

Adaptive wild bootstrap tests for a unit root with non-stationary volatility

H. PETER BOSWIJK[†] AND YANG ZU[‡]

[†]*Amsterdam School of Economics and Tinbergen Institute, University of Amsterdam,
PO Box 15867, 1001 NJ Amsterdam, the Netherlands.*

E-mail: h.p.boswijk@uva.nl

[‡]*School of Economics, University of Nottingham, Nottingham, NG7 2RD, UK.*

E-mail: yang.zu@nottingham.ac.uk

First version received: August 2016; final version accepted: July 2017

Summary Recent research has emphasized that permanent changes in the innovation variance (caused by structural shifts or an integrated volatility process) lead to size distortions in conventional unit root tests. It has been shown how these size distortions can be resolved using the wild bootstrap. In this paper, we first derive the asymptotic power envelope for the unit root testing problem when the non-stationary volatility process is known. Next, we show that under suitable conditions, adaptation with respect to the volatility process is possible, in the sense that non-parametric estimation of the volatility process leads to the same asymptotic power envelope. Implementation of the resulting test involves cross-validation and the wild bootstrap. A Monte Carlo experiment shows that the asymptotic results are reflected in finite sample properties, and an empirical analysis of real exchange rates illustrates the applicability of the proposed procedures.

Keywords: *Adaptive testing, Non-parametric estimation, Power envelope, Unit root, Wild bootstrap.*

1. INTRODUCTION

Over the past decade, a large amount of research has been devoted to the effect of heteroscedasticity on unit root tests. When the heteroscedasticity follows a stationary GARCH-type specification, such that the unconditional variance is well-defined and constant, then the invariance principle guarantees that the usual Dickey–Fuller (DF) tests remain valid asymptotically. This was illustrated using Monte Carlo simulations by Kim and Schmidt (1993). However, subsequent research has indicated that in such cases more powerful tests for a unit root may be obtained from a likelihood analysis of a model with GARCH innovations; see Seo (1999) and Ling et al. (2003), based on Ling and Li (1998), *inter alia*.

In empirical applications, the assumption that the variation in volatility effectively averages out over the relevant sample is often questionable. On the one hand, in applications involving daily financial prices (interest rates, exchange rates), the degree of mean reversion in the volatility is usually so weak that the volatility process shows persistent deviations from its mean over the relevant time-span (often ten years or less). On the other hand, in applications involving

macro-economic time series observed at a lower frequency but over a longer time-span, one often finds level shifts in the volatility, instead of volatility clustering. Intermediate cases (slowly mean-reverting volatility with changing means) may also occur.

In the presence of such persistent variation in volatility, the invariance principle cannot be expected to apply, such that the null distribution of unit root tests will be affected. The resulting size distortions have been investigated by Boswijk (2001) for the case of a near-integrated GARCH process, and by Kim et al. (2002) and Cavaliere (2004) for the case of a deterministic volatility function. Cavaliere and Taylor (2008) develop a wild bootstrap version of the standard DF tests, and show that this leads to tests with a correct asymptotic size. Cavaliere and Taylor (2007) and Beare (2016) provide two alternative solutions, in the form of non-parametric corrections that lead to statistics with the usual asymptotic null distributions.

There is no guarantee that these approaches to deliver tests with correct asymptotic size will also yield tests with the highest possible power. In particular, in the presence of heteroscedasticity we can expect higher power from a method that gives the highest weight to observations with the lowest volatility, and this is not the case for the tests discussed above.¹

In this paper, we address this issue by deriving the asymptotic power envelope; that is, the maximum possible power against a sequence of local alternatives to the unit root, for a given and known realization of the volatility process. This allows us to evaluate the power loss of various tests, and to construct a class of admissible tests, that have a point of tangency with the envelope. For the empirically more relevant case where the volatility function is not observed, we show that under suitable conditions, adaptation with respect to the volatility process is possible, in the sense that non-parametric estimation of the volatility process leads to the same asymptotic power envelope. Similar adaptivity results were obtained for stable (auto-)regressions by Hansen (1995), Xu and Phillips (2008) and Patilea and Raïssi (2012). The test statistics that come out of this analysis have an asymptotic null distribution that depends on the realization of the volatility process. Therefore, we cannot construct tables with critical values, but the null distribution and hence p -value may be recovered either by Monte Carlo simulation of the limiting distribution with estimated volatility process, or by using the wild bootstrap, analogously to Cavaliere and Taylor (2008).

The plan of the paper is as follows. In Section 2, we present the model, and we obtain some preliminary asymptotic results. In Section 3, we characterize the power envelope (conditional on the volatility process) and we illustrate the power gain possibilities in four examples. In Section 4, we discuss non-parametric estimation of the volatility process, and its use in the construction of a class of adaptive tests; we also discuss various bootstrap implementations of the tests. In Section 5, we extend the test to allow for deterministic components and short-run dynamics. The finite-sample behaviour of these tests is investigated in a Monte Carlo experiment in Section 6; simulation results are reported in the online Appendix. In Section 7, we discuss an empirical application, and in Section 8 we provide some concluding remarks. Proofs are given in Appendix A.

Throughout the paper, we use the notation $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{d} X$ to denote convergence in probability and convergence in distribution, respectively, for sequences of random variables or vectors. We let $X_n(u) \xrightarrow{d} X(u)$, $u \in [0, 1]$ denote weak convergence in $D[0, 1]$, the space of right-continuous functions with finite left limits, under the Skorohod metric, and $X_n \xrightarrow{d_p} X$

¹ An exception is Kim et al. (2002), who consider GLS-based testing for a unit root in case of a single break in the volatility.

denotes weak convergence in probability; see Giné and Zinn (1990). The notation $\lfloor x \rfloor$ is used for the largest integer less than or equal to x .

2. THE MODEL AND PRELIMINARY RESULTS

Consider the heteroscedastic first-order autoregressive model

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t, \quad t = 1, \dots, n, \quad (2.1)$$

$$\varepsilon_t = \sigma_t z_t, \quad (2.2)$$

$$E[z_t | \mathcal{F}_{t-1}] = 0, \quad E[z_t^2 | \mathcal{F}_{t-1}] = 1, \quad (2.3)$$

where $X_0 = 0$, and where $\{\mathcal{F}_t\}_{t \geq 1}$ is the filtration generated by $\{\varepsilon_t\}_{t \geq 1}$. Extensions to models with deterministic components and higher-order autoregressions are considered in Section 5. The null hypothesis of interest is the unit root hypothesis $\mathcal{H}_0 : \theta = 0$.

We assume that $\{\sigma_t\}_{t \geq 1}$ is a deterministic sequence, such that $\{\varepsilon_t\}_{t \geq 1}$ is a martingale difference sequence, with conditional (and unconditional) variance $\{\sigma_t^2\}_{t \geq 1}$ and hence volatility $\{\sigma_t\}_{t \geq 1}$. The theory developed here can be extended to allow for an exogenous stochastic volatility process, in which case the results would hold conditionally on this process. Furthermore, the analysis could be extended to the case where $\{z_t\}_{t \geq 1}$ is a stationary GARCH-type process, but this will not be considered explicitly.

If the variation in $\{\sigma_t^2\}_{t \geq 1}$ averages out over subsamples (i.e., if $(un)^{-1} \sum_{t=1}^{\lfloor un \rfloor} \sigma_t^2 \rightarrow \bar{\sigma}^2 > 0$ as $n \rightarrow \infty$, for all $u \in [0, 1]$), then under additional technical conditions, $\{\varepsilon_t\}_{t \geq 1}$ satisfies an invariance principle. This implies that conventional DF tests for a unit root will be asymptotically valid, even though more powerful tests can be obtained by explicitly modelling the volatility process; see, e.g., Seo (1999) and Ling et al. (2003).

In contrast, in this paper we are concerned with cases where the volatility displays permanent shifts or trends. We do not assume a particular parametric specification, but instead require the following.

ASSUMPTION 2.1. *In the model (2.1)–(2.3): (a) defining $\sigma_n(u) = \sigma_{\lfloor un \rfloor + 1}$ for $u \in [0, 1]$ and $\sigma_n(1) = \sigma_n$, as $n \rightarrow \infty$, $\sigma_n(\cdot) \rightarrow \sigma(\cdot)$ in $D[0, 1]$ where $\sigma(\cdot)$ is strictly positive; (b) the sequence $\{z_t\}_{t \geq 1}$ satisfies an invariance principle, i.e., as $n \rightarrow \infty$,*

$$W_n(u) := n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} z_t \xrightarrow{d} W(u), \quad u \in [0, 1], \quad (2.4)$$

where $W(\cdot)$ is a standard Brownian motion.

REMARK 2.1. Assumption 2.1(a) preserves persistent changes in the volatility as $n \rightarrow \infty$. It is closely related to the assumption $\sigma_t = \sigma(t/n)$, considered, *inter alia*, by Cavaliere (2004) and Cavaliere and Taylor (2007, 2008) (note that $\sigma_t = \sigma_n((t-1)/n)$). It implies that σ_t , and hence ε_t and X_t , are in fact triangular arrays $\{(X_{nt}, \varepsilon_{nt}, \sigma_{nt}), t = 1, \dots, n; n = 1, 2, \dots\}$. However, we suppress the double index notation for simplicity. The assumption is also similar in spirit to the analysis of Hansen (1995), who assumes that σ_t^2 is a smooth positive transformation of a near-integrated autoregression, converging to an Ornstein–Uhlenbeck process. Hansen considers

the effect of such volatility specifications on ordinary least-squares (OLS), generalized least-squares (GLS) and adaptive estimation, when the regressor is a linear process with non-stationary volatility. The analysis in this paper can be interpreted as a generalization of these results to the case of a (near-) integrated regressor.

REMARK 2.2. The invariance principle for z_t would follow if the martingale difference assumption is strengthened to an independent and identically distributed (i.i.d.) assumption, or augmented with a (conditional) Lindeberg condition.

The following lemma characterizes the limiting behaviour of the process $\{X_t\}$ under a near-integrated parameter sequence $\mathcal{H}_n : \theta_n = c/n$, with $c \in \mathbb{R}$ a fixed constant.

LEMMA 2.1. Consider the model (2.1)–(2.3) under Assumption 2.1. Under $\mathcal{H}_n : \theta_n = c/n$, and as $n \rightarrow \infty$,

$$n^{-1/2} X_{\lfloor un \rfloor} \xrightarrow{d} X_c(u) = \int_0^u e^{c(u-s)} \sigma(s) dW(s), \quad u \in [0, 1],$$

jointly with (2.4), where $X_c(\cdot)$ satisfies

$$dX_c(u) = cX_c(u)du + \sigma(u)dW(u). \quad (2.5)$$

All proofs are given in the Appendix. The lemma has direct consequences for the asymptotic properties of the conventional DF tests. In particular, let DF_n denote the t -statistic for $\theta = 0$ in the first-order autoregression $\Delta X_t = \theta X_{t-1} + \varepsilon_t$. As shown by Cavaliere (2004), Lemma 2.1 implies, under the null hypothesis $c = 0$,

$$DF_n \xrightarrow{d} \left(\int_0^1 \sigma(u)^2 du \int_0^1 X_0(u)^2 du \right)^{-1/2} \int_0^1 X_0(u) \sigma(u) dW(u). \quad (2.6)$$

The distribution of the expression on the right-hand side of (2.6) does not coincide with the usual DF null distribution, unless $\sigma(u) = \sigma$ (constant), such that $X_0(\cdot) = \sigma W(\cdot)$. Thus, the DF tests are not robust to persistent variation in σ_t , leading to a non-constant $\sigma(\cdot)$. As shown by Cavaliere and Taylor (2008), this problem can be resolved by the use of the so-called wild bootstrap. Alternatively, Cavaliere and Taylor (2007) use the fact that an Itô process such as $X_0(\cdot)$, with deterministic volatility $\sigma(\cdot)$, can be expressed as a time-deformed Brownian motion. This can be used to define a sampling scheme, where X_t is observed at a lower frequency when the volatility is low, and at a higher frequency when $\sigma(u)$ is high. The application of the DF (or Phillips–Perron) test to these skip-sampled observations leads to a statistic with the usual asymptotic null distribution (albeit with a different power function than under homoscedasticity). Yet another approach was developed by Beare (2016), who applies the DF/Phillips–Perron test to the cumulative sum of reweighted increments of X_t , i.e., to $X_t^* = \sum_{i=1}^t \Delta X_i / \hat{\sigma}_i$, where $\hat{\sigma}_i$ is obtained by kernel estimation. This again leads to a test with the same asymptotic null distribution as the DF test under homoscedasticity.

