Econometric Theory, **25**, 2009, 1228–1276. Printed in the United States of America.
doi:10.1017/S026646660809049X

HETEROSKEDASTIC TIME SERIES WITH A UNIT ROOT

GIUSEPPE CAVALIERE
University of Bologna

A.M. ROBERT TAYLOR
University of Nottingham

In this paper we provide a unified theory, and associated invariance principle, for the large-sample distributions of the Dickey–Fuller class of statistics when applied to unit root processes driven by innovations displaying nonstationary stochastic volatility of a very general form. These distributions are shown to depend on both the spot volatility and the integrated variation associated with the innovation process. We propose a partial solution (requiring any leverage effects to be asymptotically negligible) to the identified inference problem using a wild bootstrap–based approach. Results are initially presented in the context of martingale differences and are later generalized to allow for weak dependence. Monte Carlo evidence is also provided that suggests that our proposed bootstrap tests perform very well in finite samples in the presence of a range of innovation processes displaying nonstationary volatility and/or weak dependence.

1. INTRODUCTION

It is well known that the familiar Dickey–Fuller class of unit root statistics (see, among others, Dickey and Fuller, 1979, 1981) have limiting null distributions that are free of nuisance parameters in the presence of certain forms of heteroskedasticity in the innovation process. Specifically, they obtain the same limit law as was derived by Dickey and Fuller (1979) for the case of independent and identically distributed (i.i.d.) innovations provided the innovations form a martingale difference array (MDA) and satisfy a global stationarity condition. This condition is formally stated in equation (6) in Section 2, but in essence it requires the sample second moment of the innovations to be (asymptotically) stable over all possible fixed fractions of the sample. Examples where this condition holds include cases where the innovations either follow a generalized autoregressive conditionally heteroskedastic–type (GARCH-type) process satisfying certain moment/memory conditions, or are seasonally heteroskedastic, or have a variance

We are grateful to Bruce Hansen, Peter Phillips, Brendan Beare, Iliyan Georgiev, and four anonymous reviewers for helpful and constructive comments on earlier versions of this paper. Address correspondence to Robert Taylor, School of Economics, The Sir Clive Granger Building, University of Nottingham, Nottingham NG7 2RD, United Kingdom; e-mail: Robert.Taylor@nottingham.ac.uk.

 © 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

that forms a stationary and ergodic Markov chain; see Ling, Li, and McAleer (2003), Burridge and Taylor (2001), and Nelson, Piger, and Zivot (2001), respectively, for details.

A number of recent empirical studies question the reasonableness of the global homoskedasticity condition in practice. Kim and Nelson (1999) and McConnell and Perez Quiros (2000), among others, report evidence suggesting that the volatility of the U.S. gross domestic product (GDP) has declined over the past 20 years. Decreased (increased) volatility has been reported in both aggregate consumption and income, in aggregate employment, and in short-term (long-term) U.S. interest rates, in addition to for most of the real and price variables in the Stock and Watson (1999) data set; see Sensier and van Dijk (2004) and references therein. Considerable evidence against the constancy of unconditional variances in stock market returns and exchange-rate data has been reported; cf. Loretan and Phillips (1994). Hansen (1995) notes that empirical applications of autoregressive stochastic volatility (SV) models to financial data generally estimate the dominant root in the SV process to be close to one, so that volatility appears to be nonstationary.

In this paper we extend the corpus of unit root distribution theory to cover unit root processes driven by a very general class of nonstationary heteroskedastic innovations that need not satisfy the global homoskedasticity condition of (6). Among many other possibilities, our setup permits models of nonstationary SV, models with stochastic jumps in volatility, an autoregressive conditionally heteroskedastic (ARCH) model with near-integrated volatility, and nonlinear, nonstationary volatility. Although we primarily focus on stochastic models of volatility, models of deterministic volatility (such as single or multiple volatility shifts, trending volatility, smooth transition volatility) as were previously considered in the context of unit root testing by Cavaliere (2004), Cavaliere and Taylor (2007, 2008a), and Beare (2005), among others, are also covered by our setup in this paper. We show that the limiting unit root null distributions of both the normalized-bias and studentized Dickey–Fuller statistics in such cases are not free of nuisance parameters and can differ substantially (according to the underlying volatility process) from the limiting distributions that obtain for these statistics under global homoskedasticity. To solve the identified inference problem, at least within the class of SV processes considered, we discuss wild bootstrap versions of standard unit root tests. These are shown to yield (asymptotically) correctly sized inference even in the presence of stochastic, nonstationary volatility of the general form we consider.

A by-product of our analysis is to provide an invariance principle for the partial sums of such error processes. This builds upon the work of, in particular, Davidson (1994), de Jong and Davidson (2000), and Wooldridge and White (1988), who derive invariance principles for partial sums of processes whose average variance, when properly normalized, admits a deterministic limit. Our setup allows for a more general class of heteroskedastic errors including, e.g., nonstationary and nonergodic SV.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Our results also follow up on the Hansen (1995) results for regression under nonstationary volatility. Our work, however, differs from that of Hansen (1995) in at least two important respects. First, Hansen (1995) allows for some forms of time-series dependence in the regressors of a linear model but rules out autoregressive unit roots. Second, his theory deals with nonstationary, continuous SV, whereas we consider a much more general class of SV models, where, in particular, random volatility jumps are allowed. Other related work is considered by Phillips and Xu (2006), Xu and Phillips (2008), and Xu (2008), who show that nonconstant volatility can have a large impact on standard estimation and testing procedures in stable autoregressions. The paper by Xu (2008) is of particular relevance to the present study as Xu uses an assumption for the volatility process that is very similar to that considered in this paper.

It is also worth mentioning that in a recent paper, Boswijk (2005) discusses maximum likelihood estimation of mean zero AR(1) processes when the (possibly nonstationary) volatility changes smoothly over time and can be estimated consistently. In such cases, his approach represents an important complement to the wild bootstrap method proposed in this paper. However, although we allow for processes that can display a countable number of volatility shifts, Boswijk (2005) requires the volatility process to be continuous. Moreover, our analysis does not require the existence of a consistent estimator of the underlying spot volatility, nor is its applicability restricted to zero mean AR(1) processes.

The plan of the paper is as follows. In Section 2 we introduce our reference first-order autoregressive time-series model, detail the class of heteroskedastic MDA innovation processes under which we will work, and provide a number of examples of empirically important processes that lie in this class. We also briefly review the limiting behavior of the Dickey–Fuller statistics for the case where the innovations form a globally homoskedastic MDA. In Section 3 we derive generic limiting null distributions for the Dickey–Fuller normalized-bias and studentized statistics under these conditions and evaluate these for a number of important classes of nonstationary volatility. A wild bootstrap–based implementation of these tests is proposed in Section 4. We demonstrate that, providing that any leverage effects become negligible as the sample size diverges, the bootstrap unit root statistics share the same limiting null distributions and asymptotic local power functions as the original statistics and, hence, can facilitate (asymptotically) correctly sized inference on the unit root hypothesis in the presence of heteroskedastic innovations. Our main result is generalized to cover weakly dependent and heteroskedastic errors in Section 5. Section 6 concludes. Proofs are contained in a mathematical Appendix.

In what follows, $\overset{w}{\to}$ denotes weak convergence, $\overset{p}{\to}$ convergence in probability, and $\overset{w}{\to}_p$ weak convergence in probability (Giné and Zinn, 1990), in each case as the sample size diverges to positive infinity; $\mathbb{I}(\cdot)$ denotes the indicator function and $x := y$ ($x =: y$) indicates that x is defined by y (y is defined by x); $\lfloor \cdot \rfloor$ denotes the integer part of its argument; $\mathcal{D}_{\mathbb{R}^k}[0,1]$ is used to denote the space of

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

k-dimensional vectors of càdlàg functions on [0, 1], endowed with the Skorokhod topology. Finally, the notation

$$(a|b) := a(s) - \int_0^1 a(r)\,b(r)'\,dr \left(\int_0^1 b(r)\,b(r)'\,dr \right)^{-1} b(s) \tag{1}$$

is used to denote the Hilbert projection in $L_2[0, 1]$ of a onto the space orthogonal to b.

2. THE HETEROSKEDASTIC MODEL

Following Hansen (1995), we consider the random array $\{X_{T,t}, u_{T,t} : 1 \le t \le T\}$, where $X_{T,t}$ and $u_{T,t}$ are real-valued. The model of interest is the first-order autoregression:

$$(X_{T,t} - z'_{T,t}\beta) = \alpha(X_{T,t-1} - z'_{T,t-1}\beta) + u_{T,t}, \qquad t = 1, \ldots, T, \tag{2}$$

where, for some increasing array of sigma fields $\{\mathcal{I}_{T,t}, t \le T\}$ to which $X_{T,t}$ and $u_{T,t}$ are adapted, $u_{T,t}$ is a MDA, so that $\mathrm{E}\left(u_{T,t}|\mathcal{I}_{T,t-1}\right) = 0$, almost surely (a.s.). The (unobservable) conditional volatility, $\sigma_{T,t-1} := \sqrt{\mathrm{E}\left(u_{T,t}^2|\mathcal{I}_{T,t-1}\right)}$, is assumed to be strictly positive, a.s. In what follows, we will also make reference to the standardized error process, $\varepsilon_{T,t} := u_{T,t}/\sigma_{T,t-1}$. In (2), $z_{T,t}$ is a $(p \times 1)$, $p < T$, vector of deterministic components satisfying the following conditions (see Phillips and Xiao, 1998): there exist a scaling matrix δ_T and a bounded piecewise continuous function $Z(r)$ such that (a) $\delta_T z_{T,\lfloor T\cdot \rfloor} \to Z(\cdot)$ as $T \to \infty$ in the Skorokhod topology on [0, 1] and (b) $\int_0^1 Z(r)Z(r)'dr$ is positive definite. These conditions allow for a wide variety of possible forms for the deterministic component, including the pth-order trend function $z_{T,t} := (1, t, \ldots, t^p)'$, $0 \le p < \infty$. Finally, the initial value $X_{T,0}$ is assumed to be of $o_p(T^{1/2})$, unless otherwise stated.

Our focus in this paper is on testing the unit root null hypothesis $H_0 : \alpha = 1$ in (2) against the alternative $H_1 : |\alpha| < 1$. To that end let $\hat{\alpha}_T$ denote the ordinary least squares (OLS) slope estimator of α:

$$\hat{\alpha}_T := \left(\sum_{t=1}^{T} \hat{X}_{T,t}\hat{X}_{T,t-1} \right) \left(\sum_{t=1}^{T} \hat{X}_{T,t-1}^2 \right)^{-1} \tag{3}$$

with $\{\hat{X}_{T,t}\}$ the OLS residuals[1] from the regression of $X_{T,t}$ on $z_{T,t}$. As in Dickey and Fuller (1979, 1981), a test for $H_0 : \alpha = 1$ can be based on the normalized-bias statistic, $T\left(\hat{\alpha}_T - 1\right)$, and on the associated studentized statistic,

$$t_{DF} := \left(\sum_{t=1}^{T} \hat{X}_{T,t-1}^2 \right)^{1/2} \left(\hat{\alpha}_T - 1\right) / \hat{\sigma}_u,$$

where $\hat{\sigma}_u^2 := T^{-1} \sum_{t=1}^{T} (\hat{X}_{T,t} - \hat{\alpha}_T \hat{X}_{T,t-1})^2$.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Provided $\{u_{T,t}\}$ satisfies the conditions of, e.g., Theorem 3.2 of McLeish (1974, p. 625), the following MDA invariance principle (or functional central limit theorem [FCLT]) holds:

$$S_T^u(\cdot) := \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor T\cdot \rfloor} u_{T,t} \xrightarrow{w} \sigma B(\cdot), \tag{4}$$

where the finite and nonzero scalar, σ, can be either constant or random (cf. McLeish, 1974, p. 628; Hall, 1977), and under $H_0 : \alpha = 1$,

$$T\left(\hat{\alpha}_T - 1\right) \xrightarrow{w} \frac{\hat{B}_z(1)^2 - \hat{B}_z(0)^2 - 1}{2\int_0^1 \hat{B}_z(s)^2\, ds}, \quad t_{DF} \xrightarrow{w} \frac{\hat{B}_z(1)^2 - \hat{B}_z(0)^2 - 1}{2(\int_0^1 \hat{B}_z(s)^2\, ds)^{1/2}}, \tag{5}$$

where B is a standard Brownian motion and $\hat{B}_z := (B|Z)$ (cf. Phillips and Xiao, 1998).[2] The distributions in (5) are often referred to as *Dickey–Fuller* (or *unit root*) distributions.

The conditions placed before on the second moments of $\{u_{T,t}\}$ allow for some forms of heteroskedasticity but not others. Specifically, both (4) and (5) are based on, among other conditions, primitive conditions which entail that (see McLeish, 1974; Davidson, 1994, p. 450) the following *global homoskedasticity* condition holds:

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T} \sum_{t=\lfloor Ts\rfloor+1}^{\lfloor Ts'\rfloor} u_{T,t}^2 = \sigma^2\left(s' - s\right), \quad \text{all } s < s' \in [0,1]. \tag{6}$$

That is, the sample second moment of $u_{T,t}$ is required to be (asymptotically) stable over all possible fixed fractions of the data.[3]

Our interest in this paper lies in cases where $\{u_{T,t}\}$ fails the global homoskedasticity condition in (6) and what effects this has on the FCLT in (4) and the limiting distributions in (5). To that end we shall consider the following very general conditions on $\{\varepsilon_{T,t}, \sigma_{T,t}\}$.

Assumption 1. The standardized error process $\{\varepsilon_{T,t}, \mathcal{I}_{T,t}\}$ is a MDA satisfying $\sup_t \mathrm{E}\left(|\varepsilon_{T,t}|^p | \mathcal{I}_{T,t-1}\right) \le Q_\varepsilon < \infty$, a.s., for some $p > 2$.

Remark 2.1. Under Assumption 1 the normalized errors $\{\varepsilon_{T,t}\}$ form a uniformly square-integrable MDA with unit conditional variance. Consequently, the FCLT for MDAs (see, e.g., Hall and Heyde, 1980) ensures that

$$S_T(\cdot) := \frac{1}{T^{1/2}} \sum_{t=1}^{\lfloor T\cdot \rfloor} \varepsilon_{T,t} \xrightarrow{w} B(\cdot), \tag{7}$$

where B is a standard Brownian motion.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Assumption 2. For all $s \in [0, 1]$, let $\sigma_T(s) := \sigma_{T,\lfloor Ts \rfloor}$; hence, $\sigma_T(\cdot) \in \mathcal{D}_{\mathbb{R}}[0, 1]$. Then, for some strictly positive deterministic sequence $\{a_T\}$, in the Skorokhod topology on $\mathcal{D}_{\mathbb{R}}[0, 1]$

$$a_T^{-1}\sigma_T(\cdot) \xrightarrow{w} \Omega(\cdot) \in \mathcal{D}_{\mathbb{R}}[0, 1] \tag{8}$$

with $\int_0^1 \Omega(s)^2 ds > 0$, a.s.

In what follows we shall refer to series satisfying Assumptions 1 and 2 as *globally heteroskedastic* innovation processes. The class of models of heteroskedasticity satisfying Assumptions 1 and 2 is extremely wide. Indeed, the (stochastic) limiting variance process Ω is required to be only slightly more than square-integrable and bounded. Furthermore, Assumptions 1 and 2 do not necessarily require the unconditional variance $\mathrm{E}(\sigma_{T,t}^2)$ to be finite, even in the limit $T \to \infty$; cf. Example 4 later in this section. Although Assumptions 1 and 2 formally rule out some cases of globally homoskedastic sequences, such as time series with stationary GARCH or Markov-switching volatility, most of the results given in this paper continue to hold in the more general setup where $\sigma_{T,t} = \tilde{\sigma}_{T,t}\sqrt{h_{T,t}}$, with $\tilde{\sigma}_{T,t}$ satisfying Assumption 2 and $\sqrt{h_{T,t-1}}\varepsilon_{T,t} =: \tilde{\varepsilon}_{T,t}$ a globally homoskedastic MDA satisfying (6) with, without loss of generality, $\sigma = 1$. There are, however, important classes of processes that are not permitted under our assumptions, even in this more general setting. A leading example is the case where $u_{T,t}$ is in the domain of attraction of a stable law with index of stability a in the interval $(0, 2)$, denoted $u_{T,t} \in \mathcal{D}(a)$; see, among others, Chan and Tran (1989) and Phillips (1990). For instance, if $u_{T,t} \in \mathcal{D}(a)$ then even in the i.i.d. case the conditional variance $\sigma_{T,t-1}$ is not finite and $u_{T,t}$ cannot be expressed in the format $u_{T,t} = \sigma_{T,t-1}\varepsilon_{T,t}$ with $\sigma_{T,t-1}$, $\varepsilon_{T,t}$ satisfying Assumptions 1 and 2. A second example is given by the case where $u_{T,t}$ is a pure integrated GARCH process with strictly positive, fixed parameters; see, e.g. Nelson (1990a). In this case, as far as we are aware, even when $\varepsilon_{T,t}$ is Gaussian and i.i.d. the conditional volatility does not satisfy the weak convergence required in Assumption 2. As a final example, cases where the volatility is negligible in some subsample(s) relative to others are also not permitted; i.e., Assumption 2 requires the volatility to have the same order of magnitude over all possible fractions of the sample.

Some examples of empirically relevant processes that satisfy Assumption 2 but fail the global homoskedasticity condition, (6), are given subsequently, where unless stated otherwise, $a_T = 1$, for all T.[4]

Example 1. Continuous nonstationary SV models (Hansen, 1995)

Hansen's (1995) autoregressive SV model is based on the assumption that

$$\sigma_{T,t} = h\left(\omega_0 + \omega_{T1} S_{T,t}^*\right), \tag{9}$$

where $h(\cdot)$ is real and continuous ($h(\cdot) \geq b > 0$) and $S_{T,t}^* := \rho_{T1} S_{T,t-1}^* + \eta_{T,t}$, $\{\eta_{T,t}, \mathcal{I}_{T,t}\}$ being a uniformly square-integrable MDA with unit conditional

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

variance. Finally, $\omega_{T1} := \omega_1 / T^{1/2}$, $\rho_{T1} := (1 - c_1/T)$; ω_0, ω_1, and c_1 are finite, and $S^*_{T,0}$ is initialized at zero. As shown in Hansen (1995, Lem. A.3), $\sigma_{T,t}$ of (9) satisfies Assumption 2 with $\Omega(\cdot) = h\left(\omega_0 + \omega_1 J^*_c(\cdot)\right)$ where J^*_c is a standard Ornstein–Uhlenbeck (OU) process with mean reversion parameter c_1, i.e., $J^*_c(s) := \int_0^s e^{-c_1(s-r)} dB^*(r)$ (for $c_1 = 0$, J^*_c reduces to the standard Brownian motion B^*). Here, the limiting volatility process is stochastic and continuous.

Example 2. SV models with jumps

Although Hansen's SV model requires volatility to be driven by a continuous diffusion, this condition is not needed for our analysis to hold. Consider, e.g., the following nonstationary jump process for volatility: $\sigma_{T,t} = h\left(\omega_0 + \omega_1 S^*_{T,t}\right)$, $S^*_{T,t} := \sum_{i=1}^{t} \delta_{T,i}\eta_{T,i}$, $S^*_0 = 0$ a.s. where $\delta_{T,t}$ is a binary indicator that equals one if a volatility jump occurs at time t (zero otherwise), with the corresponding jump size $\eta_{T,t}$ being i.i.d. and independent of $\delta_{T,t}$. Hence, volatility is constant over time except for time periods when $\delta_{T,t} = 1$. If $\delta_{T,t}$ satisfies (Georgiev, 2008) $\sum_{t=1}^{\lfloor T\cdot \rfloor} \delta_{T,t} \to G(\cdot)$, G being a nondecreasing continuous function on $[0,1]$, then $S^*_{T,\lfloor T\cdot \rfloor}$ converges weakly to the compound Poisson process $C_G(\cdot)$ on $[0,1]$ with jump intensity parameter G; i.e., $C_G(\cdot) := \sum_{i=1}^{N(\cdot)} \eta_{T,i}$, where $N(\cdot)$ is a Poisson process on $[0,1]$ with intensity parameter G. Therefore, Assumption 2 is satisfied with $\Omega(\cdot) = h(\omega_0 + \omega_1 C_G(\cdot)) \in \mathcal{D}_{\mathbb{R}}[0,1]$, and the limiting volatility process is piecewise constant with the number of discontinuities given by $N(1)$.

Example 3. Non-Gaussian autoregressive SV models

Assume that volatility follows the autoregression $\sigma_{T,t} = \rho_{T1}\sigma_{T,t-1} + \delta_{T,t}\eta_{T,t}$, where $\delta_{T,t}$ and $\eta_{T,t}$ are as defined in Example 2. In addition, we require $\eta_{T,t}$ to be positive, a.s. If $\sigma_{T,0} = \sigma_0 \in (0,\infty)$ and ρ_{T1} is local-to-unity; i.e., $\rho_{T1} := (1 - c_1/T)$, it then follows from Georgiev (2008) that

$$\sigma_{T,\lfloor Ts \rfloor} \xrightarrow{w} \Omega(s) := \sigma_0 \exp(-c_1 s) + \int_0^s \exp(-c_1(s-r)) dC_G(r). \tag{10}$$

The right member of (10) is argued to be the leading special case of the non-Gaussian OU volatility model of Barndorff-Nielsen and Shephard (2001) where the homogeneous Lévy process driving volatility is a compound Poisson process with positive increments; see Roberts, Papaspiliopoulos, and Dellaportas (2004) and Duffie, Pan, and Singleton (2000). In contrast to the previous examples, in this case the volatility process displays sharp jumps that tail off exponentially.

