Econometric Theory, **21**, 2005, 1112–1129. Printed in the United States of America.
DOI: 10.1017/S0266466605050553

STATIONARITY TESTS UNDER TIME-VARYING SECOND MOMENTS

GIUSEPPE CAVALIERE
University of Bologna

A.M. ROBERT TAYLOR
University of Birmingham

In this paper we analyze the effects of a very general class of time-varying variances on well-known "stationarity" tests of the $I(0)$ null hypothesis. Our setup allows, among other things, for both single and multiple breaks in variance, smooth transition variance breaks, and (piecewise-) linear trending variances. We derive representations for the limiting distributions of the test statistics under variance breaks in the errors of $I(0)$, $I(1)$, and near-$I(1)$ data generating processes, demonstrating the dependence of these representations on the precise pattern followed by the variance processes. Monte Carlo methods are used to quantify the effects of fixed and smooth transition single breaks and trending variances on the size and power properties of the tests. Finally, bootstrap versions of the tests are proposed that provide a solution to the inference problem.

1. INTRODUCTION

Applied researchers have recently focused attention on the question of whether or not the variability in the shocks driving macroeconomic time series has changed over time (see, e.g., the literature review in Busetti and Taylor, 2003). The empirical evidence suggests that a decline in volatility over the past 20 years or so is a common phenomenon in many real and price variables. These findings have helped stimulate interest among econometricians in analyzing the effects of innovation variance shifts on unit root and stationarity tests. Among others, Hamori and Tokihisa (1997) and Kim, Leybourne, and Newbold (2002) have derived the implications of a single permanent variance shift in the innovations of an $I(1)$ process on the size properties of Dickey–Fuller tests. The effect of a single variance shift on the stationarity test (KPSS test) of Kwiatkowski, Phillips, Schmidt, and Shin (1992) has been analyzed independently by Busetti and Taylor (2003) and Cavaliere (2004a), who found that the test can suffer severe size distortions when there is a late (early) positive (negative) variance shift under the null.

We are grateful to Peter Phillips, a co-editor, and two anonymous referees whose comments on an earlier draft have led to a considerable improvement in the paper. Address correspondence to Robert Taylor, Department of Economics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom; e-mail: R.Taylor@bham.ac.uk.

  0266-4666/05 $12.00

We analyze the effects that a very general class of permanent variance breaks has on the behavior of the KPSS stationarity test, together with those of Lo (1991) and Xiao (2001); a brief review of these tests is given in Section 3. Our unobserved components model, introduced in Section 2, generalizes that considered in, inter alia, Kwiatkowski et al. (1992) to allow for innovation processes whose variances evolve over time according to a quite general mechanism that allows, e.g., single and multiple breaks, smooth transition breaks, and trending variances. Variance nonconstancy is allowed in both the irregular component and the errors driving the level of the process. In Sections 4 and 5 we analyze the effects of time-varying variances on the large-sample behavior of these statistics under both the $I(0)$ null and global $I(1)$ and local alternatives. In Section 6 these effects are quantified, using Monte Carlo simulation, for the aforementioned examples.

Related but different work was carried out by Hansen (2000), who shows that the Lagrange multiplier (LM) test of Nyblom (1989) for structural change in the parameters of a linear regression model (which contains the KPSS test as a special case) underrejects the $I(0)$ null when the marginal distribution of the regressors changes over time. Conversely, in this paper we show that where the variance of the errors changes over time the picture is quite different, with the KPSS (and other stationarity) tests both under- and overrejecting the null, but with a more pronounced tendency toward overrejecting. Similarly, whereas Hansen (2000) shows that Nyblom's test loses (size-unadjusted) power under structural changes in the marginal distribution of the regressors, for most of the cases we consider the KPSS test gains power when the errors are heteroskedastic. In Section 7 we adapt the heteroskedastic bootstrap of Hansen (2000) to the present problem and show that the bootstrap tests perform well in practice. Section 8 concludes. Sketch proofs are given in an Appendix; detailed proofs appear in Cavaliere and Taylor (2004).

