Econometric Theory, **33**, 2017, 755–778.
doi:10.1017/S0266466615000481

ADAPTIVE LONG MEMORY TESTING UNDER HETEROSKEDASTICITY

DAVID HARRIS AND HSEIN KEW
Monash University

This paper considers adaptive hypothesis testing for the fractional differencing parameter in a parametric ARFIMA model with unconditional heteroskedasticity of unknown form. A weighted score test based on a nonparametric variance estimator is proposed and shown to be asymptotically equivalent, under the null and local alternatives, to the Neyman-Rao effective score test constructed under Gaussianity and known variance process. The proposed test is therefore asymptotically efficient under Gaussianity. The finite sample properties of the test are investigated in a Monte Carlo experiment and shown to provide potentially large power gains over the usual unweighted long memory test.

1. INTRODUCTION

There is a large literature on statistical inference for the fractional differencing parameter in a stationary ARFIMA model. Of particular note is Robinson (1994), who derived asymptotically efficient score-based tests; see also Tanaka (1999) and Nielsen (2004). Regression based LM tests of fractional integration have been developed by Robinson (1991), Agiakloglou and Newbold (1994), Breitung and Hassler (2002), Nielsen (2005), Demetrescu, Kuzin, and Hassler (2008), and Hassler, Rodrigues, and Rubia (2009). For a Wald-type statistic, Dolado, Gonzalo, and Mayoral (2002) and Lobato and Velasco (2006, 2007) proposed a regression based testing framework; and Ling and Li (2001), Johansen and Nielsen (2010), Hualde and Robinson (2011), Nielsen (2015) and Johansen and Nielsen (in press) deal with parametric estimation of the memory parameter. All of this literature maintains an assumption of unconditional homoskedasticity. That is, while the disturbances of the model may be permitted to follow a martingale difference structure that allows for some degree of conditional heteroskedasticity, this literature does not allow for changes in the unconditional variance.

There is, however, abundant empirical evidence that macroeconomic and financial time series exhibit unconditional heteroskedasticity; see for example Pagan and Schwert (1990), Loretan and Phillips (1994), Watson (1999), McConnell and Perez-Quiros (2000), van Dijk, Osborn, and Sensier (2002), Sensier and Van Dijk (2004), Stărică and Granger (2005) and Dalla, Giraitis,

We are grateful to the co-editor and two referees for very helpful suggestions and comments that led to significant improvement on an earlier version of this paper. This research is supported by the Australian Research Council Discovery Grant DP1094010. Address correspondence to David Harris, Department of Econometrics and Business Statistics, Monash University, E-mail: david.harris3@monash.edu

and Phillips (2015). Kew and Harris (2009) and Cavaliere, Nielsen, and Taylor (2015a, hereafter CNT, 2015b) derived some implications for the size of long memory tests in the presence of such heteroskedasticity and constructed a heteroskedasticity-robust test, but they did not pursue the possibility of adapting the test to recover the power losses that unmodelled heteroskedasticity can incur. This paper takes up this point and derives a test that nonparametrically adapts to unconditional heteroskedasticity of unknown form. In particular, we first derive the infeasible asymptotically efficient score test for a known variance process and then prove the asymptotic equivalence of a feasible version of this test that estimates the variance process using a kernel-based nonparametric regression on the squares of the residuals of the model. This approach closely follows that taken by Xu and Phillips (2008) for an AR model and extends it to long memory testing in ARFIMA models.

Our paper sits within a growing literature that addresses issues of heteroskedasticity in time series models. Hamori and Tokihisa (1997), Kim, Leybourne, and Newbold (2002), Cavaliere (2004a), Cavaliere and Taylor (2007, 2008a, 2008b, 2008c, 2009), Beare (2008), Cavaliere, Harvey, Leybourne, and Taylor (2011, 2015), Smeekes and Taylor (2012) and Cavaliere, Phillips, Smeekes, and Taylor (2015) examine the effect of unconditional heteroskedasticity on unit root tests; Busetti and Taylor (2003), Cavaliere (2004b) and Cavaliere and Taylor (2005) examine the stationarity tests; Phillips and Xu (2006) and Xu and Phillips (2008) examine the stationary autoregressive models; Cavaliere and Taylor (2006), Chung and Park (2007), Cavaliere, Rahbek, and Taylor (2010a, 2010b, 2014), Kim and Park (2010), Cheng and Phillips (2012) and Cavaliere, De Angelis, Rahbek, and Taylor (2015) examine the cointegration tests; Demetrescu and Hanck (2012a, 2012b) and Westerlund (2014) examine the panel unit root tests; Xu (2013, 2015) examines the CUSUM-type statistics for structural change; and Dalla, Giraitis, and Phillips (2015) examine the variance stability statistic.

The use of kernel based methods to estimate the unconditional variance, as first suggested by Xu and Phillips (2008), has been considered, more recently, in the context of portmanteau tests and VAR models; for example Patilea and Raïssi (2014) deal with the Engle (1982) and McLeod and Li (1983) ARCH effects tests, Harris and Kew (2014) deal with the Box–Pierce (1970) autocorrelation test; and Patilea and Raïssi (2012, 2013) deal with adaptive estimation of VAR models. More generally though, constructing adaptive estimators and efficient tests that take explicit account of nonparametric heteroskedastic models has received a lot of attention in the literature; see for example, Carroll (1982), Robinson (1987), Kitamura, Tripathi, and Ahn (2004), Kuersteiner (2002), Harvey and Robinson (1988), Hansen (1995), Xu (2008a, 2008b, 2012), Xu and Phillips (2011) and Xu and Yang (2015).

In a closely related and important literature that deals with conditional heteroskedasticity, Baillie, Chung, and Tieslau (1996), Ling and Li (1997), Li, Ling, and McAleer (2002) and Ling (2003) consider efficient Maximum Likelihood estimation of an ARFIMA model in the presence of parametric GARCH models under Gaussianity. They however maintain the unconditional homoskedasticity assumption. More recently, Cavaliere, Nielsen, and Taylor (2015b) extend the consistency and asymptotic normality properties of the conditional sum-of-square estimators proposed by Hualde and Robinson (2011) to include both conditional and unconditional heteroskedasticity of a very general and unknown form.

The paper is structured as follows. In Section 2 we introduce the ARFIMA model and the general model of heteroskedasticity for the disturbances, and derive score tests for the fractional differencing parameter. The score test based on a Gaussian likelihood with known variance process is shown to be asymptotically efficient. A robust score test based on a quasi likelihood that imposes a constant variance is derived and shown to be asymptotically inefficient. In Section 3 we provide the main result of the paper, which is that a feasible test, based on re-weighting using a nonparametric variance estimator, is asymptotically equivalent to the efficient score test. This new test is shown to have superior asymptotic local power properties to the robust test, and hence to the robust tests of Kew and Harris (2009) and CNT. These properties are evaluated in finite samples in Section 4, where it is shown that the new re-weighted test can achieve substantial power gains over robust tests for certain patterns of heteroskedasticity. Section 5 concludes with some possible directions for future research. Proofs are collected in Appendix A and additional results required for these proofs are available in the online supplement (Harris and Kew (2016)). In the following, $\xrightarrow{p}$ denotes (weak) convergence in probability and $\rightsquigarrow$ denotes convergence in distribution.

2. INFEASIBLE TESTS

Suppose the observed time series z_t satisfies

$$\Delta^d z_t = y_t,$$

where d is a known differencing parameter of any value, Δ^d is the Type II fractional differencing operator[1]

$$\Delta^d = \sum_{j=0}^{t-1} \Gamma(j-d) / (\Gamma(j+1)\Gamma(-d)) L^j,$$

and y_t follows an ARFIMA process of the form

$$a(L;\psi_0)\Delta^{\theta_0} y_t = e_t, \tag{1}$$

where

$$a(L;\psi) = \sum_{j=0}^{\infty} a_j(\psi) L^j = \frac{\phi(L)}{\eta(L)} \tag{2}$$

is a rational lag polynomial defined in terms of an autoregressive component $\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^j$ and a moving average component $\eta(L) = 1 - \sum_{j=1}^{q} \eta_j L^j$

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

of known fixed orders p and q respectively. In (2) the parameter vector ψ is $\psi = \left(\phi_1, \ldots, \phi_p, \eta_1, \ldots, \eta_q\right)'$, and θ_0 and ψ_0 in (1) denote the true values of the parameters. Define the full parameter vector $\gamma = \left(\theta, \psi'\right)'$ on a parameter space satisfying the following assumption, which is the same one made in CNT.

Assumption R. The true values (θ_0, ψ_0) lie in the interior of a convex, compact parameter space $\Gamma = \Theta \times \Psi$, such that for all $\psi \in \Psi$, the polynomial functions $\phi(L)$ and $\eta(L)$ have no common roots and all their roots lie strictly outside the unit circle.

As in Robinson (1991, 1994), Tanaka (1999) and Nielsen (2004), we wish to test

$$H_0 : \theta_0 = 0,$$

against

$$H_1^L : \theta_0 < 0 \text{ or } H_1^U : \theta_0 > 0,$$

which is equivalent to testing the null hypothesis that z_t is $I(d)$ for the known value of d. Henceforth we discuss the testing problem in terms of the observable time series y_t. It can equivalently be considered as a specification test of the choice of d for the original time series z_t.

The disturbance term e_t in (1) is assumed to have the heteroskedastic specification

$$e_t = \sigma_t \varepsilon_t, \; t = 1, 2, \ldots \tag{3}$$

where σ_t^2 is the unconditional variance, with $e_t = 0$ for $t \leq 0$. We do not assume a specific parametric functional form for σ_t^2. In this section σ_t will be treated as known, with a feasible nonparametric estimator of σ_t given in the next section. For the purposes of the likelihood-based efficiency theory in this section it will be assumed that

$$\varepsilon_t \sim \text{i.i.d.} N(0, 1), \tag{4}$$

although this can be weakened for some subsequent asymptotic results, see Assumption E below.