Example 4. Near-integrated GARCH volatility

Certain conditional volatility models of the near-integrated GARCH(1,1) form also satisfy Assumption 2. Specifically, suppose that the conditional variance is generated according to the GARCH(1,1) recursion $\sigma^2_{T,t} = \omega_T + \phi_T u^2_{T,t} +$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

$\beta_T \sigma^2_{T,t-1}$, initialized at $\sigma^2_{T,0} \in (0, +\infty)$, with

$$\omega_T := \tfrac{\omega}{T}, \qquad \phi_T := \tfrac{\phi}{(2T)^{1/2}} \quad (\omega, \phi > 0),$$

$$\beta_T := 1 - \tfrac{\phi}{(2T)^{1/2}} - \tfrac{\theta}{T} \quad \left(\theta > -\tfrac{\phi^2}{2}\right) \tag{11}$$

so that $\theta = T(1 - \phi_T - \beta_T)$ controls the degree of near-integration; see Section 2 of Nelson (1990b). As discussed in Boswijk (2001), if Assumption 1 is strengthened by requiring $\{\varepsilon_{T,t}\}$ to be an i.i.d. sequence, then Theorem 3.5 in Nelson (1990b) implies that $\sigma^2_T(\cdot)$ satisfies Assumption 2 with Ω the diffusion process

$$d\Omega(s)^2 = (\omega - \theta\Omega(s)^2)\, ds + \phi\Omega(s)^2\, dB^*(s) \tag{12}$$

with $B^*(\cdot)$ a standard Brownian motion, this result holding for all $\theta > -\phi^2/2$. As in Nelson (1990b, p. 19), the unconditional distribution of $u_{T,t}$ in the near-integrated GARCH case is approximately student t with $(2 - 4\theta/\phi^2)$ degrees of freedom. Consequently, the unconditional second moment of $u_{T,t}$ does not exist unless θ is strictly positive; importantly, the result in (12) is, however, still valid where this is not the case. Notice that $\theta = 0$ where the GARCH parameters sum to unity; i.e., $\phi_T + \beta_T = 1$.

Example 5. Explosive volatility processes

Stochastically trending volatility (i.e., such that $\sigma_{T,t}$ diverges as T grows) is also allowed by choosing $\{a_T\}$ in Assumption 2 appropriately. A leading case is the polynomial "nonstationary nonlinear heteroskedastic" (NNH) time-series model of Park (2002a), where $\sigma_{T,t} := h\left(\bar{S}^*_{T,t}\right)$ with $h(x) := |\omega_0 + \omega_1 x|^\nu$ (ω_0, ω_1, and ν finite; $\nu > 0$) and $\bar{S}^*_{T,t}$ adapted to $\mathcal{I}_{T,t}$, and following the nonstationary integrated process $\bar{S}^*_{T,t} := \bar{S}^*_{T,t-1} + \eta_{T,t}$, where $\eta_{T,t}$ is a MDA with unit conditional variance. Because $T^{-1/2}\bar{S}^*_{T,\lfloor T\cdot\rfloor} \xrightarrow{w} \bar{B}^*(\cdot)$, $\bar{B}^*$ a standard Brownian motion, this model satisfies Assumption 2 with $a_T := T^{\nu/2}$ and $\Omega := \omega_1^\nu |\bar{B}^*|^\nu$. Note that our framework covers many (although not all) of the models falling within the NNH class defined in Park (2002a). For instance, the preceding result does not change if $h(\cdot)$ is an H-regular function with a polynomial limit homogeneous function; see Park and Phillips (2001).

Example 6. Deterministic volatility

In principle, any deterministic variance function that has a $\mathcal{D}_{\mathbb{R}}[0,1]$ limit as T diverges satisfies Assumption 2. This case has previously been analyzed by, among others, Cavaliere and Taylor (2007, 2008a). A leading example of this case is given by smooth transition variance (single or multiple) break models, an example of which is the function $\sigma^2_{T,t} := \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{S}_t$, $\mathbb{S}_t := (1 + \exp(-\tilde{\gamma}_T(t - \lfloor T\tau \rfloor)))^{-1}$ ($0 < \tau < 1$; $\tilde{\gamma}_T > 0$), which corresponds to a (single) smooth (logistic) transition in $\sigma^2_{T,t}$ from σ_0^2 to σ_1^2, with transition midpoint $\lfloor \tau T \rfloor$ and speed of transition $\tilde{\gamma}_T$, such that as $\tilde{\gamma}_T \to 0$, $\sigma^2_{T,t}$ changes

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

from σ_0^2 to σ_1^2 abruptly at time $\lfloor \tau T \rfloor$.[5] If the transition speed parameter follows a local drift, i.e., $\tilde{\gamma}_T := \gamma/T$, then Assumption 2 is satisfied with $\Omega(s)^2 := \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{S}(s), \mathbb{S}(s) := (1 + \exp(-\gamma(s-\tau)))^{-1}$. Also models with explosive deterministic volatility are allowed; e.g., polynomially trending volatility such as, e.g., $\sigma_{T,t} := \delta t^{\nu}$, $\nu > 0$, satisfies Assumption 2 with $a_T := T^{\nu}$ and $\Omega(s) := \delta s^{\nu}$.

3. THE HETEROSKEDASTIC UNIT ROOT DISTRIBUTION

In this section we establish the limiting unit root null distribution of the $T(\hat{\alpha}_T - 1)$ and t_{DF} statistics of Section 2 when the process (2) is driven by globally heteroskedastic innovations. To develop the asymptotic distribution theory for heteroskedastic autoregressions we need to introduce a further assumption.

Assumption 3. The weak convergence results in (7) and (8) hold jointly in the Skorokhod topology on $\mathcal{D}_{\mathbb{R}^2}[0,1]$.

Remark 3.1. Assumption 3 is satisfied, e.g., if $S_T(\cdot)$ and $\sigma_T(\cdot)$ are stochastically independent; see Billingsley (1968, pp. 26–27). However, independence is not a necessary condition for Assumption 3 to hold; cf. Example 1 in this section. Assumption 3 is satisfied by each of Examples 1–6 of Section 2.

Before presenting our main result we first state in Lemma 1 a FCLT for the partial sums of innovations that satisfy Assumptions 1–3. In what follows we use the notation $X(s-)$ to define the left limit at time s of a càdlàg process X.

LEMMA 1. *Suppose that $\{\varepsilon_{T,t}, \sigma_{T,t}\}$ satisfy Assumptions 1–3. Then the following invariance principle holds on the MDA $\{u_{T,t}\}$:*

$$\frac{1}{a_T T^{1/2}} \sum_{t=1}^{\lfloor T\cdot \rfloor} u_{T,t} \xrightarrow{w} \int_0^{\cdot} \Omega(r-)\, dB(r) =: m_{\Omega}(\cdot), \tag{13}$$

where $B(\cdot)$ and $\Omega(\cdot)$ are as defined in in (7) and (8), respectively.

Remark 3.2. The fundamental difference between the invariance principle in (13) and the conventional FCLT (see, e.g., (7)), is that the standard Brownian motion, B, appearing in the latter is replaced by the stochastic process m_{Ω}, a continuous local martingale with absolutely continuous quadratic variation, in the former. This process is well known in the probability literature and can be traced back to the so-called Dambis–Dubins–Schwartz theorem; see, among others Revuz and Yor (1991). In the SV literature (see Shephard, 2005, and references therein) the stochastic integral m_{Ω} is the standard formulation in continuous time for a univariate SV process, with $\Omega(s)$ termed the instantaneous or *spot volatility* of $m_{\Omega}(s)$ at time s and $\int_0^s \Omega(r)^2 dr$ the *integrated variation* of $m_{\Omega}(s)$. Notice that the integrated variation is bounded under Assumption 2.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Remark 3.3. Where the limiting volatility process is continuous (cf. Hansen, 1995, Assump. 2) the derivation of the FCLT in (13) can also be inferred from results in Hansen (1995, Lem. A.6). Lemma 1 therefore generalizes results in Hansen (1995) to also allow for cases where the volatility process is not continuous.

We now state our main theorem, which provides a generic representation for the limiting distributions of the $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF} statistics when applied to a random walk process driven by globally heteroskedastic innovations. This result constitutes a substantial generalization of previous results given by Cavaliere (2004) and Cavaliere and Taylor (2007) for the case of purely deterministic and nonexplosive volatility.

THEOREM 1. *Let* $\{X_{T,t}\}_0^T$ *be generated according to (2) with* $\alpha = 1$. *Then, under Assumptions 1–3* $\hat{\alpha}_T$ *of (3) and the associated studentized statistic,* t_{DF}, *satisfy*

$$T\left(\hat{\alpha}_T - 1\right) \xrightarrow{w} \frac{\hat{B}_{\Omega,z}(1)^2 - \hat{B}_{\Omega,z}(0)^2 - 1}{2\int_0^1 \hat{B}_{\Omega,z}(s)^2\, ds} =: Z_{\Omega,z,1}, \tag{14}$$

$$t_{DF} \xrightarrow{w} \frac{\hat{B}_{\Omega,z}(1)^2 - \hat{B}_{\Omega,z}(0)^2 - 1}{2\left(\int_0^1 \hat{B}_{\Omega,z}(s)^2\, ds\right)^{1/2}} =: Z_{\Omega,z,2}, \tag{15}$$

where[6] $\hat{B}_{\Omega,z} := (B_\Omega | Z)$ *with* B_Ω *the following stochastic process in* $\mathcal{D}_{\mathbb{R}}[0,1]$*:*

$$B_\Omega(s) := m_\Omega(s)\left(\int_0^1 \Omega(r)^2\, dr\right)^{-1/2}. \tag{16}$$

Remark 3.4. In general, the two processes forming B_Ω of (16), namely, Ω and B, need not necessarily be independent; see Example 1 in this section. Under independence, and conditionally on Ω, $B_\Omega(s)$ is Gaussian with zero mean and variance

$$\eta(s) := \left(\int_0^1 \Omega(r)^2 dr\right)^{-1} \int_0^s \Omega(r)^2 dr \tag{17}$$

so that B_Ω is therefore a time-change Brownian motion with directing process η; cf. Clark (1973). This attracting property does not, however, hold in general; see Shephard (2005, p. 7) for further discussion on this point.

A consequence of the FCLT in (13) is that the asymptotic distributions of the $T(\hat{\alpha}_T - 1)$ and t_{DF} statistics, under $\alpha = 1$, in (14) and (15) depend on both the spot volatility and the integrated (total) variation, reducing to the corresponding representations in (5) only when Ω is constant. We shall refer to (14) and (15)

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

as *heteroskedastic unit root* distributions in what follows. The result in (14) implies that, as in the standard case in (5), in the presence of a unit root $\hat{\alpha}_T$ is a T-consistent estimator despite the presence of global heteroskedasticity in the errors. Notice that this result holds even in the presence of trending volatility where, inter alia, $(X_{T,t} - z'_{T,t}\beta)$ is not necessarily of $O_p\left(T^{1/2}\right)$.

We now discuss some special cases of the generic representations given in Theorem 1, relating to Examples 1–6 of Section 2.

Example 1

In this case Assumption 3 holds (see Hansen, 1995), and B_Ω of Theorem 1 has the representation

$$B_\Omega(s) := \frac{\int_0^s h\left(\omega_0 + \omega_1 J_c^*(r)\right) dB(r)}{\left(\int_0^1 h\left(\omega_0 + \omega_1 J_c^*(r)\right)^2 dr\right)^{1/2}}, \tag{18}$$

where J_c^* is as defined in Section 2. Note that the Brownian motion B and the Brownian motion driving J_c^*, B^*, are not uncorrelated in general; cf. Remark 4.5 in the next section. In particular, the covariance between $B(1)$ and $B^*(1)$, say, $\overline{\omega}$, obtains as $\overline{\omega} := \lim_{T\to\infty} \sum_{t=1}^T \mathrm{E}\left(\varepsilon_{T,t}\eta_{T,t}\right)$. A sufficient condition for $\overline{\omega} = 0$ is that $\mathrm{E}\left(\varepsilon_{T,t}\eta_{T,t}|\mathcal{I}_{T,t-1}\right) = 0$, all t; i.e., current shocks to the levels and shocks to future volatility are conditionally uncorrelated. Notice also that, under $\overline{\omega} = 0$, B_Ω is a time-change Brownian motion with directing process $\eta(s) := \int_0^s h\left(\omega_0 + \omega_1 J_c^*(r)\right) dr \left(\int_0^1 h\left(\omega_0 + \omega_1 J_c^*(r)\right)^2 dr\right)^{-1}$.

Example 2

In the stochastic jump volatility model previously introduced, the weak convergence of $\sum_{t=1}^{\lfloor T\cdot\rfloor} \delta_{T,t}\eta_{T,t}$ and of $T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} \varepsilon_{T,t}$ is joint (see Georgiev, 2008), and hence Lemma 1 and Theorem 1 hold with B_Ω given as in (18) with J_c^* replaced by the compound Poisson process C_G. Notice that, although volatility displays a countable number of jumps, B_Ω is a.s. continuous.

Example 3

The non-Gaussian OU stochastic jump volatility model can be treated as in the previous example. Here B_Ω in (16) has Ω as given in (10).

Example 4

In the near-integrated GARCH case, the joint convergence requirement of Assumption 3 holds (see, e.g., Nelson, 1990b), so that Theorem 1 applies. Under the parametrization (11), the corresponding heteroskedastic unit root distribution is given by (14) with B_Ω as in (16) and Ω as given in (12), with B^* independent of B. Notice that in the GARCH framework, because B and B^* are independent, B_Ω can also be interpreted as a time-change Brownian motion.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Example 5

In the NNH explosive volatility example B_Ω is a time-change Brownian motion. Specifically, $\Omega(s) = \omega_1^{\nu} |\bar{B}^*(s)|^{\nu}$ and $B_\Omega(s) = (\int_0^1 |\bar{B}^*(r)|^{2\nu} dr)^{-1/2} \int_0^s |\bar{B}^*(r)|^{\nu} dB(r)$. In this case, the time-change Brownian motion representation holds with $\eta(s) = \left(\int_0^1 |\bar{B}^*(r)|^{2\nu} dr\right)^{-1/2} \int_0^s |\bar{B}^*(r)|^{2\nu} dr$.

Example 6

For the case of the smooth transition variance break model, the limiting functional B_Ω is the Gaussian process

$$B_\Omega(s) := c(1)^{-1/2} \left(\int_0^s \left(1 + (\kappa^2 - 1)\mathbb{S}(r)\right)^{1/2} dB(r) \right), \tag{19}$$

where $c(s) := \gamma^{-1} \ln\left(\left(1 + e^{-\gamma(s-\tau)}\right) / \left(1 + e^{\gamma\tau}\right)\right)\left(\kappa^2 - 1\right) + \kappa^2 s$ and $\kappa := \sigma_1/\sigma_0$. Notice that in this case B_Ω has quadratic variation, $c(s)/c(1)$. Moreover, as γ diverges to positive infinity (single break case), the process in (19) approaches

$$B_\Omega(s) := \frac{B(s) + (\kappa - 1)(B(s) - B(\tau))\mathbb{I}(s \geq \tau)}{\left(\tau + \kappa^2(1 - \tau)\right)^{1/2}}, \tag{20}$$

a Brownian motion with an instantaneous variance shift at time $s = \tau$.

Figures 1–3 graph the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of the heteroskedastic unit root distribution of $T(\hat{\alpha}_T - 1)$ in (14) (corresponding graphs for t_{DF} are qualitatively similar and are therefore omitted but may be obtained on request) for tests based on OLS demeaning and OLS detrending. Figure 1 relates to the smooth transition variance break model of Example 6 with both positive ($\kappa = 4$) and negative ($\kappa = \frac{1}{4}$) transitions for $\tau = \frac{3}{4}$ and $\gamma \in \{2, 10, \infty\}$. Figure 2 relates to the near-integrated SV model of Example 1 with $h(x) := \exp(x/2)$, $c \in \{0, 10, 20\}$, and $\omega_1 \in \{5, 10\}$. Finally, Figure 3 relates to the near-integrated GARCH model of Example 4 under (11), with $\phi \in \{5, 20\}$ and $\theta \in \{0, 20, 40\}$ (recall that for $\theta = 0$ the unconditional second moments of $u_{T,t}$ are infinite). In each case the dotted line is the standard (homoskedastic) unit root distribution. The reported results were obtained by direct simulation (i.e., approximating the limit functionals using partial sums of normal random variables) of the limiting functional (14). All simulations were based on 50,000 Monte Carlo replications and a discretized grid size of $T = 5{,}000$ segments and used the `rndKMn` and `rndKMgam` functions of Gauss 5.0. The p.d.f.s were estimated using a standard kernel density estimator with Epanechnikov weights, and the c.d.f.s were estimated through the corresponding empirical distribution functions.

In the smooth transition case, it is seen from Figure 1 that negative variance changes shift the distribution slightly to the left, whereas positive changes have a massive impact on the shape of the distribution, in each case relative to the standard unit root distribution. The impact of positive changes is extremely severe in the right tail of the distribution, but for large γ slightly fatter lefttails are also

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

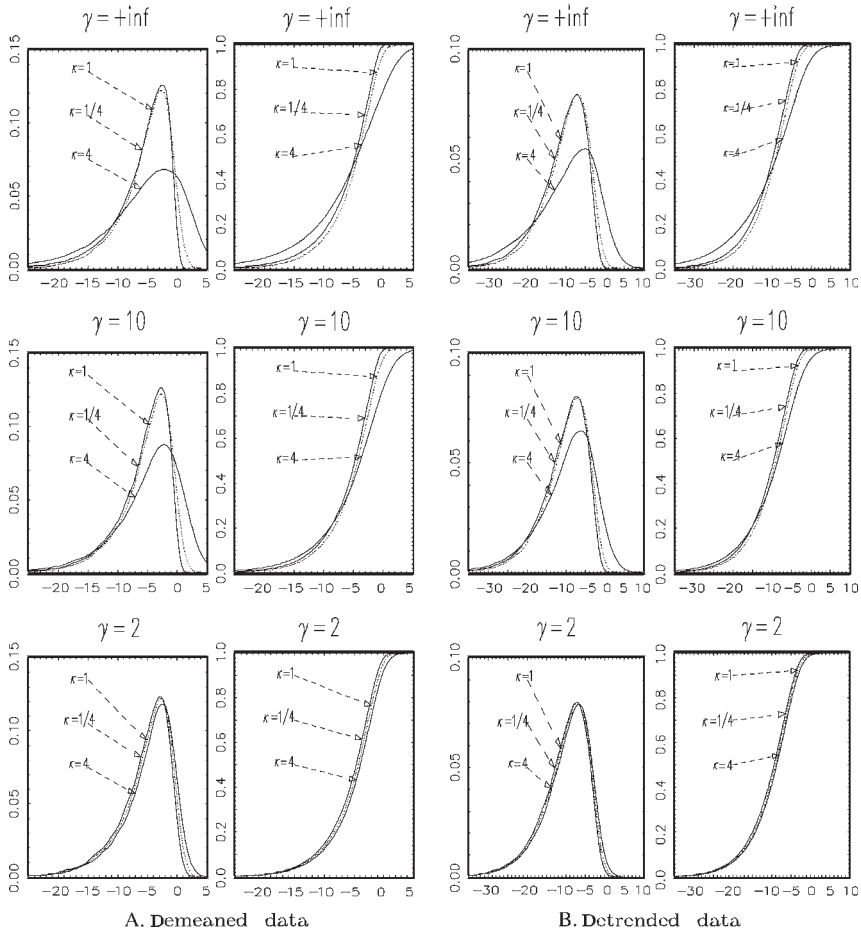


FIGURE 1. The heteroskedastic unit root distribution under smooth transition volatility ($\tau = 3/4$).

seen. In all cases the effects are greatest for the abrupt change model, $\gamma = \infty$, and are moderated as γ approaches zero, as expected. Although those results are not reported here we also considered early ($\tau = 0.25$) and midsample ($\tau = 0.50$) breaks. In the former case negative breaks shift the whole distribution strongly to the left, whereas positive breaks have rather little effect. For the latter case moderated elements of the effects from both early and late breaks are seen. In the near-integrated SV case, it is seen from Figure 2 that the distribution is flattened relative to the homoskedastic case, so that the distribution in this case is characterized by much fatter tails than under homoskedasticity. These effects are most pronounced when c equals zero, diminish as c increases, and are greater the higher is ω_1, other things being equal. Much smaller effects are seen in Figure 3 for the

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

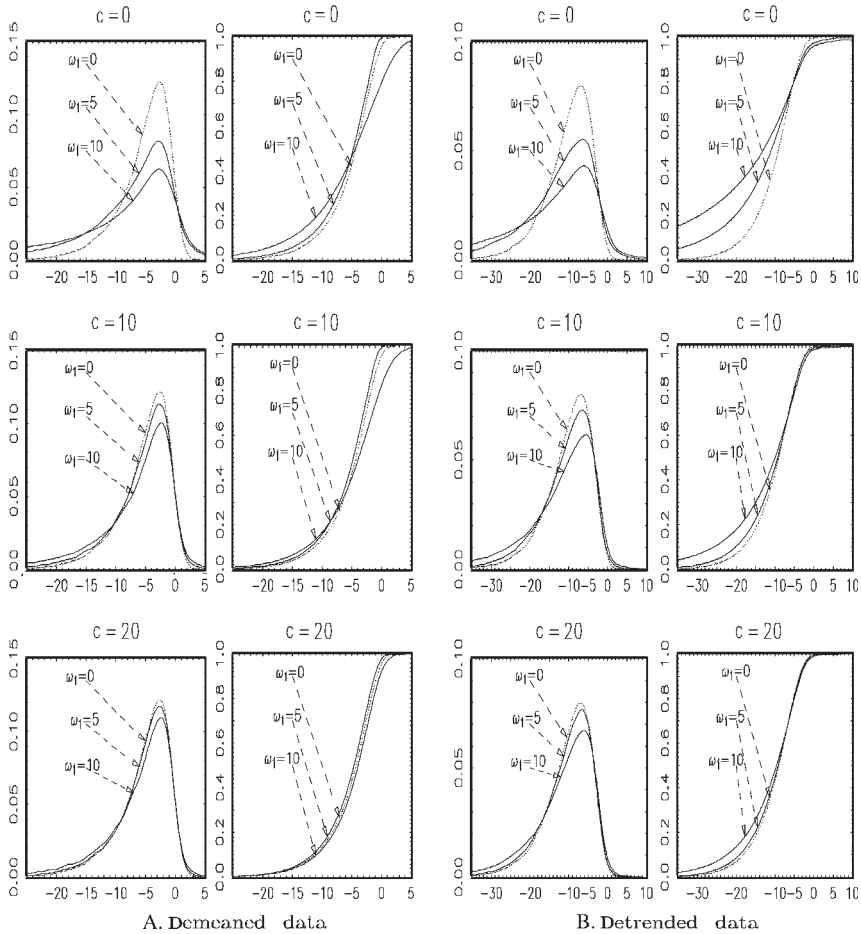


FIGURE 2. The heteroskedastic unit root distribution under (near-) integrated stochastic volatility.

near-integrated GARCH case. Here departures from the standard unit root distribution are most pronounced when ϕ is high and, as expected, are mitigated, other things being equal, as θ increases. The left tail of the distribution is, however, somewhat fatter than in the standard case.