We use $\xrightarrow{w}$, $\mathbb{I}(\cdot)$, and $\mathcal{D}$ to denote weak convergence as the sample size diverges, the indicator function, and the space of cadlag processes on $[0,1]$ endowed with the Skorohod metric, respectively, whereas $x := y$ means that x is defined by y. Finally, as in Phillips and Sun (2001), for two processes X and Y on $[0,1]$ we define the projections $P_Z X(s) := \int_0^1 X(r) \times Z(r)'\,dr(\int_0^1 Z(r)Z(r)'\,dr)^{-1}Z(s)$ and $Q_Z X(s) := \int_0^1 dX(r)Z^{(1)}(r)'(\int_0^1 Z^{(1)}(r) \times Z^{(1)}(r)'\,dr)^{-1}Z(s)$, where $^{(1)}$ denotes the first derivative.

2. THE UNOBSERVED COMPONENTS MODEL

Consider the unobservable components (UC) data generating process (DGP)

$$y_t := x_t'\beta + \mu_t + \sigma_t \varepsilon_t, \qquad t = 1,2,\ldots,T, \tag{1}$$

$$\mu_t = \mu_{t-1} + \sigma_{\eta t}\eta_t, \qquad \eta_t \sim IID(0,1), \qquad \mu_0 \overset{a.s.}{=} 0, \tag{2}$$

https://doi.org/10.1017/S0266466605050553 Published online by Cambridge University Press

under the following set of assumptions (which are taken to hold throughout the paper, except where stated otherwise).

Assumption $\mathcal{V}$. The term $\{\sigma_t\}$ satisfies $\sigma_{[sT]} := \omega(s)$, where $\omega(\cdot) \in \mathcal{D}$ is a nonstochastic 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. Similarly, except where otherwise stated, $\sigma_{\eta[sT]} := \omega_\eta(s)$, where $\omega_\eta(\cdot)$ satisfies the same conditions as $\omega(\cdot)$.

Assumption $\mathcal{E}$. The irregular component $\{\varepsilon_t\}$ is a zero-mean, unit variance, strictly stationary mixing process with $E|\varepsilon_t|^p < \infty$ 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 $r \in (2,4]$, $r \le p$. The long-run variance $\lambda_\varepsilon^2 := \sum_{k=-\infty}^{\infty} E(\varepsilon_t \varepsilon_{t+k})$ is strictly positive and finite. Furthermore, $\{\varepsilon_t\}$ is independent of $\{\eta_t\}$ at all leads and lags. As is standard, we refer to $\{\varepsilon_t\}$ as an $I(0)$ process.

Assumption $\mathcal{X}$. The component x_t is a $p \times 1$ deterministic vector satisfying the condition that there exist a scaling matrix δ_T and a bounded piecewise-continuous function $F(\cdot)$ on $[0,1]$ such that $\delta_T x_{[\cdot T]} \to F(\cdot)$ uniformly on $[0,1]$, with $\int_0^1 F(s)F(s)'\,ds$ positive definite.

From (1) and (2), observe that under Assumption $\mathcal{V}$, the variance of both the irregular component, $u_t := \sigma_t \varepsilon_t$, and the shocks to the level process $\{\mu_t\}$ are heteroskedastic. Consequently, $\{u_t\}$ is $I(0)$ provided $\{\sigma_t\}$ is constant, whereas $\{\mu_t\}$ reduces to a standard random walk if $\{\sigma_{\eta t}\}$ is constant and vanishes from (1) when $\sigma_{\eta t} = 0$, all t. Notice that the model considered here generalizes the UC model discussed in Kwiatkowski et al. (1992) by allowing both $\{\sigma_t\}$ and $\{\sigma_{\eta t}\}$ to be potentially nonconstant over time.[1]

Assumption $\mathcal{V}$ allows for a wide class of models for the variances of the errors. Models of single or multiple variance shifts satisfy Assumption $\mathcal{V}$ with $\omega(\cdot)$ piecewise constant. For example, the function $\omega(s) := \sigma_0 + (\sigma_1 - \sigma_0) \times \mathbb{I}(s \ge m)$ gives the single break model with a variance shift at time $[mT]$, $0 < m < 1$, analyzed by Busetti and Taylor (2003) and Cavaliere (2004a). If $\omega(\cdot)^2$ is an affine function, then the unconditional variance of the errors displays a linear trend. Piecewise-affine functions are also permitted, allowing for variances that follow a broken trend. Moreover, smooth transition variance shifts also satisfy Assumption $\mathcal{V}$: e.g., the function $\omega(s)^2 := \sigma_0^2 + (\sigma_1^2 - \sigma_0^2)\mathbb{S}(s)$, $\mathbb{S}(s) = (1 + \exp(-\gamma(s-m)))^{-1}$, which corresponds to a smooth (logistic) transition from σ_0^2 to σ_1^2. The parameter m determines the transition midpoint (for $t = [mT]$, $\sigma_t^2 = 0.5(\sigma_0^2 + \sigma_1^2)$) whereas $\gamma > 0$ controls the speed of transition (the fixed change-point model follows as a limiting case for $\gamma \to \infty$).

