Testing for Unit Roots in Time Series Models with Non-Stationary Volatility*

Giuseppe Cavaliere
University of Bologna

A.M.Robert Taylor
University of Birmingham

November 2004

Abstract

Many of the key macro-economic and financial variables in developed economies are characterized by permanent volatility shifts. It is known that conventional unit root tests are unreliable in the presence of such behaviour. Somewhat surprisingly then, very little work has been undertaken to develop unit root tests which are robust to the presence of permanent volatility shifts. In this paper we fill this gap in the literature by proposing a new approach to unit root testing which is valid in the presence of a quite general class of permanent variance changes which includes single and multiple (abrupt and smooth transition) volatility change processes as special cases. The new tests are based on a time transformation of the series of interest which automatically corrects their form for the presence of non-stationary volatility without the need to specify any parametric model for the volatility process. Despite their generality, the new tests perform well even in small samples. Finally, the proposed framework allows us to derive a class of tests for the null hypothesis of stationary volatility in (near) integrated time series processes. An empirical illustration of the proposed approach is given for US unemployment rate data.

Keywords: Unit root tests, integrated processes, non-stationary volatility.

J.E.L. Classifications: C30, C32.

1 Introduction

It is a well established stylized fact that many of the main macro-economic and financial variables across developed countries are characterized by the existence of significant volatility breaks; see, *inter alia*, Busetti and Taylor, (2003), Kim and Nelson (1999), Koop and Potter (2000), McConnell and Perez Quiros (2000), and van Dijk *et al.* (2002). However, it is only very recently that papers have appeared in the literature which investigate the effects of such non-constant variances on conventional methods of unit root and co-integration analysis.

Nelson *et al.* (2001) and Cavaliere (2003) have shown that volatility breaks which follow a stationary Markov switching process do not lead to significant size distortions in the standard unit root tests. Similarly, stationary time-varying conditional variances (e.g. ARCH) are also known not to impact on unit root and co-integration tests; see, for example, Hansen

*Correspondence to: Robert Taylor, Department of Economics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

and Rahbek (1998), Kim and Schmidt (1993) and Ling and MacAleer (2003). Conversely, permanent changes in volatility (so that volatility is non-stationary) can greatly affect unit root inference. For example, Hamori and Tokihisa (1997) show that a single permanent positive variance shift increases the size of Dickey-Fuller [DF]-type tests when no deterministic components are present. The effect of a single variance break on the (constant-corrected) DF t-type test is also analyzed by Kim *et al.* (2002), who show that this test can be seriously oversized when there is an early variance shift. A more general framework for investigating the effects of permanent changes in volatility is provided by Cavaliere (2004), who shows that such changes can lead both to oversized and undersized Phillips-Perron [PP]-type tests, and that the power functions of the tests are also affected.

Despite the potential for serious size distortions, there are very few contributions which propose solutions to this problem. To the best of our knowledge, the most relevant proposal thus far has been suggested by Kim *et al.* (2002) for the specific case of a single abrupt change in variance. They propose a two-stage procedure where the breakpoint together with the pre- and post-break variances are first estimated. These estimates are then employed in modified variants of the Perron (1989,1990) unit root tests.[1] Their proposed statistics have the asymptotic null distributions given in Perron (1990), which depend on the (unknown) breakpoint. The requirement of a single abrupt change in volatility is, however, not consistent with empirical evidence which suggests that data are better characterized by smooth volatility changes rather than by abrupt changes, and that multiple changes in volatility are more commonly observed. The former finding is typified by van Dijk *et al.* (2002), who discuss the need for further research which considers the possibility of smoothly changing volatility and argue that '... no specific event appears to have yet been identified that might be associated with a discrete volatility reduction in macroeconomic series.' The latter finding seems to be a particular characteristic of monetary and financial variables, especially short-term interest rates; see, for example, Watson (1999) who reports three volatility periods for US rates over the 1980-2000 period, without identifying the break dates precisely.

The approach outlined in Kim *et al.* (2002) is tied to the single abrupt change model of volatility and would not be valid against other forms of variance non-stationarity. Their approach could presumably be generalized to other models of volatility change, but this would obviously need to be done on a case-by-case basis entailing a specific choice of parametric model for volatility in each case. It is clear that such an approach is feasible for the case of multiple breaks, although the limiting null distributions of the resulting statistics would clearly depend on all of the breakpoints. It is less clear that the procedure would be feasible in more complicated models. In this paper we consider a completely different approach to the problem and develop unit root tests which are robust to a very general class of volatility changes. In doing so, rather than assuming a specific parametric model for the volatility dynamics, we do not impose any constraint on the volatility dynamics, apart from the reasonable requirement that the variance is bounded and displays a limited number of jumps. Hence, both smooth volatility changes and multiple volatility shifts are allowed within this general class.

The logic behind our proposed tests is relatively simple. As noticed in Cavaliere (2004), the consequences of non-stationary volatility are very similar to those described in Clark

[1] Boswijk (2001) proposes a unit root test which can be applied in the special case of near-integrated GARCH(1,1) volatility. In contrast to Kim *et al.* (2002), his approach is not designed to take account of volatility jumps.

(1973); that is, one observes a deformation of the time domain. The tests which we propose are based on the observation that there exists a one-to-one mapping describing such a deformation, say $f(\cdot)$, which can be consistently estimated under quite general conditions. Consequently, we propose running standard unit root statistics on the 'inverted' time transformation of the original time series, say $X_{t'}$ with[2] $t' = f^{-1}(t)$. This yields statistics which have pivotal asymptotic null distributions even in the presence of non-stationary volatility of unknown form. Notice that no new unit root test statistics are required, we simply compute the existing statistic on the (appropriately) time-transformed data. In this paper we focus attention on 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.

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 non-stationary volatility on conventional unit root tests are discussed. Section 4 develops the heteroskedasticity-robust unit root tests and their asymptotic properties, while in Section 5 some simple tests for the null hypothesis of stationary volatility are also proposed. The finite sample properties of the tests are explored through Monte Carlo simulation in Section 6, while some extensions to the case of general deterministic time trends are given in Section 7. An application to the U.S. unemployment is reported in Section 8. Section 9 concludes. All proofs are placed in the Appendix. In the following '$\xrightarrow{w}$' denotes weak convergence and '$\xrightarrow{p}$' convergence in probability 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), and $\lfloor \cdot \rfloor$ denotes the largest integer less than its argument. Finally, $\mathcal{C} := C[0,1]$ denotes the space of continuous processes on $[0,1]$, and $\mathcal{D} := D[0,1]$ denotes the space of right continuous with left limit (cadlag) processes on $[0,1]$.

2 The Heteroskedastic Model

Consider the time series process $\{X_t\}$ generated according to the model

$$X_t = \alpha X_{t-1} + u_t, \;\; t = 1, 2, ..., T \tag{1}$$

$$u_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \tag{2}$$

$$\varepsilon_t = \sigma_t e_t, \;\; e_t \sim iid(0,1). \tag{3}$$

Unless otherwise stated, we assume that the initial condition X_0 is of $O_p(1)$. The error term $\{u_t\}$ in (2) is a linear process in $\{\varepsilon_t\}$, the latter formed as the product of two components, $\{e_t\}$ and $\{\sigma_t\}$. Since $\{e_t\}$ is *iid*, conditionally on σ_t the error term ε_t has mean zero and time-varying variance σ_t^2. For the present we assume that the DGP contains no deterministic component. This is done purely to aid exposition and is later relaxed.

Through the paper the following set of assumptions will be taken to hold on (1)-(3):

Assumption $\mathcal{A}$: $\mathcal{A}_1$. *The lag polynomial satisfies $C(z) \neq 0$ for all $|z| \leq 1$, and $\sum_{j=0}^{\infty} j|c_j| < \infty$. Moreover, $E|e_t|^r < K < \infty$ for some $r \geq 4$*; $\mathcal{A}_2$. *The autoregressive coefficient α satisfies*

[2] The notation f^{-1} denotes the inverse function of f. Recall that if f^{-1} is the inverse function of f then: $f(f^{-1}(x)) = x$ for all x in the domain of f^{-1}, and $f^{-1}(f(x)) = x$ for all x in the domain of f.

$|\alpha| < 1$ *or* $\alpha = 1$; $\mathcal{A}_3$. *The volatility term* σ_t *satisfies* $\sigma_{[sT]} := \omega(s)$ *for all* $s \in [0,1]$, *where* $\omega(\cdot) \in \mathcal{D}$ *is non-stochastic and strictly positive. For* $t < 0$, $\sigma_t \leq \sigma^* < \infty$.

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 4 exist, see e.g. Chang and Park, 2002) and $\mathcal{A}_2$ are quite standard in the econometric literature[3]. The key assumption to the analysis conducted in this paper is that given in $\mathcal{A}_3$, which casts the dynamics of the innovation variance in a very general framework. The class of volatility processes which satisfy Assumption $\mathcal{A}_3$ is extremely wide. Indeed, the innovation variance is required only to be bounded and to display a limited number of jumps.

The single abrupt change model of Hamori and Tokihisa (1997) and Kim *et al.* (2002) falls within Assumption $\mathcal{A}_3$ and corresponds to the function $\omega(s) := \sigma_0 + (\sigma_1 - \sigma_0)\,\mathbb{I}(s \geq \tau)$, $0 < \tau < 1$, so that the variance shifts from σ_0^2 to σ_1^2 at time $\lfloor \tau T \rfloor$. Multiple volatility shifts can be accommodated similarly by taking $\omega(\cdot)$ piecewise constant. The case of (linearly) trending volatility is obtained for $\omega(\cdot)^2$ an affine function, e.g. $\omega(\cdot)^2 := \sigma_0^2 + \left(\sigma_1^2 - \sigma_0^2\right)s$. Polynomially trending volatility and smooth transition variance breaks are also permitted under Assumption $\mathcal{A}_3$. An example of the latter is the function $\sigma_t^2 := \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{S}_t$, $\mathbb{S}_t := (1 + \exp(-\gamma(t - \lfloor T\tau \rfloor)))^{-1}$. This specification corresponds to a smooth (logistic) transition from σ_0^2 to σ_1^2. The parameter τ, $0 < \tau < 1$, determines the transition midpoint, such that $\sigma_t^2 = 0.5(\sigma_0^2 + \sigma_1^2)$ for $t = \lfloor \tau T \rfloor$, while $\gamma > 0$ controls the speed of transition: as γ approaches zero σ_t^2 tends to $0.5(\sigma_0^2 + \sigma_1^2)$, for all t, while as γ approaches $+\infty$, σ_t^2 changes from σ_0^2 to σ_1^2 instantaneously at time $\lfloor \tau T \rfloor$, thereby containing the fixed changepoint model as a limiting case. Cyclical or periodic heteroskedasticity can also be cast within the proposed framework; here $\omega(\cdot)$ is a periodic function with period p/T, where p is the number of seasons.

Remark 2.1. Although it is not strictly necessary to require that the volatility function $\omega(\cdot)$ is non-stochastic, this does allow for a considerable simplification of the theoretical set-up. This assumption can be easily weakened simply by assuming that $\{\varepsilon_t\}$ and $\{\sigma_t\}$ are stochastically independent; in this case, $\omega(\cdot)$ must obviously have sample paths satisfying Assumption $\mathcal{A}_3$ and the results presented in the paper should then be read as *conditional* on a given realization of the (exogenous) volatility process $\omega(\cdot)$. In the stochastic volatility framework, Markov-switching variances can be obtained by assuming that $\omega(\cdot)$ is a strictly positive, continuous-time Markov chain with a finite number of states. Models of stochastic volatility also fall within the class of processes considered here for $\omega(s) = h(Z(s))$, where $Z(\cdot)$ is a diffusion process in $\mathcal{D}$ and h is a strictly positive continuous function; see Hansen (1995).

Remark 2.2. If $\sigma_t = \sigma$ for all t (i.e. $\omega(s) = \sigma$, all s), the quantity $\sigma^2 C(1)^2$ is the un-normalized spectral density at frequency zero (the so-called 'long-run variance') of $\{u_t\}$. In this case it is well-known that $T^{-1/2}\sum_{t=1}^{[\cdot T]}\varepsilon_t \xrightarrow{w} \sigma B(\cdot)$ and $T^{-1/2}\sum_{t=1}^{[\cdot T]}u_t \xrightarrow{w} \sigma C(1)\,B(\cdot)$, where $B(\cdot)$ is a standard Brownian motion in $\mathcal{C}$; see, *inter alia*, Phillips and Solo (1992). However, these invariance principles do not, in general, hold under Assumption $\mathcal{A}_3$, as will be demonstrated in Section 3.

[3] We have assumed that $\{e_t\}$ is *iid*. This assumption is mainly used to simplify the proofs and most of the results presented in the paper hold under the less restrictive assumption that $\{e_t\}$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_t := \sigma\{e_s, s \leq t\}$ with $\lim_{T\to\infty}\sum_{t=1}^{[Ts]}E(e_t^2|\mathcal{F}_{t-1}) = 1$ (all s), hence allowing for weak heterogeneity in the errors as, for example, in Phillips (1987).

Remark 2.3. Since our interest is testing $\alpha = 1$ against $|\alpha| < 1$, through Assumption $\mathcal{A}_2$ either stable or zero frequency unit roots are allowed, while explosive and seasonal unit roots are ruled out. If $\alpha = 1$, $\{X_t\}$ is a unit root process with heterogeneous increments.

Remark 2.4. Since the variance σ_t^2 depends on T, a time series generated according to (1)-(3) with σ_t satisfying Assumption $\mathcal{A}_3$ formally constitutes a triangular array of the type $\{X_{T,t} : 0 \leq t \leq T, T \geq 1\}$, where $X_{T,t}$ is recursively defined as $X_{T,t} := \alpha X_{T,t-1} + C(L)\,\sigma_{T,t}e_t$, $\sigma_{T,[sT]} := \omega(s)$. However, since the triangular array notation is not essential, for simplicity the subscript T is suppressed in the sequel.

A quantity which will play a fundamental role in what follows is given by the following function in $\mathcal{C}$:

$$\eta(s) := \left(\int_0^1 \omega(r)^2\, dr\right)^{-1} \int_0^s \omega(r)^2\, dr. \tag{4}$$

Hereafter, $\eta(\cdot)$ will be referred to as the *variance profile* of the process, since it depends solely on the time-series behaviour of the volatility. Notice that the variance profile satisfies $\eta(s) = s$ under homoskedasticity while it deviates from s in the presence of heteroskedasticity. Notice also that the quantity $\int_0^1 \omega(r)^2\, dr = ||\omega(\cdot)||_2^2 =: \omega^2$ which appears in (4), 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.

3 Unit Root Asymptotics under Non-Stationary Volatility

In this Section we discuss the impact of time-varying innovation variances on the asymptotic distributions of the M unit root test statistics of Perron and Ng (1996), Stock (1999) and Ng and Perron (2001). Corresponding results for the PP unit root tests under strong mixing assumptions are provided in Cavaliere (2004), while Kim *et al.* (2002) and Hamori and Tokihisa (1997) derive results for the ADF test under the single abrupt variance break model.

For a given sample $\{X_t\}_0^T$, the M tests are defined as:

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

and $\mathcal{MZ}_t := \mathcal{MZ}_\alpha \times \mathcal{MSB}$, where $s_{AR}^2(k)$ is an autoregressive estimator of the (non-normalized) spectral density at frequency zero of $\{u_t\}$. Specifically,

$$s_{AR}^2(k) := \widehat{\sigma}^2/(1-\widehat{\beta}(1))^2,\ \widehat{\beta}(1) := \sum_{i=1}^k \widehat{\beta}_i \tag{5}$$

where $\widehat{\beta}_i$, $i = 1, ..., k$, and $\widehat{\sigma}^2$ are, respectively, the OLS slope and variance estimators from the regression equation $\Delta X_t = \pi X_{t-1} + \sum_{i=1}^k \beta_i \Delta X_{t-i} + e_t$ or, alternatively, from the regression equation $\widehat{u}_t = \sum_{i=1}^k \beta_i \widehat{u}_{t-i} + e_t$, where $\{\widehat{u}_t\}$ denotes the residuals obtained by regressing X_t on X_{t-1}; see also Perron and Ng (1996, 1998). As is standard, we require that the lag truncation parameter, k, satisfies the following assumption (Berk, 1974; Said and Dickey, 1984):

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

Remark 3.1. An alternative long run variance estimator is given by the sum-of-covariances estimator:

$$s_{SC}^2(q) := \frac{1}{T}\sum_{t=-T+1}^{T-1}\sum_{s=-T+1}^{T-1} h\left(\frac{|t-s|}{q}\right)\widehat{u}_t\widehat{u}_s\ , \tag{6}$$

although its finite sample properties are known to often be quite poor relative to those of $s_{AR}^2(k)$; see Elliott *et al.* (1996), *inter alia.*

It is well known that if Assumptions $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{K}$ hold on (1)-(3) with $\alpha = 1$ and $\sigma_t = \sigma$, $t = 1, ..., T$, so that the errors are homoskedastic, the asymptotic (null) distributions of the M statistics are as follows:

$$\begin{aligned}
\mathcal{MZ}_\alpha &\xrightarrow{w} \frac{B(1)^2-1}{2\int_0^1 B(s)^2\,ds} =: \eta_1, \quad \mathcal{MSB} \xrightarrow{w} \left(\int_0^1 B(s)^2\,ds\right)^{1/2} =: \eta_2 \\
\mathcal{MZ}_t &\xrightarrow{w} \frac{B(1)^2-1}{2\left(\int_0^1 B(s)^2\,ds\right)^{1/2}} =: \eta_3
\end{aligned} \tag{7}$$

cf. Ng and Perron (2001) and Chang and Park (2002), *inter alia.* In the case of $\mathcal{MZ}_\alpha$ and $\mathcal{MZ}_t$, the unit root null hypothesis is rejected for large negative values of the statistics, while a test based on $\mathcal{MSB}$ rejects for small values of the statistic. Critical values for the $\mathcal{MZ}_\alpha$- and $\mathcal{MZ}_t$-based tests are provided in the first panels of Tables 10.A.1 and 10.A.2, respectively, of Fuller (1996,pp.641-642), while for the test based on $\mathcal{MSB}$, critical values are as given for the MSB^{GLS} test in the first panel of Table 1 of Ng and Perron (2001,p.1524).

In Theorem 1 we provide representations for the asymptotic null distributions of the M statistics when the variance process σ_t is now permitted to be generated by any process satisfying Assumption $\mathcal{A}_3$. As we shall see, a key role in the asymptotic distribution of the statistics is played by the process $B_\eta(s) := B(\eta(s))$ in $\mathcal{C}$, where $B(\cdot)$ is a standard Brownian motion. As $\omega(\cdot) \neq 0$ a.e., $\eta(\cdot)$ is an increasing homeomorphism on $[0, 1]$ with $\eta(0) = 0$ and $\eta(1) = 1$. Consequently $\{B_\eta(\cdot)\}$ is a 'variance-transformed Brownian motion'; see, for example, Davidson (1994) and Revuz and Yor (1991). It constitutes a Brownian motion under a modification of the time domain, since $B_\eta(\cdot)$ at time $s \in [0, 1]$ has the same distribution as the standard Brownian motion $B(\cdot)$ at time $\eta(s) \in [0, 1]$. Variance-transformed processes trace back to the concept of 'subordinate stochastic process', see Clark (1973), as $B_\eta(\cdot)$ is subordinated to $B(\cdot)$ with directing process $\eta(\cdot)$.