The focus of this paper is not on solving the size distortions caused by non-stationary volatility, but on developing tests with higher power. In the next section, we derive the maximum possible asymptotic power of any test of the unit root null against local alternatives, for the (infeasible) case where σ_t is observed, and z_t is an i.i.d. $N(0, 1)$ sequence. Next, we show that the asymptotic volatility function $\sigma(\cdot)$ is consistently estimable, and this can be used to construct a family of point optimal tests that reach the Gaussian asymptotic power envelope. The resulting

tests are adaptive, in the sense that there is no loss of asymptotic efficiency or power caused by estimating σ_t .

3. POWER ENVELOPE AND INFEASIBLE LIKELIHOOD RATIO TEST

In this section, we derive the Gaussian asymptotic power envelope for the unit root hypothesis in the model (2.1)–(2.3), with $\{\sigma_t\}_{t=1}^n$ known. This power envelope will then be compared to the asymptotic power of the DF test, and of the likelihood ratio (LR) test based on known $\{\sigma_t\}_{t=1}^n$ (which, in practice, when $\{\sigma_t\}_{t=1}^n$ is not observed, will be infeasible).

Under Gaussianity, the log-likelihood is given by

$$\ell_n(\theta) = -\frac{1}{2} \sum_{t=1}^n \left(\log 2\pi \sigma_t^2 + \frac{(\Delta X_t - \theta X_{t-1})^2}{\sigma_t^2} \right).$$

Define the log-likelihood ratio of $\theta_n = c/n$ relative to $\theta = 0$:

$$\Lambda_n(c) := \ell_n(\theta_n) - \ell_n(0) = cS_n - \frac{1}{2}c^2 J_n,$$

where

$$S_n = \frac{1}{n} \sum_{t=1}^n Z_{t-1} \frac{\Delta X_t}{\sigma_t}, \quad J_n = \frac{1}{n^2} \sum_{t=1}^n Z_{t-1}^2,$$

with $Z_{t-1} = X_{t-1}/\sigma_t$.

The envelope is based on the power of the Neyman–Pearson test in a limit experiment that provides an asymptotic approximation of the model in a neighbourhood of the null hypothesis. This limit experiment is locally asymptotically quadratic (LAQ); see, e.g. Jeganathan (1995) and Le Cam and Yang (1990). Because of the Gaussianity assumption, the log-likelihood ratio is a quadratic function. Theorem 3.1 gives its limiting behaviour under the null hypothesis and local alternatives, and characterizes the log-likelihood as locally asymptotically quadratic; see Jeganathan (1995).

THEOREM 3.1. *Consider the model (2.1)–(2.3), under Assumption 2.1. Let*

$$Z_c(u) = \sigma(u)^{-1} X_c(u) = \int_0^u e^{c(u-s)} \frac{\sigma(s)}{\sigma(u)} dW(s).$$

Under $\mathcal{H}_n : \theta_n = c/n$, we have as $n \rightarrow \infty$,

$$\begin{pmatrix} S_n \\ J_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S_c \\ J_c \end{pmatrix} = \begin{pmatrix} \int_0^1 Z_c(u) dW(u) + c \int_0^1 Z_c(u)^2 du \\ \int_0^1 Z_c(u)^2 du \end{pmatrix}, \tag{3.1}$$

and hence, for fixed $\bar{c} \in \mathbb{R}$,

$$\Lambda_n(\bar{c}) = \bar{c}S_n - \frac{1}{2}\bar{c}^2 J_n \xrightarrow{d} \bar{c}S_c - \frac{1}{2}\bar{c}^2 J_c =: \Lambda_c(\bar{c}). \tag{3.2}$$

REMARK 3.1. As usual in a likelihood analysis, the asymptotic distributions and hence power functions derived below will continue to hold when the Gaussian assumption is violated, as long as $\{z_t\}_{t \geq 1}$ satisfies an invariance principle. However, the optimality claims in the results to follow critically depend on its validity: if the actual density z_t differs from the Gaussian density, then more powerful tests can be constructed from a likelihood function derived from the actual density. In an earlier working paper version of this paper, we considered the power envelope for an arbitrary but known density $p(z)$; see Boswijk (2005).

REMARK 3.2. A similar remark applies to the possible presence of conditional heteroscedasticity in z_t , i.e., when $E[z_t^2 | \mathcal{F}_{t-1}] = h_t$, where h_t follows a stationary GARCH specification with $E[h_t] = 1$. Under suitable additional conditions, the asymptotic properties derived below will continue to hold, but more powerful testing procedures can be obtained from a likelihood analysis of the model under a parametric specification for h_t , analogous to Ling and Li (1998). A different situation arises in the case of near-integrated GARCH processes, i.e., when $h_{[un]}$, $u \in [0, 1]$ converges weakly to a stochastic process in $D[0, 1]$; see Boswijk (2001). We do not consider this case explicitly, but we conjecture that the procedures developed below will retain their validity, provided that the limiting volatility process is independent of the Brownian motion W .

REMARK 3.3. Note that in (3.2), c refers to the true data-generating process (the probability measure $P_{\theta_n, n}$ with $\theta_n = c/n$), whereas \bar{c} characterizes a chosen local alternative. Therefore, setting $c = 0$ gives the asymptotic null distribution of the Neyman–Pearson test statistic for $\mathcal{H}_0 : \theta = 0$ against $\mathcal{H}_n : \theta_n = \bar{c}/n$, whereas setting $c = \bar{c}$ gives the asymptotic distribution under local alternatives, and hence can be used to evaluate local power.

REMARK 3.4. An interpretation of Theorem 3.1 is that the model $\mathcal{E}_n = (\mathbb{R}^n, \mathcal{A}, \{P_{\theta, n}\}_{\theta \in \mathbb{R}})$ is locally approximated, for $\theta_n = c/n$, by the limit experiment $\mathcal{G} = (\mathbb{R}^2, \mathcal{B}, \{Q_c\}_{c \in \mathbb{R}})$, where \mathcal{A} and \mathcal{B} are the relevant Borel σ -fields, and where Q_c is the distribution of (S_c, J_c) , with log-likelihood ratio $\Lambda_c(\bar{c}) = \log dQ_{\bar{c}}/dQ_0$. An interpretation of this limit experiment is that we observe $X_c(u)$, $u \in [0, 1]$, generated by (2.5), to make inference on c . The limit experiment is a curved exponential model with one parameter c and two sufficient statistics (S_c, J_c) . Note that the information J_c is not ancillary, since its distribution under Q_c depends on c . This implies that the log-likelihood ratio is not locally asymptotically mixed normal (LAMN), but locally asymptotically Brownian functional (LABF); see Jeganathan (1995).

The power of the Neyman–Pearson test for $c = 0$ against $c = \bar{c}$, which rejects for large values of $\Lambda_n(\bar{c})$, defines the asymptotic power envelope (conditional on $\sigma(\cdot)$) for testing $\mathcal{H}_0 : \theta = 0$ against $\mathcal{H}_n : \theta_n = \bar{c}/n$. We evaluate this power envelope by Monte Carlo simulation, for $-c \in \{0, \dots, 20\}$, and for four different volatility functions, inspired by the simulations in Cavaliere (2004).

1. $\sigma_1(u) = 1_{[0, 0.9)}(u) + 5 \cdot 1_{[0.9, 1]}(u)$; this represents a level shift in the volatility from 1 to 5 at time $t = (9/10)n$ (i.e., late in the sample).
2. $\sigma_2(u) = 1_{[0, 0.1)}(u) + 5 \cdot 1_{[0.1, 1]}(u)$; an early level shift from 1 to 5.
3. $\sigma_3(u) = \exp((1/2)H(u))$, where $dH(u) = -10H(u)du + 10dB(u)$, with $B(\cdot)$ being a standard Brownian motion, independent of $W(\cdot)$; this represents a realization of a stochastic volatility process, with a low degree of mean-reversion and a fairly high volatility-of-volatility.

- $\sigma_4(u) = \exp((1/2)H(u))$, where $H(u) = 5B(u)$; a realization of a stochastic volatility process with no mean-reversion and a lower volatility-of-volatility.

Figure 1 depicts the volatility paths $\sigma_1(\cdot)$ – $\sigma_4(\cdot)$ that we use in our simulations. The two stochastic volatilities σ_3 and σ_4 have been obtained by discretizing the relevant Brownian motions and integrals over 5000 equidistant time points in the unit interval. Because these realizations are kept fixed in repeated draws, they can be thought of as deterministic. All computations have been performed in OX; see Doornik (2013).

The power envelopes are based on Monte Carlo simulations of $\Lambda(\bar{c})$ under Q_c , with $c \in \{0, \bar{c}\}$. The simulations of $\Lambda(\bar{c})$ under Q_0 provide 5% critical values for the test, and the rejection frequencies under $Q_{\bar{c}}$ then indicate the maximum possible power against $c = \bar{c}$. Figure 2 depicts the power envelopes for the four volatility functions, as well as the asymptotic power curves of the one-sided LR test and the DF test. The one-sided LR test rejects for small values of the signed LR test statistic for $\mathcal{H}_0 : \theta = 0$ against $\mathcal{H}_1 : \theta < 0$, given by

$$LR_n = \text{sgn}(\hat{c}_n)\sqrt{2\Lambda_n(\hat{c}_n)} = J_n^{-1/2}S_n,$$

where $\hat{c}_n = \arg \max_c \Lambda_n(c) = J_n^{-1}S_n$, so that the maximum likelihood estimator is $\hat{\theta}_n = \hat{c}_n/n$.

From these figures, we observe that the power of the LR test is close to the envelope, but not equal to it, especially in case of stochastic volatility. Furthermore, in most cases (with the exception of σ_2), the power of the DF test is substantially less than that of the LR test (and hence the envelope). Therefore, reweighting observations indeed has an important effect on the power of unit root tests. It should be emphasized that we have chosen fairly extreme volatility functions; for more realistic volatility paths, the power differences will be smaller.

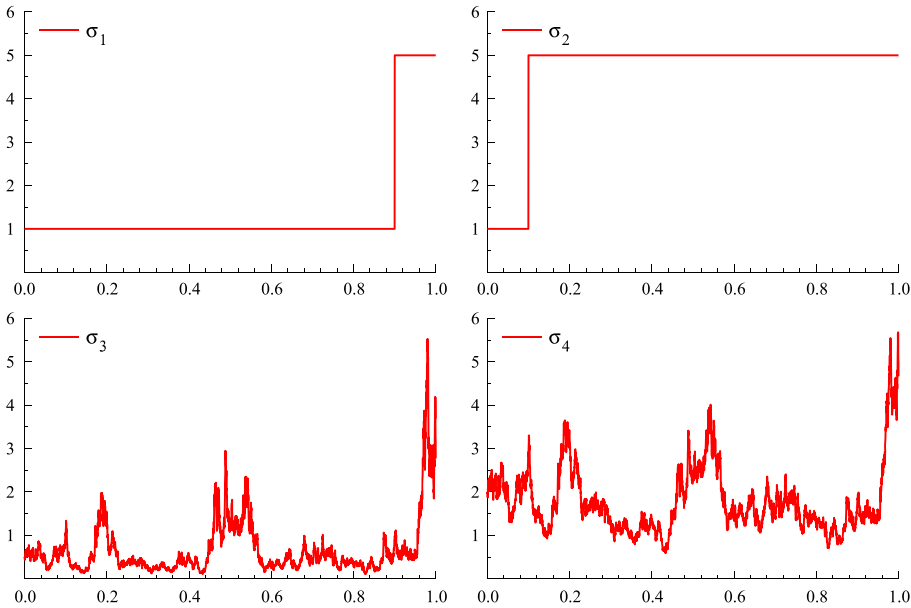


Figure 1. Realization of volatility processes σ_1 – σ_4 . [Colour figure can be viewed at wileyonlinelibrary.com]

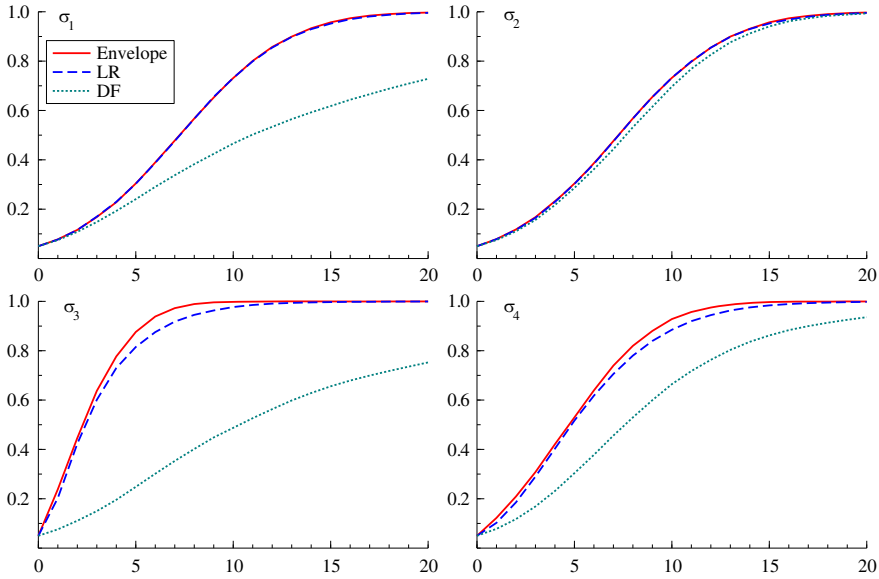


Figure 2. Asymptotic power envelope and power curves for σ_1 – σ_4 . [Colour figure can be viewed at wileyonlinelibrary.com]

4. ADAPTIVE LIKELIHOOD RATIO TEST AND BOOTSTRAP

In the previous section, we have studied the power of procedures that assume that $\{\sigma_t\}_{t=1}^n$ is known and observed. In practice, this is not the case, and σ_t will have to be estimated. One option is to specify a parametric model for σ_t , such as a GARCH model, and then to consider maximum likelihood estimation of that model. However, it is desirable to have a testing procedure that is not too sensitive to deviations from such an assumption, and that will also work well, for example, in the case of (gradual) changes in the level of the volatility.