For simplicity, the model in (1) ignores any nonstochastic variables (x_t) such as an unknown mean and trend terms. CNT Remark 2.3 provides a detailed discussion about how x_t can be taken into account; see also Robinson (1994), Tanaka (1999) and Nielsen (2004).

2.1. Scores

The log-likelihood under (4) is

$$L(\gamma) = \text{constant} + \sum_{t=1}^{T} l_t(\gamma), \tag{5}$$

where

$$l_t(\gamma) = -\frac{1}{2}\left(\frac{e_t(\gamma)}{\sigma_t}\right)^2, \quad e_t(\gamma) = a(L;\psi)\,\Delta^{\theta} y_t.$$

Denote the score vector as $s_t(\gamma) = \left(s_{\theta,t}(\gamma), s_{\psi,t}(\gamma)'\right)'$ where

$$s_{\theta,t}(\gamma) = \frac{\partial l_t(\gamma)}{\partial \theta} = -\frac{a(L;\psi)\,\Delta^{\theta}(\ln\Delta)\,y_t}{\sigma_t}\cdot\frac{e_t(\gamma)}{\sigma_t}, \tag{6}$$

$$s_{\psi,t}(\gamma) = \frac{\partial l_t(\gamma)}{\partial \psi} = -\frac{a_{\psi}(L;\psi)\,\Delta^{\theta} y_t}{\sigma_t}\cdot\frac{e_t(\gamma)}{\sigma_t}, \tag{7}$$

in which $\ln\Delta = -\sum_{j=1}^{t-1} j^{-1}L^j$ and $a_{\psi}(L;\psi) = \sum_{j=1}^{\infty} a_{\psi,j}(\psi)\,L^j$ with $a_{\psi,j}(\psi) = \partial a_j(\psi)/\partial\psi$. It is shown in Section 2.2 that an asymptotically efficient test of H_0 against H_1^U or H_1^L is based on these scores.

For comparison purposes, define the quasi log-likelihood function

$$K(\gamma) = \sum_{t=1}^{T} k_t(\gamma), \quad k_t(\gamma) = -\frac{1}{2}e_t(\gamma)^2,$$

which includes no weights to allow for heteroskedasticity. The "quasi-score" vector is similarly denoted $r_t(\gamma) = \left(r_{\theta,t}(\gamma), r_{\psi,t}(\gamma)'\right)'$ where $r_{\theta,t}(\gamma) = \partial k_t(\gamma)/\partial\theta$ and $r_{\psi,t}(\gamma) = \partial k_t(\gamma)/\partial\psi$. These unweighted scores provide the basis for the test statistics of Robinson (1994) and Tanaka (1999) derived under homoskedastic errors. CNT Theorem 1 shows that these homoskedastic score tests suffer from asymptotic size distortions in the presence of both conditional and unconditional heteroskedasticity. To resolve this problem, CNT propose a wild bootstrap method for these score tests and show that their testing procedure is robust to both conditional and unconditional heteroskedasticity of a very general and unknown form.

Our asymptotic distribution theory follows from a Central Limit Theorem for the scores, which will be shown to hold under the following assumptions on the components of e_t.

Assumption E. $\{\varepsilon_t\}$ is a martingale difference sequence that satisfies: (i) $E\left(\varepsilon_t^2\right) = 1$; (ii) $\tau_{r,s} = E\left(\varepsilon_t^2\varepsilon_{t-r}\varepsilon_{t-s}\right)$ 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 $r_1,\ldots,r_{q-2} \geq 1$, the q'th order cumulants $\kappa_q\left(t,t,t-r_1,\ldots,t-r_{q-2}\right)$ of $\left(z_t, z_t, z_{t-r_1},\ldots,z_{t-r_{q-2}}\right)$ satisfy the requirement that

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

(iv) $E\left(\varepsilon_t^2|\mathcal{F}_{t-1}\right) = 1$; and (v) $E\left(\varepsilon_t^4|\mathcal{F}_{t-1}\right) = \tau_{0,0}$ where $\mathcal{F}_t$ is the σ-field of events generated by $\varepsilon_s, s \leq t$.

Assumption S. σ_t satisfies $\sigma_t = \sigma(t/T)$, where $\sigma(.)$ is a nonstochastic function with at most a finite number of points of discontinuity; moreover $\sigma(.)$ is a measurable function on the interval $(0,1]$ such that $0 < \inf_{r\in(0,1]}\sigma(r) \le \sup_{r\in(0,1]}\sigma(r) < \infty$, and $\sigma(r)$ satisfies a (uniform) first-order Lipschitz condition except at the points of discontinuity.

Assumption E (i) to (iii) are the same as CNT Assumption $\mathcal{V}$ (b), while the remaining assumptions are made in Phillips and Xu (2006) and Xu and Phillips (2008) in the context of autoregressive models, and Hualde and Robinson (2011) and Nielsen (2015) in the context of fractionally integrated models. Assumption S, which was first introduced by Cavaliere (2004a), allows for a single structural break or multiple breaks in the volatility of the observed series z_t. It also allows for smooth transition instead of abrupt variance breaks as well as linear or nonlinear trending variances.

Define $\gamma_0 = (0, \psi_0')'$ to be the parameter vector under H_0, and define the lag polynomial $b(L;\psi) = a_\psi(L;\psi)/a(L;\psi_0) = \sum_{j=1}^{\infty} b_j(\psi) L^j$. The following Lemma gives a joint Central Limit Theorem for $s_t(\gamma_0)$ and $r_t(\gamma_0)$.

LEMMA 1. *Under H_0 and Assumptions E and S*

$$T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} s_t(\gamma_0) \\ r_t(\gamma_0)\end{pmatrix} \rightsquigarrow N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix} V & V\cdot\int_0^1 \sigma^2(s)\,ds \\ V\cdot\int_0^1 \sigma^2(s)\,ds & V\cdot\int_0^1 \sigma^4(s)\,ds\end{pmatrix}\right),$$

where

$$V = \begin{pmatrix} V_{\theta\theta} & V_{\theta\psi} \\ V_{\psi\theta} & V_{\psi\psi}\end{pmatrix} = \begin{pmatrix} \pi^2/6 & \sum_{j=1}^{\infty} j^{-1} b_j(\psi_0)' \\ \sum_{j=1}^{\infty} j^{-1} b_j(\psi_0) & \sum_{j=1}^{\infty} b_j(\psi_0) b_j(\psi_0)'\end{pmatrix}. \tag{8}$$

This lemma is fundamental to our subsequent asymptotic theory and the form of V is also required for the definitions of the effective score tests that now follow.

2.2. Effective score tests

Choi, Hall, and Schick (1996, hereafter CHS) provide a general optimality theory of hypothesis testing in likelihood-based models with unknown nuisance parameters. We follow their approach in deriving an infeasible test as a function of the nuisance parameter, but defining asymptotic efficiency in a manner that anticipates the estimation of the nuisance parameter in Section 3. CHS show that an asymptotically efficient test against a one-sided alternative uses the effective score test statistic, which in our case is

$$\xi_T = \frac{T^{-1/2}\sum_{t=1}^{T} s_{\theta|\psi,t}(\gamma_0)}{V_{\theta\theta|\psi}^{1/2}}, \tag{9}$$

where

$$s_{\theta|\psi,t}(\gamma) = s_{\theta,t}(\gamma) - s_{\psi,t}(\gamma)' V_{\psi\psi}^{-1} V_{\psi\theta}$$

is the *effective score* (as defined by CHS and previously by Hall and Mathiason (1990)) and

$$V_{\theta\theta|\psi} = V_{\theta\theta} - V_{\theta\psi} V_{\psi\psi}^{-1} V_{\psi\theta}$$

is its asymptotic variance. CHS prove that if the log-likelihood has the LAN (Locally Asymptotically Normal) property then an asymptotically efficient test is based on ξ_T. The effective quasi-score test statistic can be defined as

$$\zeta_T = \frac{T^{-1/2} \sum_{t=1}^{T} r_{\theta|\psi,t}(\gamma_0)}{\sqrt{V_{\theta\theta|\psi} \int_0^1 \sigma(s)^4 ds}}, \tag{10}$$

where

$$r_{\theta|\psi,t}(\gamma_0) = r_{\theta,t}(\gamma_0) - r_{\psi,t}(\gamma_0)' V_{\psi\psi}^{-1} V_{\psi\theta}.$$

The lack of weighting for heteroskedasticity in this statistic results in an inefficient test relative to ξ_T. This is all formalised in the following theorem.

THEOREM 2. *Define the local sequence* $\gamma_T = \gamma_0 + T^{-1/2} g$ *for a fixed finite vector* $g = \left(g_\theta, g_\psi'\right)'$.
(a) Under γ_0 *and Assumptions R, E, S, the log-likelihood* $L(\gamma)$ *admits the LAN representation*

$$L(\gamma_T) - L(\gamma_0) = g' T^{-1/2} \sum_{t=1}^{T} s_t(\gamma_0) + \frac{1}{2} g' V g + o_p(1), \tag{11}$$

and the effective score and quasi-score statistics satisfy

$$\xi_T, \zeta_T \rightsquigarrow N(0,1).$$

The score tests reject H_0 *against* H_1^U *for* $\xi_T, \zeta_T > z_\alpha$, *where* z_α *is the* $100(1-\alpha)\%$ *percentile of the standard normal distribution, and similarly for the lower tailed tests.*
(b) Under γ_T *and Assumptions R, S and* $\varepsilon_t \sim$ i.i.d.$N(0,1)$, *the statistics satisfy*

$$\xi_T \rightsquigarrow N\left(g_\theta V_{\theta\theta|\psi}^{1/2}, 1\right), \tag{12}$$

$$\zeta_T \rightsquigarrow N\left(g_\theta V_{\theta\theta|\psi}^{1/2} \nu, 1\right), \tag{13}$$

where $\nu = \int_0^1 \sigma(s)^2 ds \Big/ \sqrt{\int_0^1 \sigma(s)^4 ds}$, *and the* ξ_T *test is asymptotically efficient.*

Theorem 2(a) extends the LAN property for ARFIMA models in Proposition 1 of Hallin, Taniguchi, Serroukh, and Choy (1998) to allow for heteroskedasticity of a known form. Similarly, the rest of the Theorem extends the asymptotic efficiency results of Robinson (1994) and Tanaka (1999) for the score test to

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

include heteroskedasticity of a known form. The asymptotic distributions under local alternatives in (12) and (13) are given under $\varepsilon_t \sim$ i.i.d.$N(0,1)$ but those results could be generalised, as in CNT, to allow ε_t to satisfy Assumption E. The asymptotic efficiency property requires Gaussianity in our Theorem because the distributions under local alternatives are deduced via Le Cam's third lemma, a very convenient and elegant device that exploits the LAN property in part (a). The study of efficient tests under non-Gaussianity is left to future research.