Theorem 1 *Let* $\{X_t\}_0^T$ *be generated as in* (1)-(3) *under Assumption* $\mathcal{A}$ *with* $\alpha = 1$. *Then: (i)* $T^{-1/2}X_{\lfloor \cdot T\rfloor} \xrightarrow{w} \omega C(1)B_\eta(\cdot)$; *(ii) if Assumption* $\mathcal{K}$ *also holds,* $s_{AR}^2(k) \xrightarrow{p} \omega^2 C(1)^2$, $\mathcal{MZ}_\alpha \xrightarrow{w} 0.5(B_\eta(1)^2-1)(\int_0^1 B_\eta(s)^2\,ds)^{-1} = 0.5(B(1)^2-1)(\int_0^1 B_\eta(s)^2\,ds)^{-1}$, $\mathcal{MSB} \xrightarrow{w} (\int_0^1 B_\eta(s)^2\,ds)^{1/2}$ *and* $\mathcal{MZ}_t \xrightarrow{w} 0.5(B_\eta(1)^2-1)(\int_0^1 B_\eta(s)^2\,ds)^{-1/2} = 0.5(B(1)^2-1)(\int_0^1 B_\eta(s)^2\,ds)^{-1/2}$.

Theorem 1 contains three key results. First, under non-stationary volatility the scaled process, $T^{-1/2}X_{\lfloor \cdot T\rfloor}$, weakly converges in the presence of a unit root to a variance-transformed (rather than standard, as in the homoskedastic case) Brownian motion process. That is, heteroskedasticity induces a time-deformation in the sense of Clark (1973) in the limiting Brownian process. Second, the AR estimator, $s_{AR}^2(k)$, converges in probability to $\omega^2 C(1)^2$, where ω^2 denotes the average innovation volatility, as defined at the end of the previous section. Third, the asymptotic distributions of the statistics under the unit root null have the usual form with the exception that the standard limiting Brownian motion, $B(\cdot)$ is replaced by the corresponding variance-transformed Brownian motion, $B_\eta(\cdot)$.[4] This difference has serious implications for the size of the associated unit root tests, with the critical values from Fuller (1996, Tables 10.A.1-10.A.2,pp.641-642) and Ng and Perron (2001,Table 1.p.1524) no longer appropriate. The effects of a number of particular variance profiles on the M tests are quantified by Monte Carlo methods in Section 6.1; see also Hamori and Tokihisa (1997), Kim *et al.* (2002) and Cavaliere (2004).

Remark 3.2. The results given in Theorem 1 apply to the case where (1) contains no deterministic components. However, the stated results generalize straightforwardly to the case where $(X_t - \gamma' z_t) = \alpha(X_{t-1} - \gamma' z_{t-1}) + u_t$, $t = 1, ..., T$, with $\alpha = 1$ and u_t as previously defined, and z_t a vector of deterministic components satisfying the conditions laid out in Section 7. As noted in Ng and Perron (2001,p.1521), in this case the foregoing unit root statistics should be formed from appropriately de-trended data; that is, replace X_t throughout by $\overline{X}_{t.z}$, the OLS (one might also use pseudo-GLS de-trending as in Elliott *et al.*, 1996) residuals from the regression of X_t on z_t. In this case the results of Theorem 1 generalize (using a notation for the resulting statistics consistent with that of Section 7) to: $\mathcal{MZ}_{\alpha.z} \xrightarrow{w} 0.5(F_{B_\eta|Z}(1)^2 - 1)(\int_0^1 F_{B_\eta|Z}(s)^2 ds)^{-1}$, $\mathcal{MSB}_z \xrightarrow{w} (\int_0^1 F_{B_\eta|Z}(s)^2 ds)^{1/2}$, and $\mathcal{MZ}_{t.z} \xrightarrow{w} 0.5(F_{B_\eta|Z}(1)^2 - 1)(\int_0^1 F_{B_\eta|Z}(s)^2 ds)^{-1/2}$, where $F_{B_\eta|Z}(\cdot)$ denotes the Hilbert projection in $L_2[0,1]$ of $B_\eta(\cdot)$ onto the space orthogonal to $Z(\cdot)$ (as defined in Assumption $\mathcal{X}$ of Section 7); cf. Phillips and Xiao (1998). One might also consider, as for example in Müller and Elliott (2003), adding the term $-T^{-1}\overline{X}_{0.z}^2$ to the numerator of the resulting 'de-trended' $\mathcal{MZ}_\alpha$ statistic. In this case the limiting null distributions of the resulting $\mathcal{MZ}_{\alpha.z}$ and $\mathcal{MZ}_{t.z}$ statistics will have the term $-F_{B_\eta|Z}(0)^2$ added to the numerator of the above representations, thereby yielding de-trended Dickey-Fuller distributions under homoskedasticity. For that reason we shall use this version in what follows. In either case, provided the deterministic kernel spans the space of an intercept, we may relax the condition that X_0 be of $O_p(1)$, since the unit root statistics will be exact similar to X_0.

Remark 3.3. The results given in Theorem 1 can be readily extended to the near-integrated case where we have a local-to-unity autoregressive parameter, $\alpha := \exp(-c/T)$, $c > 0$ in (1). It is straightforward to demonstrate that Theorem 1 continues to hold but with $B_\eta(\cdot)$ replaced by the corresponding diffusion process

$$J_\eta^c(s) := \int_0^s \exp(-c(s-r))\, dB_\eta(r). \tag{8}$$

Notice that $J_\eta^c(s)$ of (8) reduces to a standard Ornstein-Uhlenbeck process for $\omega(s)$ constant. A consequence of this result is that the asymptotic local power function of the unit root tests will also be affected by non-stationary volatility. Again these effects are quantified for a number of forms of $\omega(s)$ in Section 6.2.

[4] Notice that under homoskedasticity, i.e. where $\omega(s) = \sigma$ for all s, the representations for the limiting null distributions of the $\mathcal{MZ}_\alpha$, $\mathcal{MSB}$ and $\mathcal{MZ}_t$ statistics given in part (ii) of Theorem 1 reduce to those given in (7), as should be expected.

4 Testing for Unit Roots under Non-Stationary Volatility

As demonstrated in the previous section, non-stationary volatility introduces a time deformation aspect to the limiting null distributions of unit root statistics which alters their form *vis-à-vis* the homoskedastic case. In this section we suggest a modification of the standard unit root statistics formed by applying a heteroskedasticity-dependent transformation of the observed time series which eliminates the time-deformation appearing in the asymptotic distributions. We initially consider the case of no deterministic components with $X_0 = O_p(1)$. This is later relaxed.

For the present, let us assume that the variance profile $\eta(\cdot)$ is known. Consequently its (unique) inverse function, $g(s) := \eta^{-1}(s)$, say, is also known. Now consider the time-transformed series $\{\widetilde{X}_t\}_0^T$

$$\widetilde{X}_t := X_{t'}, t' := \lfloor g(t/T)T \rfloor, \quad t = 0, ..., T. \tag{9}$$

The process $\{\widetilde{X}_t\}$ therefore corresponds to $\{X_t\}$ indexed by the non-decreasing sequence $\{t'\}$: hence, the mapping from X_t to $\widetilde{X}_t$ constitutes a time-deformation. Before moving on, let us first consider an example of these calculations based on the single abrupt change model.

Example 1 *Suppose a single break in variance from σ_0^2 to σ_1^2 occurs at time $\lfloor \tau T \rfloor$; i.e.*

$$\omega(s)^2 = \begin{cases} \sigma_0^2 \text{ if } s < \tau \\ \sigma_1^2 \text{ if } s \geq \tau. \end{cases}$$

Let $\delta := \sigma_0/\sigma_1$ and $\kappa := (1-\tau)/\delta^2 + \tau$. The corresponding variance profile is

$$\eta(s) = \begin{cases} \frac{1}{\kappa} s & \text{if } s < \tau \\ \frac{1}{\kappa\delta^2} s + \frac{\delta^2 - 1}{\kappa\delta^2}\tau & \text{if } s \geq \tau \end{cases}$$

and its unique inverse is given by the function

$$g(s) = \begin{cases} \kappa s & \text{if } s < \frac{\tau}{\kappa} \\ \kappa\delta^2 s + (1-\delta^2)\tau & \text{if } s \geq \frac{\tau}{\kappa}. \end{cases}$$

Consequently, for $t = 0, ..., T$, the time-transformed series is given by

$$\tilde{X}_t = X_{\lfloor g(t/T)T \rfloor} = \begin{cases} X_{\lfloor \kappa t \rfloor} & \text{if } t < \frac{1}{\kappa}\lfloor \tau T \rfloor \\ X_{\lfloor \kappa\delta^2 t + (1-\delta^2)\tau T \rfloor} & \text{if } t \geq \frac{1}{\kappa}\lfloor \tau T \rfloor \end{cases}$$

Notice that for $t = 0$, $\tilde{X}_0 = X_0$; for $t = T$, $\tilde{X}_T = X_T$; and for $t = \lfloor \frac{1}{\kappa}\tau T \rfloor$, $\tilde{X}_{\lfloor \frac{1}{\kappa}\tau T \rfloor} = X_{\lfloor \tau T \rfloor}$.

A very useful property of the time-transformed data, $\{\widetilde{X}_t\}$, which will be formally proved later, is that, after normalizing, it obeys weak convergence to a standard Brownian motion. To see this observe, using part (i) of Theorem 1, that

$$T^{-1/2}\widetilde{X}_{\lfloor sT \rfloor} = T^{-1/2}X_{\lfloor g(\frac{\lfloor sT \rfloor}{T})T \rfloor} \approx T^{-1/2}X_{\lfloor g(s)T \rfloor} \xrightarrow{w} \omega C(1) B_\eta(g(s)). \tag{10}$$

Notice that the right member of (10) simplifies to $\omega C(1)B(s)$, since $B_\eta(g(s)) = B(\eta(g(s))) = B(s)$. Furthermore, using the Continuous Mapping Theorem [CMT], we obtain that

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=1}^{T}\widetilde{X}_{t-1}^2 = \frac{1}{T^2}\sum_{t=1}^{T}X^2_{\lfloor g(\frac{t-1}{T})T\rfloor} &= \sum_{t=1}^{T}\int_{\frac{t-1}{T}}^{\frac{t}{T}}\left(\frac{X_{\lfloor g(\frac{\lfloor sT\rfloor}{T})T\rfloor}}{T^{1/2}}\right)^2 ds + \frac{X_T^2}{T^2} \\
&\xrightarrow{w} \omega^2 C(1)^2\int_0^1 B_\eta(g(s))^2\,ds \\
&= \omega^2 C(1)^2\int_0^1 B(s)^2\,ds
\end{aligned}$$

which depends on $\omega(\cdot)$ only through ω^2, the asymptotic average innovation variance. Again using the CMT, it follows that, for example, the quantity

$$\mathcal{MZ}_\alpha^* := \frac{T^{-1}\widetilde{X}_T^2 - \omega^2 C(1)^2}{2T^{-2}\sum_{t=1}^{T}\widetilde{X}_{t-1}^2} = \frac{T^{-1}X_T^2 - \omega^2 C(1)^2}{2T^{-2}\sum_{t=1}^{T}X^2_{\lfloor g(\frac{t-1}{T})\cdot T\rfloor}} \xrightarrow{w} \eta_1. \tag{11}$$

With knowledge of $\omega(\cdot)$ (and, hence, ω) and $C(1)$ it is therefore possible to obtain a statistic whose limiting null distribution under non-stationary volatility of an unknown form (satisfying Assumption $\mathcal{A}_3$) coincides with the (pivotal) asymptotic null distribution of the $\mathcal{MZ}_\alpha$ statistic obtained under homoskedasticity; cf. (7). Similar results hold for $\mathcal{MSB}$- and $\mathcal{MZ}_t$-type statistics.

In order to make the foregoing approach operational, consistent estimators of $\omega^2 C(1)^2$ and $g(\cdot)$ will clearly be required. These can be readily obtained as follows. First, as was demonstrated in part (ii) of Theorem 1, the long-run variance estimator $s^2_{AR}(k)$ of (5) is a consistent estimator of $\omega^2 C(1)^2$. Turning to the estimation of $g(\cdot)$, let $\{\widehat{u}_t\}$ denote the residuals of the linear regression of X_t on X_{t-1}, and consider the following estimator of $\eta(s)$:

$$\widehat{\eta}(s) := \frac{\sum_{t=1}^{\lfloor sT\rfloor}\widehat{u}_t^2 + (sT - \lfloor sT\rfloor)\widehat{u}^2_{\lfloor sT\rfloor+1}}{\sum_{t=1}^{T}\widehat{u}_t^2} \tag{12}$$

with $\widehat{\eta}(1) := 1$ and $\widehat{\eta}(0) := 0$. This estimator is strictly increasing and admits a unique inverse, say $\widehat{g}(s)$, unless $\widehat{u}_t = 0$ for some t. We rule out this possibility by imposing the following assumption, which is in general satisfied provided that the errors $\{e_t\}$ have density continuous at zero.

Assumption $\mathcal{H}$. *The functional $\widehat{\eta}(s)$ is strictly increasing in s.*

The function $g(\cdot)$ in (11) can then be replaced by its consistent estimator, $\widehat{g}(s)$, which is obtained by numerical inversion[5] of the estimated variance profile, $\widehat{\eta}(s)$. Notice that both estimators are constructed from the original data, X_t, rather than the time-transformed data.

Using the estimators $\widehat{g}(s)$ and $s^2_{AR}(k)$ we may now propose the following modified versions of the $\mathcal{MZ}_\alpha$, $\mathcal{MSB}$ and $\mathcal{MZ}_t$ unit root statistics. In what follows, just as we refer to the original tests of Perron and Ng (1996), Stock (1999) and Ng and Perron (2001) as M tests, we will refer to our proposed heteroskedastic-robust versions as M^H tests.

[5]Most mathematical packages have in-built procedures for the numerical inversion of functions. Gauss code to do this is also available from the authors on request.

Definition 1 *Our modified unit root statistics are defined as*

$$\mathcal{MZ}_{\alpha}^{H} := \frac{T^{-1}X_T^2 - s_{AR}^2(k)}{2T^{-2}\sum_{t=1}^{T} X^2_{\lfloor \widehat{g}(\frac{t-1}{T})T \rfloor}} \tag{13}$$

$$\mathcal{MSB}^{H} := \left(T^{-2}\sum_{t=1}^{T} X^2_{\lfloor \widehat{g}(\frac{t-1}{T})T \rfloor} / s_{AR}^2(k) \right)^{1/2} \tag{14}$$

$$\mathcal{MZ}_{t}^{H} := \mathcal{MZ}_{\alpha}^{H} \times \mathcal{MSB}^{H}, \tag{15}$$

where the time-transformed data, $\{\tilde{X}_t\}_{t=0}^{T}$*, are as defined in (9),* $\widehat{g}(s)$ *is the inverse of* $\widehat{\eta}(s)$ *of (12) and* $s_{AR}^2(k)$ *is as defined in (5).*

In Theorem 2 we now state the main result of the paper. Here we establish a uniform convergence result for the sample variance profile and its inverse, and show that the limiting null distributions of the modified unit root statistics of Definition 1 are pivotal.

Theorem 2 *Under the conditions of Theorem 1 and Assumption* $\mathcal{H}$*:* (i) $\widehat{\eta}(s) - \eta(s) \xrightarrow{p} 0$ *and* $\widehat{g}(s) - g(s) \xrightarrow{p} 0$*, uniformly for all* $s \in [0,1]$*; and* (ii) $\mathcal{MZ}_{\alpha}^{H} \xrightarrow{w} \eta_1$*,* $\mathcal{MSB}^{H} \xrightarrow{w} \eta_2$*, and* $\mathcal{MZ}_{t}^{H} \xrightarrow{w} \eta_3$*.*

Theorem 2 provides two powerful results. First, the variance profile $\eta(s)$ can be consistently estimated by a simple residual-based estimator. Second, this estimator allows us to implement a change-of-time transformation to the observed time series which enables us, for any innovation variance sequence satisfying Assumption $\mathcal{A}_3$, to construct the modified statistics $\mathcal{MZ}_{\alpha}^{H}, \mathcal{MSB}_{\alpha}^{H}$ and $\mathcal{MZ}_{t}^{H}$ whose asymptotic null distributions coincide with those given for the corresponding un-modified statistics under homoskedasticity in (7). Tests based on the $\mathcal{MZ}_{\alpha}^{H}$, $\mathcal{MSB}_{\alpha}^{H}$ and $\mathcal{MZ}_{t}^{H}$ statistics will therefore be (asymptotically) similar under the conditions of Theorem 2, with the critical values provided in Fuller (1996, Tables 10.A.1-10.A.2,pp.641-642) and Ng and Perron (2001,Table 1,p.1524) remaining appropriate.

Remark 4.1. The estimator, $\hat{\eta}(\cdot)$, of the variance profile $\eta(\cdot)$ suggested in (12) is based on the residuals of an estimated first order autoregression. Alternatively, one could estimate $\eta(\cdot)$ under the null hypothesis; that is, by means of

$$\widehat{\eta}_{\Delta}(s) := \frac{\sum_{t=1}^{\lfloor sT \rfloor}(\Delta X_t)^2 + (sT - \lfloor sT \rfloor)(\Delta X_{\lfloor sT \rfloor+1})^2}{\sum_{t=1}^{T}(\Delta X_t)^2}.$$

Yet another possible estimator of $\eta(\cdot)$ is given by:

$$\widehat{\eta}_{k}(s) := \frac{\sum_{t=1}^{\lfloor sT \rfloor}\widehat{e}_t^2 + (sT - \lfloor sT \rfloor)\widehat{e}^2_{\lfloor sT \rfloor+1}}{\sum_{t=1}^{T}\widehat{e}_t^2}$$

where $\{\widehat{e}_t\}$ denotes the residuals of the ADF regression of ΔX_t on $(X_{t-1}, \Delta X_{t-1}, ..., \Delta X_{t-k})'$ where k could be as used in constructing $s_{AR}^2(k)$ of (5). These three estimators of the variance profile can be shown to be asymptotically equivalent.

