Giuseppe Cavaliere

Unit root tests under time-varying variances

Serie Ricerche 2003, n.2

Dipartimento di Scienze Statistiche “Paolo Fortunati”
Università degli studi di Bologna

Unit root tests under time-varying variances

Giuseppe Cavaliere
Dipartimento di Scienze Statistiche
Università di Bologna

Abstract

The paper provides a general framework for investigating the effects of permanent changes in the variance of the errors of an autoregressive process on unit root tests. Such a framework – which is based on a novel asymptotic theory for integrated and near integrated processes with heteroskedastic errors – allows to evaluate how the variance dynamics affect the size and the power function of unit root tests. Contrary to previous studies, it is shown that under permanent variance shifts, the conventional critical values can lead both to oversized and undersized tests. The paper concludes by showing that the power function of the unit root tests is affected by non-constant variances as well.

1 Introduction

In the recent literature on integrated processes increasing attention has been paid to the effect of structural breaks on size and power of unit root tests. In his seminal papers, Perron (1989, 1990) has shown that the presence of structural breaks in the deterministic trend can reduce the power of unit root tests dramatically. These results have been extended by Leybourne and Newbold (2000a, 2000b) and by Leybourne et al. (1998), who show that the size of unit root tests is seriously affected by the presence of relatively early breaks in the slope of the deterministic trend. Further studies, e.g. on the effects of level breaks (Harvey et al., 2001) and of changes in the persistence (Kim, 2000), have been carried out.

This work extends the existing literature by focusing on the effect of permanent breaks in the variance of the errors in a linear (I(1) or I(0)) data generating process both on the size and the power of unit root tests. Although the existence of significant volatility breaks in the main macroeconomic variables across most countries has been documented in a number of papers (see van Dijk et al., 2002, and references therein), this topic has not been deeply investigated yet. Nelson et al. (2001) and Cavaliere (2003) examine the effect of variance breaks following a Markov switching process and find little evidence on size distortions. Similarly, it is well known that time-varying conditional variances (e.g. ARCH) do not have a serious impact on unit root tests, see e.g. Hansen and Rahbek (1998), Kim and Schmidt (1993). Both Markov-changing and ARCH variances, however, are not generally characterized by permanent changes. Conversely, Hamori and Tokihisa (1997) show that a permanent positive (negative) variance shift increases (decreases) the size of Dickey-Fuller (DF) type tests when no deterministic correction is employed. Their paper, however, contains an error and the conclusion is not correct. The effect of a single variance break on the (constant-corrected) DF t-type test is also analyzed by Kim, Leybourne and Newbold (2002), who report the presence of over-rejections when a (relatively early) negative variance shift occurs under the null hypothesis of a unit root

– a result which is the contrary of what has been reported by Hamori and Tokihisa (1997).

In this paper the effect of permanent variance changes on unit root tests is investigated in a very general framework. Instead of assuming a specific pattern for the variance of the errors, the reference data generating process (d.g.p.) only requires that the variance dynamics have bounded, square integrable sample paths. The commonly used models of structural breaks or (smooth/abrupt) parameter changes, e.g. the single-break models of Hamori and Tokihisa (1997) and Kim et al. (2002), fall within the class of models considered here.

The paper improves the existing literature at least in three further directions. First, the analysis of the impact of variance changes in the framework of unit root tests does not cover the size of the tests only. The effect of variance changes on the power functions is also inspected.

Second, instead of considering DF type tests only (as e.g. in Hamori and Tokihisa, 1997, and Kim et al., 2002), we extend the analysis to other unit root tests. Specifically, attention is paid to the class of unit root tests introduced by Sargan and Bhargava (1983) and later generalized by Stock (1990), and to the locally best invariant (LBI) tests, see e.g. Tanaka (1996). More important, the analysis also covers tests which employ heteroskedasticity and autocorrelation consistent (HAC) estimators of the long-run variance, such as the well-known Phillips-Perron tests or Stock's modified SB test. Furthermore, our framework allows for the presence of deterministic corrections in the construction of the test statistics; corrections for the presence of a broken trend (e.g. Perron's test) are also permitted.

Third, instead of relying on finite-sample results obtainable by Monte Carlo (MC) experiments only, variance breaks are analyzed by referring to a novel class of asymptotic results in the unit root and near-unit root framework under permanent variance changes. Such results are based on the concept of 'variance-transformed' diffusion processes, recently introduced by Davidson (1994). Standard asymptotics, as e.g. in Phillips (1987a, 1987b), follow as a special case.

Finally, since the asymptotic framework allows for a wide class of variance behaviour, we will analyze two common variance patterns, i.e. the single variance permanent break and the case of (upward or downward) trending variances.

The paper is organized as follows. In Section 2 the reference d.g.p. and unit root tests are described. In Section 3 an asymptotic theory for unit root processes under permanent variance changes is developed and the effect of non-constant variances on the size of the tests is discussed. Section 4 extends the asymptotic theory to the near-unit root case and allows to analyze the effect of variance changes on the power function of the tests. In Section 5 the large sample theory is applied to the single variance break model and to the case of trending variances. Section 6 concludes. All proofs are placed in the Appendix. In the following "$\xrightarrow{w}$" denotes weak convergence and "$\xrightarrow{p}$" convergence in probability; $\mathbb{I}(\cdot)$ is the indicator function and "$x := y$" indicates that x is defined by y. $\mathcal{D} := D[0,1]$ denotes the space of right continuous with left limit (cadlag) processes on $[0,1]$.

2 DGP and test statistics

We start by considering the following first-order autoregressive model

$$\begin{aligned} X(t) &= \alpha X(t-1) + u(t),\ t = 0, 1, ..., T &(1)\\ u(t) &= \sigma(t)\varepsilon(t) &(2)\\ X(0) &= 0 \text{ a.s.} &(3)\end{aligned}$$

where $\{\varepsilon(t)\}$ has zero unconditional mean and unit variance , i.e. $\mathrm{E}(\varepsilon(t)) = 0$, $\mathrm{E}(\varepsilon(t)^2) = 1$, all t. The error term $\{u(t)\}$ is split into the product of two components, $\{\varepsilon(t)\}$ and $\{\sigma(t)\}$, where the latter component is a multiplicative factor. If $\{\sigma(t)\}$ is assumed to be nonstochastic, $\sigma(t)^2$ can be interpreted as the time-varying unconditional variance of the errors, i.e. $\mathrm{E}(u(t)^2) = \sigma(t)^2$. Although it is not strictly necessary to require that the variance function $\sigma(t)$ is non-stochastic, see below, such an assumption allows a considerable simplification of the theoretical set-up. Through the paper we will also assume that the following set of conditions holds (unless differently specified).

Assumption $\mathcal{V}$. The variance term $\sigma(t)$ is defined according to the following equation

$$\sigma(t) := \omega(t/T) \tag{4}$$

where $\omega(\cdot) \in \mathcal{D}$ is a non-stochastic function with a finite number of points of discontinuity; moreover, $\omega(\cdot) > 0$ and satisfies a (uniform) first-order Lipschitz condition except at the points of discontinuity.

Assumption $\mathcal{E}$. $\{\varepsilon(t)\}$ is a real zero-mean, strictly stationary mixing process with $\mathrm{E}|\varepsilon(t)|^2 = 1$, $\mathrm{E}|\varepsilon(t)|^p < \infty$ for some $p > 2$ and with mixing coefficients $\{\alpha_m\}$ satisfying $\sum_{m=0}^{\infty} \alpha_m^{2(1/r-1/p)} < \infty$ for some $r \in (2,4]$, $r \leq p$. Furthermore, the long run variance $\lambda_\varepsilon^2 := \sum_{k=-\infty}^{\infty} \mathrm{E}(\varepsilon(t)\varepsilon(t+k))$ is strictly positive and finite.

Assumption $\mathcal{U}$. The autoregressive coefficient α satisfies $|\alpha| < 1$ or $\alpha = 1$.

Condition $\mathcal{V}$ describes the dynamics of the unconditional variance of the errors. Under $\mathcal{V}$, the variance function $\omega(\cdot)$ is square-integrable and bounded, i.e. $\int_0^1 \omega(s)^2 ds < \infty$, $\sup_{s\in[0,1]} \omega(s) < \infty$; moreover, it is allowed to have a finite number of jumps. Models of single or multiple variance shifts satisfy condition $\mathcal{V}$ with $\omega(\cdot)$ piecewise constant. For instance, the function $\omega(s) := \sigma_0 + (\sigma_1 - \sigma_0)\mathbb{I}(s > \tau)$ corresponds to the single break model with a variance shift at time $[\tau T]$. If $\omega(\cdot)^2$ is an affine function, then trivially the unconditional variance of the errors displays a linear trend. Seasonal heteroskedasticity is obtained by choosing $\omega(\cdot)$ as a periodic function with period p/T, where p is the number of seasons. The degenerate diffusion limit of GARCH processes, see Corradi (2000), for large T corresponds to (4) with $\omega(s)^2 := \left(\sigma_0^2 - \sigma^2\right)\exp(-\theta s) + \sigma^2$, $\theta > 0$. The assumption of a non-stochastic variance function $\omega(\cdot)$ can be easily weakened by simply assuming stochastic independence between

$\{\varepsilon(t)\}$ and $\{\sigma(t)\}$; obviously, the (stochastic) functional $\{\omega(\cdot)\}$ must have sample paths satisfying the requirements of Assumption $\mathcal{V}$.[1] In the stochastic variance framework, the results given in the following should be interpreted as *conditional* on a given realization of $\{\omega(\cdot)\}$. Finally, since the variance $\sigma^2(t)$ depends on T, a time series generated according to (1)-(4) formally constitutes a triangular array of the type $\{X_T(t) : t = 0, 1, ..., T;\ T = 0, 1, ...\}$, where $X_T(t)$ is recursively defined as $X_T(t) := X_T(t-1) + \sigma_T(t)\,\varepsilon(t)$, $\sigma_T(t) := \omega(t/T)$. However, this notation is not essential and the process will be simply denoted as $\{X(t)\}$.

