Econometric Theory, **24**, 2008, 43–71. Printed in the United States of America.
DOI: 10.1017/S0266466608080043

BOOTSTRAP UNIT ROOT TESTS FOR TIME SERIES WITH NONSTATIONARY VOLATILITY

GIUSEPPE CAVALIERE
University of Bologna

A.M. ROBERT TAYLOR
University of Nottingham

The presence of permanent volatility shifts in key macroeconomic and financial variables in developed economies appears to be relatively common. Conventional unit root tests are unreliable in the presence of such behavior, having nonpivotal asymptotic null distributions. In this paper we propose a bootstrap approach to unit root testing that is valid in the presence of a wide class of permanent variance changes that includes single and multiple (abrupt and smooth transition) volatility change processes as special cases. We make use of the so-called wild bootstrap principle, which preserves the heteroskedasticity present in the original shocks. Our proposed method does not require the practitioner to specify any parametric model for the volatility process. Numerical evidence suggests that the bootstrap tests perform well in finite samples against a range of nonstationary volatility processes.

1. INTRODUCTION

Many of the main macroeconomic and financial variables across developed countries display significant breaks in unconditional volatility; see, inter alia, Busetti and Taylor (2003), Kim and Nelson (1999), Koop and Potter (2000), McConnell and Perez Quiros (2000), van Dijk, Osborn, and Sensier (2002), Sensier and van Dijk (2004), and references therein. Sensier and van Dijk (2004), e.g., find that over 80% of the real and price variables in the Stock and Watson (1999) data set reject the null hypothesis of constant unconditional innovation variance. Considerable evidence against the constancy of unconditional variances in stock market returns and exchange-rate data has also been reported; cf. Loretan and Phillips (1994). Moreover, Hansen (1995) notes that empirical applications of autoregressive stochastic volatility (SV) models to financial data

We thank two anonymous referees, Paulo Rodrigues, Peter Phillips, and seminar participants at the URCT conference held in Faro, Portugal, September 29 to October 1, 2005, for helpful comments on previous versions of this paper. Address correspondence to Robert Taylor, School of Economics, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom; e-mail: Robert.Taylor@nottingham.ac.uk.

© 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

generally estimate the dominant root in the SV process to be close to the nonstationarity boundary at unity.

Although neither volatility breaks that follow a stationary Markov switching process nor stationary time-varying conditional variances (e.g., autoregressive conditional heteroskedasticity [ARCH]) impact on the large-sample properties of unit root and cointegration tests (for the former, see Nelson, Piger, and Zivot, 2001; Cavaliere, 2003; for the latter, see Hansen and Rahbek, 1998; Kim and Schmidt, 1993; Ling, Li, and McAleer, 2003), permanent changes in volatility can, however, greatly affect unit root inference; see Hamori and Tokihisa (1997), Kim, Leybourne, and Newbold (2002), Cavaliere (2004), and Cavaliere and Taylor (2004). A solution to this problem for the specific case of a single abrupt change in variance is given by Kim et al. (2002), who propose a two-stage procedure where the breakpoint together with the pre- and postbreak variances are first estimated. These estimates are then employed in modified variants of the Perron (1989, 1990) unit root tests. The requirement of a single abrupt change in volatility is, however, not always consistent with empirical evidence, which suggests that data are often better characterized by smooth volatility changes rather than by abrupt changes (see, inter alia, van Dijk et al., 2002) and that multiple changes in volatility are more commonly observed (see, e.g., Watson, 1999). Cavaliere and Taylor (2004) report evidence of multiple volatility breaks and trending volatility in the monthly producer price inflation series from the well-known Stock and Watson (1999) database.

Cavaliere and Taylor (2004) propose a class of modified unit root tests, based on time-transformed data, that are asymptotically robust to a wide class of nonstationary volatility processes that includes multiple and smooth transition breaks. However, because the time transformation involved entails the use of only a subset of the original sample data, simulation results in Cavaliere and Taylor (2004) show that this can lead to significant power losses relative to the corresponding (size-adjusted) standard unit root tests. In this paper we take a different approach and develop bootstrap unit root tests that are, by design, robust to the class of nonstationary volatility considered in Cavaliere and Taylor (2004). We develop bootstrap versions of the autocorrelation-robust M unit root tests of Perron and Ng (1996), Stock (1999), and Ng and Perron (2001), although the approach we outline can equally be applied to any of the commonly used unit root statistics. We focus on the local generalized least squares (GLS) detrended variants of these tests, as favored by Ng and Perron (2001).

Our proposed unit root tests are based on the wild bootstrap method, which replicates in the resampled data the pattern of heteroskedasticity present in the original shocks. An application of the wild bootstrap principle to the problem of estimating *stationary* autoregressions driven by (stationary) conditionally heteroskedastic errors has recently been shown to perform well by Gonçalves and Killian (2004). As we shall demonstrate, the wild bootstrap unit root statistics share the same limiting null distribution as the original statistics under the class of nonstationary volatility considered. Consequently, inference that is (asymp-

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

totically) robust to nonstationary volatility in the shocks can be performed without the need to specify any parametric model of volatility within the class considered. Importantly, we show that under local alternatives the bootstrap M tests achieve the asymptotic power functions of (infeasible) size-corrected versions of the standard M tests. We also show that the bootstrap tests are consistent under fixed alternatives.

The paper is organized as follows. In Section 2 we introduce our reference data generating process (DGP) and detail the class of heteroskedastic volatility processes under which we will work. In Section 3 the effects of such nonstationary volatility on the large-sample distributions of conventional unit root tests are discussed. Section 4 develops our heteroskedasticity-robust bootstrap unit root tests and their asymptotic properties. The finite-sample properties of the tests are explored through Monte Carlo simulation in Section 5. Section 6 concludes. All proofs are placed in the Appendix. In the following discussion $\xrightarrow{w}$ denotes weak convergence, $\xrightarrow{p}$ convergence in probability, and $\xrightarrow{w}_p$ weak convergence in probability (Giné and Zinn, 1990; Hansen, 1996) in each case as the sample size diverges to positive infinity; $\mathbb{I}(\cdot)$ is 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 the argument, $\mathcal{C} := C[0,1]$ denotes the space of continuous processes on $[0,1]$, and $\mathcal{D} := D[0,1]$ denotes the space of right continuous with left limit (càdlàg) processes on $[0,1]$.

2. THE HETEROSKEDASTIC MODEL

Consider the case where we have $T + 1$ observations generated according to the model

$$X_t = d_t + Y_t, \qquad t = 0, \ldots, T, \tag{1}$$

$$Y_t = \alpha Y_{t-1} + u_t, \tag{2}$$

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \tag{3}$$

$$\varepsilon_t = \sigma_t e_t \tag{4}$$

with $E(Y_0^2) < \infty$. In (1), $d_t := \gamma' z_t$ with z_t a vector of deterministic components. As in Ng and Perron (2001) we focus on the pth-order trend function, $z_t := (1, t, \ldots, t^p)'$, with special focus on the leading cases of a constant ($p = 0$) and linear trend ($p = 1$), $z_t = 1$ and $z_t = (1, t)'$, respectively, although the analysis remains valid for more general cases. The following set of assumptions will be taken to hold on (1)–(4).

Assumption $\mathcal{A}$. ($\mathcal{A}_1$) The lag polynomial satisfies $C(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{j=0}^{\infty} j|c_j| < \infty$. ($\mathcal{A}_2$) $e_t \sim iid(0,1)$ with $E|e_t|^r < K < \infty$ for some $r \geq 4$. ($\mathcal{A}_3$) The volatility term σ_t satisfies $\sigma_{[sT]} = \omega(s)$ for all $s \in [0,1]$, where

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

$\omega(\cdot) \in \mathcal{D}$ is nonstochastic and strictly positive. For $t < 0$, $\sigma_t \leq \sigma^* < \infty$. ($\mathcal{A}_4$) The autoregressive coefficient α satisfies $|\alpha| < 1$ or $\alpha = 1$.

Notice that the error term $\{u_t\}$ in (3) is a linear process in $\{\varepsilon_t\}$, the latter formed as the product of two components, $\{e_t\}$ and $\{\sigma_t\}$. Because $\{e_t\}$ is independent and identically distributed (i.i.d.), conditionally on σ_t, the error term ε_t has mean zero and time-varying variance σ_t^2.

Assumptions $\mathcal{A}_1$ (which states that $\{u_t\}$ has a stable and invertible representation in terms of the ε_t's and that moments of order four exist; see, e.g., Chang and Park, 2002) and $\mathcal{A}_4$ (which rules out explosive and seasonal unit roots) are standard in the econometric literature. Assumption $\mathcal{A}_2$ is discussed in Remark 1, which follows.

The key assumption for the purposes of this paper is $\mathcal{A}_3$, which casts the dynamics of the innovation variance in a quite general framework. The class of volatility processes that satisfy Assumption $\mathcal{A}_3$ is extremely wide. Indeed, the innovation variance is required only to be nonstochastic, to be bounded, and to display a countable number of jumps. A detailed discussion of the class of variance processes allowed under $\mathcal{A}_3$ is given by Cavaliere and Taylor (2004), who show that this includes the single abrupt change model of Hamori and Tokihisa (1997) and Kim et al. (2002), variance processes displaying multiple volatility shifts, polynomially (possibly piecewise) trending volatility, and smooth transition variance breaks, among other things.[1] The conventional assumption made in the unit root testing literature, that $\sigma_t = \sigma$, for all t, also satisfies Assumption $\mathcal{A}_3$, because here $\omega(s) = \sigma$, for all s.

A quantity that will play a key role in our analysis is given by the following function in $\mathcal{C}$:

$$\eta(s) := \left(\int_0^1 \omega(r)^2 \, dr \right)^{-1} \int_0^s \omega(r)^2 \, dr. \tag{5}$$

Hereafter, $\eta(\cdot)$ will be referred to as the *variance profile* of the process, because it depends solely on the time-series behavior of the volatility. Notice that the variance profile satisfies $\eta(s) = s$ under homoskedasticity, whereas it deviates from s in the presence of heteroskedasticity. Observe also that the quantity $\bar{\omega}^2 := \int_0^1 \omega(r)^2 \, dr$ that appears in (5) by Assumption $\mathcal{A}_3$ equals the limit of $T^{-1} \sum_{t=1}^{T} \sigma_t^2$ and may therefore be interpreted as the (asymptotic) average innovation variance.

Remark 1. Assumption $\mathcal{A}_2$ specifies that $\{e_t\}$ is i.i.d. and so consequently rules out processes of major interest such as stationary generalized autoregressive conditionally heteroskedastic (GARCH) or Markov-switching volatility processes. However, it can be shown that the results stated in this paper continue to hold under the weaker condition that for $\mathcal{F}_t := \sigma\{\varepsilon_s, s \leq t\}$, $\{e_t, \mathcal{F}_t\}$ is a martingale difference sequence (MDS) satisfying the following (standard)

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

conditions (see, e.g., Chang and Park, 2002, Assump. 1, p. 433): (a) $E(e_t^2) = \sigma_e^2 = 1$, all t; (b) $T^{-1}\sum_{t=1}^{T} e_t^2 \xrightarrow{p} \sigma_e^2 < \infty$; and (c) $E(|e_t|^r) < K < \infty$ for some $r \geq 4$, all t. Under this more general framework, e_t can, e.g., follow a covariance-stationary GARCH process (with finite fourth-order moments) or a stationary Markov-switching volatility process. Consequently, and denoting by h_t the conditional variance of e_t in the MDS case, we can write (4) as

$$\varepsilon_t = \sigma_t e_t, \qquad E(e_t^2 | \mathcal{F}_{t-1}) = h_t,$$

where h_t generates (transitory) changes in the conditional volatility and σ_t generates (permanent) changes in the (un)conditional volatility of ε_t. This setup therefore includes conditionally heteroskedastic processes with breaks (or trends) in the unconditional volatility, such as in the spline-GARCH model of Engle and Gonzalo Rangel (2004). In what follows, the i.i.d. assumption is maintained, without loss of generality, purely to simplify exposition.

Remark 2. The assumption that the volatility function $\omega(\cdot)$ is nonstochastic allows for considerable simplification of the theoretical setup. However, we conjecture that, under suitable memory and moment conditions, this assumption can be weakened to allow for cases where the innovations $\{e_t\}$ and $\omega(\cdot)$ are stochastically independent. Indeed, in such cases if the (exogenous) volatility process $\omega(\cdot)$ has sample paths satisfying Assumption $\mathcal{A}_3$, then the results presented in the paper can be thought of as holding *conditional* on a given realization of $\omega(\cdot)$. The conditioning argument used here in the context of the volatility function serves the same purpose as the exogeneity assumption used by Perron (1989, pp. 1387–1388) to permit stochastic changes in the trend function. In the (exogenous) SV framework, Markov-switching variances obtain by assuming that $\omega(\cdot)$ is a strictly positive, continuous-time Markov chain with a finite number of states, whereas nonstationary autoregressive SV models obtain for $\omega(s) = h(J(s))$, with $J(\cdot)$ a diffusion process in $\mathcal{D}$ and $h(\cdot)$ a strictly positive continuous function; see Hansen (1995).