Remark 4.2. Since the time-transformation from $\{X_t\}$ to $\{\tilde{X}_t\}$ will most likely entail some observations being dropped from the computation of the test statistics, one could consider replacing $T^{-2}\sum_{t=1}^{T} X^2_{\lfloor \widehat{g}\left(\frac{t-1}{T}\right)T\rfloor}$ with the quantity

$$\frac{1}{\lfloor hT\rfloor T}\sum_{n=h}^{\lfloor Th\rfloor} X^2_{\lfloor \widehat{g}\left(\frac{n-h}{\lfloor Th\rfloor}\right)T\rfloor},$$

where $h \geq 1$. In practice, however, we found that such a refinement did not improve the finite sample properties of the tests and so will not be discussed further.

Remark 4.3. It is straightforward to show that the asymptotic distributions of the modified unit root statistics, $\mathcal{MZ}_{\alpha}^{H}, \mathcal{MSB}_{\alpha}^{H}$ and $\mathcal{MZ}_{t}^{H}$, under near-integrated alternatives, $\alpha := 1 - c/T$ for $c > 0$, are as given in Theorem 2 but with the Brownian motion $B(s)$ replaced by $J_{\eta}^{c}(g(s))$, where $J_{\eta}^{c}(\cdot)$ is as defined in (8); see also Cavaliere (2004), Lemma 3. Therefore, even though the test statistics have pivotal limiting distributions under the null hypothesis, the asymptotic local power functions of the modified tests under non-stationary volatility will depend on the variance profile $\eta(\cdot)$.

Remark 4.4. Thus far we have assumed that (1) contains no deterministic component. Consider now the constant case where

$$(X_t - \mu) = \alpha(X_{t-1} - \mu) + u_t, \quad t = 1, ..., T \tag{16}$$

with u_t as previously defined. In the standard unit root testing problem, invariance to the level, μ, can be obtained by computing the M statistics on OLS de-meaned data; cf. Remark 3.2. In the presence of heteroskedasticity, however, the sample average $\overline{X} := (T+1)^{-1}\sum_{t=0}^{T} X_t \xrightarrow{w} \int_0^1 B_{\eta}(s)\, ds$, which depends on $\eta(\cdot)$. Consequently, tests based on applying a change-of-time transformation to the de-meaned series $X_t - \overline{X}$ would not have pivotal limiting null distributions. A de-meaning procedure which delivers statistics whose limiting null distributions do not depend on $\eta(\cdot)$ can, however, be formed from the quantity $\widehat{X}_{t.1} := X_t - (T+1)^{-1}\sum_{t=0}^{T} X_{\lfloor \widehat{g}\left(\frac{t}{T}\right)T\rfloor}$. The resulting constant corrected forms of the M^H unit root statistics are given by[6]

$$\begin{aligned}
\mathcal{MZ}_{\alpha.1}^{H} &:= \frac{T^{-1}\widehat{X}_{T.1}^2 - T^{-1}\widehat{X}_{0.1}^2 - s_{AR}^2(k)}{2T^{-2}\sum_{t=1}^{T}\widehat{X}^2_{\lfloor \widehat{g}\left(\frac{t-1}{T}\right)T\rfloor.1}}, \ \mathcal{MSB}_{1}^{H} := \left(T^{-2}\sum_{t=1}^{T}\widehat{X}^2_{\lfloor \widehat{g}\left(\frac{t-1}{T}\right)T\rfloor.1}/s_{AR}^2(k)\right)^{1/2} \\
\mathcal{MZ}_{t.1}^{H} &:= \mathcal{MZ}_{\alpha.1}^{H} \times \mathcal{MSB}_{1}^{H}
\end{aligned} \tag{17}$$

where $\widehat{\eta}(\cdot)$ is as defined in (12) but with $\{\widehat{u}_t\}$ now the OLS residuals from the regression X_t on X_{t-1} and a constant, and $s_{AR}^2(k)$ formed as in (5) but with the slope and variance estimators obtained from the regression of ΔX_t on X_{t-1}, $\{\Delta X_{t-i}\}_{i=1}^{k}$ and a constant. It is then straightforward, from previous results, to show that the following theorem holds.

[6] As noted in Remark 3.2, alternative versions of these tests could be considered with the term $-T^{-1}\widehat{X}_{0.1}^2$ omitted from the numerator of $\mathcal{MZ}_{\alpha.1}^{H}$, although the resulting $\mathcal{MZ}_{\alpha.1}^{H}$ and $\mathcal{MZ}_{t.1}^{H}$ statistics would not have de-meaned Dickey-Fuller limiting null distributions.

Theorem 3 *Let $\{X_t\}_{t=0}^{T}$ be generated as in (16),(2),(3) under Assumption $\mathcal{A}$ with $\alpha = 1$. Let Assumptions $\mathcal{H}$ and $\mathcal{K}$ also hold. Then,* $\mathcal{MZ}_{\alpha.1}^{H} \overset{w}{\rightarrow} 0.5(B_1(1)^2 - B_1(0)^2 - 1)(\int_0^1 B_1(s)^2 ds)^{-1}$, $\mathcal{MSB}_1^H \overset{w}{\rightarrow} (\int_0^1 B_1(s)^2 ds)^{1/2}$ *and* $\mathcal{MZ}_{t.1}^{H} \overset{w}{\rightarrow} 0.5(B_1(1)^2 - B_1(0)^2 - 1)(\int_0^1 B_1(s)^2 ds)^{-1/2}$, *where* $B_1(s) := B(s) - \int_0^1 B(r)\, dr$ *is a de-meaned Brownian motion.*

Again, the statistics have pivotal limiting null distributions under non-stationary volatility, these coinciding with those obtained for the corresponding standard unit root statistics under homoskedasticity with OLS de-meaning; cf. Remark 3.2 for $z_t = 1$. Relevant critical values are tabulated in, for example, the middle panels of Tables 10.A.1 and 10.A.2 of Fuller (1996,pp.641-642) for the $\mathcal{MZ}_{\alpha.1}^{H}$ and $\mathcal{MZ}_{t.1}^{H}$ tests, respectively, and in Stock (1999, Table 1,p.151) for the $\mathcal{MSB}_1^H$ test.[7] Moreover the assumption that X_0 is of $O_p(1)$ may now be dropped, as in Remark 3.2.

Remark 4.5. Although outlined for the case of the M unit tests of Perron and Ng (1996), Stock (1999) and Ng and Perron (2001), the approach used in this section can also be readily applied to any of the commonly used tests of the unit root null hypothesis, simply by applying those tests to the time-transformed series $\{\widetilde{X}_t\}$ in the same manner.

5 Tests of Stationary Volatility

The testing framework outlined in the previous sections also allows us to derive a consistent test of the null hypothesis of constant volatility which can be applied when the data generating process is either integrated or near-integrated. Although the tests outlined in this section formally focus on the null of constant volatility (homoskedasticity), we shall refer to them as tests of stationary volatility since the reference test statistics have the same asymptotic null distributions when the stochastic process driving volatility is stationary (e.g. stationary and ergodic ARCH or Markov-switching processes) as under constant volatility.

Tests of stationary volatility can be based on the sample estimate of the variance profile, $\widehat{\eta}(\cdot)$ of (12). Specifically, since $\widehat{\eta}(\cdot) \overset{p}{\rightarrow} \eta(\cdot)$, as was demonstrated in Theorem 2, if the errors are homoskedastic, so that $\eta(s) = s$, all s, then it is to be expected that $\widehat{W}(s) := \widehat{\eta}(s) - s$ will be close to zero for all $s \in [0,1]$. A formal test for the null hypothesis of stationary volatility can therefore be obtained by examining whether or not the deviations of $\widehat{W}(\cdot)$ from zero are too large to be attributable purely to sampling error. What is required to produce a formal statistical test is to choose a function which maps the random $\mathcal{C}$ function $\widehat{W}(\cdot)$ into a scalar random variable together with an appropriate scaling factor for $\widehat{W}(\cdot)$. Both of these are given in the next Theorem which states a number of possible statistics that can be used to test the null of stationary volatility.

Theorem 4 *Let the conditions of Theorem 1 hold with either $\alpha = 1$ or $\alpha := \exp(-c/T)$, $c > 0$. Moreover, let λ_v^2 denote the long-run variance of $v_t := u_t^2$ under the null hypothesis $H_0^H : \omega(s) = \omega$, all s, and assume that a strictly positive estimator $\widehat{\lambda}_v^2$ of λ_v^2 which is*

[7] Stock (1999) does not report a 1% critical value for this test: using direct simulation of the limiting functional for $5,000$ steps and 20,000 replications, we estimated this to be 0.16.

consistent under H_0^H is available[8]*. Finally, let $\widehat{\omega}^2 := T^{-1}\sum_{t=1}^{T}\widehat{u}_t^2$. Then, under H_0^H*

$$\mathcal{H}_R := \frac{T^{1/2}\widehat{\omega}^2}{\widehat{\lambda}_v}\left(\sup_{s\in[0,1]}\widehat{W}(s) - \inf_{s\in[0,1]}\widehat{W}(s)\right) \xrightarrow{w} \sup_{s\in[0,1]}V(s) - \inf_{s\in[0,1]}V(s)$$

$$\mathcal{H}_{KS} := \frac{T^{1/2}\widehat{\omega}^2}{\widehat{\lambda}_v}\left(\sup_{s\in[0,1]}|\widehat{W}(s)|\right) \xrightarrow{w} \sup_{s\in[0,1]}|V(s)|$$

$$\mathcal{H}_{CVM} := \frac{T\widehat{\omega}^4}{\widehat{\lambda}_v^2}\left(\int_0^1 \widehat{W}(s)^2\,ds\right) \xrightarrow{w} \int_0^1 V(s)^2\,ds$$

$$\mathcal{H}_{AD} := \frac{T\widehat{\omega}^4}{\widehat{\lambda}_v^2}\left(\int_0^1 \frac{\widehat{W}(s)^2}{s(1-s)}ds\right) \xrightarrow{w} \int_0^1 \frac{V(s)^2}{s(1-s)}ds$$

where $V(\cdot)$ is a standard Brownian Bridge process.

The asymptotic null distributions of the $\mathcal{H}_R$, $\mathcal{H}_{KS}$, $\mathcal{H}_{CVM}$ and $\mathcal{H}_{AD}$ statistics are well-known. Critical values for the tests which reject for large values of these statistics are as given in Shorack and Wellner (1987) for the so-called Kuiper (*ibid.*, Table 2, p.144), Kolmogorov-Smirnov (*ibid.*, Table 1, p.413), Cramer-Von Mises (*ibid.*, Table 4, p.147) and Anderson-Darling (*ibid.*, Table 5, p.148) tests, respectively.

Under the alternative hypothesis that $\omega(\cdot)$ is non-constant, the deviations of $W(s) := \eta(s) - s$ from zero are trivially seen to be of order $T^{1/2}$, while $\widehat{\lambda}_v^2$ remains strictly positive. Consequently, the tests which reject for large values of the statistics of Theorem 3 will be consistent. This result is formalized in the next Theorem, the proof of which is entirely straightforward and, hence, is omitted.

Theorem 5 *Let the conditions of Theorem 4 hold with $\omega(\cdot)$ non-constant over a non-zero Lebesgue measure set. Then, $\mathcal{H}_R$, $\mathcal{H}_{KS}$, $\mathcal{H}_{CVM}$, $\mathcal{H}_{AD}$ all diverge to positive infinity as $T \to \infty$.*

Remark 5.1. The present approach to homoskedasticity testing is not new in the literature. As can be seen from the proof of Theorem 4, what is actually being tested is that $u_t^2 - E\left(u_t^2\right)$ displays no structural mean shifts; Pagan and Schwert (1990) use a similar approach to testing the constancy of second-order moments. Their approach has also been extended by Loretan and Phillips (1994), who propose statistics equivalent to $\mathcal{H}_R$ and $\mathcal{H}_{KS}$.

Remark 5.2. Notice that the results in Theorems 4 and 5 are trivially seen to remain valid for the case where $\widehat{\eta}(s)$ used in constructing $\widehat{W}(s)$ is as in (12) but with $\{\widehat{u}_t\}$ the OLS residuals from the regression X_t on X_{t-1} and the deterministic component z_t; cf. Remarks 3.2, 4.4 and Section 7. Moreover, the stated results also hold valid for the other estimators of $\eta(\cdot)$ discussed in Remark 4.1.

6 Small sample simulations

In this section we analyze the small sample size and power properties of the heteroskedasticity-robust M^H unit root tests of Section 4 by means of Monte Carlo (MC) simulation. Precisely,

[8]Note that $\widehat{\lambda}_v^2$ can be either a sum-of-covariance estimator or an autoregressive estimator.

the $\mathcal{MZ}_\alpha^H$, $\mathcal{MZ}_t^H$ and $\mathcal{MSB}^H$ tests, and their constant corrected counterparts of Remark 4.4, are considered and compared with the standard $\mathcal{MZ}_\alpha$, $\mathcal{MZ}_t$ and $\mathcal{MSB}$ tests and their constant corrected counterparts of Remark 3.2. Results for the $\mathcal{H}_R$, $\mathcal{H}_{KS}$, $\mathcal{H}_{CVM}$ and $\mathcal{H}_{AD}$ homoskedasticity tests of Section 5 are also reported.

Data are generated according to (1)–(3), with $X_0 = 0$, for samples of size $T = 100, 250$ and 500 with $e_t \sim niid(0,1)$, using the RNDN function of Gauss 5.0. All experiments were conducted using $10,000$ MC replications. Initially, and in order to demonstrate the effects of heteroskedasticity uncontaminated by the separate effects of serial dependence, we set $C(L) = 1$ in (2), so that the error term $u_t = \sigma_t e_t$. Correspondingly, we set $k = 0$ in (5). We later extend the analysis to also allow for weak dependence in u_t. All tests were run using the nominal 0.05 level critical values for the standard M tests appropriate for $k = 0$ and independent homoskedastic Gaussian errors. Consequently, while the results for the M tests are based on an exact 0.05 nominal level, those for the M^H tests are not. As the results in Table 1 for the constant variance case show, this causes a small degree of undersizing in the M^H tests for $T = 100$. This effect should be borne in mind when assessing the results.

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

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

$$\omega_1(s)^2 := \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{I}(s \geq \tau) \ , \ s \in [0,1] \tag{18}$$

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

2. (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

$$\omega_2(s)^2 := \sigma_0^2 + \left(\sigma_1^2 - \sigma_0^2\right)\mathbb{I}(\tau_1 \leq s < \tau_2) + \left(\sigma_2^2 - \sigma_0^2\right)\mathbb{I}(\tau_2 \leq s \leq 1) \ , \ s \in [0,1]$$

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

3. (Trending volatility) Volatility follows a linear trend, between σ_0^2 for $t = 0$ and σ_1^2 for $t = T$, which corresponds to the variance function

$$\omega_3(s)^2 := \sigma_0^2 + \left(\sigma_1^2 - \sigma_0^2\right) s \ , \ s \in [0,1]. \tag{19}$$

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

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

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 1 we let the ratio $\delta := \sigma_0/\sigma_1$ vary among $\{1/5, 5\}$, 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 2 we consider the case of symmetric shifts, so that $\delta_2 := \sigma_1/\sigma_2 = 1/\delta_1$ with $\delta_1 = \delta := \sigma_0/\sigma_1$, $\tau_2 = 1 - \tau_1$ and, hence, the variance function simplifies to $\omega(s)^2 := 1 + \delta^{-2}\left(1-\delta^2\right)\mathbb{I}(\tau \le s < 1-\tau)$. Here we let δ vary among $\{1/5, 5\}$, and $\tau \in \{0.05, 0.45\}$ allowing for breaks which occur either towards the middle or the start and end of the sample. For Model 3 we let $\delta := \sigma_0/\sigma_1$ take values among $\{1/5, 5\}$ so that both positively and negatively trending variances are generated. Finally, for Model 4 we vary ν among $\{5, 10\}$ and c among $\{0, 10, 20\}$.[9] For each of Models 1-4 other combinations of parameter values were also considered, but these qualitatively add little to the reported results.

6.1 Size Properties

Tables 1 to 4 present the (empirical) size, for $\alpha = 1$ in (1), of both the standard M unit root tests and the corresponding M^H heteroskedastic-robust unit root tests of Section 4, together with the outcome of the heteroskedasticity tests. In all cases the variance profile is estimated under the null hypothesis. However, the results are little different when the variance profile is estimated without the unit root imposed.[10]

Tables 1 to 4 about here.

Let us initially consider the single break model (Table 1). As for the unit root tests considered by Kim *et al.* (2002) and Cavaliere (2004), the standard M unit root tests are generally oversized, with early negative and late positive breaks having the strongest impact. For the standard tests, undersizing is observed only for the (constant-corrected) $\mathcal{MZ}_{\alpha.1}$ and $\mathcal{MZ}_{t.1}$ tests in the case of an early positive break. In contrast, the size properties of our M^H unit root tests seem largely satisfactory.

In the no deterministics case, for late positive and early negative breaks the M^H unit root tests tend to be slightly undersized with actual size about 3% for $T = 250$ and above 4% for $T = 500$. In contrast, for late positive and early negative breaks the size of the standard tests ranges between 10% and 20%, and between 8% and 10%, respectively. Early positive and late negative breaks have much less impact on the size of the standard tests, but again the empirical size of the M^H tests is closer to the nominal level than for the corresponding standard M tests. When the tests are based on de-meaned data, the standard M tests display quite severe size biases for early negative and late positive breaks with empirical size of up to 51% on the nominal 5% significance level. In contrast, the M^H tests perform very well, although in the cases of late positive and early negative breaks sample sizes of at least 250 observations are needed to avoid significant undersizing. Interestingly, in the case of an early positive break while the $\mathcal{MZ}_{t.1}$ test tends to be somewhat undersized the corresponding $\mathcal{MZ}^H_{t.1}$ test has the correct size. It is also worth noting that the $\mathcal{MZ}^H_\alpha$, $\mathcal{MZ}^H_t$ and $\mathcal{MSB}^H$ tests display similar size properties to one another, as do the constant corrected versions of these tests; moreover, in all cases their empirical sizes rapidly converge to the nominal 5% level as T grows, as predicted by the asymptotic distribution theory; cf. Theorems 2 and 3.

The results for the double volatility break model (Table 2) are broadly comparable to those obtained for the single break model. As far as the standard M tests are concerned, when no

[9] Note that the values of c and ν calibrated on Kim *et al.*'s (1998) SV model estimates for the daily U.K. Sterling/U.S. Dollar exchange rate over the 1981–85 period ($T = 946$ observations) are given by $c = 21.172$ and $\nu = 4.866$.

[10] The full set of results is available on request.

deterministic correction is used the most severe size biases are seen for a positive followed by negative break, both occurring towards the middle of the sample (actual size is around 11%), while an early negative break followed by a late positive break raises the empirical size of standard tests based on de-meaned data to 33%. Also notice that, similarly to what obtained for the single break model, when an early positive break is followed by a late negative break, the $\mathcal{MZ}_{t.1}$ test is undersized (3% size) while the $\mathcal{MZ}_{t.1}^{H}$ test is not. Conversely, our proposed M^H tests perform well, completely avoiding the large oversizing problems seen with the M tests, although (as with the single break case) at least 250 observations are needed to avoid significant undersizing in certain cases.