In the foregoing analysis we have worked in the context of a fixed unit root model; i.e., $\alpha = 1$ in (2). Our results generalize straightforwardly to the near–unit root setting of, among others, Phillips (1987a) where $\alpha = 1 - c^*/T$, with c^* a finite constant. Here the large-sample distributions of $T(\hat{\alpha}_T - 1)$ and t_{DF}, which we denote by $Z_{\Omega,z,c,j}$, $j = 1, 2$, respectively, are given by the right members of (14) and (15) on replacing the process $\hat{B}_{\Omega,z}$ by $\hat{J}^{c^*}_{\Omega,z}$, the Hilbert projection in

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

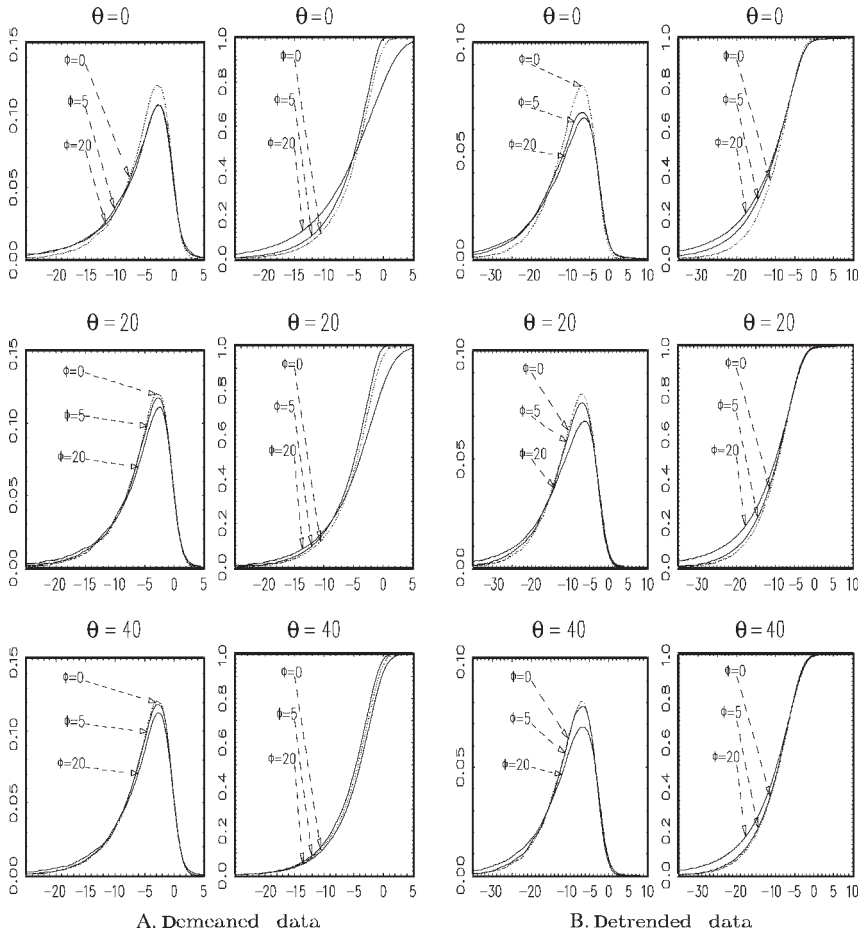


FIGURE 3. The heteroskedastic unit root distribution under (near-) integrated GARCH volatility.

$L_2[0,1]$ of $J_\Omega^{c^*}$ onto the space orthogonal to Z, where

$$J_\Omega^{c^*}(s) := \int_0^s e^{-c^*(s-r)} dB_\Omega(r) = \frac{\int_0^s e^{-c^*(s-r)} \Omega(r)\, dB(r)}{\left(\int_0^1 \Omega(r)^2\, dr\right)^{1/2}}.$$

Notice that $J_\Omega^{c^*}$ reduces to B_Ω of (16) for $c^* = 0$, and to a standard OU process only where the spot volatility, Ω, is constant with probability one. Consequently, the asymptotic local power functions of the $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF} tests at significance level ξ are given by

$$\pi(\Omega, z, c^*, j) := P(Z_{\Omega,z,c^*,j} < b(\Omega, z, j)), \tag{21}$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

$j = 1, 2$, respectively, where $b(\Omega, z, j)$ satisfies $P(Z_{\Omega,z,j} < b(\Omega, z, j)) = \xi$. Notice that the $\pi(\Omega, z, c^*, j)$, $j = 1, 2$, depend on the spot volatility and the integrated (total) variation, even though based on critical values from the correct null distributions, (14) and (15), respectively.

4. BOOTSTRAP UNIT ROOT INFERENCE

Because the heteroskedastic unit root distributions, (14) and (15), depend on both the spot volatility, Ω, and integrated (total) variation, $\int_0^1 \Omega(r)^2 dr$, inference under the unit root null hypothesis based on the $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF} statistics will not be correctly sized (even asymptotically) if p-values are retrieved on the basis of the standard Dickey–Fuller distributions. In this section we show that (asymptotically) valid p-values can be obtained in the presence of global heteroskedasticity using resampling methods,[7] provided there are no leverage effects present.

Most of the existing resampling methods for unit root tests are based on either i.i.d. (see, among others, Inoue and Kilian, 2002; Chang and Park, 2003; Park, 2003) or block (e.g., Paparoditis and Politis, 2003) resampling. Unfortunately, when the global homoskedasticity condition fails, these methods would be expected to replicate the homoskedastic unit root distributions in (5) but not the heteroskedastic unit root distributions in (14) and (15). Consequently, such tests would be expected to behave approximately the same as tests based on standard asymptotic critical values. Unreported Monte Carlo simulations support this view. A further bootstrap that can preserve heteroskadasticity in the resampled data is the pairs bootstrap; see, e.g., Gonçalves and Kilian (2007). However, as discussed in MacKinnon (2006, p. S7), this method is not amenable to hypothesis testing (where the null needs to be imposed on the bootstrap sample data), and so it cannot be adapted to the unit root framework considered in the paper.

Our proposed bootstrap algorithm draws on the "wild bootstrap" literature (see, among others, Wu, 1986; Liu, 1988; Mammen, 1993) and allows us to construct bootstrap unit root tests that are robust to nonstationary SV. In the context of the present problem, the wild bootstrap scheme is required, rather than the standard residual resampling schemes used for other bootstrap unit root tests proposed in the literature, because, unlike these, the wild bootstrap can replicate the pattern of heteroskedasticity present in the shocks; cf. Remark 4.1. The wild bootstrap has also recently been successfully applied by Gonçalves and Kilian (2004, 2007) to the problem of carrying out inference in stationary autoregressive models driven by heterogeneous shocks. Cavaliere and Taylor (2008, 2009) show, respectively, that the wild bootstrap may also be used to construct robust versions of the M unit root tests of, among others, Ng and Perron (2001) in cases where the conditional volatility of the shocks, $\sigma_{T,\lfloor T\cdot \rfloor}$, is a purely deterministic and strictly positive function in $\mathcal{D}_{\mathbb{R}}[0, 1]$, and in the cases where $u_{T,t}$ is a globally *homoskedastic* martingale difference (MD) sequence satisfying (6) (see note 3).

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

4.1. The Bootstrap Algorithm

Implementation of the bootstrap is straightforward and requires only the fitted residuals, $\hat{u}_{T,t}$, from the regression of $\hat{X}_{T,t}$ on $\hat{X}_{T,t-1}$, $t = 1, \ldots, T$. The following steps constitute our bootstrap algorithm.

ALGORITHM 1 (Wild bootstrap unit root test).

Step i. Generate T bootstrap innovations $u^*_{T,t}$, $t = 1, \ldots, T$, according to the device

$$u^*_{T,t} := \hat{u}_{T,t} w_t,$$

where $\{w_t\}_{t=1}^T$ denotes an independent $N(0, 1)$ sequence.

Step ii. Construct the bootstrap sample as the partial sum process

$$X^*_{T,t} := X^*_{T,0} + \sum_{i=1}^{t} u^*_{T,i}, \qquad t = 1, \ldots, T \tag{22}$$

for some initial value[8] $X^*_{T,0}$.

Step iii. Obtain the bootstrap test statistics,

$$T\left(\hat{\alpha}^*_T - 1\right) \quad \text{and} \quad t^*_{DF} := \left(\sum_{t=1}^{T} \hat{X}^{*2}_{T,t-1}\right)^{1/2} \left(\hat{\alpha}^*_T - 1\right) / \hat{\sigma}^*_u,$$

where

$$\hat{\alpha}^*_T := \left(\sum_{t=1}^{T} \hat{X}^*_{T,t} \hat{X}^*_{T,t-1}\right) \left(\sum_{t=1}^{T} \hat{X}^{*2}_{T,t-1}\right)^{-1}$$

and

$$\hat{\sigma}^{*2}_u := T^{-1} \sum_{t=1}^{T} \left(\hat{X}^*_{T,t} - \hat{\alpha}^*_T \hat{X}^*_{T,t-1}\right)^2$$

with $\{\hat{X}^*_{T,t}\}$ the OLS residuals from the regression of $X^*_{T,t}$ on $z_{T,t}$. If no deterministic component is included, replace $\hat{X}^*_{T,t}$ by $X^*_{T,t}$.

Step iv. Bootstrap p-values are computed as

$$p^{*,\hat{\alpha}}_T := G^{*,\hat{\alpha}}_T\left(T\left(\hat{\alpha}_T - 1\right)\right) \quad \text{and} \quad p^{*,t}_T := G^{*,t}_T\left(t_{DF}\right),$$

where $G^{*,\hat{\alpha}}_T$ and $G^{*,t}_T$ denote the conditional (on the original sample data) c.d.f.s of $T(\hat{\alpha}^*_T - 1)$ and t^*_{DF}, respectively. Notice, therefore, that bootstrap tests, run at the ξ, $0 < \xi < 1$, significance level, based on $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF}, are then defined such that they reject the unit root null hypothesis, $H_0 : \alpha = 1$ in (2) if $p^{*,\hat{\alpha}}_T < \xi$ and $p^{*,t}_T < \xi$, respectively.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Remark 4.1. Notice that the bootstrap innovations, $u^*_{T,t}$, replicate the pattern of heteroskedasticity present in the original shocks because, conditionally on $\hat{u}_{T,t}$, $u^*_{T,t}$ is independent over time with zero mean and variance $\hat{u}^2_{T,t}$.

Remark 4.2. The bootstrap samples are generated by imposing the unit root null on the resampling scheme in step ii of Algorithm 1, thereby avoiding the difficulties with the use of estimated unit roots discussed by Basawa, Mallik, McCormick, Reeves, and Taylor (1991). It is also worth noting that the arguments made in Paparoditis and Politis (2003) as to why it is preferable on power grounds not to impose the null hypothesis when constructing the residuals to be used for resampling do not apply here. This is because, as is well known, the wild bootstrap resampling scheme has the important property that, conditionally on the original data, the bootstrap innovations, $u^*_{T,t}$ from step i are serially uncorrelated. This property holds regardless of any serial correlation present in the fitted residuals $\hat{u}_{T,t}$, thereby obviating the issue raised in Paparoditis and Politis (2003).

Remark 4.3. As is well known in the wild bootstrap literature (for a review, see Davidson and Flachaire, 2008) in certain cases improved accuracy can be obtained by replacing the Gaussian distribution used for generating the pseudo-data by an asymmetric distribution with $\mathrm{E}(w_t) = 0$, $\mathrm{E}\left(w_t^2\right) = 1$, and $\mathrm{E}\left(w_t^3\right) = 1$ (Liu, 1988). A well-known example is the Mammen (1993) two-point distribution: $P(w_t = -0.5(\sqrt{5}-1)) = 0.5(\sqrt{5}+1)/\sqrt{5} =: p$, $P(w_t = 0.5(\sqrt{5}+1)) = 1-p$. Davidson and Flachaire (2008) also consider the Rademacher distribution: $P(w_t = 1) = 1/2 = P(w_t = -1)$. We found no discernible differences between the finite-sample properties of the bootstrap unit root tests based on the Gaussian, the Mammen, and the Rademacher distributions. This finding is consistent with evidence reported in Table 5 of Gonçalves and Kilian (2004, p. 105) for stationary autoregressive models.

Remark 4.4. In practice $G^{*,\hat{\alpha}}_T$ and $G^{*,t}_T$ are not known but can be approximated in the usual way through numerical simulation; cf. Hansen (1996) and Andrews and Buchinsky (2001). Using the normalized-bias statistic to illustrate, this is done by generating M (conditionally) independent bootstrap statistics, $T\left(\hat{\alpha}^*_{m:T} - 1\right)$, $m = 1, \ldots, M$, computed as before but from $X^*_{m:T,t} := X^*_{T,0} + \sum_{i=1}^{t} u^*_{m:T,i}$, $u^*_{m:T,t} := \hat{u}_{T,t} w_{m:t}$, $t = 1, \ldots, T$, for some initial value $X^*_{T,0}$ and with $\{\{w_{m:t}\}_{t=1}^{T}\}_{m=1}^{M}$ a doubly independent $N(0,1)$ sequence. The simulated bootstrap p-value is then computed as

$$\tilde{p}^{*,\hat{\alpha}}_T := M^{-1} \sum_{m=1}^{M} \mathbb{I}\left(T(\hat{\alpha}^*_{m:T} - 1) \leq T(\hat{\alpha}_T - 1)\right)$$

and is such that $\tilde{p}^{*,\hat{\alpha}}_T \overset{a.s.}{\to} p^{*,\hat{\alpha}}_T$ as $M \to \infty$. Note that an asymptotic standard error for $\tilde{p}^{*,\hat{\alpha}}_T$ is given by $(\tilde{p}^{*,\hat{\alpha}}_T(1 - \tilde{p}^{*,\hat{\alpha}}_T)/M)^{1/2}$; cf. Hansen (1996, p. 419).

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

4.2. Large-Sample Theory

The asymptotic validity of the wild bootstrap method outlined in Algorithm 1 under nonstationary SV is now established in Theorem 2. In addition to Assumptions 1–3, we also need the following condition to hold.

Assumption 4. The processes $\Omega(\cdot)$ and $B(\cdot)$ are stochastically independent.

Remark 4.5. Assumption 3 excludes the empirically relevant possibility that future volatilities are affected by the sign of the current shock, a phenomenon referred to as leverage in the GARCH and SV literatures. We have to make this assumption because the wild bootstrap is, by design, incapable of replicating in the bootstrap samples any leverage effects that may be present. This because, as noted in Remark 4.1, the bootstrap innovations have the property that, conditional on the data, they are independent over time. Examples 2–6 in Sections 2 and 3 satisfy Assumption 4. In the case of the Hansen (1995) SV model of Example 1 it constitutes an additional restriction, satisfied when $\overline{\omega} = 0$; cf. Section 3.

THEOREM 2. *Under the conditions of Theorem 1 and if Assumption 4 holds, the bootstrap statistics* $T\left(\hat{\alpha}_T^* - 1\right)$ *and* t_{DF}^*, *constructed according to Algorithm 1, are such that* $T\left(\hat{\alpha}_T^* - 1\right) \xrightarrow{w}_p Z_{\Omega,z,1}$ *and* $t_{DF}^* \xrightarrow{w}_p Z_{\Omega,z,2}$. *Moreover,* $p_T^{*,\hat{\alpha}} \xrightarrow{w} U[0,1]$ *and* $p_T^{*,t} \xrightarrow{w} U[0,1]$.

Theorem 2 demonstrates the usefulness of the wild bootstrap under the unit root null ($\alpha = 1$): as the number of observations diverges, the bootstrapped statistics have the same first-order null distribution as the original test statistics. Consequently, the bootstrap p-values are (asymptotically) uniformly distributed under the null hypothesis, leading to tests with (asymptotically) correct size.

Under the (trend-) stationary alternative, $H_1 : |\alpha| < 1$, it is easily shown that bootstrap test statistics $T\left(\hat{\alpha}_T^* - 1\right)$ and t_{DF}^* continue to admit the limiting distributions $Z_{\Omega,z,1}$ and $Z_{\Omega,z,2}$, respectively. Consequently, both of the bootstrap tests deliver consistent inference here because, as is well known, both $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF} diverge to negative infinity as the sample size diverges when $|\alpha| < 1$, so that both $p_T^{*,\hat{\alpha}}$ and $p_T^{*,t}$ converge to zero in probability.

It is quite straightforward to demonstrate that the results of Theorem 2 also hold under near unit roots where $\alpha = 1 - c^*/T$. This immediately implies that the $T\left(\hat{\alpha}_T^* - 1\right)$ and t_{DF}^* tests have the same asymptotic local power functions as the $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF} tests, respectively, provided the original tests are run using critical values from the appropriate limiting null distribution; i.e., the asymptotic local power functions in (21). The asymptotic theory therefore predicts that the bootstrap $T\left(\hat{\alpha}_T^* - 1\right)$ and t_{DF}^* tests should have power approximately equal to the size-adjusted power of the standard $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF} tests, respectively.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

4.3. Numerical Results

Tables 1 and 2 present numerical simulation evidence of the finite-sample performance of our bootstrap tests relative to the standard tests, for both the constant and linear trend cases. We generated data according to (2) with $X_{T,0} = 0$, and $u_{T,t}/\sigma_{T,t-1} =: \varepsilon_{T,t} \sim IIDN(0,1)$, for samples of size $T = 100$ and 250. Table 1 reports level results for $\alpha = 1$, and the corresponding local power results for $\alpha = 1 - c^*/T$, with $c^* = 7.0$ in the constant case and $c^* = 13.5$ in the linear trend case, are reported in Table 2. The standard tests were run using the nominal asymptotic 0.05 level critical values appropriate for the homoskedastic case for the level results in Table 1 and were size-adjusted (such that the exact level of the tests was 0.05) for the corresponding local power results in Table 2. The bootstrap tests were constructed as outlined in Algorithm 1 using $M = 499$ bootstrap replications. All simulations were run using 10,000 Monte Carlo replications and the `rndKMn` function of Gauss 5.0.