Therefore, inspired by Hansen (1995), we consider non-parametric estimation of $\{\sigma_t\}_{t=1}^n$. Let $k : [-1, 1] \rightarrow [0, 1]$ be a continuous kernel function satisfying $0 < \int_{-1}^1 k(x)dx < \infty$, and we consider the kernel estimator

$$\widehat{\sigma}_n(u) = \widehat{\sigma}_{\lfloor un \rfloor + 1}, \quad u \in [0, 1), \quad \widehat{\sigma}_n(1) = \widehat{\sigma}_n,$$

where

$$\widehat{\sigma}_t^2 = \frac{\sum_{j=-N}^N k(j/N) 1_{\{1 \leq t-j \leq n\}} \widehat{\varepsilon}_{t-j}^2}{\sum_{j=-N}^N k(j/N) 1_{\{1 \leq t-j \leq n\}}}, \quad t = 1, \dots, n.$$

Here N is a window width and $\widehat{\varepsilon}_t = \Delta X_t$ is the restricted residual.² The estimator is a weighted average of leads and lags of $\widehat{\varepsilon}_{t-j}^2$, with weights summing to 1. For $t < N$, the estimator is determined by leads more than by lags and, for $t > n - N$, the relative weight of the lags is larger.

² We could also use the OLS residual $\widehat{\varepsilon}_t = \Delta X_t - \widehat{\theta}_n X_{t-1}$ here; the asymptotic results will not change.

The estimator proposed by Hansen (1995) involves only lags, and hence can be interpreted as a filtered volatility, whereas the double-sided version considered here can be seen as the smoothed volatility. Preliminary Monte Carlo experiments have revealed that the use of the double-sided weighted average leads to better finite-sample behaviour of the adaptive test considered below.

To prove uniform consistency of $\widehat{\sigma}_n(\cdot)$, we need the following assumptions.

ASSUMPTION 4.1. $\sigma(\cdot)$ is continuous on $[0, 1]$ (i.e., $\sigma \in C[0, 1]$).

ASSUMPTION 4.2. For some $r > 2$, $\sup_t E[|z_t|^{2r}] < \infty$.

LEMMA 4.1. Consider the model (2.1)–(2.3), under Assumptions 2.1, 4.1 and 4.2. If $N = an^b$ for some a and b satisfying $0 < a < \infty$ and $b \in (2/r, 1)$, then both under $\mathcal{H}_0 : \theta = 0$ and under $\mathcal{H}_n : \theta_n = c/n$, as $n \rightarrow \infty$,

$$\sup_{u \in [0, 1]} |\widehat{\sigma}_n(u)^2 - \sigma(u)^2| \xrightarrow{P} 0.$$

REMARK 4.1. Uniform consistency of the kernel estimator requires continuity of $\sigma(\cdot)$ (Assumption 4.1). Hence we exclude level shifts in $\sigma(\cdot)$, as considered in some of the examples in the previous section. Such level shifts can be approximated arbitrarily well by a smooth transition function, such as the logistic function; but it is expected that the non-parametric estimator will perform relatively badly around the change point. As noted later in Remark 4.6, it is possible to develop the main result of this section allowing for a finite number of discontinuities in $\sigma(\cdot)$, bypassing Lemma 4.1. This is not considered explicitly, for simplicity.

REMARK 4.2. The lemma involves a trade-off between the existence of moments and the window width; for distributions with relatively fat tails, such that extreme observations occur with some frequency, more smoothing is needed to obtain consistency.

REMARK 4.3. A simple example of an implementation of the kernel estimator is that of an exponentially weighted (double-sided) moving average. Take $k(x) = e^{-5|x|}$, where the coefficient 5 is chosen such that $k(1) = k(-1) \approx 0$. Then, letting $\lambda_N = k(1/N) = e^{-5/N}$, we have $k(j/N) = \lambda_N^j$, and $\sum_{j=-N}^N k(j/N) \approx (1 + \lambda_N)/(1 - \lambda_N)$, such that $\widehat{\sigma}_t^2 \approx (1 + \lambda_N)^{-1}(1 - \lambda_N) \sum_{j=-N}^N \lambda_N^j \widehat{\varepsilon}_{t-j}^2$. For $N = 100$, this corresponds to a smoothing parameter of $\lambda_N \approx 0.95$. As the sample size increases, λ_N would have to converge to 1 to guarantee consistency, at the rate determined by Lemma 4.1.

REMARK 4.4. In practice, the window width can be chosen by a leave-one-out cross-validation procedure, which involves minimizing

$$\text{CV}(N) = \sum_{t=1}^n \left(\frac{\widehat{\varepsilon}_t^2 - \widehat{\sigma}_t^2(N)}{1 - w_{tt}(N)} \right)^2, \quad w_{tt}(N) = \frac{k(0)}{\sum_{j=-N}^N k(j/N) 1_{\{1 \leq t-j \leq n\}}},$$

over N ; see Wasserman (2006).³ As discussed by Patilea and Raïssi (2012), a formal analysis of post-selection consistency requires the result of Lemma 4.1 to hold uniformly over $N \in [N_l, N_u]$ with N_l and N_u satisfying the rate requirement of the lemma. In non-parametric estimation problems, leave-one-out cross-validation has been shown to lead to a window width that

³ Note that $\{w_{tt}(N)\}_{t=1}^n$ are the diagonal elements of the smoothing matrix that maps the vector $(\widehat{\varepsilon}_1^2, \dots, \widehat{\varepsilon}_n^2)'$ to the vector $(\widehat{\sigma}_1^2, \dots, \widehat{\sigma}_n^2)'$, corresponding to L_{ii} in Wasserman (2006).

minimizes the mean integrated squared error. However, to our knowledge no optimality theory is available when the estimated volatility is used in the construction of a test; the results to follow only require consistency of the volatility estimator. Preliminary Monte Carlo simulations have indicated that this method of selecting the window width leads to better finite sample behaviour of the resulting testing procedures than other ad hoc methods of choosing a and b in the rate $N = an^b$; therefore, we use cross-validation in the Monte Carlo simulations and the empirical application considered below.

The consistency of the kernel estimator $\widehat{\sigma}_n(\cdot)$ can be used for constructing tests for a unit root as follows. First, we can estimate the asymptotic score S_c and information J_c by

$$\widehat{S}_n = \frac{1}{n} \sum_{t=1}^n \frac{X_{t-1} \Delta X_t}{\widehat{\sigma}_t^2}, \quad \widehat{J}_n = \frac{1}{n^2} \sum_{t=1}^n \frac{X_{t-1}^2}{\widehat{\sigma}_t^2}.$$

These equations can be used to construct approximate point-optimal test statistics $\widehat{\Lambda}_n(\bar{c}) = \bar{c} \widehat{S}_n - (1/2) \bar{c}^2 \widehat{J}_n$, or a one-sided LR statistic $\widehat{LR}_n = \widehat{J}_n^{-1/2} \widehat{S}_n$. Consistency of $(\widehat{S}_n, \widehat{J}_n)$ is considered in the next theorem.

THEOREM 4.1. *Consider the model (2.1)–(2.3), under Assumptions 2.1, 4.1 and 4.2. Under $\mathcal{H}_n : \theta_n = c/n$, we have as $n \rightarrow \infty$,*

$$\begin{pmatrix} \widehat{S}_n \\ \widehat{J}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S_c \\ J_c \end{pmatrix}.$$

REMARK 4.5. This theorem implies that we may asymptotically recover (in a weak convergence sense) the likelihood ratio $\Lambda(c)$ by non-parametric estimation of the infinite-dimensional nuisance parameter $\sigma(\cdot)$, meaning that adaptive estimation and testing is possible. Note that the result applies both when $c = 0$ (i.e., under the null hypothesis) and when $c < 0$. A formal analysis of adaptivity involves finding a so-called least-favourable parametric submodel $\{P_{\theta, \phi, n}\}_{\theta \in \mathbb{R}, \phi \in \Phi}$, where $\phi \in \Phi$ is a parameter vector characterizing $\{\sigma_t\}$; see Chapter 25 of Van der Vaart (1998). Adaptivity requires block-diagonality of the information matrix in this model, which is guaranteed by the Gaussianity assumption. To see this, note that the log-likelihood of the model now becomes

$$\ell_n(\theta, \phi) = -\frac{1}{2} \sum_{t=1}^n \left(\log 2\pi \sigma_t^2(\phi) + \frac{(\Delta X_t - \theta X_{t-1})^2}{\sigma_t^2(\phi)} \right),$$

such that

$$\frac{\partial^2 \ell_n}{\partial \theta \partial \phi}(\theta, \phi) = - \sum_{t=1}^n \frac{X_{t-1} (\Delta X_t - \theta X_{t-1})}{\sigma_t^4(\phi)} \frac{\partial \sigma_t^2(\phi)}{\partial \phi},$$

and this will have mean zero when evaluated at the true value.

REMARK 4.6. Xu and Phillips (2008) show that adaptivity with respect to an unknown volatility process can also be established when the volatility process has a finite number of discontinuities. If $\sigma(u)$ is not continuous at $u = s \in (0, 1)$, then $\widehat{\sigma}_n(s)$ will converge to a weighted average of $\sigma(s-) = \lim_{u \uparrow s} \sigma(u)$ and $\sigma(s+) = \lim_{u \downarrow s} \sigma(u)$, so that the result of Lemma 4.1 will not hold. However, because this occurs at a set of points of Lebesgue measure 0, the result of Theorem 4.1 can still be shown to hold in such cases, which shows that the continuity assumption can be avoided, at the cost of a slightly more involved proof.

Theorem 4.1 implies that the limiting null distribution of adaptive tests will be affected by nuisance parameters, just like the DF test – see (2.6). This means that critical values for such tests cannot be tabulated, and should be generated on a case-by-case basis. One possibility is to simulate the asymptotic null distribution of (S_0, J_0) , with $\sigma(\cdot)$ replaced by $\widehat{\sigma}_n(\cdot)$. If the simulated continuous-time processes involved in the limiting distributions are discretized on a grid $u \in \{u_t = t/n\}_{i=0}^n$, then

$$X_0(u_t) = \int_0^{t/n} \widehat{\sigma}_n(u) dW(u) = \sum_{i=1}^t \widehat{\sigma}_i(W(u_i) - W(u_{i-1})) = n^{-1/2} X_t^*,$$

where $X_t^* = \sum_{i=1}^t \widehat{\sigma}_i z_i^*$, with $z_i^* = n^{1/2}(W(u_i) - W(u_{i-1}))$ an i.i.d. $N(0, 1)$ sequence. From this it can be shown that simulating the asymptotic null distribution based on $\widehat{\sigma}_n(\cdot)$ can be interpreted as a ‘volatility bootstrap’ procedure, where the bootstrap is based on the model under the null and uses bootstrap errors $\varepsilon_i^* = \widehat{\sigma}_i z_i^*$.