The results in Table 3 show that, in general, linear trending volatility has a lower impact on the size of the standard tests than abrupt changes. In the case of no deterministic correction, the M tests are notably oversized (around 10% size) in the presence of upward trending volatility ($\delta = 1/5$), but only slightly oversized for downward trending volatility ($\delta = 5$). For the case of de-meaned data the M tests again show size distortions with the $\mathcal{MZ}_{t.1}$ test the worst affected with significant undersizing (oversizing) in the presence of upward (downward) trending volatility. In contrast, the M^H tests display empirical sizes close to the nominal level throughout, even for $T = 100$ (bearing in mind the under-size in the M^H tests attributable to the use of critical values for the corresponding M test).

The size properties of the tests in the presence of (near-) integrated volatility are reported in Table 4. Let us briefly concentrate on standard M tests first. In the 'mildest' case, when ν is low and volatility is strongly mean reverting ($c = 20$), standard tests perform well in terms of size, which is always between 5% and 6%. However, when ν grows and/or c decreases, the impact of volatility becomes more pronounced, leading to oversized tests. When the volatility is non-stationary ($c = 0$) and ν is low, the rejection frequencies of the standard unit root tests with and without deterministic corrections are about 10% and 15%, respectively. When ν is high, such frequencies almost become double. No cases of undersized tests are observed. The performance of the M^H tests is very good under the most realistic case where ν is low, with sizes mostly between 4% and 5%, both with tests based on raw data and tests based on de-meaned data. When ν is high, the size of the tests is fair in the near-integrated volatility cases for sample sizes of $T = 250$ and above, while in the integrated case the tests are still slightly undersized for $T = 500$.

Because our proposed heteroskedasticity-robust unit root tests do not employ any information on the form of heteroskedasticity, one might have anticipated that they would show serious size distortions in small samples. Our MC results suggest, however, that overall this does not tend to be the case with the heteroskedasticity-dependent time-deformation underlying the M^H tests seemingly performing well against a diverse variety of possible volatility schemes, both in cases where the conventional unit root tests are oversized and where they are undersized. In most cases a sample size of $T = 250$ observations appears large enough to ensure good size properties. None of the three modified tests seems to dominate the others in terms of size accuracy.

6.2 Power Properties

Tables 5 to 8 report (size-adjusted) powers of the standard M unit root tests and the heteroskedastic-robust modifications thereof of Section 4. The reported results pertain to the particular alternative of $\alpha = 1 - 7/T$ in (1); qualitatively similar conclusions were drawn for other values of α. As with the simulations of Section 6.1, the variance profile is estimated under the null hypothesis; again the results did not differ when the variance profile was estimated without the unit root imposed. Size-adjusted power is reported in order to control for the unreliable size properties of the standard M tests under heteroskedastic errors; see Section 6.1.

Tables 5 to 8 about here.

Consider first the case where the errors are homoskedastic, which is reported in the first block of results in Table 5. Here the proposed M^H tests display roughly the same power as the corresponding standard M tests. Hence, appealingly, the application of the transformation based on the estimated inverse variance profile which underlies these tests does not appear to lead to any significant power loss when such a transformation is not necessary.

The effect of heteroskedasticity on the standard M tests is mixed. For instance, in the single break model a late positive break or an early negative break decreases the power of the M tests relative to the constant volatility case, while power is increased for early positive breaks in the constant-corrected case. Similar, often exacerbated, effects are seen for the corresponding M^H tests. In those cases where heteroskedasticity has significant effects on the size of the M tests, the M^H tests are usually less powerful than the corresponding M tests, although for some DGPs the reverse is observed. In constrast, in those cases where non-stationary volatility has little impact on the size of the standard M tests (e.g. Model 1 with a late negative break for both raw and de-meaned data, Model 2 with early positive and late negative breaks for raw data, and Model 4 with $c = 20$ and $\nu = 5$ for both raw and de-meaned data) we observe that the power of the M^H and corresponding M tests is quite similar. Such differences across cases are not unexpected given the observation made in Remarks 3.3 and 4.3 that the local power functions of both the M and M^H tests depend (in different ways) on the variance profile, $\eta(\cdot)$. While comparisons between the power of the M and M^H tests under heteroskedasticity are of some interest it should, of course, be remembered that, under heteroskedasticity, the M tests, unlike the corresponding M^H tests, do not display adequate size control and, hence, cannot be relied upon to provide reliable inference in practice.

Among the proposed tests, those based on raw data are generally more powerful than the corresponding test based on OLS de-meaned data. That is, the power ranking observed in the homoskedastic case appears, in general, to be preserved under a variety of heteroskedasticity patterns. Moreover, among the tests based on de-meaned data, the $(\mathcal{MSB}_1, \mathcal{MSB}_1^H)$ tests seems preferable in terms of power to the $(\mathcal{MZ}_{\alpha.1}, \mathcal{MZ}_{\alpha.1}^H)$ and the $(\mathcal{MZ}_{t.1}, \mathcal{MZ}_{t.1}^H)$ tests. Conversely, among the tests which do not involve a constant correction, the $(\mathcal{MZ}_{\alpha}, \mathcal{MZ}_{\alpha}^H)$ and $(\mathcal{MZ}_t, \mathcal{MZ}_t^H)$ tests perform marginally better than the $(\mathcal{MSB}, \mathcal{MSB}^H)$ tests.

6.3 Power of the Stationary Volatility Tests

We now briefly review the power properties of the $\mathcal{H}_R$, $\mathcal{H}_{KS}$, $\mathcal{H}_{CVM}$ and $\mathcal{H}_{AD}$ tests of the null of stationary volatility of Section 5, reported in Tables 1–8.

For the single volatility break case, the $\mathcal{H}_{AD}$ test is clearly the most powerful, followed by the $\mathcal{H}_{CVM}$ test. The $\mathcal{H}_R$ test displays the lowest power among the tests considered for all values of the volatility parameters. For the two-break case the $\mathcal{H}_R$ test now appears to be the most preferable, although the $\mathcal{H}_{AD}$ test also displays good power properties for $T = 250$

and above. For the linearly trending volatility case the $\mathcal{H}_{CVM}$ and $\mathcal{H}_{AD}$ tests perform best. Interestingly, the presence of linearly trending volatility appears to be better detected by the tests than are abrupt shifts. Finally, for the stochastic volatility model the $\mathcal{H}_R$ test now dominates. Overall, no one test dominates the others on power. Finally it is worth noting that the power properties of the stationary volatility tests do not appear to be affected to any degree by the presence/absence of a unit root.

6.4 Autocorrelated Errors

The size of the foregoing tests is now examined for data generated as outlined at the start of this section with $\alpha = 1$, but where the errors are now generated according to a linear $AR(1)$ process; that is, $u_t = \phi u_{t-1} + \epsilon_t$ so that $C(L) = (1 + \phi L + \phi^2 L^2 + \cdots)$ in (2). The AR parameter ϕ was chosen within the set $\{-0.5, 0.5\}$; results for other $ARMA$ DGPs did not differ qualitatively from those reported here. The lag truncation parameter, k, used in the computing the long-run variance estimator $s^2_{AR}(k)$ was chosen according to MIC rule of Ng and Perron (2001)[11] with $k \leq \lfloor 12(T/100)^{0.25} \rfloor$. For space constraints results are reported for only four DGPs, namely the single early negative break model ($\delta = 5$, $\tau = 0.1$), the early negative and late positive break model ($\delta_1 = 5$, $\tau = 0.05$), the upward trending volatility model ($\delta = 1/5$) and the non-stationary stochastic volatility model ($\nu = 10$, $c = 0$). The results are reported in Table 9.[12]

Table 9 about here.

When the errors are positively autocorrelated ($\phi = 0.5$), the size properties of the M^H tests are largely satisfactory, with the tests being slightly undersized in the single break model and somewhat oversized in the double break model. For the remaining cases empirical sizes are close to the nominal level. When the errors are negatively autocorrelated ($\phi = -0.5$) the tests are somewhat conservative throughout, most notably for the single break model. It is worth recalling that, under homoskedasticity, the standard M tests with automatic data-dependent lag selection rules also tend to be slightly conservative in the presence of negatively autocorrelated errors; see, in particular, Ng and Perron (2001, Table V.A). The results in Table 9 show that these results are reversed in most cases under heteroskedasticity with the standard M tests often severely oversized, especially so in the constant corrected case, regardless of whether the errors are positively or negatively autocorrelated.

7 Extension to Deterministic Time Trends

Thus far we have assumed that the process of interest either does not contain any deterministic mean components or has a constant level, as in Remark 4.4. We now generalize our results to the case where the data are generated according to $Y_t := d_t + X_t$, where X_t is as previously defined in (1)-(3) and $d_t := \gamma' z_t$, where z_t is a vector of deterministic components satisfying the following assumption (see Phillips and Xiao, 1998):

[11] Although not reported here we also considered the case where the true value of $k = 1$ was used in the computation of $s^2_{AR}(k)$. In this case the resulting sizes were slightly better than those reported in Table 9.

[12] The full set of results is available on request.

Assumption $\mathcal{X}$. *There exists a scaling matrix δ_T and a bounded piecewise continuous function $Z(r)$ such that (a) $\delta_T z_{\lfloor Tr \rfloor} \to Z(r)$ as $T \to \infty$ uniformly in $r \in [0,1]$, and (b) $\int_0^1 Z(r)Z(r)'dr$ is positive definite.*

Assumption $\mathcal{X}$ allows for a wide variety of possible forms for the deterministic component, including the leading case of the pth order trend function, $z_t := (1, t, ..., t^p)'$. The broken intercept and trend functions of Perron (1989,1990), as well as smooth transition intercept and trend break functions are also permitted.

Let $\widehat{\eta}(\cdot)$ and $\widehat{g}(\cdot)$ denote estimators of the variance profile and its inverse respectively, see Section 4, constructed from the OLS residuals from the regression of Y_t on Y_{t-1} and z_t. Moreover, let $\widetilde{Y}_t := Y_{\lfloor \widehat{g}(t/T)T \rfloor}$ and $\widetilde{z}_t := z_{\lfloor \widehat{g}(t/T)T \rfloor}$, and denote by $\widehat{\gamma}$ the estimator of γ obtained from the OLS regression of $\widetilde{Y}_t$ on $\widetilde{z}_t$. Finally, let $\widehat{X}_{t.z}$ denote the resulting de-trended time series $\widehat{X}_{t.z} := Y_t - \widehat{\gamma}' z_t$. The unit root statistics corrected for the deterministic component $\{z_t\}$, which we denote with an obvious notation by $\mathcal{MZ}^H_{\alpha.z}$, $\mathcal{MSB}^H_z$ and $\mathcal{MZ}^H_{t.z}$, are then defined as in (17) but with $\widehat{X}_{t.1}$ replaced by $\widehat{X}_{t.z}$, throughout, and with $s^2_{AR}(k)$ formed as in (5) but with the slope and variance estimators obtained from the regression of ΔX_t on X_{t-1}, $\{\Delta X_{t-i}\}_{i=1}^k$ and z_t. It is then straightforward to show from previous results that the following theorem holds.

Theorem 6 *Let $\{Y_t\}_{t=0}^T$ be such that $Y_t := \gamma' z_t + X_t$, with $\{X_t\}_{t=0}^T$ generated as in* (1)-(3) *under Assumption $\mathcal{A}$ with $\alpha = 1$. Then, under Assumptions $\mathcal{H}$ and $\mathcal{X}$:* (i) $\widehat{\eta}(s) - \eta(s) \xrightarrow{p} 0$, $\widehat{g}(s) - g(s) \xrightarrow{p} 0$ *uniformly for all $s \in [0,1]$ and, letting Assumption $\mathcal{K}$ also hold,* (ii) $\mathcal{MZ}^H_{\alpha.z} \xrightarrow{w} 0.5(F_{B|\breve{Z}}(1)^2 - F_{B|\breve{Z}}(0)^2 - 1)(\int_0^1 F_{B|\breve{Z}}(s)^2 ds)^{-1}$, $\mathcal{MSB}^H_z \xrightarrow{w} (\int_0^1 F_{B|\breve{Z}}(s)^2 ds)^{1/2}$ *and* $\mathcal{MZ}^H_{t.z} \xrightarrow{w} 0.5(F_{B|\breve{Z}}(1)^2 - F_{B|\breve{Z}}(0)^2 - 1)(\int_0^1 F_{B|\breve{Z}}(s)^2 ds)^{-1/2}$, *where*

$$F_{B|\breve{Z}}(s) := B(s) - \breve{Z}(s)' \left(\int_0^1 \breve{Z}(r) \breve{Z}(r)' dr \right)^{-1} \int_0^1 \breve{Z}(r) B(r) dr \tag{20}$$

and $\breve{Z}(s) := Z(g(s))$.

Theorem 6 shows that the limiting null distributions have the usual form except that they are defined in terms of the Hilbert projection in $L_2[0,1]$ of the standard Brownian motion $B(\cdot)$ onto the space orthogonal to $\breve{Z}(\cdot)$. In general, such distributions are not pivotal; however, since $g(\cdot)$ can be consistently estimated and $\breve{Z}(\cdot)$ depends on a standard Brownian motion, their quantiles can be easily retrieved by MC simulation.

Remark 7.1 For part (ii) of Theorem 6 to hold we technically also require conditions analogous to those stated for z_t in Assumption $\mathcal{X}$ to hold on $\tilde{z}_t$. These conditions rule out possible degeneracies induced by the change-of-time transformation. These conditions are satisfied, for example, in the polynomial trend case, $z_t := (1, t, ..., t^p)'$.

Remark 7.2. As in Remark 3.2, the requirement that X_0 is of $O_p(1)$ may be relaxed provided $\{z_t\}$ spans the space of an intercept.

Remark 7.3. For $z_t = 1$, all t, $\breve{Z}(s) = 1$, all s, and $F_{B|\breve{Z}}(s) = B(s) - \int_0^1 B(s) ds$, so that the results of Theorem 3 are obtained, as should be expected.

Remark 7.4 Interestingly, in one prominent case asymptotically pivotal unit root statistics are obtained in the presence of deterministic time trends. This is the 'ARCH-in-mean' trend case, which corresponds to $\Delta z_t = \mu\sigma_t^2$, so that under the null hypothesis $\Delta Y_t = \mu\sigma_t^2 + \sigma_t\varepsilon_t$ is an ARCH-in-mean process (see Engle *et al.*, 1987). In this case, d_t can be written as $d_t = \gamma' z_t$ with $z_t = \left(1, \sum_{i=1}^{t} \sigma_i^2\right)'$ and $\gamma := (\gamma_0, \gamma_1)'$. Therefore, $\delta_T := diag\left(1, \omega^{-1}T^{-1/2}\right)$, and so $\delta_T z_{\lfloor sT \rfloor} \to (1, \eta(s))' =: Z(s)$ and, hence, $\breve{Z}(s) = Z(g(s)) = (1, \eta(g(s)))' = (1, s)'$. Consequently, (20) reduces to $F_{B|Z}(s) := B(s) - Z(s)' \left(\int_0^1 Z(r) Z(r)' dr\right)^{-1} \int_0^1 Z(r) B(r) dr$, $Z(s) := (1, s)'$, so that the asymptotic null distributions of the statistics in this case are defined in terms of a (pivotal) de-trended Brownian motion.

Remark 7.5 The vector γ can also be estimated by applying pseudo-GLS (as in Elliott *et al.*, 1996) to the time-changed variables. For example, when $z_t := (1, t)'$ and pseudo-GLS detrending is used, it is straightforward to show that Theorem 6 holds for $F_{B|\breve{Z}}(s) := B(s) - g(s)\left[\lambda_{\bar{c}} B(1) + 3(1 - \lambda_{\bar{c}})\left(\int_0^1 g(r) B(r)\, dr\right)^{-1}\right]$, $\lambda_{\bar{c}} := (1 + \bar{c})/(1 + \bar{c} + \bar{c}^2/3)$, $\bar{c} \geq 0$. Similarly, using pseudo-GLS de-meaning for $z_t = 1$, $F_{B|\breve{Z}}(s)$ reduces to $B(s)$, so that the resulting M^H tests in this case will have the same limiting null distributions as given in Theorem 2. In this case, these tests are also asymptotically equivalent to running the tests on the deviations from the initial value, $\{X_t - X_0\}$.

We conclude this section by reporting small sample size results for the M^H and M tests based on linearly de-trended data (so that $z_t = (1,\ t)'$) for the four DGPs discussed in Section 6.4 above.[13] These results are reported in Table 10. The M^H tests were run using 0.05 level asymptotic critical values, obtained by direct simulation of the limiting functionals in Theorem 6, using the estimated inverse variance profile, $\widehat{g}(s)$, for T steps and $1,000$ replications. Exact critical values are again used in the case of the M tests. The reported size and power results were again based on $10,000$ replications.

Table 10 about here.

The most obvious feature of Table 10 is that the M^H tests completely avoid the severe over-rejections of the null often seen with the standard M^H tests in the linear trend case. However, and particularly so in the case of abrupt variance changes (Models 1 and 2), the M^H tests tend to be rather undersized. Relative to the tests run on raw or de-meaned data, and as might be expected, larger sample sizes are needed in the linear trend case for empirical size to approach the nominal level to a given degree. For Models 3 and 4, $T = 250$ seems to provide reasonable size properties for the tests.

8 Empirical Illustration

In order to illustrate how the proposed tests can be applied in practice, in this section we analyze the postwar U.S. unemployment rate. In the literature it has been documented (see, for example, Warnock and Warnock, 2000) that U.S. unemployment has experienced a significant volatility reduction during the 1980s. Conventional tests for a unit root in the U.S. unemployment rate might, therefore, be affected by the presence non-constant volatility.

[13] Again the full set of results is available on request. We also computed power results for the linear trend case and found the relative behaviour of the tests to be qualitatively similar to (but reduced relative to) that observed for tests run on either raw or de-meaned data (Tables 5 to 8).

We consider the monthly U.S. unemployment rate among adult males observed over the period January, 1950 through August, 1999. These data have recently been analyzed by Caner and Hansen (2001), from where a detailed description of the data can be obtained, over the sub-sample period 1956–1999. The data are plotted in Figure 1, both in levels (Panel (a)) and in first differences (Panel (b)).[14]

Initially, and in order to assess the time-series behaviour of volatility we report in Figure 1, Panel (c), the estimated variance profile of the series considered. Notice that the 45° line corresponds to a constant variance process. The variance profile is estimated under the unit root null hypothesis, although estimates obtained without imposing a unit root (see Remark 4.2), do not change substantially. A graphical inspection of the estimated variance profile shows that it is difficult to identify a single volatility break in US unemployment. More specifically, it appears that the first and third quarters of the sample are characterized by higher volatility than the second and fourth quarters. This pattern can also be seen to some extent in Panel (b) of Figure 1. The U.S. male unemployment rate therefore seems to be better characterized by multiple (smooth or abrupt) volatility shifts than by a single abrupt shift. Consequently, tests such as those proposed by Kim *et al.* (2002), which assume a single abrupt volatility shift, might not be useful in this context.

