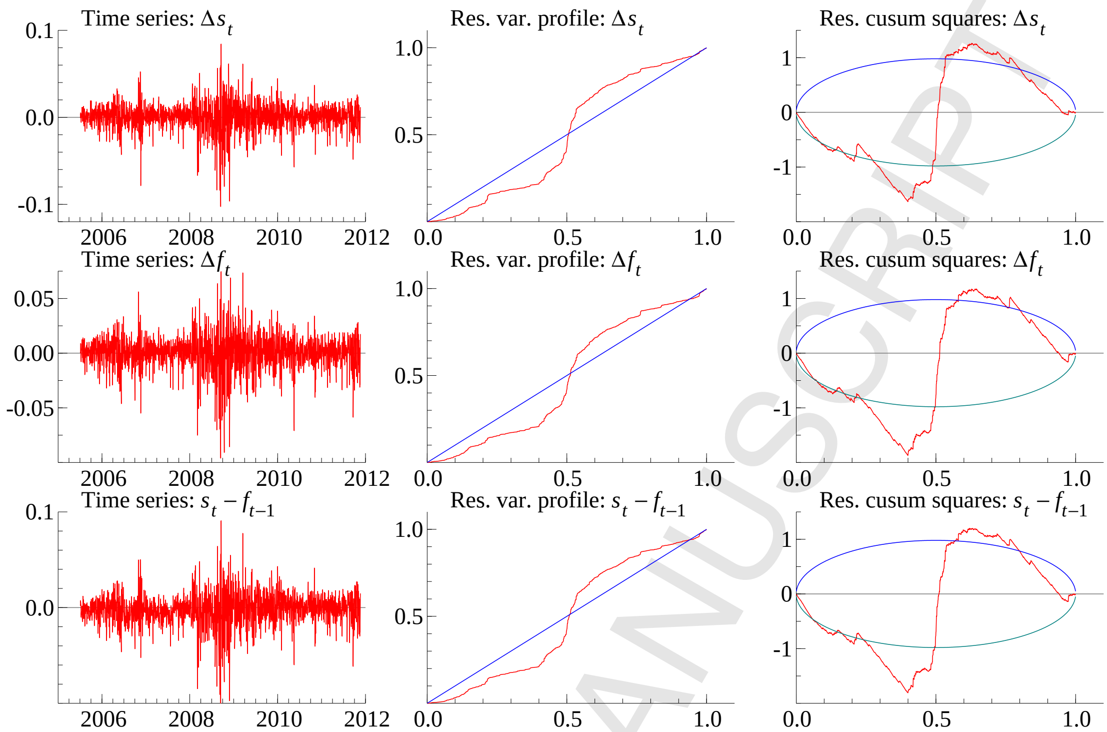
squared residuals. In large samples, $\hat{\eta}\,(u) \approx (\int_0^1 \sigma\,(s)\,ds)^{-1} \int_0^u \sigma\,(s)\,ds =: \eta\,(u)$, which equals u when the unconditional volatility is constant; that is, when there is no unconditional heteroskedasticity. Consequently, under conditional homoskedasticity or – more generally – under stationary conditional heteroskedasticity, $\hat{V}_T(u)$ 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[7], are reported in the right-hand panels of Figures 2-5. Correspondingly, in the lower panel of Table 5 we also report the associated stationary volatility tests of Cavaliere and Taylor (2008*b*, pp. 311–312). With the exception of silver, there is strong evidence of unconditional heteroskedasticity (non-stationary volatility) in all of the commodities. This evidence is manifested, and to similar extents, in both the spot and futures prices, as well as in the associated forward premium. Notice also that clear changes in the variance profile with associated significant values of the cumulated sum of squared residuals are apparent (even to some extent for silver) at around the time of the financial crisis, as might be expected. Given the strength of these rejections it is therefore quite striking that most empirical studies (including that of Westerlund and Narayan, 2013) are based on the maintained assumption of (un)conditional homoskedasticity.

We now turn to testing the main implications of the EMH; that is, conditions (i)–(iii) and (iv') discussed in section 1. As stated in condition (i), under the assumption that spot prices are I(1), futures prices should also be I(1). We test both claims in the first two columns of Table 6, where we present results for the LM test of the null hypothesis $H_0 : d = 0$ for Δs_t and Δf_t, respectively (note this is equivalent to testing $H_0 : d = 1$ in the levels). For each series, we report the (QML)

[7]The confidence bands are obtained as suggested by Cavaliere and Taylor (2008*b*) and Cheng and Phillips (2012). This requires estimation of the long-run variance of u_t^2 under the null hypothesis, which is done here using a sums-of-covariances estimator with the Bartlett kernel and a lag truncation of five.

Figure 4. Graphics for platinum



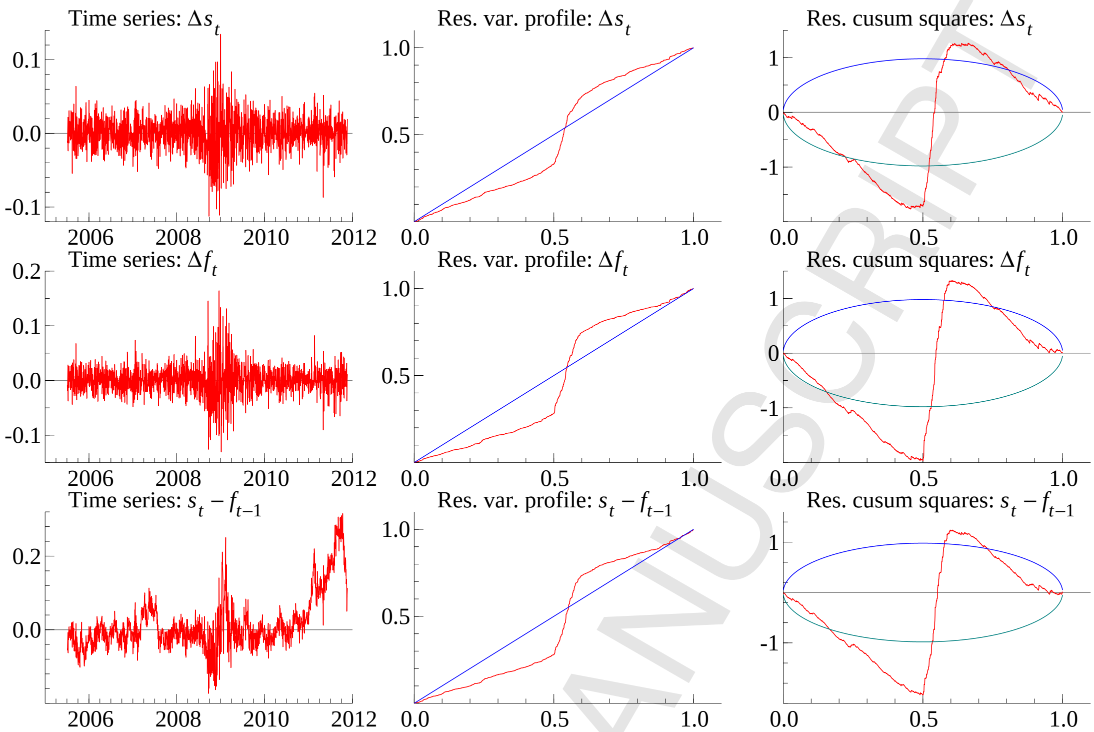
Note: Left panels show time series plots of Δs_t, Δf_t, $s_t - f_{t-1}$, middle panels show the residual variance profiles, $\hat{\eta}(u)$, and right panels show the residual cusum of squares process with 95% confidence bands.

estimate of the fractional parameter d, the two-sided LM test statistic S_{2T} of $H_0 : d = 0$, along with the corresponding asymptotic p-values together with the wild bootstrap p-values, computed as in Algorithm 1 using $B = 9999$ bootstrap replications.

For gold, silver and crude oil, the null hypothesis, $H_0 : d = 0$, cannot be rejected at any conventional significance level using any of the tests, with p-values all above 20% (30% using the wild bootstrap), leading us to conclude that the spot and future prices are indeed both I(1); moreover, the lag lengths selected by the BIC then suggests that these series both follow random walks. On the other hand, for the data on platinum the tests lead to quite different conclusions. When using the asymptotic test, the null hypothesis is rejected at the 1% level for both spot and futures prices. However, based on the results from Table 5 where the hypothesis of constant (un)conditional variance is strongly rejected for the platinum spot and futures prices, our Monte Carlo results in section 5 would suggest that asymptotic tests for $d = 0$ are likely to be unreliable. This standpoint is supported by the corresponding results for the wild bootstrap test. Specifically, when the wild bootstrap is employed, the null hypothesis is now *not* rejected at the 5% level for both the spot and futures prices (p-values are 7.9% and 5.4%, respectively). Hence, the strong heteroskedasticity characterising both spot and futures prices for platinum might explain why the asymptotic test leads to the rejection of the I(1) hypothesis for spot and futures prices. However, by using a test which is robust to heteroskedasticity we are able to accept the hypotheses that both the spot and futures prices for platinum are I(1). Overall, at least when the heteroskedasticity-robust wild bootstrap tests are employed, requirement (i) of the EMH is seen to be consistent with the data.

We now analyse (the weak form of) the EMH, i.e. requirements (ii), (iii), and (iv') jointly, based on the spreads, $s_t - f_{t-1}$, for each of the four commodities considered. For gold, the hypothesis $d = 0$ is easily rejected with p-values less than 1% for the asymptotic test. Using the wild bootstrap test

Figure 5. Graphics for crude oil



Note: Left panels show time series plots of Δs_t, Δf_t, $s_t - f_{t-1}$, middle panels show the residual variance profiles, $\hat{\eta}(u)$, and right panels show the residual cusum of squares process with 95% confidence bands.

the evidence against the null is not as strong but it can still be rejected at the 5% level. Importantly, however, these are left-tail rejections meaning that the I(0) null is being rejected because the estimated value of d is negative. That is, while fractional dynamics appear to exist in the forward premium for gold, there is nonetheless significant evidence of (fractional) co-integration.

The results for the silver and platinum forward premia are qualitatively similar to one another. For both of these commodities the estimate of d is relatively close to zero (slightly negative for silver and slightly positive for platinum), and all reported tests do not reject the null hypothesis, $H_0 : d = 0$, at any conventional significance level. Again, this supports the hypothesis that spot and futures prices are co-integrated with co-integrating vector $(1, -1)'$. Unlike gold, however, the results for these two series suggest that the spread is a standard (non-fractional) I(0) process. As a result, using our heteroskedastic fractionally integrated model we are able to conclude that all of the requirements in (i)–(iii), as well as (iv'), of (the weak form of) the EMH are consistent with the price data for the gold, silver and platinum markets. Our results also highlight that fractional behaviour and/or heteroskedasticity are present in these data which may help to explain why some previous studies have struggled to find support for the EMH in these commodities.