If $\omega(\cdot)$ is not constant, then the process is unconditionally heteroskedastic. Furthermore, conditional heteroskedasticity is also permitted, since condition $\mathcal{E}$ is compatible with the presence of time-varying conditional variances, see e.g. Hansen (1992b). Condition $\mathcal{E}$ has been used extensively in the econometric literature, as it allows $\{\varepsilon(t)\}$ to belong to a wide class of weakly dependent stationary processes. Moments of order $p > 2$ are assumed to exist, and the restriction on the mixing coefficients embodies the well-known trade-off between the moments and the memory of the process. The restriction on the long-run variance λ_ε^2 rules out non-invertibility. The strict stationarity assumption is mainly used to simplify the proofs of the results of the paper; it can be removed, hence allowing weak heterogeneity of the errors (as e.g. in Phillips, 1987a).

Condition $\mathcal{U}$ assumes that the process is generated either through a stable autoregression ($|\alpha| < 1$) or through a difference equation with a unit root. Specifically, if $\alpha = 1$, $\{X(t)\}$ is a unit root process with heterogeneous increments. Integration at non-zero frequencies and explosive roots are ruled out.

In order to assess the impact of time-varying variances on unit root tests we initially consider the standard Dickey-Fuller type unit

[1] In the stochastic variance framework, one could assume that $\{\omega(\cdot)\}$ is generated by the diffusion $d\omega(s)^2 = (\omega_0 - \theta\omega(s))ds + \sigma\omega(s)\,dB(s)$, $\{B(\cdot)\}$ being a standard Brownian motion. This pattern corresponds to the non-degenerate large sample limit of GARCH processes, see Corradi (2000). Note, however, that in order to cover this case one should relax the Lipschitz continuity restriction.

root tests

$$\mathrm{DF}(\widehat{\rho}) := T(\widehat{\rho}-1) \tag{5}$$

$$\mathrm{DF}(t) := \frac{\widehat{\rho}-1}{s.e.(\widehat{\rho})} \tag{6}$$

where $\widehat{\rho} := (\sum_{t=1}^{T} X(t-1)^2)^{-1} \sum_{t=1}^{T} X(t) X(t-1)$ is the first order sample autoregressive coefficient of $\{X(t)\}$ and $s.e.(\widehat{\rho}) := \widehat{\sigma}_u (\sum_{t=1}^{T} X(t-1)^2)^{-1/2}$ is the (OLS) standard error of $\widehat{\rho}$, with $\widehat{\sigma}_u^2$ denoting the sample variance of the residuals, i.e. $\widehat{\sigma}_u^2 := (1/T) \sum_{t=1}^{T} \widehat{u}(t)^2$, $\widehat{u}(t) := X(t) - \widehat{\rho} X(t-1)$.

When constant or trend corrected tests are used, the previous statistics do not change except for $X(t)$ being substituted for by $X_c(t) := X(t) - \overline{X}$ (constant-corrected test) or by the residuals $X_{c,t}(t)$ of the OLS projection of $X(t)$ on $(1,t)'$ (constant and trend-corrected test). We denote intercept-corrected and trend-corrected Dickey-Fuller tests with $\mathrm{DF}_c(\widehat{\rho}), \mathrm{DF}_c(t)$ and with $\mathrm{DF}_{c,t}(\widehat{\rho}), \mathrm{DF}_{c,t}(t)$ respectively. In order to assess whether variance breaks can be confused with structural breaks in the trend component, we also consider Perron's test, which is based on statistics (5)-(6) with $X(t)$ replaced by the residuals $X_{c,t;\tau}(t)$ of the OLS projection of $X(t)$ on $(1, t, t\mathbb{I}(t > [\tau T]))'$, where $[\tau T]$, $0 < \tau < 1$, is the break date. We denote these tests with $\mathrm{DF}_{c,t;\tau}(\widehat{\rho})$ and $\mathrm{DF}_{c,t;\tau}(t)$.[2]

Furthermore, it is interesting to consider the locally best invariant test (LBI, see Tanaka, 1996)

$$\mathrm{LBI} := \frac{(X(T) - X(0))^2}{T\widehat{\sigma}_u^2} \tag{7}$$

and the (modified) Sargan-Bhargava test

$$\mathrm{MSB} := \frac{\sum_{t=0}^{T} X(t)^2}{T\widehat{\sigma}_u^2} \tag{8}$$

[2] For space constraints we will not focus on unit root tests based on GLS detrending, see Elliott et al. (1996), which can be treated in the same manner as the tests considered here.

The LBI and MSB tests reject the unit root null hypothesis for small values of (7). Like the DF tests, the MSB statistics can be corrected either for a constant, a constant and a linear trend or a constant and a broken linear trend by substituting $X(t)$ in (8) with $X_c(t)$, $X_{c,t}(t)$ and $X_{c,t;\tau}(t)$ respectively.

Along with the previously mentioned tests, since the error component $\{\varepsilon(t)\}$ can display short memory we also consider autocorrelation-corrected unit root tests, namely the Phillips-Perron coefficient-based and t-based tests[3]. Such tests employ a sum-of-covariances estimator of the long-run variance of the form

$$\widehat{\lambda}_u^2 := \sum_{k=-T+1}^{T-1} h\left(\frac{k}{q_T}\right)\widehat{\gamma}_u(k)\,,\ \widehat{\gamma}_u(k) := \frac{1}{T}\sum_{t=|k|+1}^{T}\widehat{u}(t)\,\widehat{u}(t-|k|) \tag{9}$$

and are given by

$$\begin{aligned}\mathrm{PP}(\widehat{\rho}) &:= \mathrm{DF}(\widehat{\rho}) - \frac{(\widehat{\lambda}_u^2 - \widehat{\sigma}_u^2)/2}{T^{-2}\sum_{t=1}^{T} X_{t-1}^2} \\ \mathrm{PP}(t) &:= \frac{\widehat{\sigma}_u}{\widehat{\lambda}_u}\mathrm{DF}(t) - \frac{(\widehat{\lambda}_u^2 - \widehat{\sigma}_u^2)/2}{\widehat{\lambda}_u(T^{-2}\sum X_{t-1}^2)^{1/2}}\end{aligned}$$

In the following, it is convenient to assume that the lag truncation/bandwidth parameter q_T and the kernel function $h(\cdot)$ satisfy the condition (de Jong, 2000):

Assumption $\mathcal{K}$. (K_1) For all $x \in \mathbb{R}$, $|h(x)| \leq 1$, $h(x) = h(-x)$; $h(x)$ is continuous at 0 and for almost all $x \in \mathbb{R}$; $\int |h(x)|dx < \infty$; (K_2) $q_T \uparrow \infty$ as $T \uparrow \infty$, and $q_T = \mathrm{o}(T^{\gamma})$, $\gamma \leq 1/2 - 1/r$.

This condition is general enough for our purposes since it is satisfied by the most commonly employed kernel functions, see Andrews (1991) and Hansen (1992a).

[3]It is preferable to consider semiparametric unit root tests instead of fully-parametric augmented Dickey-Fuller type tests since the latter class of tests requires a more involved asymptotic theory without displaying substantial differences from the former class.

We also consider the autocorrelation-corrected LBI and MSB tests, which are based on statistics (7) and (8) with $\widehat{\sigma}_u^2$ replaced by $\widehat{\lambda}_u^2$.

3 Unit root asymptotics under changing variances

In this section an asymptotic theory for heteroskedastic autoregressive process will be derived under the assumption of a unit root. Such a theory allows to derive the large sample distribution of the unit root test statistics when the underlying d.g.p. is integrated with heteroskedastic errors, consequently allowing to evaluate the size properties of the tests.

Assume that the d.g.p. satisfies (1)–(3) with $\alpha = 1$. As we will show in the following, even if $\{X(t)\}$ has a unit root, due to the presence of heteroskedasticity it is not I(1) in the usual sense[4]. In order to develop the necessary asymptotic theory it is useful to introduce the following Brownian functional

$$B_\omega(s) := \overline{\sigma}^{-1} \int_0^s \omega(r)\, dB(r) \ , \qquad \overline{\sigma}^2 := \int_0^1 \omega(r)^2\, dr \tag{10}$$

where $B(\cdot)$ is a standard Brownian motion in $\mathcal{D}$. Since $\overline{\sigma}^2$ is the limit of $(1/T)\sum_{t=1}^T \sigma_t^2$, it can be naturally interpreted as the (asymptotic) average variance of the errors. The process $\{B_\omega(\cdot)\}$ is Gaussian with quadratic variation $\int_0^s \omega(r)^2\, dr / \int_0^1 \omega(r)^2\, dr$.[5] It corresponds to the diffusion process generated by the stochastic differential equation $dB_\omega(s) = \overline{\sigma}^{-1}\omega(s)\, dB(s)$ with initial condition $B_\omega(0) = 0$ a.s. For $\omega(\cdot)$ constant, $B_\omega(\cdot)$ becomes a standard Brownian motion.

The next lemma contains the basic results required to derive the asymptotic distributions of the test statistics in the heteroskedastic unit root framework.

---

[4]That is, it does not satisfies the invariance principle $T^{-1/2}X([\cdot T]) \xrightarrow{w} \lambda B(\cdot)$, where $\{B(\cdot)\}$ is a standard Brownian motion.

[5]If $\{\omega(\cdot)\}$ is stochastic, then the marginal distribution of $B_\omega(s)$ is mixed Gaussian with centre 0 and mixing parameter $\int_0^s \omega(r)^2\, dr / \int_0^1 \omega(r)^2\, dr$.

Lemma 1 *If conditions* (1)-(3) *are fulfilled with* $\alpha = 1$ *then, as* $T \uparrow \infty$*, the approximant* $X_T(s) := T^{-1/2}X([sT])$ *satisfies* $X_T(\cdot) \overset{w}{\to} \overline{\sigma}\lambda_\varepsilon B_\omega(\cdot)$.

According to Lemma 1, the time series $\{X(\cdot)\}$, properly scaled, converges to a continuous transformation of a standard Brownian motion. Trivially, for $\omega(\cdot)$ constant the lemma becomes a standard invariance principle, see e.g. Billingsley (1968). The implications for unit root tests are reported in the next theorem, where $N_{a|b}(s) := a(s) - (\int a(r)b(r)'dr)(\int b(r)b(r)'dr)^{-1}b(s)$ if $b(\cdot) \neq 0$ a.s., $N_{a|b}(s) := a(s)$ otherwise, all integrals running from 0 to 1.