Remark 3. Because the variance σ_t^2 depends on T, a time series generated according to (1)–(4) with σ_t satisfying Assumption $\mathcal{A}_3$ formally constitutes a triangular array of the type $\{X_{T,t} := d_{T,t} + Y_{T,t} : 0 \leq t \leq T, T \geq 1\}$, where $Y_{T,t}$ is recursively defined as $Y_{T,t} := \alpha Y_{T,t-1} + C(L)\sigma_{T,t} e_t$, $\sigma_{T,[sT]} := \omega(s)$. However, because the triangular array notation is not essential, for simplicity the subscript T is suppressed in what follows.

3. THE *M* CLASS OF UNIT ROOT TESTS

In this section we outline the M unit root tests of Perron and Ng (1996), Stock (1999), and Ng and Perron (2001) and discuss the impact of time-varying innovation variances on the asymptotic distributions of these statistics.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

As in Elliott, Rothenberg, and Stock (1996) denote by $\hat{X}_t$ the local GLS residuals from the regression of X_t on z_t; i.e., $\hat{X}_t := X_t - \hat{\gamma}' z_t$, $t = 0, \ldots, T$, where $\hat{\gamma}$ is obtained from the ordinary least squares (OLS) regression of

$$\mathbf{X}_{\bar{c}} \equiv [X_0, X_1 - (1 - \bar{c}/T)X_0, \ldots, X_T - (1 - \bar{c}/T)X_{T-1}]' \tag{6}$$

on

$$\mathbf{z}_{\bar{c}} \equiv [z_0, z_1 - (1 - \bar{c}/T)z_0, \ldots, z_T - (1 - \bar{c}/T)z_{T-1}]', \tag{7}$$

and the M unit root statistics are given by

$$\mathcal{MZ}_\alpha := \frac{T^{-1}\hat{X}_T^2 - s_{AR}^2(k)}{2T^{-2}\sum_{t=1}^{T}\hat{X}_{t-1}^2}, \qquad \mathcal{MSB} := \left(T^{-2}\sum_{t=1}^{T}\hat{X}_{t-1}^2/s_{AR}^2(k)\right)^{1/2},$$

$$\mathcal{MZ}_t := \mathcal{MZ}_\alpha \times \mathcal{MSB}, \tag{8}$$

where $s_{AR}^2(k)$ is an autoregressive estimator of the (nonnormalized) spectral density at frequency zero of $\{u_t\}$. Specifically,

$$s_{AR}^2(k) := \hat{\sigma}^2/(1 - \hat{\beta}(1))^2, \qquad \hat{\beta}(1) := \sum_{i=1}^{k}\hat{\beta}_{i,k}, \tag{9}$$

where $\hat{\beta}_{i,k}$, $i = 1, \ldots, k$, and $\hat{\sigma}^2$ are, respectively, the OLS slope and variance estimators from the augmented Dickey–Fuller (ADF) regression equation

$$\Delta\hat{X}_t = \pi\hat{X}_{t-1} + \sum_{i=1}^{k}\beta_{i,k}\Delta\hat{X}_{t-i} + \hat{u}_{t,k}. \tag{10}$$

As is standard, in all of the preceding statistics we require that the lag truncation parameter, k, satisfies the following assumption (Berk, 1974; Said and Dickey, 1984).

Assumption $\mathcal{K}$. As $T \to \infty$, $1/k + k^3/T \to 0$.

In the case of $\mathcal{MZ}_\alpha$, $\mathcal{MZ}_t$, the unit root null hypothesis ($\alpha = 1$) is rejected for large negative values of the statistics, whereas a test based on $\mathcal{MSB}$ rejects for small values of the statistic. All three tests are therefore left-tailed. The value of $\bar{c}$ in (6) and (7) is specified according to the form of the deterministic vector z_t and the desired significance level; see Elliott et al. (1996) for details. For $p = 0$, they suggest $\bar{c} = 7$, whereas for $p = 1$, $\bar{c} = 13.5$, in each case for tests run at the 0.05 significance level.

If Assumptions $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{K}$ hold on (1)–(4) and $\sigma_t = \sigma$, $t = 1, \ldots, T$ (so that the errors are homoskedastic), then under the near-integrated autoregres-

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

sive model, $H_c: \alpha := 1 - c/T$, with $0 \leq c < \infty$, the asymptotic distributions of the preceding unit root statistics are well known:

$$\mathcal{MZ}_\alpha \xrightarrow{w} \frac{B_c^p(1,\bar{c})^2 - 1}{2\int_0^1 B_c^z(s,\bar{c})^2\,ds} =: \xi_1^{c,\bar{c},p},$$

$$\mathcal{MSB} \xrightarrow{w} \left(\int_0^1 B_c^p(s,\bar{c})^2\,ds\right)^{1/2} =: \xi_2^{c,\bar{c},p}, \tag{11}$$

$$\mathcal{MZ}_t \xrightarrow{w} \frac{B_c^p(1,\bar{c})^2 - 1}{2\left(\int_0^1 B_c^p(s,\bar{c})^2\,ds\right)^{1/2}} =: \xi_3^{c,\bar{c},p},$$

where in the constant case ($p = 0$) $B_c^0(s,\bar{c}) = B_c(s)$, whereas in the linear trend case ($p = 1$) $B_c^1(s,\bar{c}) = B_c(s) - s(\lambda_{\bar{c}} B_c(1) + 3(1 - \lambda_{\bar{c}})\int_0^1 rB_c(r)\,dr)$, $\lambda_{\bar{c}} := (1 + \bar{c})/(1 + \bar{c} + \bar{c}^2/3)$, with $B_c(s) := \int_0^s \exp(-c(s - r))\,dB(r)$, a standard Ornstein–Uhlenbeck process and $B(s)$ a standard Brownian motion; cf. Elliott et al. (1996), Ng and Perron (2001), and Chang and Park (2002).

Representations for the limiting null distributions of the statistics therefore obtain on setting $c = 0$ in (11). Selected critical values for the tests for $p = 0$ and $p = 1$ are provided in Table I of Elliott et al. (1996, p. 825) and Table 1 of Ng and Perron (2001, p. 1524).

For $c > 0$ the representations can be used to delineate the asymptotic local power functions of the tests, in each case indexed by the noncentrality parameter c. Specifically, for tests run at significance level ε these are given by

$$\pi_j(c,\bar{c},p) := P(\xi_j^{c,\bar{c},p} < b_j(\bar{c},p)), \qquad j = 1,2,3, \tag{12}$$

where $b_j(\bar{c},p)$ satisfies $P(\xi_j^{0,\bar{c},p} < b_j(\bar{c},p)) = \varepsilon$, $j = 1,2,3$; cf. Elliott et al. (1996, p. 818). Ng and Perron (2001) demonstrate numerically that the $\mathcal{MZ}_\alpha$, $\mathcal{MZ}_t$, and $\mathcal{MSB}$ tests have asymptotic local power functions each of which is virtually indistinguishable from the local power envelope for both $p = 0$ and $p = 1$.

In cases where the volatility process σ_t is permitted to be generated by any process satisfying Assumption $\mathcal{A}_3$, we may state the following theorem.

THEOREM 1. *Let $\{X_t\}_0^T$ be generated as in (1)–(4) under Assumptions $\mathcal{A}$ and $\mathcal{K}$ and let $H_c: \alpha := 1 - c/T$, $0 \leq c < \infty$, hold. Then, $\mathcal{MZ}_\alpha \xrightarrow{w} 0.5(B_{\eta,c}^p(1,\bar{c})^2 - 1)(\int_0^1 B_{\eta,c}^p(s,\bar{c})^2\,ds)^{-1} =: \tilde{\xi}_1^{c,\bar{c},p,\eta}$, $\mathcal{MSB} \xrightarrow{w} (\int_0^1 B_{\eta,c}^p(s,\bar{c})^2\,ds)^{1/2} =: \tilde{\xi}_2^{c,\bar{c},p,\eta}$, and $\mathcal{MZ}_t \xrightarrow{w} 0.5(B_{\eta,c}^p(1,\bar{c})^2 - 1)(\int_0^1 B_{\eta,c}^p(s,\bar{c})^2\,ds)^{-1/2} = \tilde{\xi}_1^{c,\bar{c},p,\eta} \times \tilde{\xi}_2^{c,\bar{c},p,\eta} =: \tilde{\xi}_3^{c,\bar{c},p,\eta}$, where in the constant case ($p = 0$), $B_{\eta,c}^0(s,\bar{c}) := B_{\eta,c}(s)$, whereas in the linear trend case ($p = 1$) $B_{\eta,c}^1(s,\bar{c}) := B_{\eta,c}(s) -$*

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

$s(\lambda_{\bar{c}} B_{\eta,c}(1) + 3(1 - \lambda_{\bar{c}}) \int_0^1 r B_{\eta,c}(r)\, dr)$, $\lambda_{\bar{c}} := (1 + \bar{c})/(1 + \bar{c} + \bar{c}^2/3)$, *with* $B_{\eta,c}(s) := \int_0^s \exp(-c(s - r))\, dB(\eta(r))$.

As can be seen from the representations given in Theorem 1, the asymptotic null distributions (those pertaining to $c = 0$) of the unit root statistics have the usual form with the exception that the standard limiting Brownian motion, $B_0(s)$, is replaced by the Brownian functional $B_{\eta,0}(s)$, the latter reducing to the former only where the process is homoskedastic, i.e., where $\omega(\cdot)$ is a constant function. The process $B_{\eta,0}(s)$ is known as a *variance-transformed* Brownian motion, i.e., a Brownian motion under a modification of the time domain; see, e.g., Davidson (1994). This difference has serious implications for the size of the associated unit root tests, with the standard tabulated critical values (cited before) no longer appropriate. The effects of a number of particular variance profiles on the M tests are quantified by Monte Carlo in Section 5.1; see also Hamori and Tokihisa (1997), Kim et al. (2002), and Cavaliere (2004).

It is also clear from the representations in Theorem 1 for $c > 0$ that the asymptotic local power functions of the tests will also be affected by nonstationary volatility (even if critical values from the correct null distributions were used) because, as with the null case, it is only where $\omega(\cdot)$ is constant that the representations given in Theorem 1 reduce to the corresponding representations given in (11). Specifically, the asymptotic local power functions of the tests at significance level ε are given by

$$\pi_j(c, \bar{c}, p, \eta) := P(\xi_j^{c,\bar{c},p,\eta} < b_j(\bar{c}, p, \eta)), \qquad j = 1, 2, 3, \tag{13}$$

where $b_j(\bar{c}, p, \eta)$ satisfies $P(\xi_j^{0,\bar{c},p,\eta} < b_j(\bar{c}, p, \eta)) = \varepsilon$, $j = 1, 2, 3$. Moreover, unlike in the homoskedastic case, these cannot necessarily be expected to lie arbitrarily close to the local asymptotic power envelope.

4. BOOTSTRAP UNIT ROOT TESTS

As demonstrated in the previous section, nonstationary volatility introduces a time deformation aspect to the limiting distributions of the unit root statistics under $H_c: \alpha = 1 - c/T$, which alters their form vis-à-vis the homoskedastic case. In this section we suggest a wild bootstrap algorithm that allows us to retrieve asymptotically correct p-values in the presence of both heteroskedasticity of the form described in Section 2 and serial correlation in the shocks. The wild bootstrap can be successfully implemented in our framework because whereas the asymptotic distributions of the M unit root statistics are affected by the heteroskedasticity-dependent variance profile $\eta(\cdot)$ under H_c, they are not affected by the short memory properties of the shocks; cf. Theorem 1.

In Section 4.1 we outline our bootstrap algorithm. The asymptotic properties of the bootstrapped M tests are subsequently established in Section 4.2.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

4.1. The Bootstrap Algorithm

Let $\hat{u}_{t,k}$ denote the OLS residuals from the ADF-type regression, (10), based on the local GLS detrended time series $\hat{X}_t$. Next, generate the bootstrap residuals u_t^b according to the device

$$u_t^b := \tilde{u}_t w_t, \tag{14}$$

$$\tilde{u}_t := \hat{u}_{t,k}, \tag{15}$$

where $\{w_t\}_{t=1}^T$ denotes an independent $N(0,1)$ sequence. In (14) and (15) the multiplicative factor $\tilde{u}_t$ is what distinguishes the wild bootstrap from a simple parametric bootstrap. It serves to replicate the pattern of heteroskedasticity present in the original shocks in the bootstrap shocks because, conditionally on $\hat{u}_{t,k}$, u_t^b is independent over time with zero mean and variance $\hat{u}_{t,k}^2$. The bootstrap sample is then obtained by constructing the partial sum process

$$X_t^b := X_0^b + \sum_{i=1}^{t} u_i^b, \qquad t = 1, \ldots, T, \tag{16}$$

for some initial value X_0^b. Because the asymptotic distributions of the considered tests are not affected by the initial condition as long as $E(X_0^b)^2 < \infty$ (cf. Section 2), we simply set $X_0^b = 0$.