The picture is, however, somewhat different for the forward premium for crude oil. The point estimate of d is 0.78 which is clearly much higher than the estimates of d obtained for the other three commodities. Consequently, we do not present results for the hypothesis $d = 0$ (it is overwhelmingly rejected in any case) and instead present results for one-sided tests of $H_0 : d \leq 1/2$ and $H_0 : d \geq 1$. The former is a test of the null of stationarity of the spreads and the latter is a test of the null of no (fractional) co-integration with co-integrating vector $(1, -1)'$. Firstly, the null hypothesis $H_0 : d \geq 1$ is very easily rejected by all of the tests. This result provides evidence in favour of the existence of the $(1, -1)'$ co-integrating relationship between spot and futures prices. Secondly, the spread does not

Table 5. Conditional and unconditional heteroskedasticity tests

	Gold			Silver		
	s_t	f_t	$s_t - f_{t-1}$	s_t	f_t	$s_t - f_{t-1}$
ARCH(5)	48.576^a	84.805^a	57.507^a	72.408^a	72.752^a	81.799^a
ARCH(21)	193.896^a	150.103^a	142.410^a	161.138^a	99.837^a	168.663^a
$\mathcal{H}_{KS}$	1.469^b	1.345^c	1.384^b	0.850	1.206	1.042
$\mathcal{H}_{K}$	2.197^a	2.239^a	2.170^a	1.337	1.402	1.455
$\mathcal{H}_{CvM}$	0.432^c	0.380^c	0.376^c	0.173	0.327	0.241
$\mathcal{H}_{AD}$	2.837^b	2.452^c	2.423	0.995	1.847	1.364
	Platinum			Crude oil		
	s_t	f_t	$s_t - f_{t-1}$	s_t	f_t	$s_t - f_{t-1}$
ARCH(5)	232.795^a	233.656^a	228.694^a	294.233^a	307.836^a	257.533^a
ARCH(21)	322.834^a	338.595^a	341.414^a	441.169^a	454.992^a	440.536^a
$\mathcal{H}_{KS}$	1.633^b	1.871^a	1.809^a	1.767^a	1.965^a	2.037^a
$\mathcal{H}_{K}$	2.897^a	3.046^a	3.010^a	3.031^a	3.292^a	3.278^a
$\mathcal{H}_{CvM}$	0.837^a	0.931^a	0.910^a	0.972^a	1.135^a	1.090^a
$\mathcal{H}_{AD}$	5.357^a	5.981^a	5.826^a	6.445^a	7.560^a	7.350^a

Notes: ARCH(k) denotes the LM test for ARCH(k) based on a AR(k) regression fitted to the squared residuals, and $\mathcal{H}_{KS}$, $\mathcal{H}_{K}$, $\mathcal{H}_{CvM}$ and $\mathcal{H}_{AD}$ denote the stationary volatility tests proposed in Cavaliere and Taylor (2008*b*, pp. 311–312). The superscripts a,b and c denote significance at the 1%, 5% and 10% nominal (asymptotic) levels, respectively.

appear to be I(0), as noted above, but rather the spread appears to be fractionally integrated. Indeed, stationarity of the spread, $H_0 : d \leq 1/2$, is strongly rejected by the asymptotic test and by both bootstrap tests. As a result, the statistical evidence for oil suggests the existence of co-integration in the spread, but that the associated linear combination $(1, -1)'$, does not decrease the order of integration sufficiently to render the spread stationary. That is, the EMH, even in its weaker form (iv'), does not appear to hold in the case of the crude oil market. This result is not at odds with recent empirical evidence that underlines the inefficiency of the futures crude oil market, see, for example, the discussions on this point in Narayan, Huson and Narayan (2012) and Westerlund and Narayan (2013). However, it is worth noting that these authors, using the more restrictive I(0)/I(1) paradigm, reject the hypothesis that the oil spread constitutes a co-integrated relationship.

We complete our empirical analysis by considering a brief examination of the time (in)stability of the results obtained for the four spreads. This is mainly motivated by the recent financial crisis. Westerlund and Narayan (2013) also investigate the stability of their results across the crisis by splitting the sample into two sub-samples at September 12, 2008. Rather than split the data at an arbitrary time point in this way, we choose instead to repeat our full sample anaysis reported above across rolling subsamples of the data. To that end, in Figure 6 we report rolling subsample estimates of d for the four spreads. These are obtained using a rolling window of length approximately equal to one year (each estimate is based on 260 consecutive observations), where estimates are updated on a weekly basis (every five observations). The AR and MA orders of the baseline ARFIMA models are those obtained by BIC on the full sample, see Table 6. Overall, the estimates of d are seen to be fairly stable over the selected period. These fluctuate around 0 for gold, silver and platinum, and around 0.8 for crude oil. For the latter, the estimate of d increases slightly when the rolling window starts after the third quarter of 2009, which may be a reflection of some instability due to the financial crisis.

In Figure 7 we report the associated rolling subsample p-values for the tests of $H_0 : d = 0$ against $H_1 : d \neq 0$. Again, the results are pretty much in line with what was reported for the full sample above. The wild bootstrap p-values associated with the subsample tests for silver and platinum almost never fall below 5%, while for gold, the subsample wild bootstrap p-values for $d = 0$ fall below 5% for a significant fraction of the rolling windows considered (but as with the full sample results this is due to anti-persistence in the gold spread, see the first panel of Figure 6). Finally, the p-values

Table 6. Application to unbiasedness hypothesis in commodity futures markets

	Δs_t	Δf_t	$s_t - f_{t-1}$	
Panel A: gold				
ARMA order (p, q)	$(0,0)$	$(0,0)$	$(0,1)$	
Estimate of d	-0.025	-0.004	-0.084	
Hypothesis tests	$H_0 : d = 0$	$H_0 : d = 0$	$H_0 : d = 0$	
Test statistic	$S_{2T} = 1.523$	$S_{2T} = 0.033$	$S_{2T} = 8.897$	
p-value, asymptotic	21.7%	85.6%	0.3%	
p-value, wild bootstrap	30.1%	89.1%	2.4%	
Panel B: silver				
ARMA order (p, q)	$(0,0)$	$(0,0)$	$(0,1)$	
Estimate of d	-0.018	-0.005	-0.017	
Hypothesis tests	$H_0 : d = 0$	$H_0 : d = 0$	$H_0 : d = 0$	
Test statistic	$S_{2T} = 0.858$	$S_{2T} = 0.071$	$S_{2T} = 0.338$	
p-value, asymptotic	35.4%	79.0%	56.1%	
p-value, wild bootstrap	56.3%	86.0%	68.8%	
Panel C: platinum				
ARMA order (p, q)	$(0,0)$	$(0,0)$	$(0,1)$	
Estimate of d	0.054	0.057	0.038	
Hypothesis tests	$H_0 : d = 0$	$H_0 : d = 0$	$H_0 : d = 0$	
Test statistic	$S_{2T} = 7.786$	$S_{2T} = 8.281$	$S_{2T} = 2.248$	
p-value, asymptotic	0.5%	0.4%	13.4%	
p-value, wild bootstrap	7.9%	5.4%	28.7%	
Panel D: crude oil				
ARMA order (p, q)	$(0,0)$	$(0,0)$	$(0,1)$	
Estimate of d	-0.014	-0.029	0.780	
Hypothesis tests	$H_0 : d = 0$	$H_0 : d = 0$	$H_0 : d \leq 1/2$	$H_0 : d \geq 1$
Test statistic	$S_{2T} = 0.645$	$S_{2T} = 2.466$	$S_{1T} = 6.361$	$S_{1T} = -16.05$
p-value, asymptotic	42.2%	11.6%	0.0%	0.0%
p-value, wild bootstrap	56.0%	32.2%	0.0%	0.1%

Notes: The table shows point estimates of d, LM test statistics, and corresponding asymptotic and wild bootstrap p-values. For each of the four commodities we analyze: (i) spot returns, Δs_t, (ii) futures returns, Δf_t, (iii) spread, $s_t - f_{t-1}$. The ARMA orders are chosen based on the BIC. Bootstrap p-values are based on $B = 9999$ bootstrap replications.

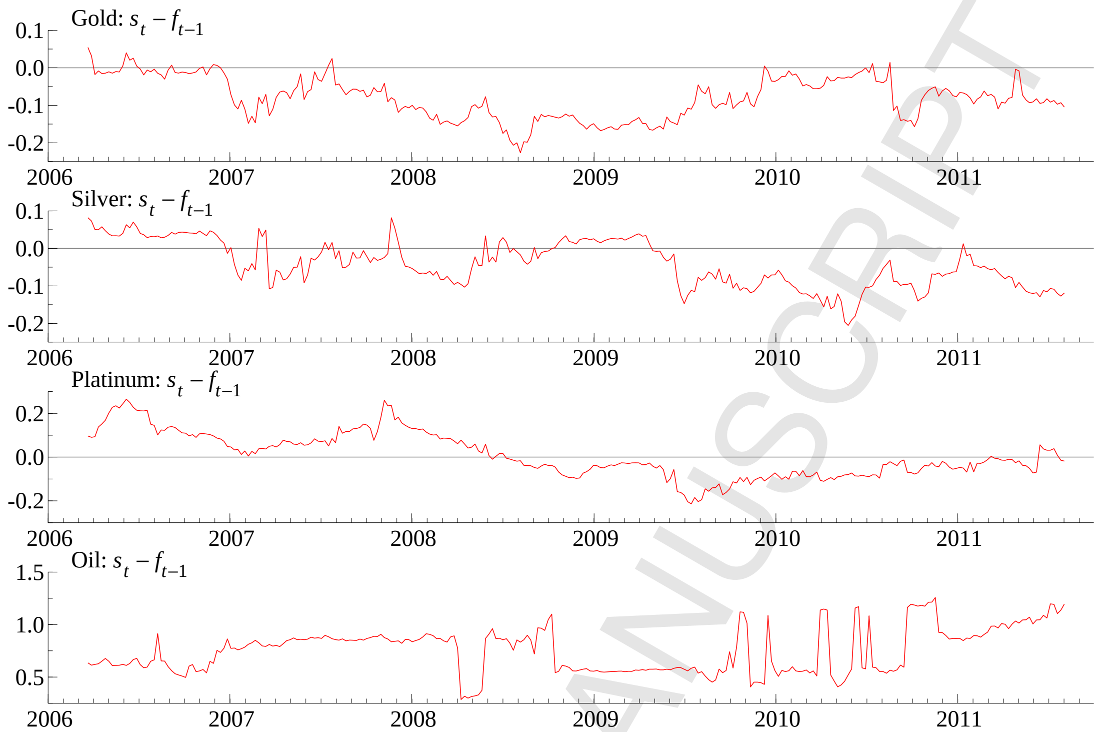
for the sub-sample rolling tests on the crude oil spread lie well below 5% throughout the sample. To summarise, the rolling sample results suggest firstly that the acceptance of (the weak form of) the EMH for gold, silver and platinum prices is robust as to whether the data sample used includes the recent financial crisis period or not, and secondly that the failure to accept the EMH for the case of crude oil cannot simply be attributed to the financial crisis.