Theorem 2 *Let the d.g.p. satisfy conditions* (1)-(3) *with* $\alpha = 1$*. Then, if* $\lambda_\varepsilon^2 = 1$ *the asymptotic distributions of the unit root test statistics are*

$$\mathrm{DF}(\widehat{\rho}), \mathrm{PP}(\widehat{\rho}) \overset{w}{\to} \frac{N_{B_\omega|F}(1)^2 - N_{B_\omega|F}(0)^2 - 1}{2\int_0^1 N_{B_\omega|F}(s)^2\,ds} \tag{11}$$

$$\mathrm{DF}(t), \mathrm{PP}(t) \overset{w}{\to} \frac{N_{B_\omega|F}(1)^2 - N_{B_\omega|F}(0)^2 - 1}{2\left(\int_0^1 N_{B_\omega|F}(s)^2\,ds\right)^{1/2}} \tag{12}$$

$$\mathrm{LBI} \overset{w}{\to} \left(N_{B_\omega|F}(1) - N_{B_\omega|F}(0)\right)^2 \tag{13}$$

$$\mathrm{MSB} \overset{w}{\to} \int_0^1 N_{B_\omega|F}(s)^2\,ds \tag{14}$$

where F *depends on the deterministic terms included in the estimation. Specifically:*

1. *if no deterministics are introduced in the estimation, then* $F(s) := 0$*, all* s*, and* $N_{B_\omega|F}(s) = B_\omega(s)$*;*
2. *if a constant is appended to* X_{t-1} *in the estimation, then* $F(s) := 1$*, all* s*;*
3. *if a constant and a trend are appended to* X_{t-1} *in the estimation, then* $F(s) := (1, s)'$*;*

4. if a constant and a broken trend are appended to X_{t-1} in the estimation, then $F(s) := (1, s, s\mathbb{I}(s > \tau))'$.

If $\lambda_\varepsilon^2 \neq 1$, the results do not change provided that the unit root tests are autocorrelation-corrected and satisfy condition $\mathcal{K}$.

The results of the theorem do not change if seasonal dummies are included both in the DGP and in the model, provided that a constant is appended to X_{t-1} in the estimation and the dummies are orthogonalized with respect to the constant, i.e. they are of the type $d_i(t) := \mathbb{I}(t \bmod p = i) - p^{-1}$, where p is the number of seasons. Moreover, the results 2–4 do not change if a constant is added to the DGP; the results 3–4 do not change if a linear trend is added to the DGP; result 4 does not change if a linear trend with slope varying at time $[\tau T]$ is added to the DGP.

In order to understand the distributional implications of changing variances, let us initially consider the Dickey-Fuller DF $(\widehat{\rho})$/Phillips-Perron PP $(\widehat{\rho})$ test without deterministic corrections. Trivially, when $\omega(s)^2 = \sigma^2 > 0$, $B_\omega(s) = B(s)$ and if the errors are not autocorrelated (i.e. $\lambda_\varepsilon^2 = 1$) the test statistics are asymptotically Dickey-Fuller distributed, i.e.

$$\mathrm{DF}(\widehat{\rho}), \mathrm{PP}(\widehat{\rho}) \xrightarrow{w} \frac{(B(1)^2 - 1)/2}{\int_0^1 B(s)^2 \, ds} \tag{15}$$

while, in the presence of heteroskedasticity, the distribution becomes

$$\mathrm{DF}(\widehat{\rho}), \mathrm{PP}(\widehat{\rho}) \xrightarrow{w} \frac{(B_\omega(1)^2 - 1)/2}{\int_0^1 B_\omega(s)^2 \, ds} \tag{16}$$

The difference between (15) and (16) can be appreciated by noticing that the limit process in Theorem 1 can be expressed in a slightly different way, as explained in the following proposition.

Proposition 3 *The limit Brownian functional* (10) *satisfies the distributional equality*

$$B_\omega(s) \stackrel{d}{=} B_\eta(s) := B(\eta(s))$$

where $\eta(s) := (\int_0^1 \omega(r)^2 dr)^{-1} \int_0^s \omega(r)^2 dr$ *and* $B(\cdot)$ *is a standard Brownian motion.*

Since $\omega(\cdot) \neq 0$ a.e., $\eta(\cdot)$ is an increasing homeomorphism on $[0,1]$ with $\eta(0) = 0$, $\eta(1) = 1$; consequently $\{B_\eta(\cdot)\}$ is a 'variance-transformed' Brownian motion, see Davidson (1994). It represents a Brownian motion under 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 $\eta(s) \in [0,1]$. A consequence of Proposition 3 and of the equality $B_\eta(1) = B(1)$ is that in the absence of deterministic corrections the asymptotic distributions of all the considered tests can be expressed as combinations of two random variables, namely $(B(1)^2, \int_0^1 B(s)^2 ds + Z)$, where the random element $Z := \int_0^1 (B_\eta(s)^2 - B(s)^2) ds$ equals zero in the standard homoskedastic unit root framework. The following asymptotic representation is therefore obtained for the simple DF $(\widehat{\rho})$ / PP $(\widehat{\rho})$ statistic

$$\text{DF}(\widehat{\rho}), \text{PP}(\widehat{\rho}) \overset{w}{\to} \frac{(B(1)^2 - 1)/2}{\int_0^1 B_\eta(s)^2 ds} = \frac{(B(1)^2 - 1)/2}{\int_0^1 B(s)^2 ds + Z} \tag{17}$$

where the term Z depends on the presence of heteroskedastic errors. Similarly, the following representation can be given for the other tests considered:

$$\text{DF}(t), \text{PP}(\widehat{\rho}) \overset{w}{\to} \frac{(B(1)^2 - 1)/2}{\left(\int_0^1 B(s)^2 ds + Z\right)^{1/2}} \tag{18}$$

$$\text{LBI} \overset{w}{\to} B(1)^2 \tag{19}$$

$$\text{MSB} \overset{w}{\to} \int_0^1 B(s)^2 ds + Z \tag{20}$$

The above results allow to sketch some size implications of heteroskedasticity. The presence of the random variable Z represents the leading cause of size distortions, and a test which put weight on it (like all the commonly used unit root tests) might not reliable. The LBI statistic without deterministic corrections, which asymptotically depends on $B_\eta(1)^2 = B(1)^2$ and not on $\int_0^1 B_\eta^2(s)\,ds$, is not affected by the presence of heteroskedasticity, i.e. it is $\chi^2(1)$-distributed as in the homoskedastic framework[6]. The other tests (as well as the tests based on deterministic corrections) do not fulfil the same property.

Note that since Z can be both positive and negative, it is not possible to establish if heteroskedasticity leads to either oversized or undersized tests, unless specific forms of $\omega(\cdot)$, hence of $\eta(\cdot)$, are analyzed. In general the unconditional expectation of Z is $-0.5 + \int \eta(s)\,ds$, which lies between -0.5 and 0.5; values greater (lower) than 0 might indicate that the test statistic tends to take on average higher (lower) values than in the homoskedastic case. However, the effect of Z on the *tails* of the distribution can be different, see Section 5.

Finally, it is worth noting that the asymptotic distribution of autocorrelation-corrected tests (such as e.g. the PP tests) are not affected by any nuisance parameters depending on the short run memory of the errors, in particular the long-run variance λ_ε^2. This interesting property holds (although heteroskedasticity affects the distribution of the test statistics as shown in equations (11)–(14)) since estimators of the long-run variance as provided in formula (9) under condition $\mathcal{K}$ remain consistent even under heteroskedasticity of the form considered here.

4 Near-unit root asymptotics under changing variances

In this section the implications of heteroskedastic errors in the near-unit root framework are investigated. This issue is related to the analysis of the power properties of unit root tests. Since it is straightforward to show that the consistency of the tests is not affected by

[6]Hence, even if its (asymptotic) power function is generally below the asymptotic power envelope (see Elliott et al., 1996), the LBI test has the correct size under non-constant variances.

heteroskedasticity of the form considered in this paper[7], a more interesting exercise is to assess how the asymptotic local power function of the tests depends on the presence and on the form of heteroskedasticity.

Suppose that the d.g.p. satisfies (1)-(3) in the local-to-unity framework:

$$\alpha = \exp\left(-c/T\right) \approx 1 - \frac{c}{T},\ c > 0. \tag{21}$$

This specification generalizes the class formed by the so-called near-integrated processes since heteroskedasticity of the errors is allowed. As in the standard near-unit root framework, the non-negative parameter c controls the speed of mean reversion.

As in Lemma (1), it is possible to derive a weak convergence result for the continuous-time approximant $X_T\left(s\right) := T^{-1/2}X\left(\left[sT\right]\right)$. For this purpose the following Brownian functional needs to be introduced

$$J_\omega\left(s\right) := \int_0^s e^{-c(s-r)} dB_\omega\left(r\right) = \overline{\sigma}^{-1}\int_0^s e^{-c(s-r)}\omega\left(r\right) dB\left(r\right) \tag{22}$$

where $B_\omega\left(\cdot\right)$ is the diffusion process with heteroskedastic increments defined in (10) and $B\left(\cdot\right)$ is a standard Brownian motion in $\mathcal{D}$. Process (22) is an Ornstein-Uhlenbeck (OU) process with heteroskedastic increments; it satisfies the stochastic differential equation $dJ_\omega\left(s\right) := -cJ_\omega\left(s\right) ds + \overline{\sigma}^{-1}\omega\left(s\right) dB\left(s\right)$ with the usual initial condition $J_\omega\left(0\right) = 0$. The standard OU process is obtained as a special case for $\omega\left(\cdot\right) = \overline{\sigma}$ a.e. An alternative representation of $\left\{J_\omega\left(\cdot\right)\right\}$ which does not require stochastic integration can be given, namely $J_\omega\left(s\right) = B_\omega\left(s\right) - c\int_0^s e^{-c(s-r)}B_\omega\left(r\right) dr$. Moreover, by recalling the relation between $\left\{B_\omega\left(\cdot\right)\right\}$ and the variance-transformed Brownian motion $\left\{B_\eta\left(\cdot\right)\right\}$ the following distribution equality is also obtained

$$J_\omega\left(s\right) \overset{d}{=} \int_0^s e^{-c(s-r)} dB_\eta\left(r\right) = B_\eta\left(s\right) - c\int_0^s e^{-c(s-r)}B_\eta\left(r\right) dr\ .$$

The linkage between $\left\{X_T\left(\cdot\right)\right\}$ and the functional $\left\{J_\omega\left(\cdot\right)\right\}$ is given in the next lemma.