More formal tests against non-stationary volatility are reported in the first panel of Table 11, where the outcome of the tests of Section 5 is given. The test statistics are based on the estimate of the variance profile in Panel (c) of Figure 1. Moreover, the estimate of λ_v^2 used was obtained by using an autoregressive spectral density estimator based on an $AR(k)$ model fit to the data with number of lags chosen according to the BIC criterion. With the exception of $\mathcal{H}_R$, the tests confirm the visual evidence from Figure 1 and reject the stationary volatility hypothesis at the (asymptotic) 5% significance level.

The outcome of the standard M unit root tests is presented in the second panel of Table 11 using both raw and OLS de-meaned data. The lag truncation, k, used in the long-run variance estimate, $s^2_{AR}(k)$, was chosen according to the MIC criterion of Ng and Perron (2001) with $k \leq \lfloor 12(T/100)^{0.25} \rfloor$, as in Section 6.4. All of the standard tests reject the null hypothesis at the (asymptotic) 5% significance level, with the exception of the $\mathcal{MZ}_{t.1}$ test which rejects the null at the 10% level. Due to the presence of non-stationary volatility, the outcomes of the tests are potentially unreliable. In the third panel of Table 11 we therefore present the outcomes of the heteroskedasticity-robust M^H unit root test statistics of Section 4. These were computed using the estimate of the variance profile from Panel (c) of Figure 1 and with the long-run variance estimate, $s^2_{AR}(k)$, that used for the corresponding standard unit root tests. All of the M^H tests reject the unit root null at the (asymptotic) 5% significance level. Hence, the results obtained using the M^H tests of Section 4 suggest that the rejection of the unit root null hypothesis in the postwar U.S. unemployment rate obtained from standard unit root tests, which do not control for the observed non-stationary volatility in the series, appears safe.

[14]Since the unemployment rate is bounded between 0 and 100, the empirical analysis has also been carried out using alternative specifications of the dependent variable. Specifically, we have found that the results presented in this Section do not substantially change when preliminary transformations of the unemployment rate which are bounded (see Caner and Hansen, 2001, p. 1581) in either or both directions are used.

9 Conclusions

The presence of non-stationary volatility has potentially serious implications for the reliability of unit root tests, effecting tests whose true size can be substantially different from (either above or below) the nominal significance level. To rectify this problem, we have proposed an entirely new approach to testing for a unit root in the presence of non-stationary volatility. Unlike previous tests proposed in the literature, our approach requires only that the innovation volatility is bounded and has a countable number of jumps. Our proposed unit root tests have the considerable advantage that they are not tied to a given parametric model of non-stationary volatility. This is a particularly attractive feature of the tests since 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. Our proposed tests do not require new tables of critical values when the null hypothesis is level-stationarity, while critical values can be easily computed by Monte Carlo simulation for general deterministic time trends.

While we have motivated our approach through the autocorrelation-robust M unit root tests of Perron and Ng (1996), Stock (1999) and Ng and Perron (2001), because of their desirable finite sample properties, the idea underlying our proposed procedure (that is, to correct for heteroskedasticity by applying a proper time-transformation of the time series of interest) can in principle be applied to almost all unit root tests. Moreover, the approach could be extended to the multivariate setting where standard co-integration tests would be similarly modified by constructing them from the appropriately time-transformed data.

A Mathematical Appendix

A.1 Preliminary Lemmata

We first establish some preparatory lemmata which are essential for the proofs of our main theorems.

Lemma 1 *(Weak convergence under a change of time). Let $\mathcal{D}^0$ denote the space of those elements of $\mathcal{D}$ which are nondecreasing and bounded between* 0 *and* 1*. Let $X(\cdot)$, $X_T(\cdot)$ and $g(\cdot)$ be three stochastic processes in $\mathcal{C}$, $\mathcal{D}$ and $\mathcal{C} \cap \mathcal{D}^0$ respectively; moreover let $X^g(s) := X(g(s)) \in \mathcal{D}$. If $X_T(\cdot) \xrightarrow{w} X(\cdot)$, then $X_T^g(\cdot) \xrightarrow{w} X^g(\cdot) := X(g(\cdot))$. Furthermore, for $g_T(\cdot) \in \mathcal{D}^0$ such that $g_T(\cdot) \xrightarrow{w} g(\cdot)$, then $X_T^{g_T}(\cdot) := X_T(g_T(\cdot)) \xrightarrow{w} X^g(\cdot)$.*

Proof. The proof follows straightforwardly from Billingsley (1968), pp. 144–5. □

Lemma 2 *(SLLN for weighted iid triangular arrays). Let $\eta_t \sim iid(0,1)$ and $\{a_{t,T}\}$ a non-stochastic triangular array with $0 < |a_{t,T}| \leq a_U < \infty$, all t, T. Then, the weighted average $T^{-1}\sum_{t=1}^{T} a_{t,T}\eta_t$ converges in L^2.*

Proof. Convergence in L^2 holds since as $E(T^{-1}\sum_{t=1}^{T} a_{t,T}\eta_t)^2 \leq T^{-1}a_U^2 E\left(\eta_t^2\right) \to 0$ as $T \to \infty$. □

Lemma 3 *If $\mathcal{A}_3$ holds, as $T \to \infty$, $T^{-1}\sum_{t=k}^{T} \sigma_t\sigma_{t-k} - T^{-1}\sum_{t=k}^{T} \sigma_t^2 \to 0$ for any $k = o(T)$.*

PROOF. For $k = o(T)$

$$
\begin{aligned}
\left|\frac{1}{T}\sum_{t=k}^{T}\left(\sigma_t\sigma_{t-k}-\sigma_t^2\right)\right| &= \left|\frac{1}{T}\sum_{t=k}^{T}\left(\sigma_t-\sigma_{t-k}\right)\sigma_t\right| \\
&\le \frac{1}{T}\sum_{t=k}^{T}\left|\sigma_t-\sigma_{t-k}\right|\sigma_t \\
&\le \left(\max_{t=1,\dots,T}\sigma_t\right)\frac{1}{T}\sum_{t=k}^{T}\left|\sigma_t-\sigma_{t-k}\right| \\
&\le \left(\sup_{s\in[0,1]}\omega(s)\right)\frac{1}{T}\sum_{t=0}^{T-|k|}\left|\omega\left(\frac{t}{T}+\frac{|k|}{T}\right)-\omega\left(\frac{t}{T}\right)\right|
\end{aligned}
$$

It is straightforwardly seen that the last term can be made arbitrarily small. Specifically, since $\omega(\cdot) \in \mathcal{D}$ then for every $\varepsilon > 0$ there exists a positive constant δ such that $|\omega(x) - \omega(x+\delta)| < \varepsilon$. Since $|k| = o(T)$, for T sufficiently large $|k|/T < \delta$ so that $T^{-1}\sum_{t=0}^{T-|k|}|\omega(t/T + |k|/T) - \omega(t/T)| \le T^{-1}(T-|k|)\varepsilon \le \varepsilon$. □

Lemma 4 (SLLN FOR HETEROSKEDASTIC ARRAYS). *Let* $u_t := C(L)\varepsilon_t = C(L)\sigma_t e_t$ *and* $\widetilde{u}_t := \omega C(L)e_t$.[15] *Then, under Assumptions* $\mathcal{A}_1$, $\mathcal{A}_3$, *as* $T \to \infty$ (i) $T^{-1}\sum_{t=1}^{T}u_t - T^{-1}\sum_{t=1}^{T}\widetilde{u}_t \xrightarrow{L^2} 0$ *and* (ii) $T^{-1}\sum_{t=k}^{T}u_t u_{t-k} - T^{-1}\sum_{t=k}^{T}\widetilde{u}_t\widetilde{u}_{t-k} \xrightarrow{L^2} 0$ *when* k *is* $o(T)$.

PROOF. Part (i). By Assumption $\mathcal{A}_1$, the standard BN decomposition for linear processes, see Phillips and Solo (1992) [PS], implies that $u_t = C(1)\varepsilon_t + \widetilde{\varepsilon}_{t-1} - \widetilde{\varepsilon}_t$, $\widetilde{\varepsilon}_t := \widetilde{C}(L)\varepsilon_t = \sum_{j=0}^{\infty}\widetilde{c}_j\varepsilon_{t-j}$, where the condition $\sum_{j=0}^{\infty}j|c_j| < \infty$ ensures that $\sum_{j=0}^{\infty}\widetilde{c}_j^2 < \infty$. Similarly, $\widetilde{u}_t = \omega C(1)e_t + \omega\widetilde{e}_{t-1} - \omega\widetilde{e}_t$, $\widetilde{e}_t := \widetilde{C}(L)e_t = \sum_{j=0}^{\infty}\widetilde{c}_j e_{t-j}$. It then follows that

$$
\frac{1}{T}\sum_{t=1}^{T}u_t - \frac{1}{T}\sum_{t=1}^{T}\widetilde{u}_t = \frac{C(1)}{T}\sum_{t=1}^{T}\left(\varepsilon_t - \omega e_t\right) + \frac{1}{T}\left(\widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_T + \omega\widetilde{e}_0 - \omega\widetilde{e}_T\right). \tag{21}
$$

As $E(\widetilde{e}_t^2) = \sum_{j=0}^{\infty}\widetilde{c}_j^2 < \infty$, all t, both $T^{-1}\widetilde{e}_0$ and $T^{-1}\widetilde{e}_T$ (and hence $T^{-1}(\widetilde{e}_0 - \widetilde{e}_T)$ for the triangle inequality) converge in L^2 to zero; the same result holds for $T^{-1}(\widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_T)$ since $E(T^{-1}\widetilde{\varepsilon}_t)^2 = T^{-2}\sum_{j=0}^{\infty}\widetilde{c}_{t-j}^2\sigma_{t-j}^2 E(e_{t-j}^2) = T^{-2}\sum_{j=0}^{t}\widetilde{c}_{t-j}^2\sigma_{t-j}^2 + T^{-2}\sum_{j=1}^{\infty}\widetilde{c}_{t-j}^2\sigma_{-j}^2 \le T^{-2}(\max_t\sigma_t^2\sum_{j=0}^{T}\widetilde{c}_{t-j}^2 + (\sigma^*)^2\sum_{j=1}^{\infty}\widetilde{c}_{t-j}^2) \to 0$ since the $\widetilde{c}_t$'s are square summable and $\{\sigma_t\}$ is bounded. Regarding the first term in the right member of (21), since $\sum_{t=1}^{T}(\varepsilon_t - \omega e_t) = \sum_{t=1}^{T}(\sigma_t - \omega)e_t$, where $\sigma_t - \omega$ is bounded for all t, Lemma 2 implies L^2 convergence to zero.

Part (ii). For space constraints we consider only the case of $k = 0$; the proof extends to the general case straightforwardly. To prove the Lemma we make extensive use of the BN decomposition for squared linear processes introduced by PS. Specifically, $u_t^2 = X_t^a + 2X_t^b$, where

$$
\begin{aligned}
X_t^a &:= f_0(1)\varepsilon_t^2 + Y_{t-1}^0 - Y_t^0,\ Y_t^0 := \widetilde{f}_0(L)\varepsilon_t^2 = \sum_{j=0}^{\infty}\widetilde{f}_{0j}\varepsilon_{t-j}^2 \\
X_t^b &:= \sum_{r=1}^{\infty}f_r(1)\varepsilon_t\varepsilon_{t-r} + \sum_{r=1}^{\infty}(Y_{t-1}^r - Y_t^r),\ Y_t^r := \widetilde{f}_r(L)\varepsilon_t\varepsilon_{t-r} = \sum_{j=0}^{\infty}\widetilde{f}_{rj}\varepsilon_{t-r}\varepsilon_{t-r-j}
\end{aligned}
$$

[15] Note that $\widetilde{u}_t$ is linear in $\omega e_t, \omega e_{t-1}, \dots$ while u_t is linear in $\sigma_t e_t, \sigma_{t-1}e_{t-1}, \dots$

where $\sum_{j=0}^{\infty} jc_j^2 < \infty$ implies $\sum_{j=0}^{\infty} \widetilde{f}_{rj}^2 < \infty$, all r, see PS, p. 979. The corresponding decomposition for $\widetilde{u}_t^2$ is $\widetilde{u}_t^2 = \widetilde{X}_t^a + 2\widetilde{X}_t^b$, where $\widetilde{X}_t^a, \widetilde{X}_t^b$ are defined as X_t^a, X_t^b but with ωe_t replacing ε_t. It then follows that

$$\frac{1}{T}\sum_{t=1}^{T} u_t^2 - \frac{1}{T}\sum_{t=1}^{T} \widetilde{u}_t^2 = \frac{1}{T}\sum_{t=1}^{T}(X_t^a - \widetilde{X}_t^a) + \frac{2}{T}\sum_{t=1}^{T}(X_t^b - \widetilde{X}_t^b).$$

We now show that $T^{-1}\sum_{t=1}^{T}(X_t^a - \widetilde{X}_t^a)$ converges to zero in L^2; in the same manner it can be shown that $T^{-1}\sum_{t=1}^{T}(X_t^b - \widetilde{X}_t^b)$ also converges to zero. First, simple substitutions lead to the equality $T^{-1}\sum_{t=1}^{T}(X_t^a - \widetilde{X}_t^a) = f_0(1)A_T^1 + A_T^2$, $A_T^1 := T^{-1}\sum_{t=1}^{T}(\varepsilon_t^2 - \omega^2 e_t^2)$ and $A_T^2 := T^{-1}(Y_0^0 - Y_T^0 + \widetilde{Y}_0^0 - \widetilde{Y}_T^0)$. Letting $\zeta_t := e_t^2 - 1$, we have that

$$A_T^1 = \frac{1}{T}\sum_{t=1}^{T}(\sigma_t^2 e_t^2 - \omega^2 e_t^2) = \frac{1}{T}\sum_{t=1}^{T}(\sigma_t^2 - \omega^2)\zeta_t + \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 - \omega^2.$$

As $T^{-1}\sum_{t=1}^{T}\sigma_t^2 \to \omega^2$ and $T^{-1}\sum_{t=1}^{T}(\sigma_t^2 - \omega^2)\zeta_t \xrightarrow{L^2} 0$, see Lemma 2, $A_T^1 \xrightarrow{L^2} 0$. We now show L^2 convergence for A_T^2. For all t, $Y_t^0 - \widetilde{Y}_t^0 = \sum_{j=0}^{\infty} f_{0j}(\sigma_{t-j}^2 - \omega^2)e_{t-j}^2$ and, hence,

$$E\left(\frac{1}{T}(Y_t^0 - \widetilde{Y}_t^0)\right)^2 = \frac{1}{T^2}\sum_{j=0}^{\infty}\sum_{j'=0}^{\infty}\widetilde{f}_{0j}\widetilde{f}_{0j'}(\sigma_{t-j}^2 - \omega^2)(\sigma_{t-j'}^2 - \omega^2) \leq \frac{4\max_{t\leq T}\sigma_t^4}{T^2}\left(\sum_{j=0}^{\infty}|\widetilde{f}_{0j}|\right)^2.$$

Since both $\{\sigma_t\}$ and $\sum_{j=0}^{\infty}|\widetilde{f}_{0j}|$ are bounded (see PS, p. 991), L^2 convergence of A_T^2 follows immediately. □

Lemma 5 (IP for heteroskedastic independent arrays). *Let* $Y_T(s) := T^{-1/2}\sum_{t=1}^{\lfloor sT\rfloor}\sigma_t e_t$, *where* $\{\sigma_t\}$ *and* $\{e_t\}$ *satisfy Assumptions* $\mathcal{A}_1$ *and* $\mathcal{A}_3$. *Then, as* $T\to\infty$, $Y_T(\cdot) \xrightarrow{w} \omega B_\eta(\cdot)$, *where* $B_\eta(\cdot)$ *is a variance-transformed Brownian motion with directing process* $\eta(s) := (1/\omega^2)\int_0^s \omega(r)^2 dr$.

Proof. Since the e_t's are independent with bounded fourth moments, $T^{-1}\sum_{t=\lfloor sT\rfloor}^{\lfloor s'T\rfloor}\sigma_t^2$ is bounded away from zero uniformly for all s, s' $(s < s')$, $E(Y(s)) = 0$ and $E(Y(s)^2) \to \eta(s)^2$, convergence of the finite dimensional distributions follows straightly from a standard (Liapunov) CLT for heterogeneous independent random variables. It is left to prove that the sequence of probability measures associated to $Y(\cdot)$ is tight. A sufficient condition for tightness (see Theorem 15.6 of Billingsley, 1968) is that $E((Y(s) - Y(s_1))^2(Y(s_2) - Y(s_1))^2) \leq K(s_2 - s_1)^2, s_1 \leq s \leq s_2$ for some finite constant K. But since $Y(\cdot)$ has independent increments, the left member of the previous inequality is $(\omega^2 T)^{-2}\sum_{t=[s_1T]+1}^{[sT]}\sigma_t^2\sum_{t=[sT]+1}^{[s_2T]}\sigma_t^2 \leq KT^{-2}([sT] - [s_1T])([s_2T] - [sT]) \leq K\left(T^{-1}[s_2T] - T^{-1}[s_1T]\right)^2$, $K := \sup_r \omega(r)^4/\omega^4 < \infty$, which proves the desired result (see Billingsley, 1968, p. 138).