7 Conclusions

In this paper we have proposed wild bootstrap implementations of the asymptotic score (one-sided) and LM (two-sided) tests for the order of integration of a fractionally integrated time series. The wild bootstrap was shown to yield tests which are robust to both conditional and unconditional heteroskedasticity of quite general and unknown forms in the shocks. This property was shown not to be shared by the asymptotic tests.

A simulation study highlighted both the potential for severe size distortions with the standard asymptotic LM test in the presence of heteroskedastic shocks and the excellent job done by the wild bootstrap test in controlling finite sample sizes in these cases. Moreover, the bootstrap test was also shown to deliver considerably more reliable finite sample inference than the asymptotic LM test in the homoskedastic case, particularly so where weak dependence was present in the shocks. The simulation study also compared the finite sample properties of using a bootstrap algorithm where the bootstrap sample data were generated using model estimates obtained under the null hypothesis (restricted) with one where they were estimated unrestrictedly. The former was shown to deliver tests with considerably

Figure 6. Rolling window estimates of d for spreads



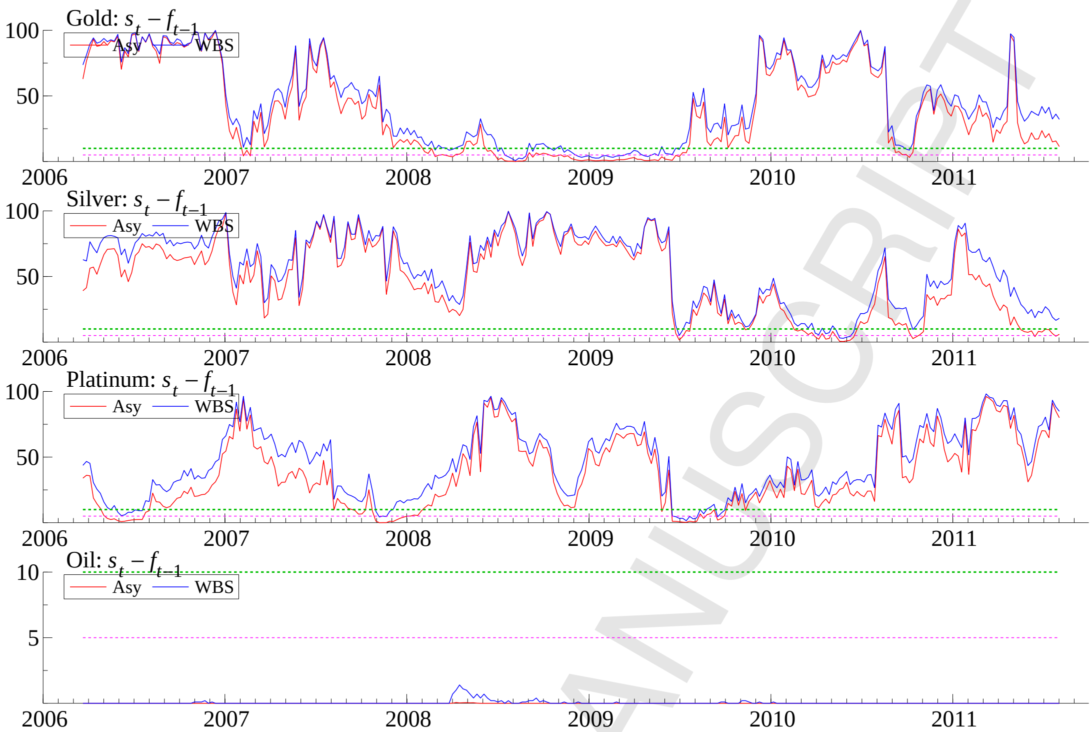
Note: The figure shows estimates of d for rolling windows of length 260.

more robust finite sample size properties whilst at the same time not sacrificing finite sample power to the latter under the alternative. We also compared our proposed tests with the White standard error corrected implementation of the LM test of Agiakoglu and Newbold (1994) and Breitung and Hassler (2002), developed in Kew and Harris (2009). The bootstrap tests were seen to display significantly better finite sample size control than the Kew and Harris (2009) test but to have lower finite sample (size-adjusted) power, suggesting a useful complementarity between the two approaches.

We applied our new bootstrap tests to investigate the price dynamics in four commodity spot and futures markets: gold, silver, platinum and crude oil. Using daily trading data for the period 2005–2011, we found that when fractional integration together with conditional and/or unconditional heteroskedasticity of very general forms are allowed, the evidence in favour of co-integration in the spread between spot and futures prices for these commodities is markedly stronger than had been found in previous work based on more restrictive (usually) homoskedastic I(0)/I(1) models; see Figuera-Ferretti and Gonzalo (2010), Westerlund and Narayan (2013), and reference therein. Moreover, (the weak form of) the efficient market hypothesis is accepted for all markets but oil. Our results were also little altered by whether the data samples used included the recent financial crisis or not, further illustrating the robustness of our proposed tests to large volatility breaks.

We conclude with a topic for further research suggested by an anonymous referee. An alternative approach to the bootstrap tests suggested in this paper would be to attempt to use non-parametric methods to estimate the nuisance parameters λ, ω, and ϖ arising from the heteroskedasticity and weak dependence in the data, and to then use these estimates to non-parameterically correct the LM and score statistics in order to obtain asymptotically pivotal inference. Moreover, bootstrap (using either i.i.d. or wild re-sampling) tests could also be developed based on these statistics which would be expected to improve their finite-sample properties. It would be interesting to develop such tests and to compare them with the tests considered in this paper.

Figure 7. Rolling window tests of $H_0 : d = 0$ for spreads



Notes: The figure shows asymptotic, i.i.d. bootstrap, and wild bootstrap p-values of two-tailed tests of $H_0 : d = 0$ for rolling windows of length 260. The bootstrap tests are based on $B = 999$ replications.

A Appendix

Recall that $\xi_j = (-j^{-1}, c_j')'$, where c_j decays exponentially under Assumption $\mathcal{R}$. This implies the bound $||\xi_j|| \leq Kj^{-1}$ for some $K < \infty$, which we use throughout the proofs without special reference.

A.1 Proof of Theorem 1

We begin with a proof of consistency of the maximum likelihood estimator under the null given in (3.4). This is somewhat more delicate than usual because of the presence of the parameter $\bar{d}$, which is not equal to, but local to, the true value, d_0.

Lemma A.1 *Define* $r(\psi) := \lim_{T\to\infty} ET^{-1}\sum_{t=1}^{T}(c(L,\psi)\, c(L,\psi_0)^{-1}\varepsilon_t)^2$ *and let the assumptions of Theorem 1 be satisfied. Then*

$$\sup_{\psi\in\Psi}\left|T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\left(\bar{d},\psi\right)^2 - r(\psi)\right| \xrightarrow{p} 0 \text{ as } T \to \infty, \tag{A.1}$$

$$\inf_{\psi\in\Psi\cap\{\psi:||\psi-\psi_0||\geq\epsilon\}} r(\psi) > r(\psi_0) \text{ for all } \epsilon > 0. \tag{A.2}$$

It follows that $\tilde{\psi}$ is consistent, i.e., $\tilde{\psi} \xrightarrow{p} \psi_0$ as $T \to \infty$.

Proof. Consistency of $\tilde{\psi}$ follows from (A.1) and (A.2) by Theorem 5.7 of van der Vaart (1998). Let $e_t(\psi) := c(L,\psi)\, c(L,\psi_0)^{-1}\varepsilon_t =: \sum_{n=0}^{\infty}\varphi_n(\psi)\varepsilon_{t-n}$, where $\varphi_0(\psi) = 1$ and $\varphi_n(\psi)$ decays exponentially for all ψ under Assumption $\mathcal{R}$. We can thus assume, for example, that $|\varphi_n(\psi)| \leq Kn^{-1}$ for all $\psi \in \Psi$, but also that $\sum_{n=0}^{\infty}|\varphi_n(\psi)| < \infty$, and we shall use both in this proof.

To show (A.2) first note that

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

As in the proof of Lemma A.3 of CNT, let $q_T = \lfloor T^{\chi}\rfloor$ for some $\chi \in (0,1)$. Then the last term is bounded as

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\sum_{n=0}^{\infty}\varphi_n(\theta)^2(\sigma_{t-n}^2-\sigma_t^2) &\le \sum_{n=0}^{q_T}\varphi_n(\theta)^2 T^{-1}\sum_{t=1}^{T}|\sigma_{t-n}^2-\sigma_t^2| && \text{(A.3)}\\
&\quad + \sum_{n=q_T+1}^{\infty}\varphi_n(\theta)^2 T^{-1}\sum_{t=1}^{T}|\sigma_{t-n}^2-\sigma_t^2|. && \text{(A.4)}
\end{aligned}
$$

Because $\sup_{n=1,\ldots,q_T} T^{-1}\sum_{t=1}^{T}|\sigma_{t-n}^2-\sigma_t^2| \to 0$ by Lemma A.1 in Cavaliere and Taylor (2009) and $\sum_{n=0}^{q_T}\varphi_n(\theta)^2 \le \sum_{n=0}^{\infty}\varphi_n(\theta)^2 < \infty$ for all $\psi \in \Psi$, it holds that $|(A.3)| \to 0$. Next, by Assumption $\mathcal{V}$(a) we have $\sup_t \sigma_t^2 \le M < \infty$ such that $\sup_t T^{-1}\sum_{t=1}^{T}|\sigma_{t-n}^2-\sigma_t^2| \le 2M$, and by Assumption $\mathcal{R}$ we have $\sum_{n=q_T+1}^{\infty}\varphi_n(\theta)^2 \to 0$ for all $\psi\in\Psi$ (because it is the tail of a convergent sum). Therefore $|(A.4)| \to 0$, showing that $T^{-1}\sum_{t=1}^{T}E(e_t^2) = T^{-1}\sum_{t=1}^{T}\sigma_t^2\sum_{n=0}^{\infty}\varphi_n(\psi)^2 + o(1)$. Since $T^{-1}\sum_{t=1}^{T}\sigma_t^2 \to \int_0^1\sigma(s)^2ds$ by Assumption $\mathcal{V}$(a) and the continuous mapping theorem, we have $r(\psi) = \int_0^1\sigma(s)^2ds\sum_{n=0}^{\infty}\varphi_n(\psi)^2$. Under Assumption $\mathcal{R}$, $\varphi_0(\psi) = 1$ for all ψ and $\sum_{n=0}^{\infty}\varphi_n(\psi)^2 = 1 + \sum_{n=1}^{\infty}\varphi_n(\psi)^2 \ge 1$ with equality if and only if $\psi = \psi_0$, which proves (A.2).