Assumption $\mathcal{X}$ is standard and allows for a wide variety of possible forms for the deterministic component, including the pth-order trend function $x_t := (1, t, \ldots, t^p)'$, $0 \le p < \infty$. The broken intercept and broken intercept and trend functions considered, e.g., in Busetti and Harvey (2001) are obtained by specifying $x_t'\beta := \sum_{j=0}^{i} \beta_j t^j + \sum_{j=0}^{i} \beta_{m,j} t_m^j$ for $i = 0,1$ respectively, in (1), t_m^j being defined as $t_m^j := (t-m)^j \mathbb{I}(t > m)$ and m satisfying $\lim_{T\to\infty}(m/T) = \mu \in (0,1)$ (see Phillips and Xiao, 1998, p. 448).

Remark 1. If $\omega(\cdot)$ is not constant then the irregular component, $\{u_t\}$, is unconditionally heteroskedastic. Conditional heteroskedasticity is also permitted through Assumption $\mathcal{E}$ (see, e.g., Hansen, 1992b). Assumption $\mathcal{E}$ has been used extensively in the econometric literature as it allows $\{\varepsilon_t\}$ to belong to a wide class of weakly dependent stationary processes. The strict stationarity assumption is made without loss of generality and may be weakened to allow for weak heterogeneity of the errors, as in, e.g., Phillips (1987).

Remark 2. The assumption of nonstochastic variance functions $\{\omega(\cdot), \omega_\eta(\cdot)\}$ can be easily weakened simply by assuming stochastic independence between $\{\varepsilon_t, \eta_t\}$ and $\{\sigma_t, \sigma_{\eta t}\}$, given that the stochastic functionals $\{\omega(\cdot), \omega_\eta(\cdot)\}$ must have sample paths satisfying the requirements of Assumption $\mathcal{V}$. In the stochastic variance framework, the results given in this paper hold *conditionally* on a given realization of $\{\omega(\cdot), \omega_\eta(\cdot)\}$.

3. STATIONARITY TESTS

Kwiatkowski et al. (1992) focus on testing the $I(0)$ null hypothesis, $H_0: \sigma_\eta^2 = 0$, against the $I(1)$ alternative hypothesis, $H_1: \sigma_\eta^2 > 0$, under the ancillary assumption that $\sigma_t = \sigma, \sigma_{\eta t} = \sigma_\eta$, all t, so that, under H_0, $\{y_t\}$ reduces to the $I(0)$ process $y_t = x_t'\beta + u_t$, $t = 1, \ldots, T$. Kwiatkowski et al. (1992) propose the test that rejects H_0 for large values of the statistic

$$\mathcal{KPSS} := \frac{\sum_{t=1}^{T} \hat{S}_t^2}{T^2 \hat{\lambda}^2}, \tag{3}$$

where $\hat{S}_t := \sum_{i=1}^t \hat{u}_i$, $\hat{u}_t$, the ordinary least squares (OLS) residuals from the regression of y_t on x_t, $t = 1, \ldots, T$; $\hat{\lambda}^2$ is a consistent estimator of the long-run variance of $\{u_t\}$ under H_0 and has the form $\hat{\lambda}^2 := \sum_{j=-T+1}^{T-1} k(jq_T^{-1})\hat{\gamma}(j)$, $\hat{\gamma}(j) := T^{-1}\sum_{t=|j|+1}^{T} \hat{u}_t \hat{u}_{t-|j|}$, q_T being a bandwidth parameter and $k(\cdot)$ a weighting function. Kwiatkowski et al. (1992) assume $1/q_T + T^{-1/2} q_T \to 0$ as $T \to \infty$ and $k(x) := 1 - |x|\mathbb{I}(|x| \le 1)$ (Bartlett weights). However, because we are dealing with mixing errors (see Assumption $\mathcal{E}$), throughout the paper we will require that q_T and $k(\cdot)$ satisfy the following assumption (de Jong, 2000).