Results are reported for the cases where $\sigma_{T,t-1}$ (a) is constant, $\sigma_{T,t-1} = \sigma_0$, for all t; (b) displays a single abrupt volatility shift from σ_0 to σ_1 at time $\lfloor \tau T \rfloor$, as in Example 6; and (c) follows an exponential (near-) integrated autoregressive SV model as in Example 1, with $h(x) = \exp(x/2)$. In all cases $\sigma_0 = 1$ without loss of generality. For (b) we vary σ_1 among $\{4.00, 0.25\}$ and the break date among $\tau \in \{0.25, 0.50, 0.75\}$, so that early, midsample, and late breaks, which may be positive ($\sigma_1 > 1$) or negative ($\sigma_1 < 1$), are allowed. For (c) we vary ω_1 among $\{5, 10\}$ and c among $\{0, 10, 20\}$. In the context of (c) we assume that $\{\eta_{T,t}, \varepsilon_{T,t}\}$ form a zero mean, unit variance independent Gaussian sequence with $\mathrm{E}\left(\eta_{T,t}\varepsilon_{T,t}\right) = \overline{\omega}$. We report results for both (i) the no leverage case, $\overline{\omega} = 0$, and (ii) the case where strong leverage is present with $\overline{\omega} = -0.5$.[9]

Consider first the results for $\alpha = 1$ in Table 1. For the single break model the standard unit root tests are badly oversized, with no amelioration as T increases, in the presence of early and midsample negative and late positive breaks. In the case of the t_{DF} test, e.g., empirical size can be as high as 37% under early negative breaks. The t_{DF} test is also oversized under late negative breaks. For the (near-) integrated SV model, regardless of the leverage parameter $\overline{\omega}$, when ω_1 is low ($\omega_1 = 5$) and volatility is strongly mean reverting ($c = 20$), the standard tests perform reasonably well in terms of size, which is rarely in excess of 8%. However, when ω_1 grows and/or c decreases, the impact of volatility becomes more pronounced, leading to strongly oversized tests. When the volatility is nonstationary ($c = 0$) and ω_1 is low, rejection frequencies for the standard tests range between about 14% and 20%. When ω_1 is high, these frequencies are considerably increased. Significant oversize is also seen when ω_1 is high but the process is mean reverting ($c = 10, 20$). In contrast, the size properties of the corresponding bootstrap tests are largely satisfactory throughout with empirical sizes converging well to the nominal level as T increases. Although not theoretically valid in the presence of leverage, it can be seen from a comparison of the final two panels of Table 1 that the leverage coefficient, $\overline{\omega}$, appears to have very little impact on the

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

TABLE 1. Level of bootstrap and standard unit root tests: (a) constant volatility; (b) single volatility shift model; (c) near-integrated autoregressive stochastic volatility (AR-SV) model

			Demeaned case				Detrended case			
		T	$T(\hat{\alpha}_T^* - 1)$	t_{DF}^*	$T(\hat{\alpha}_T - 1)$	t_{DF}	$T(\hat{\alpha}_T^* - 1)$	t_{DF}^*	$T(\hat{\alpha}_T - 1)$	t_{DF}
Case (a) Constant volatility model										
		100	5.0	5.3	4.4	5.7	5.1	5.4	3.8	6.1
		250	5.0	5.1	4.6	5.1	5.2	5.1	4.8	5.6
σ_1	τ		Case (b) Single volatility shift model							
4.00	0.25	100	5.3	5.1	4.2	2.9	5.2	5.2	4.0	4.4
		250	5.1	5.1	4.2	2.5	5.1	5.0	4.5	4.0
4.00	0.50	100	5.6	5.3	6.7	5.0	5.9	5.9	6.2	6.5
		250	5.3	5.2	7.4	5.0	5.6	5.4	6.7	5.7
4.00	0.75	100	6.1	6.1	11.0	8.6	6.9	6.9	10.9	10.6
		250	5.5	5.5	11.5	8.4	5.8	5.7	11.3	9.7
0.25	0.25	100	6.4	7.4	18.8	37.1	6.8	8.4	15.8	29.2
		250	6.0	6.4	21.2	37.3	6.2	6.5	18.5	29.0
0.25	0.50	100	5.5	6.4	10.1	18.4	5.6	6.0	8.1	13.4
		250	5.2	5.4	11.6	17.9	5.3	5.5	9.5	13.6
0.25	0.75	100	5.5	5.5	6.2	9.3	5.4	5.8	5.3	9.0
		250	5.0	5.0	6.5	8.7	5.5	5.4	6.6	8.6
ω_1	c		Case (c)(i) Near-integrated AR-SV model with no leverage ($\varpi = 0$)							
5	0	100	6.7	6.9	16.1	19.6	7.1	7.2	14.1	18.5
		250	5.6	5.8	17.6	20.1	5.8	6.1	15.9	18.6
5	10	100	5.5	5.5	6.8	8.3	5.8	6.0	6.1	9.1
		250	5.2	5.3	7.4	7.6	5.5	5.7	7.4	8.3
5	20	100	5.5	5.5	5.6	7.1	5.5	5.8	5.1	7.8
		250	5.2	5.3	6.1	6.4	5.3	5.4	6.1	7.2
10	0	100	8.1	8.0	26.8	29.0	8.9	9.0	24.2	28.1
		250	6.7	6.6	28.7	30.4	6.9	7.0	27.4	29.5
10	10	100	6.5	5.9	11.6	13.2	6.9	6.9	11.8	14.7
		250	5.9	5.6	13.2	13.4	6.3	6.2	14.6	14.9
10	20	100	6.2	5.7	8.6	10.0	6.3	6.5	8.5	11.2
		250	5.7	5.5	9.5	9.6	5.8	5.9	10.5	11.2
ω_1	c		Case (c)(ii) Near-integrated AR-SV model with leverage ($\varpi = -0.5$)							
5	0	100	6.4	7.4	14.6	18.8	6.7	7.5	13.0	17.7
		250	5.5	6.4	15.4	18.8	6.4	6.3	15.1	18.6
5	10	100	5.6	5.8	6.7	8.2	6.6	6.9	7.6	10.4
		250	5.2	5.1	6.9	6.7	5.6	5.0	7.7	8.4
5	20	100	6.0	5.6	5.8	6.8	6.4	6.8	6.3	9.3
		250	4.9	4.9	5.7	5.8	5.8	5.5	6.7	7.3
10	0	100	7.1	7.8	23.6	27.2	8.3	9.1	22.7	27.0
		250	5.6	6.5	26.8	29.4	7.5	7.7	28.7	30.3
10	10	100	6.0	5.5	11.2	12.3	7.9	7.9	14.3	17.0
		250	6.2	6.2	13.0	12.6	7.8	7.7	16.2	16.2
10	20	100	5.9	6.1	8.4	9.7	7.6	7.9	10.7	14.0
		250	6.2	5.6	9.5	9.3	7.4	7.0	12.5	13.2

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

TABLE 2. Power of bootstrap and standard unit root tests: (a) constant volatility; (b) single volatility shift model; (c) near-integrated autoregressive stochastic volatility (AR-SV) model

			Demeaned case: Power, $c^* = 7$				Detrended case: Power, $c^* = 13.5$			
		T	$T(\hat{\alpha}_T^* - 1)$	t_{DF}^*	$T(\hat{\alpha}_T - 1)$	t_{DF}	$T(\hat{\alpha}_T^* - 1)$	t_{DF}^*	$T(\hat{\alpha}_T - 1)$	t_{DF}
Case (a) Constant volatility model										
		100	29.1	18.4	28.3	16.8	40.5	33.0	37.2	29.9
		250	28.7	18.1	28.9	17.1	40.7	31.4	41.3	31.6
σ_1	τ		Case (b) Single volatility shift model							
4.00	0.25	100	33.1	31.7	31.4	30.2	39.6	37.6	36.7	34.9
		250	32.7	31.6	32.0	30.9	38.5	36.3	40.0	37.7
4.00	0.50	100	26.3	25.1	24.7	23.2	39.5	37.8	36.7	35.6
		250	25.3	23.8	25.0	23.6	37.5	35.8	37.3	35.4
4.00	0.75	100	25.1	23.0	21.5	19.4	36.5	34.6	30.1	29.1
		250	23.8	21.6	22.4	20.6	33.8	31.9	32.3	29.9
0.25	0.25	100	14.9	4.3	12.5	2.1	26.1	16.3	22.4	10.6
		250	13.9	3.1	11.9	2.1	23.9	12.6	21.6	10.4
0.25	0.50	100	18.1	8.0	15.3	6.1	29.8	22.0	27.7	19.1
		250	16.9	7.4	16.3	5.9	27.8	19.5	26.6	19.0
0.25	0.75	100	25.7	14.8	23.7	13.2	34.1	26.2	30.5	23.9
		250	24.8	13.7	26.0	13.6	32.1	24.2	32.0	24.1
ω_1	c		Case (c)(i) Near-integrated AR-SV model with no leverage ($\varpi = 0$)							
5	0	100	21.9	17.2	10.6	3.8	30.8	26.5	18.4	10.4
		250	20.5	16.1	11.1	4.7	28.7	24.1	18.1	9.5
5	10	100	27.1	19.5	24.2	16.1	36.8	30.9	31.4	26.8
		250	27.4	20.3	24.9	17.1	35.7	30.8	34.5	28.3
5	20	100	27.6	18.9	26.6	16.7	37.9	31.4	34.0	27.3
		250	28.3	19.7	26.9	17.4	37.4	31.3	36.4	30.1
10	0	100	19.1	16.8	5.7	2.8	25.2	23.1	9.3	5.0
		250	17.2	14.8	5.7	2.1	22.5	20.1	8.6	4.1
10	10	100	25.0	20.6	18.0	13.2	31.8	28.8	22.1	19.1
		250	24.7	20.7	18.4	14.8	29.1	26.8	22.4	19.7
10	20	100	27.0	20.7	21.3	14.6	34.5	30.5	26.5	21.0
		250	26.9	21.3	23.4	16.8	33.4	29.6	30.2	24.5
ω_1	c		Case (c)(ii) Near-integrated AR-SV model with leverage ($\varpi = -0.5$)							
5	0	100	23.3	18.5	14.9	5.0	32.2	27.8	21.2	12.3
		250	21.0	17.0	12.4	2.6	28.4	24.4	17.4	7.9
5	10	100	31.1	22.8	29.1	19.9	39.8	34.2	31.8	27.0
		250	29.2	21.7	26.7	17.6	37.3	31.8	31.9	25.5
5	20	100	30.8	21.8	30.6	19.8	41.5	34.5	34.8	30.0
		250	29.8	20.0	29.2	19.2	38.9	32.3	34.0	26.3
10	0	100	19.7	17.6	6.5	2.7	26.9	25.1	10.5	5.4
		250	17.5	15.2	5.8	1.8	23.0	21.1	8.2	2.9
10	10	100	30.9	25.4	23.1	17.5	37.6	33.8	22.9	20.6
		250	29.1	24.2	21.5	18.2	33.6	30.4	19.9	17.7
10	20	100	32.0	26.0	27.2	18.8	40.5	36.4	27.4	21.8
		250	31.5	24.5	27.5	19.8	38.4	33.6	25.1	21.1

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

behavior of the bootstrap tests. The same appears to also be true with respect to the standard tests.

Turning to the results for $\alpha = 1 - c^*/T$ in Table 2, we see that the prediction from the local asymptotic theory holds true in almost all cases; i.e., each of the bootstrap tests has local power that is approximately equal to the size-adjusted power of the corresponding standard test. The only significant exceptions occur for the SV model when $c = 0$, particularly so when ω_1 is high. Here the bootstrap tests have much higher local power than the size-adjusted power of the corresponding standard tests. Overall, the local power functions of the $T(\hat{\alpha}_T - 1)$ and $T(\hat{\alpha}_T^* - 1)$ tests appear to be much less dependent on the form of the volatility process, $\sigma_{T,t-1}$, than the t_{DF} and t_{DF}^* tests, which in some cases display rejection frequencies below the nominal level. Finally, in the case where $\sigma_{T,t-1}$ is constant the bootstrap test has almost identical power to the size-adjusted power of the standard tests, indicating that there is no loss in finite-sample power attached to using the bootstrap versions of the tests when volatility is constant. For both the standard and bootstrap unit root tests leverage again appears to have rather little impact on the power properties of the tests.

5. WEAK DEPENDENCE

Thus far we have assumed that $\{u_{T,t}\}$ is a MDA. In this section we weaken this to allow for weak dependence in $\{u_{T,t}\}$. As we show subsequently, this can introduce additional nuisance parameters into the asymptotic distributions of the unit root statistics. To introduce dependence we assume that the shocks in (2) are a linear process in a MDA; specifically, we require $u_{T,t} = c(L)e_{T,t}$ with $c(z) := \sum_{k=0}^{\infty} c_k z^k$ and $e_{T,t} = \sigma_{T,t-1}\varepsilon_{T,t}$, where $\mathrm{E}\left(e_{T,t}|\mathcal{I}_{T,t-1}\right) = 0$, $\mathrm{E}(e_{T,t}^2|\mathcal{I}_{T,t-1}) = \sigma_{T,t-1}^2$ (a.s.), and we replace Assumption 1 with the following assumption.

Assumption 1′. (a) $\{\varepsilon_{T,t}\}$ satisfies Assumption 1; (b) $\sum_{k=0}^{\infty} k|c_k| < \infty$; (c) $\bar{c}_1 := c(1) \neq 0$; (d) for all $T > 0$, $\sup_{t<0} a_T^{-1}\sigma_{T,t} \leq K < \infty$, a.s.

We now generalize the results in Lemma 1 and Theorem 1 to allow for weak dependence of the form given in Assumption 1′.

THEOREM 3. *Suppose $\{X_{T,t}\}$ is generated according to (2) with $\alpha = 1$. Then, under Assumptions 1′, 2, and 3,*

$$\frac{1}{a_T T^{1/2}} \sum_{t=1}^{\lfloor T\cdot \rfloor} u_{T,t} \xrightarrow{w} \bar{c}_1 m_\Omega(\cdot) \tag{23}$$

and $\hat{\alpha}_T$ of (3) and the associated studentized statistic, t_{DF}, satisfy

$$T\left(\hat{\alpha}_T - 1\right) \xrightarrow{w} \frac{\hat{B}_{\Omega,z}(1)^2 - \hat{B}_{\Omega,z}(0)^2 - \varrho}{2\int_0^1 \hat{B}_{\Omega,z}(s)^2\, ds} =: Z_{\Omega,\varrho,z,1}, \tag{24}$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

$$t_{DF} \xrightarrow{w} \frac{\hat{B}_{\Omega,z}(1)^2 - \hat{B}_{\Omega,z}(0)^2 - \varrho}{2\left(\varrho \int_0^1 \hat{B}_{\Omega,z}(s)^2 ds\right)^{1/2}} =: Z_{\Omega,\varrho,z,2}, \tag{25}$$

where $\varrho := \bar{c}_2/\bar{c}_1^2$, *with* $\bar{c}_2 := \sum_{k=0}^{\infty} c_k^2$, *and* $\hat{B}_{\Omega,z}$ *is as defined in Theorem 1.*

Remark 5.1. The results in (24) and (25) extend the well-known unit root distributions for globally homoskedastic, weakly dependent errors to the case of globally heteroskedastic, weakly dependent errors. If the error process, $\{u_{T,t}\}$, satisfies Assumptions 1′ with $\{\sigma_{T,t}\}$ constant, the results in (24) and (25) reduce to the expressions given in parts (c) and (e), respectively, of Theorem 3.1 of Phillips (1987b, p. 282). Notice that $\hat{\alpha}_T$ is T-consistent under the unit root null, even though the error process can display both serial correlation and nonstationary, possibly explosive volatility.

Provided Assumption 4 holds, (asymptotically) correctly sized inference can again be obtained by using the wild bootstrap algorithm of Section 4. The main difference is that, instead of considering the simple unit root statistics of Section 3, autocorrelation robust test statistics need to be employed. To that end, consider the well-known statistic of Phillips (1987b) and Phillips and Perron (1988),

$$\hat{Z}_{\hat{\alpha}_T} := T\left(\hat{\alpha}_T - 1\right) - 0.5(\hat{\lambda}^2 - \hat{\sigma}^2)(T^{-2}\textstyle\sum_{t=1}^{T} \hat{X}_{T,t-1}^2)^{-1}$$

and the associated studentized statistic $\hat{Z}_t$ (see, e.g., Phillips, 1987b, eqn. (22)), where $\hat{\lambda}^2$ is a sum-of-covariances (SC) estimator of the form

$$\hat{\lambda}^2 := \sum_{h=-T+1}^{T-1} \omega(hq_T^{-1})\hat{\gamma}_h, \qquad \hat{\gamma}_h := T^{-1}\sum_{t=1}^{T-|h|} \hat{u}_{T,t}\hat{u}_{T,t+|h|}$$

with kernel function $\omega(\cdot)$ satisfying, e.g., the general conditions reported in Jansson (2002, Assump. A3) and the bandwidth parameter $q_T \in (0, \infty)$ satisfying $q_T^{-1} + T^{-1/2}q_T \to 0$ (which corresponds to Assumption A4 of Jansson, 2002). If the bootstrap samples and the bootstrap statistics are constructed according to Algorithm 1, then bootstrap p-values computed as $p_T^{**,\hat{\alpha}} := G_T^{*,\hat{\alpha}}(\hat{Z}_{\hat{\alpha}_T})$ and $p_T^{**,t} := G_T^{*,t}(\hat{Z}_t)$ are asymptotically valid, as is shown in the next theorem. To establish this result, we require a consistent estimator of the serial correlation nuisance parameter ϱ. To do so, as is standard, we need to impose a supplementary moment restriction (which implies the finiteness of the unconditional second moment of $u_{T,t}$) to hold on the volatility process.

THEOREM 4. *Suppose* $\{X_{T,t}\}_0^T$ *is generated according to (2) with* $\alpha = 1$. *Then, under Assumptions 1′ and 2–4, and if, for some* $q > 2$, $\sup_{t \le T} \mathrm{E}((a_T^{-1}\sigma_{T,t})^q) \le Q_\sigma < \infty$, *then* $p_T^{**,\hat{\alpha}} \xrightarrow{w} U[0,1]$ *and* $p_T^{**,t} \xrightarrow{w} U[0,1]$.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Notice that because the bootstrap samples have (conditionally) uncorrelated increments (see Remark 4.2), and the limiting null distributions of $\hat{Z}_{\hat{\alpha}_T}$ and $\hat{Z}_t$ do not depend on serial correlation nuisance parameters, the bootstrap statistics can still be computed as in Algorithm 1 and do not require a correction for serial correlation. However, an improved finite-sample approximation in the presence of serially correlated errors might be anticipated from computing the bootstrap analogues of the $\hat{Z}_{\hat{\alpha}_T}$ and $\hat{Z}_t$ statistics, say, $\hat{Z}^*_{\hat{\alpha}_T}$ and $\hat{Z}^*_t$, respectively, in step iii of Algorithm 1 and then computing the p-values in step iv using the c.d.f.s of $\hat{Z}^*_{\hat{\alpha}_T}$ and $\hat{Z}^*_t$. The results in Theorem 4 would be unaltered.

In an attempt to improve finite-sample performance under weak dependence, the basic bootstrap principle outlined here can be further extended to include a recoloring (or sieve) component, without altering the large-sample theory given in Theorem 4.[10] Following the approach of Ferretti and Romo (1996), Park (2002b), and Chang and Park (2003), among others, this involves rebuilding stationary serial correlation into the bootstrap innovations. This can be done by using the estimated stationary lag dynamics obtained from fitting an augmented Dickey–Fuller (ADF) regression of the form[11]

$$\Delta \hat{X}_{T,t} = \pi \hat{X}_{T,t-1} + \sum_{i=1}^{k} \beta_i \Delta \hat{X}_{T,t-i} + u_{T,t;k}. \tag{26}$$

Specifically, and denoting $\hat{\beta}_k(z) := \sum_{i=1}^{k} \hat{\beta}_i z^i$, the resampling scheme (22) in step ii of Algorithm 1 is then replaced with the recoloring recursion

$$\hat{\beta}_k(L)(1-L)X^*_{T,t} = u^*_{T,t}, \qquad t = 1, \ldots, T, \tag{27}$$

again initialized at $\mathbf{0}$. The scheme in (27) differs from that in (22) in that the estimated autoregressive lag polynomial, $\hat{\beta}_k(L)$ from (26), is incorporated into the resampling scheme to recolor the bootstrap innovations.[12] In small samples there is the possibility that the resulting estimated lag polynomial could have one or more unit or explosive roots, and we found that the performance of the bootstrap was improved if any such root was shrunk to have modulus less than unity; in our experiments reported in Section 5.1 we scaled such estimated roots to have modulus equal to 0.99.

In what follows we shall denote the lag truncation used in estimating (26) for the purposes of recoloring as $k = k_{rc}$. Notice that k_{rc} need not satisfy the condition that it diverges to infinity with the sample size because the recoloring device is motivated from purely finite-sample concerns. The large-sample results established in Theorem 4 also apply to the corresponding recolored bootstrap statistics. Additionally one might also consider a wild bootstrap procedure based on the ADF statistic, denoted t_{ADF}, for testing $\pi = 0$ against $\pi < 0$ in (26). In this case we would need to use a lag length $k = k_T$ in (26) satisfying the standard condition (see, e.g., Chang and Park, 2002) that $1/k_T + k_T^3/T \to 0$ as $T \to \infty$. The wild bootstrap analogue of the ADF statistic, denoted t^*_{ADF}, is then obtained

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

as the ADF statistic from the bootstrap ADF regression,

$$\Delta \hat{X}^*_{T,t} = \pi^* \hat{X}^*_{T,t-1} + \sum_{i=1}^{k_*} \beta_i \Delta \hat{X}^*_{T,t-i} + u^*_{T,t;k_*}, \tag{28}$$

where k_* need only satisfy the condition that $k_*/T^{1/3} \to 0$ as $T \to \infty$ and, where recoloring is used, is such that $k_* \geq k_{rc}$ (notice that k_{rc} does not need to be the same as k_T), used in the recoloring scheme (27). We conjecture that the results stated in Theorem 4 relating to the wild bootstrap $\hat{Z}^*_t$ statistic will also hold for t^*_{ADF}, although stronger restrictions on the short-run dynamics of the process would need to be imposed.[13]

5.1. Numerical Results

Tables 3 and 4 report results analogous to those in Tables 1 and 2, respectively, for the $\hat{Z}_t$ and t_{ADF} tests outlined earlier, together with their bootstrap analogues, $\hat{Z}^*_t$ and t^*_{ADF}, respectively, for data generated according to the data generating procedure (DGP)

$$X_{T,t} = \alpha X_{T,t-1} + u_{T,t}, \qquad t = 1, \ldots, T,$$
$$u_{T,t} = \rho u_{T,t-1} + e_{T,t} - \vartheta e_{T,t-1},$$

with $X_{T,0} = 0$ and $e_{T,t} = \sigma_{T,t-1}\varepsilon_{T,t}$, with $\varepsilon_{T,t} \sim NIID(0, 1)$ and $\sigma_{T,t-1}$ generated according to either case (a) (constant volatility); case (b) (one-time change in volatility) with $\sigma_1 = 4$ and $\tau = 0.75$; or case (c) (near-integrated autoregressive SV) with $\omega_1 = 10$ and $c = 10$, with either (i) $\overline{\omega} = 0.0$ (no leverage) or (ii) $\overline{\omega} = -0.5$ (strong leverage), from Tables 1 and 2.[14] Results are reported for $\rho = \vartheta = 0$ and for $\rho = \pm 0.5$, $\vartheta = 0$, and for $\vartheta = \pm 0.5$, $\rho = 0$. In all cases, $\sigma_{T,t-1} = 1$ for $t \leq 0$ and $u_{T,0}$ is drawn from the corresponding stationary distribution assuming a unit conditional variance.