It follows from (13) that the effective quasi-score test ζ_T is asymptotically inefficient, with loss of power relative to ξ_T determined by the constant $\upsilon \leq 1$, with $\upsilon = 1$ under homoskedasticity. The smaller the value of υ, for a given value of g_θ, the ζ_T test suffers the larger power loss and this is illustrated in Kew and Harris (2009) Corollary 2 and Figure 2, CNT Figures 2 and 3, and CNT Remark 3.1. Xu and Phillips (2008) derive explicit expressions for υ under a single shift variance model (see their Example 1) and a trending variance model (Example 2). Under a single downward shift variance model, it is clear that the quasi-score ζ_T test suffers from substantial asymptotic power loss relative to the efficient score ξ_T test when this downward shift occurs early in the sample. Finite-sample power loss due to this variance model is also reflected in our Monte Carlo simulation results presented in Section 4.

3. FEASIBLE TESTS

In this section we propose feasible versions of the ξ_T and ζ_T tests. For the quasi-score ζ_T test in (10), ψ_0 is unknown and so we define the quasi-MLE $\hat{\psi}$ under the null as

$$\hat{\psi} = \arg \max_{\psi \in \Psi,\, \theta = 0} K(\gamma),$$

and $\hat{\gamma} = \left(0, \hat{\psi}'\right)'$. That is, $\hat{\psi}$ is the standard ARMA coefficient estimator assuming a constant variance, shown in CNT Lemma A.1 to be consistent under the null. The first-order condition for $\hat{\psi}$ is $\sum_{t=1}^{T} r_{\psi,t}\left(\hat{\gamma}\right) = 0$ and hence $T^{-1/2}\sum_{t=1}^{T} r_{\theta|\psi,t}\left(\hat{\gamma}\right) = T^{-1/2}\sum_{t=1}^{T} r_{\theta,t}\left(\hat{\gamma}\right)$ when substituting $\hat{\gamma}$ for γ_0 in the numerator of ζ_T. A feasible denominator for ζ_T is found by defining the estimated variance matrix

$$\hat{W} = \begin{pmatrix} \hat{W}_{\theta\theta} & \hat{W}_{\theta\psi} \\ \hat{W}_{\psi\theta} & \hat{W}_{\psi\psi} \end{pmatrix} = T^{-1}\sum_{t=1}^{T} r_t\left(\hat{\gamma}\right) r_t\left(\hat{\gamma}\right)',$$

and $\hat{W}_{\theta\theta|\psi} = \hat{W}_{\theta\theta} - \hat{W}_{\theta\psi}\hat{W}_{\psi\psi}^{-1}\hat{W}_{\psi\theta}$. The feasible quasi-score statistic is then

$$\hat{\zeta}_T = \frac{T^{-1/2}\sum_{t=1}^{T} r_{\theta,t}\left(\hat{\gamma}\right)}{\hat{W}_{\theta\theta|\psi}^{1/2}}.$$

Theorem 3 below establishes that $\hat{\zeta}_T$ is asymptotically equivalent to its infeasible counterpart.

Turning to the efficient score ξ_T test in (9), both σ_t^2 and ψ_0 are unknown. Following the approach of Xu and Phillips (2008) we estimate σ_t^2 nonparametrically and then adaptively estimate ψ_0. Xu and Phillips (2008) deal with an unconditionally heteroskedastic AR model and propose an Adaptive Least Squares estimator that has the same asymptotic distribution as the infeasible Generalised Least Squares estimator. We show that their method can be extended to our ARFIMA testing framework. Specifically we construct, under H_0, $\hat{e}_t = a\left(L;\hat{\psi}\right)y_t$ and define the nonparametric variance estimator as

$$\hat{\sigma}_t^2 = \sum_{i=1}^{T} w_{ti}\hat{e}_i^2, \tag{14}$$

where $w_{ti} = \left(\sum_{i=1}^{T} K_{ti}\right)^{-1} K_{ti}$ and $K_{ti} = K\left(\frac{t-i}{Tb}\right)$, with $K(.)$ is a bounded nonnegative continuous kernel function defined on the real line such that $\int_{-\infty}^{\infty} K(z)\,dz = 1$, and b is a bandwidth parameter. Following Xu and Phillips (2008), we define $K_{ti} = 0$ if $t = i$, leaving out the t^{th} observation of $\hat{e}_t^2$ when estimating $\hat{\sigma}_t^2$. We use the cross validation method to select b; i.e., we calculate $CV(b) = T^{-1}\sum_{t=1}^{T}\left(\hat{e}_t^2 - \hat{\sigma}_t^2\right)^2$ for a range of values of b and select b^* such that $CV(b)$ is minimised.

The feasible log-likelihood is then defined by replacing σ_t in (5) with $\hat{\sigma}_t$ to give

$$\hat{L}(\gamma) = \text{constant} + \sum_{t=1}^{T}\hat{l}_t(\gamma), \quad \hat{l}_t(\gamma) = -\frac{1}{2}\left(\frac{e_t(\gamma)}{\hat{\sigma}_t}\right)^2.$$

Similarly the score vector $\hat{s}_t(\gamma) = \left(\hat{s}_{\theta,t}(\gamma), \hat{s}_{\psi,t}(\gamma)'\right)'$ is defined by replacing σ_t in (6) and (7) with $\hat{\sigma}_t$. Define the feasible MLE $\tilde{\psi}$ under the null as

$$\tilde{\psi} = \arg\max_{\psi\in\Psi,\,\theta=0} \hat{L}(\gamma),$$

giving $\tilde{\gamma} = \left(0, \tilde{\psi}'\right)'$. A feasible version of ξ_T is constructed similarly to $\hat{\zeta}_T$, exploiting $\sum_{t=1}^{T}\hat{s}_{\psi,t}(\tilde{\gamma}) = 0$ in the numerator of the statistic. A variance matrix estimator may be defined as

$$\tilde{V} = \begin{pmatrix} \tilde{V}_{\theta\theta} & \tilde{V}_{\theta\psi} \\ \tilde{V}_{\psi\theta} & \tilde{V}_{\psi\psi} \end{pmatrix} = T^{-1}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_t^2}\left.\frac{\partial e_t(\gamma)}{\partial\gamma}\right|_{\gamma=\tilde{\gamma}}\left.\frac{\partial e_t(\gamma)}{\partial\gamma'}\right|_{\gamma=\tilde{\gamma}},$$

based on the information equality holding once the likelihood has been weighted appropriately (asymptotically). An "outer product of gradients" estimator $\tilde{V} = T^{-1}\sum_{t=1}^{T}\hat{s}_t(\tilde{\gamma})\hat{s}_t(\tilde{\gamma})'$ can also be shown to be consistent. In either case we define $\tilde{V}_{\theta\theta|\psi} = \tilde{V}_{\theta\theta} - \tilde{V}_{\theta\psi}\tilde{V}_{\psi\psi}^{-1}\tilde{V}_{\psi\theta}$, and the feasible score statistic is

$$\tilde{\xi}_T = \frac{T^{-1/2}\sum_{t=1}^{T}\hat{s}_{\theta,t}(\tilde{\gamma})}{\tilde{V}_{\theta\theta|\psi}^{1/2}}.$$

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

To establish the asymptotic equivalence of $\tilde{\xi}_T$ with its infeasible counterpart, we require the following Assumption B, which is from Xu and Phillips (2008).

Assumption B. As $T \to \infty, b + 1/Tb^2 \to 0$.

The following theorem gives the main result of our paper.

THEOREM 3. *(a) Under H_0 and Assumptions S, R, E, and B,*

$$\hat{\zeta}_T - \zeta_T = o_p(1) \quad \textit{and} \quad \tilde{\xi}_T - \xi_T = o_p(1).$$

(b) These asymptotic equivalences also hold under γ_T and Assumptions S, R, B and $\varepsilon_t \sim$ i.i.d.$N(0,1)$.

The implication of this theorem is that the feasible tests $\hat{\zeta}_T$ and $\tilde{\xi}_T$ inherit the asymptotic properties of ζ_T and ξ_T respectively. In particular, the nonparametrically variance-weighted test $\tilde{\xi}_T$ is asymptotically efficient in the Gaussian model, and retains the same asymptotic properties as the correctly weighted test when ε_t is not Gaussian.

The asymptotic efficiency of $\tilde{\xi}_T$ that has been shown for unconditional heteroskedasticity is not expected to hold under conditional heteroskedasticity (that is if Assumption E(iv) were relaxed), and adaptation to the latter remains an open question in this context. Also CNT have shown that the wild bootstrap provides robust inference on long memory in the presence of conditional heteroskedasticity and the combination of their bootstrap with the kernel re-weighting developed here could be a productive topic for future research.