Assumption $\mathcal{K}$. (K_1) For all $x \in \mathbb{R}$, $|k(x)| \le 1$, $k(x) = k(-x)$; $k(0) = 1$; $k(x)$ is continuous at 0 and for almost all $x \in \mathbb{R}$; $\int_{-\infty}^{\infty}|k(x)|dx < \infty$; $|k(x)| \le l(x)$, where $l(x)$ is a nonincreasing function such that $\int_{-\infty}^{\infty}|x||l(x)|dx < \infty$; (K_2) $q_T \uparrow \infty$ as $T \uparrow \infty$, and $q_T = o(T^\gamma)$, $\gamma \le 1/2 - 1/r$, where r is given in $\mathcal{E}$.

Assumption $\mathcal{K}$ is sufficiently general for our purposes as it is satisfied by many of the most commonly employed kernels (see Hansen, 1992a; Jansson, 2002).

Remark 3. The $\mathcal{KPSS}$ statistic maps the sequence $\{\hat{S}_t\}$ onto [0,1] by averaging the squared values of the sequence. Other stationarity tests can be obtained by taking different mappings. For example, the supremum of $\{|\hat{S}_t|\}$ and range of $\{\hat{S}_t\}$ deliver, respectively, the test of Xiao (2001) and the rescaled range (RS) test of Lo (1991), which reject H_0 for large values of the statistics $\mathcal{KS} := \max_t|\hat{S}_t|/(\hat{\lambda}T^{1/2})$ and $\mathcal{RS} := (\max_t \hat{S}_t - \min_t \hat{S}_t)/(\hat{\lambda}T^{1/2})$, respectively.

4. ASYMPTOTIC SIZE

Under the null hypothesis considered by Kwiatkowski et al. (1992), $H_0: \sigma^2_{\eta t} = \sigma^2_\eta = 0$, all t, if $\{\sigma_t\}$ is constant across the sample, it is well known that (e.g., Kwiatkowski et al., 1992, pp. 164–165) $\mathcal{KPSS} \xrightarrow{w} \int_0^1 V(s)^2\,ds$, where $V(s) := B(s) - Q_{\mathbb{F}}B(s)$, $\mathbb{F}(s) := \int_0^s F(r)\,dr$, with $B(\cdot)$ a standard Brownian motion. For example, if $x_t := (1,t,\ldots,t^{p-1})'$, then $F(s) := (1,s,\ldots,s^{p-1})'$ and $V(\cdot)$ is a pth-level Brownian bridge.

Now, assume that H_0 holds but that σ_t is not necessarily constant over time; rather it satisfies Assumption $\mathcal{V}$. Then, the asymptotic distribution of the $\mathcal{KPSS}$ statistic assumes the form detailed in the following theorem.

THEOREM 1. *Under* $H_0: \sigma^2_{\eta t} = \sigma^2_\eta = 0$, *all* t, $\mathcal{KPSS} \xrightarrow{w} \int_0^1 V_\omega(s)^2\,ds$, *where* $V_\omega(s) := B_\omega(s) - Q_{\mathbb{F}}B_\omega(s)$ *and where* $B_\omega(s) := \bar{\omega}^{-1}\int_0^s \omega(r)\,dB(r)$, $\bar{\omega} := (\int_0^1 \omega(s)^2\,ds)^{1/2}$.

Consequently, with respect to the homoskedastic case, the asymptotic distribution of the $\mathcal{KPSS}$ statistic has the usual structure but with $B(\cdot)$ replaced by $B_\omega(\cdot)$. It is only where $\omega(\cdot)$ is constant throughout the sample that $B_\omega(\cdot)$ reduces to a standard Brownian motion and, hence, that $\mathcal{KPSS}$ has the standard limiting distribution.

Remark 4. The process $B_\omega(\cdot)$ is a diffusion corresponding to the stochastic differential equation $dB_\omega(s) = (\omega(s)/\bar{\omega})dB(s)$ with initial condition $B_\omega(0) = 0$. Because $B_\omega(\cdot)$ has zero mean, variance $E((1/\bar{\omega})\int_0^s \omega(r)\,dB(r))^2 = (1/\bar{\omega}^2) \times \int_0^s \omega(r)^2\,dr =: \Lambda_\omega(s) \in [0,1]$ (where $\Lambda_\omega(\cdot)$ is an increasing homeomorphism on [0,1]) and has independent increments, Corollary 29.10 of Davidson (1994) implies that $B_\omega(\cdot)$ is distributed as $B(\Lambda_\omega(\cdot))$, and therefore at time $s \in [0,1]$, $B_\omega(\cdot)$ has the same distribution as the standard Brownian motion $B(\cdot)$ at time $\Lambda_\omega(s) \in [0,1]$. That is, $B_\omega(\cdot)$ is a Brownian motion under modification of the time domain (see, e.g., Revuz and Yor, 1991, p. 170).