An alternative is given by the wild bootstrap; see Liu (1988) and Mammen (1993). The application of the wild bootstrap to unit root tests in the presence of non-stationary volatility was analysed first by Cavaliere and Taylor (2008); in subsequent research it has been shown to be widely applicable in unit root and cointegration analysis. The particular implementation we propose is to generate bootstrap samples from

$$X_t^* = \sum_{i=1}^t \varepsilon_i^* = \sum_{i=1}^t \widehat{\varepsilon}_i z_i^*,$$

where $\widehat{\varepsilon}_i = \Delta X_i$ (the restricted residual) and $z_i^* \sim$ i.i.d. $N(0, 1)$, independent of the data. Next, the bootstrap test statistics are constructed from

$$\widehat{S}_n^* = \frac{1}{n} \sum_{t=1}^n \frac{X_{t-1}^* \Delta X_t^*}{\widehat{\sigma}_t^2}, \quad \widehat{J}_n^* = \frac{1}{n^2} \sum_{t=1}^n \frac{X_{t-1}^{*2}}{\widehat{\sigma}_t^2}.$$

Note that we do not propose to re-estimate σ_t^2 for each bootstrap replication, for the simple reason that $\varepsilon_t^{*2} = \widehat{\varepsilon}_t^2 + \widehat{\varepsilon}_t^2(z_t^{*2} - 1)$, so that a kernel estimator applied to ε_t^{*2} would have an additional noise term relative to $\widehat{\sigma}_t^2$.

Let $\widehat{LR}_n = \widehat{J}_n^{-1/2} \widehat{S}_n$, the adaptive one-sided LR statistic, and let $\widehat{LR}_n^* = (\widehat{J}_n^*)^{-1/2} \widehat{S}_n^*$, its bootstrap version (either the volatility bootstrap or the wild bootstrap).

THEOREM 4.2. *Consider the model (2.1)–(2.3), under Assumptions 2.1, 4.1 and 4.2. Under both $\mathcal{H}_0 : \theta = 0$ and $\mathcal{H}_n : \theta_n = c/n$, we have as $n \rightarrow \infty$,*

$$\begin{pmatrix} \widehat{S}_n^* \\ \widehat{J}_n^* \end{pmatrix} \xrightarrow{d_p} \begin{pmatrix} \int_0^1 Z_0(u) dW(u) \\ \int_0^1 Z_0(u)^2 du \end{pmatrix},$$

so that

$$\widehat{LR}_n^* \xrightarrow{d_p} \left(\int_0^1 Z_0(u)^2 du \right)^{-1/2} \int_0^1 Z_0(u) dW(u).$$

The theorem implies that the (wild or volatility) bootstrap is asymptotically valid, in the sense that the bootstrap p -value, defined as $\Pr[\widehat{\text{LR}}_n^* < \widehat{\text{LR}}_n | \{X_t\}_{t=1}^n]$, is asymptotically uniformly distributed on the unit interval under the null hypothesis. Because $\widehat{\text{LR}}_n^*$ has the same limiting null distribution under \mathcal{H}_n with $c \neq 0$ as under \mathcal{H}_0 , it follows that the bootstrap test has the same asymptotic power function as the test based on the true but unknown critical values.

5. EXTENSIONS

The first-order autoregression with a known zero mean and a zero starting value is too restrictive in many empirical applications. Therefore, in this section we discuss how the adaptive test derived in the previous section can be extended in these directions.

Suppose, first, that the observed data are

$$Y_t = \mu' d_t + X_t, \quad t = 1, \dots, n, \quad (5.1)$$

where X_t satisfies the same assumptions as in the previous sections, and d_t is a vector of deterministic functions of t , with μ being a conformable parameter vector. As usual in the unit root literature, we focus on the cases $d_t = 1$ (constant mean μ) and $d_t = (1, t)'$ (linear trend $\mu' d_t = \mu_1 + \mu_2 t$). Maintaining the assumption that $X_0 = 0$, this implies that the point-optimal invariant test for $\theta = 0$ against $\theta_n = \bar{c}/n$, with observed $\{\sigma_t\}_{t=1}^n$, follows as a straightforward extension of the analysis of Elliott et al. (1996). In particular, let $\widehat{\mu}(\bar{c})$ be the OLS estimator of μ in the regression

$$\begin{aligned} \frac{Y_1}{\sigma_1} &= \mu' \frac{d_1}{\sigma_1} + z_1, \\ \frac{\Delta Y_t - (\bar{c}/n)Y_{t-1}}{\sigma_t} &= \mu' \frac{\Delta d_t - (\bar{c}/n)d_{t-1}}{\sigma_t} + z_t, \quad t = 2, \dots, n. \end{aligned}$$

Using the notational convention $Y_0 = 0$ and $d_0 = 0$, so that $\Delta Y_1 = Y_1$ and $\Delta d_1 = d_1$, we have

$$\begin{aligned} \widehat{\mu}(\bar{c}) &= \left(\sum_{t=1}^n \frac{1}{\sigma_t^2} \left(\Delta d_t - \frac{\bar{c}}{n} d_{t-1} \right) \left(\Delta d_t - \frac{\bar{c}}{n} d_{t-1} \right)' \right)^{-1} \\ &\quad \times \sum_{t=1}^n \frac{1}{\sigma_t^2} \left(\Delta d_t - \frac{\bar{c}}{n} d_{t-1} \right) \left(\Delta Y_t - \frac{\bar{c}}{n} Y_{t-1} \right). \end{aligned} \quad (5.2)$$

Next, let $X_t^d(\bar{c}) = Y_t - \widehat{\mu}(\bar{c})' d_t$. Then the point-optimal invariant test rejects for large values of

$$\begin{aligned} \Lambda_n^d(\bar{c}) &= -\frac{1}{2} \sum_{t=1}^n \frac{(\Delta X_t^d(\bar{c}) - (\bar{c}/n)X_{t-1}^d(\bar{c}))^2}{\sigma_t^2} + \frac{1}{2} \sum_{t=1}^n \frac{\Delta X_t^d(0)^2}{\sigma_t^2} \\ &= \bar{c} S_n^d(\bar{c}) - \frac{1}{2} \bar{c}^2 J_n^d(\bar{c}) - \frac{1}{2} A_n^d(\bar{c}), \end{aligned}$$

where, defining $Z_{t-1}^d(\bar{c}) = X_{t-1}^d(\bar{c})/\sigma_t$,

$$S_n^d(\bar{c}) = \frac{1}{n} \sum_{t=1}^n Z_{t-1}^d(\bar{c}) \frac{\Delta X_t^d(\bar{c})}{\sigma_t},$$

$$J_n^d(\bar{c}) = \frac{1}{n^2} \sum_{t=1}^n Z_{t-1}^d(\bar{c})^2,$$

$$A_n^d(\bar{c}) = \sum_{t=1}^n \frac{\Delta X_t^d(\bar{c})^2 - \Delta X_t^d(0)^2}{\sigma_t^2}.$$

The asymptotic distribution of $\Lambda_n^d(\bar{c})$ is given next.

THEOREM 5.1. *Consider the model defined by (2.1)–(2.3) and (5.1), under Assumptions 2.1, 4.1 and 4.2. (a) If $d_t = 1$ (constant mean), then under $\mathcal{H}_n : \theta_n = c/n$, as $n \rightarrow \infty$,*

$$\begin{pmatrix} S_n^d(\bar{c}) \\ J_n^d(\bar{c}) \\ A_n^d(\bar{c}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S_c^d(\bar{c}) \\ J_c^d(\bar{c}) \\ A_c^d(\bar{c}) \end{pmatrix} = \begin{pmatrix} S_c \\ J_c \\ 0 \end{pmatrix},$$

where (S_c, J_c) are as in Theorem 3.1; (b) if $d_t = (1, t)'$ (linear trend), then under $\mathcal{H}_n : \theta_n = c/n$, as $n \rightarrow \infty$,

$$\begin{pmatrix} S_n^d(\bar{c}) \\ J_n^d(\bar{c}) \\ A_n^d(\bar{c}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S_c^d(\bar{c}) \\ J_c^d(\bar{c}) \\ A_c^d(\bar{c}) \end{pmatrix} = \begin{pmatrix} \int_0^1 \sigma(u)^{-2} X_{c,\bar{c}}^d(u) dX_{c,\bar{c}}^d(u) \\ \int_0^1 \sigma(u)^{-2} X_{c,\bar{c}}^d(u)^2 du \\ (M(\bar{c}) - M(0))^2 \int_0^1 \sigma(u)^{-2} du \end{pmatrix},$$

where $X_{c,\bar{c}}^d(u) = X_c(u) - M(\bar{c})u$, and

$$M(\bar{c}) = \left(\int_0^1 \left(\frac{1 - \bar{c}u}{\sigma(u)} \right)^2 du \right)^{-1} \int_0^1 \frac{1 - \bar{c}u}{\sigma(u)^2} (dX_c(u) - \bar{c}X_c(u)du). \tag{5.3}$$

Analogously to Elliott et al. (1996), for a given choice of \bar{c} we can define a GLS-detrended LR test statistic as $LR_n(\bar{c}) = J_n^d(\bar{c})^{-1/2} S_n^d(\bar{c})$. As indicated by Jansson and Nielsen (2012), we can also optimize over the parameter \bar{c} , i.e., use $\min_{\bar{c} \leq 0} LR_n(\bar{c})$ as test statistic. Adaptive bootstrap versions (with σ_t estimated) are defined entirely analogously to the case with no deterministic components discussed in the previous section.

In general, the first-order autoregressive model for Y_t might be misspecified. Therefore, the testing procedures can be extended to higher-order dynamics as follows. Suppose that we maintain (5.1) for the observed time series Y_t , but now (2.1) is replaced by

$$\Delta X_t = \theta X_{t-1} + \phi(L)^{-1} \varepsilon_t, \quad t = 1, \dots, n, \tag{5.4}$$

with $X_0 = 0$, where L is the lag operator and $\phi(z) = 1 - \sum_{j=1}^{p-1} \phi_j z^j$ has all roots outside the unit circle. This corresponds to the AR(p) model

$$\phi(L)(1 - (1 + \theta)L)(Y_t - \mu' d_t) = \varepsilon_t, \tag{5.5}$$

where the errors ε_t are still assumed to satisfy (2.2)–(2.3) and Assumption 2.1.

Generalizing the approach of Elliott et al. (1996), we can (from an asymptotic point of view) ignore the short-term dynamics in constructing the GLS-detrended time series. Therefore, let

$X_t^d(\bar{c}) = Y_t - \widehat{\mu}(\bar{c})'d_t$ with $\widehat{\mu}(\bar{c})$ as defined in (5.2). Then, a test that rejects for small values of the t -statistic of δ in the weighted least-squares regression

$$\frac{\Delta X_t^d(\bar{c})}{\sigma_t} = \delta \frac{X_{t-1}^d(\bar{c})}{\sigma_t} + \sum_{j=1}^{p-1} \gamma_j \frac{\Delta X_{t-j}^d(\bar{c})}{\sigma_t} + z_t, \quad t = p + 1, \dots, n, \quad (5.6)$$

is a natural extension of the DF–GLS test of Elliott et al. (1996). In combination with a non-parametric volatility estimator and the wild bootstrap, this leads to the following testing procedure.

ALGORITHM 5.1. (ADAPTIVE WILD BOOTSTRAP UNIT ROOT LR TEST IN AR(p) MODEL)

- STEP 1. Estimate σ_t based on OLS residuals $\widehat{\varepsilon}_t$ in an AR($p - 1$) for ΔY_t (i.e. an AR(p) for Y_t with a unit root imposed), including a constant if $d_t = (1, t)'$.
- STEP 2. Construct $X_t^d(\bar{c}) = Y_t - \widehat{\mu}(\bar{c})'d_t$, with $\widehat{\mu}(\bar{c})$ as in (5.2), with σ_t replaced by $\widehat{\sigma}_t$.
- STEP 3. Calculate the t -statistic $\widehat{\text{LR}}_n$ for $\delta = 0$ in (5.6) with σ_t replaced by $\widehat{\sigma}_t$.
- STEP 4. Construct bootstrap errors $\varepsilon_t^* = \widehat{\varepsilon}_t z_t^*$, and generate bootstrap observations Y_t^* from the same estimated AR(p) model under the unit root restriction as in Step 1 (using starting values $(Y_1^*, \dots, Y_p^*) = (Y_1, \dots, Y_p)$); construct bootstrap statistics $\widehat{\text{LR}}_n^*$ by applying Steps 2 and 3 to Y_t^* , and use these to calculate the bootstrap p -value.