The bootstrap M unit root test statistics, which we will generically refer to as the M^b class of tests, then obtain as follows. Let $\hat{X}_t^b$ denote the local GLS residuals from the regression of X_t^b on z_t; i.e., $\hat{X}_t^b := X_t^b - \hat{\gamma}^{b\prime} z_t$, $t = 0, \ldots, T$, where $\hat{\gamma}^b$ is obtained from the OLS regression of

$$\mathbf{X}_{\bar{c}}^b \equiv [X_0^b, X_1^b - (1 - \bar{c}/T)X_0^b, \ldots, X_T^b - (1 - \bar{c}/T)X_{T-1}^b]' \tag{17}$$

on $\mathbf{z}_{\bar{c}}$; cf. (7). The wild bootstrap analogues of the unit root statistics corresponding to $\mathcal{MZ}_\alpha$, $\mathcal{MZ}_t$, and $\mathcal{MSB}$ are then computed on the bootstrap sample; i.e.,

$$\mathcal{MZ}_\alpha^b := \frac{T^{-1}(\hat{X}_T^b)^2 - (\hat{\sigma}_0^b)^2}{2T^{-2}\sum_{t=1}^{T}(\hat{X}_{t-1}^b)^2}, \qquad \mathcal{MSB}^b := \left(T^{-2}\sum_{t=1}^{T}(\hat{X}_{t-1}^b)^2/(\hat{\sigma}_0^b)^2\right)^{1/2}, \tag{18}$$

and $\mathcal{MZ}_t^b := \mathcal{MZ}_\alpha^b \times \mathcal{MSB}^b$, where $(\hat{\sigma}_0^b)^2 := T^{-1}\sum_{t=1}^{T}(\hat{u}_t^b)^2$, where $\hat{u}_t^b$ are the OLS residuals from the regression of $\hat{X}_t^b$ on $\hat{X}_{t-1}^b$ $(t = 1, \ldots, T)$. Notice that this is a standard, rather than long-run, residual variance estimator, the bootstrap sample having uncorrelated increments.

The bootstrap p-values finally obtain as $p_{T,\mathcal{MZ}_\alpha}^b := G_T^{\mathcal{MZ}_\alpha^b}(\mathcal{MZ}_\alpha)$, $p_{T,\mathcal{MZ}_t}^b := G_T^{\mathcal{MZ}_t^b}(\mathcal{MZ}_t)$, and $p_{T,\mathcal{MSB}}^b := G_T^{\mathcal{MSB}^b}(\mathcal{MSB})$, where $G_T^{\mathcal{MZ}_\alpha^b}$, $G_T^{\mathcal{MZ}_t^b}$, and

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

$G_T^{\mathcal{MSB}^b}$ denote the cumulative distribution functions (c.d.f.s) of $\mathcal{MZ}_\alpha^b$, $\mathcal{MZ}_t^b$, and $\mathcal{MSB}^b$, respectively. In practice these c.d.f.s are, of course, unknown and so may be approximated by repeating the wild bootstrap resampling procedure a large number of times; cf. Hansen (1996) and Andrews and Buchinsky (2001). Specifically, and taking the $\mathcal{MSB}$ statistic to illustrate, from the doubly independent $N(0,1)$ sequence$\{\{w_{j,t}\}_{t=1}^T\}_{j=1}^N$, N conditionally independent bootstrap statistics $\mathcal{MSB}_j^b$, $j = 1,\ldots,N$, are computed as before but from $X_{j,t}^b := X_{j,0}^b + \sum_{i=1}^t u_{j,i}^b$, $u_{j,t}^b := \tilde{u}_t w_{j,t}$, $t = 1,\ldots,T$. The simulated bootstrap p-value is then computed as the percentage of bootstrap statistics that do not exceed the actual statistic; i.e., $\tilde{p}_{T,\mathcal{MSB}}^b := N^{-1}\sum_{j=1}^N \mathbb{I}(\mathcal{MSB}_j^b < \mathcal{MSB})$. Note that an asymptotic standard error for $\tilde{p}_{T,\mathcal{MSB}}^b$ is given by $(\tilde{p}_{T,\mathcal{MSB}}^b(1 - \tilde{p}_{T,\mathcal{MSB}}^b)/N)^{1/2}$; cf. Hansen (1996, p. 419). Applying the same approach to the $\mathcal{MZ}_\alpha$ and $\mathcal{MZ}_t$ statistics yields the corresponding quantities $\tilde{p}_{T,\mathcal{MZ}_\alpha}^b$ and $\tilde{p}_{T,\mathcal{MZ}_t}^b$.

Remark 4. Although, as mentioned previously, the bootstrap statistics do not require a long-run variance estimator, one could equivalently construct the bootstrap p-values by using the statistics

$$\mathcal{MZ}_\alpha^b := \frac{T^{-1}(\hat{X}_T^b)^2 - s_{AR}^b(k^b)^2}{2T^{-2}\sum_{t=1}^T (\hat{X}_{t-1}^b)^2}, \qquad \mathcal{MSB}^b := \left(T^{-2}\sum_{t=1}^T (\hat{X}_{t-1}^b)^2/s_{AR}^b(k^b)^2\right)^{1/2},$$

and $\mathcal{MZ}_t^b := \mathcal{MZ}_\alpha^b \times \mathcal{MSB}^b$, where $s_{AR}^b(k^b)^2$ is the long-run variance estimator

$$s_{AR}^b(k^b)^2 := (\hat{\sigma}^b)^2/(1 - \hat{\beta}^b(1))^2, \qquad \hat{\beta}^b(1) := \sum_{i=1}^{k^b} \hat{\beta}_{i,k^b}^b \tag{19}$$

with $\hat{\beta}_{i,k^b}^b$, $i = 1,\ldots,k^b$, and $(\hat{\sigma}^b)^2$ the OLS slope and variance estimators, respectively, from a standard ADF regression on $\hat{X}_t^b$ using k^b lags. Because the bootstrap sample has uncorrelated increments, k^b in (19) need not diverge to infinity with T, unlike k in (9). Notice that $s_{AR}^b(0) = \hat{\sigma}_0^b$.

Remark 5. In principle, the bootstrap integrated time series (16) does not need to be generated by using the residuals from an ADF regression based on local GLS detrended data. For example, one could obtain the residuals by imposing a unit root in the data. This procedure would not alter the large-sample properties of the bootstrap procedure. More specifically, the asymptotic properties of the wild bootstrap do not change if $\tilde{u}_t$ of (14) is alternatively defined through any one of the following procedures:

(a) $\hat{u}_{t,k}^0$, the residuals from a standard k-lags autoregression run on $\Delta\hat{X}_t$ using k lags (i.e., by imposing a unit root);
(b) $\hat{u}_{t,0}$, the residuals from a simple Dickey–Fuller regression on the local GLS detrended data, i.e., of $\hat{X}_t$ on $\hat{X}_{t-1}$;

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

(c) $\hat{u}_t^0$, obtained by differencing the local GLS detrended data, i.e., $\hat{u}_t^0 := \hat{X}_t - \hat{X}_{t-1}$;
(d) u_t^0, obtained as $X_t - X_{t-1}$, corrected for Δz_t.

Although the large-sample properties of the wild bootstrap are not affected by the choice of $\tilde{u}_t$, numerical evidence suggested that procedure (d) delivered the best finite-sample size and power properties of the five suggested procedures.

Remark 6. As is well known in the wild bootstrap literature (for a review, see Davidson and Flachaire, 2001) in certain cases better bootstrap accuracy can be obtained by replacing the Gaussian distribution used for generating the pseudo-data by an asymmetric distribution with $E(w_t) = 0$, $E(w_t^2) = 1$, and $E(w_t^3) = 1$ (Liu, 1988). A well-known example of this is the Mammen (1993) two-point distribution: $P\big(w_t = -0.5\big(\sqrt{5} - 1\big)\big) = 0.5\big(\sqrt{5} + 1\big)/\sqrt{5} =: \pi$, $P\big(w_t = 0.5\big(\sqrt{5} + 1\big)\big) = 1 - \pi$. We found no discernible differences between the finite-sample properties of the bootstrap unit root tests based on the Gaussian or Mammen distribution.

4.2. Asymptotic Properties

In this section we discuss the asymptotic properties of the wild bootstrap unit root tests. We show that under the unit root null, and for any volatility process satisfying Assumption $\mathcal{A}_3$, the bootstrap statistics converge to the same asymptotic distribution as the test statistics computed on the original data; i.e., the wild bootstrap allows us to replicate the correct first-order asymptotic null distribution of the test statistics of interest. We also demonstrate the asymptotic properties of the bootstrap statistics under both near-integrated and stable autoregressive alternatives.

Recall from Section 2 that in the presence of nonstationary volatility satisfying Assumption $\mathcal{A}_3$, $\tilde{\xi}_1^{c,\bar{c},z,\eta}$, $\tilde{\xi}_2^{c,\bar{c},z,\eta}$, and $\tilde{\xi}_3^{c,\bar{c},z,\eta}$ denote the asymptotic distributions of $\mathcal{MZ}_\alpha$, $\mathcal{MSB}$, $\mathcal{MZ}_t$, respectively, under $H_c: \alpha = 1 - c/T$. The usefulness of the wild bootstrap in the present framework is given in the following result, which shows that our bootstrap procedure allows us to retrieve the correct asymptotic null distribution and hence that the p-values based on $G_T^{\mathcal{MZ}_\alpha^b}$, $G_T^{\mathcal{MZ}_t^b}$, and $G_T^{\mathcal{MSB}^b}$ are all asymptotically pivotal.

THEOREM 2. *Let the conditions of Theorem 1 hold and let the bootstrap sample be generated as detailed in Section 4.1. Then* $\mathcal{MZ}_\alpha^b \xrightarrow{w}_p \tilde{\xi}_1^{0,\bar{c},p,\eta}$, $\mathcal{MSB}^b \xrightarrow{w}_p \tilde{\xi}_2^{0,\bar{c},p,\eta}$, *and* $\mathcal{MZ}_t^b \xrightarrow{w}_p \tilde{\xi}_3^{0,\bar{c},p,\eta}$. *Moreover, if* $\alpha = 1$, $p_{T,\mathcal{MZ}_\alpha}^b \xrightarrow{w} U[0,1]$, $p_{T,\mathcal{MZ}_t}^b \xrightarrow{w} U[0,1]$, *and* $p_{T,\mathcal{MSB}}^b \xrightarrow{w} U[0,1]$.

Theorem 2 proves the usefulness of the wild bootstrap under the null hypothesis: as the number of observations diverges, the bootstrapped statistics have the same null distribution as the original test statistics, and consequently the

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

bootstrap p-values are uniformly distributed under the null hypothesis, leading to tests with asymptotically correct size. As Theorem 2 also shows, these limiting distributions also hold for the bootstrapped statistics under local alternatives. This immediately implies that the bootstrap unit root tests will have the same asymptotic local power functions as the original unit root tests, provided the latter were run using critical values from the appropriate limiting null distribution, i.e., the asymptotic local power functions given in (13). The asymptotic theory therefore predicts that the bootstrap tests should have power approximately equal to the size-adjusted power of the standard tests.

We now turn to the behavior of the bootstrap statistics under stable autoregressive alternatives. The distribution of the bootstrap statistics is derived under the assumption that α of (2) is bounded by $|\alpha| < 1$, and it is shown that, as a result, our bootstrap unit root tests are consistent.

THEOREM 3. *Let $\{X_t\}_0^T$ be generated as in (1)–(4) under Assumptions $\mathcal{A}$ and $\mathcal{K}$, with $|\alpha| < 1$, and let the bootstrap sample be generated as in Section 4.1. Then, $\mathcal{MZ}_\alpha^b \xrightarrow{w}_p \tilde{\xi}_1^{\bar{c},p,\eta}$, $\mathcal{MSB}^b \xrightarrow{w}_p \tilde{\xi}_2^{\bar{c},p,\eta}$, $\mathcal{MZ}_t^b \xrightarrow{w}_p \tilde{\xi}_3^{\bar{c},p,\eta}$. Moreover, $\mathcal{MZ}_\alpha$ and $\mathcal{MZ}_t$ diverge to minus infinity whereas $\mathcal{MSB} \xrightarrow{p} 0$; hence, $p^b_{T,\mathcal{MZ}_\alpha} \xrightarrow{p} 0$, $p^b_{T,\mathcal{MZ}_t} \xrightarrow{p} 0$, $p^b_{T,\mathcal{MSB}} \xrightarrow{p} 0$.*

The finite-sample behavior of the standard and bootstrap M tests is explored in the next section.