Remark 5. Under the conditions of Theorem 1, $\mathcal{KS} \xrightarrow{w} \sup_{s\in[0,1]}|V_\omega(s)|$ and $\mathcal{RS} \xrightarrow{w} \sup_{s,s'\in[0,1]}|V_\omega(s) - V_\omega(s')|$. Interestingly, in the case of no deterministic terms (i.e., $x_t'\beta = 0$), because $V_\omega(s) \stackrel{d}{=} B(\Lambda_\omega(s))$ (see Remark 4), it holds

https://doi.org/10.1017/S0266466605050553 Published online by Cambridge University Press

that $\sup_{s\in[0,1]}|V_\omega(s)| \overset{d}{=} \sup_{s\in[0,1]}|B(\Lambda(s))| = \sup_{s\in[0,1]}|B(s)|$, $\sup_{s,s'\in[0,1]}|V_\omega(s) - V_\omega(s')| \overset{d}{=} \sup_{s,s'\in[0,1]}|B(\Lambda(s)) - B(\Lambda(s'))| = \sup_{s,s'\in[0,1]}|B(s) - B(s')|$, and the asymptotic sizes of the $\mathcal{KS}$ and $\mathcal{RS}$ tests are not affected by variance changes that satisfy Assumption $\mathcal{V}$. Simulation evidence reported in Cavaliere and Taylor (2004) suggests that this invariance property also holds reasonably well in small samples.

5. ASYMPTOTIC POWER

In this section we investigate the impact of time-varying variances in the irregular component in (1), and/or the error driving the level equation, (2), on both the consistency and local asymptotic power properties of the tests.

5.1. Consistency

It is well known (e.g., Kwiatkowski et al., 1992, eqn. (25)) that if $\sigma_{\eta t}^2 = \sigma_\eta^2 > 0$, then

$$\frac{\bar{k}q_T}{T}\,\mathcal{KPSS} \xrightarrow{w} \frac{\displaystyle\int_0^1\left(\int_0^s W(s)\,ds\right)^2 ds}{\displaystyle\int_0^1 W(s)^2\,ds}, \tag{4}$$

where $\bar{k} := \int_{-\infty}^{\infty} k(s)\,ds$, $W(s) := B_0(s) - P_F B_0(s)$, and $\{B_0(\cdot)\}$ is a standard Brownian motion independent of $B(\cdot)$. Because $q_T/T \to 0$, (4) implies that $\mathcal{KPSS}$ diverges to $+\infty$ at rate $O_p(T/q_T)$ under the $I(1)$ alternative. In addition to this result, note that if the $\{u_t\}$ component has a time-varying variance, $\mathcal{KPSS}$ is still distributed as in (4), because as $T \to \infty$, the $I(1)$ component $\{\eta_t\}$ dominates.

Now, consider the general case where $\sigma_{\eta t}^2 \neq 0$ but is not necessarily constant, satisfying Assumption $\mathcal{V}$. Here the following result holds.

THEOREM 2. *If $\sigma_{\eta t}^2 \neq 0$, all t, the weak convergence (4) holds with $W(\cdot)$ replaced by $W_{\omega_\eta}(\cdot)$, where $W_{\omega_\eta}(s) := B_{\omega_\eta}(s) - P_F B_{\omega_\eta}(s)$ with $B_{\omega_\eta}(s) := \bar{\omega}_\eta^{-1}\int_0^s \omega_\eta(r)\,dB_0(r)$, $\bar{\omega}_\eta := (\int_0^1 \omega_\eta(s)^2\,ds)^{1/2}$.*

Consequently, as in the case of constant variances, because $q_T/T \to 0$, Theorem 2 implies that $\mathcal{KPSS}$ diverges to $+\infty$ at rate $O_p(T/q_T)$ under global $I(1)$ alternatives.