A.2 Proofs of the Main Theorems

Proof of Theorem 1. The proof follows by adapting Cavaliere (2004), Theorem 2, to the case of linear processes. First, we prove the result in part (i) that $X_T(s) := T^{-1/2}X_{\lfloor sT\rfloor}$

weakly converges to $\omega C(1)B_{\eta}(s)$. Using the BN decomposition, we have that

$$X_T(s) = \frac{1}{T^{1/2}}X_0 + S_T(s),\ S_T(s) := \frac{1}{T^{1/2}}\sum_{t=1}^{\lfloor sT\rfloor} u_t = \frac{C(1)}{T^{1/2}}\sum_{t=1}^{\lfloor sT\rfloor}\sigma_t e_t + \frac{1}{T^{1/2}}\left(\widetilde{\varepsilon}_0 - \widetilde{\varepsilon}_{\lfloor sT\rfloor}\right)$$

where $\widetilde{\varepsilon}_t$ is defined in the proof of Lemma 4. By Lemma 5 and Theorem 4.1 in Billingsley (1968), $S_T(\cdot) \xrightarrow{w} \omega C(1)B_{\eta}(\cdot)$ provided $\max_{0\leq t\leq T}\left(T^{-1/2}\widetilde{\varepsilon}_t\right)$ is of $o_p(1)$. To show this property it suffices to prove finiteness of the fourth moments of $\widetilde{\varepsilon}_t$, all t, since for any $\delta > 0$, the Bonferroni and Markov inequalities guarantee that

$$\Pr\left(\max_{0\leq t\leq T} T^{-1/2}|\widetilde{\varepsilon}_t| > \delta\right) \leq \sum_{t=1}^{T}\Pr\left(|\widetilde{\varepsilon}_t| > \delta T^{1/2}\right) \leq \sum_{t=1}^{T}\frac{E(\widetilde{\varepsilon}_t^4)}{\delta^4 T^2} \leq \frac{\sup_{0\leq t\leq T}E(\widetilde{\varepsilon}_t^4)}{\delta^4 T} \to 0$$

if $\sup_t E(\widetilde{\varepsilon}_t^4)$ is bounded. But this is certainly true as (see, for example, Tanaka, 1996, pp. 501–2)

$$|\widetilde{\varepsilon}_t| = \left|\sum_{j=0}^{\infty}\widetilde{c}_j\sigma_{t-j}e_{t-j}\right| \leq \sum_{j=0}^{\infty}|\widetilde{c}_j|^{3/4}\left(|\widetilde{c}_j|^{1/4}\sigma_{t-j}|e_{t-j}|\right) \leq \max_t \sigma_t\left(\sum_{j=0}^{\infty}|\widetilde{c}_j|\right)^{3/4}\left(\sum_{j=0}^{\infty}|\widetilde{c}_j||e_{t-j}|^4\right)^{1/4}$$

implies $\sup_{0\leq t\leq T}E(\widetilde{\varepsilon}_t^4) \leq \max_t \sigma_t\left(\sum_{j=0}^{\infty}|\widetilde{c}_j|\right)^4 E\left(e_t^4\right) < \infty$ (under Assumption $\mathcal{A}_1$, $\sum_{j=0}^{\infty}|\widetilde{c}_j| < \infty$, see PS, and $E\left(e_t^4\right) < \infty$, all t). As X_0 is of $O_p(1)$, we have also established that $X_T(\cdot) \xrightarrow{w} \omega C(1)B_{\eta}(\cdot)$ and, by a standard continuous mapping argument, that $(T^{-1}X_T^2, T^{-2}\sum_{t=1}^{T}X_{t-1}^2) \xrightarrow{w} \omega^2 C(1)^2(B_{\eta}(1), \int_0^1 B_{\eta}(s)^2 ds)$.

We now establish part (ii) of the Theorem. Consider first $s_{AR}^2(k)$ and suppose, initially, that $s_{AR}^2(k)$ is computed under the null hypothesis, so that it is based on $\Delta X_t = u_t$; we denote this estimator by $\widetilde{s}_{AR}^2(k)$. Later we will consider the AR estimator obtained by fitting the regression $\widehat{u}_t = \sum_{i=1}^{k}\beta_i\widehat{u}_{t-i} + e_t$; the same result will also hold for the AR estimator obtained from the regression equation $\Delta X_t = \pi X_{t-1} + \sum_{i=1}^{k}\beta_i\Delta X_{t-i} + e_t$. Let

$$\begin{aligned} x_{t,k} &:= (u_{t-1}, ..., u_{t-k})' \\ \widetilde{x}_{t,k} &:= (\widetilde{u}_{t-1}, ..., \widetilde{u}_{t-k})' \end{aligned}$$

and $\widehat{M}_{xx,k} := T^{-1}\sum_t x_{t,k}x_{t,k}'$, $\widehat{M}_{xu,k} := T^{-1}\sum_t x_{t,k}u_t$, $\widehat{M}_{\widetilde{x}\widetilde{x},k} := T^{-1}\sum_t \widetilde{x}_{t,k}\widetilde{x}_{t,k}'$ and $\widehat{M}_{\widetilde{x}\widetilde{u},k} := T^{-1}\sum_t \widetilde{x}_{t,k}\widetilde{u}_t$. The AR estimator of $1/C(1)$ based on $\Delta X_t = u_t$ is then given by $\widehat{\beta}(1) := 1 - \iota_k'\widehat{\beta}_k$, $\widehat{\beta}_k := \widehat{M}_{xx,k}^{-1}\widehat{M}_{xu,k}$, while the AR estimator based on $\{\widetilde{u}_t\}$ is given by $\widetilde{\beta}(1) := 1 - \iota_k'\widetilde{\beta}_k$, $\widetilde{\beta}_k := \widehat{M}_{\widetilde{x}\widetilde{x},k}^{-1}\widehat{M}_{\widetilde{x}\widetilde{u},k}$, ι_k denoting the unit vector in $\mathbb{R}^k$. It is possible to show that the two estimators are asymptotically equivalent. To show this property, let $A_1 := ||\widehat{M}_{xx,k}^{-1} - \widehat{M}_{\widetilde{x}\widetilde{x},k}^{-1}||$, $A_2 := ||\widehat{M}_{xu,k} - \widehat{M}_{\widetilde{x}\widetilde{u},k}||$, $B_1 := ||\widehat{M}_{\widetilde{x}\widetilde{u},k}||$, $B_2 := ||\widehat{M}_{xx,k}^{-1}||$ and note that

$$|\widehat{\beta}(1) - \widetilde{\beta}(1)| = |\iota_k'\widehat{\beta}_k - \iota_k'\widetilde{\beta}_k| \leq k^{1/2}||\widehat{\beta}_k - \widetilde{\beta}_k|| \leq k^{1/2}(A_1B_1 + B_2A_2) \tag{22}$$

where $\|\cdot\|$ denotes the (matrix) Euclidean norm.[16] We therefore need to prove that the right member of (22) is of $o_p(1)$. First, note that B_1 is $O_p(1)$, as $B_1 \leq ||E(\widehat{M}_{\widetilde{x}\widetilde{u},k})|| +$

[16] That is for a matrix A, $\|A\| := \sup_{||x||<1}||Ax||$, using the Euclidean norm for the vector x, i.e. $||x||^2 := x'x$.

$||\widehat{M}_{\widetilde{x}\widetilde{u},k} - E(\widehat{M}_{\widetilde{x}\widetilde{u},k})||$, where $||E(\widehat{M}_{\widetilde{x}\widetilde{u},k})||^2 \leq \sum_{j=1}^{\infty} \left(E(\widetilde{u}_t\widetilde{u}_{t+j})\right)^2 \leq \sum_{j=0}^{\infty}(\sum_{i=0}^{\infty} c_i c_{i+j})^2 < \infty$ (as implied by the summability condition in $\mathcal{A}_1$, see PS) and $||\widehat{M}_{\widetilde{x}\widetilde{u},k} - E(\widehat{M}_{\widetilde{x}\widetilde{u},k})|| = o_p(1)$ if $k/T \to 0$, see Berk (1974), eq. (2.20). Second, from Berk (1974) it follows that A_1 if of the same order as $||\widehat{M}_{xx,k} - \widehat{M}_{\widetilde{x}\widetilde{x},k}||$, provided that $B_3 := ||\widehat{M}^{-1}_{\widetilde{x}\widetilde{x},k}||$ is bounded in probability, which is indeed the case; see, Chang and Park (2002, Lemma 3.2a). Now consider the inequality

$$E\left\|\widehat{M}_{xx,k} - \widehat{M}_{\widetilde{x}\widetilde{x},k}\right\|^2 \leq k^2 \max_{0\leq i,j\leq k} E\left|\frac{1}{T}\sum_{t=k}^{T}(u_{t-i}u_{t-j} - \widetilde{u}_{t-i}\widetilde{u}_{t-j})\right|^2.$$

Since, for all i, j, $E\left|T^{-1}\sum_{t=k}^{T}(u_{t-i}u_{t-j} - \widetilde{u}_{t-i}\widetilde{u}_{t-j})\right|^2$ is of $O\left(T^{-1}\right)$, see Lemma 4, we have that $kE(A_1^2) \to 0$ as $k^3/T \to 0$ (Assumption $\mathcal{K}$). Under this condition, $k^{1/2}A_1B_1 \xrightarrow{p} 0$. Third, in a similar way it can be proved that $kA_2 \xrightarrow{p} 0$ if $k^3/T \to 0$, and hence (by referring to the inequality $B_2 \leq A_1 + B_3$) it easily follows that $k^{1/2}A_2B_2$ also converges in probability to zero.

Taken together, these results imply that $k^{1/2}(A_1B_1 + B_2A_2)$ and, hence, that $\widehat{\beta}(1) - \widetilde{\beta}(1)$ converges in probability to zero. Consequently,

$$(1 - \widehat{\beta}(1))^2 \xrightarrow{p} 1/C(1)^2 \tag{23}$$

since $(1 - \widetilde{\beta}(1))^2 \xrightarrow{p} 1/C(1)^2$ (see Chang and Park, 2002, Lemma 3.4, and the subsequent discussion).

Now, consider $\widehat{\sigma}^2 = (T-k)^{-1}\sum_{t=k+1}^{T}(u_t - \widehat{\beta}_k' x_{t,k})^2$ and set $\widetilde{\sigma}^2 = (T-k)^{-1}\sum(\widetilde{u}_t - \widetilde{\beta}_k'\widetilde{x}_{t,k})^2$, where $\widetilde{\sigma}^2 = \omega^2 + o_p(1)$ (Chang and Park, 2002, Lemma 3.3). Similarly to what was shown for $\widehat{\beta}(1)$, it can be proved that $\widehat{\sigma}^2 - \widetilde{\sigma}^2 \xrightarrow{p} 0$. To prove this result, notice first that simple computations allow us to write

$$\begin{aligned}
\widehat{\sigma}^2 - \widetilde{\sigma}^2 &= \frac{1}{T-k}\sum_{t=k+1}^{T}(u_t^2 - \widetilde{u}_t^2) - (\widehat{M}_{ux,k}\widehat{\beta}_k - \widehat{M}_{\widetilde{u}\widetilde{x},k}\widetilde{\beta}_k) \\
&= \frac{1}{T-k}\sum_{t=k+1}^{T}(u_t^2 - \widetilde{u}_t^2) - (\widehat{M}_{ux,k} - \widehat{M}_{\widetilde{u}\widetilde{x},k})(\widehat{\beta}_k - \widetilde{\beta}_k) \qquad (24) \\
&\quad -(\widehat{M}_{ux,k} - \widehat{M}_{\widetilde{u}\widetilde{x},k})\widetilde{\beta}_k - \widehat{M}_{\widetilde{u}\widetilde{x},k}(\widehat{\beta}_k - \widetilde{\beta}_k).
\end{aligned}$$

Using Lemma 4 and the above results it is straightforward to see that the four terms constituting the right member of (24) all converge in probability to zero. This result, taken together with the convergence result in (23), ensures that $\widetilde{s}^2_{AR}(k) \xrightarrow{p} \omega^2 C(1)^2$.

Finally, tedious but straightforward computations based on the above results allow us to show that the difference between $\widehat{\beta}_k$ and the estimator based on the residuals $\{\widehat{u}_t\}$ of the (OLS) projection of X_t on X_{t-1}, say $\widehat{\beta}_k^* := \widehat{M}^{-1}_{\widehat{x}\widehat{x},k}\widehat{M}_{\widehat{x}\widehat{u},k}$ ($\widehat{M}_{\widehat{x}\widehat{x},k} := T^{-1}\sum_t \widehat{x}_{t,k}\widehat{x}_{t,k}'$, $\widehat{M}_{\widehat{x}\widehat{u},k} := T^{-1}\sum_t \widehat{x}_{t,k}'\widehat{u}_t$, with $\widehat{x}_{t,k} := (\widehat{u}_{t-1}, ..., \widehat{u}_{t-k})'$), satisfies $k^{1/2}||\widehat{\beta}_k^* - \widehat{\beta}_k|| = o_p(1)$ under Assumption $\mathcal{K}$, that $(T-k)^{-1}\sum_{t=k+1}^{T}(\widehat{u}_t^2 - u_t^2) = o_p(1)$ and, hence, that the AR estimator of $\omega^2 C(1)^2$ based on the $\widehat{u}_t$'s, $s^2_{AR}(k)$, is consistent.

Proof of Theorem 2. Part (i). As $\widehat{\eta}(\cdot)$ and $\eta(\cdot)$ are both continuous, bounded and monotone, in order to prove the uniform convergence of $\widehat{\eta}(\cdot)$ to $\eta(\cdot)$ it is sufficient to establish pointwise convergence. Initially, consider the estimator of $\eta(\cdot)$ based on the error term $\{u_t\}$:

$$\widetilde{\eta}(s) := \frac{\sum_{t=1}^{\lfloor sT \rfloor} u_t^2 + (sT - \lfloor sT \rfloor) u_{\lfloor sT \rfloor + 1}^2}{\sum_{t=1}^{T} u_t^2}$$

with $\widetilde{\eta}(0) := 0$ and $\widetilde{\eta}(1) := 1$. Now, since $\sup_{s\in[0,1]} |\widetilde{\eta}(s) - (\sum_{t=1}^{T} u_t^2)^{-1} \sum_{t=1}^{\lfloor sT \rfloor} u_t^2| \leq (T^{-1}\sum_{t=1}^{T} u_t^2)^{-1} \max_{1\leq t\leq T} T^{-1}u_t^2$, where $\max_{1\leq t\leq T} T^{-1}u_t^2$ is of $o_p(1)$ (see the proof of Lemma 4), to prove that $\widetilde{\eta}(s) \xrightarrow{p} \eta(s)$ pointwise we need to show that

$$\frac{1}{T}\sum_{t=1}^{\lfloor sT \rfloor} u_t^2 - E\left(\frac{1}{T}\sum_{t=1}^{\lfloor sT \rfloor} u_t^2\right) \xrightarrow{p} 0 \tag{25}$$

and that

$$E\left(\frac{1}{T}\sum_{t=1}^{\lfloor sT \rfloor} u_t^2\right) \to \int_0^s \omega(r)^2 \, dr \tag{26}$$

uniformly. But (25) follows simply by modifying Lemma 4 to account for the fact that the summation is now taken over $t = 1, ..., \lfloor sT \rfloor$, rather than $t = 1, ..., T$, while (26) holds since $E(T^{-1}\sum_{t=1}^{\lfloor sT \rfloor} u_t^2) = T^{-1}\sum_{t=1}^{\lfloor sT \rfloor} \omega(t/T)^2 \to \int_0^s \omega(r)^2 \, dr$ as $\omega(\cdot)$ is square integrable on all subintervals of $[0, 1]$.

In order to prove the consistency of $\widehat{g}(s)$, it suffices to notice that if $\widehat{\eta}(s) \xrightarrow{p} \eta(s)$ uniformly for all s, and if $\widehat{g}(s)$ and $g(s)$ are the unique inverse functions of $\widehat{\eta}(s)$ and $\eta(s)$, respectively, then it necessarily holds that $\widehat{g}(s) \xrightarrow{p} g(s)$. To see this, suppose that, for some s, $\widehat{g}(s) \xrightarrow{p} g^*(s) \neq g(s)$. Then, the CMT would imply that $\widehat{\eta}(\widehat{g}(s)) \xrightarrow{p} \eta(g^*(s)) \neq s$. But since $\widehat{g}(s)$ is also the unique inverse of $\widehat{\eta}(s)$, i.e. $\widehat{\eta}(\widehat{g}(s)) = s$, this contradicts the previous result.

We conclude by showing that the foregoing results continue to hold if $\{u_t^2\}$ are replaced by the OLS residuals, $\{\widehat{u}_t^2\}$. Specifically, since under the conditions of the Theorem $\widehat{u}_t := X_t - \widehat{\alpha}X_{t-1} = u_t - (\widehat{\alpha} - 1)X_{t-1}$, it holds that $\widehat{u}_t^2 = u_t^2 + (\widehat{\alpha} - 1)^2 X_{t-1}^2 - 2(\widehat{\alpha} - 1)u_t X_{t-1}$, so that

$$\sup_{s\in[0,1]} \left| \frac{1}{T}\sum_{t=1}^{\lfloor sT \rfloor} \widehat{u}_t^2 - \frac{1}{T}\sum_{t=1}^{\lfloor sT \rfloor} u_t^2 \right| \leq T(\widehat{\alpha}-1)^2 \sup_{s\in[0,1]} \left| \frac{1}{T^2}\sum_{t=1}^{\lfloor sT \rfloor} X_{t-1}^2 \right| + 2|\widehat{\alpha}-1| \sup_{s\in[0,1]} \left| \frac{1}{T}\sum_{t=1}^{\lfloor sT \rfloor} u_t X_{t-1} \right|.$$

From Theorem 1 and the CMT we have that $\widehat{\alpha}-1$ is of $O_p(T^{-1})$, $\sup_{s\in[0,1]} |T^{-2}\sum_{t=1}^{\lfloor sT \rfloor} X_{t-1}^2|$ is of $O_p(1)$ and $\sup_{s\in[0,1]} |T^{-1}\sum_{t=1}^{\lfloor sT \rfloor} u_t X_{t-1}|$ is of $O_p(1)$. Consequently, $\sup_{s\in[0,1]} |T^{-1}\sum_{t=1}^{\lfloor sT \rfloor} \widehat{u}_t^2 - T^{-1}\sum_{t=1}^{\lfloor sT \rfloor} u_t^2|$ is of $o_p(1)$.

Part (ii). We first show that under Assumption $\mathcal{A}$, if $\alpha = 1$ then

$$X_T^{\widehat{g}}(\cdot) := T^{-1/2} X_{\lfloor \widehat{g}(\frac{\lfloor \cdot T \rfloor}{T})T \rfloor} \xrightarrow{w} \omega C(1) B(\cdot). \tag{27}$$

Recall from Theorem 1 that $X_T(\cdot) := T^{-1/2} X_{\lfloor \cdot T \rfloor} \in \mathcal{D}$ converges weakly to $B_\eta(\cdot) \in \mathcal{C}$, and consider the approximant $X_T^{g_T}(\cdot) := T^{-1/2} X_{\lfloor \widehat{g}(\cdot) T \rfloor} \in \mathcal{D}$, where $\widehat{g}(\cdot) \in \mathcal{D}^0$. As $\widehat{g}(\cdot) \to g(\cdot) \in \mathcal{C} \cap \mathcal{D}^0$ uniformly in probability, it also weakly converges. Consequently, all of the conditions of Lemma 1 are satisfied and we therefore obtain that $X_T^{g_T}(s) = X_T(\widehat{g}(s)) \xrightarrow{w}$

$\omega C(1) B_\eta (g(s)) = \omega C(1) B(\eta(g(s))) = \omega C(1) B(s)$. It remains to be shown that the same result holds for $X_T^{\widehat{g}}(s) := T^{-1/2} X_{\lfloor \widehat{g}(\lfloor sT \rfloor/T)T \rfloor}$. From the uniform continuity of $\widehat{g}(\cdot)$ it follows that $\widehat{g}([sT]/T) - \widehat{g}(s)$ converges uniformly in probability to zero, from which the desired result follows using Theorem 4.1 of Billingsley (1968).