5. FINITE-SAMPLE SIMULATIONS

In this section we analyze the finite-sample size and power properties of the bootstrap M^b unit root tests of Section 4 by means of Monte Carlo simulation. Precisely, the $\mathcal{MZ}_\alpha^b$, $\mathcal{MZ}_t^b$, and $\mathcal{MSB}^b$ tests are considered and compared with the standard $\mathcal{MZ}_\alpha$, $\mathcal{MZ}_t$, and $\mathcal{MSB}$ tests. Results are presented for the constant ($p = 0$) case. Results for the linear trend ($p = 1$) case are qualitatively similar and are available on request.

We generated data according to (1)–(4), with $X_0 = 0$, for samples of size $T =$ 100, 250, and 500 with $e_t \sim IN(0,1)$, using the Kiss+Monster function of Gauss 5.0. All experiments were conducted using 10,000 Monte Carlo replications. The standard M tests were computed as in (8) and were run using the nominal asymptotic 0.05 level critical values appropriate for the homoskedastic case. The bootstrap M^b tests were constructed as outlined in Section 4 with $k^b = k$, using the first difference residual bootstrap (scheme (d) outlined in Remark 5) and Gaussian pseudo-data. The number of bootstrap replications used was $N = 499$ throughout.

Initially, and to demonstrate the effects of heteroskedasticity on the finite-sample size (Section 5.1) and power (Section 5.2) properties of the standard M and corresponding bootstrap M^b tests uncontaminated by the separate effects

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

of serial dependence, we set $C(L) = 1$ in (3), so that the error term $u_t = \sigma_t e_t$. Correspondingly, we set $k = 0$ in (9) and $k^b = 0$ in (19). In Section 5.3 we extend the analysis to also allow for weak dependence in u_t, reporting results for a subset of the cases reported in Sections 5.1 and 5.2.

Results are reported for the following four models for the time-series behavior of the volatility term σ_t^2:

A. Single volatility shift. A single abrupt shift in volatility from σ_0^2 to σ_1^2 occurring at time $\lfloor \tau T \rfloor$, which corresponds to the variance function

$$\omega_1(s)^2 := \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{I}(s \geq \tau), \qquad s \in [0,1] \tag{20}$$

with $\tau \in (0,1)$.

B. Two volatility shifts. Two volatility shifts, the first from σ_0^2 to σ_1^2 occurring at time $\lfloor \tau_1 T \rfloor$, the second from σ_1^2 to σ_0^2 occurring at time $\lfloor \tau_2 T \rfloor$, which corresponds to the variance function

$$\omega_2(s)^2 := \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{I}(\tau_1 \leq s < \tau_2), \qquad s \in [0,1]$$

for $\tau_1, \tau_2 \in (0,1)$, $\tau_1 < \tau_2$.

C. Trending volatility. Volatility follows a linear trend, between σ_0 for $t = 0$ and σ_1 for $t = T$, which corresponds to the volatility function

$$\omega_3(s) := \sigma_0 + (\sigma_1 - \sigma_0)s, \qquad s \in [0,1]. \tag{21}$$

D. Exponential (near-) integrated stochastic volatility. Standard models of (near-) integrated SV can be approximated through the variance function (see the discussion in Hansen, 1995, p. 1116),

$$\omega_4(s)^2 := \sigma_0^2 \exp(\nu J_c(s)), \qquad s \in [0,1]$$

for $\nu > 0$, with $J_c(\cdot)$ either an Ornstein–Uhlenbeck process ($c > 0$) or a Brownian motion ($c = 0$).

Without loss of generality we set $\sigma_0^2 = 1$ in all cases. For Model A we let the ratio $\delta := \sigma_0/\sigma_1$ vary among $\{\frac{1}{5}, 5\}$, and the break date among $\tau \in \{0.1, 0.9\}$, so that both early and late breaks, which may be positive ($\delta < 1$) or negative ($\delta > 1$), are allowed. For Model B we vary the ratio $\delta := \sigma_0/\sigma_1$ among $\{\frac{1}{5}, 5\}$ and consider symmetric shifts, such that $\tau_2 = 1 - \tau_1$, with $\tau \in \{0.05, 0.45\}$, thereby allowing for breaks that occur either toward the middle or the start and end of the sample. For Model C we let $\delta := \sigma_0/\sigma_1$ take values among $\{\frac{1}{5}, 5\}$ so that both positively and negatively trending variances are generated. Finally, for Model D we vary ν among $\{4, 9\}$ and c among $\{0, 10, 20\}$. It should stressed that, formally, Model D does not satisfy Assumption $\mathcal{A}_3$; however, see Remark 2.

For each of Models A–D other combinations of parameter values were also considered, including asymmetric shifts in the case of Model B, but these qual-

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

itatively add little to the reported results. In the case of Models A and B we also considered smooth (logistic) transition shifts in variance. These results were qualitatively similar to those reported subsequently, with the effects of the volatility breaks diminishing somewhat as the speed of transition parameter is reduced, other things being equal, as would be expected. We also considered the cases where the volatility process follows either a covariance stationary autoregressive SV process of the form $\sigma_t^2 = \sigma_0^2 \exp(\nu h_t)$, $h_t = \gamma h_{t-1} + e_t$, or a covariance stationary autoregressive conditionally heteroskedastic GARCH(1,1) process of the form $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ where in both cases e_t is as before. In each case, both the standard and bootstrap tests behaved almost identically to the constant volatility case here, as expected; cf. Remark 1. All of these results are available on request.

5.1. Size Properties

Tables 1 and 2 present the (empirical) size, for $\alpha = 1$ in (2), of both the standard M unit root tests and the corresponding M^b heteroskedastic-robust bootstrap unit root tests of Section 4.

Let us initially consider the single break model (Table 1A). As for the unit root tests considered by Kim et al. (2002) and Cavaliere (2004), the standard M unit root tests are badly oversized in the presence of early negative and late positive breaks. For example, the $\mathcal{MSB}$ statistic shows empirical size of around 20% in the presence of late positive breaks. In contrast, the size properties of our M^b bootstrap unit root tests seem largely satisfactory throughout. It is also worth noting that in the homoskedastic case, the bootstrap yields tests with finite-sample size closer to the nominal level than the standard tests based on asymptotic critical values.

The results for the double volatility break model (Table 1B) are broadly comparable to those obtained for the single break model. As far as the standard M tests are concerned, the most severe size biases are seen for a positive followed by a negative break, both occurring toward the middle of the sample (actual size varies between around 12% and 15%), whereas a negative followed by a positive break again toward the middle of the sample also causes significant oversizing in the standard tests. Conversely, our proposed M^b tests perform well, completely avoiding the oversizing problems seen with the standard M tests.

The results in Table 1C show that, in general, linear trending volatility has a lower impact on the size of the standard tests than abrupt changes. The M tests are notably oversized for $p = 0$ (around 10% size) in the presence of upward trending volatility ($\delta = \frac{1}{5}$). Downward trending volatility ($\delta = 5$) appears to have rather less impact. Again the M^b tests display empirical sizes very close to the nominal level throughout.

The size properties of the tests in the presence of (near-) integrated volatility are reported in Table 2. In the "mildest" case, when ν is low ($\nu = 4$) and vol-

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

TABLE 1. Volatility Models A–C: Size of bootstrap and standard unit root tests

Volatility model	T	$\mathcal{MZ}_\alpha^b$	$\mathcal{MZ}_t^b$	$\mathcal{MSB}^b$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$
A. Single break							
Constant	100	4.8	4.8	4.7	5.9	5.9	5.3
	250	5.4	5.3	5.3	5.9	5.9	5.4
	500	5.0	5.1	4.8	5.2	5.2	4.9
$\delta = \frac{1}{5}$, $\tau = 0.1$	100	4.9	4.9	5.0	7.2	7.3	6.9
	250	5.2	5.2	5.0	6.9	6.6	6.3
	500	5.0	4.9	5.1	6.1	6.1	6.0
$\delta = \frac{1}{5}$, $\tau = 0.9$	100	5.5	5.6	4.8	16.7	14.6	23.5
	250	5.2	5.2	5.2	15.8	14.0	20.8
	500	5.5	5.4	5.2	15.1	14.0	19.3
$\delta = 5$, $\tau = 0.1$	100	6.1	6.1	6.1	11.2	11.2	10.2
	250	5.4	5.1	5.4	9.3	9.5	8.4
	500	5.3	5.4	5.2	9.8	9.8	9.0
$\delta = 5$, $\tau = 0.9$	100	4.7	4.8	4.8	5.7	5.8	4.5
	250	5.3	5.4	5.1	5.9	5.9	4.7
	500	4.8	4.8	4.8	5.2	5.3	4.4
B. Double break							
$\delta = \frac{1}{5}$, $\tau = 0.05$	100	5.0	5.0	4.9	6.3	6.4	5.4
	250	5.3	5.4	5.2	6.1	6.3	5.3
	500	4.6	4.6	4.7	5.2	5.3	4.7
$\delta = \frac{1}{5}$, $\tau = 0.45$	100	5.6	5.6	5.5	13.7	13.8	12.3
	250	5.2	5.2	5.1	12.9	13.0	12.0
	500	5.3	5.1	5.1	12.8	12.8	11.8
$\delta = 5$, $\tau = 0.45$	100	5.5	5.7	5.2	10.0	8.9	14.0
	250	5.7	5.6	5.5	9.8	9.1	12.6
	500	5.2	5.2	4.9	9.2	8.2	11.5
C. Trending							
$\delta = \frac{1}{5}$	100	5.1	5.2	5.1	9.7	9.0	10.5
	250	5.0	5.1	5.1	9.2	8.7	10.0
	500	4.7	5.0	4.6	8.3	8.0	8.9
$\delta = 5$	100	5.1	5.0	5.1	6.6	6.8	5.5
	250	5.6	5.5	5.5	6.3	6.7	5.3
	500	4.9	4.9	4.9	5.7	6.0	4.7

atility is strongly mean reverting ($c = 20$), the standard M tests perform well in terms of size, which is rarely in excess of 7%. However, when ν 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 ν is

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

TABLE 2. Exponential (near-) integrated stochastic volatility model: Size of bootstrap and standard unit root tests

Volatility model	T	$\mathcal{MZ}_\alpha^b$	$\mathcal{MZ}_t^b$	$\mathcal{MSB}^b$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$
$c = 0$, $\nu = 4$	100	5.5	5.5	5.5	12.8	12.4	13.4
	250	5.5	5.4	5.3	11.8	11.4	12.1
	500	4.9	4.9	4.6	10.8	10.3	11.2
$c = 10$, $\nu = 4$	100	5.4	5.5	5.4	7.8	7.8	7.5
	250	5.4	5.4	5.5	7.4	7.3	6.9
	500	5.0	5.1	5.0	6.5	6.2	6.1
$c = 20$, $\nu = 4$	100	5.5	5.5	5.6	7.2	7.0	6.7
	250	5.6	5.5	5.5	6.8	6.8	6.3
	500	5.2	5.1	5.1	5.7	5.7	5.4
$c = 0$, $\nu = 9$	100	6.3	6.3	5.9	20.5	19.6	23.0
	250	5.6	5.5	5.8	20.3	19.6	22.7
	500	5.2	5.1	5.1	18.9	18.0	21.6
$c = 10$, $\nu = 9$	100	6.1	5.9	5.9	12.0	11.9	12.1
	250	5.4	5.3	5.4	11.6	11.1	11.7
	500	4.9	4.9	4.9	11.2	10.8	11.0
$c = 20$, $\nu = 9$	100	6.0	6.0	5.9	9.6	9.4	9.1
	250	5.6	5.5	5.5	8.5	8.2	8.3
	500	5.0	5.1	5.1	7.8	7.6	7.4

low, the rejection frequencies of the standard unit root tests range between about 8% and 14%. When ν is high, such frequencies all but double. Significant oversizing is also seen when ν is high but the process is mean reverting ($c = 10, 20$). The performance of the M^b tests is very good, with empirical size never in excess of 6.5%, supportive of our conjecture in Remark 2.