To show the result in (A.1) note, by the mean value theorem that, $\hat{\varepsilon}_t\left(\bar{d},\psi\right) = \Delta_+^{\bar{d}-d_0}e_t(\psi) = e_t(\psi)$ $+\frac{\delta}{\sqrt{T}}\sum_{m=1}^{t-1}m^{-1}e_{t-m}(\psi)(1+o_p(1))$, where the $o_p(1)$ term is uniform in t and ignored in the (pointwise) proof of convergence. Thus,

$$
\begin{aligned}
T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\left(\bar{d},\psi\right)^2 - T^{-1}\sum_{t=1}^{T}E(e_t(\psi)^2) &= T^{-1}\sum_{t=1}^{T}\left(e_t(\psi)^2 - T^{-1}\sum_{s=1}^{T}E(e_s(\psi)^2)\right) && \text{(A.5)}\\
&\quad + 2T^{-1}\sum_{t=1}^{T}e_t(\psi)^2\frac{\delta}{\sqrt{T}}\sum_{m=1}^{t-1}m^{-1}e_{t-m}(\psi)^2 && \text{(A.6)}\\
&\quad + T^{-1}\sum_{t=1}^{T}\frac{\delta^2}{T}\sum_{m=1}^{t-1}m^{-1}e_{t-m}(\psi)^2\sum_{j=1}^{t-1}j^{-1}e_{t-j}(\psi)^2. && \text{(A.7)}
\end{aligned}
$$

First write (A.6) as $\sum_{n=0}^{\infty}\varphi_n(\psi)\frac{\delta}{\sqrt{T}}\sum_{m=1}^{T-1}m^{-1}\sum_{k=0}^{\infty}\varphi_k(\psi)T^{-1}\sum_{t=m+1}^{T}\varepsilon_{t-n}\varepsilon_{t-m-k}$ and note that $T^{-1}\sum_{t=m+1}^{T}\varepsilon_{t-n}\varepsilon_{t-m-k} = O_p(1)$ under Assumption $\mathcal{V}$. Then,

$$
(A.6) = O_p\left(\left(\sum_{n=0}^{\infty}|\varphi_n(\psi)|\right)^2\frac{\delta}{\sqrt{T}}\sum_{m=1}^{T-1}m^{-1}\right) = O_p(T^{-1/2}(\log T))
$$

since $\sum_{n=0}^{\infty}|\varphi_n(\psi)| < \infty$ for all $\psi\in\Psi$ under Assumption $\mathcal{R}$. The same argument shows that $(A.7) = O_p(T^{-1}(\log T)^2)$. Next, (A.5) clearly has mean zero. The second moment is

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

where the expectations are zero unless the two highest subscripts are equal, see Lemma A.2 of CNT. By symmetry, we only need to consider three cases.

Case 1) $t-n_1 = t-n_2 = s-m_1 = s-m_2$: In this case the expectations and the σ_t's are uniformly bounded using Assumption $\mathcal{V}$ and we find the contribution $KT^{-2}\sum_{t=1}^{T}\left(\sum_{n=0}^{\infty}\varphi_n(\psi)^2\right)^2 \leq KT^{-1} \to 0$, because $|\varphi_n(\psi)| \leq Kn^{-1}$ for all $\psi \in \Psi$ under Assumption $\mathcal{R}$.

Case 2) $t-n_1 = t-n_2 > s-m_1 \geq s-m_2$: The contribution is bounded by (a constant times)

$$T^{-2}\sum_{t,s=1}^{T}\sum_{n=0}^{\infty}\sum_{m_1=\max(0,s-t+n+1)}^{\infty}\sum_{m_2=m_1}^{\infty}\varphi_n(\psi)^2|\varphi_{m_1}(\psi)||\varphi_{m_2}(\psi)||\kappa_4(t-n,t-n,s-m_1,s-m_2)|$$

$$= T^{-2}\sum_{t\leq s=1}^{T}\sum_{n=0}^{\infty}\sum_{m_1=s-t+n+1}^{\infty}\sum_{m_2=m_1}^{\infty}\varphi_n(\psi)^2|\varphi_{m_1}(\psi)||\varphi_{m_2}(\psi)||\kappa_4(t-n,t-n,s-m_1,s-m_2)| \quad (A.8)$$

$$+ T^{-2}\sum_{t>s=1}^{T}\sum_{n=t-s}^{\infty}\sum_{m_1=s-t+n+1}^{\infty}\sum_{m_2=m_1}^{\infty}\varphi_n(\psi)^2|\varphi_{m_1}(\psi)||\varphi_{m_2}(\psi)||\kappa_4(t-n,t-n,s-m_1,s-m_2)| \quad (A.9)$$

$$+ T^{-2}\sum_{t>s=1}^{T}\sum_{n=0}^{t-s-1}\sum_{m_1=0}^{\infty}\sum_{m_2=m_1}^{\infty}\varphi_n(\psi)^2|\varphi_{m_1}(\psi)||\varphi_{m_2}(\psi)||\kappa_4(t-n,t-n,s-m_1,s-m_2)| \quad (A.10)$$

For (A.8) we note that $|\varphi_{m_1}(\psi)| \leq Km_1^{-1} \leq K(s-t+1)^{-1}$ such that $\sum_{s=t}^{T}|\varphi_{m_1}(\psi)| \leq K(\log T)$ showing that $|(A.8)| = O(T^{-1}(\log T))$ because the summations over m_1, m_2 of $\kappa_4(\cdot)$ are bounded using Assumption $\mathcal{V}$(b)(iii) and the summation over n of $\varphi_n(\psi)^2$ is bounded using Assumption $\mathcal{R}$. For (A.9) we note that $|\varphi_{m_1}(\psi)| \leq Km_1^{-1} \leq K(s-t+n)^{-1}$ such that $\sum_{n=t-s}^{\infty}\varphi_n(\psi)^2|\varphi_{m_1}(\psi)| \leq K\sum_{n=1}^{\infty}n^{-1}(t-s+n)^{-1} \leq K(t-s)^{-1+\eta}$ for some $\eta \in (0,1)$. Since the summations over m_1, m_2 of $\kappa_4(\cdot)$ are bounded using Assumption $\mathcal{V}$(b)(iii), this shows that $|(A.9)| = O(T^{\eta-1})$. Finally, we obtain

$$\begin{aligned}
|(A.10)| &\leq KT^{-2}\sum_{t>s=1}^{T}\sum_{n=0}^{t-s-1}\varphi_n(\psi)^2\sum_{m_1=0}^{\infty}\sum_{m_2=m_1}^{\infty}|\kappa_4(t-n,t-n,s-m_1,s-m_2)| \\
&= KT^{-2}\sum_{t=2}^{T}\sum_{s=1}^{\sqrt{t}}\sum_{n=0}^{t-s-1}\varphi_n(\psi)^2\sum_{m_1=0}^{\infty}\sum_{m_2=m_1}^{\infty}|\kappa_4(t-n,t-n,s-m_1,s-m_2)| \\
&+ KT^{-2}\sum_{t=2}^{T}\sum_{s=\sqrt{t}+1}^{t-1}\sum_{n=0}^{t-s-1}\varphi_n(\psi)^2\sum_{m_1=0}^{\infty}\sum_{m_2=m_1}^{\infty}|\kappa_4(t-n,t-n,s-m_1,s-m_2)|,
\end{aligned}$$

where the first term is $O(T^{-1/2})$ and the second term is $o(1)$ because $\sum_{m_1=0}^{\infty}\sum_{m_2=m_1}^{\infty}|\kappa_4(t-n,t-n,s-m_1,s-m_2)|$ is the tail of the convergent sum $\sum_{m_1=n+s-t}^{\infty}\sum_{m_2=m_1}^{\infty}|\kappa_4(t-n,t-n,s-m_1,s-m_2)|$ when $t-s \geq t-\sqrt{t}-1 \to \infty$, see Assumption $\mathcal{V}$(b)(iii).

Case 3) $t-n_1 = s-m_1 > t-n_2 \geq s-m_2$: The contribution is

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

for some $\eta \in (0,1)$, where the first inequality is by Assumptions $\mathcal{V}$(a),(b)(iii) and $\mathcal{R}$. This shows that the convergence in (A.1) holds pointwise for all $\psi \in \Psi$.

The pointwise convergence in probability thus established can be strengthened to uniform convergence in probability by showing that $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t(d,\psi)^2$ is stochastically equicontinuous (or tight). From Newey (1991, Corollary 2.2) this holds if the derivative is dominated, uniformly in (d,ψ), by a random variable $B_T = O_p(1)$. From Lemma C.3 of CNT it holds that $B_T = \sup|\frac{\partial}{\partial\gamma}T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t(d,\psi)^2|$ $= O_p(1)$, where the supremum is taken over $(d,\psi)\in\{(d,\psi): d-d_0\geq -1/2+c, \psi\in\Psi\}$ for some small $c>0$ such that $u_1=u_2=d-d_0\geq -1/2+c$ and $a=2c>0$. This shows that $T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t(d,\psi)^2$ is stochastically equicontinuous (on a fixed set) and hence that the convergence holds uniformly. ■

Let Υ_0, Ξ_0 and $\xi_{0,j}$ denote Υ, Ξ and ξ_j, respectively, evaluated at the true value γ_0.

Lemma A.2 *Let Assumptions $\mathcal{R}$ and $\mathcal{V}$ be satisfied. Then,*

$$\sqrt{T}\left.\frac{\partial\hat{\sigma}^2(d,\psi)}{\partial\gamma}\right|_{\gamma=\gamma_0} \xrightarrow{w} N\left(0, 4\Upsilon_0\int_0^1\sigma^4(s)ds\right), \tag{A.11}$$

$$\left.\frac{\partial^2\hat{\sigma}^2(d,\psi)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\breve{\gamma}} \xrightarrow{p} 2\Xi_0\int_0^1\sigma^2(s)ds \text{ for any } \breve{\gamma}\xrightarrow{p}\gamma_0. \tag{A.12}$$

Proof. The first and second derivatives of (3.3) are $\sqrt{T}\frac{\partial\hat{\sigma}^2(d,\psi)}{\partial\gamma} = 2T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t(d,\psi)\sum_{j=1}^{t-1}\xi_j\hat{\varepsilon}_{t-j}(d,\psi)$ and

$$\frac{\partial^2\hat{\sigma}^2(d,\psi)}{\partial\gamma\partial\gamma'} = 2T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1}\xi_j\hat{\varepsilon}_{t-j}(d,\psi)\sum_{k=1}^{t-1}\xi_k'\hat{\varepsilon}_{t-k}(d,\psi) + 2T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t(d,\psi)\sum_{j=1}^{t-2}\sum_{k=1}^{t-j-1}\xi_j\xi_k'\hat{\varepsilon}_{t-j-k}(d,\psi).$$