[7] Provided that the residual variance which enters the DF(t), the LBI and the MSB statistics is computed under the alternative hypothesis, see Stock (1994).

Lemma 4 *If conditions* (1)-(3) *are fulfilled with* $\alpha = \exp(-c/T)$, $c > 0$, *then, as* $T \uparrow \infty$, *the approximant* $X_T(s) := T^{-1/2} X([sT])$ *satisfies* $X_T(s) \xrightarrow{w} \overline{\sigma} \lambda_\varepsilon J_\omega(s)$.

The implications for the asymptotic distributions of the unit root tests are summarized below.

Theorem 5 *Let the d.g.p. satisfy conditions* (1)-(3) *with* $\alpha = \alpha_T := \exp(-c/T)$, $c > 0$. *Then, if* $\lambda_\varepsilon^2 = 1$ *the asymptotic distributions of the unit root test statistics are*

$$\begin{aligned} \mathrm{DF}(\widehat{\rho}), \mathrm{PP}(\widehat{\rho}) &\xrightarrow{w} \frac{N_{J_\omega|F}(1)^2 - N_{J_\omega|F}(0)^2 - 1}{2\int_0^1 N_{J_\omega|F}(s)^2 ds} \\ \mathrm{DF}(t), \mathrm{PP}(t) &\xrightarrow{w} \frac{N_{J_\omega|F}(1)^2 - N_{J_\omega|F}(0)^2 - 1}{2\left(\int_0^1 N_{J_\omega|F}(s)^2 ds\right)^{1/2}} \\ \mathrm{LBI} &\xrightarrow{w} \left(N_{J_\omega|F}(1) - N_{J_\omega|F}(0)\right)^2 \\ \mathrm{MSB} &\xrightarrow{w} \int_0^1 N_{J_\omega|F}(s)^2 ds \end{aligned}$$

where F *depends on the deterministic terms included in the estimation, see Theorem* 2. *If* $\lambda_\varepsilon^2 \neq 1$, *the results do not change provided that the unit root tests are autocorrelation-corrected and satisfy condition* $\mathcal{K}$.

Theorem 5 shows how the local-to-unity asymptotics are non-standard in the heteroskedastic case. The asymptotic distributions (and consequently the asymptotic local power functions of the tests) are affected by both the location parameter c and the structure of heteroskedasticity. Note that it is not possible to establish whether heteroskedasticity leads to higher or lower power with respect to the standard homoskedatic framework, unless a specific form is chosen for the variance $\omega(\cdot)$ functional. However, substantial modifications of the power functions of the tests are likely to be found. This issue is examined in the next section.

5 Further insights on size and power

In this section we show how the results obtained so far can be used to analyze two special models of heteroskedasticity. In the first case, the variance changes at a given point in time, namely at time $[\tau T]$, $\tau \in (0,1)$. Since for most macroeconomic series it is difficult to identify specific events that might cause an abrupt volatility change (see van Dijk et al. 2002), we also consider the effect of a smoothly changing variance. Specifically, the second case deals with both upward and downward trending variances. For ease of notation, the asymptotics are presented under the assumption of no deterministics, while Monte Carlo estimates of the asymptotic size and power functions are presented for all the four deterministic corrections previously considered.

5.1 Single variance shift

The case of a single variance shift which occurs at time $[\tau T]$ corresponds to the variance function

$$\omega(s)^2 = \sigma([sT])^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{I}(s > \tau), \; s \in [0,1], \; \tau \in (0,1) \tag{23}$$

where τ represents the date of the variance break, expressed as the ratio between the pre-break sample size and the total sample size. The parameters σ_0^2, σ_1^2 are the pre-break innovation variance and the post-break innovation variance respectively. The asymptotic average variance is given by $\overline{\sigma}^2 := \int_0^1 \omega(r)^2 dr = \sigma_0^2\tau + \sigma_1^2(1-\tau)$. The same d.g.p. is analyzed by Kim et al. (2002), who focus on the $\mathrm{DF}_c(t)$ test; the behavior of $\mathrm{DF}(\widehat{\rho})$ (no deterministics) under an observationally equivalent d.g.p. is considered by Hamori and Tokihisa (1997). We extend their analysis by covering different types of deterministic corrections and the various unit root tests previously mentioned.

Let $\delta^2 := (\sigma_0/\sigma_1)^2$ denote the ratio between the pre-break and the post-break variances. In the case of no deterministic corrections the next result follows from Theorem 2.

Corollary 6 *Let the d.g.p. satisfy the conditions of Theorem 2 with* $\omega(\cdot)$ *defined as in* (23). *Then, the asymptotic distributions of the* *unit root test statistics in the absence of deterministic corrections are given by* (17)–(20) *with* Z *replaced by*

$$Z_{\tau,\delta} := \frac{1-\delta^2}{\delta^2}\left((1-\tau)\int_0^{a(\delta,\tau)} B(r)^2\,dr - \tau\delta^2\int_{a(\delta,\tau)}^1 B(r)^2\,dr\right) \tag{24}$$

and $a(\delta,\tau) := \tau\delta^2/\left(\delta^2\tau + 1 - \tau\right)$.

Hence, the limiting distribution of the DF/PP and MSB statistics is affected by the random variable $Z_{\tau,\delta}$, which vanishes only for $\tau = 0$, $\tau = 1$ or $\delta = 1$. Note that the term $a(\delta,\tau)$ can be rewritten as the ratio between $\sigma_0^2\tau$ and the average variance $\overline{\sigma}^2 = \sigma_0^2\tau + \sigma_1^2(1-\tau)$.

The asymptotic size of the Dickey-Fuller DF $(\widehat{\rho})$/Phillips-Perron PP $(\widehat{\rho})$ tests is plotted in Figure 1 under a 5% significance level. The break date τ takes values within the set $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ while δ spans the interval $[0.01, 100]$ in order to evaluate also the impact of very large variance changes. The asymptotic sizes were simulated by discretizing the limit Brownian functional over $T = 5,000$ segments and using $50,000$ replications. Four cases are considered depending on the chosen deterministic corrections: no deterministics (panel a), constant (panel b), constant and trend (panel c), constant and broken trend (panel d), the last case corresponding to the Perron's test with break date set to τ.

The inspection of the figure reveals the following facts which have not been reported in the existing literature. First, a single variance shift mostly causes an increase in the size of the tests. Variance shifts imply size reductions only in a small subset of the parameter space; moreover, such reductions are usually negligible. Second, a positive variance shift ($\delta < 1$) which takes place near the end of the sample (i.e. τ is close to 1) has a stronger impact on the test size than a positive shift which occurs at the beginning of the sample (τ close to 0). Conversely, negative variance shifts ($\delta > 1$) have a stronger impact when τ is close to 0.[8] Third, by adding deterministic

[8] It should be noticed that the latter result is reported by Kim et al. (2002) when the $\text{DF}_c(t)$ test is considered. These authors, however, have not emphasized

corrections the size does not necessarily increase. Generally, however, in the case of variance reductions the inclusion of extra deterministic terms inflates the test size. Fourth, Perron's test, as well as the other tests, can reject the null hypothesis too often in the presence of a variance shift. Hence, the presence of a single variance break can be erroneously interpreted by Perron's test as evidence of stationarity around a broken trend.

Are all unit root tests equally affected by variance changes? The results of Section 3 suggest that a variance shift might not have the same effect on all tests. We briefly tackle this issue by showing in Figure 2 the asymptotic size of the DF $(\widehat{\rho})$ / PP $(\widehat{\rho})$, DF (t) / PP (t) and MSB tests by fixing $\delta = \sigma_0/\sigma_1$ and varying λ between 0 and 1; specifically, $\delta = 0.2$ when a positive variance shift takes place (panel a) and $\delta = 5$ when a negative variance shift occurs (panel b). Due to space constraints, only the tests corrected for a constant and a linear trend are considered. Interestingly, in the case of a negative variance shift, the DF (t) test is the most seriously affected by size distortions, followed by the DF $(\widehat{\rho})$ test and by the MSB test. This picture changes when positive variance shifts are considered: in such a case, the MSB test is the most oversized test, followed by the DF $(\widehat{\rho})$ test and by the DF (t) test. Hence, no test dominates the other tests in terms of size accuracy. Interestingly, if a constant is included in the estimation, the size of the MSB test does not change when δ and τ are replaced with $1/\delta$ and $1-\tau$ respectively[9].

We can now investigate the effect of a variance shift on the asymptotic power of the tests. As for the size analysis, 5% significance level tests are considered for different values of δ and τ. An MC simulation has been employed, where the limit Brownian functional (22) has been simulated over $T = 5,000$ segments and using $50,000$ replications. The (autoregressive) location parameter c is set to 10.

that also positive variance breaks can significantly affect the size properties of the tests.

[9]This result is a special case of the following property, which holds for any heteroskedasticity function $\omega(\cdot)$: the asymptotic distributions of the MSB_c, $\text{MSB}_{c,t}$,$\text{MSB}_{c,t;\tau}$ statistics do not change if $\{\omega(s)\}$ is replaced by $\{\omega(1-s)\}$ and τ is replaced by $1-\tau$.

Results are illustrated in Figure 3. Generally, the tests tend to have higher power with respect to the homoskedastic case. Similarly to what obtained for the size analysis, the presence of deterministic corrections inflates the power especially when a positive variance shift takes place near the end of the sample, or when a negative shift takes place at the beginning of the sample. An interesting exception is the test without deterministics, which loses power when the variance increases ($\delta < 1$).