5.2. Power Properties

Tables 3 and 4 report the finite-sample powers of the bootstrap M^b unit root tests of Section 4, together with the (size-adjusted) powers of the corresponding standard M unit root tests. The reported results are for the same models as considered in Section 5.1 except that we replace $\alpha = 1$ with the local alternative $\alpha = 1 - 7/T$ in (2). Under constant volatility, therefore, power would be expected to be close to 50%, and this should be viewed as the relevant benchmark when reading the results. Qualitatively similar conclusions were drawn for other values of α. In the case of the standard M tests size-adjusted power is reported to control for the unreliable size properties of these tests under nonstationary volatility; see Section 5.1.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

TABLE 3. Volatility Models A–C: Empirical power of bootstrap and size-adjusted power of standard tests

Volatility model	T	$\mathcal{MZ}_\alpha^b$	$\mathcal{MZ}_t^b$	$\mathcal{MSB}^b$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$
A. Single break							
Constant	100	43.5	44.3	40.5	42.1	43.0	40.2
	250	46.9	47.6	44.6	47.7	49.0	45.1
	500	47.2	48.0	44.5	47.6	47.9	44.8
$\delta = \frac{1}{5}, \tau = 0.1$	100	42.8	43.7	39.5	41.2	41.8	38.8
	250	43.8	44.3	41.7	42.9	42.9	42.0
	500	43.6	44.4	41.7	44.2	45.0	42.3
$\delta = \frac{1}{5}, \tau = 0.9$	100	34.9	33.4	33.4	29.9	28.8	30.1
	250	36.2	34.6	37.5	34.4	31.7	35.9
	500	35.1	34.1	37.5	34.9	33.1	35.9
$\delta = 5, \tau = 0.1$	100	23.4	23.9	22.3	19.9	20.3	19.0
	250	29.0	30.1	27.7	27.1	27.4	25.5
	500	32.7	33.6	31.4	31.5	32.2	31.1
$\delta = 5, \tau = 0.9$	100	43.5	45.2	40.6	43.4	45.8	40.4
	250	47.7	49.4	44.8	48.3	50.0	45.3
	500	47.5	49.7	44.5	48.3	50.3	44.6
B. Double break							
$\delta = \frac{1}{5}, \tau = 0.05$	100	46.6	48.5	43.0	44.6	46.6	41.8
	250	47.8	49.5	44.8	49.9	51.0	46.2
	500	46.4	48.2	43.4	47.2	48.2	44.5
$\delta = \frac{1}{5}, \tau = 0.45$	100	36.8	38.0	34.8	32.1	33.5	30.7
	250	36.9	38.0	35.1	35.3	36.6	33.8
	500	37.3	38.3	36.0	36.4	37.1	35.0
$\delta = 5, \tau = 0.45$	100	22.4	22.4	19.6	19.9	19.4	19.6
	250	28.7	28.2	27.6	27.1	26.3	27.3
	500	33.5	32.9	33.3	31.2	30.7	31.6
C. Trending							
$\delta = \frac{1}{5}$	100	37.5	36.9	35.4	35.7	35.1	34.5
	250	39.1	38.8	38.3	40.0	38.9	38.1
	500	38.7	37.8	37.7	39.0	38.0	38.6
$\delta = 5$	100	37.1	38.3	35.0	36.3	37.9	34.5
	250	42.9	44.5	40.7	43.7	45.4	41.2
	500	44.6	46.1	42.1	43.1	44.4	40.5

Consider first the case where the errors are homoskedastic, which is reported in Table 3A. Here the bootstrap M^b tests display roughly the same power as the corresponding standard M tests. Hence, appealingly, the bootstrap does not appear to lead to any significant power loss when the data are homoskedastic.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

TABLE 4. Exponential (near-) integrated stochastic volatility model: Empirical power of bootstrap and size-adjusted power of standard tests

Volatility model	T	$\mathcal{MZ}_\alpha^b$	$\mathcal{MZ}_t^b$	$\mathcal{MSB}^b$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$
$c = 0$, $\nu = 4$	100	32.2	32.3	30.7	24.8	26.2	22.8
	250	35.2	35.5	33.7	26.8	27.3	24.1
	500	34.6	34.9	33.6	29.6	30.8	27.7
$c = 10$, $\nu = 4$	100	41.7	42.4	39.3	39.4	39.8	36.8
	250	44.1	44.5	42.3	41.3	41.5	39.3
	500	44.1	44.8	42.5	44.6	44.9	42.8
$c = 20$, $\nu = 4$	100	43.3	43.9	40.3	43.7	44.1	41.1
	250	45.7	46.1	43.5	43.0	43.7	40.6
	500	46.5	46.7	44.0	45.4	46.1	42.9
$c = 0$, $\nu = 9$	100	23.7	24.4	22.3	13.5	14.8	11.7
	250	25.2	25.6	24.5	15.0	16.2	13.1
	500	24.7	25.1	24.3	15.5	16.6	14.0
$c = 10$, $\nu = 9$	100	36.5	37.5	34.3	29.4	29.9	28.5
	250	38.1	38.7	36.1	30.9	32.0	28.7
	500	37.2	37.7	35.6	33.7	34.9	32.4
$c = 20$, $\nu = 9$	100	41.0	41.7	38.1	35.7	36.9	33.2
	250	42.2	42.7	40.5	37.8	38.6	36.1
	500	41.1	42.0	39.1	40.4	41.2	38.4

Under heteroskedasticity (Tables 3 and 4) the standard M tests tend to experience a decrease in size-adjusted power, relative to the constant volatility case, in cases where the tests are significantly oversized under the null. This effect can be most clearly seen in the single break model with a late positive break or early negative break, in the double break model where a positive followed by a negative break or a negative followed by a positive break occur toward the middle of the sample, and in the (near-) integrated volatility case when either $c = 0$ or ν is large. As anticipated by the asymptotic theory (see the discussion after Theorem 2), the bootstrap M^b tests display power that is very similar to the size-adjusted power of the corresponding M tests throughout. The only exception seems to be the case of integrated volatility ($c = 0$), especially when ν is large, where the M^b tests display power somewhat in excess of the size-adjusted power of the corresponding M tests.

Overall, our bootstrap M^b tests perform very well in practice. They all but eliminate the serious size distortions seen with the standard M tests across a range of models of nonstationary volatility, yet display almost identical finite-sample power properties to (infeasible) size-adjusted M tests under near-integrated alternatives.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

5.3. Serial Correlation

Consider now the case where we allow for both nonstationary volatility and serial dependence in the shocks, so that $C(L) \neq 1$ in (3). Because serial correlation is now permitted the M and M^b tests are constructed as detailed at the start of this section except that now, as in Ng and Perron (2001, p. 1536), we set the maximum lag $k = k^b = \lfloor 12(T/100)^{1/4} \rfloor$ in (9) and (19). The preferred modified Akaike information criterion (MAIC) lag selection criterion of Ng and Perron (2001) was employed throughout. We also computed the corresponding experiments for $k^b = 0$: these results were little different from those reported.

Table 5 reports the size of the M and M^b tests with data-dependent lag selection for data generated as outlined at the start of this section with $\alpha = 1$, but where the errors are now generated according to either an MA(1) process, i.e., $u_t = \varepsilon_t + \theta\varepsilon_{t-1}$, or an AR(1) process, i.e., $u_t = \phi u_{t-1} + \varepsilon_t$ so that $C(L) = (1 + \phi L + \phi^2 L^2 + \ldots)$, in (3). The moving average parameter θ was chosen within the set $\{0, \pm 0.5, \pm 0.8\}$, and the autoregressive parameter within the set $\{0, \pm 0.8\}$. For space constraints results are reported for two of the volatility models, namely, the constant volatility and late negative single volatility shift ($\delta = \frac{1}{5}$, $\tau = 0.9$) models,[2] for $T = 100, 250$.

Consider first the results for the constant volatility model in Table 5. Here we can compare the finite-sample performance of the bootstrap M^b and corresponding standard M tests in the presence of serially correlated shocks. Overall there is really little to choose between the standard and bootstrap tests. Both do, overall, an excellent job in controlling size. However, consistent with the findings of Ng and Perron (2001), significant undersize is seen for the large negative autoregressive model ($\phi = -0.8$) and for the large negative moving average model ($\theta = -0.8$), whereas some oversize is seen with the large positive autoregressive model ($\phi = 0.8$) and to a lesser extent positive moving average models.

Turning to the nonstationary volatility model, we see that for the standard M tests the size distortions observed for this case with serially uncorrelated shocks in Table 1 and the size distortions noted previously for the serially correlated but homoskedastic case appear to combine. In general the oversize due to the nonconstant variance process is further increased in cases where the serial correlation in the shocks causes oversize: this is particularly pronounced in the case of either large positive autoregressive ($\phi = 0.8$) or moving average ($\theta = 0.8$) models. The reverse holds true for cases where serial correlation causes undersizing. The bootstrap M^b tests are much less badly affected than the standard tests, although significant finite-sample oversize is again seen for the large positive autoregressive and moving average models. These distortions are, however, significantly reduced between $T = 100$ and $T = 250$, as predicted by the asymptotic theory.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

TABLE 5. Single volatility shift model with ARMA ($u_t = \phi u_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$) shocks: Size of bootstrap and standard tests

Volatility model	T	ϕ	θ	$\mathcal{MZ}_\alpha^b$	$\mathcal{MZ}_t^b$	$\mathcal{MSB}^b$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$
Constant	100	0.0	0.0	4.8	4.6	4.6	4.4	4.4	3.9
		−0.8	0.0	1.0	1.1	1.0	0.9	1.0	0.8
		0.8	0.0	7.5	7.2	7.9	7.5	7.1	7.3
		0.0	0.8	6.7	6.6	6.7	6.5	6.3	6.2
		0.0	0.5	5.9	5.7	5.6	5.6	5.5	4.9
		0.0	−0.5	5.7	5.6	5.3	5.2	5.3	4.4
		0.0	−0.8	3.7	3.8	3.2	3.5	3.6	3.0
	250	0.0	0.0	5.5	5.5	5.4	5.4	5.4	4.6
		−0.8	0.0	3.2	3.6	3.2	3.1	3.2	2.6
		0.8	0.0	6.3	6.0	6.2	5.9	5.8	5.7
		0.0	0.8	6.4	6.4	6.4	6.2	6.2	5.8
		0.0	0.5	6.0	5.7	5.9	5.7	5.7	5.4
		0.0	−0.5	6.4	6.6	5.8	5.9	6.1	5.3
		0.0	−0.8	2.5	2.5	2.3	2.3	2.4	1.7
$\delta = \frac{1}{5}$, $\tau = 0.9$	100	0.0	0.0	6.1	6.0	6.2	16.8	14.6	22.2
		−0.8	0.0	6.3	5.9	6.6	11.4	10.4	15.2
		0.8	0.0	12.7	12.8	12.6	27.5	25.3	31.3
		0.0	0.8	10.2	10.2	10.1	21.5	19.5	26.3
		0.0	0.5	8.4	8.3	8.3	19.9	17.4	25.4
		0.0	−0.5	4.5	4.4	4.8	11.5	10.0	16.3
		0.0	−0.8	2.8	2.7	2.8	6.4	6.0	8.8
	250	0.0	0.0	5.3	5.3	5.3	18.4	16.0	22.9
		−0.8	0.0	6.3	5.9	6.7	14.1	12.5	18.3
		0.8	0.0	9.8	9.9	9.7	24.6	22.5	28.8
		0.0	0.8	7.4	7.5	7.4	20.1	18.3	24.7
		0.0	0.5	6.4	6.4	6.3	19.7	17.3	24.4
		0.0	−0.5	3.9	3.7	4.3	14.1	12.1	18.6
		0.0	−0.8	1.0	1.0	1.0	4.8	4.3	7.0
$\delta = 5$, $\tau = 0.1$	100	0.0	0.0	6.2	6.1	6.2	9.5	9.5	8.6
		−0.8	0.0	2.1	2.3	1.8	2.7	2.8	2.2
		0.8	0.0	2.1	2.0	2.0	2.7	2.6	2.5
		0.0	0.8	1.8	1.8	1.9	2.6	2.8	2.2
		0.0	0.5	3.5	3.5	3.6	5.4	5.4	4.5
		0.0	−0.5	5.6	5.8	5.6	8.6	8.7	7.6
		0.0	−0.8	3.2	3.3	3.1	4.4	4.7	3.9
	250	0.0	0.0	4.2	4.3	4.2	2.8	2.8	2.5
		−0.8	0.0	3.4	3.6	3.1	2.0	2.1	1.7
		0.8	0.0	2.6	2.6	2.6	1.8	1.8	1.6
		0.0	0.8	3.6	3.4	3.3	2.3	2.4	1.9
		0.0	0.5	3.6	3.7	3.6	2.5	2.6	2.1
		0.0	−0.5	3.8	3.9	3.6	2.5	2.6	2.2
		0.0	−0.8	2.0	2.1	1.7	1.3	1.4	1.0

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

6. CONCLUSIONS

The presence of nonstationary volatility has potentially serious implications for the reliability of unit root tests, affecting tests whose true size can be significantly in excess of the nominal significance level. To rectify this problem, we have proposed a wild bootstrap–based approach to testing for a unit root. Unlike previous tests that have appeared in the literature, our approach requires only that the innovation volatility is nonstochastic, is bounded, and has a countable number of jumps. Our proposed bootstrap unit root tests have the considerable advantage that they are not tied to a given parametric model of volatility within the class of nonstationary volatility processes considered. This is a particularly attractive feature because economic and financial series appear to be subject to both smooth volatility changes and multiple volatility changes, the break dates associated with which can be hard to identify in practice. As with any bootstrap procedure, no tables of critical values are required as the procedure automatically delivers a *p*-value for the hypothesis being tested. The asymptotic validity of our proposed bootstrap tests within the class of nonstationary volatility considered was demonstrated. Monte Carlo evidence was also reported that suggested that the proposed bootstrap unit root tests perform well in finite samples, avoiding the large oversize problems that can occur with the standard tests in the presence of a variety of forms of nonstationary volatility, yet emulating the finite-sample power properties of (infeasible) size-adjusted implementations of the standard tests.