Remark 6. Under the conditions of Theorem 2, $(\bar{k}q_T/T)^{1/2}\mathcal{KS} \xrightarrow{w} (\sup_{s\in[0,1]}|\int_0^s W_{\omega_\eta}(r)\,dr|)\ (\int_0^1 W_{\omega_\eta}(s)^2\,ds)^{-1/2}$ and $(\bar{k}q_T/T)^{1/2}\mathcal{RS} \xrightarrow{w} (\sup_{s,s'\in[0,1]}|\int_{s'}^s W_{\omega_\eta}(r)\,dr|)(\int_0^1 W_{\omega_\eta}(s)^2\,ds)^{-1/2}$, which imply that both $\mathcal{KS}$ and $\mathcal{RS}$ also diverge to $+\infty$, at rate $O_p((T/q_T)^{1/2})$.

5.2. Asymptotic Local Power

We now focus attention on the limiting behavior of the $\mathcal{KPSS}$ statistic under the local alternative (see also Busetti and Taylor, 2003, p. 513):

$$H_c : \sigma_{\eta t}^2 = \frac{c^2}{T^2}\,\omega_\eta\left(\frac{t}{T}\right)^2\left(\frac{\lambda_\varepsilon\bar{\omega}}{\bar{\omega}_\eta}\right)^2, \qquad t = 1,\dots,T, \tag{5}$$

where $c \geq 0$ is a noncentrality parameter and $\lambda_\varepsilon\bar{\omega}/\bar{\omega}_\eta > 0$ is a scale factor that simplifies the representation of the asymptotic distributions. Notice that $\bar{\omega}/\bar{\omega}_\eta = 1$ if $\sigma_t = \sigma_{\eta t}$, $t = 1,\dots,T$; i.e., if the pattern of time variation is common to the variances of the irregular component in (1) and the error driving the level in (2). Moreover, where $\sigma_t = \sigma$ and $\sigma_{\eta t} = \sigma_\eta$, $t = 1,\dots,T$, H_c reduces to the local alternative considered by, inter alia, Stock (1994, p. 2799).

The following theorem details the large-sample behavior of $\mathcal{KPSS}$ under H_c.

THEOREM 3. *Under H_c of (5),*

$$\mathcal{KPSS} \xrightarrow{w} \int_0^1\left[V_\omega(s) + c\int_0^s W_{\omega_\eta}(r)\,dr\right]^2 ds, \tag{6}$$

where the (independent) processes $V_\omega(\cdot)$ and $W_{\omega_\eta}(\cdot)$ are as previously defined.

Remark 7. Notice from (6) that the asymptotic local power of $\mathcal{KPSS}$ is affected by heteroskedasticity in both the irregular component and the errors driving the level process. Moreover, because the limiting processes relating to these components enter the asymptotic distribution in different forms ($W_{\omega_\eta}(\cdot)$ is integrated whereas $V_\omega(\cdot)$ is not), it is anticipated that heteroskedasticity will have different effects in these two cases.

Remark 8. Under the homoskedastic condition that $\sigma_t^2 = \sigma^2$ and $\sigma_{\eta t}^2 = \sigma_\eta^2$, for all t, the local alternative simplifies to $H_c : \sigma_\eta^2 = (c^2/T^2)\sigma^2\lambda_\varepsilon^2$, and the right member of (6) reduces to $\int_0^1[V(s) + c\int_0^s W(r)\,dr]^2\,ds$ (cf. Busetti and Taylor, 2003, p. 513).

Remark 9. Under the conditions of Theorem 3, $\mathcal{KS} \xrightarrow{w} \sup_{s\in[0,1]}|V_\omega(s) + c\int_0^s W_{\omega_\eta}(r)\,dr|$ and $\mathcal{RS} \xrightarrow{w} \sup_{s,s'\in[0,1]}|V_\omega(s) - V_\omega(s') + c\int_{s'}^s W_{\omega_\eta}(r)\,dr|$.

6. NUMERICAL RESULTS

In this section we use Monte Carlo methods to quantify the finite-sample size and power properties of $\mathcal{KPSS}$, $\mathcal{RS}$, and $\mathcal{KS}$ of (3) and Remark 3, for the DGP (1)–(2) with $\beta = 0$ and $(\varepsilon_t, \eta_t)' \sim NIID(\mathbf{0}, \mathbf{I}_2)$, where $\{\sigma_t^2\}$ and/or $\{\sigma_{\eta t}^2\}$ vary according to Assumption $\mathcal{V}$. We focus on the following three particular cases, where $f(s)$ can be either $\omega(s)$ or $\omega_\eta(s)$:

https://doi.org/10.1017/S0266466605050553 Published online by Cambridge University Press