The second derivative is tight (stochastically equicontinuous) by Newey (1991, Corollary 2.2) if its derivative is dominated uniformly in (d,ψ) by a random variable $B_T = O_p(1)$. From Lemma C.3 of CNT this is satisfied uniformly in any small neighborhood of (d_0,ψ_0), see also Nielsen (2013), showing that the second derivative is tight in this neighborhood. This result, together with $\breve{\gamma}\xrightarrow{p}\gamma_0$, implies by Lemma A.3 of Johansen and Nielsen (2012) that the second derivative can be evaluated at the true value, i.e., $\left.\frac{\partial^2\hat{\sigma}^2(d,\psi)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\breve{\gamma}} \xrightarrow{p} \left.\frac{\partial^2\hat{\sigma}^2(d,\psi)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\gamma_0}$. The second derivative, evaluated at the true value, is

$$\left.\frac{\partial^2\hat{\sigma}^2(d,\psi)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\gamma_0} = 2T^{-1}\sum_{t=1}^{T}\sum_{j,k=1}^{t-1}\xi_{0,j}\xi_{0,k}'\varepsilon_{t-j}\varepsilon_{t-k} + 2T^{-1}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-2}\sum_{k=1}^{t-j-1}\xi_{0,j}\xi_{0,k}'\varepsilon_{t-j-k}. \tag{A.13}$$

The first term on the right-hand side has mean $2T^{-1}\sum_{t=1}^{T}\sum_{j,k=1}^{t-1}\xi_{0,j}\xi_{0,k}'\sigma_{t-j}\sigma_{t-k}E(z_{t-j}z_{t-k}) = 2T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1}\xi_{0,j}\xi_{0,j}'\sigma_{t-j}^2 \to 2\Xi_0\int_0^1\sigma^2(s)ds$, by Assumption $\mathcal{V}$(b)(i) and Lemma A.2 of CNT. The variance of the (m,n)'th element is

$$\begin{aligned}
&4T^{-2}\sum_{t,s=1}^{T}\sum_{i,j=1}^{s-1}\sum_{k,l=1}^{t-1}(\xi_{0,i})_m(\xi_{0,j})_n(\xi_{0,k})_m(\xi_{0,l})_n\sigma_{s-i}\sigma_{s-j}\sigma_{t-k}\sigma_{t-l}\\
&\times[E(z_{s-i}z_{s-j}z_{t-k}z_{t-l}) - E(z_{s-i}z_{s-j})E(z_{t-k}z_{t-l})]\\
&\leq KT^{-2}\sum_{t=1}^{T}\sum_{s=t}^{T}\sum_{i,j=1}^{s-1}\sum_{k,l=1}^{t-1}i^{-1}j^{-1}k^{-1}l^{-1}\left|E(z_{s-i}z_{s-j}z_{t-k}z_{t-l}) - E(z_{s-i}z_{s-j})E(z_{t-k}z_{t-l})\right|,
\end{aligned}$$

which converges to zero by exactly the argument for (A.5) in the proof of Lemma A.1. Thus, the first term on the right-hand side of (A.13) converges in L_2-norm, and hence in probability, to $2\Xi_0\int_0^1\sigma^2(s)ds$.

The (m,n)'th element of the second term on the right-hand side of (A.13) has second moment

$$\begin{aligned}
&4T^{-2}\sum_{t=1}^{T}\sigma_t^2E\left(z_t^2\left(\sum_{j=1}^{t-2}\sum_{k=1}^{t-j-1}(\xi_{0,j})_m(\xi_{0,k})_n\sigma_{t-j-k}z_{t-j-k}\right)^2\right)\\
&\leq KT^{-2}\sum_{t=1}^{T}\left(\sum_{j=1}^{t-1}j^{-1}\right)^4 \leq KT^{-1}(\log T)^4 \to 0,
\end{aligned}$$

using Assumptions $\mathcal{V}$(a) and $\mathcal{V}$(b)(ii), so that the second term on the right-hand side of (A.13) converges to zero in L_2-norm, and hence in probability, which proves (A.12).

The first derivative, evaluated at the true value, is $\sqrt{T}\left.\frac{\partial\hat{\sigma}^2(d,\psi)}{\partial\gamma}\right|_{\gamma=\gamma_0} = \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t\sum_{j=1}^{t-1}\xi_{0,j}\varepsilon_{t-j} = \sum_{t=1}^{T}x_{Tt}$, where $x_{Tt} := 2T^{-1/2}\varepsilon_t\sum_{j=1}^{t-1}\xi_{0,j}\varepsilon_{t-j} = 2T^{-1/2}\sigma_t z_t\sum_{j=1}^{t-1}\xi_{0,j}\sigma_{t-j}z_{t-j}$ is a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t$, the sigma-field generated by $\{z_s\}_{s\leq t}$, see Assumption $\mathcal{V}$(b). To apply the central limit theorem, we first verify the Lindeberg condition via Lyapunov's sufficient condition that $\sum_{t=1}^{T}E||x_{Tt}||^{2+\epsilon}\to 0$ for some $\epsilon>0$. Thus,

$$E||x_{Tt}||^{2+\epsilon} = E\left((2T^{-1/2})^{2+\epsilon}|\sigma_t z_t|^{2+\epsilon}||\sum_{j=1}^{t-1}\xi_{0,j}\sigma_{t-j}z_{t-j}||^{2+\epsilon}\right) \leq KT^{-1-\epsilon/2}E\left(|z_t|^{2+\epsilon}(\sum_{j=1}^{t-1}j^{-1}|z_{t-j}|)^{2+\epsilon}\right)$$

by Assumptions $\mathcal{R}$ and $\mathcal{V}$(a). From Minkowski's inequality we find the bound $E(\sum_{j=1}^{t-1}|z_t|j^{-1}|z_{t-j}|)^{2+\epsilon} \leq (\sum_{j=1}^{t-1}(E(|z_t|j^{-1}|z_{t-j}|)^{2+\epsilon})^{1/(2+\epsilon)})^{2+\epsilon}$ such that $E||x_{Tt}||^{2+\epsilon}$ is bounded by

$$KT^{-1-\epsilon/2}\left(\sum_{j=1}^{t-1}\left(E(|z_t|j^{-1}|z_{t-j}|)^{2+\epsilon}\right)^{1/(2+\epsilon)}\right)^{2+\epsilon} \leq KT^{-1-\epsilon/2}\left(\sum_{j=1}^{t-1}j^{-1}\right)^{2+\epsilon} \leq KT^{-1-\epsilon/2}(\log T)^{2+\epsilon},$$

where the first inequality is due to Assumption $\mathcal{V}$(b)(iii) provided ϵ is chosen as $2\epsilon+4\leq 8$. Therefore,

$$\sum_{t=1}^{T}E||x_{Tt}||^{2+\epsilon} \leq KT^{-\epsilon/2}(\log T)^{2+\epsilon}\to 0. \tag{A.14}$$

The sum of squares of x_{Tt} is equal to

$$4T^{-1}\sum_{t=1}^{T}\sigma_t^2 z_t^2\sum_{j,k=1}^{t-1}\xi_{0,j}\xi_{0,k}'\sigma_{t-j}\sigma_{t-k}z_{t-j}z_{t-k}$$
$$= 4T^{-1}\sum_{t=1}^{T}\sigma_t^2\sum_{j,k=1}^{t-1}\xi_{0,j}\xi_{0,k}'\sigma_{t-j}\sigma_{t-k}E(z_t^2 z_{t-j}z_{t-k}) \tag{A.15}$$
$$+ 4T^{-1}\sum_{t=1}^{T}\sigma_t^2\sum_{j,k=1}^{t-1}\xi_{0,j}\xi_{0,k}'\sigma_{t-j}\sigma_{t-k}(z_t^2 z_{t-j}z_{t-k} - E(z_t^2 z_{t-j}z_{t-k})). \tag{A.16}$$

By Lemma A.3 of CNT, (A.15) is

$$4T^{-1}\sum_{t=1}^{T}\sigma_t^4\sum_{j,k=1}^{t-1}\xi_{0,j}\xi_{0,k}'\tau_{jk}(1+o(1))$$
$$= 4T^{-1}\sum_{t=1}^{T}\sigma_t^4\sum_{j,k=1}^{T-1}\xi_{0,j}\xi_{0,k}'\tau_{jk}(1+o(1)) - 4T^{-1}\sum_{t=1}^{T}\sigma_t^4\sum_{j,k=t}^{T-1}\xi_{0,j}\xi_{0,k}'\tau_{jk}(1+o(1)),$$

where the first term converges to $4\Upsilon_0\int_0^1\sigma^4(s)ds$. The second term is bounded by $KT^{-1}\sum_{t=1}^{T}\sum_{j,k=t}^{T-1}j^{-1}k^{-1}\tau_{jk}$ $\leq KT^{-1}\sum_{t=1}^{T}t^{-2}\sum_{j,k=t}^{T-1}\tau_{jk}$, which converges to zero by Assumption $\mathcal{V}$(b)(iii).

The second moment of the (m,n)'th element of (A.16) is

$$16T^{-2}\sum_{t,s=1}^{T}\sigma_t^2\sigma_s^2\sum_{i,j=1}^{s-1}\sum_{k,l=1}^{t-1}(\xi_{0,i})_m(\xi_{0,j})_n(\xi_{0,k})_m(\xi_{0,l})_n\sigma_{s-i}\sigma_{s-j}\sigma_{t-k}\sigma_{t-l}Cov(z_t^2 z_{t-k}z_{t-l}, z_s^2 z_{s-i}z_{s-j})$$
$$\leq KT^{-2}\sum_{t,s=1}^{T}\sum_{i,j=1}^{s-1}\sum_{k,l=1}^{t-1}i^{-1}j^{-1}k^{-1}l^{-1}|Cov(z_t^2 z_{t-k}z_{t-l}, z_s^2 z_{s-i}z_{s-j})|$$
$$= KT^{-2}\sum_{t=1}^{T}\sum_{i,j=1}^{t-1}\sum_{k,l=1}^{t-1}i^{-1}j^{-1}k^{-1}l^{-1}|Cov(z_t^2 z_{t-i}z_{t-j}, z_t^2 z_{t-k}z_{t-l})| \tag{A.17}$$
$$+ KT^{-2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{i,j=1}^{s-1}\sum_{k,l=1}^{t-1}i^{-1}j^{-1}k^{-1}l^{-1}|Cov(z_t^2 z_{t-k}z_{t-l}, z_s^2 z_{s-i}z_{s-j})|. \tag{A.18}$$

For (A.17) we find the simple bound, $KT^{-2}\sum_{t=1}^{T}\left(\sum_{k=1}^{t-1}k^{-1}\right)^4 \leq KT^{-1}(\log T)^4 \to 0$, because z_t has finite eighth order moments by Assumption $\mathcal{V}$(b)(iii). The covariance in (A.18) is a combination of the cumulants of z_t up to order eight. For the eighth order cumulant we find

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

by Assumption $\mathcal{V}$(b)(iii). There are no seventh order cumulants in (A.18) because they would be multiplied by a first order cumulant, which is zero. For products of sixth and second order cumulants we find, for example,

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

by Assumption $\mathcal{V}$(b)(iii). Another example is

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

using Lemma A.2 of CNT. The second term is clearly $O(T^{-1})$ by Assumption $\mathcal{V}$(b)(iii) and the first term is