Although the class of nonstationary volatility processes considered in this paper is wide, containing many of the forms of nonstationarity argued for in the empirical literature, we have imposed the condition that the volatility process is nonstochastic. We conjecture that the wild bootstrap–based unit root tests outlined in this paper will remain asymptotically valid in the presence of (exogenous) stochastic volatility under suitable moment and memory conditions. A proof of this conjecture is beyond the scope of the present paper but is currently under investigation by the authors. Monte Carlo results presented in the paper for the empirically important case of near-integrated autoregressive stochastic volatility appeared supportive of our conjecture.

NOTES

1. The effects of heteroskedastic errors satisfying $\mathcal{A}_3$ on stable autoregressions recently have been analyzed by Phillips and Xu (2006).

2. Corresponding experiments were conducted for the other models of nonstationary volatility considered in Sections 5.1 and 5.2, but qualitatively these neither added to nor contradicted the reported results.

REFERENCES

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.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.
Busetti, F. & A.M.R. Taylor (2003) Testing against stochastic trend in the presence of variance shifts. *Journal of Business & Economic Statistics* 21, 510–531.
Cavaliere, G. (2003) Asymptotics for unit root tests under Markov-regime switching. *Econometrics Journal* 6, 193–216.
Cavaliere, G. (2004) Unit root tests under time-varying variances. *Econometric Reviews* 23, 259–292.
Cavaliere, G. & A.M.R. Taylor (2004) Testing for Unit Roots in Time Series Models with Non-stationary Volatility. Working paper 04–24, Department of Economics, University of Birmingham.
Cavaliere, G. & A.M.R. Taylor (2007) Testing for unit roots in time series models with non-stationarity volatility. *Journal of Econometrics*, forthcoming.
Chang, Y. & J.Y. Park (2002) On the asymptotics of ADF tests for unit roots. *Econometric Reviews* 21, 431–447.
Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.
Davidson, R. & E. Flachaire (2001) The Wild Bootstrap, Tamed at Last. Working paper IER 1000, Queens University.
Elliott, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.
Engle, R.F. & J. Gonzalo Rangel (2004) The Spline GARCH Model for Unconditional Volatility and Its Global Macroeconomic Causes. Mimeo, New York University.
Giné, E. & J. Zinn (1990) Bootstrapping general empirical measures. *Annals of Probability* 18, 851–869.
Gonçalves, S. & L. Killian (2004) Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89–120.
Hamori, S. & A. Tokihisa (1997) Testing for a unit root in the presence of a variance shift. *Economics Letters* 57, 245–253.
Hansen, B.E. (1995) Regression with nonstationary volatility. *Econometrica* 63, 1113–1132.
Hansen, B.E. (1996) Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64, 413–430.
Hansen, B.E. (2000a) Sample splitting and threshold estimation. *Econometrica* 68, 575–603.
Hansen, B.E. (2000b) Testing for structural change in conditional models. *Journal of Econometrics* 97, 93–115.
Hansen, E. & A. Rahbek (1998) Stationarity and Asymptotics of Multivariate ARCH Time Series with an Application to Robustness of Cointegration Analysis. Preprint 12/1998, Department of Theoretical Statistics, University of Copenhagen.
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.
Kim, K. & P. Schmidt (1993) Unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 59, 287–300.
Kim, T.H., S. Leybourne, & P. Newbold (2002) Unit root tests with a break in innovation variance. *Journal of Econometrics* 109, 365–387.
Koop, G. & S.M. Potter (2000) Nonlinearity, structural breaks or outliers in economic time series? In W.A. Barnett, D.F. Hendry, S. Hylleberg, T. Terasvirta, D. Tjostheim, & A.H. Wurtz (eds.), *Nonlinear Econometric Modelling in Time Series Analysis*, pp. 61–78. Cambridge University Press.
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.
Lopez, J.H. (1997) The power of the ADF test. *Economics Letters* 57, 5–10.
Loretan, M. & P.C.B. Phillips (1994) Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datsets. *Journal of Empirical Finance* 1, 211–248.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

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.
Nelson, C.R., J. Piger, & E. Zivot (2001) Markov regime-switching and unit root tests. *Journal of Business & Economics Statistics* 19, 404–415.
Ng, S. & P. Perron (2001) Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519–1554.
Park, J.Y. (2003) Bootstrap unit root tests. *Econometrica* 71, 1845–1895.
Perron, P. (1989) The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–1401.
Perron, P. (1990) Testing for a unit root in a time series with a changing mean. *Journal of Business & Economic Statistics* 8, 153–162.
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. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.
Phillips, P.C.B. & K.-L. Xu (2006) Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis* 27, 289–308.
Said, S.E. & D.A. Dickey (1984) Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* 89, 1420–1437.
Sensier, M. & D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *Review of Economics and Statistics* 86, 833–839.
Stock, J.H. (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. (1999) A class of tests for integration and cointegration. In R.F. Engle & H. White (eds.), *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, pp. 137–167. Oxford University Press.
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.
van Dijk, D., D.R. Osborn, & M. Sensier (2002) Changes in Variability of the Business Cycle in the G7 Countries. Econometric Institute Report EI 2002–28, Erasmus University Rotterdam.
Watson, M.W. (1999) Explaining the increased variability in long-term interest rates. *Federal Reserve Bank of Richmond Economic Quarterly* 85, 71–96.

MATHEMATICAL APPENDIX

In the following discussion we assume that $\gamma = 0$ without loss of generality. Some of the proofs are based on the following lemmas. More specifically, in Lemmas 2 and 3 two uniform consistency results for the squared residuals $\hat{u}_{t,k}^2$ from (10) are given.

LEMMA 1. *Let* $\omega_{T,k}^2(s) := T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor}\sigma_t^2$ *and let* $f_j(z) := \sum_{i=0}^{\infty} c_i c_{i+j} z^i$. *Then, under Assumption* $\mathcal{A}$, *as* $T \to \infty$, *and uniformly for all* $j = 0,\dots,k = o(T)$: *(i)* $E(T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor} u_t u_{t-j} - f_j(1)\omega_{T,k}^2(s))^2 = O(T^{-1})$ *and (ii)* $T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor} u_t u_{t-j} \xrightarrow{L^2} f_j(1)\bar{\omega}^2\eta(s)$, *where* $\xrightarrow{L^2}$ *denotes convergence in* L^2 *as* $T \to \infty$.

Proof. The proof is similar to the proof of Lemma 3 in Cavaliere and Taylor (2007) and is therefore omitted in the interest of brevity. ■

LEMMA 2. *Let the conditions of Theorem 1 hold. Then,* $T^{-1}\sum_{t=1}^{\lfloor sT\rfloor}\hat{u}_{t,k}^2 \xrightarrow{p} \int_0^s \omega(r)^2\,dr$ *uniformly for all* $s \in [0,1]$.

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

Proof of Lemma 2. We assume, without loss of generality (see Chang and Park, 2002; Park, 2003), that $c = 0$ and that the residuals $\hat{u}_{t,k}$ are obtained under the null hypothesis, i.e., by estimating the regression $\Delta\hat{X}_t = \sum_{i=1}^{k}\beta_{i,k}\Delta\hat{X}_{t-i} + \hat{u}_{t,k}$. In this case, if $\hat{X}_t$ is replaced by X_t, so that the estimated regression is $u_t = \sum_{i=1}^{k}\breve{\beta}_{i,k}u_{t-i} + \breve{u}_{t,k}$, in Cavaliere and Taylor (2007) it is shown that $T^{-1}\sum_{t=1}^{\lfloor sT\rfloor}\breve{u}_{t,k}^2 \xrightarrow{p} \int_0^s \omega(r)^2\,dr$. We now generalize this result to the case of local GLS residuals. Initially, we show that the difference between $\breve{\beta}_k := (\breve{\beta}_{1,k},\ldots,\breve{\beta}_{k,k})' := \hat{M}_{xx,k}^{-1}\hat{M}_{xu,k}$ (where $\hat{M}_{xx,k} := T^{-1}\sum_t x_{t,k}x_{t,k}'$, $\hat{M}_{xu,k} := T^{-1}\sum_t x_{t,k}\Delta\hat{X}_t$, with $x_{t,k} := (\Delta X_{t-1},\ldots,\Delta X_{t-k})' = (u_{t-1},\ldots,u_{t-k})'$) and the estimator based on the residuals $\{\hat{X}_t\}$ from the local GLS projection of X_t on X_{t-1}, $\hat{\beta}_k := \hat{M}_{\hat{x}\hat{x},k}^{-1}\hat{M}_{\hat{x}\hat{u},k}$ (where $\hat{M}_{\hat{x}\hat{x},k} := T^{-1}\sum_t \hat{x}_{t,k}\hat{x}_{t,k}'$, $\hat{M}_{\hat{x}\hat{u},k} := T^{-1}\sum_t \hat{x}_{t,k}\Delta\hat{X}_t$, with $\hat{x}_{t,k} := (\Delta\hat{X}_{t-1},\ldots,\Delta\hat{X}_{t-k})'$), satisfies $k^{1/2}\|\hat{\beta}_k - \breve{\beta}_k\| = o_p(1)$ under Assumption $\mathcal{K}$, where $\|\cdot\|$ denotes the (matrix) euclidean norm. (That is for a matrix A, $\|A\| := \sup_{\|x\|<1}\|Ax\|$, using the euclidean norm for the vector x, i.e., $\|x\|^2 := x'x$.) As $\|\hat{M}_{xu,k}\|$ and $\|\hat{M}_{\hat{x}\hat{x},k}^{-1}\|$ are bounded in probability, as shown in Cavaliere and Taylor (2004), this result follows from the inequality $\|\hat{\beta}_k - \breve{\beta}_k\| \le \|\hat{M}_{\hat{x}\hat{x},k}^{-1} - \hat{M}_{xx,k}^{-1}\|\,\|\hat{M}_{xu,k}\| + \|\hat{M}_{\hat{x}\hat{x},k}^{-1}\|\,\|\hat{M}_{\hat{x}\hat{u},k} - \hat{M}_{xu,k}\|$, if we can prove the sufficient conditions: (a) $\|\hat{M}_{\hat{x}\hat{x},k}^{-1} - \hat{M}_{xx,k}^{-1}\| = o_p(kT^{-1/2})$ and (b) $\|\hat{M}_{\hat{x}\hat{u},k} - \hat{M}_{xu,k}\| = O_p(k^{1/2}T^{-1/2})$. To see that (a) holds, as in Berk (1974, p. 493), set $q := \|\hat{M}_{\hat{x}\hat{x},k}^{-1} - \hat{M}_{xx,k}^{-1}\|$, $Q := \|\hat{M}_{\hat{x}\hat{x},k} - \hat{M}_{xx,k}\|$, $p := \|\hat{M}_{xx}^{-1}\|$ so that $q := \|\hat{M}_{\hat{x}\hat{x},k}^{-1} - \hat{M}_{xx,k}^{-1}\| \le \|\hat{M}_{\hat{x}\hat{x},k}^{-1}\|\,\|\hat{M}_{\hat{x}\hat{x},k} - \hat{M}_{xx,k}\|\,\|\hat{M}_{xx,k}^{-1}\| \le (p+q)Qp$, which implies $q \le p^2Q/(1-pQ)$ for $pQ < 1$. As $p = \|\hat{M}_{xx}^{-1}\| = O_p(1)$, $q \to 0$ in probability if $Q = o_p(1)$. To prove the last condition, note that $Q^2 \le k^2\max_{i,j}\sum T^{-1}(-\hat{\gamma}'\Delta d_{t-i}u_{t-j} - \hat{\gamma}'\Delta d_{t-j}u_{t-i} + \hat{\gamma}'\Delta d_{t-j}\Delta d_{t-i}'\hat{\gamma})$; under local GLS demeaning, $\Delta d_t = 0$, and hence $Q = 0$, as expected; under local GLS linear detrending, $\Delta d_t'\hat{\gamma} = O_p(T^{-3/2})$ (see the proof of Theorem 2, which follows); furthermore, using $\sup_t|u_t| = o_p(T^{1/2})$, we easily obtain $Q^2 = o_p(k^2/T) \to 0$ under $\mathcal{K}$. For (b), similarly to (a), $\|\hat{M}_{\hat{x}\hat{u},k} - \hat{M}_{xu,k}\|^2 \le k\max_i\sum T^{-1}(-\hat{\gamma}'\Delta d_{t-i}u_t - \hat{\gamma}'\Delta d_t u_{t-i} + \hat{\gamma}'\Delta d_{t-i}\Delta d_t'\hat{\gamma}) = o_p(k/T)$.