5.2 Trending variances

In the framework of the d.g.p. (1)–(3), a simple specification which allows for trending variances is obtained by setting

$$\omega(s)^2 := \sigma_0^2 + \left(\sigma_1^2 - \sigma_0^2\right) s, \tag{25}$$

which implies that the variance varies linearly from σ_0^2 ($s = 0$) to σ_1^2 ($s = 1$); the average variance is $\overline{\sigma}^2 = 0.5(\sigma_0^2 + \sigma_1^2)$.

As in the previous section let $\delta := \sigma_0/\sigma_1$. It is straightforward to derive the following result from Theorem 2.

Corollary 7 *Let the d.g.p. satisfy the conditions of Theorem 2 with* $\omega(\cdot)$ *defined as in* (25)*. Then, the asymptotic distributions of the unit root test statistics in the absence of deterministic corrections are given by* (17)–(20) *with* Z *replaced by* $Z_\delta := \int_0^1 f(s) B(s)^2 ds$*, where* $f(\cdot)$ *is defined as*

$$f(s) := \frac{(1+\delta^2)/2 - \sqrt{\left(\delta^4 + (1-\delta^4)s\right)}}{\sqrt{\left(\delta^4 + (1-\delta^4)s\right)}}\ . \tag{26}$$

For $\delta = 1$ (constant variance), $f(s) = 0$, all s, and the usual asymptotic distributions arise.

Figures 4 and 5 report the asymptotic size and local power of the DF $(\widehat{\rho})$/PP $(\widehat{\rho})$ tests with a 5% nominal size and different deterministic components; see the previous section for details on the MC design.

Together with the variance model (25), the more general specification is also considered:

$$\omega(s)^r := \sigma_0^r + (\sigma_1^r - \sigma_0^r)\, s \tag{27}$$

where for the sake of simplicity r is assumed to be strictly positive. Equation (27) implies that the variance varies monotonically from σ_0^2 to σ_1^2, but not necessarily in a linear fashion[10]. The linear case of equation (25) corresponds to $r = 2$. The value of r is chosen within the set $\{0.1, 0.5, 1, 2, 10\}$.

Conclusions similar to the case of single variance shifts can be drawn: upward or downward trending variances usually lead to oversized tests (although cases of slightly undersized tests cannot be ruled out). With respect to the case of a single abrupt shift, trending variances seem to determine lower size distortions and lower power changes. The strongest size deviations from the nominal level occur when the variance increases at growing rates or decreases at declining rates ($r = 0.1$ in the figures).

We now turn our attention to the asymptotic local power function. In the presence of deterministic corrections, the power is generally higher than in the homoskedastic case. Again, the test without deterministics is an interesting exception since heteroskedasticity mainly leads to a power loss, especially when the variance is trending upward ($\delta < 1$). Finally, note that the smaller r, the stronger the impact of heteroskedasticity on the asymptotic power function of the test.

6 Conclusions

The theoretical results presented in this paper allow to analyze in detail whether (and how) permanent changes in the variance of the errors of a (possibly integrated) autoregressive process affect the size

[10]Setting $r = 1$ generates the hypothesis that the unconditional *standard deviation* of $u(t)$ – not the variance – is a linear function of t. In general, if $\delta < 1$ (i.e. $\sigma_1 > \sigma_0$) the variance increases at growing rates when $r < 2$. If $\delta > 1$ (i.e. $\sigma_1 < \sigma_0$) the variance decreases at growing rates when $r > 2$.

and the power function of unit root tests. It is shown that variance changes usually lead to oversized tests, although the tests can also be undersized. In general, (abrupt or smooth) early negative variance changes and (abrupt or smooth) late positive variance changes have a strong impact on the size of the tests. The power functions of the tests are also affected by the presence of heteroskedasticity.

The implications of our results are twofold. First, the constancy of the second moments of the errors becomes a crucial issue which should always be investigating before interpreting the outcome of unit root tests. Second, since the direction of size distortions depends on nuisance parameters, new unit root tests which are invariant to the presence of heteroskedastic errors should be developed. To our knowledge, this issue has been tackled only by Kim et al. (2002). However, only the case of a single variance shift is covered by their proposal. Therefore, since for most macroeconomic time series it is difficult to identify specific dates that could be connected to an abrupt volatility change, it is certainly of interest to robustify unit root tests for the presence of different and unknown forms of variance dynamics. The results given in this paper provide the basic steps to investigate this topic.

A Mathematical Appendix

A.1 Preliminary lemmas

The following lemmas are useful to prove the results of the paper.

Lemma 8 *Let* $\{X_t\}$ *be a mixing process of size* $-\varphi$*. Then* $Y_{t,T} := \omega(t/T)\,X_t$*, where* $\omega()$ *is a deterministic function, is mixing of the same size.*

Lemma 9 *Let* $\lambda_{u,T}^2 := \sum_{k=-T+1}^{T-1} h(k/q_T)\,\gamma_{u,T}(k)$*, where, for* $k \geq 0$*,* $\gamma_{u,T}(k) := (1/T)\sum_{t=1}^{T-k} \mathrm{E}(u(t)\,u(t+k))$*, while* $\gamma_{u,T}(k) = \gamma_{u,T}(-k)$ *for* $k < 0$*. Then, under Assumptions* $\mathcal{V}$ *and* $\mathcal{K}$*,* $\lambda_{u,T}^2 - \overline{\sigma}^2\lambda_{\varepsilon}^2 \to 0$ *as* $T \uparrow \infty$*.*

Lemma 10 *Let $\widetilde{\lambda}_u^2$ be an estimator of the long-run variance of $\{u(t)\}$, where the $u(t)$'s are used instead of the $\widehat{u}(t)$'s. Then, under Assumptions $\mathcal{V}$ and $\mathcal{K}$, $\widetilde{\lambda}_u^2 - \lambda_{u,T}^2 \xrightarrow{p} 0$ as $T \uparrow \infty$.*

Lemma 11 *Let $\widehat{\lambda}_u^2$ be an estimator of the long-run variance of $\{u(t)\}$, based on the $\widehat{u}(t)$'s. Then, under Assumptions $\mathcal{V}$ and $\mathcal{K}$, $\widetilde{\lambda}_u^2 - \lambda_{u,T}^2 \xrightarrow{p} 0$ as $T \uparrow \infty$.*

Lemmas 9, 10 and 11 have the following corollary.

Corollary 12 *Under the assumptions of Lemma 10, $\widehat{\lambda}_u^2 - \overline{\sigma}^2 \lambda_\varepsilon^2 \xrightarrow{p} 0$ as $T \to \infty$.*

A.2 Proofs

Proof of Lemma 1. Under the assumptions of the theorem it is straightforward to see that $(\sigma([sT]), X_T(s))' = (\omega(s), X_T(s))'$ satisfies the requirements for convergence to stochastic integral. Specifically, since $\sigma([sT]) \to \omega(s)$ in the Skorohod topology, we can apply e.g. Hansen (1992b), Theorem 3.1 to obtain the following weak convergence

$$T^{-1/2} \sum_{t=1}^{[sT]} \omega(t/T)\, u(t) - C_T(s) \xrightarrow{w} \lambda_\varepsilon \int_0^s \omega(r)\, dB(r) = \overline{\sigma} \lambda_\varepsilon B_\omega(s)$$

where

$$C_T(s) := \frac{1}{T^{1/2}} \sum_{t=1}^{[sT]} (\sigma(t) - \sigma(t-1)) z([sT]) - \frac{1}{T^{1/2}} \sigma([sT])\, z([sT]+1), \tag{28}$$

and $z(t) := \sum_{k=1}^{\infty} \mathrm{E}_t(\varepsilon(t+k))$. To complete the proof we need to show that $C_T(s)$ is $o_p(1)$. Consider the second term of the r.h.s. of (28). Since $\omega(\cdot)$, hence $\sigma(\cdot)$, is bounded, for some $K < \infty$ $|T^{-1/2}\sigma([sT])\, z([sT]+1)| \leq K\ T^{-1/2} \sup_{s\in[0,1]} |z([sT])| \xrightarrow{p} 0$ since

$\max_t |z([sT])|$ is $o_p(T^{1/2})$, see Hansen (1992b), formula A.3. Regarding the first term we can consider the inequality

$$\left| \frac{1}{T^{1/2}} \sum_{t=1}^{[sT]} (\sigma(t) - \sigma(t-1)) z([sT]) \right| \leq \frac{\max_t |z(t)|}{T^{1/2}} \sum_{t=1}^{[sT]} |\sigma(t) - \sigma(t-1)| \ .$$

Since $T^{-1/2} \max_t |z(t)| = o_p(1)$, it is only left to prove that $\sum |\sigma(t) - \sigma(t-1)|$ is asymptotically bounded. But since $\omega(\cdot)$ is first-order Lipschitz continuous except at $N < \infty$ points of discontinuity,

$$\begin{aligned} \sum_{t=1}^{[sT]} |\sigma(t) - \sigma(t-1)| &= \sum_{t=1}^{[sT]} \left| \omega\left(\frac{t}{T}\right) - \omega\left(\frac{t-1}{T}\right) \right| \leq 2KN + \sum_{t=1}^{[sT]} M \frac{1}{T} \\ &= 2KN + M \frac{[sT]}{T} \leq 2KN + M. \ \blacksquare \end{aligned}$$

Proof of Theorem 2. We prove the theorem for the DF $(\widehat{\rho})$ / PP $(\widehat{\rho})$ tests in the case of no deterministic corrections. The proof easily extends to the remaining tests.

Following Phillips (1987a) we can write DF $(\widehat{\rho})$ as

$$\mathrm{DF}(\widehat{\rho}) := \frac{T^{-1} X(T)^2 - T^{-1} \sum_{t=1}^{T} \Delta X(t)^2}{2T^{-2} \sum_{t=1}^{T} X(t-1)^2}$$

The weak convergence of Lemma 1 and the continuous mapping theorem ensure that $T^{-1}(X([sT])^2, \sum_{t=1}^{T} X(t-1)^2)$ converges weakly to $\overline{\sigma}^2 \lambda_\varepsilon^2 (B_\omega(s)^2, \int_0^1 B_\omega(s)^2 ds)$. Regarding the term $(1/T) \sum_{t=1}^{T} \Delta X(t)^2 = (1/T) \sum_{t=1}^{T} \sigma(t)^2 \varepsilon(t)^2 =: \widehat{\sigma}_0^2$, it is sufficient to refer to Lemma 8 and to apply a triangular array version of McLeish's strong law of large numbers (see e.g. White, 1984) to get $(1/T) \sum_{t=1}^{T} \sigma(t)^2 \varepsilon(t)^2 - (1/T) \sum_{t=1}^{T} \mathrm{E}(\sigma(t)^2 \varepsilon(t)^2) \overset{a.s.}{\to} 0$; since $(1/T) \sum_{t=1}^{T} \mathrm{E}(\sigma(t)^2 \varepsilon(t)^2) = (1/T) \sum_{t=1}^{T} \sigma(t)^2 \to \overline{\sigma}^2$ the desired result follows.