4. SIMULATION EVIDENCE

This section compares the finite sample size and power properties of the various tests described in Theorem 3 when σ_t^2 follows a one-time structural break model with $\sigma_t^2 = \beta_1^2$ for $t \le \lfloor \tau T \rfloor$ and $\sigma_t^2 = \beta_2^2$ for $t > \lfloor \tau T \rfloor$ for some $\tau \in (0,1)$. We set, without loss of generality, $\beta_1 = 1$. Let $\delta = \beta_2/\beta_1$ measure the size of the shift and, following Cavaliere (2004) and Cavaliere and Taylor (2007), we set $\delta = 1/3$ (downward variance shift) and $\tau = 0.2$ (early shift) and $\tau = 0.8$ (late shift). Simulation results for $\delta = 3$ (upward variance shift) are omitted since they are quite similar. For comparison purposes, we also give results for the homoskedastic case where $\delta = 1$. The innovation ε_t is generated using the `rndn` routine in Gauss. The sample sizes $T = 100, 400$ and the number of replications is 50000.

Following Tanaka (1999), the data generating process for y_t is $(1 - \phi_0 L)\Delta^{\theta_0} y_t = e_t$. We test $H_0 : \theta_0 = 0$ vs $H_1^L : \theta_0 < 0$ or $H_1^U : \theta_0 > 0$ and we report the null rejection percentages based on a 5% nominal level. We follow Tanaka (1999) and set the values for the AR coefficient $\phi_0 = 0, 0.6$ and -0.8. If $\phi_0 = 0$ we let θ_0 range between -0.2 and 0.2 in steps of 0.05 and if $\phi_0 \neq 0$ we let θ_0 range between -0.4 and 0.4 in steps of 0.1.

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

In the simulations, the feasible quasi $\hat{\zeta}_T$ test and efficient $\tilde{\xi}_T$ test are computed assuming that the true orders p and q are known. In practice CNT Remark 2.5 suggests that the orders p and q can be selected by employing the usual Schwarz information criterion. As for the $\tilde{\xi}_T$ test, the estimator $\hat{\sigma}_t^2$ in (14) is computed using the Gaussian kernel and the estimator $\tilde{V}$ is computed via the "outer product of gradients" method[2].

Table 1 reports the case when no autocorrelation is present (i.e., $\phi_0 = 0$). It shows that when the errors are homoskedastic ($\delta = 1$) the $\hat{\zeta}_T$ and $\tilde{\xi}_T$ tests display acceptable size properties. The efficient $\tilde{\xi}_T$ test does not yield any power gains over the quasi $\hat{\zeta}_T$ test and this is expected since ν in Theorem 2 is equal to 1. Also as expected, the empirical power of each test increases as T increases for a given θ_0, and the power increases as $|\theta_0|$ becomes large for a given T.

When σ_t is not constant because of an early downward variance shift with $\tau = 0.2$, both our proposed $\hat{\zeta}_T$ and $\tilde{\xi}_T$ tests display relatively good size properties. In all cases, the powers of the efficient score $\tilde{\xi}_T$ test clearly exceed those of the quasi-score $\hat{\zeta}_T$ test and these observed power gains are expected since, by Xu and Phillips (2008) Example 1, $\nu = 0.63$, which is far less than 1. By comparison, we consider a late variance shift with $\tau = 0.8$. The efficient score $\tilde{\xi}_T$ test no longer yields significant power gains over the quasi-score $\hat{\zeta}_T$ test and this too is expected since now $\nu = 0.92$, which is close to 1.

Table 2 reports the case when first order autocorrelation is present. Again $\tilde{\xi}_T$ continues to yield substantial power gains over $\hat{\zeta}_T$ under an early downward variance shift with $\tau = 0.2$. Under homoskedasticity ($\delta = 1$) there are very small differences in terms of size and power between $\tilde{\xi}_T$ and $\hat{\zeta}_T$. Results for $\tau = 0.8$ are

TABLE 1. Empirical size and power of tests when $\phi_0 = 0$

			$H_1: \theta_0 < 0$					$H_1: \theta_0 > 0$				
		T/θ_0	0	−0.05	−0.10	−0.15	−0.20	0	0.05	0.10	0.15	0.20
$\delta = 1$	$\hat{\zeta}_T$	100	5.60	14.95	31.48	52.90	74.26	3.88	14.40	33.92	56.41	75.96
		400	5.47	34.30	79.86	98.27	99.96	4.43	35.02	80.09	97.49	99.90
	$\tilde{\xi}_T$	100	5.55	14.73	31.16	52.42	73.75	3.97	14.26	33.21	55.39	74.78
		400	5.48	34.21	79.56	98.21	99.95	4.42	34.81	79.75	97.44	99.89
$\tau = 0.2$	$\hat{\zeta}_T$	100	5.62	11.56	19.63	30.47	43.38	3.57	10.23	22.01	38.97	56.88
		400	5.56	20.75	46.84	74.95	92.20	4.00	19.43	50.79	81.12	95.83
	$\tilde{\xi}_T$	100	5.57	14.47	30.07	50.05	70.57	3.96	14.60	33.94	56.68	75.55
		400	5.44	33.31	77.69	97.58	99.92	4.47	34.58	79.06	97.15	99.86
$\tau = 0.8$	$\hat{\zeta}_T$	100	5.58	14.03	28.56	47.81	68.14	3.84	13.71	31.46	53.15	72.60
		400	5.53	30.99	73.74	96.37	99.84	4.27	31.61	74.90	95.86	99.68
	$\tilde{\xi}_T$	100	5.50	14.34	29.62	49.63	70.30	4.02	14.31	33.04	55.08	74.15
		400	5.48	32.81	76.98	97.52	99.91	4.32	34.22	78.40	96.96	99.84

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

Table 2. Empirical size and power of tests when $\phi_0 \neq 0$

			$H_1 : \theta_0 < 0$					$H_1 : \theta_0 > 0$			
		T/θ_0	0	−0.1	−0.2	−0.3	−0.4	0	0.1	0.2	0.3	0.4
							$\phi_0 = 0.6$					
$\delta = 1$	$\hat{\zeta}_T$	100	7.16	11.14	17.01	25.34	36.96	3.55	7.27	10.88	16.04	25.07
		400	6.25	18.40	41.62	70.72	90.14	3.58	14.55	28.62	39.69	54.72
	$\tilde{\xi}_T$	100	7.23	11.21	16.93	25.22	36.54	3.59	7.26	10.72	15.55	24.16
		400	6.19	18.37	41.50	70.57	90.16	3.60	14.56	28.55	39.34	54.09
$\tau = 0.2$	$\hat{\zeta}_T$	100	5.04	8.70	13.92	20.31	26.92	4.17	10.29	14.09	16.60	20.98
		400	5.53	14.89	25.85	40.28	57.28	3.78	13.28	22.21	27.75	35.13
	$\tilde{\xi}_T$	100	8.16	12.44	17.82	24.58	33.16	3.91	7.95	10.74	13.25	18.37
		400	6.37	18.54	40.15	66.81	87.66	3.67	16.35	30.61	36.14	43.52
							$\phi_0 = -0.8$					
$\delta = 1$	$\hat{\zeta}_T$	100	6.41	31.44	71.61	94.67	99.41	3.48	28.73	68.93	91.44	98.27
		400	5.84	76.51	99.89	100.00	100.00	4.16	74.13	99.63	100.00	100.00
	$\tilde{\xi}_T$	100	6.51	31.11	71.11	94.38	99.37	3.54	28.40	68.07	90.74	98.01
		400	5.82	76.36	99.89	100.00	100.00	4.16	73.84	99.60	100.00	100.00
$\tau = 0.2$	$\hat{\zeta}_T$	100	7.33	23.43	46.33	68.55	84.16	3.68	20.50	51.60	78.94	92.78
		400	6.44	46.80	90.16	99.55	99.99	3.70	45.42	92.75	99.85	100.00
	$\tilde{\xi}_T$	100	6.62	30.00	67.85	92.19	98.85	3.74	30.00	69.97	91.65	98.19
		400	5.76	74.42	99.82	100.00	100.00	4.23	73.49	99.54	100.00	100.00

not reported since, like the previous $\phi_0 = 0$ case in Table 1 and as expected, there are very small differences in size and power between the two tests.

5. CONCLUSION

This paper proposes adaptive testing for the memory parameter under a parametric specification for the short memory component. A flexible alternative framework is the semi-parametric model, whereby the short memory component is estimated nonparametrically. Thus, in the presence of unconditional heteroskedasticity, the developments of robust and adaptive inference procedures for semi-parametric models of the Log-Periodogram (LP) regression (Robinson (1995a)), pooled LP (Shimotsu and Phillips (2002)), nonlinear LP (Sun and Phillips (2003)), Local Whittle (LW) estimator (Robinson (1995b), Velasco (1999), Phillips and Shimotsu (2004)), a modified LW estimator (Phillips (1999), Shimotsu and Phillips (2000)), exact LW estimator (Shimotsu and Phillips (2005, 2006), Shimotsu (2010)), local polynomial Whittle with noise estimator (Frederiksen, Nielsen, and Nielsen (2012)) and multivariate LW estimator (Shimotsu (2007) and Nielsen and Shimotsu (2007)) are important topics for the future.

NOTES

1. See equation (4) of Tanaka (1999) for the computation of this operator. Jensen and Nielsen (2014) propose a fast algorithm, based on a discrete Fourier transform, for computing this operator.

2. We do not report results for the homoskedastic S'_{T1} test in Tanaka (1999) because CNT demonstrate that, under a single downward varaince shift model, this S'_{T1} test, as expected, is severely oversized even when the sample size increases.

REFERENCES

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

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

Beare, B. (2008) Unit Root Testing with Unstable Volatility. Nuffield College Economics working papers, No. 2008–W06, University of Oxford.

Box, G.E. & D.A. Pierce (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509–1526.

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

Busetti, F. & A.M.R. Taylor (2003) Testing against stochastic trend in the presence of variance shifts. *Journal of Business and Economic Statistics* 21, 510–531.