The standard tests were run using the nominal asymptotic 0.05 level critical values appropriate for the homoskedastic case for the level results ($\alpha = 1$) in Table 3 and were again size-adjusted (such that the exact level of the tests was 0.05) for the corresponding local power results in Table 4 ($\alpha = 1 - c^*/T$). For $\hat{Z}_t$ and $\hat{Z}^*_t$, we employed the Andrews (1991) SC long-run variance estimator with quadratic spectral kernel and AR(1)-automatic bandwidth selection. For t_{ADF} and t^*_{ADF}, the lag truncation lengths in (26) and (28) were (individually) determined using the modified Akaike information criterion (MAIC) of Ng and Perron (2001) with in each case a maximum lag of $\lfloor 12(T/100)^{1/4} \rfloor$ used throughout. The bootstrap $\hat{Z}^*_t$ and t^*_{ADF} tests were constructed as outlined previously, using the estimated c.d.f.s of $\hat{Z}^*_t$ and t^*_{ADF}, respectively, for $M = 499$ bootstrap replications. The bootstrap samples were constructed using either a fixed recoloring lag length of 4, $k_{rc} = 4$, or by setting the recoloring lag length equal to the lag length chosen according to the MAIC criterion used in constructing

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

TABLE 3. Level of bootstrap and standard unit root tests, dependent errors: (a) constant volatility model; (b) single volatility shift model; (c) near-integrated autoregressive stochastic volatility (AR-SV) model

			Demeaned case						Detrended case					
					$k_{rc}=4$		$k_{rc}=MAIC$				$k_{rc}=4$		$k_{rc}=MAIC$	
ρ	ϑ	T	t_{ADF}	$\hat{Z}_t$	t^*_{ADF}	$\hat{Z}^*_t$	t^*_{ADF}	$\hat{Z}^*_t$	t_{ADF}	$\hat{Z}_t$	t^*_{ADF}	$\hat{Z}^*_t$	t^*_{ADF}	$\hat{Z}^*_t$
Case (a) Constant volatility model														
0.0	0.0	100	4.1	5.9	6.7	6.7	5.2	4.8	4.3	7.0	11.2	12.0	5.6	4.8
0.0	0.0	250	3.8	5.0	5.4	5.7	5.0	4.4	3.9	5.6	7.2	7.3	5.1	4.4
0.5	0.0	100	3.8	2.5	6.5	6.9	5.3	5.8	4.3	1.2	9.9	9.3	5.8	5.9
0.5	0.0	250	4.1	2.6	5.5	5.5	4.9	4.8	3.8	1.5	6.3	6.4	4.8	4.9
−0.5	0.0	100	3.8	29.8	6.0	10.7	5.0	6.5	3.9	52.5	9.6	21.3	5.1	10.3
−0.5	0.0	250	4.0	24.1	5.4	7.6	4.9	5.2	3.5	43.9	6.7	10.6	4.8	6.2
0.0	0.5	100	3.4	3.3	5.8	6.6	4.8	5.4	3.0	1.9	7.5	9.9	4.4	5.8
0.0	0.5	250	3.5	3.3	4.7	5.0	4.5	4.5	2.8	2.7	4.9	6.2	4.0	5.0
0.0	−0.5	100	5.5	47.9	6.1	15.3	5.7	12.0	7.9	74.5	10.4	30.1	7.7	23.4
0.0	−0.5	250	5.0	40.0	5.9	10.7	5.8	8.8	4.7	67.1	6.1	15.7	5.6	12.5
Case (b) Single volatility shift model														
0.0	0.0	100	1.7	7.7	5.1	8.4	4.3	2.1	2.5	10.2	8.0	13.5	4.1	2.9
0.0	0.0	250	2.8	7.9	5.0	6.1	4.5	3.0	2.3	9.8	5.4	7.9	4.6	2.9
0.5	0.0	100	1.6	1.0	5.2	7.2	4.5	3.2	1.8	0.5	7.5	9.7	4.4	2.9
0.5	0.0	250	2.7	2.2	4.9	6.2	4.5	3.7	2.7	1.6	5.4	7.4	4.8	3.7
−0.5	0.0	100	1.6	37.2	4.7	13.0	3.9	4.8	2.2	50.3	7.5	20.9	3.8	7.3
−0.5	0.0	250	2.5	32.3	4.8	8.2	4.2	4.5	2.3	47.0	5.2	12.1	4.4	5.0
0.0	0.5	100	1.5	2.4	4.9	8.0	4.0	3.0	1.4	1.9	6.2	11.3	3.9	3.1
0.0	0.5	250	2.4	4.2	4.7	5.5	4.2	3.2	2.1	3.7	4.8	6.7	4.4	3.5
0.0	−0.5	100	2.9	53.0	5.6	17.5	4.6	8.3	5.2	67.5	9.5	28.6	6.6	15.1
0.0	−0.5	250	2.9	48.8	4.8	11.5	4.6	6.2	2.8	65.5	5.3	16.5	4.9	8.2

Case (c)(i) Near-integrated AR-SV model with no leverage ($\overline{\omega} = 0$)														
0.0	0.0	100	6.2	12.1	5.6	5.9	4.7	3.0	5.7	14.5	7.7	9.7	5.1	3.1
0.0	0.0	250	6.4	12.6	5.0	5.5	4.7	2.9	4.6	14.6	5.9	7.1	5.0	2.9
0.5	0.0	100	5.9	4.3	5.7	5.1	5.3	3.5	4.0	2.4	6.6	7.4	4.9	3.0
0.5	0.0	250	6.0	5.6	4.9	5.7	4.6	3.9	4.6	3.9	5.9	7.0	5.1	4.0
−0.5	0.0	100	6.3	37.4	5.4	9.5	4.7	4.7	5.4	54.5	7.4	16.4	4.7	7.9
−0.5	0.0	250	6.6	33.2	5.0	6.4	4.8	3.5	4.8	49.6	6.1	9.7	5.5	4.1
0.0	0.5	100	5.8	6.3	5.4	5.3	4.8	3.4	3.5	4.5	5.9	8.1	4.9	3.4
0.0	0.5	250	6.4	8.2	5.1	5.7	4.9	3.7	4.5	6.5	5.3	7.1	4.7	3.7
0.0	−0.5	100	7.9	51.8	5.5	14.3	5.3	8.7	8.4	71.1	9.3	25.3	7.1	16.5
0.0	−0.5	250	7.3	47.0	5.3	9.6	5.1	5.2	5.8	66.7	6.3	14.5	5.6	7.7
Case (c)(ii) Near-integrated AR-SV model with leverage ($\overline{\omega} = -0.5$)														
0.0	0.0	100	6.6	12.9	5.9	6.3	5.3	3.3	6.1	16.7	9.7	11.9	6.2	4.5
0.0	0.0	250	5.9	12.5	5.3	5.6	5.1	3.0	5.4	16.7	7.1	8.4	6.3	3.3
0.5	0.0	100	5.4	4.5	5.7	5.5	5.0	3.4	3.8	2.7	7.5	7.9	5.0	3.4
0.5	0.0	250	5.9	5.0	5.1	5.0	4.7	3.4	4.9	4.0	6.5	7.6	5.9	4.4
−0.5	0.0	100	6.5	38.6	5.9	10.0	4.9	5.1	5.9	60.0	8.6	19.9	5.6	9.9
−0.5	0.0	250	6.1	33.9	5.5	7.1	5.3	4.1	5.3	55.4	6.6	11.9	6.0	4.8
0.0	0.5	100	5.7	6.4	5.4	5.7	4.8	3.4	4.0	5.7	7.0	9.2	5.0	3.6
0.0	0.5	250	6.0	7.7	5.1	5.2	4.9	3.5	5.2	7.8	6.4	7.7	6.0	4.2
0.0	−0.5	100	7.9	54.2	6.5	14.4	6.1	9.5	9.5	76.0	10.7	28.7	8.2	19.9
0.0	−0.5	250	6.5	48.6	5.4	10.6	5.2	5.7	6.4	73.2	6.8	17.7	6.5	8.9

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

TABLE 4. Power of bootstrap and standard unit root tests, dependent errors: (a) constant volatility model; (b) single volatility shift model; (c) near-integrated autoregressive stochastic volatility (AR-SV) model.

			Demeaned case: Power, $c^* = 13.5$						Detrended case: Power, $c^* = 13.5$					
					$k_{rc} = 4$		$k_{rc} = MAIC$				$k_{rc} = 4$		$k_{rc} = MAIC$	
ρ	ϑ	T	t_{ADF}	$\hat{Z}_t$	t^*_{ADF}	$\hat{Z}^*_t$	t^*_{ADF}	$\hat{Z}^*_t$	t_{ADF}	$\hat{Z}_t$	t^*_{ADF}	$\hat{Z}^*_t$	t^*_{ADF}	$\hat{Z}^*_t$
Case (a) Constant volatility model														
0.0	0.0	100	15.3	18.8	18.8	22.5	15.0	15.7	26.5	33.0	41.7	46.8	27.5	27.1
0.0	0.0	250	18.1	20.7	18.0	20.6	16.3	16.7	29.3	33.8	35.0	38.2	27.9	26.9
0.5	0.0	100	13.0	6.0	16.4	10.7	13.7	7.6	13.5	12.5	25.9	21.9	16.9	9.8
0.5	0.0	250	16.7	11.4	16.7	12.0	15.3	10.5	24.4	18.9	29.4	23.3	24.2	17.6
−0.5	0.0	100	14.9	30.1	17.1	47.1	13.6	34.7	25.2	41.5	35.8	72.8	24.6	55.4
−0.5	0.0	250	17.0	32.1	18.0	38.3	15.9	31.3	29.1	41.2	33.1	57.8	27.2	44.7
0.0	0.5	100	12.6	11.2	13.8	15.2	11.1	11.0	15.2	21.0	21.9	30.7	13.6	14.7
0.0	0.5	250	15.1	15.6	13.3	15.6	12.3	12.9	22.2	25.0	21.9	28.0	18.3	21.2
0.0	−0.5	100	16.3	29.2	19.1	58.0	17.5	48.7	21.6	38.9	32.2	84.1	27.3	75.0
0.0	−0.5	250	16.2	33.2	18.2	48.9	17.6	41.5	26.2	41.6	28.4	71.6	26.6	61.4
Case (b) Single volatility shift model														
0.0	0.0	100	16.1	22.1	16.6	30.7	13.5	11.5	26.2	31.3	32.9	51.5	22.8	19.8
0.0	0.0	250	16.4	22.2	16.0	25.1	14.6	13.2	23.2	32.3	23.9	39.0	21.4	18.8
0.5	0.0	100	14.0	18.9	14.2	22.0	11.8	9.2	16.1	28.8	20.9	34.2	14.0	10.0
0.5	0.0	250	15.6	20.0	15.5	21.3	14.0	13.1	20.9	29.0	22.1	32.2	19.7	18.2
−0.5	0.0	100	15.8	23.1	15.7	42.9	13.0	20.1	26.3	28.7	30.0	67.6	21.2	39.5
−0.5	0.0	250	16.8	22.9	15.3	32.2	14.0	19.1	23.5	31.2	23.9	50.2	21.3	29.6
0.0	0.5	100	13.8	21.1	13.2	25.4	11.5	11.4	19.1	31.6	22.3	41.6	15.2	13.6
0.0	0.5	250	17.0	21.5	15.3	22.1	14.1	13.7	22.2	31.6	21.5	34.1	19.5	19.9
0.0	−0.5	100	16.4	22.8	17.6	52.0	15.4	30.9	22.8	27.6	32.1	78.4	25.6	59.1
0.0	−0.5	250	16.0	22.8	16.0	40.2	14.8	24.7	23.6	30.7	24.6	61.4	22.4	41.2

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Case (c)(i) Near-integrated AR SV model with no leverage ($\overline{\omega} = 0$)														
0.0	0.0	100	8.8	14.7	18.1	22.2	14.9	11.8	18.1	20.5	31.4	40.5	21.0	19.1
0.0	0.0	250	8.2	16.2	16.1	20.9	15.3	11.1	16.4	20.9	22.1	30.2	21.1	16.0
0.5	0.0	100	5.7	5.2	16.3	14.8	13.8	8.5	13.5	13.8	23.6	24.7	15.9	8.5
0.5	0.0	250	7.2	9.8	15.5	16.7	14.6	10.8	14.9	16.4	21.1	25.1	20.6	15.9
−0.5	0.0	100	8.5	18.7	17.0	36.1	13.7	21.5	18.0	20.5	29.8	59.6	20.7	39.0
−0.5	0.0	250	7.9	19.0	15.5	29.3	14.9	17.1	15.9	21.2	21.8	40.0	20.2	24.3
0.0	0.5	100	5.7	9.5	15.8	17.8	13.6	10.5	15.5	17.9	23.6	31.6	16.1	12.4
0.0	0.5	250	7.2	12.6	14.9	18.4	14.2	12.0	13.8	19.6	20.7	27.2	19.9	16.8
0.0	−0.5	100	10.4	18.6	19.5	48.1	17.4	35.4	18.5	18.8	32.3	72.9	26.6	58.9
0.0	−0.5	250	8.3	18.4	16.0	38.1	15.6	22.9	15.8	21.0	22.2	51.7	21.6	36.1
Case (c)(ii) Near-integrated AR SV model with leverage ($\overline{\omega} = -0.5$)														
0.0	0.0	100	11.6	17.2	21.4	26.5	18.4	15.2	20.1	19.8	34.9	44.2	25.1	22.3
0.0	0.0	250	11.3	19.6	18.9	23.8	18.3	13.9	17.3	21.1	26.1	34.8	23.5	18.1
0.5	0.0	100	8.7	7.3	17.9	17.3	15.0	9.3	14.7	14.8	24.5	27.7	16.8	8.7
0.5	0.0	250	10.5	13.0	18.5	19.3	17.3	12.9	17.2	17.9	25.7	28.1	23.1	17.9
−0.5	0.0	100	10.7	19.7	20.0	40.9	16.8	26.6	18.6	18.6	32.4	63.6	23.4	43.9
−0.5	0.0	250	11.9	21.8	18.6	32.0	17.8	18.7	18.3	23.0	25.1	44.8	23.1	26.7
0.0	0.5	100	8.2	11.1	17.3	21.1	15.4	11.9	15.8	15.9	25.0	34.5	17.5	13.5
0.0	0.5	250	10.3	15.2	17.5	21.3	16.8	14.1	16.6	20.5	24.4	31.1	22.6	19.7
0.0	−0.5	100	11.5	19.1	21.5	52.6	19.5	39.3	20.1	17.7	34.9	76.0	29.1	63.3
0.0	−0.5	250	11.2	22.0	19.3	42.2	18.8	25.9	18.1	22.9	26.3	57.1	24.2	38.8

t_{ADF}; the latter case will be denoted by $k_{rc} = MAIC$. All simulations were again run using 10,000 Monte Carlo replications and the `rndKMn` function of Gauss 5.0.

Consider first the results for the constant volatility case in panel (a) of Tables 3 and 4. As has been well documented in the unit root literature (see, e.g., Perron and Ng, 1996), the $\hat{Z}_t$ test shows poor finite-sample level control relative to t_{ADF}, displaying very bad over (under) sizing for large negative (positive) values of ρ and ϑ. However, a striking feature of the results in Table 3 is how much better behaved is the bootstrap $\hat{Z}_t^*$ test, especially for $k_{rc} = MAIC$. As an example, with detrended data, constant volatility, and $\rho = -0.5$, $\vartheta = 0$, the $\hat{Z}_t$ test rejects the unit root null around 44% of the time even for $T = 250$ whereas for $\hat{Z}_t^*$ with $k_{rc} = MAIC$ this figure is down to only 6%. Although the wild bootstrap clearly delivers a vast improvement in the behavior of the $\hat{Z}_t$ test, it is important to note that $\hat{Z}_t^*$ is still not as well behaved as the bootstrap t_{ADF}^* test. As with $\hat{Z}_t^*$, t_{ADF}^* displays somewhat better size control for $k_{rc} = MAIC$ than for $k_{rc} = 4$. Indeed, the only case where t_{ADF}^* displays empirical size in excess of 6% when using $k_{rc} = MAIC$ is for detrended data with $\vartheta = -0.5$, where it has size of 7.7%. In the same example, $\hat{Z}_t$ and $\hat{Z}_t^*$ have sizes of 75% and 23%, respectively. Finite-sample power (Table 4) for given values of ρ and ϑ tends to be slightly lower for both $\hat{Z}_t^*$ and t_{ADF}^* when using $k_{rc} = MAIC$, vis-à-vis $k_{rc} = 4$, reflecting a not unexpected finite-sample size-power trade-off in the choice of recoloring lag length. Notice that the power of the $\hat{Z}_t$ and t_{ADF} tests, although size-adjusted, does vary with ρ and ϑ, again consistent with what is found in Perron and Ng (1996). Greater variation in power across ρ and ϑ is seen both in and between the bootstrap t_{ADF}^* and $\hat{Z}_t^*$ tests. Recall that the bootstrap tests have not been size-adjusted.

Turning to the results in panels (b), (c)(i), and (c)(ii) of Tables 3 and 4, we see first from Table 3 that the bootstrap $\hat{Z}_t^*$ and t_{ADF}^* tests that use $k_{rc} = MAIC$ are remarkably well behaved in terms of their empirical size throughout. The dependence of the size of the tests on the form of heteroskedasticity present is much weaker than was seen in Table 1. Moreover, the size distortions due to serial correlation noted in panel (a) tend to be ameliorated in the presence of heteroskedasticity with, if anything, a tendency to undersize in the $\hat{Z}_t^*$ test. This latter effect is, however, not evident when using the fixed recoloring lag length. Like the results in panel (a) of Table 4, power tends to be slightly lower for both $\hat{Z}_t^*$ and t_{ADF}^* when using $k_{rc} = MAIC$, vis-à-vis $k_{rc} = 4$, under heteroskedasticity. As with the results in panel (a), the reported powers (size-adjusted in the case of $\hat{Z}_t$ and t_{ADF}) for a given test can vary significantly across ρ and ϑ, although interestingly the dependence of the power of the tests on the form of heteroskedasticity present is much less strong than was seen in Table 2, consistent with the impact on test size under heteroskedasticity, noted previously. As with the results in Tables 1 and 2, a comparison of the results in (c)(i) and (c)(ii) in both Tables 3 and 4 suggests that leverage again has rather little impact on the behavior of either the standard or bootstrap unit root tests.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

6. CONCLUSIONS

In this paper we have analyzed the large-sample behavior of the Dickey–Fuller and Phillips–Perron classes of unit root statistics when the series under analysis is a (near-) unit root process driven by globally heteroskedastic innovations, empirically important examples of which include nonstationary autoregressive stochastic volatility, models with random volatility jumps, near-integrated GARCH processes, and explosive, nonstationary volatility. Deterministic models of volatility are also permitted within our framework. Assuming that any leverage effects become negligible as the sample size diverges, a wild bootstrap–based solution was shown to solve the associated inference problem. We recommend that the basic wild bootstrap be used in conjunction with a recoloring device as our simulation results show that this greatly improves upon the finite-sample size properties of the unit root tests when the innovations display weak dependence. Numerical evidence presented suggested that our proposed bootstrap procedure works very well in practice, even in the presence of near-integrated autoregressive SV processes with leverage. A formal solution to the inference problem when leverage effects are present but that does not presume a specific parametric model for the volatility process is beyond the scope of this paper but represents an important direction for further research to focus on.

NOTES

1. Here we use OLS detrending, as in Dickey and Fuller (1979). Other methods of detrending, such as first-difference detrending or local–generalized least squares (local-GLS) detrending (see Schmidt and Phillips, 1992 and Elliott, Rothenberg, and Stock, 1996, respectively), can also be used without changing the key findings in this paper—all that would be required is to replace in Theorems 1–4 the OLS-based projection in (1) with the appropriate projection relevant to the method of detrending employed. We focus our attention on OLS-based tests because they are by far the most popular tests used in applied work.

Provided $\{z_{T,t}\}$ spans the space of an intercept, the statistics $T\left(\hat{\alpha}_T - 1\right)$ and t_{DF} will be (exact) similar with respect to $X_{T,0}$. Consequently, in such cases, the asymptotic null distributions given for these statistics in the text remain valid if the assumption that $X_{T,0}$ is of $o_p(T^{1/2})$ is dropped.

2. If $\beta = 0$ in (2) and $\hat{X}_{T,t}$ is consequently replaced by $X_{T,t}$ in calculating $\hat{\alpha}_T$ and t_{DF}, then $\hat{B}_z$ reduces to B in (5).

3. Some authors, e.g., Stock (1994, p. 2745), replace (6) by the condition that the average conditional variance, $T^{-1}\sum_{t=\lfloor Ts\rfloor+1}^{\lfloor Ts'\rfloor} \mathrm{E}\left(u_{T,t}^2|\mathcal{I}_{T,t-1}\right)$, converges a.s. to $\sigma^2(s'-s)$ as $T \to \infty$. The two formulations are, however, equivalent under rather general conditions (see the discussion in Hall and Heyde, 1980) and may be considered interchangeable for our purposes.