In practice, the first step will have to be preceded by a lag order selection procedure, based on information criteria, residual autocorrelation tests, or a combination of both. In the next theorem, we assume that this has led to a selected autoregressive order p that is (larger than or) equal to the true order.

THEOREM 5.2. Consider the model defined by (5.4), (2.2), (2.3) and (5.1), under Assumptions 2.1, 4.1 and 4.2. Then, under $\mathcal{H}_n : \theta_n = c/n$, as $n \rightarrow \infty$,

$$\widehat{\text{LR}}_n \xrightarrow{d} \begin{cases} J_c^{-1/2} S_c, & d_t = 1, \\ J_c^d(\bar{c})^{-1/2} S_c^d(\bar{c}), & d_t = (1, t)', \end{cases}$$

where (S_c, J_c) are as in Theorem 3.1, and $(S_c^d(\bar{c}), J_c^d(\bar{c}))$ are as in Theorem 5.1. Under both $\mathcal{H}_0 : \theta = 0$, and $\mathcal{H}_n : \theta_n = c/n$, we have as $n \rightarrow \infty$,

$$\widehat{\text{LR}}_n^* \xrightarrow{d,p} \begin{cases} J_0^{-1/2} S_0, & d_t = 1, \\ J_0^d(\bar{c})^{-1/2} S_0^d(\bar{c}), & d_t = (1, t)', \end{cases}$$

so that bootstrap p -values are asymptotically uniformly distributed on $[0, 1]$ under \mathcal{H}_0 .

6. MONTE CARLO RESULTS

In this section, we compare the finite-sample behaviour of the adaptive one-sided LR test for a unit root with that of the DF-type t -test in a Monte Carlo experiment. We consider four data-generating processes, corresponding to the volatility functions $\sigma_1(\cdot)$ – $\sigma_4(\cdot)$ considered in Section 3; i.e., for any sample size n , we set $\sigma_{jt} = \sigma_j((t - 1)/n)$ for $j = 1, \dots, 4$. The innovations $\{z_t\}_{t=1}^n$ are generated as i.i.d. $N(0, 1)$. The lag length is fixed at 1, both in the data-generating process and in the test regressions. Both tests allow for an unknown mean, removed

by GLS demeaning with $\bar{c} = -7$ – so that DF is in fact the DF–GLS^μ test of Elliott et al. (1996). The sample sizes considered are $n \in \{100, 250, 500\}$, and we use the exponential kernel $k(x) = e^{-5|x|}$, with the window width N selected by leave-one-out cross-validation. The choice of kernel function is arbitrary, in this case inspired by its relation to double-sided exponential smoothing as discussed in Remark 4.3. As usual in kernel-based estimation, we expect the choice of the window width to have a much bigger impact on the results than the choice of the kernel.

The volatility smoother uses restricted residuals in all the scenarios. For both test statistics, we consider two approaches to obtain their critical values. For the adaptive LR statistic, we compare the wild bootstrap implementation to a test based on simulated asymptotic critical values, replacing the unknown $\sigma(\cdot)$ with the estimate $\hat{\sigma}_n(\cdot)$ (i.e., the volatility bootstrap). For the DF statistic, we compare the wild bootstrap implementation with a version using the standard asymptotic critical values (which are valid only in case of unconditional homoscedasticity). All results are based on 10,000 Monte Carlo replications and 999 bootstrap replications (which is also the number of replications used for simulating the asymptotic p -value of the LR test). Both the wild bootstrap DF test and the wild bootstrap LR test use the restricted OLS residuals to generate the bootstrap samples.

The simulation results are provided in the online Appendix. They show that the wild bootstrap is an effective way of correcting size distortions. As the sample size grows, the adaptive LR test realizes an increasing part of the power gain potential over the DF test, as predicted by the power envelope. Although our assumptions do not allow for it, the adaptive LR test has good power in case of a discontinuous change in the volatility process. This supports our claim that such abrupt changes in volatility are not a problem for the test in practice.

7. EMPIRICAL APPLICATION: EU REAL EXCHANGE RATES, 1973–2015

In this section, we apply the adaptive LR test developed in this paper to study the validity of the purchasing power parity (PPP) hypothesis in 16 EU countries. The PPP hypothesis states that in a well-functioning world market, foreign currencies should have the same purchasing power in the long run, which implies that the real exchange rate should exhibit stationary, mean-reverting properties. Macro-economists often use classical unit root tests with real exchange rate data to test the PPP hypothesis, where a rejection of the unit root hypothesis is used as evidence to support the PPP hypothesis; see, e.g., Froot and Rogoff (1995). However, empirical studies often fail to reject the unit root hypothesis, which is sometimes attributed to the low power of classical unit root tests; see Taylor et al. (2001). Because macro-economic time series often display persistent changes in volatility, we conjecture that our LR test designed to account for non-stationary volatility could provide stronger evidence for the PPP hypothesis.

For 16 EU member countries, we analyse their real effective exchange rate (REER), i.e., the average of the bilateral real exchange rates of their trading partners, weighted by the respective trade shares of each partner. The use of REERs provides a test of the multi-country version of PPP; rejection of the unit root hypothesis based on REERs can be viewed as stronger evidence for PPP to hold than tests using bilateral rates; see Bahmani-Oskooee et al. (2007). We focus on the post-Bretton Woods floating exchange rate period, and apply the tests to monthly log-REER observations, 1973:1–2015:12, obtained from the web site of the Bank for International Settlements.⁴

⁴ <http://www.bis.org/statistics/eer.htm>

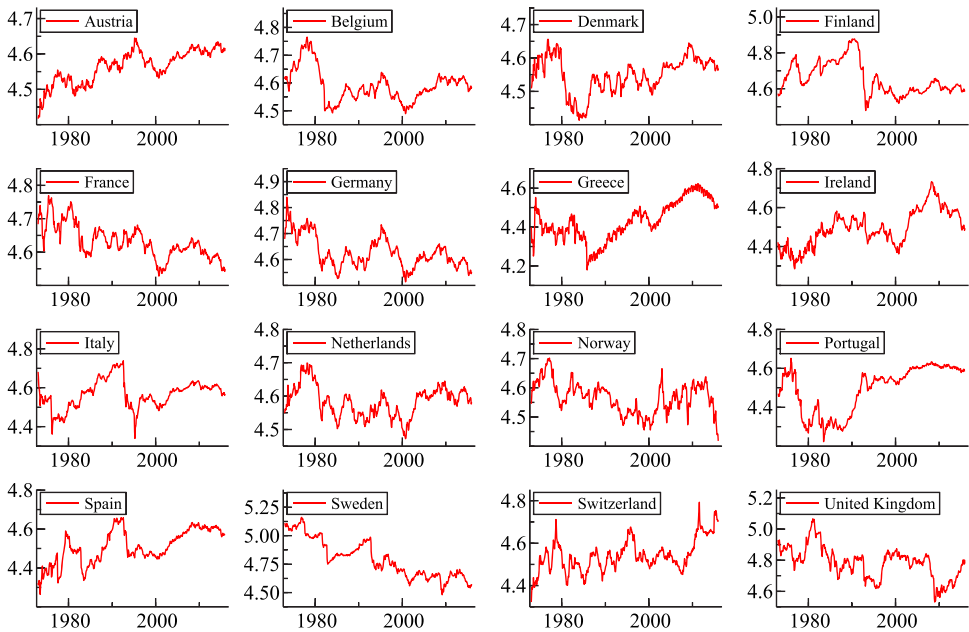


Figure 3. Log-real effective exchange rates for 16 EU countries, 1973:1–2015:12. [Colour figure can be viewed at wileyonlinelibrary.com]

The data are depicted in Figure 3. For almost all countries, we do not observe a clear pattern of strong mean reversion: the REERs can persistently deviate from their mean for a large number of years. This illustrates the common empirical difficulty to find strong evidence supporting PPP.

The results are based on $AR(p)$ models with an unknown mean for each of the 16 time series. The autoregressive orders p have been chosen to obtain residuals with no significant autocorrelation.

Figure 4 displays the non-parametric kernel estimate of the volatility (monthly percentage standard deviation) of the 16 real exchange rate series, where we have used the exponential kernel $k(x) = e^{-5|x|}$, with the window width N selected by the leave-one-out cross-validation method. The volatility estimator is based on the OLS residuals from the selected $AR(p)$ model under the unit root restriction, in agreement with Algorithm 5.1.

It is observed that the volatility of most series decreases gradually, although with different patterns, in the sample period considered; exceptions are Norway, Switzerland and the UK, which display an increase volatility around the financial crisis. The volatility paths suggest that the constant volatility assumption in classical unit root tests might be violated, and it seems reasonable to entertain the possibility of non-stationary volatility.

Table 1 reports wild bootstrap p -values of the DF and adaptive LR tests. For comparison, the asymptotic p -values of the DF test (valid only in case of unconditional homoscedasticity) are also provided; the difference with the bootstrap p -values is small in most cases. We observe that the p -values of the adaptive LR test may be both lower and higher than those of the DF test, and are often in the same order of magnitude. Most remarkable is the result for Italy: using a 5% significance level, the unit root hypothesis is not rejected based on the DF test, but application of

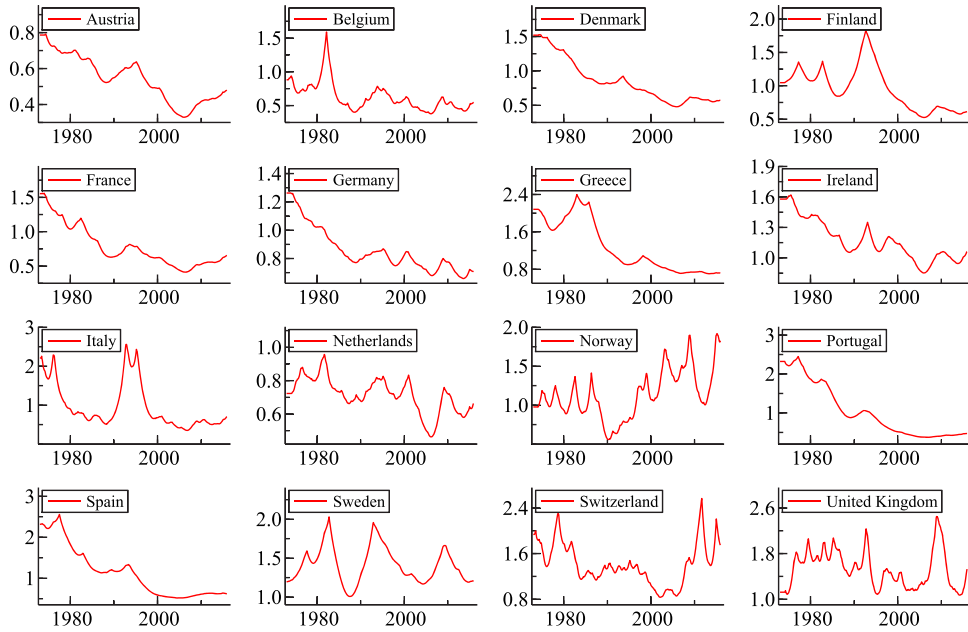


Figure 4. Non-parametric volatility estimate (percentage) for 16 EU countries, 1973:1–2015:12. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 1. Asymptotic and wild bootstrap p -values of DF and adaptive LR test.

| | DF asy p -value | DF WB p -value | LR WB p -value | p |
|----------------|-------------------|------------------|------------------|-----|
| Austria | 0.697 | 0.718 | 0.799 | 12 |
| Belgium | 0.086 | 0.101 | 0.060 | 1 |
| Denmark | 0.050 | 0.062 | 0.059 | 12 |
| Finland | 0.046 | 0.034 | 0.299 | 11 |
| France | 0.140 | 0.141 | 0.252 | 10 |
| Germany | 0.060 | 0.056 | 0.069 | 10 |
| Greece | 0.160 | 0.166 | 0.219 | 12 |
| Ireland | 0.197 | 0.208 | 0.338 | 13 |
| Italy | 0.215 | 0.229 | 0.038 | 1 |
| Netherlands | 0.009 | 0.005 | 0.006 | 12 |
| Norway | 0.020 | 0.019 | 0.022 | 1 |
| Portugal | 0.097 | 0.123 | 0.161 | 12 |
| Spain | 0.519 | 0.492 | 0.799 | 1 |
| Sweden | 0.757 | 0.785 | 0.752 | 1 |
| Switzerland | 0.604 | 0.628 | 0.440 | 1 |
| United Kingdom | 0.131 | 0.130 | 0.062 | 12 |

Note: The table reports asymptotic and wild bootstrap p -values for the DF–GLS $^{\mu}$, and wild bootstrap p -values for the adaptive LR tests, for the real effective exchange rate of 16 EU countries; p refers to the autoregressive order used in the test regressions.

the adaptive LR test leads to a clear rejection, with a p -value of around 4%. To a lesser extent, similar conclusions apply to Belgium and the UK. In summary, the example illustrates that the use of the more powerful adaptive LR test can indeed provide stronger evidence for the PPP hypothesis than using conventional tests, which confirms its useful role in the empirical analysis of macro-economic data.