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

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

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

$$4T^{-1}\sum_{t=1}^{T}\sigma_t^2 z_t^2 \sum_{j,k=1}^{t-1}\xi_{0,j}\xi_{0,k}'\sigma_{t-j}\sigma_{t-k}z_{t-j}z_{t-k} \xrightarrow{p} 4\Upsilon_0\int_0^1 \sigma^4(s)ds. \tag{A.19}$$

Now (A.11) follows by the martingale central limit theorem of McLeish (1974), see his Theorem 2.3 and the comments in the two paragraphs following it. ■

By consistency of the estimator of γ under the null, i.e. $\tilde{\gamma} = (\bar{d}, \tilde{\psi}')'$, see Lemma A.1, we have the following expansion of the likelihood (with subscripts denoting the relevant blocks of the derivatives),

$$D_{Td}(\tilde{\gamma}) = D_{Td}(\gamma_0) + H_{Td\psi}(\check{\gamma})(\tilde{\psi} - \psi_0) + H_{Tdd}(\check{\gamma})(\bar{d} - d_0),$$
$$0 = D_{T\psi}(\tilde{\gamma}) = D_{T\psi}(\gamma_0) + H_{T\psi\psi}(\check{\gamma})(\tilde{\psi} - \psi_0) + H_{T\psi d}(\check{\gamma})(\bar{d} - d_0),$$

where $\check{\gamma}$ denotes an intermediate point between $\tilde{\gamma}$ and γ_0 (different for each row of the Hessian, although this is not important for the subsequent analysis). Using (3.8), this implies, in particular, that

$$\tilde{\psi} - \psi_0 = -H_{T\psi\psi}(\check{\gamma})^{-1} D_{T\psi}(\gamma_0) - H_{T\psi\psi}(\check{\gamma})^{-1} H_{T\psi d}(\check{\gamma})\delta T^{-1/2} \tag{A.20}$$

and thus

$$T^{-1/2} D_{Td}(\tilde{\gamma}) = [1, -H_{Td\psi}(\check{\gamma})H_{T\psi\psi}(\check{\gamma})^{-1}]T^{-1/2} D_T(\gamma_0) + T^{-1}\left(H_{Tdd}(\check{\gamma}) - H_{Td\psi}(\check{\gamma})H_{T\psi\psi}(\check{\gamma})^{-1}H_{T\psi d}(\check{\gamma})\right)\delta. \tag{A.21}$$

Here we note that, by Lemma A.2 combined with $\hat{\sigma}^2(d_0, \psi_0) = T^{-1}\sum_{t=1}^{T}\varepsilon_t^2 \xrightarrow{p} \int_0^1 \sigma(s)^2 ds$, we have $T^{-1/2}D_T(\gamma_0) \xrightarrow{w} N(0, \Upsilon_0\lambda)$ and $T^{-1}H_T(\check{\gamma}) \xrightarrow{p} -\Xi_0$ as $T \to \infty$. Thus, by the partitioned matrix inverse formula, $T^{-1/2}D_{Td}(\tilde{\gamma}) \xrightarrow{w} [1, -(\Xi_0)_{d\psi}(\Xi_0)^{-1}_{\psi\psi}]N(0, \Upsilon_0\lambda) - (\Xi_0^{-1})^{-1}_{dd}\delta$, and

$$S_{1T} = T^{-1/2}D_{Td}(\tilde{\gamma})\sqrt{-TH_T^{-1}(\tilde{\gamma})_{11}} \xrightarrow{w} \sqrt{(\Xi_0^{-1})_{dd}}[1, -(\Xi_0)_{d\psi}(\Xi_0)^{-1}_{\psi\psi}]N(0, \Upsilon_0\lambda) - (\Xi_0^{-1})^{-1/2}_{dd}\delta,$$

which shows (3.9) because $(\Xi_0^{-1})_{dd}[1, -(\Xi_0)_{d\psi}(\Xi_0)^{-1}_{\psi\psi}]\Upsilon_0[1, -(\Xi_0)_{d\psi}(\Xi_0)^{-1}_{\psi\psi}]' = (\Xi_0^{-1}\Upsilon_0\Xi_0^{-1})_{dd}(\Xi_0^{-1})^{-1}_{dd} = \frac{\varpi^2}{\omega^2}$, by another application of the partitioned matrix inverse formula. (3.10) then follows immediately.

A.2 Proof of Theorem 2

Throughout, we use P^* and E^*, respectively, to denote the probability and expectation conditional on the realization of the original sample. Moreover, for a given sequence X_T^* computed on the bootstrap data, with the notations $X_T^* = o_p^*(1)$, in probability, and $X_T^* \xrightarrow{p^*} X$, in probability, we mean that $P^*(|X_T^*| > \epsilon) \to 0$ in probability and $P^*(|X_T^* - X|) > \epsilon) \to 0$ in probability, respectively, for any $\epsilon > 0$ as $T \to \infty$. We first present a lemma with the asymptotic distribution of the restricted estimator.

Lemma A.3 *Let Assumptions $\mathcal{R}$ and $\mathcal{V}$ be satisfied and let $\tilde{\psi}$ denote the restricted estimator (3.4) obtained under (3.1). Then $\sqrt{T}(\tilde{\psi} - \psi_0) \xrightarrow{w} N(-\Phi_0^{-1}\kappa_0\delta, \lambda\Phi_0^{-1}\Lambda_0\Phi_0^{-1})$, where Λ_0 corresponds to $\Lambda := \sum_{j,k=1}^{\infty} c_j c_k' \tau_{j,k}$ evaluated at the true value γ_0.*

Proof. Consistency was shown in Lemma A.1 and the asymptotic distribution follows from (A.20) combined with Lemma A.2. ■

We next present versions of Lemmas A.1 and A.2 for the bootstrap data. The bootstrap objective function is $\hat{\sigma}_*^2(d, \psi) := T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^*(d, \psi)^2$, where $\hat{\varepsilon}_t^*(d, \psi) := c(L, \psi)\Delta_+^d y_t^*$ and y_t^* is defined in (4.1).

Lemma A.4 *Let Assumptions $\mathcal{R}$ and $\mathcal{V}$ be satisfied and let γ_0^* denote the bootstrap true value; i.e., $\gamma_0^* := (\bar{d}, \tilde{\psi}')'$. Let the estimator of ψ for the bootstrap data be given by $\tilde{\psi}^* := \arg\min_{\psi\in\Psi}\hat{\sigma}_*^2(\bar{d}, \psi)$. Then $\tilde{\psi}^* - \tilde{\psi} \xrightarrow{p^*} 0$, in probability, and therefore $(\bar{d}, \tilde{\psi}^{*\prime})' - \gamma_0^* \xrightarrow{p^*} 0$, in probability.*

Proof. The result follows by standard arguments if it is shown that

$$\sup_{\psi\in\Psi}\left|\hat{\sigma}_*^2(\bar{d}, \psi) - \hat{\sigma}^2(\bar{d}, \psi)\right| \xrightarrow{p^*} 0, \text{ in probability.} \tag{A.22}$$

We give the pointwise proof and note that this can be extended to uniform convergence by a similar argument to that in the proof of (A.1).

Define $r^*(\psi) := E^*\hat{\sigma}_*^2(\bar{d},\psi) = E^*T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^*\left(\bar{d},\psi\right)^2$. We then show (A.22) (pointwise) by showing

$$\left|\hat{\sigma}_*^2(\bar{d},\psi) - r^*(\psi)\right| \xrightarrow{p^*} 0, \text{ in probability,} \tag{A.23}$$

$$\left|\hat{\sigma}^2(\bar{d},\psi) - r^*(\psi)\right| \xrightarrow{p} 0. \tag{A.24}$$

Let $c\,(z,\psi)\,c(z,\tilde{\psi})^{-1} =: \sum_{n=0}^{\infty}\tilde{\varphi}_n(\psi)z^n$, where the coefficients $\tilde{\varphi}_n(\psi)$ are exponentially declining under Assumption $\mathcal{R}$. Conditionally on the original data, $\varepsilon_t^* = \tilde{\varepsilon}_{c,t}w_t$ is an independent sequence, see Algorithm 1(ii), so that $\hat{\varepsilon}_t^*\left(\bar{d},\psi\right) = c(L,\psi)c(L,\tilde{\psi})^{-1}\varepsilon_t^* = \sum_{n=0}^{t-1}\tilde{\varphi}_n(\psi)\varepsilon_{t-n}^*$ is a linear process with independent innovations and exponentially declining coefficients. Because, conditionally on the data, the fourth moments of ε_t^* are bounded uniformly in t by Assumption $\mathcal{V}$ and the properties of w_t, the law of large numbers implies that (A.23) holds pointwise for all $\psi \in \Psi$.

Since, conditionally on the data, ε_t^* is uncorrelated and $E^*\varepsilon_t^{*2} = \tilde{\varepsilon}_{c,t}^2$, we have that $r^*(\psi) = T^{-1}\sum_{t=1}^{T}\sum_{n=0}^{t-1}\tilde{\varphi}_n(\psi)^2\tilde{\varepsilon}_{c,t-n}^2$. Also, $\hat{\varepsilon}_t(\bar{d},\psi) = e_t(\psi) + \delta T^{-1/2}\sum_{m=1}^{t-1}m^{-1}e_{t-m}(\psi)(1+o_p(1))$, where the $o_p(1)$ term is uniform in t,m and can be ignored, and thus the left-hand side of (A.24) is

$$\hat{\sigma}^2(\bar{d},\psi) - r^*(\psi) = T^{-1}\sum_{t=1}^{T}\sum_{n=0}^{t-1}\left(\varphi_n(\psi)^2\varepsilon_{t-n}^2 - \tilde{\varphi}_n(\psi)^2\tilde{\varepsilon}_{c,t-n}^2\right) + T^{-1}\sum_{t=1}^{T}\sum_{n=t}^{\infty}\varphi_n(\psi)^2\varepsilon_{t-n}^2$$
$$+ T^{-1}\sum_{t=1}^{T}\frac{\delta^2}{T}\sum_{m=1}^{t-1}\sum_{n=1}^{t-1}m^{-1}n^{-1}e_{t-m}(\psi)e_{t-n}(\psi) + 2T^{-1}\sum_{t=1}^{T}\frac{\delta}{T^{1/2}}\sum_{m=1}^{t-1}m^{-1}e_{t-m}(\psi)e_t(\psi).$$