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

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

Cavaliere, G. (2004b) Testing stationarity under a permanent variance shift. *Economics Letters* 82, 403–408.

Cavaliere, G., L. De Angelis, A. Rahbek, & A.M.R. Taylor (2015) A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic VAR models. *Oxford Bulletin of Economics and Statistics* 77, 106–128.

Cavaliere, G., D.I. Harvey, S.J. Leybourne, & A.M.R. Taylor (2011) Testing for unit roots in the presence of a possible break in trend and nonstationary volatility. *Econometric Theory* 27, 957–991.

Cavaliere, G., D.I. Harvey, S.J. Leybourne, & A.M.R. Taylor (2015) Testing for unit roots under multiple possible trend breaks and non-stationary volatility using bootstrap minimum Dickey-Fuller statistics. *Journal of Time Series Analysis* 36, 603–629.

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

Cavaliere, G., M.Ø. Nielsen, & A.M.R. Taylor (2015b) Quasi-maximum likelihood estimation and bootstrap inference in fractional time series models with heteroskedasticity of unknown form. Working papers 1324, Queen's University, Department of Economics.

Cavaliere, G., P.C.B. Phillips, S. Smeekes, & A.M.R. Taylor (2015) Lag length selection for unit root tests in the presence of nonstationary volatility. *Econometric Reviews* 34, 512–536.

Cavaliere, G., A. Rahbek, & A.M.R. Taylor (2010a) Testing for co-integration in vector autoregressions with non-stationary volatility. *Journal of Econometrics* 158, 7–24.

Cavaliere, G., A. Rahbek, & A.M.R. Taylor (2010b) Cointegration rank testing under conditional heteroskedasticity. *Econometric Theory* 26, 1719–1760.

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

Cavaliere, G. & A.M.R. Taylor (2005) Stationarity tests under time-varying variances. *Econometric Theory* 21, 1112–1129.

Cavaliere, G. & A.M.R. Taylor (2006) Testing the null of co-integration in the presence of variance breaks. *Journal of Time Series Analysis* 27, 613–636.

Cavaliere, G. & 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. & A.M.R. Taylor (2008a) Testing for a change in persistence in the presence of non-stationary volatility. *Journal of Econometrics* 147, 84–98.

Cavaliere, G. & A.M.R. Taylor (2008b) Bootstrap unit root tests for time series with nonstationary volatility. *Econometric Theory* 24, 43–71.

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

Cavaliere, G. & A.M.R. Taylor (2009) Heteroskedastic time series with a unit root. *Econometric Theory* 25, 1228–1276.

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

Choi, S., W.J. Hall, & A. Schick (1996) Asymptotically uniformly most powerful tests in parametric and semiparametric models. *Annals of Statistics* 24, 841–861.

Chung, H. & J.Y. Park (2007) Nonstationary nonlinear heteroskedasticity in regression. *Journal of Econometrics* 137, 230–259.

Dalla, V., L. Giraitis, & P.C.B. Phillips (2015) Testing Mean Stability of Heteroskedastic Time Series. Cowles Foundation Discussion paper, No. 2006, Yale University.

Demetrescu, M. & C. Hanck (2012a) Unit root testing in heteroscedastic panels using the Cauchy estimator. *Journal of Business and Economic Statistics* 30, 256–264.

Demetrescu, M. & C. Hanck (2012b) A simple nonstationary-volatility robust panel unit root test. *Economics Letters* 117, 10–13.

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

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

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Frederiksen, P., F.S. Nielsen, & M.Ø. Nielsen (2012) Local polynomial Whittle estimation of perturbed fractional processes. *Journal of Econometrics* 167, 426–447.

Hall, W.J. & D.J. Mathiason (1990) On large-sample estimation and testing in parametric models. *International Statistical Review* 58, 77–97.

Hamori, S. & A. Tokihisa (1997) Testing for a unit root in the presence of a variance shift. *Economics Letters* 57, 245–253.

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

Harris, D. & H. Kew (2014) Portmanteau autocorrelation tests under q-dependence and heteroskedasticity. *Journal of Time Series Analysis* 35, 203–217.

Harris, D. & H. Kew (2016) Online Supplement to 'Adaptive Long Memory Testing under Heteroskedasticity'.

Harvey, A.C. & P.M. Robinson (1988) Efficient estimation of nonstationary time series regression. *Journal of Time Series Analysis* 9, 201–214.

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

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

Jensen, A.N. & M.Ø. Nielsen (2014) A fast fractional difference algorithm. *Journal of Time Series Analysis* 35, 428–436.

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

Johansen, S. & M.Ø. Nielsen (in press) The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory*.

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

Kim, T.H., S. Leybourne, & P. Newbold (2002) Unit root tests with a break in innovation variance. *Journal of Econometrics* 109, 365–387.

Kim, C.S. & J.Y. Park (2010) Cointegrating regressions with time heterogeneity. *Econometric Reviews* 29, 397–438.

Kitamura, Y., G. Tripathi, & H. Ahn (2004) Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72, 1667–1714.

Kuersteiner, G.M. (2002) Efficient IV estimation for autoregressive models with conditional heteroskedasticity. *Econometric Theory* 18, 547–583.

Li, W.K., S. Ling, & M. McAleer (2002) Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys* 16, 245–269.

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. & 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.

Ling, S. & W.K. Li (2001) Asymptotic inference for nonstationary fractionally integrated autoregressive moving-average models. *Econometric Theory* 17, 738–764.

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

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

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

McConnell, M.M. & G. Perez-Quiros (2000) Output fluctuations in the United States: What has changed since the early 1980s? *American Economic Review* 90, 1464–1476.

McLeod, A.I. & W.K. Li (1983) Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* 4, 269–73.

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

Nielsen, M.Ø. (2005) Multivariate Lagrange multiplier tests for fractional integration. *Journal of Financial Econometrics* 3, 372–398.

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

Nielsen, M.Ø. & K. Shimotsu (2007) Determining the cointegrating rank in nonstationary fractional systems by the exact local Whittle approach. *Journal of Econometrics* 141, 574–596.

Pagan, A.R. & G.W. Schwert (1990) Testing for covariance stationarity in stock market data. *Economics Letters* 33, 165–170.

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

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

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

Phillips, P.C.B. (1999) Discrete Fourier transforms of fractional processes. Cowles Foundation Discussion paper, No. 1243, Yale University.

Phillips, P.C.B. & K. Shimotsu (2004) Local Whittle estimation in nonstationary and unit root cases. *The Annals of Statistics* 32, 656–692.

Phillips, P.C.B. & K.-L. Xu (2006) Inference in Autoregression under heteroskedasticity. *Journal of Time Series Analysis* 27, 289–308.

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

Robinson, P.M. (1991) Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* 47, 67–84.

Robinson, P.M. (1994) Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420–1437.

Robinson, P.M. (1995a) Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* 23, 1048–1072.

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

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

Shimotsu, K. (2007) Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics* 137, 277–310.

Shimotsu, K. (2010) Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory* 26, 501–540.

Shimotsu, K. & P.C.B. Phillips (2000) Modified local Whittle estimation of the memory parameter in the nonstationary case. Cowles Foundation Discussion paper, No. 1265, Yale University.

Shimotsu, K. & P.C.B. Phillips (2002) Pooled log periodogram regression. *Journal of Time Series Analysis* 23, 57–93.

Shimotsu, K. & P.C.B. Phillips (2005) Exact local Whittle estimation of fractional integration. *The Annals of Statistics* 33, 1890–1933.

Shimotsu, K. & P.C.B. Phillips (2006) Local Whittle estimation of fractional integration and some of its variants. *Journal of Econometrics* 130, 209–233.

Smeekes, S. & A.M.R. Taylor (2012) Bootstrap union tests for unit roots in the presence of nonstationary volatility. *Econometric Theory* 28, 422–456.

Stărică, C. & C. Granger (2005) Nonstationarities in stock returns. *Review of Economics and Statistics* 87, 503–522.

Sun, Y. & P.C.B. Phillips (2003) Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics* 115, 355–389.

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

van Dijk, D., D.R. Osborn, & M. Sensier (2002) Changes in variability of the business cycles in the G7 countries. Econometric Institute Report EI 2002–28.

Velasco, C. (1999) Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis* 20, 87–127.

Watson, M.W. (1999) Explaining the increased variability in long-term interest rates. *Federal Reserve Bank of Richmond Economic Quarterly* 85, 71–96.

Westerlund, J. (2014) Heteroscedasticity robust panel unit root tests. *Journal of Business and Economic Statistics* 32, 112–135.

Xu, K.-L. (2008a) Testing against nonstationary volatility in time series. *Economics Letters* 101, 288–292.

Xu, K.-L. (2008b) Bootstrapping autoregression under non-stationary volatility. *The Econometrics Journal* 11, 1–26.

Xu, K.-L. (2012) Robustifying multivariate trend tests to nonstationary volatility. *Journal of Econometrics* 169, 147–154.

Xu, K.-L. (2013) Powerful tests for structural changes in volatility. *Journal of Econometrics* 173, 126–142.

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

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

Xu, K.-L. & J.-C. Yang (2015) Towards uniformly efficient trend estimation under weak/strong correlation and non-stationary volatility. *Scandinavian Journal of Statistics* 42, 63–86.
Xu, K.-L. (2015) Testing for structural change under non-stationary variances. *The Econometrics Journal* 18, 274–305.