8. DISCUSSION

In this paper, we have demonstrated that substantial power differences of unit root tests can arise in models with non-stationary volatility. We have shown that it is possible to construct a class of tests that have asymptotic power close to the envelope. The tests are based on non-parametric volatility estimation, and therefore do not require very specific assumptions on the parametric form of the volatility process. This approach can be extended in various directions.

First, for uniform consistency of the non-parametric volatility estimator, the volatility process needs to have continuous sample paths. This means that sudden level shifts are excluded. In practice, one might argue that these can be approximated arbitrarily well by smooth transition functions; furthermore, as shown by Xu and Phillips (2008), adaptive testing might be possible even in the presence of a finite number of discontinuities. The Monte Carlo experiment in this paper suggests that the procedure might perform quite well for level shifts in the volatility.

Secondly, the analysis is based on a deterministic volatility sequence. The asymptotic theory and the bootstrap method could be extended to allow for an exogenous volatility process, as long as it is independent of the Brownian motion defined from the standardized innovations. Hence, this excludes non-stationary volatility processes with statistical leverage effects, which are relevant in applications to equity prices. Note that our approach does not allow for stationary (GARCH-type) conditional heteroscedasticity, with or without leverage effects. It would be of interest to extend the analysis in this direction, leading to further possibilities for higher power.

The analysis in this paper can be extended to the multivariate case. The non-parametric volatility estimator has a very obvious extension to an estimator of a time-varying variance matrix; as long as the same kernel and window width is used for all variances and covariances, the resulting estimator will be positive semi-definite by construction. This can be used to construct more efficient cointegration tests or adaptive estimators of cointegrating vectors in the presence of non-stationary volatility. We are currently exploring this possibility; see Boswijk and Zu (2016).

ACKNOWLEDGEMENTS

We would like to thank the Co-Editor, Michael Jansson, and an anonymous referee for helpful comments and suggestions. Comments on earlier versions from Rob Taylor, Oliver Linton, Peter Phillips, Anders Rahbek, Ulrich Müller and Barbara Rossi are also gratefully acknowledged.

REFERENCES

Bahmani-Oskooee, M., A. M. Kutan and S. Zhou (2007). Testing PPP in the non-linear STAR framework. *Economics Letters* 94, 104–10.

- Beare, B. K. (2016). Unit root testing with unstable volatility. Working paper, University of California, San Diego. Available at <http://econweb.ucsd.edu/bbeare/pdfs/unitroot.pdf>.
- Boswijk, H. P. (2001). Testing for a unit root with near-integrated volatility. Tinbergen Institute Discussion Paper 01-077/4. Available at <http://papers.tinbergen.nl/01077.pdf>.
- Boswijk, H. P. (2005). Adaptive testing for a unit root with nonstationary volatility. UvA-Econometrics Discussion Paper 2005/07. Available at <http://dare.uva.nl/document/228023>.
- Boswijk, H. P. and Y. Zu (2016). Adaptive testing for cointegration with nonstationary volatility. Working paper, University of Amsterdam.
- Cavaliere, G. (2004). Unit root tests under time-varying variances. *Econometric Reviews* 23, 259–92.
- Cavaliere, G. and A. M. R. Taylor (2007). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis* 29, 300–30.
- Cavaliere, G. and A. M. R. Taylor (2008). Bootstrap unit root tests for time series with non-stationary volatility. *Econometric Theory* 24, 43–71.
- Doornik, J. A. (2013). *An Object-oriented Matrix Programming Language – Ox 7*. London: Timberlake Consultants Ltd (<http://www.doornik.com>).
- Elliott, G., T. J. Rothenberg and J. H. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–36.
- Froot, K. A. and K. Rogoff (1995). Perspectives on PPP and long-run real exchange rates. In G. M. Grossman and K. Rogoff (Eds.), *Handbook of International Economics, Volume 3*, 1647–1688. Amsterdam: North-Holland.
- Giné, E. and J. Zinn (1990). Bootstrapping general empirical measures. *Annals of Probability* 18, 851–69.
- Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489–500.
- Hansen, B. E. (1995). Regression with nonstationary volatility. *Econometrica* 63, 1113–32.
- Jansson, M. and M. Ø. Nielsen (2012). Nearly efficient likelihood ratio tests of the unit root hypothesis. *Econometrica* 80, 2321–32.
- Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models. *Econometric Theory* 11, 818–87.
- Kim, T.-H., S. Leybourne and P. Newbold (2002). Unit root tests with a break in innovation variance. *Journal of Econometrics* 109, 365–87.
- Kim, K. and P. Schmidt (1993). Unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 59, 287–300.
- Le Cam, L. and G. L. Yang (1990). *Asymptotics in Statistics: Some Basic Concepts*. New York, NY: Springer.
- Ling, S. and W. K. Li (1998). Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. *Annals of Statistics* 26, 84–125.
- Ling, S., W. K. Li and M. McAleer (2003). Estimation and testing for unit root processes 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, 1696–708.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255–85.
- Patilea, V. and H. Raïssi (2012). Adaptive estimation of vector autoregressive models with time-varying variance: Application to testing linear causality in mean. *Journal of Statistical Planning and Inference* 142, 2891–912.
- Phillips, P. C. B. and K.-L. Xu (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis* 27, 289–308.

- Seo, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. *Journal of Econometrics* 91, 113–44.
- Taylor, M.P., D. A. Peel and L. Sarno (2001). Nonlinear mean-reversion in real exchange rates: toward a solution to the purchasing power parity puzzles. *International Economic Review* 42, 1015–42.
- Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.
- Wasserman, L. (2006). *All of Nonparametric Statistics*. New York, NY: Springer.
- Xu, K.-L. and P. C. B. Phillips (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics* 142, 265–80.

APPENDIX A: PROOFS OF RESULTS

Proof of Lemma 2.1: Let $\alpha_n = 1 + \theta_n$, and note that $X_t = \alpha_n X_{t-1} + \sigma_t z_t$, such that $X_t = \sum_{i=0}^{t-1} \alpha_n^i \sigma_{t-i} z_{t-i}$ (as $X_0 = 0$) and hence

$$n^{-1/2} X_{\lfloor un \rfloor} = f_n(u) \int_0^u g_n(s) dW_n(s), \quad (\text{A.1})$$

where $f_n(u) = \alpha_n^{\lfloor un \rfloor}$ and $g_n(u) = \alpha_n^{-\lfloor un \rfloor - 1} \sigma_n(u)$. It follows that $f_n(u) = (1 + c/n)^{\lfloor un \rfloor} \rightarrow e^{cu}$, and $g_n(u) \rightarrow e^{-cu} \sigma(u)$, both in $D[0, 1]$. The required result

$$n^{-1/2} X_{\lfloor un \rfloor} \xrightarrow{d} \int_0^s e^{c(u-s)} \sigma(s) dW(s)$$

then follows from the continuous mapping theorem (as the integrand is non-stochastic).

The stochastic differential equation for $X_c(u)$ follows from the fact that $Y_c(u) = e^{-cu} X_c(u)$ satisfies $dY_c(u) = e^{-cu} \sigma(u) dW(u)$, and applying Itô's lemma to $X_c(u) = e^{cu} Y_c(u) = f(u, Y_c(u))$, leading to

$$dX_c(u) = ce^{cu} Y_c(u) du + e^{cu} dY_c(u) = cX_c(u) du + \sigma(u) dW(u). \quad (\text{A.2})$$

□

Proof of Theorem 3.1: Write J_n as

$$J_n = \frac{1}{n^2} \sum_{t=1}^n \sigma_t^{-2} X_{t-1}^2 = \int_0^1 \sigma_n(u)^{-2} X_n(u)^2 du,$$

where $X_n(u) = n^{-1/2} X_{\lfloor un \rfloor}$. Lemma 2.1, Assumption 2.1 (including strict positivity of $\sigma(\cdot)$) and the continuous mapping theorem together imply

$$J_n \xrightarrow{d} J_c = \int_0^1 \sigma(u)^{-2} X_c(u)^2 du = \int_0^1 Z_c(u)^2 du.$$

For S_n , we have

$$S_n - cJ_n = \frac{1}{n} \sum_{t=1}^n \sigma_t^{-1} X_{t-1} z_t = \int_0^1 \sigma_n(u)^{-1} X_n(u) dW_n(u).$$

Using $(\sigma_n(\cdot), X_n(\cdot), W_n(\cdot)) \xrightarrow{d} (\sigma(\cdot), X_c(\cdot), W(\cdot))$, and $E[z_t^2] = 1$, it follows from Theorem 2.1 of Hansen (1992) that

$$S_n \xrightarrow{d} \int_0^1 \sigma(u)^{-1} X_c(u) dW(u) + cJ_c = \int_0^1 Z_c(u) dW(u) + cJ_c.$$

□

Proof of Lemma 4.1: The proof is adapted from Hansen (1995), Theorem 2.1. We first show the result:

$$\max_{1 \leq t \leq n} |\widehat{\sigma}_t^2 - \sigma_t^2| \xrightarrow{p} 0. \tag{A.3}$$

Let

$$w_{jN} = \left(\sum_{j=-N}^N k\left(\frac{j}{N}\right) 1_{\{1 \leq t-j \leq n\}} \right)^{-1} k\left(\frac{j}{N}\right) 1_{\{1 \leq t-j \leq n\}},$$

such that $\widehat{\sigma}_t^2 = \sum_{j=-N}^N w_{jN} \widehat{\varepsilon}_{t-j}^2$, with $\sum_{j=-N}^N w_{jN} = 1$. We have

$$\widehat{\sigma}_t^2 - \sigma_t^2 = R_t^a + \sigma_t^2 R_t^b + R_t^c + R_t^d, \tag{A.4}$$

where

$$\begin{aligned} R_t^a &= \sum_{j=-N}^N w_{jN} (\sigma_{t-j}^2 - \sigma_t^2), & R_t^b &= \sum_{j=-N}^N w_{jN} (z_{t-j}^2 - 1), \\ R_t^c &= \sum_{j=-N}^N w_{jN} (\sigma_{t-j}^2 - \sigma_t^2) (z_{t-j}^2 - 1), & R_t^d &= \sum_{j=-N}^N w_{jN} (\widehat{\varepsilon}_{t-j}^2 - \varepsilon_{t-j}^2). \end{aligned}$$

Hansen's proof that $\max_{1 \leq t \leq n} |R_t^a| \xrightarrow{p} 0$, $\max_{1 \leq t \leq n} |\sigma_t^2 R_t^b| \xrightarrow{p} 0$ and $\max_{1 \leq t \leq n} |R_t^c| \xrightarrow{p} 0$ can be directly extended to the present case. For the fourth term, we note that $\widehat{\varepsilon}_t = \varepsilon_t + (c/n)X_{t-1}$. Therefore,

$$\begin{aligned} \left| \sum_{j=-N}^N w_{jN} (\widehat{\varepsilon}_{t-j}^2 - \varepsilon_{t-j}^2) \right| &\leq 2|c| \left| \frac{1}{n} \sum_{j=-N}^N w_{jN} \varepsilon_{t-j} X_{t-1-j} \right| \\ &\quad + c^2 \left| \frac{1}{n^2} \sum_{j=-N}^N w_{jN} X_{t-1-j}^2 \right|. \end{aligned} \tag{A.5}$$

When $c = 0$, the right-hand side is identically zero; when $c \neq 0$, analogous to Hansen (1995), p. 1130, it follows that

$$\max_{1 \leq t \leq n} \left| \frac{1}{n} \sum_{j=-N}^N w_{jN} \varepsilon_{t-j} X_{t-1-j} \right| \xrightarrow{p} 0, \tag{A.6}$$

$$\max_{1 \leq t \leq n} \left| \frac{1}{n^2} \sum_{j=-N}^N w_{jN} X_{t-1-j}^2 \right| = O_p\left(\frac{N^2}{n^2}\right) \xrightarrow{p} 0, \tag{A.7}$$

such that $\max_{1 \leq t \leq n} |R_t^d| \xrightarrow{p} 0$. This proves (A.3).