- Case (a): Single Break: $f(s)^2 = f_0^2 + (f_1^2 - f_0^2)\mathbb{I}(s \geq m)$
- Case (b): Smooth Transition: $f(s)^2 = f_0^2 + (f_1^2 - f_0^2)\mathbb{S}(s)$, $\mathbb{S}(s) = (1 + \exp(-\gamma(s - m)))^{-1}$
- Case (c): Piecewise-Linear Trend: $f(s)^2 = f_0^2 + (f_1^2 - f_0^2)(s - m) \times (1 - m)^{-1}\mathbb{I}(s \geq m)$.

Without loss of generality, in each case we set $f_0 = 1$ and vary the ratio $d = f_0/f_1$ among $d \in \{0.25,4\}$. A positive (negative) variance shift obtains for $d < 1$ ($d > 1$). In both Cases (a) and (b) we vary the parameter m among $m \in \{0.1,0.5,0.9\}$. In Case (b) we report results setting the speed of transition parameter $\gamma = 10$. Under Case (c) we consider $m \in \{0.0,0.5,0.9\}$. For $m = 0.0$ the variance process follows a linear trend between f_0^2 for $s = 0$ and f_1^2 for $s = 1$. When $m > 0$ the variance is fixed at f_0^2 up until time $[mT]$ after which time it follows a linear trend path until $s = 1$ where it equals f_1^2. Other parameter values were considered but add little to what is reported.[2]

We have set both $\{\varepsilon_t\}$ and $\{\eta_t\}$ to be serially uncorrelated Gaussian sequences as the effects we are looking to quantify are those caused by nonconstant variances rather than serial correlation. The latter are already well documented in the literature; (see, inter alia, Kwiatkowski et al., 1992, pp. 169–172). Accordingly, we use a Bartlett kernel with $q_T = 1$. Samples of sizes $T = 50$ and 250 are considered; all tests were run at the nominal 5% level using critical values, obtained in the same fashion, under $\sigma_t = 1$ and $\sigma_{\eta t} = 0$, $t = 1,\dots,T$.[3]

6.1. Size Properties

Table 1 reports empirical rejection frequencies of the $\mathcal{KPSS}$, $\mathcal{RS}$, and $\mathcal{KS}$ tests when $\sigma_{\eta t} = 0$, $t = 1,\dots,T$, and $\omega(s)^2$, $0 \leq s \leq 1$, satisfies either Case (a), (b), or (c) with $\sigma_j = f_j$, $j = 0,1$, for the range of parameter values outlined before. Results are reported for the cases where $\hat{u}_t$ are the OLS residuals from the regression of y_t on $x_t = 1$ (a constant) or $x_t = (1, t)'$ (a constant and linear trend), $t = 1,\dots,T$.

Consider first the results for the single break model. For early breaks ($m = 0.1$) the $\mathcal{KPSS}$ test is (over-) undersized when ($d = 4$) $d = 0.25$. For late breaks this pattern is reversed. For the constant case, $\mathcal{KS}$ displays the largest size distortions in most cases, whereas there seems to be little to choose between the $\mathcal{KPSS}$ and $\mathcal{RS}$ tests overall: $\mathcal{KPSS}$ is better behaved (with only slight oversizing) than $\mathcal{RS}$ for $m = 0.5$, but the reverse is true for both $m = 0.1$ and $m = 0.9$. Where significant size distortions occur in the $\mathcal{KS}$ and $\mathcal{RS}$ tests for the constant case, they worsen considerably for the linear trend case, especially so in the case of $\mathcal{KS}$. In the trend case the $\mathcal{KPSS}$ test is noticeably better behaved than the other tests, behaving similarly to the constant case. Finally, for $m = 0.5$ the degree of oversizing seen in each of the three tests does not vary significantly between $d = 4$ and $d = 0.25$.