Taken together, these results prove the theorem for the DF $(\widehat{\rho})$ statistic in the absence of deterministic corrections and for $\lambda_\varepsilon^2 =$

1. Note that in order to extend the proof to the remaining tests it is needed to prove that $\widehat{\sigma}_u^2 \xrightarrow{p} \overline{\sigma}^2\lambda_\varepsilon^2$. But this convergence follows immediately since $\widehat{\sigma}_u^2 = \widehat{\sigma}_0^2 + \mathrm{O}_p\left(T^{-1}\right)$.

To prove that the asymptotic distribution of the PP $(\widehat{\rho})$ statistic does not depend on λ_ε^2, consider the equality

$$\mathrm{PP}\left(\widehat{\rho}\right) = \frac{T^{-1}X\left(T\right)^2 - \widehat{\sigma}_0^2 - (\widehat{\lambda}_u^2 - \widehat{\sigma}_u^2)}{2T^{-2}\sum_{t=1}^{T} X\left(t-1\right)^2} \ .$$

Since (i) $(T^{-1}X\left(T\right)^2, T^{-2}\sum_{t=1}^{T} X\left(t-1\right)^2)' \xrightarrow{w} \overline{\sigma}^2\lambda_\varepsilon^2 \left(B\left(1\right)^2, \int B\left(s\right)^2 ds\right)'$, (ii) $\widehat{\sigma}_u^2, \widehat{\sigma}_0^2 \xrightarrow{p} \overline{\sigma}^2$ and (iii) $\widehat{\lambda}_u^2 \xrightarrow{p} \overline{\sigma}^2\lambda_\varepsilon^2$, see Lemma 12, the desired result immediately follows. ∎

Proof of Lemma 4. We follow the proof of Lemma 1 and Phillips (1987b). By recursive substitution $X_T\left(s\right)$ can be expressed as

$$\begin{aligned} X_T\left(s\right) &= \frac{1}{T^{1/2}} X[sT] = \frac{1}{T^{1/2}} \sum_{t=1}^{[sT]} e^{-c([sT]-t)/T} \sigma\left(t\right) \varepsilon\left(t\right) \\ &= \int_0^s e^{-c(s-r)} dS_T\left(r\right) + \mathrm{O}_p\left(T^{-1/2}\right), \end{aligned}$$

where $S_T\left(s\right) := T^{-1/2}\sum_{t=0}^{[sT]} \sigma\left(t/T\right)\varepsilon\left(t\right)$ has bounded variation, see Phillips (1987b, p.539). Integration by parts leads to the equality

$$\int_0^s e^{-c(s-r)} dS_T\left(r\right) = S_T\left(s\right) + c\int_0^s e^{-c(s-r)} S_T\left(r\right) dr$$

and, since $S_T\left(s\right) \xrightarrow{w} \overline{\sigma}\lambda_\varepsilon B_\omega\left(s\right)$, see Lemma 1, the continuous mapping theorem ensures the weak convergence

$$X_T\left(s\right) \xrightarrow{w} \overline{\sigma}\lambda_\varepsilon B_\omega\left(s\right) - c\overline{\sigma}\lambda_\varepsilon \int_0^s e^{-c(s-r)} B_\omega\left(r\right) dr$$

which can be equivalently written as the stochastic integral

$$X_T\left(s\right) \xrightarrow{w} \overline{\sigma}\lambda_\varepsilon \int_0^s e^{-c(s-r)} dB_\omega\left(r\right) = \overline{\sigma}\lambda_\varepsilon \int_0^s e^{-c(s-r)} \omega\left(s\right) dB\left(r\right) \ . \ ∎$$

PROOF OF THEOREM 5. By direct consequence of Lemma 4 and the continuous mapping theorem, in order to prove the theorem we can follow the proof of Theorem 2, with the only exception of considering $\{J_\omega(\cdot)\}$ as the limit Brownian functional. Regarding $\widehat{\sigma}_0^2$ and $\widehat{\sigma}_u^2$, since $\widehat{\phi} = \exp(-c/T) + \mathrm{O}_p(T^{-1}) = 1 + \mathrm{O}_p(T^{-1})$, then $\widehat{\sigma}_u^2 = \widehat{\sigma}_0^2 + \mathrm{O}_p(T^{-1})$ and $\widehat{\sigma}_0^2 = T^{-1}\sum_{t=1}^T \sigma(t)^2 \varepsilon(t)^2 + \mathrm{O}_p(T^{-1})$. As $T^{-1}\sum_{t=1}^T \sigma(t)^2 \varepsilon(t)^2 \xrightarrow{p} \overline{\sigma}^2$, see Theorem 2, the proof is completed. ∎

PROOF OF PROPOSITION 3. Since (i) $\mathrm{E}(B_\omega(s)) = 0$, (ii) $\mathrm{E}(B_\omega(s)^2) = \overline{\sigma}^{-2}\,\mathrm{E}((\int_0^s \omega(s)\,dB(s))^2)$
$= \overline{\sigma}^{-2}\int_0^s \omega(s)^2\,ds =: \eta(s)$, (iii) $\{B_\omega(\cdot)\}$ has independent increments and (iv) $\eta(\cdot)$ is an increasing homeomorphism on $[0,1]$ with $\eta(0) = 0$, $\eta(1) = 1$, from Corollary 29.10 of Davidson (1994) it follows that $\{B_\omega(\cdot)\}$ is distributed as $\{B_\eta(\cdot)\}$. ∎

PROOF OF COROLLARY 6. To prove the corollary we refer to Theorem 2 and to the equality $B_\omega(s) \stackrel{d}{=} B(\eta(s))$, see Proposition 3. Under the assumption of the corollary, $\eta(\cdot)$ is given by the piecewise linear function

$$\eta(s) = \frac{\delta^2 s + (1-\delta^2)(s-\tau)\,\mathbb{I}(s>\tau)}{\delta^2\tau + 1 - \tau}\,, \tag{29}$$

which implies

$$\int_0^1 B(\eta(s))^2\,ds = \int_0^\tau B\left(\tfrac{\delta^2 s}{\delta^2\tau+1-\tau}\right)^2 ds + \int_\tau^1 B\left(\tfrac{s-\tau+\delta^2\tau}{\delta^2\tau+1-\tau}\right)^2 ds$$

$$= \left(\delta^2\tau + 1 - \tau\right)\left(\frac{1}{\delta^2}\int_0^{a(\delta,\tau)} B(r)^2\,dr + \int_{a(\delta,\tau)}^1 B(r)^2\,dr\right)$$

with $a(\delta,\tau) := \delta^2\tau/\left(\delta^2\tau+1-\tau\right)$. By adding and subtracting $\int_0^1 B(r)^2\,dr$, after some algebra the desired result is obtained. ∎

PROOF OF COROLLARY 7. As in the previous proof it is useful to consider the equality $B_\omega(s) \stackrel{d}{=} B(\eta(s))$, where $\{B(\cdot)\}$ is a standard

Brownian motion. In this case, some algebra allows to get $\eta(s) = \left(1+\delta^2\right)^{-1}(2\delta^2 s + (1-\delta^2)s^2)$, and a simple application of the change-of-variables rule yields

$$\begin{aligned}
\int_0^1 B(\eta(s))^2\, ds &= \int_0^1 B\left(\frac{(2\delta^2 + (1-\delta^2)s}{1+\delta^2} s\right)^2 ds \\
&= \frac{1}{2}\int_0^1 \frac{1+\delta^2}{\sqrt{(\delta^4 + (1-\delta^4)s)}} B(s)^2\, dr
\end{aligned}$$

The proof is completed by adding and subtracting $\int_0^1 B(s)^2\, ds$, rearranging and defining $f(\cdot)$ as in (26). ∎

Proof of Lemma 8. It follows by simply extending White (1984), Lemma 6.18, to the case of mixing triangular arrays. ∎

Proof of Lemma 9. let $\lambda^2_{\varepsilon,T} := \sum_{k=-T+1}^{T-1} h(k/q_T)\gamma_\varepsilon(k)$, $\gamma_\varepsilon(k) := \mathrm{E}(\varepsilon(t)\varepsilon(t+|k|))$. Since $\lambda^2_{\varepsilon,T} \to \lambda^2_\varepsilon$, see de Jong (2000), we only need to prove that $\lambda^2_{u,T} - \overline{\sigma}^2\lambda^2_{\varepsilon,T}$ is $\mathrm{o}(1)$.