APPENDIX A. Proofs of main results

Some additional results required for the following proofs are available in the online supplement (Harris and Kew (2016)). We first define some lag polynomials. Recalling $\psi = \left(\phi_1, \ldots, \phi_p, \eta_1, \ldots, \eta_q\right)'$, and the short run lag polynomial $a(L;\psi) = \phi(L)/\eta(L)$, the first derivative vector is the $p+q$ dimensional vector

$$a_\psi(L;\psi) = \frac{\partial a(L;\psi)}{\partial \psi} = \left(\frac{\partial a(L;\psi)}{\partial \phi'} \;\; \frac{\partial a(L;\psi)}{\partial \eta'}\right)'$$

where $\phi = \left(\phi_1, \ldots, \phi_p\right)'$ and $\eta = \left(\eta_1, \ldots, \eta_q\right)'$ and

$$\frac{\partial a(L;\psi)}{\partial \phi_k} = \frac{-1}{1 - \sum_{j=1}^{q} \eta_j L^j} \cdot L^k, \quad \frac{\partial a(L;\psi)}{\partial \eta_k} = \frac{1 - \sum_{j=1}^{p} \phi_j L^j}{\left(1 - \sum_{j=1}^{q} \eta_j L^j\right)^2} \cdot L^k.$$

The second derivative is the $(p+q) \times (p+q)$ matrix

$$a_{\psi\psi}(L;\psi) = \frac{\partial^2 a(L;\psi)}{\partial \psi \partial \psi'} = \begin{pmatrix} \frac{\partial^2 a(L;\psi)}{\partial \phi \partial \phi'} & \frac{\partial^2 a(L;\psi)}{\partial \phi \partial \eta'} \\ \frac{\partial^2 a(L;\psi)}{\partial \eta \partial \phi'} & \frac{\partial^2 a(L;\psi)}{\partial \eta \partial \eta'} \end{pmatrix}$$

in which

$$\frac{\partial^2 a(L;\psi)}{\partial \phi_k \partial \phi_h} = 0, \quad \frac{\partial^2 a(L;\psi)}{\partial \phi_k \partial \eta_h} = \frac{-1}{\left(1 - \sum_{j=1}^{q} \eta_j L^j\right)^2} \cdot L^{k+h},$$

$$\frac{\partial^2 a(L;\psi)}{\partial \eta_k \partial \eta_h} = \frac{2\left(1 - \sum_{j=1}^{p} \phi_j L^j\right)}{\left(1 - \sum_{j=1}^{q} \eta_j L^j\right)^3} \cdot L^{k+h}.$$

Then we can define

$$c_0(L;\gamma) = \frac{a(L;\psi)}{a(L;\psi_0)} \Delta^\theta$$

$$c_1(L;\gamma) = \frac{\partial c_0(L;\gamma)}{\partial \gamma} = \begin{pmatrix} \frac{a(L;\psi)}{a(L;\psi_0)} \Delta^\theta \ln \Delta \\ \frac{a_\psi(L;\psi)}{a(L;\psi_0)} \Delta^\theta \end{pmatrix}$$

$$c_2(L;\gamma) = \frac{\partial^2 c_0(L;\gamma)}{\partial \gamma \partial \gamma'} = \begin{pmatrix} \frac{a(L;\psi)}{a(L;\psi_0)} \Delta^\theta (\ln \Delta)^2 & \frac{a_\psi(L;\psi)'}{a(L;\psi_0)} \Delta^\theta \ln \Delta \\ \frac{a_\psi(L;\psi)}{a(L;\psi_0)} \Delta^\theta \ln \Delta & \frac{a_{\psi\psi}(L;\psi)}{a(L;\psi_0)} \Delta^\theta \end{pmatrix}.$$

Proof of Lemma 1. Under H_0 we have $y_t = a(L;\psi_0)^{-1} e_t$ and hence $e_t(\gamma) = c_0(L;\gamma) e_t$. Similarly $\partial e_t(\gamma)/\partial\gamma = c_1(L;\gamma) e_t$, giving

$$s_t(\gamma) = -\frac{1}{\sigma_t^2}\frac{\partial e_t(\gamma)}{\partial\gamma} e_t(\gamma) = -\frac{c_1(L;\gamma) e_t}{\sigma_t}\cdot\frac{c_0(L;\gamma) e_t}{\sigma_t}$$

$$r_t(\gamma) = -\frac{\partial e_t(\gamma)}{\partial\gamma} e_t(\gamma) = -c_1(L;\gamma) e_t \cdot c_0(L;\gamma) e_t.$$

Evaluating at $\gamma = \gamma_0$ gives $c_0(L;\gamma_0) = 1$ and

$$c_1(L;\gamma_0) = \left(\ln\Delta\ \frac{a_\psi(L;\psi_0)'}{a(L;\psi_0)}\right)' = \left(\ln\Delta\ b(L;\psi_0)'\right)'.$$

The Central Limit Theorem (CLT) in (A.11) of Lemma A.2 of CNT applies directly to $r_t(\gamma_0) = c_1(L;\gamma_0) e_t \cdot e_t$ but not to the weighted version $s_t(\gamma_0) = \sigma_t^{-1} c_1(L;\gamma_0) e_t \cdot \varepsilon_t$. However we can define

$$s_t^{\#}(\gamma_0) = c_1(L;\gamma_0)\varepsilon_t \cdot \varepsilon_t,$$

so that the reasoning leading to CNT's result (A.11) can be immediately applied jointly to $\left(s_t^{\#}(\gamma_0), r_t(\gamma_0)\right)$. It then remains (i) to check the form of the asymptotic variance in the CLT and (ii) to prove that

$$T^{-1/2}\sum_{t=1}^{T}\left(s_t^{\#}(\gamma_0) - s_t(\gamma_0)\right) \xrightarrow{p} 0. \tag{A.1}$$

(i) To derive the form of V, use

$$E\left[\begin{pmatrix} s_t^{\#}(\gamma_0) \\ r_t(\gamma_0)\end{pmatrix}\begin{pmatrix} s_t^{\#}(\gamma_0) \\ r_t(\gamma_0)\end{pmatrix}'\right] = \begin{pmatrix} \sum_{j=1}^{t-1} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)' & \sum_{j=1}^{t-1} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'\sigma_{t-j}\sigma_t \\ \sum_{j=1}^{t-1} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'\sigma_{t-j}\sigma_t & \sum_{j=1}^{t-1} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'\sigma_{t-j}^2\sigma_t^2 \end{pmatrix}.$$

Assumption S and similar arguments to Phillips and Xu (2006) Lemma A gives

$$T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)' \to \sum_{j=1}^{\infty} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'$$

$$T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'\sigma_{t-j}\sigma_t \to \sum_{j=1}^{\infty} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'\cdot\int_0^1 \sigma(s)^2\,ds$$

$$T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{t-1} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'\sigma_{t-j}^2\sigma_t^2 \to \sum_{j=1}^{\infty} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)'\cdot\int_0^1 \sigma(s)^4\,ds$$

in which, as required,

$$V = \sum_{j=1}^{\infty} c_{1,j}(\gamma_0) c_{1,j}(\gamma_0)' = \sum_{j=1}^{\infty}\begin{pmatrix} j^{-1} \\ b_j(\psi_0)\end{pmatrix}\begin{pmatrix} j^{-1} \\ b_j(\psi_0)\end{pmatrix}'.$$

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

(ii) To show (A.1), we write

$$T^{-1/2}\sum_{t=1}^{T}\left(s_t^{\#}(\gamma_0)-s_t(\gamma_0)\right)=T^{-1/2}\sum_{t=2}^{T}\left(\varepsilon_t\sum_{j=1}^{t-1}c_{1,j}(\gamma_0)\left(\frac{\sigma_{t-j}}{\sigma_t}-1\right)\varepsilon_{t-j}\right),$$

which is shown to satisfy (A.1) if we prove the generic result

$$r_{s,T}=T^{-1/2}\sum_{t=2}^{T}\left(\varepsilon_t\sum_{j=1}^{t-1}c_j\left(\frac{\sigma_{t-j}}{\sigma_t}-1\right)\varepsilon_{t-j}\right)\xrightarrow{p}0$$

for any coefficients c_j satisfying $\sum_{j=1}^{\infty}c_j^2<\infty$. Using that $\varepsilon_t\sim$ i.i.d. (0, 1), it follows that

$$\begin{aligned}E\left(r_{s,T}^2\right)&=T^{-1}\sum_{t=2}^{T}\sum_{j=1}^{t-1}c_j^2\left(\frac{\sigma_{t-j}}{\sigma_t}-1\right)^2\\&\le\frac{1}{\inf_r\sigma(r)^2}\sum_{j=1}^{T-1}c_j^2T^{-1}\sum_{t=j+1}^{T}\left(\sigma\left(\frac{t}{T}\right)-\sigma\left(\frac{t-j}{T}\right)\right)^2\\&=\frac{1}{\inf_r\sigma(r)^2}\sum_{j=1}^{T-1}c_j^2T^{-1}\sum_{t=j+1}^{T}\left(\sum_{i=t-j}^{t-1}\sigma\left(\frac{i+1}{T}\right)-\sigma\left(\frac{i}{T}\right)\right)^2.\end{aligned}\tag{A.2}$$