Next, (A.3) can be strengthened to uniform consistency of $\widehat{\sigma}_n(\cdot)^2$ as follows. Assumption 2.1 implies that $\sup_{u \in [0,1]} |\sigma_n(u)^2 - \sigma(u)^2| \rightarrow 0$ as $n \rightarrow \infty$, where $\sigma_n(u) = \sigma_{\lfloor un \rfloor + 1}$ for $u \in [0, 1)$ and $\sigma_n(1) = \sigma_n$. This definition implies that $\widehat{\sigma}_n(u)^2 - \sigma_n(u)^2 = \widehat{\sigma}_t^2 - \sigma_t^2$ for $u \in [(t-1)/n, t/n]$. This in turn implies that

$$\begin{aligned} \sup_{u \in [0,1]} |\widehat{\sigma}_n(u)^2 - \sigma(u)^2| &\leq \sup_{u \in [0,1]} |\widehat{\sigma}_n(u)^2 - \sigma_n(u)^2| + \sup_{u \in [0,1]} |\sigma_n(u)^2 - \sigma(u)^2| \\ &= \max_{1 \leq t \leq n} |\widehat{\sigma}_t^2 - \sigma_t^2| + \sup_{u \in [0,1]} |\sigma_n(u)^2 - \sigma(u)^2| \\ &\xrightarrow{p} 0. \end{aligned}$$

□

Proof of Theorem 4.1: Consistency of \widehat{J}_n follows directly from Lemma 2.1 and 4.1 and the continuous mapping theorem. For \widehat{S}_n , we use

$$\widehat{S}_n - c\widehat{J}_n = \frac{1}{n} \sum_{t=1}^n \widehat{\sigma}_t^{-2} X_{t-1} \sigma_t z_t = \int_0^1 \widehat{\sigma}_n(u)^{-2} \sigma_n(u) dU_n(u),$$

where

$$U_n(u) := n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} X_{t-1} z_t = \int_0^u X_n(s) dW_n(s).$$

Because $(X_n(\cdot), W_n(\cdot)) \xrightarrow{d} (X_c(\cdot), W(\cdot))$, and $E[z_t^2] = 1$, it follows from Theorem 2.1 of Hansen (1992) that $U_n(u) \xrightarrow{d} U(u) := \int_0^u X_c(s) dW(s)$. Combining this with Lemma 4.1, and the fact that $\sigma(\cdot)$ is strictly positive and non-stochastic, we find

$$\widehat{S}_n \xrightarrow{d} \int_0^1 \sigma(u)^{-1} dU(u) + cJ_c = \int_0^1 Z_c(u) dW(u) + cJ_c = S_c.$$

□

Proof of Theorem 4.2: Define $X_n^*(u) = n^{-1/2} X_{\lfloor un \rfloor}^*$ and

$$\widehat{W}_n^*(u) = n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} \frac{\Delta X_t^*}{\widehat{\sigma}_t}.$$

We prove the following joint convergence:

$$\begin{pmatrix} X_n^*(u) \\ \widehat{W}_n^*(u) \\ \widehat{\sigma}_n(u) \end{pmatrix} \xrightarrow{d}_p \begin{pmatrix} X_0(u) \\ W(u) \\ \sigma(u) \end{pmatrix}, \tag{A.8}$$

with $X_0(u) = \int_0^u \sigma(s) dW(s)$. Using the continuous mapping theorem, it then follows that

$$\widehat{J}_n^* = \int_0^1 \widehat{\sigma}_n(u)^{-2} X_n^*(u)^2 du \xrightarrow{d}_p \int_0^1 \sigma(u)^{-2} X_0(u)^2 du = \int_0^1 Z_0(u)^2 du = J_0, \tag{A.9}$$

and analogously to the proof of Theorem 4.1,

$$\widehat{S}_n = \int_0^1 \widehat{\sigma}_n(u)^{-1} X_n^*(u) d\widehat{W}_n^*(u) \xrightarrow{d}_p \int_0^1 Z_0(u) dW(u),$$

jointly with (A.9). This will prove the theorem, noting that the result for $\widehat{LR}_n^* = (\widehat{J}_n^*)^{-1/2} \widehat{S}_n^*$ follows easily from the continuous mapping theorem.

To prove (A.8) for the wild bootstrap, note that

$$\begin{pmatrix} X_n^*(u) \\ \widehat{W}_n^*(u) \end{pmatrix} = n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} h_t z_t^*, \quad h_t = \begin{pmatrix} \widehat{\varepsilon}_t \\ \widehat{\varepsilon}_t / \widehat{\sigma}_t \end{pmatrix}. \tag{A.10}$$

Because $\{z_t^*\}_{t \geq 1}$ is i.i.d. $N(0, 1)$, independent of the data, it follows that conditionally on the data, this is a bivariate Gaussian process with covariance kernel $\widehat{C}_n(u, s) = \widehat{V}_n(u \wedge s)$, where $\widehat{V}_n(u) := n^{-1} \sum_{t=1}^{\lfloor un \rfloor} h_t h_t'$. Defining $h(s) = (\sigma(s), 1)'$, we prove that

$$\widehat{V}_n(u) \xrightarrow{p} \int_0^u h(s) h(s)' ds =: V(u), \tag{A.11}$$

uniformly in $u \in [0, 1]$. This will imply

$$\begin{pmatrix} X_n^*(u) \\ \widehat{W}_n^*(u) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} X_0(u) \\ W(u) \end{pmatrix},$$

because the right-hand side is a Gaussian process with covariance kernel $C(u, s) = V(u \wedge s)$. The joint convergence (A.8) then holds trivially, because of Lemma 4.1 and because conditionally on the data, $\widehat{\sigma}_n(u)$ has a degenerate distribution.

To prove (A.11), we start with the first diagonal element $\widehat{V}_{11,n}(u)$ of $\widehat{V}_n(u)$:

$$\widehat{V}_{11,n}(u) = \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} \widehat{\varepsilon}_t^2 = \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t^2 + \frac{c^2}{n^3} \sum_{t=1}^{\lfloor un \rfloor} X_{t-1}^2 + 2 \frac{c}{n^2} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t X_{t-1}.$$

For the first term,

$$\frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t^2 = \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} \sigma_t^2 z_t^2 \xrightarrow{p} \int_0^u \sigma^2(s) ds = V_{11}(u),$$

uniformly in $u \in [0, 1]$, which is (a special case of) the classical uniform convergence in probability result for the quadratic variation of a semimartingale. The second and third terms are both 0 under $\mathcal{H}_0 : \theta = 0$; under $\mathcal{H}_n : \theta_n = c/n$, they are both $O_p(n^{-1}) = o_p(1)$, as

$$\left(n^{-2} \sum_{t=1}^{\lfloor un \rfloor} X_{t-1}^2, n^{-1} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t X_{t-1} \right) \xrightarrow{d} \left(\int_0^u X_c(s)^2 ds, \int_0^u X_c(s) \sigma(s) dW(s) \right)$$

by the continuous mapping theorem and weak convergence to the stochastic integral, respectively. Next,

$$\widehat{V}_{22,n}(u) = \frac{1}{n} \sum_{t=1}^{\lfloor un \rfloor} \frac{\widehat{\varepsilon}_t^2}{\widehat{\sigma}_t^2} = \int_0^u \widehat{\sigma}_n(s)^{-2} d\widehat{V}_{11,n}(s) \xrightarrow{p} \int_0^u \sigma(s)^{-2} dV_{11}(s) = u = V_{22}(u),$$

by the continuous mapping theorem. Analogously,

$$\widehat{V}_{12,n}(u) = n^{-1} \sum_{t=1}^{\lfloor un \rfloor} \widehat{\varepsilon}_t^2 / \widehat{\sigma}_t \xrightarrow{p} \int_0^u \sigma(s) ds = V_{12}(u).$$

This proves (A.11) and hence (A.8), and hence the theorem.

An analogous result obtains for the volatility bootstrap, based on $\varepsilon_t^* = \widehat{\sigma}_t z_t^*$. In this case, h_t in (A.10) is replaced by $h_t = (\widehat{\sigma}_t, 1)$, so that Lemma 4.1 directly implies (A.11). \square

Proof of Theorem 5.1: Consider, first, the case of a constant mean, $d_t = 1$. Because $\Delta d_1 = 1$ and $\Delta d_t = 0$ for $t = 2, \dots, n$, it follows that under \mathcal{H}_n ,

$$\widehat{\mu}(\bar{c}) = \left(\frac{1}{\sigma_1^2} + \frac{\bar{c}^2}{n^2} \sum_{t=2}^n \frac{1}{\sigma_t^2} \right)^{-1} \left(\frac{Y_1}{\sigma_1^2} - \frac{\bar{c}}{n} \sum_{t=2}^n \frac{\Delta Y_t - (\bar{c}/n) Y_{t-1}}{\sigma_t^2} \right) = Y_1 + O_p(n^{-1/2}),$$

and hence $X_t^d(\bar{c}) = Y_t - Y_1 + O_p(n^{-1/2}) = X_t - X_1 + O_p(n^{-1/2})$. From this, and the proof of Theorem 3.1, it easily follows that

$$\begin{pmatrix} S_n^d(\bar{c}) \\ J_n^d(\bar{c}) \end{pmatrix} \xrightarrow{d} \begin{pmatrix} S_c \\ J_c \end{pmatrix},$$

with (S_c, J_c) as defined in Theorem 3.1. Furthermore, as $\Delta X_t^d(\bar{c}) = \Delta X_t^d(0) = \Delta X_t$ for $t = 2, \dots, n$ and $\Delta X_1^d(\bar{c}) = \Delta X_1^d(0) + O_p(n^{-1/2})$, it follows that

$$A_n^d(\bar{c}) = \sum_{t=1}^n \frac{\Delta X_t^d(\bar{c})^2 - \Delta X_t^d(0)^2}{\sigma_t^2} = \frac{\Delta X_1^d(\bar{c})^2 - \Delta X_1^d(0)^2}{\sigma_1^2} \xrightarrow{p} 0.$$

In case of a linear trend, we have

$$\Delta d_1 = d_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \Delta d_t = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad t = 2, \dots, n,$$

which yields

$$\begin{aligned} \widehat{\mu}(\bar{c}) - \mu &= \left(\frac{1}{\sigma_1^2} d_1 d_1' + \sum_{t=2}^n \frac{1}{\sigma_t^2} \begin{pmatrix} (\bar{c}/n)^2 & -(\bar{c}/n)(1 - (\bar{c}/n)(t-1)) \\ -(\bar{c}/n)(1 - (\bar{c}/n)(t-1)) & (1 - (\bar{c}/n)(t-1))^2 \end{pmatrix} \right)^{-1} \\ &\times \left(\frac{X_1}{\sigma_1^2} d_1 + \sum_{t=2}^n \frac{1}{\sigma_t^2} \begin{pmatrix} -(\bar{c}/n)(\Delta X_t - (\bar{c}/n)X_{t-1}) \\ (1 - (\bar{c}/n)(t-1))(\Delta X_t - (\bar{c}/n)X_{t-1}) \end{pmatrix} \right). \end{aligned}$$

Next,

$$n^{-1/2} X_{[un]}^d(\bar{c}) = n^{-1/2} X_{[un]} - n^{-1/2} (\widehat{\mu}(\bar{c}) - \mu)' D_n D_n^{-1} d_{[un]},$$

where $D_n = \text{diag}(1, n^{1/2})$ is chosen such that $D_n(\widehat{\mu}(\bar{c}) - \mu)$ has a non-degenerate distribution:

$$\begin{aligned} D_n(\widehat{\mu}(\bar{c}) - \mu) &= \begin{pmatrix} 1/\sigma_1^2 & 0 \\ 0 & n^{-1} \sum_{t=2}^n (1 - (\bar{c}/n)(t-1))^2 / \sigma_t^2 \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} X_1 / \sigma_1^2 \\ n^{-1/2} \sum_{t=2}^n (1 - (\bar{c}/n)(t-1)) (\Delta X_t - (\bar{c}/n)X_{t-1}) / \sigma_t^2 \end{pmatrix} + o_p(1) \\ &\xrightarrow{d} \begin{pmatrix} X_1 \\ M(\bar{c}) \end{pmatrix}, \end{aligned}$$

with $M(\bar{c})$ defined in (5.3). Noting that $n^{-1/2} D_n^{-1} d_{[un]} \rightarrow (0, u)'$, this yields $n^{-1/2} X_{[un]}^d(\bar{c}) \xrightarrow{d} X_{c,\bar{c}}^d(u)$, and hence the limit results for $S_n^d(\bar{c})$ and $J_n(\bar{c})$. Finally, with $\widehat{\mu}_2(\bar{c})$ denoting the second component of $\widehat{\mu}(\bar{c})$, we have

$$\begin{aligned} A_n^d(\bar{c}) &= \sum_{t=1}^n \frac{\Delta X_t^d(\bar{c})^2 - \Delta X_t^d(0)^2}{\sigma_t^2} \\ &= \sum_{t=2}^n \frac{(\Delta X_t - \widehat{\mu}_2(\bar{c}))^2 - (\Delta X_t - \widehat{\mu}_2(0))^2}{\sigma_t^2} \\ &= n(\widehat{\mu}_2(\bar{c})^2 - \widehat{\mu}_2(0)^2) n^{-1} \sum_{t=2}^n \frac{1}{\sigma_t^2} - 2n^{1/2}(\widehat{\mu}_2(\bar{c}) - \widehat{\mu}_2(0)) n^{-1/2} \sum_{t=2}^n \frac{\Delta X_t}{\sigma_t^2} \\ &\xrightarrow{d} (M(\bar{c})^2 - M(0)^2) \int_0^1 \sigma(u)^{-2} du - 2(M(\bar{c}) - M(0)) \int_0^1 \sigma(u)^{-2} dX_c(u) \\ &= (M(\bar{c}) - M(0))^2 \int_0^1 \sigma(u)^{-2} du. \end{aligned}$$

□

Proof of Theorem 5.2: We will follow the steps of Algorithm 5.1.