The last three terms are easily shown to be $o_p(1)$ by L_1-convergence using that $E|\varepsilon_{t-n}^2|$, $E|e_{t-m}(\psi)e_{t-n}(\psi)|$, and $E|e_{t-m}(\psi)e_t(\psi)|$ are bounded for all $\psi \in \Psi$. For the first term we use $\tilde{\varepsilon}_{c,t} = \hat{\varepsilon}_t(\bar{d},\tilde{\psi}) - T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t(\bar{d},\tilde{\psi})$, and by L_1-convergence the contributions of all the terms in $\tilde{\varepsilon}_{c,t-n}^2$ to $\hat{\sigma}^2(\bar{d},\psi) - r^*(\psi)$ are easily shown to be $o_p(1)$, except that involving $e_{t-n}(\tilde{\psi})^2$, which leaves $\hat{\sigma}^2(\bar{d},\psi) - r^*(\psi) = T^{-1}\sum_{t=1}^{T}\sum_{n=0}^{t-1}(\varphi_n(\psi)^2\varepsilon_{t-n}^2 - \tilde{\varphi}_n(\psi)^2e_{t-n}(\tilde{\psi})^2) + o_p(1)$. This term can be shown to be $o_p(1)$ using standard methods since it does not depend on the bootstrap data and only the weak dependence parameter ψ is now involved (and not d), and this concludes the proof. ■

Lemma A.5 *Let Assumptions $\mathcal{R}$ and $\mathcal{V}$ be satisfied and let γ_0^* denote the bootstrap true value; i.e., $\gamma_0^* := (\bar{d},\tilde{\psi}')'$. Then, defining $\Upsilon_0^\dagger := \sum_{j=1}^{\infty}\xi_{0,j}\xi_{0,j}'\tau_{j,j}$,*

$$\sqrt{T}\left.\frac{\partial\hat{\sigma}_*^2\,(d,\psi)}{\partial\gamma}\right|_{\gamma=\gamma_0^*} \xrightarrow{w}_p N(0, 4\Upsilon_0^\dagger\int_0^1\sigma^4(s)ds), \tag{A.25}$$

$$\left.\frac{\partial^2\hat{\sigma}_*^2\,(d,\psi)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\check{\gamma}} \xrightarrow{p^*} 2\Xi_0\int_0^1\sigma^2(s)ds, \text{ in probability, for any } \check{\gamma} \text{ such that } \check{\gamma}-\gamma_0^* \xrightarrow{p^*} 0, \text{ in probability.} \tag{A.26}$$

Proof. As in Lemma A.2 it holds that $\sqrt{T}\frac{\partial\hat{\sigma}_*^2(d,\psi)}{\partial\gamma} = 2T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^*\,(d,\psi)\sum_{j=1}^{t-1}\xi_j\hat{\varepsilon}_{t-j}^*\,(d,\psi)$ and

$$\frac{\partial^2\hat{\sigma}_*^2(d,\psi)}{\partial\gamma\partial\gamma'} = 2T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1}\xi_j\hat{\varepsilon}_{t-j}^*\,(d,\psi)\sum_{k=1}^{t-1}\xi_k'\hat{\varepsilon}_{t-k}^*\,(d,\psi) + 2T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^*\,(d,\psi)\sum_{j=1}^{t-2}\sum_{k=1}^{t-j-1}\xi_j\xi_k'\hat{\varepsilon}_{t-j-k}^*\,(d,\psi).$$

We first provide the proof for the weak convergence in (A.25). We have that $\sqrt{T}\left.\frac{\partial\hat{\sigma}_*^2(d,\psi)}{\partial\gamma}\right|_{\gamma=\gamma_0^*} = \frac{2}{\sqrt{T}}\sum_{t=1}^{T}\varepsilon_t^*\sum_{j=1}^{t-1}\tilde{\xi}_j\varepsilon_{t-j}^*$, where $\tilde{\xi}_j$ denotes ξ_j evaluated at $(\bar{d},\tilde{\psi}')'$. Conditional on the original data, $x_{Tt}^* := 2T^{-1/2}\varepsilon_t^*\sum_{j=1}^{t-1}\tilde{\xi}_j\varepsilon_{t-j}^*$ is a martingale difference array with respect to the filtration $\mathcal{F}_t^*$, i.e. the sigma-field generated by $\{\varepsilon_t^*,\ldots,\varepsilon_1^*\}$. First we find the probability limit of $\sum_{t=1}^{T}x_{Tt}^{*2}$ and then we show that the Lindeberg condition is satisfied.

The sum of squares of x_{Tt}^* is

$$4T^{-1}\sum_{t=1}^{T}\varepsilon_t^{*2}\sum_{j,k=1}^{t-1}\tilde{\xi}_j\tilde{\xi}_k'\varepsilon_{t-j}^*\varepsilon_{t-k}^* = 4T^{-1}\sum_{t=1}^{T}\varepsilon_t^{*2}\sum_{j=1}^{t-1}\tilde{\xi}_j\tilde{\xi}_j'\varepsilon_{t-j}^{*2} \tag{A.27}$$
$$+ 4T^{-1}\sum_{t=1}^{T}\varepsilon_t^{*2}\sum_{j=1}^{t-1}\sum_{k=1,k\neq j}^{t-1}\tilde{\xi}_j\tilde{\xi}_k'\varepsilon_{t-j}^*\varepsilon_{t-k}^* =: A_{1T}^* + A_{2T}^*,$$

where we now show that $A_{1T}^* \xrightarrow{p^*} 4\Upsilon_0^\dagger \int_0^1 \sigma(s)^4 ds$ and $A_{2T}^* \xrightarrow{p^*} 0$, in probability, respectively. Recall that $\varepsilon_t^* := \tilde{\varepsilon}_{c,t} w_t$ such that, under the wild bootstrap probability measure, we have that $E^*\left(\varepsilon_t^{*2}\right) = \tilde{\varepsilon}_{c,t}^2 = \left(\tilde{\varepsilon}_t - \bar{\tilde{\varepsilon}}_T\right)^2$, where $\bar{\tilde{\varepsilon}}_T := T^{-1}\sum_{t=1}^T \tilde{\varepsilon}_t$ and $\tilde{\varepsilon}_t := \hat{\varepsilon}_t(\bar{d}, \tilde{\psi})$ denotes the restricted residuals.

Consider A_{1T}^* first. By setting $\eta_t^* := \tilde{\varepsilon}_{c,t}^2\left(w_t^2 - 1\right)$ we can rearrange A_{1T}^* as

$$A_{1T}^* = 4T^{-1}\sum_{t=1}^T \tilde{\varepsilon}_{c,t}^2 \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \tilde{\varepsilon}_{c,t-j}^2 + 4T^{-1}\sum_{t=1}^T \tilde{\varepsilon}_{c,t}^2 \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \eta_{t-j}^* + 4T^{-1}\sum_{t=1}^T \eta_t^* \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \varepsilon_{t-j}^{*2}. \quad (A.28)$$

We first examine the first term of (A.28). By the mean value theorem, $\tilde{\varepsilon}_t = \hat{\varepsilon}_t(\bar{d}, \tilde{\psi}) = \Delta_+^{\bar{d}} c(L, \tilde{\psi}) y_t = \varepsilon_t + (\tilde{\gamma} - \gamma_0)' \sum_{m=1}^{t-1} \tilde{\xi}_m \varepsilon_{t-m}(1 + o_p(1))$, where the $o_p(1)$ term is uniform in t and ignored in the following. From Lemma A.3 and because $E|\varepsilon_t| < \infty$ uniformly in t, $(\tilde{\gamma} - \gamma_0)' \sum_{m=1}^{t-1} \tilde{\xi}_m \varepsilon_{t-m} = O_p(T^{-1/2}(\log T))$ and $\bar{\tilde{\varepsilon}}_T = O_p(T^{-1/2}(\log T))$ uniformly in t. Then $\tilde{\varepsilon}_t - \bar{\tilde{\varepsilon}}_T = \varepsilon_t + a_{tT}$, where $a_{tT} = O_p(T^{-1/2}(\log T))$ uniformly in t, so the first term of A_{1T}^* is

$$\begin{aligned}
&T^{-1}\sum_{t=1}^T (\varepsilon_t + a_{tT})^2 \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' (\varepsilon_{t-j} + a_{t-j,T})^2 - T^{-1}\sum_{t=1}^T \varepsilon_t^2 \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \varepsilon_{t-j}^2 \\
&= T^{-1}\sum_{t=1}^T \varepsilon_t^2 \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \left(a_{t-j,T}^2 + 2\varepsilon_{t-j} a_{t-j,T}\right) + T^{-1}\sum_{t=1}^T \left(a_{tT}^2 + 2\varepsilon_t a_{tT}\right) \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \varepsilon_{t-j}^2 \\
&+ T^{-1}\sum_{t=1}^T \left(a_{tT}^2 + 2\varepsilon_t a_{tT}\right) \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \left(a_{t-j,T}^2 + 2\varepsilon_{t-j} a_{t-j,T}\right),
\end{aligned}$$

where each of the three terms on the right-hand side converge to zero in L_1-norm. Then, by Lemma A.3 and the delta method, $T^{-1}\sum_{t=1}^T \varepsilon_t^2 \sum_{j=1}^{t-1} \tilde{\xi}_j \tilde{\xi}_j' \varepsilon_{t-j}^2 = T^{-1}\sum_{t=1}^T \varepsilon_t^2 \sum_{j=1}^{t-1} \xi_{0,j} \xi_{0,j}' \varepsilon_{t-j}^2 + o_p(1)$, so that we are left with $4T^{-1}\sum_{t=1}^T \varepsilon_t^2 \sum_{j=1}^{t-1} \xi_{0,j}\xi_{0,j}' \varepsilon_{t-j}^2 = 4T^{-1}\sum_{t=1}^T \sigma_t^2 z_t^2 \sum_{j=1}^{t-1} \xi_{0,j}\xi_{0,j}' \sigma_{t-j}^2 z_{t-j}^2 \xrightarrow{p} 4\Upsilon_0^\dagger \int_0^1 \sigma^4(s) ds$, following the same arguments as in the proof of (A.19).