Set $\Delta_T(k) := |\gamma_{u,T}(k) - \overline{\sigma}^2\gamma_\varepsilon(k)|$. Since $\mathrm{E}(u(t)u(t+k))$ can be written as $\sigma(t)\sigma(t+k)\mathrm{E}(\varepsilon(t)\varepsilon(t+k))$, the following inequality holds

$$\begin{aligned}
\Delta_T(k) &\leq |\mathrm{E}(\varepsilon(t)\varepsilon(t+|k|))| \left|\frac{1}{T}\sum_{t=1}^{T-|k|}\left(\sigma(t)\sigma(t+|k|) - \overline{\sigma}^2\right)\right| \\
&= |\mathrm{E}(\varepsilon(t)\varepsilon(t+|k|))| c(T,k)
\end{aligned}$$

where $c(T,k) := \left|(1/T)\sum_{t=1}^{T-|k|}\left(\sigma(t)\sigma(t+|k|) - \overline{\sigma}^2\right)\right|$. This result implies that

$$\begin{aligned}
\left|\lambda^2_{u,T} - \overline{\sigma}^2\lambda^2_\varepsilon\right| &= \left|\sum_{k=-T+1}^{T-1} h\left(\frac{k}{q_T}\right)\Delta_T(k)\right| \\
&\leq \left|\sum_{k=-T+1}^{T-1} h\left(\frac{k}{q_T}\right) c(T,k)\,|\mathrm{E}(\varepsilon(t)\varepsilon(t+|k|))|\right|
\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{-T<k<T} c(T,k)\, h\left(\frac{k}{q_T}\right) \sum_{k=-T+1}^{T-1} |\operatorname{E}(\varepsilon(t)\,\varepsilon(t+|k|))| \\
&\leq A \sup_{-T<k<T} c(T,k)\, h\left(\frac{k}{q_T}\right)
\end{aligned}
$$

since for some $a > 0$ $\sum_{k=0}^{\infty} |\operatorname{E}(\varepsilon(t)\,\varepsilon(t+|k|))| < a \sum_{k=0}^{\infty} \alpha_m^{1-2/p} < a \sum_{k=0}^{\infty} \alpha_m^{2(1/r-1/p)} =: A < \infty$, where the first inequality follows e.g. from Davidson (1994, Corollary 14.3) while the second follows from the inequality $1-2/p > 2(1/r-1/p)$, always true under Assumption $\mathcal{E}$. Hence, what is left to be proved is that $\lim_{T\to\infty} \sup_k c(T,k)\, h(k/q_T) = 0$. Since as $T \to \infty$, $c(T,k)$ is bounded for all admissible values of k while $h(k/q_T) \to 0$ if k/q_T is not $\operatorname{O}(1)$, it is sufficient to consider the supremum over all k which are $\operatorname{O}(q_T)$. For $k \geq 0$

$$
\begin{aligned}
c(T,k) &= \left| \frac{1}{T} \sum_{t=1}^{T-|k|} \left(\sigma(t)\,\sigma(t+|k|) - \overline{\sigma}^2\right) \right| \\
&= \left| \frac{1}{T} \sum_{t=1}^{T-|k|} \sigma(t)\,\sigma(t+|k|) - \frac{1}{T} \sum_{t=1}^{T-|k|} \sigma(t)^2 \right. \\
&\quad \left. + \frac{1}{T} \sum_{t=1}^{T-|k|} \sigma(t)^2 - \frac{T-|k|}{T} \overline{\sigma}^2 \right| \\
&= \left| \frac{1}{T} \sum_{t=1}^{T-|k|} \sigma(t)\,\sigma(t+|k|) - \frac{1}{T} \sum_{t=1}^{T-|k|} \sigma(t)^2 \right| \\
&\quad + \left| \frac{1}{T} \sum_{t=1}^{T-|k|} \sigma(t)^2 - \frac{T-|k|}{T} \overline{\sigma}^2 \right| \qquad (30)
\end{aligned}
$$

For $|k| = \operatorname{O}(q_T) = \operatorname{o}(T)$ and $(1/T) \sum_{t=1}^{T} \sigma(t)^2 \to \overline{\sigma}^2$, the term in (30) converges to 0. Concerning the first term, the following inequalities hold

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

which converges to 0 since $\sup \omega(s)$ is bounded and $\sum_{t=1}^{T-|k|} |\omega(t/T) - \omega(t/T + |k|/T)|$ is $\mathrm{O}(1)$, see the proof of Lemma 1 for $k = 1$. This proves that $(1/T) \sum_{t=1}^{T-|k|} \sigma(t)\,\sigma(t+|k|) \to \overline{\sigma}^2$ for $|k| = \mathrm{O}(q_T)$ and consequently that $\limsup c(T,k)\,h(k/q_T) = 0$, thus completing the proof of the Lemma ∎

Proof of Lemma 10. Since $\{u(t)\}$ is mixing of the same size of $\{\varepsilon(t)\}$, see Lemma 8, it is sufficient to adapt the proof of Theorem 2 in de Jong (2000). ∎

Proof of Lemma 11. It can be easily obtained by adapting Theorem 3 in Hansen (1992a). ∎

References

Andrews D.W.K. (1991), *Heteroskedasticity and autocorrelation consistent covariance matrix estimation*, Econometrica 59, 817–58.

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

Cavaliere G. (2003), *Asymptotics for unit root tests under Markov-regime switching*, Econometrics Journal 6, 193–216.

Corradi V. (2000), *Reconsidering the continuous time limit of the GARCH(1,1) process*, Journal of Econometrics 96, 145–153.

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.

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. (1992a), *Consistent covariance matrix estimation for dependent heterogeneous processes*, Econometrica 60, 967–972.

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

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.

Harvey D.I., Leybourne S.J. and P. Newbold (2001), *Innovational outlier unit root tests with an endogenously determined break in level*, Oxford Bulletin of Economics and Statistics 63, 559–575.

de Jong R. (2000), *A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators*, Econometric Theory 16, 262–268.

Kim J.Y. (2000), *Detection of change in persistence of a linear time series*, Journal of Econometrics 95, 97–116.

Kim K. and P. Schmidt (1993), *Unit root tests with conditional heteroskedasticity*, Journal of Econometrics 59, 287–300.

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

Leybourne S.J., T.C. Mills and P. Newbold (1998), *Spurious rejections by Dickey–Fuller tests in the presence of a break under the null*, Journal of Econometrics 87, 191–203.

Leybourne S.J. and P. Newbold (2000a), *Behaviour of the standard and symmetric Dickey–Fuller-type tests when there is a break under the null hypothesis*, Econometrics Journal 3, 1–15.

Leybourne S.J. and P. Newbold (2000b), *Behaviour of Dickey–Fuller t-tests when there is a break under the alternative hypothesis*, Econometric Theory 16, 779–789.

McLeish D.L. (1975), *A maximal inequality and dependent strong laws*, Annals of Probability 3, 829–839.

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.

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.

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

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

Sargan J.D. and A. Bhargava (1983), *Testing residuals from least squares regression for being generated by the Gaussian random walk*, Econometrica 51, 153–174.

Stock J.H. (1990), *A class of tests for integration and cointegration*, Manuscript, Harvard University.

Stock J.H. (1994), *Unit roots, structural breaks, and trends*, in Engle, R.F. and D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2740-2843.

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.

White H. (1984), *Asymptotic theory for econometricians*, New York: Academic Press.

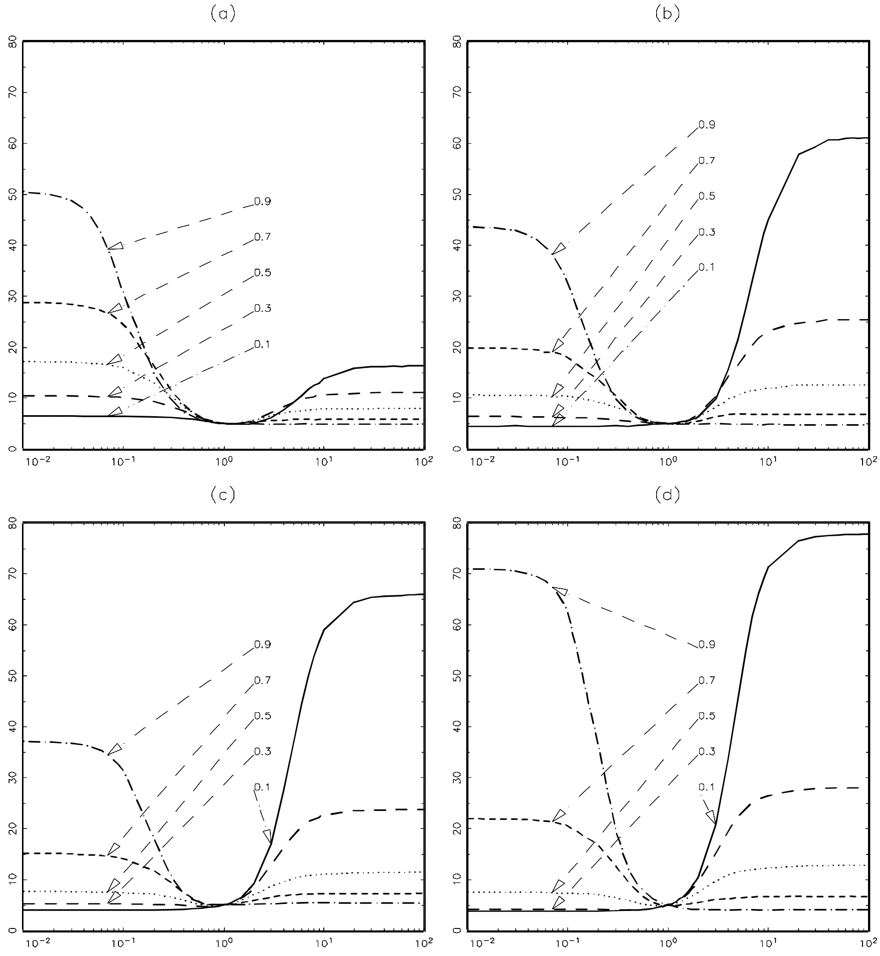

Figure 1: Asymptotic size of the DF $(\hat{\rho})$ / PP $(\hat{\rho})$ tests under a single variance shift for $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and for δ between 1/100 and 100. Nominal size: 0.05. Panel (a): no deterministics correction; Panel (b): constant-corrected test; Panel (c): constant and trend-corrected test; Panel (d): constant and broken trend-corrected (Perron) test.