Having established that $k^{1/2}\|\hat{\beta}_k - \breve{\beta}_k\| = o_p(1)$, it is straightforward to see that the residual-based estimators $T^{-1}\sum_{t=1}^{\lfloor sT\rfloor}\hat{u}_{t,k}^2$ and $T^{-1}\sum_{t=1}^{\lfloor sT\rfloor}\breve{u}_{t,k}^2$ converge in probability to the same limiting functional. By Lemma A.10 of Hansen (2000a) the convergence is also uniform over all $s \in [0,1]$, because $T^{-1}\sum_{t=1}^{\lfloor sT\rfloor}\hat{u}_{t,k}^2$ is monotonically increasing in s and the limit function is continuous (see also Hansen, 2000b, p. 112). ■

LEMMA 3. *Let the conditions of Theorem 3 hold. Then* $(1/T)\sum_{t=1}^{\lfloor sT\rfloor}\hat{u}_{t,k}^2 \xrightarrow{p} \int_0^s \omega(r)^2\,dr$ *uniformly for all* $s \in [0,1]$.

Proof of Lemma 3. Under the conditions of Theorem 3, the process may be written as $X_t = u_t^* := (1-\alpha L)^{-1}u_t = C^*(L)\varepsilon_t = \sum_{j=0}^{\infty}c_j^*\varepsilon_{t-j}$, $c_j^* := \sum_{i=0}^{j}c_i\alpha^{j-i}$; as $|\alpha| < 1$, it is straightforward to see that $\sum_{j=0}^{\infty}j|c_j| < \infty$ also implies the summability condition $\sum_{j=0}^{\infty}j|c_j^*| < \infty$. Moreover, the process admits the AR(∞) representation

$$X_t = \sum_{i=1}^{\infty}\beta_{i,\infty}^*X_{t-i} + \varepsilon_t, \tag{A.1}$$

which can be approximated through the $(k+1)$th order autoregression (for the homoskedastic case, see Chang and Park, 2002):

$$X_t = \sum_{i=1}^{k+1}\beta_{i,\infty}^*X_{t-i} + u_{t,k} \tag{A.2}$$

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

with $(\beta^*_{1,\infty},\ldots,\beta^*_{k+1,\infty})' =: \beta^*_k$. Assume, for the present, that the auxiliary ADF regression (10) is estimated using X_t instead of the local GLS detrended series $\hat{X}_t$. As in Lopez (1997), the residuals from (10) coincide with the OLS residuals from (A.2), say,

$$X_t = \sum_{i=1}^{k+1} \hat{\beta}^*_{i,k} X_{t-i} + \hat{u}^*_{t,k}. \tag{A.3}$$

Hence, using the results in Cavaliere and Taylor (2007, proof of Thm. 1(ii)), it is straightforward to obtain that under nonstationary heteroskedasticity $k^{1/2}\|\hat{\beta}^*_k - \beta^*_k\| = o_p(1)$; cf. Chang and Park (2002) for the homoskedastic case.

Let us define the following functionals on $[0,1]$: $\hat{M}^*_{xx,k}(s) := T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor} x^*_{t,k}x^{*\prime}_{t,k}$, $\hat{M}^*_{xu,k}(s) := T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor} x^*_{t,k}u^*_{t,k}$, where $x^*_{t,k} := (X_{t-1},\ldots X_{t-k-1})'$. Now, observe that $T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor}\hat{u}^2_{t,k} = T^{-1}\sum_{t=k+1}^{T}(u^*_t - \hat{\beta}^{*\prime}_k x^*_{t,k})^2$ and consider the equality

$$\frac{1}{T}\sum_{t=k+1}^{\lfloor sT\rfloor}\hat{u}^2_{t,k} - \frac{1}{T}\sum_{t=k+1}^{\lfloor sT\rfloor}u^{*2}_{t,k} = (\hat{\beta}^*_k - \beta^*_k)'\hat{M}^*_{xx,k}(s)(\hat{\beta}^*_k - \beta^*_k) - 2(\hat{\beta}^*_k - \beta^*_k)'\hat{M}^*_{ux,k}(s)$$

$$=: g_1(s) + g_2(s).$$

We now show that $g_1(s)$ and $g_2(s)$ are both of $o_p(1)$. Regarding $g_1(s)$, consider the inequalities

$$\sup_{s\in[0,1]}\|g_1(s)\| \le \|\hat{\beta}^*_k - \beta^*_k\|^2 \sup_{s\in[0,1]}\|\hat{M}^*_{xx,k}(s)\| = \|\hat{\beta}^*_k - \beta^*_k\|^2\|\hat{M}^*_{xx,k}\|.$$

As demonstrated previously, $\|\hat{\beta}^*_k - \beta^*_k\| = o_p(k^{-1/2})$. Moreover, because $\|\hat{M}^*_{xx,k}\| \le \|\hat{M}^*_{xx,k} - M^*_{xx,k}\| + \|M^*_{xx,k}\|$, with $M^*_{xx,k} = \omega^2_{T,k}(1)F^*_k$, $\omega^2_{T,k}(1)$ as defined in Lemma 1, and $F_k = (f^*_{|i-j|})_{i,j=1,\ldots,k}$, $f^*_j := \sum_{i=0}^{\infty} c^*_i c^*_{i+j}$, we obtain $\|\hat{M}^*_{xx,k}\| = O_p(1)$ as $\|\hat{M}^*_{xx,k} - M^*_{xx,k}\| = O_p(k/T^{1/2})$ (see Cavaliere and Taylor, 2007) and $\|M^*_{xx,k}\| = \omega^2_{T,k}(1)\|F^*_k\| = O(1)$ (the last equality following from $\omega^2_{T,k}(1)$ being bounded and from $\|F^*_k\|$ being of $O(1)$ under the stated assumptions; cf. Berk, 1974). Hence, $\|g_1(s)\| = o_p(k^{-1})$ uniformly for all s in $[0,1]$.

Results for $g_2(s)$ follow similarly. Specifically, because

$$\sup_{s\in[0,1]}\|g_2(s)\| \le 2\|\hat{\beta}^*_k - \beta^*_k\| \sup_{s\in[0,1]}\|\hat{M}^*_{ux,k}(s)\| = 2\|\hat{\beta}^*_k - \beta^*_k\|\,\|\hat{M}^*_{ux,k}\|,$$

where $\|\hat{\beta}^*_k - \beta^*_k\| = o_p(k^{-1/2})$ and $\|\hat{M}^*_{ux,k}\| = o_p(k^{-1/2})$ (see Cavaliere and Taylor, 2007), the desired result immediately follows.

Because $\sup_{s\in[0,1]}|T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor}\hat{u}^2_{t,k} - T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor}u^{*2}_{t,k}| = o_p(1)$, to prove the lemma it is left to be shown that $T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor}u^{*2}_{t,k} \xrightarrow{p} \int_0^s \omega(r)^2\,dr$, uniformly. As $T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor}u^{*2}_{t,k} = T^{-1}\sum_{t=k+1}^{\lfloor sT\rfloor}\varepsilon^2_t + o_p(1)$ (cf. Chang and Park, 2002), pointwise convergence follows straightforwardly using Lemmas 1 and 2 in Cavaliere and Taylor (2007). The proof that the convergence is also uniform then follows using the same arguments as were employed in the proof of Lemma 2.

Finally, tedious but straightforward algebra allows us to show that the preceding results also hold if $\hat{\beta}_k$ is based on the residuals of the local GLS projection of X_t on $\mathbf{z}_t$. ■

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

Proof of Theorem 1. The theorem can be proved by extending the proof of Theorem 1 in Cavaliere and Taylor (2007) to the case of local GLS detrended data. This extension is straightforward and is therefore omitted for brevity. ■

Proof of Theorem 2. Recall from Section 4 that the bootstrap time series X_t^b satisfies $X_{\lfloor sT \rfloor}^b = \sum_{t=1}^{\lfloor sT \rfloor} w_t \tilde{u}_t = \sum_{t=1}^{\lfloor sT \rfloor} w_t \hat{u}_{t,k}$, $s \in [0,1]$. Because $\{w_t\}$ is Gaussian, then, conditionally on $\hat{u}_{1,k}, \ldots, \hat{u}_{T,k}$, $T^{-1/2} X_{\lfloor sT \rfloor}^b$ is Gaussian with covariance kernel $\widetilde{M}_T(s,s') = M_T(s \wedge s')$, where $M_T(s) := T^{-1} \sum_{t=1}^{\lfloor sT \rfloor} \hat{u}_{t,k}^2$. Using Lemma 2, we have the result that $M_T(s) \xrightarrow{p} \int_0^s \omega(r)^2 \, dr = \bar{\omega}^2 \eta(s)$ uniformly for all s. Furthermore, $\bar{\omega}^2 \eta(s)$ is the kernel function of the variance-transformed Brownian motion $\bar{\omega} B_\eta(s) = \bar{\omega} B(\eta(s))$, which implies the weak convergence in probability result

$$\frac{1}{T^{1/2}} X_{\lfloor \cdot T \rfloor}^b \xrightarrow{w}_p \bar{\omega}^2 B_\eta(\cdot). \tag{A.4}$$

Note that this convergence holds irrespective of the value taken by the local-to-unity noncentrality parameter, c. Standard continuous mapping arguments then apply to the convergence (A.4), implying, e.g., the following results:

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} X_{t-1}^b \xrightarrow{w}_p \bar{\omega} \int_0^1 B_\eta(s) \, ds, \tag{A.5}$$

$$\frac{1}{T^2} \sum_{t=1}^{T} (X_{t-1}^b)^2 \xrightarrow{w}_p \bar{\omega}^2 \int_0^1 B_\eta^2(s) \, ds, \tag{A.6}$$

$$\frac{1}{T^{5/2}} \sum_{t=1}^{T} (t-1) X_{t-1}^b \xrightarrow{w}_p \bar{\omega} \int_0^1 s B_\eta(s) \, ds, \tag{A.7}$$

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} (t-1) u_t^b = \frac{1}{T^{1/2}} X_T^b - \frac{1}{T^{3/2}} \sum_{t=1}^{T} X_t^b \xrightarrow{w}_p \bar{\omega} \left(B_\eta(1) - \int_0^1 B_\eta(s) \, ds \right) \tag{A.8}$$

jointly. We now establish a weak convergence in probability result for the local GLS detrended series $\hat{X}_t^b$. Let

$$\begin{aligned} \mathbf{X}_{\bar{c}}^b &\equiv [X_0^b, X_1^b - (1 - \bar{c}/T) X_0^b, \ldots, X_T^b - (1 - \bar{c}/T) X_{T-1}^b]' \\ &= [0, X_1^b, \ldots, X_T^b - (1 - \bar{c}/T) X_{T-1}^b]'. \end{aligned} \tag{A.9}$$

Suppose $p = 0$ first. In this case, $\mathbf{z}_{\bar{c}}' \mathbf{z}_{\bar{c}} \to 1$, $\mathbf{z}_{\bar{c}}' \mathbf{X}_{\bar{c}}^b = (\bar{c}/T) X_T^b + (\bar{c}^2/T^2) \sum_{t=1}^{T} X_{t-1}^b = o_p(1)$ by (A.5) and (A.6). Hence, $\hat{\gamma}^b = (\mathbf{z}_{\bar{c}}' \mathbf{z}_{\bar{c}})^{-1} \mathbf{z}_{\bar{c}}' \mathbf{X}_{\bar{c}}^b = o_p(1)$ and

$$\frac{1}{T^{1/2}} \hat{X}_{\lfloor sT \rfloor}^b = \frac{1}{T^{1/2}} (X_{\lfloor sT \rfloor}^b - \hat{\gamma}) = \frac{1}{T^{1/2}} X_{\lfloor sT \rfloor}^b + o_p(1) \xrightarrow{w}_p \bar{\omega} B_\eta(s).$$

Consider next the case $p = 1$, i.e., $z_t = (1, t)'$, and note that $\hat{X}_t^b = X_t^b - (\delta_T z_t)' \delta_T^{-1} \hat{\gamma}^b$, where $\delta_T := \mathrm{diag}(1, T)$. Obviously, $\delta_T \hat{\gamma}^b = (\delta_T^{-1} \mathbf{z}_{\bar{c}}' \mathbf{z}_{\bar{c}} \delta_T^{-1})^{-1} \delta_T^{-1} \mathbf{z}_{\bar{c}}' \mathbf{X}_{\bar{c}}$. As in Elliott et al. (1996), $(\delta_T^{-1} \mathbf{z}_{\bar{c}}' \mathbf{z}_{\bar{c}} \delta_T^{-1}) \to \mathrm{diag}(1, 1 + \bar{c} + \bar{c}^2/3)$, and, after some algebra based on (A.4) and (A.8), one obtains that

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

$$\frac{1}{T^{1/2}}\,\delta_T^{-1}\mathbf{z}_{\bar{c}}'\mathbf{X}_{\bar{c}} \xrightarrow{w}_p \bar{\omega}\begin{bmatrix} 0 \\ (1+\bar{c})B_\eta(1) + \bar{c}^2\int_0^1 sB_\eta(s)\,ds \end{bmatrix}.$$