The proof that this disappears under Assumption S is given allowing for a single discontinuity in $\sigma(.)$ to illustrate, with extension to a finite number of discontinuities following identically. Suppose there is a single discontinuity at $\tau\in(0,1)$ such that $\lim_{r\downarrow\tau}\sigma(r)-\sigma(\tau)=\delta$, $0<\delta<\infty$. It follows that $\limsup_T\left|\sigma\left(\frac{\lfloor\tau T\rfloor+1}{T}\right)-\sigma\left(\frac{\lfloor\tau T\rfloor}{T}\right)\right|=\delta$, while for $i\neq\lfloor\tau T\rfloor$ the Lipschitz condition imposed in Assumption S implies that $\left|\sigma\left(\frac{i+1}{T}\right)-\sigma\left(\frac{i}{T}\right)\right|\le\frac{\ell}{T}$ for some $\ell<\infty$. Thus

$$\begin{aligned}&\sum_{t=j+1}^{T}\left(\sum_{i=t-j}^{t-1}\sigma\left(\frac{i+1}{T}\right)-\sigma\left(\frac{i}{T}\right)\right)^2\\&=\sum_{t=j+1}^{T}\left(\sum_{\substack{i=t-j\\i\neq\lfloor\tau T\rfloor}}^{t-1}\left(\sigma\left(\frac{i+1}{T}\right)-\sigma\left(\frac{i}{T}\right)\right)+1\left(t-j\le\lfloor\tau T\rfloor\le t\right)\left(\sigma\left(\frac{\lfloor\tau T\rfloor+1}{T}\right)-\sigma\left(\frac{\lfloor\tau T\rfloor}{T}\right)\right)\right)^2\\&\le2\sum_{t=j+1}^{T}\left(\sum_{\substack{i=t-j\\i\neq\lfloor\tau T\rfloor}}^{t-1}\left(\sigma\left(\frac{i+1}{T}\right)-\sigma\left(\frac{i}{T}\right)\right)\right)^2+2\left(\sigma\left(\frac{\lfloor\tau T\rfloor+1}{T}\right)-\sigma\left(\frac{\lfloor\tau T\rfloor}{T}\right)\right)^2\sum_{t=j+1}^{T}1_{t-j\le\lfloor\tau T\rfloor\le t}\\&\le2\sum_{t=j+1}^{T}\left(\frac{j\ell}{T}\right)^2+2j\left(\sigma\left(\frac{\lfloor\tau T\rfloor+1}{T}\right)-\sigma\left(\frac{\lfloor\tau T\rfloor}{T}\right)\right)^2\end{aligned}$$

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

Using this bound in (A.2) gives

$$E\left(r_{s,T}^2\right) \le \frac{2}{\inf_r \sigma(r)^2} \sum_{j=1}^{T-1} c_j^2 T^{-1} \left(\sum_{t=j+1}^{T} \left(\frac{j\ell}{T}\right)^2 + j\left(\sigma\left(\frac{\lfloor \tau T \rfloor + 1}{T}\right) - \sigma\left(\frac{\lfloor \tau T \rfloor}{T}\right)\right)^2 \right)$$
$$\le 2\left(T^{-1}\sum_{j=1}^{T-1} j c_j^2\right)\left(\ell^2 + \left(\sigma\left(\frac{\lfloor \tau T \rfloor + 1}{T}\right) - \sigma\left(\frac{\lfloor \tau T \rfloor}{T}\right)\right)^2\right) \to 0,$$

since $\sum_{j=1}^{\infty} c_j^2 < \infty$ implies the Cesaro sum $T^{-1}\sum_{j=1}^{T-1} j c_j^2 \to 0$. ■

Proof of Theorem 2. (a) The LAN representation is based on the standard mean value expansion

$$\lambda_T(g) = L(\gamma_T) - L(\gamma_0) = g' T^{-1/2} \sum_{t=1}^{T} s_t(\gamma_0) + \frac{1}{2} g' T^{-1} \sum_{t=1}^{T} h_t\left(\gamma_T^*\right) g,$$

where γ_T^* is a convex combination of γ_T and γ_0 and

$$h_t(\gamma) = \frac{\partial^2 l_t(\gamma)}{\partial\gamma\,\partial\gamma'} = -\frac{1}{\sigma_t^2}\left(e_t(\gamma)\frac{\partial^2 e_t(\gamma)}{\partial\gamma\,\partial\gamma'} + \frac{\partial e_t(\gamma)}{\partial\gamma}\frac{\partial e_t(\gamma)}{\partial\gamma'}\right).$$

Given Lemma 1, it remains to show that

$$T^{-1}\sum_{t=1}^{T} h_t\left(\gamma_T^*\right) \xrightarrow{p} -V.$$

Under H_0 we write $\partial^2 e_t(\gamma)/\partial\gamma\,\partial\gamma' = c_2(L;\gamma)e_t$, and hence

$$h_t(\gamma) = -\frac{1}{\sigma_t^2}\left(c_0(L;\gamma)e_t \cdot c_2(L;\gamma)e_t + (c_1(L;\gamma)e_t)(c_1(L;\gamma)e_t)'\right).$$

We define

$$h_t^{\#}(\gamma) = -\left(c_0(L;\gamma)\varepsilon_t \cdot c_2(L;\gamma)\varepsilon_t + (c_1(L;\gamma)\varepsilon_t)(c_1(L;\gamma)\varepsilon_t)'\right)$$

and show

$$T^{-1}\sum_{t=1}^{T}\left(h_t^{\#}(\gamma_T) - h_t(\gamma_T)\right) \xrightarrow{p} 0, \tag{A.3}$$

so that (A.12) of CNT applies to conclude the required convergence

$$T^{-1}\sum_{t=1}^{T} h_t^{\#}(\gamma_T) \xrightarrow{p} -V.$$

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

To show (A.3), we show that

$$T^{-1}\sum_{t=1}^{T}\left(h_t^{\#}(\gamma)-h_t(\gamma)\right)=-T^{-1}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}c_{0,j}(\gamma)\varepsilon_{t-j}\cdot\sum_{j=0}^{t-1}c_{2,j}(\gamma)\varepsilon_{t-j}\right.$$

$$\left.-\sum_{j=0}^{t-1}c_{0,j}(\gamma)\frac{\sigma_{t-j}}{\sigma_t}\varepsilon_{t-j}\cdot\sum_{j=0}^{t-1}c_{2,j}(\gamma)\frac{\sigma_{t-j}}{\sigma_t}\varepsilon_{t-j}\right) \tag{A.4}$$

$$-T^{-1}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}c_{1,j}(\gamma)\varepsilon_{t-j}\cdot\left(\sum_{j=0}^{t-1}c_{1,j}(\gamma)\varepsilon_{t-j}\right)'\right.$$

$$\left.-\sum_{j=0}^{t-1}c_{1,j}(\gamma)\frac{\sigma_{t-j}}{\sigma_t}\varepsilon_{t-j}\cdot\left(\sum_{j=0}^{t-1}c_{1,j}(\gamma)\frac{\sigma_{t-j}}{\sigma_t}\varepsilon_{t-j}\right)'\right)\xrightarrow{p}0 \tag{A.5}$$

uniformly on $\Gamma_h=\Theta_h\times\Psi$, where $\Theta_h=\left[-1,\frac{1}{2}-\varepsilon\right]$ for any $\varepsilon>0$, and Ψ satisfies Assumption R. This parameter space is large enough to accommodate γ_T as required in (A.3), at least for large enough T. We note that each element of the coefficients of $c_0(L;\gamma)$, $c_1(L;\gamma)$ and $c_2(L;\gamma)$ are square summable uniformly on Γ_h. Therefore each element of each term in (A.4) and (A.5) will be shown to satisfy the general convergence

$$T^{-1}\sum_{t=1}^{T}\left[\left(\sum_{j=0}^{t-1}c_{1,j}\left(\frac{\sigma_{t-j}}{\sigma_t}\right)\varepsilon_{t-j}\right)\left(\sum_{j=0}^{t-1}c_{2,j}\left(\frac{\sigma_{t-j}}{\sigma_t}\right)\varepsilon_{t-j}\right)\right.$$

$$\left.-\left(\sum_{j=0}^{t-1}c_{1,j}\varepsilon_{t-j}\right)\left(\sum_{j=0}^{t-1}c_{2,j}\varepsilon_{t-j}\right)\right]\xrightarrow{p}0, \tag{A.6}$$

where $c_{0,j}(\gamma)$ and the individual elements of $c_{1,j}(\gamma)$ and $c_{2,j}(\gamma)$ are represented as generic scalar coefficients $c_{1,j}$ and $c_{2,j}$ that satisfy $\sum_{j=0}^{\infty}c_{1,j}^2<\infty$ and $\sum_{j=0}^{\infty}c_{2,j}^2<\infty$ (the γ can be dropped from the generic notation because of the uniform square summability of the coefficients on Γ_h). This is sufficient for (A.5) and hence (A.3).

In (A.6) the convergence in probability of $T^{-1}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}c_{1,j}\varepsilon_{t-j}\right)\left(\sum_{j=0}^{t-1}c_{2,j}\varepsilon_{t-j}\right)$ to some limit is standard, while that of $T^{-1}\sum_{t=1}^{T}\left(\sum_{j=0}^{t-1}c_{1,j}\left(\frac{\sigma_{t-j}}{\sigma_t}\right)\varepsilon_{t-j}\right)\left(\sum_{j=0}^{t-1}c_{2,j}\left(\frac{\sigma_{t-j}}{\sigma_t}\right)\varepsilon_{t-j}\right)$ follows by similar arguments while exploiting $0<\inf_r\sigma(r)$ and $\sup_r\sigma(r)<\infty$. The proof of (A.6) therefore consists of verifying that both terms have the same probability limit. This follows by using

$$E\left(\sum_{j=0}^{t-1}c_{1,j}\left(\frac{\sigma_{t-j}}{\sigma_t}\right)\varepsilon_{t-j}\right)\left(\sum_{j=0}^{t-1}c_{2,j}\left(\frac{\sigma_{t-j}}{\sigma_t}\right)\varepsilon_{t-j}\right)=\sum_{j=0}^{t-1}c_{1,j}c_{2,j}\left(\frac{\sigma_{t-j}}{\sigma_t}\right)^2,$$

$$E\left(\sum_{j=0}^{t-1}c_{1,j}\varepsilon_{t-j}\right)\left(\sum_{j=0}^{t-1}c_{2,j}\varepsilon_{t-j}\right)=\sum_{j=0}^{t-1}c_{1,j}c_{2,j},$$