Next, recalling that $\eta_t^* := \tilde{\varepsilon}_{c,t}^2\left(w_t^2 - 1\right)$, the (m,n)'th element of the second term of (A.28) is given by $4T^{-1}\sum_{t=1}^T \tilde{\varepsilon}_{c,t}^2 \sum_{j=1}^{t-1} (\tilde{\xi}_j)_m (\tilde{\xi}_j)_n \eta_{t-j}^* = 4T^{-1}\sum_{t=1}^{T-1} \eta_t^* \sum_{j=1}^{T-t} (\tilde{\xi}_j)_m (\tilde{\xi}_j)_n \tilde{\varepsilon}_{c,t+j}^2$. Conditional on the data and with $\zeta_4 := E((w_t^2 - 1)^2)$, the second moment of this term is $16\zeta_4 T^{-2}\sum_{t=1}^{T-1} \tilde{\varepsilon}_{c,t}^4 (\sum_{j=1}^{T-t} (\tilde{\xi}_j)_m (\tilde{\xi}_j)_n \tilde{\varepsilon}_{c,t+j}^2)^2 \xrightarrow{L_1} 0$, under the 8th-order moment condition implied by Assumption $\mathcal{V}$(b)(iii). Thus, the second term of (A.28) is $o_p^*(1)$, in probability. Similarly, conditional on the data, the second moment of the (m,n)'th element of the third term of (A.28) is $16\zeta_4 T^{-2}\sum_{t=2}^T \tilde{\varepsilon}_{c,t}^4 \sum_{j,k=1}^{t-1} E(w_{t-j}^2 w_{t-k}^2)(\tilde{\xi}_j)_m (\tilde{\xi}_j)_n (\tilde{\xi}_k)_m (\tilde{\xi}_k)_n \tilde{\varepsilon}_{c,t-j}^2 \tilde{\varepsilon}_{c,t-k}^2 \xrightarrow{L_1} 0$, such that third term of (A.28) is $o_p^*(1)$, in probability. Hence $A_{1T}^* \xrightarrow{p^*} 4\Upsilon_0^\dagger \int_0^1 \sigma(s)^4 ds$, in probability

Next, consider $A_{2T}^* = 4T^{-1}\sum_{t=1}^{T-1}\sum_{s=1, s\neq t}^{T-1} w_t w_s a_{t,s}$, where $a_{t,s} := \sum_{j=\max(t,s)+1}^T \tilde{\varepsilon}_{c,j}^2 \tilde{\xi}_{j-t} \tilde{\xi}_{j-s}' \tilde{\varepsilon}_{c,t} \tilde{\varepsilon}_{c,s}$ depends only on the original data. Thus, conditional on the original data, A_{2T}^* is zero mean and the variance of its (m,n)'th element is $E^*((A_{2T}^*)_{m,n}^2) = 16T^{-2}\sum_{t=1}^{T-1}\sum_{s=1,s\neq t}^{T-1} E(w_t^2)E(w_s^2)(a_{t,s})_{m,n}^2 = 16T^{-2}\sum_{t=1}^{T-1}\sum_{s=1,s\neq t}^{T-1} (a_{t,s})_{m,n}^2$. As above, apart from $o_p(1)$ terms, $a_{t,s} = \sum_{j=\max(t,s)+1}^T \varepsilon_j^2 \xi_{0,j-t} \xi_{0,j-s}' \varepsilon_t \varepsilon_s$, and we therefore examine $T^{-2}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T-1}\sum_{j=s+1}^T\sum_{k=s+1}^T \varepsilon_j^2 \varepsilon_k^2 (\xi_{0,j-t})_m (\xi_{0,j-s})_n \varepsilon_t^2 \varepsilon_s^2 (\xi_{0,k-t})_m (\xi_{0,k-s})_n$, with expected absolute value bounded by $KT^{-2}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T-1}\sum_{j=s+1}^T\sum_{k=s+1}^T (j-t)^{-1}(j-s)^{-1}(k-t)^{-1}(k-s)^{-1} \leq KT^{-2}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T-1} (\log T)^2 (s-t)^{-2} \leq K(\log T)^2 T^{-1}$, using $||\xi_{0,j}|| \leq Kj^{-1}$ for all $j \geq 1$, so that A_{2T}^* converges to zero in L_1-norm, and thus in probability.

For the Lindeberg condition we verify Lyapunov's sufficient condition. Conditional on the original data and for any arbitrary conforming vector ν,

$$\begin{aligned}
T^{-2}\sum_{t=1}^T E^*\left(\nu'\left(\varepsilon_t^* \sum_{j=1}^{t-1} \tilde{\xi}_j \varepsilon_{t-j}^*\right)\right)^4 &= T^{-2}\sum_{t=1}^T E^*\left(\varepsilon_t^{*4}\right) \sum_{j,k=1}^{t-1} \left(\nu'\tilde{\xi}_j \nu'\tilde{\xi}_k\right)^2 E^*\left(\varepsilon_{t-j}^{*2} \varepsilon_{t-k}^{*2}\right) \\
&= E(w_t^4) T^{-2}\sum_{t=1}^T \tilde{\varepsilon}_{c,t}^4 \sum_{j,k=1}^{t-1} E(w_{t-j}^2 w_{t-k}^2)\left(\nu'\tilde{\xi}_j \nu'\tilde{\xi}_k\right)^2 \tilde{\varepsilon}_{c,t-j}^2 \tilde{\varepsilon}_{c,t-k}^2,
\end{aligned}$$

where the first equality is because the ε_t^* are independent conditional on the original data. By exactly the same methods as applied in the analysis of the sum of squares of x_{Tt}^* above, the L_1-norm of the right-hand side is bounded by $KT^{-2}\sum_{t=1}^{T}(\sum_{j=1}^{t-1}(\nu'\tilde{\xi}_j)^2)^2 = O(T^{-1})$ under the 8th-order moment condition in Assumption $\mathcal{V}$(b)(iii), so that the right-hand side converges to zero in probability. Thus, the Lindeberg condition is satisfied, which completes the proof of (A.25).

We finally show (A.26). By the same argument as in the proof of (A.12) in Lemma A.2, the second derivative can be evaluated at the bootstrap true value, γ_0^*. Thus,

$$\left.\frac{\partial^2 \hat{\sigma}_*^2(d,\psi)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\gamma_0^*} = 2T^{-1}\sum_{t=1}^{T}\sum_{j,k=1}^{t-1}\tilde{\xi}_j\tilde{\xi}_k'\varepsilon_{t-j}^*\varepsilon_{t-k}^* + 2T^{-1}\sum_{t=1}^{T}\varepsilon_t^*\sum_{j=1}^{t-2}\sum_{k=1}^{t-j-1}\tilde{\xi}_j\tilde{\xi}_k'\varepsilon_{t-j-k}^* =: B_{1T}^* + B_{2T}^*.$$

First, by the same reasoning used for (A.27), $B_{1T}^* \xrightarrow{p^*} 2\Xi_0\int_0^1\sigma^2(s)ds$, in probability. Second, also by the same reasoning as applied above, $\varepsilon_t^*\sum_{j=1}^{t-2}\sum_{k=1}^{t-j-1}\tilde{\xi}_j\tilde{\xi}_k'\varepsilon_{t-j-k}^*$ is a martingale difference sequence with respect to $\mathcal{F}_t^*$, and B_{2T}^* is therefore $o_p^*(1)$, in probability, because of the normalization by T^{-1}. ■

In view of Lemmas A.4 and A.5, the proof of the theorem is completed as in the proof of Theorem 1. We note that, under Assumption $\mathcal{V}'$, $\Upsilon_0 = \sum_{j=1}^{\infty}\xi_{0,j}\xi_{0,j}'\tau_{j,j} = \Upsilon_0^\dagger$.

A.3 Proof of Corollary 2

Theorem 2 implies that, uniformly in probability, $G_T^*(\cdot) \to F_1\left(\frac{\cdot}{\lambda\varpi^2\omega^{-2}}, 0\right)$, with F_1 as defined in section 3. This implies that, under the null hypothesis, P_T^* converges weakly to $U[0,1]$, see Hansen (2000, proof of Theorem 5).

References

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

Baillie, R.T. and T. Bollerslev (1994), The long memory of the forward premium, *Journal of International Money and Finance* 13, 565–571.

Baillie, R.T. and T. Bollerslev (2000), The forward premium anomaly is not as bad as you think, *Journal of International Money and Finance* 19, 471–488.

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.

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

Cavaliere, G., M.Ø. Nielsen and A.M.R. Taylor (2013), Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price dynamics in commodity spot and futures markets, QED working paper 1309, Queen's University.

Cavaliere, G., M.Ø. Nielsen and A.M.R. Taylor (2014), Quasi-maximum likelihood estimation of heteroskedastic fractional time series models, QED working paper 1324, Queen's University.

Cavaliere, G., A. Rahbek and A.M.R. Taylor (2014), Bootstrap determination of the co-integration rank in heteroskedastic VAR models, *Econometric Reviews* 33, 606–650.

Cavaliere, G. and A.M.R. Taylor (2007), Testing for unit roots in time series models with non-stationary volatility, *Journal of Econometrics* 140, 919–947.

Cavaliere, G. and A.M.R. Taylor (2008*a*), Bootstrap unit root tests for time series with non-stationary volatility, *Econometric Theory* 24, 43–71.

Cavaliere, G. and A.M.R. Taylor (2008*b*), Time-change unit root tests for time series 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.

Cheng, X. and P.C.B. Phillips (2012), Cointegrating rank selection in models with time-varying variance, *Journal of Econometrics* 169, 155–165.

Davidson, R. and E. Flachaire (2008), The wild bootstrap, tamed at last, *Journal of Econometrics* 146, 162–169.

Davidson, R. and J. MacKinnon (2000), Bootstrap tests: how many bootstraps? *Econometric Reviews* 19, 55–68.

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

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

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

Engle, R.F. (1990), Discussion: stock market volatility and the crash of '87, *Review of Financial Studies* 3, 103–106.

Figuerola-Ferretti, I. and J. Gonzalo (2010), Modelling and measuring price discovery in commodity markets, *Journal of Econometrics* 158, 95–107.

Gil-Alana, L.A. and P.M. Robinson (1997), Testing of unit root and other non-stationary hypotheses in macroeconomic time series, *Journal of Econometrics* 80, 241–268.

Glosten, L.R., R. Jaganathan and D.E. Runkle (1993), On the relation between the expected value and the volatility of nominal excess returns on stocks, *Journal of Finance* 48, 1779–1801.

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

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

Hansen, B.E. (2000), Sample splitting and threshold estimation, *Econometrica* 68, 575–603.

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

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

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

Johansen, S. and M.Ø. Nielsen (2014), The role of initial values in nonstationary fractional time series models, QED working paper 1300, Queen's University.

Kew, H. and D. Harris (2009), Heteroskedasticity-robust testing for a fractional unit root, *Econometric Theory* 25, 1734–1753.

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

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.

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

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

Loretan, M. and P.C.B. Phillips (1994), Testing the covariance stationarity of heavy-tailed time series: an overview of the theory with applications to several financial datasets, *Journal of Empirical Finance* 1, 211–248.

Luo, G.Y. (1998), Market efficiency and natural selection in a commodity futures market, *Review of Financial Studies* 11, 647–674.

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

Maynard, A. and P.C.B. Phillips (2001), Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly, *Journal of Applied Econometrics* 16, 671–708.

McLeish, D.L. (1974), Dependent central limit theorems and invariance principles, *Annals of Probability* 2, 620–628.

Narayan, P.K., A.A. Huson, and S. Narayan (2012), Do momentum-based trading strategies work in commodity markets. Working paper, Deakin University.

Nelson, D.B. (1991), Conditional heteroskedasticity in asset returns: a new approach, *Econometrica* 59, 347–369.

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

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

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.

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

Sensier, M. and D. van Dijk (2004), Testing for volatility changes in U.S. macroeconomic time series, *Review of Economics and Statistics* 86, 833–839.

Sin, C.-Y. and H. White (1996), Information criteria for selecting possibly misspecified parametric models, *Journal of Econometrics* 71, 207–225.

Stock, J.H. and M.W. Watson (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in R.F. Engle and H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger,* Oxford: Oxford University Press, 1–44.

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

Westerlund, J. and P. Narayan (2013), Testing the efficient market hypothesis in conditionally heteroskedastic futures markets, *Journal of Futures Markets* 33, 1024–1045.

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