As a consequence of (27) and the CMT, $T^{-1} X_T^2 = X_T^{\widehat{g}}(1)^2 \xrightarrow{w} \omega^2 C(1)^2 B_\eta(g(1))^2 = \omega^2 C(1)^2 B(1)^2$ and

$$\begin{aligned}
\frac{1}{T^2} \sum_{t=1}^{T} X^2_{\lfloor \widehat{g}(\frac{t}{T}) \cdot T \rfloor} &= \sum_{t=1}^{T} \int_{\frac{t-1}{T}}^{\frac{t}{T}} X_T^{\widehat{g}}(s)^2 \, ds + \frac{X_T^2}{T^2} = \int_0^1 X_T^{\widehat{g}}(s)^2 \, ds + O_p\left(T^{-1}\right) \\
&\xrightarrow{w} \omega^2 C(1)^2 \int_0^1 B_\eta(g(s))^2 \, ds = \omega^2 C(1)^2 \int_0^1 B(s)^2 \, ds \, .
\end{aligned}$$

Finally, since by Theorem 1 $s^2_{AR}(k)$ converges in probability to $\omega^2 C(1)^2$, the desired result follows immediately.

Proof of Theorem 4. Consider the statistic $\widehat{W}_T(s) := T^{1/2} \widehat{W}(s) = T^{1/2}(\widehat{\eta}(s) - s)$. Simple algebra yields that

$$\begin{aligned}
\widehat{\omega}^2 \widehat{W}_T(s) &= \frac{1}{T^{1/2}} \left(\sum_{t=1}^{\lfloor sT \rfloor} \widehat{\upsilon}_t - s \sum_{t=1}^{T} \widehat{\upsilon}_t \right) + \frac{1}{T^{1/2}} (sT - \lfloor sT \rfloor) \widehat{\upsilon}_{\lfloor sT \rfloor + 1} \qquad (28) \\
&= \widehat{V}_T(s) + \frac{1}{T^{1/2}} (sT - \lfloor sT \rfloor) \widehat{\upsilon}_{\lfloor sT \rfloor + 1}.
\end{aligned}$$

where $\widehat{V}_T(s)$ is defined implicitly. Let us first prove that $\widehat{V}_T(\cdot)$ satisfies an invariance principle. Consider then the idealized functional $V_T(s) := T^{-1/2}(\sum_{t=1}^{\lfloor sT \rfloor} \upsilon_t - s \sum_{t=1}^{T} \upsilon_t) = T^{-1/2}(\sum_{t=1}^{\lfloor sT \rfloor} (\upsilon_t - \omega^2) - s \sum_{t=1}^{T} (\upsilon_t - \omega^2))$ where, under the null hypothesis, $\omega^2 = E(\upsilon_t)$, all t. Under the conditions of the Theorem, we can apply PS, Theorem 3.8, to obtain that $T^{-1/2} \sum_{t=1}^{[\cdot T]} (\upsilon_t - \omega^2) \xrightarrow{w} \lambda_\upsilon B(\cdot)$, where λ_υ is bounded and is also bounded away from zero. By the CMT it then follows that $V_T(s) \xrightarrow{w} \lambda_\upsilon (B(s) - sB(1)) = \lambda_\upsilon V(s)$, where $V(\cdot)$ is a standard Brownian bridge. To see that the same result holds for $\widehat{V}_T(s)$, notice that

$$\sum_{t=1}^{\lfloor sT \rfloor} \widehat{\upsilon}_t - s \sum_{t=1}^{T} \widehat{\upsilon}_t = \sum_{t=1}^{\lfloor sT \rfloor} \upsilon_t - s \sum_{t=1}^{T} \upsilon_t + \left(\sum_{t=1}^{\lfloor sT \rfloor} n_t - s \sum_{t=1}^{T} n_t \right)$$

with $n_t := (\widehat{\alpha} - \alpha)^2 X_{t-1}^2 - 2(\widehat{\alpha} - \alpha) u_t X_{t-1}$. Consequently,

$$\begin{aligned}
\sup_{0 \le s \le 1} |\widehat{V}_T(s) - V_T(s)| &\le \frac{1}{T^{1/2}} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor sT \rfloor} n_t - s \sum_{t=1}^{T} n_t \right| \\
&\le \frac{1}{T^{1/2}} \sup_{s \in [0,1]} \left| \sum_{t=1}^{\lfloor sT \rfloor} n_t \right| + \frac{1}{T^{1/2}} \left| \sum_{t=1}^{T} n_t \right| \\
&\le \frac{2(\widehat{\alpha} - \alpha)^2}{T^{1/2}} \sum_{t=1}^{T} X_{t-1}^2 + \frac{4|\widehat{\alpha} - \alpha|}{T^{1/2}} \sup_{s \in [0,1]} \left| \sum_{t=1}^{[sT]} u_t X_{t-1} \right| \, .
\end{aligned}$$

From Theorem 1 and Remark 3.1, $\alpha = \exp(-c/T)$ $(c \geq 0)$ implies that $\widehat{\alpha} - \alpha$ is of $O_p(T^{-1})$, $\sum_{t=1}^{T} X_{t-1}^2$ is of $O_p(T^2)$ and $\sup_{s\leq 1} |\sum_{t=1}^{[sT]} u_t X_{t-1}|$ is also of $O_p(T)$. Consequently, $\sup_s |\widehat{W}_T(s) - W_T(s)|$ is of $o_p(1)$, and Theorem 4.1 of Billingsley (1968) ensures weak convergence of $\widehat{W}_T(\cdot)$ to $\lambda_v V(\cdot)$.

Consider now the second term in the right member of (28). For all s we have that

$$\begin{aligned}\left|T^{-1/2}(sT - \lfloor sT \rfloor)\widehat{\upsilon}_{\lfloor sT \rfloor+1}\right| &\leq \left|T^{-1/2}(sT - \lfloor sT \rfloor)\upsilon_{\lfloor sT \rfloor+1}\right| + \left|T^{-1/2}(sT - \lfloor sT \rfloor) n_{\lfloor sT \rfloor+1}\right| \\ &\leq T^{-1/2}\max_{1\leq t\leq T}|\upsilon_t| + T^{-1/2}\max_{1\leq t\leq T}|n_t| = o_p(1)\end{aligned}$$

as $\max_{1\leq t\leq T} T^{-1/2}|n_t| = o_p(1)$, which follows from the arguments above, while $T^{-1/2}\max_{1\leq t\leq T}|\upsilon_t| = T^{-1/2}\max_{1\leq t\leq T}|u_t^2| = o_p(1)$, which follows as in the proof of Theorem 3.8 in PS.

Finally, since $\widehat{\lambda}_v^2$ is a consistent estimator of the non-normalized spectral density at zero of $\{\upsilon_t\}$ under the null hypothesis, the Theorem follows by applying Theorem 4.1 of Billingsley (1968).

References

Berk K.N. (1974), Consistent autoregressive spectral estimates, *Annals of Statistics* 2, 489–502.

Billingsley P. (1968), *Convergence of probability measures*, New York: Wiley.

Boswijk H.P. (2001), Testing for a unit root with near-integrated volatility, Tinbergen Institute Discussion Paper # 01-077/4.

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

Caner M. and B.E. Hansen (2001), Threshold autoregression with a unit root, *Econometrica* 69, 1555–1596.

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.

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

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

Davidson J. (1994), *Stochastic limit theory*, Oxford: Oxford University Press.

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

Engle R.F., D.M. Lilien and R.P. Robins (1987), Estimating time-varying risk premia in the term structure: the ARCH-M model, *Econometrica* 55, 391–407.

Hamori S. and 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 E. and Rahbek A. (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. and 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. and P. Schmidt (1993), Unit root tests with conditional heteroskedasticity, *Journal of Econometrics* 59, 287–300.

Kim S., N. Shepard and S. Chib (1998), Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies* 65, 361–393.

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

Koop, G. and 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 and A.H. Wurtz (eds.), *Nonlinear Econometric Modelling in Time Series Analysis*, Cambridge: Cambridge University Press, 61–78.

Ling, S. and 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.

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

MacKinnon J. G., A. A. Huag and L. Michelis (1999), Numerical distribution functions of likelihood ratio tests for cointegration, *Journal of Applied Econometrics* 14, 563-577.

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

Müller, U.K. and G. Elliott (2003), Tests for unit roots and the initial condition, *Econometrica* 71, 1269-1286

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

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

Pagan A.R. and G.W. Schwert (1990), Testing for covariance stationarity in stock market data, *Economics Letters* 33, 165–70.

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 and Economic Statistics* 8, 153-162.

Perron P. and 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.

Perron P. and S. Ng (1998), An autoregressive spectral density estimator at frequency zero for nonstationarity tests, *Econometric Theory* 14, 560–603.

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

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

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

Revuz D. and M. Yor (1991), *Continuous martingales and Brownian motion*, Berlin: Springer-Verlag.

Said S.E. and D.A. Dickey (1984), Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* 89, 1420–1437.

Shorack, G.R. and J.A. Wellner (1987), *Empirical processes and their applications to statistics*, New York: John Wiley and Sons.

Stock J.H. (1999), A class of tests for integration and cointegration, In Engle, R.F., and White, H. (Eds.), *Cointegration, Causality and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Oxford: Oxford University Press, 137–167.

Tanaka K. (1996), *Time series analysis*, New York: Wiley.

van Dijk D., D.R. Osborn and M. Sensier (2002), Changes in variability of the business cycle in the G7 countries, Erasmus University Rotterdam, Econometric Institute Report EI 2002-28.

Warnock M.V.C. and F.E. Warnock (2000), The declining volatility of US unemployment: was Arthur Burns right?, International Finance Discussion Paper No. 677, Board of Governors of the Federal Reserve System.

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

Figures and Tables

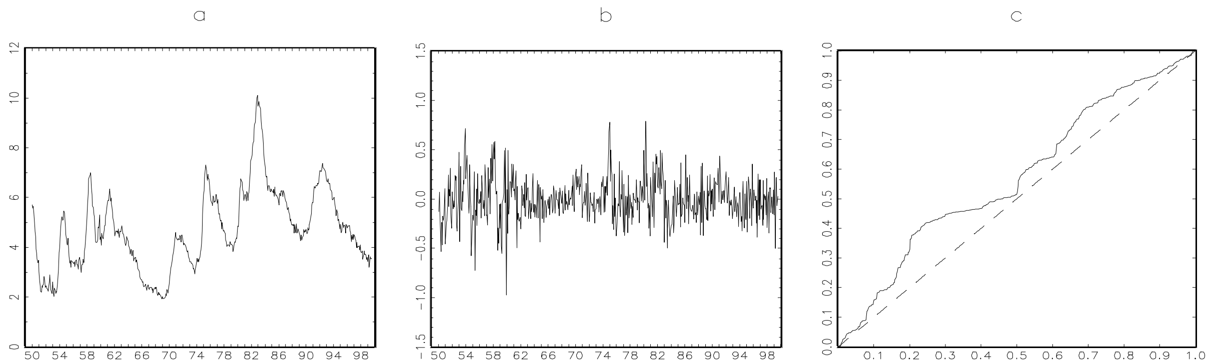

Figure 1: Monthly U.S. unemployment rate among adult males from January, 1950 through August, 1999. Panel (a): levels; Panel (b): first differences; Panel (c); estimated variance profile.

volatility model	T	no deterministic correction						constant correction						stationary volatility tests			
		$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
	100	4.4	4.4	4.6	5.0	5.0	5.0	4.4	4.1	4.3	5.0	5.0	5.0	5.0	5.0	5.0	5.0
constant	250	4.9	4.9	5.0	5.0	5.0	5.0	4.6	4.8	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	500	4.9	5.0	4.9	5.0	5.0	5.0	4.8	4.8	4.7	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	100	4.3	4.3	4.3	6.1	6.2	6.2	4.3	3.9	4.3	3.9	1.9	4.8	13.4	11.6	16.4	19.8
$\delta = \frac{1}{5}, \tau = 0.1$	250	5.3	5.2	5.3	6.5	6.7	6.7	5.0	4.8	5.3	4.2	2.3	5.3	31.4	38.3	44.8	60.5
	500	5.4	5.4	5.4	6.7	6.6	7.0	5.0	4.6	5.0	3.8	2.0	4.7	76.6	99.0	83.1	97.9
	100	1.3	1.4	1.2	13.3	11.7	18.4	0.4	1.1	0.4	14.1	8.8	20.8	81.3	97.1	98.8	99.6
$\delta = \frac{1}{5}, \tau = 0.9$	250	3.2	3.4	3.3	14.0	12.9	19.6	1.7	2.5	1.9	14.9	10.3	21.4	99.9	100.0	100.0	100.0
	500	4.2	4.4	4.3	15.3	13.9	20.7	3.3	3.4	3.4	15.6	10.3	22.1	100.0	100.0	100.0	100.0
	100	1.3	1.4	1.0	7.8	7.9	7.5	0.4	0.4	0.5	33.8	49.5	17.3	74.2	94.6	97.3	99.2
$\delta = 5, \tau = 0.1$	250	3.2	3.4	3.1	8.9	9.0	8.7	1.6	1.9	1.8	35.6	51.0	20.5	99.8	100.0	100.0	100.0
	500	4.1	4.4	4.0	9.7	9.8	9.6	2.9	2.9	2.9	36.8	51.3	21.5	100.0	100.0	100.0	100.0
	100	4.3	4.4	4.2	4.9	5.1	4.3	4.0	4.0	3.9	5.2	6.2	4.3	15.4	10.0	13.7	15.6
$\delta = 5, \tau = 0.9$	250	4.9	4.9	4.9	4.9	5.2	4.5	4.8	4.9	5.0	5.3	6.2	4.9	34.7	36.6	42.3	58.6
	500	5.7	5.6	5.6	5.6	5.8	5.2	4.8	4.4	4.9	5.3	5.7	4.7	80.0	98.9	82.7	98.0

Table 1: Single volatility shift model: size of the heteroskedasticy-robust M^H tests, size of the standard M tests and power of the stationary volatility tests.

volatility model	T	no deterministic correction						constant correction						stationary volatility tests			
		$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
$\delta = \frac{1}{5}, \tau = 0.05$	100	4.3	4.5	4.2	5.1	5.3	4.6	3.8	4.0	3.9	4.0	3.3	3.8	11.5	3.4	3.8	3.9
	250	4.7	4.7	4.8	5.1	5.3	5.0	4.8	5.2	4.9	4.2	3.2	4.4	29.1	5.3	6.8	13.3
	500	5.4	5.4	5.5	5.9	6.0	5.7	5.1	4.6	4.9	4.0	2.8	4.4	73.1	9.7	14.9	53.0
$\delta = \frac{1}{5}, \tau = 0.45$	100	1.8	1.9	1.5	10.8	10.9	10.2	0.6	1.3	0.5	11.0	7.8	11.4	86.4	2.8	6.6	2.1
	250	3.5	3.7	3.3	12.0	12.1	11.7	2.0	2.6	1.9	11.5	8.6	12.3	99.9	77.0	90.9	82.1
	500	4.2	4.3	4.1	11.7	11.8	11.5	3.1	3.0	3.0	11.6	8.4	12.1	100.0	99.8	100.0	99.9
$\delta = 5, \tau = 0.05$	100	1.3	1.4	1.1	6.8	6.1	9.5	0.5	0.4	0.6	24.6	26.3	22.2	63.5	38.5	34.8	90.6
	250	3.5	3.6	3.3	9.0	8.2	11.7	1.8	2.0	2.1	28.6	30.2	26.3	99.8	91.7	93.9	100.0
	500	4.7	4.8	4.8	9.2	8.5	11.8	3.1	3.0	3.4	31.7	32.9	27.5	100.0	100.0	100.0	100.0

Table 2: Double volatility shift model: size of the heteroskedasticy-robust M^H tests, size of the standard M tests and power of the stationary volatility tests.

volatility model	T	no deterministic correction						constant correction						stationary volatility tests			
		$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
$\delta = \frac{1}{5}$	100	4.2	4.2	4.0	8.1	7.8	8.9	3.1	3.5	3.5	4.9	2.5	6.9	89.9	98.7	99.8	99.8
	250	5.0	5.0	5.1	8.9	8.5	9.8	4.5	4.5	4.9	5.4	3.2	7.5	100.0	100.0	100.0	100.0
	500	5.7	5.7	5.7	9.0	8.8	10.4	5.0	4.5	4.9	5.4	3.0	7.3	100.0	100.0	100.0	100.0
$\delta = 5$	100	3.8	3.8	3.4	5.4	5.6	4.8	3.4	3.2	3.3	8.8	16.4	6.3	90.7	98.2	99.7	99.5
	250	4.6	4.9	4.7	5.7	6.1	5.2	4.5	4.4	4.6	9.2	16.6	6.9	100.0	100.0	100.0	100.0
	500	5.1	5.1	5.2	5.9	6.2	5.5	4.7	4.7	4.5	9.5	16.1	7.0	100.0	100.0	100.0	100.0

Table 3: Trending volatility model: size of the heteroskedasticy-robust M^H tests, size of the standard M tests and power of the stationary volatility tests.

volatility model	T	no deterministic correction						constant correction						stationary volatility tests			
		$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
	100	3.3	3.2	3.1	10.5	10.1	11.3	2.1	2.1	2.0	13.9	15.9	12.8	79.4	77.3	77.5	77.6
$c = 0, \nu = 5$	250	3.8	3.8	3.7	10.4	10.0	11.6	3.2	3.8	3.3	14.2	16.7	13.3	96.8	94.2	93.8	95.1
	500	4.3	4.4	4.3	10.1	10.0	11.0	3.9	3.9	3.9	15.0	16.8	13.4	99.8	99.1	98.5	99.1
	100	4.0	4.1	4.0	6.2	5.9	6.2	3.1	3.2	3.2	6.2	6.5	6.3	46.3	38.5	35.4	34.7
$c = 10, \nu = 5$	250	4.9	5.1	4.9	6.4	6.4	6.5	4.1	4.8	4.3	6.3	6.6	6.6	82.8	70.9	67.2	69.3
	500	4.6	4.7	4.8	6.0	6.0	6.3	4.4	4.2	4.5	6.7	6.5	6.5	97.0	90.8	86.6	88.3
	100	4.3	4.2	4.1	5.6	5.5	5.6	3.4	3.8	3.6	5.6	5.8	5.8	23.6	19.3	17.1	17.3
$c = 20, \nu = 5$	250	4.9	4.8	4.8	5.7	5.6	5.8	4.2	4.8	4.4	5.3	6.1	5.4	58.0	45.3	40.2	41.9
	500	4.7	4.7	4.8	5.5	5.4	5.5	4.6	4.5	4.5	5.8	5.6	5.5	85.6	72.5	64.7	67.7
	100	1.2	1.2	1.1	16.6	15.5	19.8	0.4	0.8	0.4	25.6	26.7	24.0	67.2	69.7	72.0	71.7
$c = 0, \nu = 10$	250	2.6	2.7	2.4	17.9	16.8	21.1	1.4	2.2	1.4	27.4	28.7	26.5	93.2	91.3	92.9	92.5
	500	3.1	3.2	3.0	18.2	17.1	21.4	2.2	2.5	2.2	28.2	28.6	27.4	99.2	98.5	98.8	98.6
	100	2.3	2.3	2.1	9.8	9.3	10.1	1.0	1.6	1.0	11.3	11.4	11.9	52.2	43.6	42.4	41.3
$c = 10, \nu = 10$	250	3.1	3.1	3.2	10.1	9.9	10.7	2.1	3.1	2.1	11.6	11.7	12.1	87.7	76.8	75.6	75.9
	500	3.8	3.9	3.9	9.6	9.5	10.3	3.0	3.5	3.0	11.7	11.5	12.1	98.5	94.5	93.2	93.8
	100	3.3	3.2	3.0	7.5	7.2	7.5	1.6	2.1	1.7	8.8	9.1	8.6	35.3	28.4	26.6	27.1
$c = 20, \nu = 10$	250	3.7	3.7	3.9	7.5	7.3	7.8	2.4	3.3	2.6	8.2	8.7	8.6	77.0	62.9	58.7	59.9
	500	4.1	4.1	4.3	7.4	7.3	7.7	4.1	4.2	3.8	9.2	9.2	9.5	96.4	87.5	83.0	84.8