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

and arguing that the average difference between these converges to zero:

$$
\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}\sum_{j=0}^{t-1}c_{1,j}c_{2,j}\left(\sigma_{t-j}^2-\sigma_t^2\right)\\
&\le \frac{1}{\inf_r \sigma(r)^2}T^{-1}\sum_{j=0}^{T-1}\left|c_{1,j}\right|\left|c_{2,j}\right|\sum_{t=j+1}^{T}\left|\left(\sigma\left(\frac{t}{T}\right)^2-\sigma\left(\frac{t-j}{T}\right)^2\right)\right|\\
&\le \frac{2\sup_r \sigma(r)}{\inf_r \sigma(r)^2}T^{-1}\sum_{j=0}^{T-1}\left|c_{1,j}\right|\left|c_{2,j}\right|\sum_{t=j+1}^{T}\sum_{i=t-j}^{t-1}\left|\left(\sigma\left(\frac{i+1}{T}\right)-\sigma\left(\frac{i}{T}\right)\right)\right|\\
&\le \frac{2\sup_r \sigma(r)}{\inf_r \sigma(r)^2}T^{-1}\sum_{j=0}^{T-1}\left|c_{1,j}\right|\left|c_{2,j}\right|\sum_{t=j+1}^{T}\left(\left(\frac{j\ell}{T}\right)+j\left|\sigma\left(\frac{\lfloor\tau T\rfloor+1}{T}\right)-\sigma\left(\frac{\lfloor\tau T\rfloor}{T}\right)\right|\right)\\
&\le \frac{2\sup_r \sigma(r)}{\inf_r \sigma(r)^2}\left(T^{-1}\sum_{j=0}^{T-1}jc_{1,j}^2\right)^{1/2}\left(T^{-1}\sum_{j=0}^{T-1}jc_{2,j}^2\right)^{1/2}\left(\ell+\left|\sigma\left(\frac{\lfloor\tau T\rfloor+1}{T}\right)-\sigma\left(\frac{\lfloor\tau T\rfloor}{T}\right)\right|\right)\\
&\to 0.
\end{aligned}
$$

(b) Define the shorthand notation $\omega_\xi^2 = V_{\theta\theta|\psi}$, $\upsilon_2 = \int_0^1 \sigma(s)^2\, ds$ and $\upsilon_4 = \int_0^1 \sigma(s)^4\, ds$. The definitions of ξ_T and ζ_T and the LAN representation in (a) give

$$
\begin{pmatrix}\xi_T\\ \zeta_T\\ \lambda_T(g)-\frac{1}{2}g'Vg\end{pmatrix}
=\begin{pmatrix}\frac{1}{\omega_\xi} & -\frac{V_{\theta\psi}V_{\psi\psi}^{-1}}{\omega_\xi} & 0 & 0\\ 0 & 0 & \frac{1}{\omega_\xi\upsilon_4^{1/2}} & -\frac{V_{\theta\psi}V_{\psi\psi}^{-1}}{\omega_\xi\upsilon_4^{1/2}}\\ g_\theta & g_\psi' & 0 & 0\end{pmatrix}
T^{-1/2}\sum_{t=1}^{T}\begin{pmatrix}s_{\theta,t}(\gamma_0)\\ s_{\psi,t}(\gamma_0)\\ r_{\theta,t}(\gamma_0)\\ r_{\psi,t}(\gamma_0)\end{pmatrix}+o_p(1)
$$

$$
\rightsquigarrow N\left(\begin{pmatrix}0\\0\\ \frac{1}{2}g'Vg\end{pmatrix},\begin{pmatrix}1 & \upsilon_2/\upsilon_4^{1/2} & g_\theta\omega_\xi\\ \upsilon_2/\upsilon_4^{1/2} & 1 & g_\theta\omega_\xi\left(\upsilon_2/\upsilon_4^{1/2}\right)\\ g_\theta\omega_\xi & g_\theta\omega_\xi\left(\upsilon_2/\upsilon_4^{1/2}\right) & g'Vg\end{pmatrix}\right)
$$

by Lemma 1. The null distributions of ξ_T and ζ_T follow immediately. The distributions of ξ_T and ζ_T under γ_T then follow from Le Cam's third lemma. The asymptotic efficiency of the test based on ξ_T follows from Theorem 1 of CHS. ∎

Proof of Theorem 3. (a) The proof for $\hat{\zeta}_T$ is essentially a special case (with $\hat{\sigma}_t^2 = 1$) of that for $\hat{\xi}_T$, so we focus on the efficient test. It is shown in Lemma B of the online supplement that $\tilde{\psi} \xrightarrow{p} \psi_0$. For clarity we write (θ, ψ) for γ in the rest of the proof of this Theorem. Define the Hessian

$$
\hat{h}_t(\theta,\psi)=\begin{pmatrix}\hat{h}_{\theta\theta,t}(\theta,\psi) & \hat{h}_{\theta\psi,t}(\theta,\psi)\\ \hat{h}_{\psi\theta,t}(\theta,\psi) & \hat{h}_{\psi\psi,t}(\theta,\psi)\end{pmatrix}
$$

as the partitioned matrix of second derivatives of $\hat{l}_t(\theta,\psi)$. The mean value equality

$$
\hat{s}_{\psi,t}\left(0,\tilde{\psi}\right)=\hat{s}_{\psi,t}(0,\psi_0)+\hat{h}_{\psi\psi,t}(0,\psi^*)\left(\tilde{\psi}-\psi_0\right)
$$

in the first order conditions $\sum_{t=1}^{T} \hat{s}_{\psi,t}\left(0, \tilde{\psi}\right) = 0$ gives

$$\sqrt{T}\left(\tilde{\psi} - \psi_0\right) = -\left(T^{-1}\sum_{t=1}^{T}\hat{h}_{\psi\psi,t}\left(0,\psi^{*}\right)\right)^{-1} T^{-1/2}\sum_{t=1}^{T}\hat{s}_{\psi,t}\left(0,\psi_0\right).$$

Another mean value expansion in $\hat{s}_{\theta,t}\left(0, \tilde{\psi}\right)$ gives

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\hat{s}_{\theta,t}\left(0,\tilde{\psi}\right)\\
&\quad = T^{-1/2}\sum_{t=1}^{T}\hat{s}_{\theta,t}\left(0,\psi_0\right) + T^{-1}\sum_{t=1}^{T}\hat{h}_{\theta\psi,t}\left(0,\psi^{**}\right)\cdot\sqrt{T}\left(\tilde{\psi}-\psi_0\right)\\
&\quad = T^{-1/2}\sum_{t=1}^{T}\hat{s}_{\theta,t}\left(0,\psi_0\right) - T^{-1}\sum_{t=1}^{T}\hat{h}_{\theta\psi,t}\left(0,\psi^{**}\right)\left(T^{-1}\sum_{t=1}^{T}\hat{h}_{\psi\psi,t}\left(0,\psi^{*}\right)\right)^{-1}\\
&\qquad \times T^{-1/2}\sum_{t=1}^{T}\hat{s}_{\psi,t}\left(0,\psi_0\right).
\end{aligned}$$

In Lemma B of the online supplement, we show that

$$\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\left(\hat{s}_t\left(0,\psi_0\right) - s_t\left(0,\psi_0\right)\right) \xrightarrow{p} 0,\\
&\sup_{\psi\in\Psi} T^{-1}\sum_{t=1}^{T}\left(\hat{h}_t\left(0,\psi\right) - h_t\left(0,\psi\right)\right) \xrightarrow{p} 0.
\end{aligned} \tag{A.7}$$

Also the Hessian is stochastically equicontinuous. To see this, for any $\breve{\psi} \xrightarrow{p} \psi_0$ we have

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T}\left(h_t\left(0,\breve{\psi}\right) - h_t\left(0,\psi_0\right)\right)\\
&\quad = T^{-1}\sum_{t=1}^{T}\left(h_t\left(0,\breve{\psi}\right) - h_t^{\#}\left(0,\breve{\psi}\right)\right)\\
&\qquad + T^{-1}\sum_{t=1}^{T}\left(h_t^{\#}\left(0,\breve{\psi}\right) - h_t^{\#}\left(0,\psi_0\right)\right) - T^{-1}\sum_{t=1}^{T}\left(h_t\left(0,\psi_0\right) - h_t^{\#}\left(0,\psi_0\right)\right).
\end{aligned}$$

The first and third terms $\xrightarrow{p} 0$ follow from equation (A.5) and the second term $\xrightarrow{p} 0$ follows from CNT equation A.12. These, together with $\psi^{*}, \psi^{**} \xrightarrow{p} \psi_0$, are sufficient to conclude

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press

that

$$
\begin{aligned}
T^{-1/2}\sum_{t=1}^{T}\hat{s}_{\theta,t}\left(0,\tilde{\psi}\right) &= T^{-1/2}\sum_{t=1}^{T}s_{\theta,t}(0,\psi_0) \\
&\quad - T^{-1}\sum_{t=1}^{T}h_{\theta\psi,t}(0,\psi_0)\left(T^{-1}\sum_{t=1}^{T}h_{\psi\psi,t}(0,\psi_0)\right)^{-1} \\
&\quad \times T^{-1/2}\sum_{t=1}^{T}s_{\psi,t}(0,\psi_0)+o_p(1) \\
&= T^{-1/2}\sum_{t=1}^{T}s_{\theta|\psi,t}(0,\psi_0)+o_p(1).
\end{aligned}
$$

The consistency of $\tilde{V}$ is implied by the arguments that lead to (A.7). The "outer product of gradients" estimator is consistent by similar lengthy algebra. This proves asymptotic equivalence of the feasible and infeasible statistics under H_0.

(b) Le Cam's third lemma implies equivalence under γ_T for both $\hat{\zeta}_T$ and $\hat{\xi}_T$. That is, $\hat{\zeta}_T - \zeta_T \xrightarrow{p} 0$ and $\hat{\xi}_T - \xi_T \xrightarrow{p} 0$ imply that the joint distributions of $\left(\hat{\xi}_T, \hat{\zeta}_T, \lambda_T(g)\right)$ and $\left(\xi_T, \zeta_T, \lambda_T(g)\right)$ are asymptotically equivalent, so the conclusions of Theorem 2 apply to $\left(\hat{\xi}_T, \hat{\zeta}_T, \lambda_T(g)\right)$. ■

https://doi.org/10.1017/S0266466615000481 Published online by Cambridge University Press