The latter result implies that $T^{-1/2}z_{\lfloor sT\rfloor}'\hat{\gamma}^b = (z_{\lfloor sT\rfloor}'\delta_T^{-1})T^{-1/2}\delta_T\hat{\gamma}^b \xrightarrow{w}_p \bar{\omega}(\lambda_{\bar{c}}B_\eta(1) + 3(1-\lambda_{\bar{c}})\int_0^1 rB_\eta(r)\,dr)s$, from which it follows that $T^{-1/2}\hat{X}^b_{\lfloor sT\rfloor} \xrightarrow{w}_p \bar{\omega}(B_\eta(s) - s(\lambda_{\bar{c}}B_\eta(1) + 3(1-\lambda_{\bar{c}})\int_0^1 rB_\eta(r)\,dr)) = \bar{\omega}B^1_{\eta,0}(s,\bar{c})$. This establishes that the local GLS residuals, $\hat{X}_t^b$, obtained from the regression of X_t^b on z_t, satisfy the weak convergence in probability result, $T^{-1/2}\hat{X}^b_{\lfloor\cdot T\rfloor} \xrightarrow{w}_p \bar{\omega}B^p_{\eta,0}(\cdot,\bar{c})$. Consequently, from a standard continuous mapping argument,

$$\left(T^{-1}(\hat{X}_T^b)^2, T^{-2}\sum_{t=1}^T(\hat{X}^b_{t-1})^2\right) \xrightarrow{w}_p \bar{\omega}^2\left(B^p_{\eta,0}(1,\bar{c})^2, \int_0^1 B^p_{\eta,0}(s,\bar{c})^2\,ds\right). \tag{A.10}$$

Now consider the bootstrap residual variance estimator, $(\hat{\sigma}_0^b)^2$, which can be written as

$$(\hat{\sigma}_0^b)^2 = T^{-1}\sum_{t=1}^T(\Delta\hat{X}^b_{t-1})^2 - T^{-1}(\hat{\alpha}_b - 1)^2\sum_{t=1}^T(\hat{X}^b_{t-1})^2, \tag{A.11}$$

where $\hat{\alpha}_b := \sum_{t=1}^T\hat{X}^b_{t-1}\hat{X}^b_t/\sum_{t=1}^T(\hat{X}^b_{t-1})^2$. Consider first $T^{-1}\sum_{t=1}^T(\Delta\hat{X}^b_{t-1})^2$. It is straightforward to see that for $p=0$, $T^{-1}\sum_{t=1}^T(\Delta\hat{X}^b_{t-1})^2 = T^{-1}\sum_{t=1}^T(u_t^b)^2$, whereas, for $p=1$, we may write $T^{-1}\sum_{t=1}^T(\Delta\hat{X}^b_{t-1})^2 = T^{-1}\sum_{t=1}^T(u_t^b)^2 + T^{-1}\hat{\gamma}_b'\sum_{t=1}^T\Delta z_t\Delta z_t'\hat{\gamma}_b - 2(T^{-1}\sum_{t=1}^T u_t^b)([0,1]\hat{\gamma}_b)$. In the latter case, because $\delta_T\hat{\gamma}_b$ is of $O_p(T^{1/2})$ and $T^{-1/2}\sum_{t=1}^T u_t^b \xrightarrow{w}_p \bar{\omega}B_\eta(1)$ (see the preceding discussion), $T^{-1}\hat{\gamma}_b'\sum_{t=1}^T\Delta z_t\Delta z_t'\hat{\gamma}_b$ and $(T^{-1}\sum_{t=1}^T u_t^b)([0,1]\hat{\gamma}_b)$ are both of $o_p(1)$. Hence, provided that

$$T^{-1}\sum_{t=1}^T(u_t^b)^2 \xrightarrow{w}_p \bar{\omega}^2 \tag{A.12}$$

we immediately have the result that

$$T^{-1}\sum_{t=1}^T(\Delta\hat{X}^b_{t-1})^2 \xrightarrow{w}_p \bar{\omega}^2, \tag{A.13}$$

as required. To establish the validity of (A.12) consider the following equality:

$$\frac{1}{T}\sum_{t=1}^T(u_t^b)^2 = \frac{1}{T}\sum_{t=1}^T\hat{u}_{t,k}^2w_t^2 = \frac{1}{T}\sum_{t=1}^T\hat{u}_{t,k}^2 + \frac{1}{T}\sum_{t=1}^T\hat{u}_t^2\xi_t =: f_1 + f_2$$

with $\xi_t := w_t^2 - 1$ an independent sequence of centered $\chi^2(1)$ random variables and f_1, f_2 implicitly defined. By Lemma 2, $f_1 \xrightarrow{p} \bar{\omega}^2$. Moreover, conditionally on the sample,

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

$$E\left(\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{t,k}^2\xi_t\right)^2 = \frac{1}{T^2}E\left(\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{u}_{t,k}^2\hat{u}_{s,k}^2\xi_t\xi_s\right)^2 = \frac{1}{T^2}\left(\sum_{t=1}^{T}\hat{u}_{t,k}^4E(\xi_t^2)\right)^2$$

$$= \frac{4}{T}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{u}_{t,k}^4\right)^2 = O_p(T^{-1})$$

because $T^{-1}\sum_{t=1}^{T}\hat{u}_{t,k}^4$ is of $O_p(1)$ under $E(e_t^4) < \infty$.

Now, consider the random variable $T(\hat{\alpha}_b - 1)$, which can be written (see, e.g., Phillips, 1987; Stock, 1994) as $T^{-1}((\hat{X}_T^b)^2 - (\hat{X}_0^b)^2 - \sum_t(\Delta\hat{X}_t^b)^2)/(2T^{-2}\sum_t(\hat{X}_{t-1}^b)^2)$. By using (A.10), (A.13), and a continuous mapping argument we obtain that

$$T(\hat{\alpha}_b - 1) \xrightarrow[p]{w} \frac{B_{\eta,0}^p(1,\bar{c})^2 - 1}{2\int_0^1 B_{\eta,0}^p(s,\bar{c})^2\,ds},$$

which, taken together with (A.6), yields the result that $T^{-1}(\hat{\alpha}_b - 1)^2\sum_{t=1}^{T}(\hat{X}_{t-1}^b)^2 = o_p(1)$. Hence, from (A.11) we have that $(\hat{\sigma}_0^b)^2 \xrightarrow[p]{w} \bar{\omega}^2$, which, taken together with (A.10), yields that

$$\mathcal{MZ}_\alpha^b \xrightarrow[p]{w} \frac{B_{\eta,0}^p(1,\bar{c})^2 - 1}{2\int_0^1 B_{\eta,0}^p(s,\bar{c})^2\,ds} = \tilde{\xi}_1^{0,\bar{c},p,\eta}.$$

This result also implies that, uniformly in probability, $G_T^{\mathcal{MZ}_\alpha^b}(\cdot) \to G^{\tilde{\xi}_1^{0,\bar{c},p,\eta}}(\cdot)$, with $G^{\tilde{\xi}_i^{0,\bar{c},p,\eta}}(\cdot)$ denoting the c.d.f. of $\tilde{\xi}_i^{0,\bar{c},p,\eta}$, $i = 1,\ldots,3$. This implies that under the null hypothesis $p_{T,\mathcal{MZ}_\alpha}^b$ converges weakly to $U[0,1]$.

The remainder of the proof is identical to the proof of Theorem 5 in Hansen (2000b). The corresponding results for $\mathcal{MZ}_t^b$ and $\mathcal{MSB}^b$ follow similarly. ■

Proof of Theorem 3. We first need to establish that the bootstrap statistics have the same asymptotic distribution as under the null hypothesis. This result will also imply that, uniformly in probability, the c.d.f.s of $\mathcal{MZ}_\alpha^b$, $\mathcal{MSB}^b$, and $\mathcal{MZ}_t^b$, say, $G_T^{\mathcal{MZ}_\alpha^b}$, $G_T^{\mathcal{MSB}^b}$, $G_T^{\mathcal{MZ}_t^b}$, satisfy $G_T^{\mathcal{MZ}_\alpha^b}(\cdot) \to G^{\tilde{\xi}_1^{0,\bar{c},p,\eta}}(\cdot)$, $G_T^{\mathcal{MSB}^b}(\cdot) \to G^{\tilde{\xi}_2^{0,\bar{c},p,\eta}}(\cdot)$, $G_T^{\mathcal{MZ}_t^b}(\cdot) \to G^{\tilde{\xi}_3^{0,\bar{c},p,\eta}}(\cdot)$. This result can be shown by following exactly the same steps as in the proof of Theorem 2, with the only difference that now we need to call Lemma 3, which is designed for fixed alternatives, instead of Lemma 2, which is valid under the null hypothesis and local alternatives.

It is left to prove that, under the conditions of the theorem, $\mathcal{MZ}_\alpha$ and $\mathcal{MZ}_t$ diverge to negative infinity whereas $\mathcal{MSB}$ converges to zero, in each case as the sample size diverges. By adapting Stock (1999) to the present framework, a sufficient condition for $\mathcal{MZ}_\alpha$, $\mathcal{MZ}_t \to -\infty$ and $\mathcal{MSB} \to 0$ is that $Ts_{AR}^2(k)$ diverges as $T \to \infty$. We now prove that this requirement holds by showing that $1/(Ts_{AR}^2(k))$ converges to zero in probability. As in the proof of Lemma 2 we clarify exposition by assuming, without loss of generality, that the auxiliary ADF regression, (10), is estimated using X_t instead of the local GLS detrended series $\hat{X}_t$.

First, notice that the OLS estimates from (10) are related to those from the AR$(k+1)$ regression (A.3) via $\hat{\pi} = \sum_{i=1}^{k+1}\hat{\beta}_{i,k}^* - 1$ and $\hat{\beta}_{i,k} = -\sum_{j=i+1}^{k+1}\hat{\beta}_{j,k}^*$, $i = 1,\ldots,k$. Conse-

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press

quently, because $\sum_{i=1}^{k}\hat{\beta}_{i,k} = -\sum_{i=1}^{k}\sum_{j=i+1}^{k+1}\hat{\beta}^{*}_{j,k} = \sum_{i=2}^{k+1}\hat{\beta}^{*}_{i,k}(i-1)$, $s^2_{AR}(k)$ may be written as

$$s^2_{AR}(k) = \frac{\hat{\sigma}^2}{\left(1-\sum_{i=1}^{k}\hat{\beta}_{i,k}\right)^2} = \frac{\hat{\sigma}^2}{\left(1-\sum_{i=2}^{k+1}\hat{\beta}^{*}_{i,k}(i-1)\right)^2} = \frac{\hat{\sigma}^2}{(1-c_k'\hat{\beta}^{*}_{k})^2}$$

with $c_k := (1,2,\ldots,k)'$, and, hence,

$$\frac{1}{Ts^2_{AR}(k)} = \frac{(1-c_k'(\hat{\beta}^{*}_{k}-\beta^{*}_{k})+c_k'\beta^{*}_{k})^2}{T\hat{\sigma}^2}. \tag{A.14}$$

Using the results in the proof of Lemma 2, and because $\|c_k\|^2 = \frac{1}{6}k(2k^2+3k+1)$, we have $\|c_k'(\hat{\beta}^{*}_{k}-\beta^{*}_{k})\| \le \|c_k\|\,\|\hat{\beta}^{*}_{k}-\beta^{*}_{k}\| = O(k^{3/2})o_p(k^{-1/2}) = o_p(k)$, which implies that $T^{-1}(c_k'(\hat{\beta}^{*}_{k}-\beta^{*}_{k}))^2 = o_p(1)$. It is also easy to verify that $c_k'\beta^{*}_{k}$ is bounded; specifically, $\mathcal{A}_1$ implies that the coefficients of autoregressive representation (A.1) satisfy $\sum_{i=1}^{\infty}|i|\beta^{*}_{i,\infty} < \infty$ (see Chang and Park, 2002, p. 434, and references therein). This condition suffices for $|c_k'\beta^{*}_{k}| = |\sum_{i=2}^{k+1}\beta^{*}_{i,\infty}(i-1)|$ to be bounded as k diverges. Hence, the numerator in (A.14) is bounded in probability. Finally, because from the proof of Lemma 3 we know that $\hat{\sigma}^2 = (T-k)^{-1}\sum_{t=k+1}^{T}\hat{u}^2_{t,k} \xrightarrow{p} \bar{\omega}^2$, we obtain that $1/(Ts^2_{AR}(k)) \xrightarrow{p} 0$.

It then follows immediately that $p^{b}_{T,\mathcal{MZ}_\alpha} = G^{\xi_1^{0,\bar{c},p,\eta}}(\mathcal{MZ}_\alpha) + o_p(1) \xrightarrow{p} 0$, $p^{b}_{T,\mathcal{MSB}} = G^{\xi_2^{0,\bar{c},p,\eta}(\mathcal{MSB})} + o_p(1) \xrightarrow{p} 0$, and $p^{b}_{T,\mathcal{MZ}_t} = G^{\xi_3^{0,\bar{c},p,\eta}}(\mathcal{MZ}_t) + o_p(1) \xrightarrow{p} 0$, which completes the proof of the theorem. ■

https://doi.org/10.1017/S0266466608080043 Published online by Cambridge University Press