Table 4: Exponential (near-) integrated stochastic volatility model: size of the heteroskedasticy-robust M^H tests, size of the standard M tests and power of the stationary volatility tests.

volatility		no deterministic correction						constant correction						stationary volatility tests			
model	T	$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
	100	41.5	42.1	38.0	43.7	44.0	42.2	25.6	13.7	31.1	27.8	15.3	32.7	5.3	5.2	5.4	5.4
constant	250	45.2	46.4	43.4	47.3	47.2	44.9	26.6	15.5	32.3	27.3	16.6	32.5	5.0	5.1	5.0	4.9
	500	47.1	47.4	46.2	47.4	47.2	46.0	27.1	16.8	33.1	27.6	17.3	33.5	5.2	5.1	5.1	5.2
	100	43.2	44.3	40.6	44.9	45.1	42.5	22.4	11.2	27.9	34.5	30.1	35.6	14.9	13.2	17.7	21.4
$\delta = \frac{1}{5}, \tau = 0.1$	250	40.9	41.4	38.9	40.8	41.6	38.9	24.9	14.8	29.0	35.0	32.1	35.4	31.3	37.7	42.7	58.6
	500	42.4	42.5	40.5	43.6	43.4	41.7	24.8	14.4	29.6	35.3	32.3	35.2	74.1	97.8	82.0	97.9
	100	16.2	15.9	16.9	32.5	30.3	35.3	7.9	2.2	12.0	22.3	19.2	25.4	83.0	97.5	98.9	99.6
$\delta = \frac{1}{5}, \tau = 0.9$	250	16.9	16.1	17.8	34.2	33.4	34.8	9.6	3.7	12.8	22.6	19.8	25.2	99.9	100.0	100.0	100.0
	500	17.6	17.4	17.5	34.9	33.5	36.6	10.0	4.9	13.3	22.8	20.5	24.9	100.0	100.0	100.0	100.0
	100	13.4	13.9	11.7	19.2	19.3	18.5	10.1	5.3	11.0	9.2	0.0	16.2	74.7	95.0	97.8	99.3
$\delta = 5, \tau = 0.1$	250	17.2	17.9	15.2	26.9	27.9	25.7	11.6	6.1	13.2	9.4	0.1	18.2	99.8	100.0	100.0	100.0
	500	18.7	19.7	16.1	32.3	33.4	30.8	11.2	6.6	12.6	10.2	0.1	17.5	100.0	100.0	100.0	100.0
	100	37.6	37.8	34.9	45.2	47.2	42.3	23.1	11.4	27.1	28.4	15.3	33.8	15.9	10.2	14.0	15.9
$\delta = 5, \tau = 0.9$	250	42.3	43.3	37.6	47.7	49.8	44.9	24.3	14.1	27.6	27.2	14.8	32.4	35.1	34.6	40.2	56.4
	500	44.5	46.9	39.4	48.7	50.8	45.1	25.7	15.4	29.1	27.4	14.7	32.6	75.3	97.8	79.5	97.0

Table 5: Single volatility shift model: size-adjusted power of the heteroskedasticy-robust M^H tests, size-adjusted power of the standard M tests and power of the stationary volatility tests.

volatility model	T	no deterministic correction						constant correction						stationary volatility tests			
		$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
	100	42.7	42.6	41.5	48.6	50.7	46.0	23.6	11.5	28.8	32.4	22.0	35.9	12.4	3.8	4.5	4.5
$\delta = \frac{1}{5}, \tau = 0.05$	250	42.8	43.3	40.6	46.1	47.4	44.4	23.6	13.0	28.1	31.6	24.4	34.3	30.0	5.0	6.5	12.8
	500	44.4	44.9	41.7	46.8	48.5	44.7	24.0	13.8	28.6	32.6	26.3	35.2	69.4	9.2	14.1	51.8
	100	17.6	19.1	14.3	34.2	35.5	33.0	9.2	2.5	11.6	24.7	22.9	25.4	86.4	3.8	8.7	3.3
$\delta = \frac{1}{5}, \tau = 0.45$	250	18.4	20.3	15.8	34.7	35.8	33.9	10.8	4.5	12.6	24.1	22.7	24.9	99.9	78.0	91.4	83.3
	500	19.0	20.9	16.3	36.3	37.3	35.1	11.6	5.7	13.5	23.7	22.1	24.4	100.0	99.8	100.0	100.0
	100	13.1	12.6	13.8	18.3	17.8	18.3	10.4	5.9	12.3	7.9	0.8	15.4	63.5	38.0	35.7	91.8
$\delta = 5, \tau = 0.05$	250	16.4	16.7	17.0	25.8	24.9	25.9	10.6	5.0	13.6	8.3	0.8	16.0	99.7	91.0	92.5	100.0
	500	19.2	19.3	19.4	31.3	30.0	30.9	11.3	5.8	14.0	8.7	0.7	17.6	100.0	100.0	100.0	100.0

Table 6: Double volatility shift model: size-adjusted power of the heteroskedasticy-robust M^H tests, size-adjusted power of the standard M tests and power of the stationary volatility tests.

volatility model	T	no deterministic correction $\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	constant correction $\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	stationary volatility tests $\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
	100	32.8	32.6	32.0	38.4	38.1	38.0	16.5	7.2	22.9	29.2	25.9	31.1	90.5	98.6	99.9	99.8
$\delta = \frac{1}{5}$	250	32.0	32.1	31.9	37.9	37.4	38.0	19.0	10.4	24.1	29.8	28.1	31.0	100.0	100.0	100.0	100.0
	500	33.4	33.1	33.4	38.3	37.2	37.8	20.4	11.5	24.9	30.7	28.8	31.8	100.0	100.0	100.0	100.0
	100	29.2	31.1	24.8	36.9	37.8	34.7	20.1	11.2	21.9	19.4	3.9	27.0	89.9	98.0	99.6	99.6
$\delta = 5$	250	35.0	38.0	30.5	43.8	45.7	41.2	22.7	13.2	25.1	20.3	4.5	28.4	100.0	100.0	100.0	100.0
	500	34.7	37.9	30.1	45.8	47.9	42.8	21.7	13.1	24.2	19.4	4.5	26.6	100.0	100.0	100.0	100.0

Table 7: Trending volatility model: size-adjusted power of the heteroskedasticy-robust M^H tests, size-adjusted power of the standard M tests and power of the stationary volatility tests.

volatility model	T	no deterministic correction						constant correction						stationary volatility tests			
		$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
$c = 0, \nu = 5$	100	31.5	34.0	26.9	40.1	41.1	38.1	22.4	10.5	24.8	20.2	3.0	28.0	24.7	48.3	56.1	62.8
	250	29.1	31.5	24.7	41.9	42.9	40.1	16.4	9.3	18.5	26.9	22.4	29.0	100.0	100.0	100.0	100.0
	500	33.4	32.4	31.2	40.8	41.0	40.1	20.1	12.1	23.3	17.1	5.5	24.2	100.0	100.0	100.0	100.0
$c = 10, \nu = 5$	100	36.0	36.4	34.1	44.6	45.7	42.2	19.6	9.7	23.6	27.0	18.1	31.1	27.9	15.7	9.7	8.8
	250	29.8	29.3	30.3	39.3	37.9	39.5	17.7	9.1	21.3	22.6	15.7	26.5	99.8	100.0	99.9	100.0
	500	40.0	40.5	39.6	44.5	44.7	42.0	23.6	14.0	28.5	27.7	20.3	31.1	96.6	73.9	68.5	81.5
$c = 20, \nu = 5$	100	39.9	42.2	36.9	46.6	47.4	43.6	25.9	12.8	28.7	27.3	13.1	33.6	15.6	14.6	15.5	12.4
	250	39.4	38.5	39.9	43.7	43.2	42.9	24.2	14.2	29.4	26.9	15.1	31.7	34.7	28.0	24.4	33.7
	500	43.2	42.9	41.7	45.2	45.0	44.0	26.0	15.5	30.3	28.6	19.8	32.4	53.7	63.4	58.5	58.5
$c = 0, \nu = 10$	100	16.7	19.1	12.9	32.6	33.0	31.4	8.8	2.2	11.0	17.1	12.1	19.1	92.3	95.4	96.2	92.0
	250	22.6	22.1	22.7	28.4	27.8	28.7	12.4	5.8	15.3	25.3	25.0	25.3	98.1	99.7	100.0	100.0
	500	20.8	22.2	17.9	37.4	38.9	35.0	12.1	6.4	13.4	17.8	10.7	21.6	99.5	99.4	99.2	99.4
$c = 10, \nu = 10$	100	29.2	30.4	25.0	44.3	45.4	41.4	15.9	7.8	20.1	17.8	5.6	24.5	35.2	43.9	36.3	34.3
	250	24.6	27.1	21.9	40.4	42.1	38.2	14.4	6.9	16.4	23.2	20.1	25.2	95.8	84.3	74.8	69.3
	500	23.3	24.2	21.3	32.9	33.2	31.5	12.7	6.7	14.7	23.9	23.7	24.2	99.9	94.5	92.9	93.0
$c = 20, \nu = 10$	100	22.4	22.2	21.8	41.9	41.1	38.9	12.6	4.1	16.0	27.5	21.2	29.0	9.3	4.9	5.4	10.8
	250	32.4	33.6	30.3	41.3	41.9	39.7	19.0	10.9	22.4	28.6	24.9	30.8	72.3	79.6	65.8	57.6
	500	24.3	25.4	21.4	40.0	42.2	38.7	13.9	7.0	16.1	22.0	15.3	24.4	99.7	99.9	100.0	100.0

Table 8: Exponential (near-) integrated stochastic volatility model: size-adjusted power of the heteroskedasticy-robust M^H tests, size-adjusted power of the standard M tests and power of the stationary volatility tests.

	volatility		no deterministic correction						constant correction						stationary volatility tests			
ϕ	model	T	$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
	Model 1	100	3.0	3.0	2.9	5.1	5.1	4.9	4.0	2.8	4.0	28.4	48.7	12.5	99.6	99.6	99.6	99.7
	$\delta = 5, \tau = 0.1$	250	1.9	2.0	1.9	2.9	3.1	2.7	2.5	2.5	2.6	26.5	45.9	11.8	74.1	85.2	89.7	93.3
		500	1.9	2.1	1.7	5.2	5.4	5.1	2.0	1.9	1.8	31.7	47.8	15.7	33.2	52.5	63.8	73.4
	Model 2	100	8.5	8.1	8.8	9.7	8.8	12.1	10.9	8.7	11.8	29.9	35.3	25.8	57.5	50.3	49.9	64.3
	$\delta = 5, \tau = 0.05$	250	6.3	6.1	6.9	8.3	7.2	11.1	7.6	5.8	8.5	30.5	34.2	26.9	55.1	44.3	44.1	60.0
0.5		500	5.8	5.5	6.3	7.7	6.8	10.4	6.4	5.5	6.9	32.6	35.0	29.2	44.4	23.6	24.3	51.4
	Model 3	100	6.7	6.8	6.9	9.8	9.3	11.3	7.7	7.4	8.4	7.7	4.7	10.3	31.3	58.2	77.9	75.8
	$\delta = \frac{1}{5}$	250	5.4	5.5	5.5	8.8	8.6	10.3	5.0	5.4	5.7	5.4	3.1	7.7	74.8	90.4	96.1	95.8
		500	5.6	5.6	5.9	9.3	8.7	10.6	4.6	4.4	4.8	5.2	2.6	7.1	90.5	95.8	98.5	98.4
	Model 4	100	5.9	5.6	5.8	9.6	9.6	10.3	6.7	6.9	7.0	12.9	16.9	12.2	35.5	41.0	45.2	43.9
	$c = 0, \nu = 5$	250	5.1	5.1	5.1	9.9	9.9	10.7	5.0	5.5	5.3	13.3	16.5	12.3	39.8	46.4	54.0	52.7
		500	4.5	4.6	4.6	10.0	9.6	10.9	4.3	4.5	4.5	14.5	16.3	13.2	47.1	52.9	60.9	60.3
	Model 1	100	0.2	0.3	0.1	5.7	5.8	5.2	0.1	1.3	0.1	31.3	47.4	14.5	99.3	99.5	99.4	99.6
	$\delta = 5, \tau = 0.1$	250	0.4	0.5	0.4	3.2	3.2	3.1	0.5	2.5	0.5	27.6	44.5	12.4	75.6	86.2	90.5	93.4
		500	0.8	0.9	0.6	5.7	5.8	5.5	0.6	2.0	0.5	32.5	47.6	16.7	33.0	52.7	63.9	73.5
	Model 2	100	2.3	2.2	2.5	5.3	4.5	8.2	2.3	2.7	2.6	18.5	23.3	16.1	56.5	48.9	48.5	63.4
	$\delta = 5, \tau = 0.05$	250	2.0	1.9	2.4	6.3	5.5	9.4	1.9	3.4	2.3	24.4	28.1	22.0	55.1	44.6	44.5	59.9
-0.5		500	3.1	2.9	3.4	7.0	6.1	9.8	3.0	4.1	3.4	29.8	32.7	26.7	44.3	23.3	24.3	51.1
	Model 3	100	1.9	1.8	1.9	5.8	5.7	6.6	1.0	1.7	1.2	3.4	1.7	4.7	31.6	58.8	77.8	75.8
	$\delta = \frac{1}{5}$	250	2.8	3.0	2.7	7.4	7.3	8.3	1.1	2.0	1.3	3.6	1.8	5.4	73.3	89.3	95.6	95.5
		500	3.8	4.0	3.9	8.1	7.8	9.3	2.0	2.3	2.3	4.3	1.9	5.9	90.4	95.9	98.7	98.6
	Model 4	100	1.6	1.6	1.6	7.1	6.6	8.0	1.4	2.7	1.5	9.4	13.6	8.4	34.7	39.9	44.0	42.9
	$c = 0, \nu = 5$	250	2.1	2.2	2.1	8.5	8.4	9.4	1.4	2.6	1.5	11.1	14.5	10.3	39.9	46.5	53.2	52.7
		500	2.7	2.8	2.7	9.9	9.6	10.9	1.7	2.8	1.8	13.3	15.4	12.3	47.0	53.0	61.4	60.7

Table 9: Size of the heteroskedasticy-robust M^H tests, size of the standard M tests and power of the stationary volatility tests, $AR(1)$ data generating process.

volatility model	T	constant and trend correction						stationary volatility tests			
		$\mathcal{MZ}^H_{\alpha.z}$	$\mathcal{MZ}^H_{t.z}$	$\mathcal{MSB}^H_z$	$\mathcal{MZ}_{\alpha.z}$	$\mathcal{MZ}_{t.z}$	$\mathcal{MSB}_z$	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
	100	3.4	3.5	3.4	5.0	5.0	5.0	5.0	5.0	5.0	5.0
constant	250	4.0	4.0	4.2	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	500	4.7	4.5	4.7	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Model 1	100	0.0	0.1	0.1	33.1	52.9	19.1	74.6	95.4	97.4	99.2
$\delta = 5, \tau = 0.1$	250	0.6	0.7	0.6	36.7	54.2	22.9	99.7	100.0	100.0	100.0
	500	2.2	2.2	2.2	36.7	54.0	23.4	100.0	100.0	100.0	100.0
Model 2	100	0.1	0.1	0.1	27.3	30.3	23.3	63.1	40.5	35.2	91.1
$\delta = 5, \tau = 0.05$	250	0.5	0.6	0.5	33.7	37.0	29.6	99.7	92.7	94.4	100.0
	500	1.9	1.9	1.9	35.8	39.5	30.3	100.0	100.0	100.0	100.0
Model 3	100	1.9	2.5	1.9	5.3	3.0	6.9	89.2	98.8	99.8	99.8
$\delta = \frac{1}{5}$	250	3.5	3.8	3.7	6.2	4.3	7.8	100.0	100.0	100.0	100.0
	500	4.6	4.4	4.6	6.2	4.2	7.5	100.0	100.0	100.0	100.0
Model 4	100	0.9	1.5	0.8	13.2	15.0	12.3	78.0	77.5	77.8	78.3
$c = 0, \nu = 5$	250	2.6	2.9	2.4	14.2	15.5	13.6	97.0	94.1	93.5	94.6
	500	3.6	3.9	3.7	14.0	15.1	12.8	99.7	99.0	98.4	98.9

Table 10: Size of the heteroskedasticy-robust M^H tests, size of the standard M tests and power of the stationary volatility tests, OLS detrended data.

Tests for the null hypothesis of stationary volatility					
	$\mathcal{H}_R$	$\mathcal{H}_{KS}$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$	
	1.439	1.404	0.471	2.544	
Standard tests for a unit root					
$\mathcal{MZ}_\alpha$	$\mathcal{MZ}_t$	$\mathcal{MSB}$	$\mathcal{MZ}_{\alpha.1}$	$\mathcal{MZ}_{t.1}$	$\mathcal{MSB}_1$
−9.686	−2.132	0.220	−15.559	−2.785	0.179
Heteroskedasticity-robust tests for a unit root					
$\mathcal{MZ}_\alpha^H$	$\mathcal{MZ}_t^H$	$\mathcal{MSB}^H$	$\mathcal{MZ}_{\alpha.1}^H$	$\mathcal{MZ}_{t.1}^H$	$\mathcal{MSB}_1^H$
−11.754	−2.337	0.199	−14.299	−2.633	0.184

Table 11: Unit root and stationary volatility tests for the U.S. unemployment rate among adult males from January, 1950 through August, 1999.