4. Where appropriate, elements of the examples in Section 2 could occur simultaneously. For example, combining Examples 1 and 2, yields a jump-diffusion volatility process.

5. That is, the abrupt single variance break model, where $\sigma_{T,t} = \sigma_0 + (\sigma_1 - \sigma_0)\mathbb{I}(t > \lfloor \tau T \rfloor)$, $\mathbb{I}(\cdot)$ denoting the indicator function, and, hence, $\Omega(s) := \sigma_0 + (\sigma_1 - \sigma_0)\mathbb{I}(s \geq \tau)$.

6. Paralleling note 2,$(B_\Omega|Z)$ reduces to B_Ω when $\beta = 0$ in (2) and, commensurately, $X_{T,t}$ replaces $\hat{X}_{T,t}$ in calculating $\hat{\alpha}_T$ and t_{DF}.

7. Hansen (1995) provides a theory of nonparametric (adaptive) variance estimation. Where volatility displays a (possibly countable) number of jumps, as is allowed under the class of volatility processes considered in this paper, then, so far as we are aware, nonparametric (adaptive) variance

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

estimation is infeasible. If, however, the limiting volatility process has continuous sample paths, Hansen's theory may provide an alternative way of estimating the quantiles of the heteroskedastic unit root distribution.

8. Provided $z_{T,t}$ spans the space of an intercept, the bootstrap samples may be initialized at $X^*_{T,0} = 0$. Where this is not the case, one may set $X^*_{T,0} = X_{T,0}$.

9. The value $\overline{\omega} = -0.5$ appears to be reasonably large, relative to observed levels of leverage. For example, Omori, Chibo, Shephard, and Nakajimaa (2007) report a 95% interval for leverage in the Tokyo Stock Exchange Price Index (5 January 1998, through 30 December 2002) of $\overline{\omega} \in [-0.59, -0.11]$.

Other values of c^* and other combinations of the parameter values given in Section 4.3 for cases (b) and (c) were also considered, as were experiments for smooth transition volatility shifts and near-integrated GARCH processes but qualitatively add little to the reported results.

10. Recoloring has also been used in the context of the $\hat{Z}_{\hat{\alpha}_T}$ and $\hat{Z}_t$ statistics by Lee and Phillips (1994), who construct an estimate of the long-run variance using prewhitened data and then recolor the resulting estimator. They show that this can improve upon the finite-sample size properties of the $\hat{Z}_{\hat{\alpha}_T}$ and $\hat{Z}_t$ tests relative to using standard kernel estimators of the long-run variance.

11. Chang and Park (2003) suggest estimating the stationary lag dynamics under the unit root null hypothesis, i.e., by fitting the autoregression $\Delta \hat{X}_{T,t} = \sum_{i=1}^{k} \beta_i \Delta \hat{X}_{T,t-i} + u_{T,t;k}$. This equation differs from the usual ADF regression (26) through the omission of $\hat{X}_{T,t-1}$ from the set of regressors. However, in a number of Monte Carlo experiments we found that the inclusion of $\hat{X}_{T,t-1}$ generally led to improved finite-sample behavior in the resulting tests.

12. Notice that the recursion in (22) reduces to the scheme in (27) when $k = 0$.

13. For instance, Chang and Park (2003) and Cavaliere and Taylor (2007) require that the lag polynomial satisfies $c(z) \neq 0$ for all $|z| \leq 1$. This condition is likely to be violated in the case of seasonally adjusted data, where the lag polynomial will usually be noninvertible at some or all of the seasonal frequencies. Conversely, this condition is not required for the $\hat{Z}_{\hat{\alpha}_T}$ and $\hat{Z}_t$ tests or for their bootstrap analogues.

14. Results for the other parameter configurations of cases (b) and (c) from Tables 1 and 2 were qualitatively similar to those reported and, hence, are omitted.

REFERENCES

Andrews, D. W. K. (1988) Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory* 4, 458–467.

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. & M. Buchinsky (2001) Evaluation of a three-step method for choosing the number of bootstrap repetitions. *Journal of Econometrics* 103, 345–386.

Barndorff-Nielsen, O.E. & N. Shephard (2001) Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society, Series B* 63, 167–241.

Basawa, I.V., A.K. Mallik, W.P. McCormick, J.H. Reeves, & R.L. Taylor (1991) Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics* 19, 1098–1101.

Beare, B. (2005) Robustifying Unit Root Tests to Permanent Changes in Innovation Variance. Mimeo, Yale University.

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

Boswijk, H.P. (2001) Testing for a Unit Root with Near Integrated Volatility. Tinbergen Institute Discussion paper 2001-077/4.

Boswijk, H.P. (2005) Adaptive Testing for a Unit Root with Nonstationary Volatility. UvA-Econometrics Discussion paper 2005/07.

Burridge, P. & A.M.R. Taylor (2001) On regression-based tests for seasonal unit roots in the presence of periodic heteroscedasticity. *Journal of Econometrics* 104, 91–117.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

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

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 (2008) Bootstrap unit root tests for time series with non-stationary volatility. *Econometric Theory* 24, 43–71.

Cavaliere, G. & A.M.R. Taylor (2009) Bootstrap *M* unit root tests. *Econometric Reviews* 28, 393–421.

Chan, N.H. & L.T. Tran (1989) On the first order autoregressive process with infinite variance. *Econometric Theory* 5, 354–362.

Chang, Y. & J.Y. Park (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431–447.

Chang, Y. & J.Y. Park (2003) A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis* 24, 379–400.

Clark, P.K. (1973) A subordinated stockastic process model with finite variance for speculative prices. *Econometrica* 41, 135–155.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

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

de Jong, R. & J. Davidson (2000) The functional central limit theorem and weak convergence to stochastic integrals, part 1. *Econometric Theory* 16, 621–642.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Dickey, D.A. & W.A. Fuller (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49, 1057–1072.

Duffie, D., J. Pan, & K. Singleton (2000) Transformation analysis and asset pricing for affine jump-diffusions. *Econometrica* 68, 1343–1376.

Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Ferretti, N. & J. Romo (1996) Bootstrap tests for unit root AR(1) models. *Biometrika* 84, 849–860.

Georgiev, I. (2008) Asymptotics for cointegrated processes with infrequent stochastic level shifts and outliers. *Econometric Theory* 24, 587–615.

Giné, E. & J. Zinn (1990) Bootstrapping general empirical measures. *Annals of Probability* 18, 851–869.

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

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

Hall, P. (1977) Martingale invariance principles. *Annals of Probability* 5, 875–887.

Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications.* Academic Press.

Hansen, B.E. (1992a) Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.

Hansen, B.E. (1992b) Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.

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

Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.

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

Inoue, A. & L. Kilian (2002) Bootstrapping autoregressive processes with possible unit roots. *Econometrica* 70, 377–391.

Jansson, M. (2002) Consistent covariance matrix estimation for linear processes. *Econometric Theory* 18, 1449–1459.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Kim, C.-J. & C.R. Nelson (1999) Has the US economy become more stable? A Bayesian approach based on a Markov-switching model of the business cycle. *Review of Economics and Statistics* 81, 608–616.

Kurtz, T.G. & P. Protter (1991) Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability* 19, 1035–1070.

Lee, C.C. & P.C.B. Phillips (1994) An ARMA Pre-whitened Long-Run Variance Estimator. Manuscript, Yale University.

Ling, S., W.K. Li, & M. McAleer (2003) Estimation and testing for unit root process with GARCH(1,1) errors: Theory and Monte Carlo evidence. *Econometric Reviews* 22, 179–202.

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

Loretan, M. & P.C.B. Phillips (1994) Testing covariance stationarity under moment condition failure with an application to common stock returns. *Journal of Empirical Finance* 1, 211–248.

MacKinnon, J. (2006) Bootstrap methods in econometrics. *Economic Record* 82, S2–S18.

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

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.

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

Nelson, C.R., J. Piger, & E. Zivot (2001) Markov regime switching and unit-root tests. *Journal of Business & Economic Statistics* 19, 404–415.

Nelson, D.B. (1990a) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.

Nelson, D.B. (1990b) ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–38.

Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.

Omori, Y., S. Chib, N. Shephard, & J. Nakajimaa (2007) Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics* 140, 425–449.

Paparoditis, E.P. & D.N. Politis (2003) Residual-based block bootstrap for unit root testing. *Econometrica* 71, 813–855.

Park, J.Y. (2002a) Nonstationary nonlinear heteroskedasticity. *Journal of Econometrics* 110, 383–415.

Park, J.Y. (2002b) An invariance principle for sieve bootstrap in time series. *Econometric Theory* 18, 469–490.

Park, J.Y. (2003) Bootstrap unit root tests. *Econometrica* 71, 1845–1895.

Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69, 117–161.

Perron, P. & S. Ng (1996) Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63, 435-463.

Phillips, P.C.B. (1987a) Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (1987b) Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. (1990) Time series regression with a unit root and infinite variance errors. *Econometric Theory* 6, 44–62.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Phillips, P.C.B. & Z. Xiao (1998) A primer on unit root testing. *Journal of Economic Surveys* 12, 423–470.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

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

Revuz, D. & M. Yor (1991) *Continuous Martingales and Brownian Motion*. Springer-Verlag.

Roberts, G.O., O. Papaspiliopoulos, & P. Dellaportas (2004) Bayesian inference for non-Gaussian Ornstein–Uhlenbeck stochastic volatility processes. *Journal of the Royal Statistical Society, Series B* 66, 369–393.

Schmidt, P. & P.C.B. Phillips (1992) LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54, 257–287.

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

Shephard, N. (2005) *Stochastic Volatility: Selected Readings*. Oxford University Press.

Stock, J.H. (1994) Unit roots, structural breaks and trends. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2739–2840. Elsevier Science.

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

Wooldridge, J.M. & H. White (1988) Some invariance principles and central limit theorems for dependent heterogeneous processes. *Econometric Theory* 4, 210–230.

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

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

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

APPENDIX

To simplify exposition, results are proved for the case of no deterministic variables. The results stated in the main text hold simply by replacing B_Ω by its detrended counterpart, $\hat{B}_\Omega$, throughout. Moreover, without loss of generality we set $X_{T,0} = 0$, as the proofs trivially extend to the case where $X_{T,0}$ is of $o_p(T^{1/2})$.

Before moving to the main proofs, we first state and prove some preparatory results.

LEMMA A.1. *If* $X_T(\cdot) := X_{T,\lfloor T\cdot\rfloor} \xrightarrow{w} X(\cdot)$ *in the space* $\mathcal{D}_{\mathbb{R}}[0,1]$ *endowed with the Skorokhod topology, and if* $0 < q \leq BT^{\gamma}$ *where* $0 < B < \infty$ *and* $0 \leq \gamma < 1$, *then*

$$\sup_{k=1,\ldots,q} \frac{1}{T} \sum_{t=1}^{T-q} \left| X_{T,t+k} - X_{T,t} \right| \xrightarrow{p} 0. \tag{A.1}$$

Proof. When $X(\cdot)$ has continuous sample paths a.s., the proof follows from Lemma A.1 in Hansen (1995) using the inequality $\sup_k T^{-1} \sum_{t=1}^{T-q} |X_{T,t-k} - X_{T,t}| \leq \max_{t\leq T;k\leq q} \left| X_{T,t-k} - X_{T,t} \right|$. We now prove the result in the general case.

Because $X(\cdot)$ lies in $\mathcal{D}_{\mathbb{R}}[0,1]$ and $X_T(\cdot) := X_{T,\lfloor T\cdot\rfloor}$ converges weakly, the associated sequence of probability measures is tight. A necessary condition for tightness in the space $\mathcal{D}_{\mathbb{R}}[0,1]$ endowed with the Skorokhod topology (cf. Billingsley, 1968, Thm. 15.2) is that for all $\varepsilon > 0$ and $\eta > 0$ there exist a $\delta \in (0,1)$ and an integer T_0 such that, for

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

all $T \geq T_0$,

$$P\left(\inf_{\Pi_\delta}\left\{\max_{i=1,\ldots,r}\left\{\sup_{s,s'\in[s_{i-1},s_i)}\left|X_T(s)-X_T\left(s'\right)\right|\right\}\right\}\geq\frac{\varepsilon}{2}\right)\leq\frac{\eta}{2}, \tag{A.2}$$

where Π_δ denotes a partition of $[0,1]$ of the form $\{s_1,\ldots,s_r\}$ with $r\leq\lfloor 1/\delta\rfloor$ and $\min_i\delta_i>\delta$, $\delta_i:=s_i-s_{i-1}$ (we set $s_0=0$ and $s_r=1$ in what follows). Now, let δ be such that (A.2) is satisfied for large T and let $T_i:=\lfloor s_iT\rfloor$, $i=0,\ldots,r$. We have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T-q}\left|X_{T,t+k}-X_{T,t}\right|&=\frac{1}{T}\sum_{i=1}^{r}\sum_{t=T_{i-1}+1}^{T_i\wedge(T-q)}\left|X_{T,t+k}-X_{T,t}\right|\\&=\frac{1}{T}\sum_{i=1}^{r}\sum_{t=T_{i-1}+1}^{T_i-q}\left|X_{T,t+k}-X_{T,t}\right|\\&\quad+\frac{1}{T}\sum_{i=1}^{r}\sum_{t=T_i-q+1}^{T_i\wedge(T-q)}\left|X_{T,t+k}-X_{T,t}\right|\\&=R_{1T}(k)+R_{2T}(k)\end{aligned}$$

with $R_{1T}(k)$ and $R_{2T}(k)$ defined implicitly. Moreover,

$$\sup_{k=0,\ldots,q}R_{1T}(k)\leq\max_{i=1,\ldots,r}\left\{\sup_{s,s'\in[s_{i-1},s_i)}\left|X_T(s)-X_T\left(s'\right)\right|\right\},$$

$$\sup_{k=0,\ldots,q}R_{2T}(k)\leq\frac{2q}{\delta T}\sup_{s\in[0,1]}\left|X_T(s)\right|,$$

which implies

$$\begin{aligned}\sup_{k=0,\ldots,q}\frac{1}{T}\sum_{t=1}^{T-q}\left|X_{T,t-k}-X_{T,t}\right|&\leq\max_{i=1,\ldots,r}\left\{\sup_{s,s'\in[s_{i-1},s_i)}\left|X_T(s)-X_T\left(s'\right)\right|\right\}\\&\quad+\frac{2q}{\delta T}\sup_{s\in[0,1]}\left|X_T(s)\right|\end{aligned}$$

for sufficiently large T. By the arbitrariness of Π_δ,

$$\begin{aligned}&\sup_{k=0,\ldots,q}\frac{1}{T}\sum_{t=1}^{T-q}\left|X_{T,t+k}-X_{T,t}\right|\\&\quad\leq\inf_{\Pi_\delta}\left\{\max_{i=1,\ldots,r}\left\{\sup_{s,s'\in[s_{i-1},s_i)}\left|X_T(s)-X_T\left(s'\right)\right|\right\}+\frac{2q}{\delta T}\sup_{s\in[0,1]}\left|X_T(s)\right|\right\}\\&\quad=\inf_{\Pi_\delta}\left\{\max_{i=1,\ldots,r}\left\{\sup_{s,s'\in[s_{i-1},s_i)}\left|X_T(s)-X_T\left(s'\right)\right|\right\}\right\}+\frac{2q}{\delta T}\sup_{s\in[0,1]}\left|X_T(s)\right|,\end{aligned}$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

and, hence,

$$P\left(\sup_{k=0,\ldots,q}\frac{1}{T}\sum_{t=1}^{T-q}\left|X_{T,t+k}-X_{T,t}\right|\geq\varepsilon\right)$$

$$\leq P\left(\inf_{\Pi_\delta}\left\{\max_{i=1,\ldots,r}\left\{\sup_{s,s'\in[s_{i-1},s_i)}\left|X_T(s)-X_T\left(s'\right)\right|\right\}\right\}\geq\frac{\varepsilon}{2}\right)$$

$$+P\left(\frac{2q}{\delta T}\sup_{s\in[0,1]}|X_T(s)|\geq\frac{\varepsilon}{2}\right)$$

$$\leq\frac{\eta}{2}+o(1)\leq\eta$$

for large T. ■

Remark A.1. Lemma A.1 is a basic ingredient needed for proving Theorems A.1 and A.2 in this Appendix. Theorem A.1 generalizes an important result given in Hansen (1992a, Thm. 3.3), demonstrating that L^1 mixingales are (asymptotically) uncorrelated with arrays having a limit in the space $\mathcal{D}_{\mathbb{R}}[0,1]$ endowed with the Skorokhod topology, even if the limit $X(\cdot)$ is *not* a.s. continuous (the continuous case is covered by Hansen's theorem). Theorem A.2 then shows that the squared innovation process is useful for estimating the integrated variation of the process $\int_0^{\cdot}\Omega(r-)\,dB(r)$.

THEOREM A.1. *Suppose that* $X_T(\cdot):=X_{T,\lfloor T\cdot\rfloor}\xrightarrow{w}X(\cdot)$ *in the space* $\mathcal{D}_{\mathbb{R}}[0,1]$ *endowed with the Skorokhod topology. Moreover, for a random array* $\{e_{T,t}\}$ *and a sequence of nondecreasing sigma fields* $\{\mathcal{F}^e_{T,t}\}$ *to which* $\{e_{T,t}\}$ *is adapted, assume that* $\sup_t \mathrm{E}\left|\mathrm{E}\left(e_{T,t}|\mathcal{F}^e_{T,t-m}\right)\right|\to 0$ *as* $m\to\infty$. *Then*

$$\sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor}X_{T,t}e_{T,t}\right|\xrightarrow{p}0.$$

Proof. The proof uses some of the arguments employed to prove Lemma A.1. Specifically, recall that (using the same notation as in the proof of Lemma A.1) for given ε,η, there exists a $\delta\in(0,1)$ such that

$$P\left(\inf_{\Pi_\delta}\left\{\max_{i=1,\ldots,r}\left\{\sup_{s,s'\in[s_{i-1},s_i)}\left|X_T(s)-X_T\left(s'\right)\right|\right\}\right\}\geq\frac{\varepsilon}{2}\right)\leq\frac{\eta}{2}$$

for T large enough. Moreover, let $T_{si}:=T_i\wedge\lfloor Ts\rfloor$ and $r_s:=\min\{i\in\{1,\ldots,r\}:T_i\geq\lfloor Ts\rfloor\}$. Then, similarly to Hansen (1992a, proof of Thm. 3.3) we may write

$$\sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{t=1}^{\lfloor Ts\rfloor}X_{T,t}e_{T,t}\right|=\sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{i=1}^{r_s}\sum_{t=T_{i-1}+1}^{T_{si}}X_{T,t}e_{T,t}\right|$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

$$
\begin{aligned}
&\leq \sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{i=1}^{r_s} X_{T,T_{i-1}+1}\sum_{t=T_{i-1}+1}^{T_{si}} e_{T,t}\right| \\
&\quad+\sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{i=1}^{r_s}\sum_{t=T_{i-1}+1}^{T_{si}}\left(X_{T,t}-X_{T,T_{i-1}+1}\right)e_{T,t}\right| \\
&\leq \sup_{s\in[0,1]}\sum_{i=1}^{r_s}\left|X_{T,T_{i-1}+1}\right|\left|\frac{1}{T}\sum_{t=T_{i-1}+1}^{T_{si}} e_{T,t}\right| \\
&\quad+\sup_{s\in[0,1]}\frac{1}{T}\sum_{i=1}^{r_s}\sum_{t=T_{i-1}+1}^{T_{si}}\left|X_{T,t}-X_{T,T_{i-1}+1}\right|\left|e_{T,t}\right| \\
&\leq \sum_{i=1}^{r}\left|X_{T,T_{i-1}+1}\right|\left|\frac{1}{T}\sum_{t=T_{i-1}+1}^{T_i} e_{T,t}\right| \\
&\quad+\left|X_{T,T_{r_s-1}+1}\right|\sup_{s\in[0,1]}\left|\frac{1}{T}\sum_{t=T_{r_s-1}+1}^{\lfloor Ts\rfloor} e_{T,t}\right| \\
&\quad+\frac{1}{T}\sum_{i=1}^{r}\sum_{t=T_{i-1}+1}^{T_i}\left|X_{T,t}-X_{T,T_{i-1}+1}\right|\left|e_{T,t}\right| \\
&\leq \sup_{s\in[0,1]}\left|X_T(s)\right| \\
&\quad\times\left(\sum_{i=1}^{r}\left|\frac{1}{T}\sum_{t=T_{i-1}+1}^{T_i} e_{T,t}\right|+2\max_{t=1,\ldots,T}\left|\frac{1}{T}\sum_{i=1}^{t} e_{T,t}\right|\right) \\
&\quad+\frac{1}{T}\sum_{i=1}^{r}\sum_{t=T_{i-1}+1}^{T_i}\left|X_{T,t}-X_{T,T_{i-1}+1}\right|\left|e_{T,t}\right|. \qquad \textbf{(A.3)}
\end{aligned}
$$

Because $X_T(\cdot)$ converges weakly, $\sup_{s\in[0,1]}|X_T(s)| = O_p(1)$. Moreover, $\max_t \left|T^{-1}\sum_{i=1}^{t} e_{T,t}\right| \to 0$ in probability (Hansen, 1992a, Cor. to Thm. 3.3) and $T^{-1}\sum_{i=1}^{r}\left|\sum_{t=T_{i-1}+1}^{T_i} e_{T,t}\right| \xrightarrow{p} 0$. To see the latter result notice that, as in Andrews (1988, proof of Thm. 1), for any $\varepsilon > 0$ there exists a constant $B < \infty$ such that

$$
\begin{aligned}
\mathrm{E}\left(\frac{1}{T}\sum_{i=1}^{r}\left|\sum_{t=T_{i-1}+1}^{T_i} e_{T,t}\right|\right) &= \frac{1}{T}\sum_{i=1}^{r}\lfloor\delta_i T\rfloor\,\mathrm{E}\left|\frac{\sum_{t=T_{i-1}+1}^{T_i} e_{T,t}}{\lfloor\delta_i T\rfloor}\right| \\
&\leq \frac{1}{T}\sum_{i=1}^{r}\lfloor\delta_i T\rfloor\left(\frac{B}{\lfloor\delta_i T\rfloor^{1/2}}+\frac{\varepsilon}{2}\right) \\
&= \frac{B}{T}\sum_{i=1}^{r}\lfloor\delta_i T\rfloor^{1/2}+\frac{\varepsilon}{2},
\end{aligned}
$$

which, for T sufficiently large, is less than ε.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Finally, to show that the last term on the right-hand side of (A.3) also converges to zero in probability, consider the inequality

$$\frac{1}{T}\sum_{i=1}^{r}\sum_{t=T_{i-1}+1}^{T_i}\left|X_{T,t}-X_{T,T_{i-1}+1}\right|\left|e_{T,t}\right|$$

$$\leq \max_{i=1,\ldots,r}\max_{t=T_{i-1}+1,\ldots,T_i}\left|X_{T,t}-X_{T,T_{i-1}+1}\right|\frac{1}{T}\sum_{t=1}^{T}\left|e_{T,t}\right|.$$

By the arbitrariness of Π_δ, the term $\max_i \max_t \left|X_{T,t}-X_{T,T_{i-1}+1}\right|$ can be made arbitrarily small, and hence the desired result follows as, in addition, $T^{-1}\sum_{t=1}^{T}\left|e_{T,t}\right| = O_p(1)$ because $e_{T,t}$ is uniformly integrable. ■

THEOREM A.2. *Under Assumptions 1–3,*

$$V_T(\cdot) := \frac{1}{a_T^2 T}\sum_{t=1}^{\lfloor T\cdot\rfloor} u_{T,t}^2 \xrightarrow{w} V(\cdot), \tag{A.4}$$

where $V(s) := \int_0^s \Omega(r)^2\,dr$.