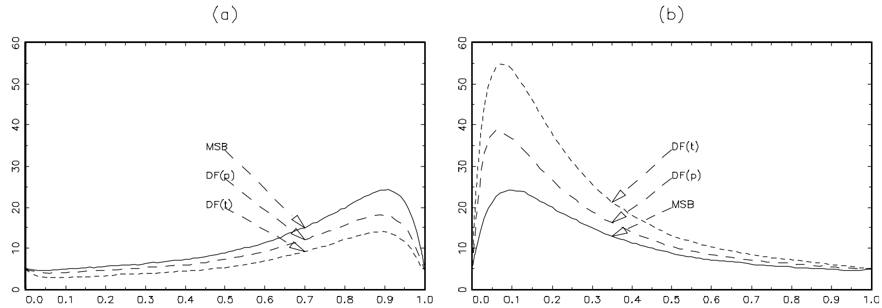

Figure 2: Asymptotic size of the DF(ρ), DF(t) and MSB tests under a single variance shift for τ between 0 and 1. Panel (a): $\delta = 0.2$ (positive variance shift); Panel (b): $\delta = 5$ (negative variance shift). Constant and trend-corrected tests. Nominal size: 0.05

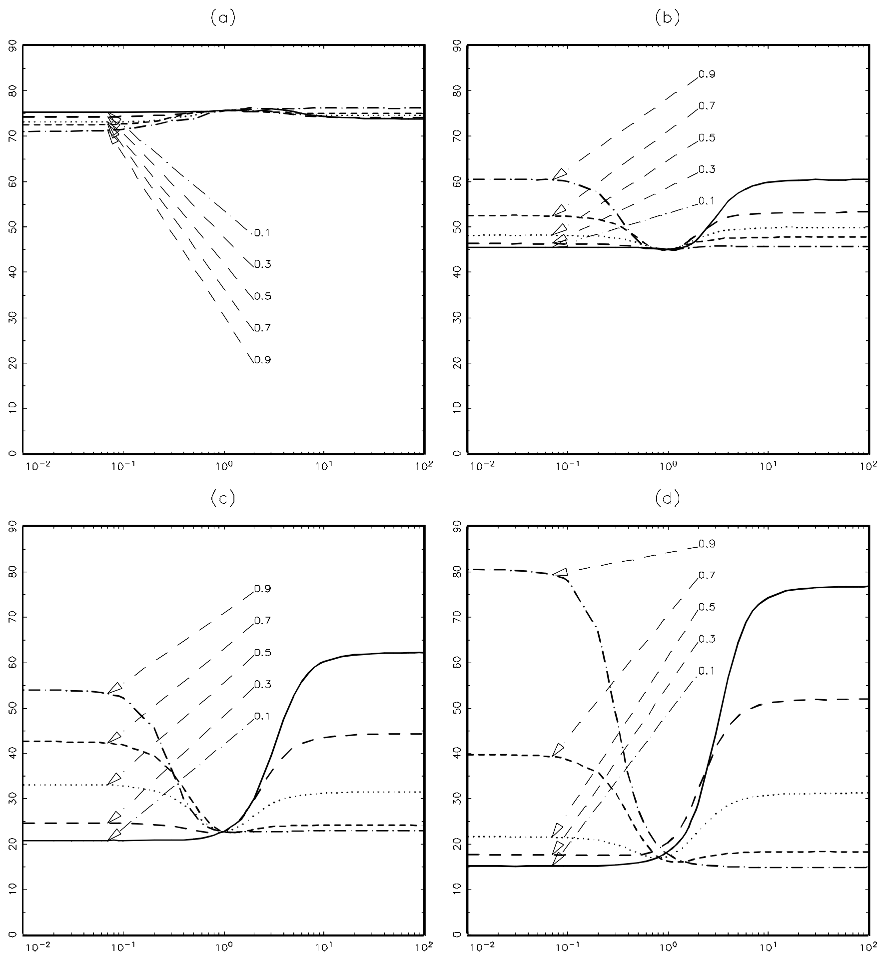

Figure 3: Asymptotic power function of the DF $(\hat{\rho})$ / PP $(\hat{\rho})$ tests under a single variance shift for $\tau \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and for δ between 1/100 and 100. The location parameter c is set to 10. Nominal size: 0.05. Panel (a): no deterministics correction; Panel (b): constant-corrected test; Panel (c): constant and trend-corrected test; Panel (d): constant and broken trend-corrected (Perron) test.

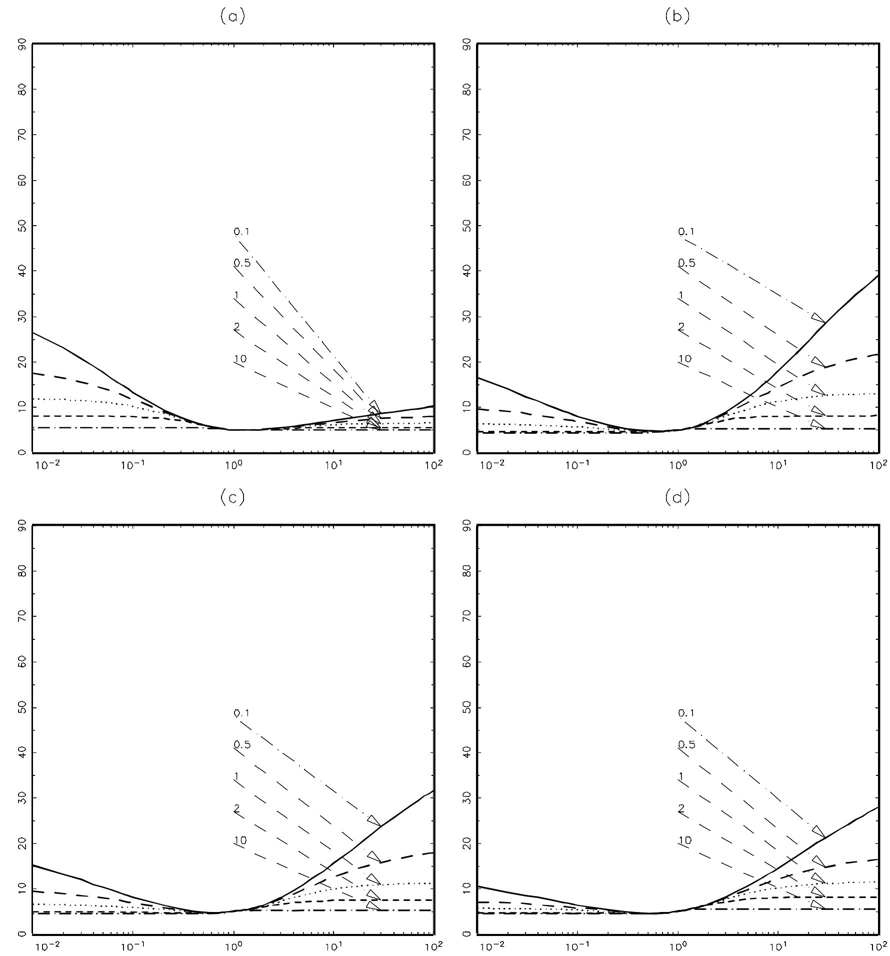

Figure 4: Asymptotic size of the DF $(\hat{\rho})$ / PP $(\hat{\rho})$ tests under trending variances for $r \in \{0.1, 0.5, 1, 2, 10\}$ and for δ between 1/100 and 100. Nominal size: 0.05. Panel (a): no deterministics correction; Panel (b): constant-corrected test; Panel (c): constant and trend-corrected test; Panel (d): constant and broken trend-corrected (Perron) test with $\tau = 0.5$.

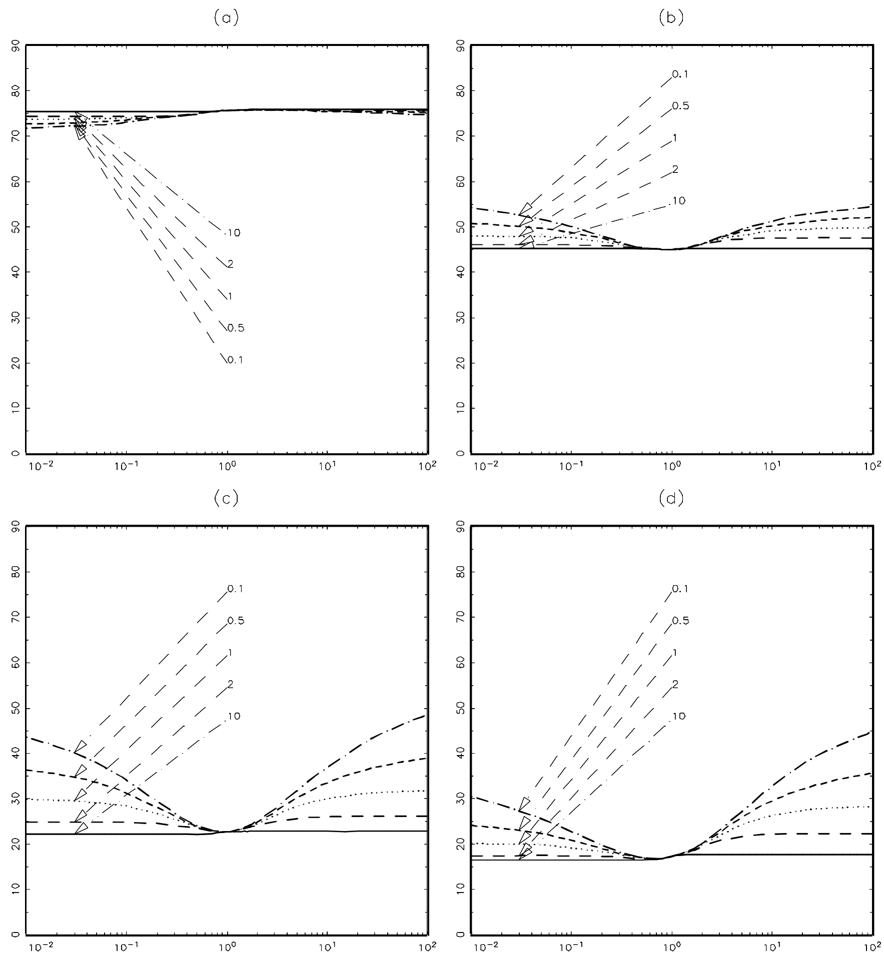

Figure 5: Asymptotic power function of the DF $(\hat{\rho})$ / PP $(\hat{\rho})$ tests under trending variances for $r \in \{0.1, 0.5, 1, 2, 10\}$ and for δ between 1/100 and 100. The location parameter c is set to 10. Nominal size: 0.05. Panel (a): no deterministics correction; Panel (b): constant-corrected test; Panel (c): constant and trend-corrected test; Panel (d): constant and broken trend-corrected (Perron) test with $\tau = 0.5$.