STEP 1. *Consistency of $\widehat{\sigma}_t$.* The restricted residuals $\widehat{\varepsilon}_t$ are obtained from the OLS regression

$$\begin{aligned} \Delta Y_t &= (\gamma_0 +) \gamma_1 \Delta Y_t + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + \varepsilon_t \\ &= \gamma' \mathbb{Y}_{t-1} + \varepsilon_t, \end{aligned}$$

where the intercept γ_0 is only included in the model with a linear trend, $d_t = (1, t)'$, and where the regressor vector \mathbb{Y}_{t-1} and coefficient vector γ are implicitly defined. Under \mathcal{H}_0 , this regression is of the type studied in Phillips and Xu (2006), and it follows directly from their Theorem 1 that $(\widehat{\gamma}_n - \gamma) = O_p(n^{-1/2})$. Under \mathcal{H}_n , the regression error needs to absorb the omitted variable, so that the error becomes $\varepsilon_t + (c/n)X_{t-1}$. Here, we use the fact that in the above regression, we can replace ΔY_{t-j} by ΔX_{t-j} because they are identical in the model with only a constant, and they differ by a constant in the model with a trend, which in that case is corrected for by the intercept. Because $\sum_{t=p+1}^n \mathbb{Y}_{t-1} \varepsilon_t = O_p(n^{1/2})$ and $\sum_{t=p+1}^n \mathbb{Y}_{t-1} (c/n)X_{t-1} = O_p(n^{1/2})$, this does not change the rate of consistency of $\widehat{\gamma}_n$. Using $\widehat{\varepsilon}_t = \varepsilon_t + (c/n)X_{t-1} - (\widehat{\gamma}_n - \gamma)' \mathbb{Y}_{t-1}$, it now follows analogously to (A.5)–(A.7) that

$$\sum_{j=-N}^N w_{jN} \widehat{\varepsilon}_{t-j}^2 = \sum_{j=-N}^N w_{jN} \varepsilon_{t-j}^2 + o_p(1),$$

which leads to uniform consistency following the proof of Lemma 4.1.

STEP 2. *Limiting representation of adaptive GLS-detrended series.* Define $\psi(z) := \phi(z)^{-1} = \sum_{j=0}^{\infty} \psi_j z^j$, and use the decomposition $\psi(z) = \psi(1) + \psi^*(z)(1-z)$, with $\psi^*(z) = \sum_{j=0}^{\infty} \psi_j^* z^j$, $\psi_j^* = -\sum_{i=j+1}^{\infty} \psi_i$. This leads to

$$u_t := \phi(L)^{-1} \varepsilon_t = \psi(1)\varepsilon_t + \eta_t - \eta_{t-1},$$

where $\eta_t = \psi^*(L)\varepsilon_t$. Because the coefficients ψ_j^* are exponentially decaying, it follows that

$$n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} u_t = \psi(1)n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} \varepsilon_t + n^{-1/2}(\eta_{\lfloor un \rfloor} - \eta_0) \xrightarrow{d} \psi(1) \int_0^u \sigma(s) dW(s).$$

Now, because $\Delta X_t = (c/n)X_{t-1} + u_t$, we find that $n^{-1/2} X_{\lfloor un \rfloor} \xrightarrow{d} \psi(1)X_c(u)$, analogously to the proof of Lemma 2.1, where $X_c(\cdot)$ is the same as in that lemma. The steps of the proof of Theorem 5.1 can now be followed to show that with known σ_t^2 ,

$$n^{-1/2} X_{\lfloor un \rfloor}^d(\bar{c}) \xrightarrow{d} \begin{cases} \psi(1)X_c(u), & d_t = 1, \\ \psi(1)X_{c,\bar{c}}^d(u), & d_t = (1, t)', \end{cases}$$

where $X_{c,\bar{c}}^d(\cdot)$ is the same as defined in Theorem 5.1. Finally, the consistency of $\widehat{\sigma}_t^2$ as analysed above implies that the same limit applies when the estimated σ_t^2 are used in the construction of the GLS-detrended time series.

STEP 3. *Limiting distribution of adaptive t-statistic \widehat{LR}_n .* Assume first that σ_t is known, so that the test is based on the regression $y_t = \beta' x_t + z_t$, where $y_t = \Delta X_t^d(\bar{c})/\sigma_t$,

$$x_t' = \left(\frac{X_{t-1}^d(\bar{c})}{\sigma_t}, \frac{\Delta X_{t-1}^d(\bar{c})}{\sigma_t}, \dots, \frac{\Delta X_{t-p+1}^d(\bar{c})}{\sigma_t} \right) = (x_{1t}, x'_{2t}),$$

and $\beta = (\delta, \gamma_1, \dots, \gamma_{p-1})' = (\delta, \gamma)'$. Let $D_n = \text{diag}(n, n^{1/2}I_{p-1})$, and use the representation

$$D_n(\widehat{\beta}_n - \beta) = \left(\frac{n(\widehat{\delta}_n - \delta)}{n^{1/2}(\widehat{\gamma}_n - \gamma)} \right) = \left(D_n^{-1} \sum_{t=p+1}^n x_t x_t' D_n^{-1} \right)^{-1} D_n^{-1} \sum_{t=p+1}^n x_t z_t.$$

For $d_t = 1$, we have

$$D_n^{-1} \sum_{t=p+1}^n x_t x_t' D_n^{-1} \xrightarrow{d} \begin{pmatrix} \psi(1)^2 \int_0^1 \sigma(u)^{-2} X_c(u)^2 du & 0 \\ 0 & \Sigma \end{pmatrix}, \tag{A.12}$$

$$D_n^{-1} \sum_{t=p+1}^n x_t z_t \xrightarrow{d} \begin{pmatrix} \psi(1) \int_0^1 \sigma(u)^{-1} X_c(u) dW(u) \\ U \end{pmatrix}, \tag{A.13}$$

where Σ is a positive definite stochastic matrix, and U is a random vector. The first diagonal element of (A.12) and the first element of (A.13) follow from the continuous mapping theorem and the results derived in Step 2. An expression for Σ and U can be obtained from the proof of Theorem 1 of Xu and Phillips (2008). Note that Xu and Phillips (2008) consider a stable autoregression, and therefore their results are directly applicable only under \mathcal{H}_0 ; but under \mathcal{H}_n , the additional terms entering x_{2t} will be of lower order such that $n^{-1} \sum_{t=p+1}^n x_{2t} x_{2t}' \xrightarrow{d} \Sigma$ and $n^{-1} \sum_{t=p+1}^n x_{2t} z_t \xrightarrow{d} U$. The off-diagonal block in (A.12) follows from the other results as

$$\begin{aligned} n^{-3/2} \sum_{t=p+1}^n X_{t-1}^d(\bar{c}) \Delta X_{t-j}^d(\bar{c}) &= n^{-3/2} \sum_{t=p+1}^n X_{t-j-1}^d(\bar{c}) \Delta X_{t-j}^d(\bar{c}) \\ &\quad + n^{-3/2} \sum_{i=1}^{j-1} \sum_{t=p+1}^n \Delta X_{t-i}^d(\bar{c}) \Delta X_{t-j}^d(\bar{c}) \\ &= o_p(1). \end{aligned}$$

For the linear trend case $d_t = (1, t)'$ it is easy to see that (A.12) and (A.13) continue to hold but with $X_c(u)$ replaced by $X_{c,\bar{c}}^d(u)$. Consistency of $\hat{\sigma}_t^2$ implies that the same results hold when σ_t^2 is replaced by its estimate.

Because the errors in the adaptive GLS regression already have variance 1, the t -statistic can be defined as $\widehat{\text{LR}}_n = \hat{\delta}/s_{\hat{\delta}}$, where $s_{\hat{\delta}}$ is the square root of the first diagonal element of $(D_n^{-1} \sum_{t=p+1}^n x_t x_t' D_n^{-1})^{-1}$. Realizing that the autoregressive polynomial satisfies

$$(1 - z) - \delta z - \sum_{j=1}^{p-1} \gamma_j z^j (1 - z) = \phi(z)(1 - (1 + c/n)z),$$

we find that $\delta = \delta_n = \phi(1)c/n$, so that (using $\psi(1) = 1/\phi(1)$)

$$\begin{aligned} \widehat{\text{LR}}_n &\xrightarrow{d} \frac{\psi(1) \int_0^1 \sigma(u)^{-1} X_c(u) dW(u)}{\sqrt{\psi(1)^2 \int_0^1 \sigma(u)^{-2} X_c(u)^2 du}} + \phi(1)c \frac{\psi(1)^2 \int_0^1 \sigma(u)^{-2} X_c(u)^2 du}{\sqrt{\psi(1)^2 \int_0^1 \sigma(u)^{-2} X_c(u)^2 du}} \\ &= \frac{\int_0^1 \sigma(u)^{-2} X_c(u) dX_c(u)}{\sqrt{\int_0^1 \sigma(u)^{-2} X_c(u)^2 du}} = \frac{S_c}{\sqrt{J_c}}, \end{aligned}$$

for the case $d = 1$; in the linear trend model, the same steps lead to $\widehat{\text{LR}}_n \xrightarrow{d} J_c^d(\bar{c})^{-1/2} S_c^d(\bar{c})$.

STEP 4. Bootstrap validity. The bootstrap observations are generated from

$$\Delta Y_t^* = (\hat{\gamma}_0 +) \hat{\gamma}_1 \Delta Y_{t-1}^* + \dots + \hat{\gamma}_{p-1} \Delta Y_{t-p+1}^* + \hat{\varepsilon}_t z_t^*,$$

and hence $Y_t^* = Y_1 + \sum_{i=2}^t \Delta Y_i^*$. As before, the intercept is included only in the model with a linear trend; that would generate a linear trend in the bootstrap data, but because we apply GLS detrending to obtain $X_t^{d*}(\bar{c})$, this trend will be eliminated from the observations. Define $X_n^{d*}(u) = n^{-1/2} X_{[un]}^{d*}(\bar{c})$ (the dependence

of the left-hand side on \bar{c} is left implicit for notational convenience), and $\widehat{W}_n(u) = n^{-1/2} \sum_{t=1}^{\lfloor un \rfloor} \widehat{\varepsilon}_t z_t^* / \widehat{\sigma}_t$. Consistency in the least-squares estimators $\widehat{\gamma}_j$ implies that

$$\begin{pmatrix} X_n^{d*}(u) \\ \widehat{W}_n(u) \end{pmatrix} \xrightarrow{d,p} \begin{pmatrix} \psi(1)X_{0,\bar{c}}^d(u) \\ W(u) \end{pmatrix},$$

where $X_{0,\bar{c}}^d(u)$ reduces to $X_0(u)$ if $d_t = 1$. To prove this result, we follow the same steps as the proof of Step 2 given above, in combination with the proof of Theorem 4.2. Using this fact, and the fact that ΔY_{t-j}^* follows a stable autoregression, we obtain the required result, noting that $\delta = 0$ in the bootstrap data. \square

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix
Replication files