Proof. Let $\xi_{T,t} := \varepsilon_{T,t}^2 - 1$ and write $V_T(s) = a_T^{-2}T^{-1}\sum_{t=1}^{\lfloor Ts\rfloor}\sigma_{T,t-1}^2\varepsilon_{T,t}^2 = V_T^a(s) + V_T^b(s)$, where $V_T^a(s) := a_T^{-2}T^{-1}\sum_{t=1}^{\lfloor Ts\rfloor}\sigma_{T,t-1}^2$ and $V_T^b(s) := a_T^{-2}T^{-1}\sum_{t=1}^{\lfloor Ts\rfloor}\sigma_{T,t-1}^2\xi_{T,t}$ $= a_T^{-2}T^{-1}\sum_{t=1}^{\lfloor Ts\rfloor}\sigma_{T,t}^2\xi_{T,t+1} - a_T^{-2}T^{-1}\sigma_{T,\lfloor Ts\rfloor}^2\xi_{T,\lfloor Ts\rfloor+1} + a_T^{-2}T^{-1}\sigma_{T,0}^2\xi_{T,1}$, where we put $\xi_{T,T+1} := 0$. By Assumption 2, $a_T^{-1}\sigma_T(\cdot) \xrightarrow{w} \Omega(\cdot)$, and an application of the continuous mapping theorem (CMT) yields $V_T^a(s) = a_T^{-2}\int_0^s \sigma_T(r)^2\,dr \xrightarrow{w} \int_0^s \Omega(r)^2\,dr$. Regarding $V_T^b(s)$, we have

$$\sup_{s\in[0,1]}\left|a_T^{-2}T^{-1}\sigma_{T,\lfloor Ts\rfloor}^2\xi_{T,\lfloor Ts\rfloor+1} - a_T^{-2}T^{-1}\sigma_{T,0}^2\xi_{T,1}\right|$$

$$\leq 2\sup_{s\in[0,1]}\left|a_T^{-2}\sigma_{\lfloor Ts\rfloor}^2\right|\sup_{s\in[0,1]}\left|T^{-1}\xi_{T,\lfloor Ts\rfloor}\right| = o_p(1)$$

because from Assumption 1, $\sup_{s\in[0,1]} a_T^{-2}\sigma_{T,\lfloor Ts\rfloor}^2 = O_p(1)$ and, from Assumption 2, $\{\xi_{T,t}\}$ is a uniformly integrable MDA. Moreover, $\{\xi_{T,t}\}$ is also a uniformly integrable L^1 mixingale (as shown, e.g., in Hall and Heyde, 1980, p. 19); hence, as $a_T^{-2}\sigma_T(\cdot)^2 \xrightarrow{w} \Omega(\cdot)^2$ we can make use of Theorem A.1 to obtain

$$\sup_{s\in[0,1]}\left|V_T^b(s)\right| = \sup_{s\in[0,1]}\left|\frac{1}{a_T^2 T}\sum_{t=1}^{\lfloor Ts\rfloor}\sigma_{T,t-1}^2\xi_{T,t}\right|$$

$$= \sup_{s\in[0,1]}\left|\frac{1}{a_T^2 T}\sum_{t=1}^{\lfloor Ts\rfloor}\sigma_{T,t}^2\xi_{T,t+1}\right| + o_p(1) \xrightarrow{p} 0,$$

as required. ■

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

A useful corollary of Theorem A.2 is the following result for the sample second moment of $\{u_{T,t}\}$.

COROLLARY A.1. *For* $\hat{\sigma}^2 := T^{-1}\sum_{t=1}^{T} u_{T,t}^2$, *we have the weak convergence result* $a_T^{-2}\hat{\sigma}^2 = V_T(1) \xrightarrow{w} V(1) = \int_0^1 \Omega(r)^2\,dr =: \bar{\Omega}^2$.

Proof of Lemma 1. The proof of the FCLT given in the lemma can basically be obtained as in Hansen (1995, Lem. A.6), although the continuity of the limiting volatility process required under his Assumption 2 need not in fact be assumed. To see this fact, notice first that $X_T(s) = a_T^{-1}T^{-1/2}\sum_{t=1}^{\lfloor Ts\rfloor} u_{T,t} = T^{-1/2}\sum_{t=1}^{\lfloor Ts\rfloor} a_T^{-1}\sigma_{T,t-1}\varepsilon_{T,t}$, where $\{\varepsilon_{T,t}\}$ is a uniformly square integrable MDA with unit conditional variance. This partial sum can therefore be written as the stochastic integral

$$X_T(s) = a_T^{-1}\int_0^s \sigma_T(r)\,dS_T(r),$$

where $S_T(\cdot)$ is defined in (7). Because both $\sigma_T(s)$ and $S_T(s)$ are adapted to $\mathcal{I}_{T,\lfloor Ts\rfloor}$, and under Assumptions 1–3 the joint convergence $\left(a_T^{-1}\sigma_T(\cdot), S_T(\cdot)\right) \xrightarrow{w} (\Omega(\cdot), B(\cdot))$ holds, we may proceed as in Hansen (1992a, Thm. 2.1) where it is shown that if the following condition holds:

$$\sup_{s\in[0,1]} T^{-1}\sum_{t=1}^{\lfloor Ts\rfloor} \mathrm{E}(\varepsilon_{T,t}^2) < \infty \tag{A.5}$$

then we may apply Theorem 4.6 of Kurtz and Protter (1991) to obtain that on $\mathcal{D}_{\mathbb{R}^3}[0,1]$

$$\left(a_T^{-1}\sigma_T(\cdot), S_T(\cdot), X_T(\cdot)\right) \xrightarrow{w} \left(\Omega(\cdot), B(\cdot), \int_0^{\cdot}\Omega(r-)\,dB(r)\right). \tag{A.6}$$

Condition (A.5) is trivially verified because, by construction, the conditional expectation $\mathrm{E}(\varepsilon_{T,t}^2|\mathcal{I}_{T,t-1})$ equals unity with probability one.

The joint convergence result in (A.6) finally implies that

$$X_T(\cdot) \xrightarrow{w} \int_0^{\cdot}\Omega(r-)\,dB(r) = \left(\int_0^1 \Omega(r)^2\,dr\right)^{1/2} B_\Omega(\cdot),$$

which establishes the invariance principle in (13). ■

Proof of Theorem 1. Following Phillips (1987b), write the normalized first-order sample autoregressive coefficient as

$$T\left(\hat{\alpha}_T - 1\right) = \frac{X_T(1)^2 - X_T(0)^2 - a_T^{-2}\hat{\sigma}^2}{2\int_0^1 X_T(s)^2\,ds}, \tag{A.7}$$

where $X_T(s) := a_T^{-1}T^{-1/2}X_{T,\lfloor Ts\rfloor}$ and $\hat{\sigma}^2 := T^{-1}\sum_{t=1}^{T} u_{T,t}^2$. By Lemma 1,

$$X_T(\cdot) \xrightarrow{w} m_\Omega(\cdot) := \left(\int_0^1 \Omega(s)^2\,ds\right)^{1/2} B_\Omega(\cdot).$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Moreover, from (A.6) and applications of the CMT,

$$\begin{pmatrix} X_T(1)^2 - X_T(0)^2 \\ \int_0^1 X_T(s)^2\,ds \end{pmatrix} \xrightarrow{w} \int_0^1 \Omega(s)^2\,ds \begin{pmatrix} B_\Omega(1)^2 - B_\Omega(0)^2 \\ \int_0^1 B_\Omega(s)^2\,ds \end{pmatrix}. \tag{A.8}$$

Next, consider the term $\hat{\sigma}^2$. Because $\hat{\sigma}^2 = a_T^2 V_T(1)$ we can apply Corollary A.1 to obtain $a_T^{-2}\hat{\sigma}^2 \xrightarrow{w} V(1) = \int_0^1 \Omega(s)^2\,ds$. Combining this result with (A.8) and appealing to the CMT we therefore have from (A.7) that

$$\begin{aligned} T(\hat{\alpha}_T - 1) &\xrightarrow{w} \frac{B_\Omega(1)^2 \int_0^1 \Omega(s)^2\,ds - B_\Omega(0)^2 \int_0^1 \Omega(s)^2\,ds - \int_0^1 \Omega(s)^2\,dr}{2\int_0^1 B_\Omega(s)^2\,ds \int_0^1 \Omega(s)^2\,dr} \\ &= \frac{B_\Omega(1)^2 - B_\Omega(0)^2 - 1}{2\int_0^1 B_\Omega(s)^2\,ds} \end{aligned}$$

as required.

To prove the result for the studentized statistic, t_{DF}, observe first from (A.8) and the CMT that

$$a_T^{-1}\left(\sum_{t=1}^T X_{T,t-1}^2\right)^{1/2}(\hat{\alpha}_T - 1) \xrightarrow{w} \left(\int_0^1 \Omega(s)^2\,ds\right)^{1/2} \frac{\left(B_\Omega(1)^2 - B_\Omega(0)^2 - 1\right)}{2\left(\int_0^1 B_\Omega(s)^2\,ds\right)^{1/2}}.$$

Consequently, all that remains is to show that $a_T^{-2}\hat{\sigma}_u^2$ weakly converges to $\int_0^1 \Omega(s)^2\,ds$. Now, because $a_T^{-2}\hat{\sigma}_u^2 = a_T^{-2}\hat{\sigma}^2 + O_p(T^{-1})$ (see the proof of Theorem 2, which follows), the required result obtains from the fact that $a_T^{-2}\hat{\sigma}^2 \xrightarrow{w} \int_0^1 \Omega(s)^2\,ds$, as demonstrated previously. ■

Proof of Theorem 2. We present a proof of the results relating only to the $T(\hat{\alpha}_T^* - 1)$-based test; the extension of the proof to cover the studentized test is straightforward and, hence, omitted. As in the proof of Theorem 1, it is useful to write the bootstrap unit root statistic as

$$T(\hat{\alpha}_T^* - 1) = \frac{X_T^*(1)^2 - X_T^*(0)^2 - a_T^{-2}(\hat{\sigma}^*)^2}{2\int_0^1 X_T^*(s)^2\,ds}, \tag{A.9}$$

where $X_T^*(\cdot) := a_T^{-1}T^{-1/2}X_{T,\lfloor T\cdot\rfloor}^* = a_T^{-1}T^{-1/2}\sum_{t=1}^{\lfloor T\cdot\rfloor} u_{T,t}^*$ and $(\hat{\sigma}^*)^2 := T^{-1}\sum_{t=1}^T (u_{T,t}^*)^2$, $u_{T,t}^* := \hat{u}_{T,t}w_t$.

We first establish a weak convergence (in probability) result for the scaled bootstrap partial sum $X_T^*(\cdot)$. To this aim we make use of the convergence $\hat{V}_T(\cdot) \xrightarrow{w} V(\cdot)$, which follows from Theorem A.2 if

$$\sup_{s\in[0,1]}\left|\hat{V}_T(s) - a_T^{-2}T^{-1}\sum_{t=1}^{\lfloor Ts\rfloor} u_{T,t}^2\right| = \sup_{s\in[0,1]}\left|\hat{V}_T(s) - V_T(s)\right| \xrightarrow{p} 0. \tag{A.10}$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

To establish the validity of (A.10) it suffices to consider the standard equality $\hat{u}_{T,t} = u_{T,t} - \left(\hat{\alpha}_T - 1\right) X_{T,t-1}$, which implies that

$$\sup_{s\in[0,1]} \left|\hat{V}_T(s) - V_T(s)\right| \leq T\left(\hat{\alpha}_T - 1\right)^2 \left(\frac{1}{a_T^2 T^2} \sum_{t=1}^{T} X_{T,t-1}^2\right) + 2T\left|\hat{\alpha}_T - 1\right| \times \sup_{t\leq T}\left|\frac{X_{T,t}}{a_T T^{1/2}}\right| \sup_{t\leq T}\left(\frac{\sigma_{T,t}}{a_T}\right) \frac{1}{T^{3/2}} \sum_{t=1}^{T} \left|\varepsilon_{T,t}\right| = o_p(1)$$

because (a) $\left(\hat{\alpha}_T - 1\right)$ is of $O_p\left(T^{-1}\right)$ (Theorem 1), (b) $a_T^{-1} \sup_{t\leq T} T^{-1/2}|X_{T,t}|$ and $a_T^{-2}T^{-2}\sum_{t=1}^{T} X_{T,t-1}^2$ are all of $O_p(1)$ (Lemma 1 and the CMT), (c) $\sup_t a_T^{-1}\sigma_{T,t} = O_p(1)$ (Assumption 2), and (d) $T^{-3/2}\sum_{t=1}^{T}\left|\varepsilon_{T,t}\right| = o_p(1)$ (from Assumption 1).

Because w_t is Gaussian, conditionally on the sample, $X_T^*(\cdot)$ is a zero mean Gaussian process with covariance kernel $\mathcal{V}_T(s,s') := \mathrm{E}^*\left(X_T^*(s) X_T^*\left(s'\right)\right) = \hat{V}_T(\min(s,s'))$, where $\hat{V}_T(\cdot) := a_T^{-2}T^{-1}\sum_{t=1}^{\lfloor T\cdot\rfloor} \hat{u}_{T,t}^2$ and where E^* is used to denote expectation taken under P^*, the bootstrap probability measure. Now, as $\hat{V}_T(\cdot) \xrightarrow{w} V(\cdot)$ (see the preceding discussion), where under Assumption 4, V is the covariance kernel of m_Ω of (13), it immediately follows that $X_T^*(\cdot) \xrightarrow{w} m_\Omega(\cdot)$ in probability, (cf. Hansen, 1996, among others).

Notice that from the CMT we also have the joint weak convergence result, $\left(X_T^*(1)^2 - X_T^*(0)^2, \int_0^1 X_T^*(s)^2 ds\right) \xrightarrow{w}_p \left(m_\Omega(1)^2 - m_\Omega(0)^2, \int_0^1 m_\Omega(s)^2 ds\right)$.

It is left to demonstrate that $a_T^{-2}(\hat{\sigma}^*)^2$ converges weakly in probability to $\int_0^1 \Omega(s)^2 ds$. Similarly to the proof of Theorem A.2, letting $\zeta_t := w_t^2 - 1$ we can write

$$\frac{1}{a_T^2}(\hat{\sigma}^*)^2 = \frac{1}{a_T^2 T}\sum_{t=1}^{T} \hat{u}_{T,t}^2 w_t^2 = \frac{1}{a_T^2 T}\sum_{t=1}^{T} \hat{u}_{T,t}^2 + \frac{1}{a_T^2 T}\sum_{t=1}^{T} \hat{u}_{T,t}^2 \zeta_t = A_T^a + A_T^b$$

with A_T^a and A_T^b defined implicitly. Because, $A_T^a := \hat{V}_T(1) \xrightarrow{w} V(1)$ and because for all $\varepsilon > 0$,

$$P^*\left(\left|A_T^b - \frac{1}{a_T^2 T}\sum_{t=1}^{T} u_{T,t}^2 \zeta_t\right| > \varepsilon\right) = P^*\left(\left|\frac{1}{a_T^2 T}\sum_{t=1}^{T}\left(\hat{u}_{T,t}^2 - u_{T,t}^2\right)\zeta_t\right| > \varepsilon\right) \leq \frac{\sup_{t\leq T} a_T^{-2}|\hat{u}_{T,t}^2 - u_{T,t}^2|}{\varepsilon}\mathrm{E}^*\left(\frac{1}{T}\sum_{t=1}^{T}|\zeta_t|\right) \xrightarrow{p} 0,$$

we simply need to establish that $a_T^{-2}T^{-1}\sum_{t=1}^{T} u_{T,t}^2\zeta_t$ converges to zero in probability. Because $u_{T,t}^2\zeta_t$ is conditionally independent, conditionally on the sample and for $\tilde{p}'$ such that $1 < \tilde{p}' < \min(2, p/2)$, as in Davidson (1994, pp. 172–173) we have $\mathrm{E}^*\left|T^{-1}\sum_{t=1}^{T} a_T^{-2}u_{T,t}^2\zeta_t\right|^{\tilde{p}} \leq 2T^{-\tilde{p}}\sum_{t=1}^{T}\left|a_T^{-1}u_{T,t}\right|^{2\tilde{p}}\mathrm{E}|\zeta_t|^{\tilde{p}} \leq 2BT^{-\tilde{p}}\sum_{t=1}^{T}\left|a_T^{-1}u_{T,t}\right|^{2\tilde{p}}$, $B := \mathrm{E}|\zeta_t|^{\tilde{p}} < \infty$. But $T^{-\tilde{p}}\sum_{t=1}^{T}\left|a_T^{-1}u_{T,t}\right|^{2\tilde{p}} \leq \left(\sup_{t\leq T}\left|a_T^{-1}\sigma_{T,t}\right|\right)^{2\tilde{p}}\left(T^{-1}\sum_{t=1}^{T}\left|\varepsilon_{T,t}\right|^{2\tilde{p}}\right) T^{1-\tilde{p}} \xrightarrow{p} 0$ because $\sup_{t\leq T}\left|a_T^{-1}\sigma_{T,t}\right| = O_p(1)$, $T^{-1}\sum_{t=1}^{T}\left|\varepsilon_{T,t}\right|^{2\tilde{p}} = O_p(1)$ (as $\mathrm{E}\left|\varepsilon_{T,t}\right|^{2\tilde{p}} < \sup_{t\leq T}\mathrm{E}\left|\varepsilon_{T,t}\right|^{p} < \infty$) and $\tilde{p} > 1$.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

From the foregoing results and applications of the CMT it follows immediately that $T(\hat{\alpha}_T^* - 1) \xrightarrow{w}_p Z_{\Omega,z,1}$ and, hence, that $G_T^{*,\hat{\alpha}}$ converges uniformly in probability to $G_{\Omega,z,1}$, the c.d.f. of $Z_{\Omega,z,1}$. The proof of the result that $p_T^{*,\hat{\alpha}} \xrightarrow{w} U[0,1]$ then follows using exactly the same arguments as are made in the proof of Theorem 5 of Hansen (2000). ■

Proof of Theorem 3. The proof is qualitatively similar to the proof of Theorem 1; we therefore sketch the additional results needed for the $T\left(\hat{\alpha}_T - 1\right)$ statistic only. First, notice that under Assumption $1'$ we are allowed to use the standard decomposition $u_{T,t} = \bar{c}_1 e_{T,t} + \tilde{e}_{T,t-1} - \tilde{e}_{T,t}$, where $\tilde{e}_{T,t} := \sum_{k=0}^{\infty} \tilde{c}_k e_{T,t-k}$, $\tilde{c}_k := \sum_{i=k+1}^{\infty} c_i$. To derive the FCLT in (23), notice that the partial sum $X_T(s) = a_T^{-1} T^{-1/2} \sum_{t=1}^{\lfloor Ts \rfloor} u_{T,t}$ now equals $X_T(s) = a_T^{-1} T^{-1/2} \bar{c}_1 \sum_{t=1}^{\lfloor Ts \rfloor} \sigma_{T,t-1} \varepsilon_{T,t} + a_T^{-1} T^{-1/2} \tilde{e}_{T,0} - a_T^{-1} T^{-1/2} \tilde{e}_{T,\lfloor Ts \rfloor}$ where, as in Theorem 1, the first term on the right-hand side converges weakly to $\bar{c}_1 m_\Omega(s)$. If $\sup_s \left| X_T(s) - a_T^{-1} T^{-1/2} \bar{c}_1 \sum_{t=1}^{\lfloor Ts \rfloor} \sigma_{T,t-1} \varepsilon_{T,t} \right| \leq a_T^{-1} T^{-1/2} \tilde{e}_{T,0} + \sup_{t=1,\ldots,T} \left| a_T^{-1} T^{-1/2} \tilde{e}_{T,t} \right|$ is of $o_p(1)$, then (23) follows by standard arguments. To prove this result, consider the inequality

$$\left| \frac{1}{a_T T^{1/2}} \tilde{e}_{T,t} \right| = \frac{1}{a_T T^{1/2}} \left| \sum_{k=0}^{\infty} \tilde{c}_k e_{T,t-k} \right| \leq \frac{1}{a_T T^{1/2}} \left| \sum_{k=0}^{t} \tilde{c}_k e_{T,t-k} \right| + \frac{1}{T^{1/2}} r_{T,t}$$

with $r_{T,t} := a_T^{-1} \left| \sum_{k=t+1}^{\infty} \tilde{c}_k e_{T,t-k} \right|$. Now, for all $t = 1, \ldots, T$,

$$a_T^{-1} T^{-1/2} \sup_{t=1,\ldots,T} \left| \sum_{k=0}^{t} \tilde{c}_k e_{T,t-k} \right|$$

$$\leq \left| \sum_{k=0}^{\infty} \tilde{c}_k \right| \sup_{t=1,\ldots,T} \left(a_T^{-1} \sigma_{T,t} \right) T^{-1/2} \sup_{t=1,\ldots,T} |\varepsilon_{T,t-k}|,$$

which converges to zero in probability. Regarding the second term, because for t, $e_{T,t}$ is a uniformly integrable MDA we may proceed, e.g., as in the proof of Theorem 3.15 in Phillips and Solo (1992) to obtain that $T^{-1/2} \sup_{t=1,\ldots,T} |r_{T,t}| \to 0$ in probability.

It is left to show that $a_T^{-2} \hat{\sigma}^2 \to \bar{c}_2 \int_0^1 \Omega(r)^2 dr$ (when the limiting volatility process is continuous this result follows from Lemma A.5 in Hansen, 1995, with $w_{T,t} = 1$). As in Hansen (1995, eqn. (23)), $a_T^{-2} \hat{\sigma}^2 = a_T^{-2} T^{-1} \sum_{t=1}^{T} f_{T,t} + a_T^{-2} T^{-1} \sum_{t=1}^{T} g_{T,t}$ where $g_{T,t} := \sum_{k=1}^{\infty} \sum_{j=0}^{\infty} c_j c_{j+k} e_{T,t-j} e_{T,t-k-j}$ and, setting $\xi_{T,t} := \varepsilon_{T,t}^2 - 1$,

$$f_{T,t} := \sum_{k=0}^{\infty} c_k^2 e_{T,t-k}^2 = \bar{c}_2 \sigma_{T,t-1}^2 + \sigma_{T,t-1}^2 \underbrace{\sum_{k=0}^{\infty} c_k^2 \xi_{T,t-k}}_{f_{T,t}^a}$$

$$+ \underbrace{\sum_{k=0}^{\infty} c_k^2 \left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2 \right)}_{f_{T,t}^b} + \underbrace{\sum_{k=0}^{\infty} c_k^2 \left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2 \right) \xi_{T,t-k}}_{f_{T,t}^c}.$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Because $a_T^{-2}T^{-1}\sum_{t=1}^{T}\sigma_{T,t-1}^2 \xrightarrow{w} \bar{\Omega}^2$ by Assumption 2 and the CMT, we need to show that $a_T^{-2}T^{-1}\sum_{t=1}^{T} g_{T,t}$, $T^{-1}\sum_{t=1}^{T} f_{T,t}^a$, and $a_T^{-2}T^{-1}\sum_{t=1}^{T} f_{T,t}^i$, $i = b, c$, each converge in probability to zero. To see that these convergence results hold, first notice that because $\sup_t \mathrm{E}\left|\mathrm{E}\left(f_{T,t}^a|\mathcal{F}_{T,t-m}^e\right)\right| \leq 2\sum_{k=m}^{\infty}|c_k| \to 0$ as $m \to \infty$, by Theorem A.1, $T^{-1}\sum_{t=1}^{T} f_{T,t}^a \xrightarrow{p} 0$. Second, by taking $m = T^{\upsilon}$ for $0 < \upsilon < 1$ we have

$$\begin{aligned}\frac{1}{a_T^2 T}\sum_{t=1}^{T} f_{T,t}^b &= \frac{1}{a_T^2 T}\sum_{t=1}^{T}\sum_{k=0}^{m} c_k^2\left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2\right)\\ &\quad + \frac{1}{a_T^2 T}\sum_{t=1}^{T}\sum_{k=m+1}^{\infty} c_k^2\left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2\right)\\ &\leq \left(\sum_{k=0}^{\infty} c_k^2\right)\sup_{k=0,\ldots,m}\frac{1}{a_T^2 T}\sum_{t=1}^{T}\left|\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2\right|\\ &\quad + 2\sup_{t\leq T} a_T^{-2}\sigma_{T,t}^2\sum_{k=m+1}^{\infty} c_k^2 \xrightarrow{p} 0\end{aligned}$$

by Theorem A.1, Assumption 2, and Assumption 1′(b). Similarly, we may write, for $m = T^{\upsilon}$, $0 < \upsilon < 1$,

$$\begin{aligned}\left|\frac{1}{a_T^2 T}\sum_{t=1}^{T} f_{T,t}^c\right| &\leq \left|\frac{1}{a_T^2 T}\sum_{t=1}^{T}\sum_{k=0}^{m} c_k^2\left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2\right)\xi_{T,t-k}\right|\\ &\quad + \left|\frac{1}{a_T^2 T}\sum_{t=1}^{T}\sum_{k=m+1}^{\infty} c_k^2\left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2\right)\xi_{T,t-k}\right|\\ &\leq \left|\frac{1}{a_T^2 T}\sum_{t=1}^{T}\sum_{k=0}^{m} c_k^2\left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2\right)\xi_{T,t-k}\right|\\ &\quad + 2\sup_{t\leq T}\frac{\sigma_{T,t}^2}{a_T^2}\sum_{k=m+1}^{\infty} c_k^2\frac{1}{T}\sum_{t=1}^{T}\left|\xi_{T,t-k}\right|\\ &= \left|\frac{1}{a_T^2 T}\sum_{t=1}^{T}\sum_{k=0}^{m} c_k^2\left(\sigma_{T,t-k-1}^2 - \sigma_{T,t-1}^2\right)\xi_{T,t-k}\right| + o_p(1)\end{aligned}$$

as

$$\begin{aligned}\mathrm{E}\left(\sum_{k=m+1}^{\infty} c_k^2 T^{-1}\sum_{t=1}^{T}\left|\xi_{T,t-k}\right|\right) &= \left(\sum_{k=m+1}^{\infty} c_k^2\right)T^{-1}\sum_{t=1}^{T}\mathrm{E}\left|\xi_{T,t-k}\right|\\ &\leq 2\sum_{k=m+1}^{\infty} c_k^2 \to 0.\end{aligned}$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

Moreover,

$$\left|\frac{1}{a_T^2 T}\sum_{t=1}^{T}\sum_{k=0}^{m} c_k^2\left(\sigma_{T,t-k-1}^2-\sigma_{T,t-1}^2\right)\xi_{T,t-k}\right|$$

$$\leq\left(\sum_{k=0}^{\infty}c_k^2\right)\sup_{k\leq m}\frac{1}{a_T^2 T}\sum_{t=1}^{T}\left|\left(\sigma_{T,t-k-1}^2-\sigma_{T,t-1}^2\right)\xi_{T,t-k}\right|,$$

where the last term can be made arbitrarily small, as in the proof of Lemma A.1. This establishes that $\left|a_T^{-2}T^{-1}\sum_{t=1}^{T} f_{T,t}^c\right|$ is of $o_p(1)$, as required. The proof that $a_T^2 T^{-1}\sum_{t=1}^{T} g_{T,t}=o_p(1)$ is entirely straightforward using the same argument as before in conjunction with the decomposition employed by Hansen (1995, p. 1127). ■

Proof of Theorem 4. To save space, we prove the results for the $\hat{Z}_{\hat{\alpha}}$ test only; the proof extends straightforwardly to the $\hat{Z}_t$ test. In what follows $\omega(hq_T^{-1})$ will be abbreviated as ω_h.

To prove the theorem, we need to demonstrate that both $\hat{Z}_{\hat{\alpha}}$ and the bootstrap statistic $T(\hat{\alpha}_T^*-1)$ weakly converge to the heteroskedastic unit root distribution, $Z_{\Omega,z,1}$. To show that $T(\hat{\alpha}_T^*-1)\xrightarrow{w} Z_{\Omega,z,1}$ we can simply adapt the arguments used for proving Theorem 2 (the proof is straightforward, given previous results, and is therefore omitted in the interest of brevity). To prove the convergence $\hat{Z}_{\hat{\alpha}}\xrightarrow{w} Z_{\Omega,z,1}$, because $a_T^{-2}\hat{\sigma}^2\xrightarrow{w}\bar{\Omega}^2$ (see the proof of Theorem 1), if we can prove that $a_T^{-2}\hat{\lambda}^2\xrightarrow{w}\bar{c}_1^2\bar{\Omega}^2$ then the convergence $\hat{Z}_{\hat{\alpha}}\xrightarrow{w} Z_{\Omega,z,1}$ follows from previously proved results using a simple application of the CMT, as shown in Phillips (1987b). Hence, it is only left to derive the limit of $\hat{\lambda}^2$, which can be done by adapting the proof of Theorem 2 in Jansson (2002), which deals with the constant volatility case, to the present nonstationary volatility framework. We prove the result by assuming that the errors are observed, i.e., $\hat{u}_{T,t}=u_{T,t}$; the proof for the AR(1) residuals then follows by Theorem 3 in Hansen (1992b).

First, let $\tilde{\gamma}_h:=T^{-1}\sum_{t=1}^{T-|h|}\sum_{k=0}^{\infty}c_k c_{k+|h|}\sigma_{T,t-k-1}^2$. Because

$$\hat{\gamma}_h=\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}c_k c_j\frac{1}{T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}\sigma_{T,t+|h|-j-1}\varepsilon_{T,t-k}\varepsilon_{T,t+|h|-j}$$

$$=\sum_{k=0}^{\infty}c_k c_{k+|h|}\frac{1}{T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}^2\varepsilon_{T,t-k}^2$$

$$+\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}c_k c_j\mathbb{I}(j\neq k+|h|)\frac{1}{T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}\sigma_{T,t+|h|-j-1}\varepsilon_{T,t-k}\varepsilon_{T,t+|h|-j}$$

we may write

$$a_T^{-2}\left(\hat{\gamma}_h-\tilde{\gamma}_h\right)=\sum_{k=0}^{\infty}c_k c_{k+h}\frac{1}{a_T^2 T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}^2\xi_{T,t-k}$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

$$+\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}c_kc_j\mathbb{I}(j\neq k+|h|)$$

$$\times\frac{1}{a_T^2T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}\sigma_{T,t+|h|-j-1}\varepsilon_{T,t-k}\varepsilon_{T,t+|h|-j} \tag{A.11}$$

with, as before, $\xi_{T,t}:=\varepsilon_{T,t}^2-1$. Regarding the first term on the right-hand side, as shown in Jansson (2002, p. 1454) we have that, for any $\varepsilon>0$,

$$\mathrm{E}\left|\sum_{k=0}^{\infty}c_kc_{k+|h|}\frac{1}{a_T^2T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}^2\xi_{T,t-k}\right|$$

$$\leq\sum_{k=0}^{\infty}\left|c_kc_{k+|h|}\right|\mathrm{E}\left|\frac{1}{a_T^2T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}^2\xi_{T,t-k}\right|\leq\left(T^{-1/2}C_\varepsilon+\varepsilon\right)\sum_{k=0}^{\infty}\left|c_kc_{k+|h|}\right|$$

for a finite constant C_ε, independent of h and T (the last inequality holds by proceeding as in the proof of Theorem 2.22 in Hall and Heyde, 1980, in the special case $p=1$, using the fact that under the conditions of the theorem, $a_T^{-2}\sigma_{T,t-1}^2\xi_{T,t}$ is a uniformly integrable MDA). Regarding the second term on the right-hand side of (A.11), we can again proceed as in Jansson (2002, p. 1455) and notice that, for all h,

$$\mathrm{E}\left|\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}c_kc_j\mathbb{I}(j\neq k+|h|)\frac{1}{a_T^2T}\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}\sigma_{T,t+|h|-j-1}\varepsilon_{T,t-k}\varepsilon_{T,t+|h|-j}\right|$$

$$\leq\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\left|c_kc_j\right|\left(\mathrm{E}\left(\mathbb{I}(j\neq k-|h|)\frac{1}{a_T^2T}\right.\right.$$

$$\left.\left.\sum_{t=1}^{T-|h|}\sigma_{T,t-k-1}\sigma_{T,t+|h|-j-1}\varepsilon_{T,t-k}\varepsilon_{T,t+|h|-j}\right)^2\right)^{1/2}$$

$$\leq\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\left|c_kc_j\right|\left(\frac{1}{a_T^4T^2}\sum_{t=1}^{T-|h|}\mathrm{E}\left(\mathbb{I}(j\neq k-|h|)\sigma_{T,t-k-1}^2\right.\right.$$

$$\left.\left.\times\sigma_{T,t+|h|-j-1}^2\varepsilon_{T,t-k}^2\varepsilon_{T,t+|h|-j}^2\right)\right)^{1/2}$$

$$\leq\sum_{k=0}^{\infty}\sum_{j=0}^{\infty}\left|c_kc_j\right|\left(\frac{T-|h|}{T^2}Q_\sigma^{4/q}\right)^{1/2}\leq\frac{1}{T^{1/2}}\left(\sum_{k=0}^{\infty}|c_k|\right)^2Q_\sigma^{2/q}$$

because, for $j>k+|h|$, $\mathrm{E}(a_T^{-4}\sigma_{T,t-k-1}^2\sigma_{T,t+|h|-j-1}^2\varepsilon_{T,t-k}^2\varepsilon_{T,t+|h|-j}^2)=\mathrm{E}(\mathrm{E}(a_T^{-2}\sigma_{T,t-k-1}^2\varepsilon_{T,t-k}^2|\mathcal{I}_{T,t+|h|-j})\,a_T^{-2}\sigma_{T,t+h-j-1}^2\varepsilon_{T,t+h-j}^2)\leq\sup_t\mathrm{E}\left(a_T^{-2}\sigma_{T,t}^2\right)^2\leq Q_\sigma^{4/q}$.

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

By setting $\tilde{\lambda}^2 := \sum_h \omega_h \tilde{\gamma}_h$ we hence have that

$$a_T^{-2}(\hat{\lambda}^2 - \tilde{\lambda}^2) \le \sum_{h=-T+1}^{T-1} |\omega_h| a_T^{-2} \left|\hat{\gamma}_h - \tilde{\gamma}_h\right|$$

$$\le \sum_{h=-T+1}^{T-1} |\omega_h| \left(\left(T^{-1/2} C_\varepsilon + \varepsilon\right) \sum_{k=0}^{\infty} \left|c_k c_{k+|h|}\right| + \frac{1}{T^{1/2}} \left(\sum_{k=0}^{\infty} |c_k| \right)^2 Q_\sigma^{2/q} \right)$$

$$\le \sup_{x \ge 0} |\omega(x)| \left(T^{-1/2} C_\varepsilon + \varepsilon\right) \sum_{h=-\infty}^{\infty} \sum_{k=0}^{\infty} \left|c_k c_{k-|h|}\right| + \sum_{h=-\infty}^{\infty} |\omega_h| \frac{1}{T^{1/2}} \left(\sum_{k=0}^{\infty} |c_k| \right)^2 Q_\sigma^{2/q},$$

which can be made arbitrarily small because $\sup_{x \ge 0} |\omega(x)| < \infty$, $\sum_{k=0}^{\infty} |c_k| < \infty$, and $q_T = o\left(T^{1/2}\right)$. The proof is hence completed if we show that $a_T^{-2} \tilde{\lambda}^2 := \sum_{h=-T+1}^{T-1} \omega_h \tilde{\gamma}_h \xrightarrow{w} \bar{\Omega}^2 (\sum_{k=0}^{\infty} c_k)^2$. Consider the equality

$$a_T^{-2} \sum_{h=-T+1}^{T-1} \omega_h \tilde{\gamma}_h = \sum_{h=-T+1}^{T-1} \omega_h \frac{1}{a_T^2 T} \sum_{t=1}^{T-|h|} \sum_{k=0}^{\infty} c_k c_{k+|h|} \sigma^2_{T,t-k-1}$$

$$= - \sum_{h=-T+1}^{T-1} \omega_h \frac{1}{a_T^2 T} \sum_{t=1}^{T-|h|} \sum_{k=0}^{\infty} c_k c_{k+|h|} \left(\sigma^2_{T,t-1} - \sigma^2_{T,t-k-1} \right) + \sum_{h=-T+1}^{T-1} \omega_h \frac{1}{a_T^2 T} \sum_{t=1}^{T-|h|} \sum_{k=0}^{\infty} c_k c_{k+|h|} \sigma^2_{T,t-1}. \tag{A.12}$$

Under the stated assumptions, the last term in (A.12) satisfies

$$\sum_{h=-T-1}^{T-1} \omega_h \frac{1}{a_T^2 T} \sum_{t=1}^{T-|h|} \sum_{k=0}^{\infty} c_k c_{k+|h|} \sigma^2_{T,t-1} = \frac{1}{a_T^2 T} \sum_{t=1}^{T} \sigma^2_{T,t-1} \sum_{h=-T+1}^{T-1} \omega_h \sum_{k=0}^{\infty} c_k c_{k+|h|} - \sum_{h=-T+1}^{T-1} \omega_h \sum_{k=0}^{\infty} c_k c_{k+|h|} \frac{1}{a_T^2 T} \sum_{t=T-|h|+1}^{T} \sigma^2_{T,t-1}$$

$$\xrightarrow{w} \bar{\Omega}^2 \left(\sum_{k=0}^{\infty} c_k \right)^2 = \bar{\Omega}^2 \bar{c}_1^2$$

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press

because $a_T^{-2}T^{-1}\sum_t \sigma_{T,t-1}^2 \xrightarrow{w} \bar{\Omega}^2$ (Corollary A.1), $\sum_h \omega_h \sum_k c_k c_{k+|h|} \to \bar{c}_1^2$ (Jansson, 2002, Lem. 6), and

$$\left|\sum_{h=-T+1}^{T-1}\omega_h \sum_{k=0}^{\infty} c_k c_{k+|h|}\frac{1}{a_T^2 T}\sum_{t=T-|h|+1}^{T}\sigma_{T,t-1}^2\right|$$

$$\leq \sup_{t=1,\ldots,T}\left(\frac{\sigma_{T,t}^2}{a_T^2}\right)\sup_{x\geq 0}|\omega(x)|\frac{1}{T}\sum_{h=-T+1}^{T-1}|h|\sum_{k=0}^{\infty}|c_k c_{k+|h|}|,$$

the right member of which converges to zero under the summability condition in Assumption 1′(c). Regarding the first term on the right-hand side of (A.12), for $m = BT^{\gamma}$, $0 < \gamma < 1$, we may write

$$\left|\sum_{h=-T+1}^{T-1}\omega_h \sum_{k=0}^{\infty} c_k c_{k-|h|}\frac{1}{a_T^2 T}\sum_{t=1}^{T-|h|}\left(\sigma_{T,t-1}^2-\sigma_{T,t-k-1}^2\right)\right|$$

$$\leq \left|\sum_{h=-T+1}^{T-1}\omega_h \sum_{k=0}^{m} c_k c_{k+|h|}\frac{1}{a_T^2 T}\sum_{t=1}^{T-|h|}\left(\sigma_{T,t-1}^2-\sigma_{T,t-k-1}^2\right)\right|$$

$$+\left|\sum_{h=-T+1}^{T-1}\omega_h \sum_{k=m+1}^{\infty} c_k c_{k+|h|}\frac{1}{a_T^2 T}\sum_{t=1}^{T-|h|}\left(\sigma_{T,t-1}^2-\sigma_{T,t-k-1}^2\right)\right|$$

$$\leq \sum_{h=-T+1}^{T-1}|\omega_h|\sum_{k=0}^{m}\left|c_k c_{k+|h|}\right|\sup_{h\leq m}\frac{1}{a_T^2 T}\sum_{t=1}^{T-|h|}\left|\sigma_{T,t-1}^2-\sigma_{T,t-k-1}^2\right|$$

$$+2\sup_{t\leq T}\left(a_T^{-2}\sigma_{T,t}^2\right)\sup_{x\geq 0}|\omega(x)|\sum_{k=m+1}^{\infty}|c_k|\sum_{h=-T+1}^{T-1}\left|c_{k+|h|}\right|$$

$$\leq \sup_{x\geq 0}|\omega(x)|\left(\sum_{k=0}^{\infty}|c_k|\right)^2\sup_{h\leq m}\frac{1}{a_T^2 T}\sum_{t=1}^{T-|h|}\left|\sigma_{T,t-1}^2-\sigma_{T,t-k-1}^2\right|+o_p(1)=o_p(1)$$

because by Lemma A.1, $\sup_{h\leq m} a_T^{-2}T^{-1}\sum_{t=1}^{T-|h|}\left|\sigma_{T,t-1}^2-\sigma_{T,t-k-1}^2\right| \xrightarrow{p} 0$. ■

https://doi.org/10.1017/S026646660809049X Published online by Cambridge University Press