The results for the smooth transition break model largely mirror those for the single break but with the distortions somewhat ameliorated. This result is

TABLE 1. Empirical size of stationarity tests: Heteroskedastic errors

			$\mathcal{KPSS}$						$\mathcal{KS}$						$\mathcal{RS}$					
			$m = m^*$		$m = 0.5$		$m = 0.9$		$m = m^*$		$m = 0.5$		$m = 0.9$		$m = m^*$		$m = 0.5$		$m = 0.9$	
x_t	T	d	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4
Case (a)																				
1	50		4.2	13.8	5.3	6.1	15.1	3.8	5.0	10.3	9.2	10.4	13.4	5.1	5.0	4.5	7.8	8.2	6.4	4.9
	250		4.1	14.6	5.7	5.7	14.9	4.1	5.1	13.8	10.4	10.3	14.3	5.1	5.1	9.2	9.1	9.1	9.9	5.1
$(1, t)'$	50		4.2	15.2	6.8	6.5	13.8	4.2	5.2	25.4	14.7	16.0	28.4	5.9	5.4	16.2	7.9	7.6	15.9	5.6
	250		4.3	13.4	6.8	6.6	13.5	4.1	5.6	30.9	16.4	16.5	31.5	5.6	5.5	17.4	9.0	8.8	17.6	5.5
Case (b), $\gamma = 10$																				
1	50		4.2	10.6	5.4	5.6	10.8	4.1	5.1	11.2	8.2	8.7	11.4	5.1	5.2	6.8	6.7	6.6	7.3	5.0
	250		4.2	10.7	5.7	5.6	10.9	4.2	5.2	11.9	9.0	8.8	12.2	5.1	5.2	8.2	7.5	7.2	8.7	5.1
$(1, t)'$	50		4.9	8.2	6.0	6.0	8.0	4.8	5.6	18.0	11.7	12.1	18.5	5.7	5.5	8.9	7.4	7.4	8.8	5.5
	250		4.9	7.8	6.1	5.8	8.0	4.7	5.8	19.7	13.3	12.7	20.3	5.7	5.6	9.5	8.0	7.9	9.5	5.6
Case (c)																				
1	50		5.2	5.3	9.9	3.8	12.3	4.2	6.6	7.0	12.3	5.6	9.0	4.9	5.4	5.6	8.2	5.3	4.8	5.0
	250		5.6	5.2	9.8	3.8	12.4	4.3	7.3	6.8	13.0	5.9	10.3	4.9	5.9	5.5	9.5	5.5	7.1	4.8
$(1, t)'$	50		5.5	5.5	7.0	5.2	14.0	4.4	8.6	9.0	19.3	7.3	22.7	5.3	6.4	6.7	7.6	6.3	14.7	5.2
	250		5.3	5.5	6.8	5.3	13.3	4.3	8.6	9.3	20.7	7.3	24.0	4.8	6.4	6.9	8.9	6.5	15.5	4.9

Note: (i) In the column headed x_t, 1 and $(1, t)'$ indicate that $\hat{u}_t$, $t = 1, \ldots, T$, are the OLS residuals from the regression of y_t on a constant, and a constant and linear time trend respectively; (ii) for Cases (a) and (b), $m^* = 0.1$, whereas for Case (c), $m^* = 0.0$.

perhaps not surprising given that the logistic function used in Case (b) smooths the break across the sample. Although we report results for a relatively slow transition speed, $\gamma = 10$, we computed experiments for a range of values of γ and found the differences across γ quite small with results tending toward those for the single break model as γ increased. For example, by $\gamma = 50$ these results were indistinguishable.

Turning to the results for trending variances, for $m = 0$ the size of the $\mathcal{KPSS}$ test is not substantially affected in either the constant or constant and trend cases, whereas the size distortions seen in the constant and trend cases for the $\mathcal{KS}$ and $\mathcal{RS}$ tests are roughly the same throughout for $d = 4$ and $d = 0.25$. Again the $\mathcal{KS}$ test displays the worst size distortions. For all of the tests linear trending variances seem in most cases to have a lesser impact on size than either fixed or smooth transition breaks. The patterns of size distortions for the piecewise-linear trend ($m = 0.5$ and $m = 0.9$) exaggerate (dampen) those seen in the same setting when $m = 0$ and $d = 0.25$ ($d = 4$).

6.2. Local Power Properties

Table 3 reports empirical rejection frequencies of the $\mathcal{KPSS}$, $\mathcal{RS}$, and $\mathcal{KS}$ tests under a local alternative for each of Cases (a), (b) and (c). For each case, results are reported where either only $\{\sigma_t^2\}$ (labeled "shift in $I(0)$ only") or only $\{\sigma_{\eta t}^2\}$ (labeled "shift in $I(1)$ only") vary through time and for the case where both vary. The range of values for the parameters is as in Section 6.1, excepting the case where both components vary through time where $\{\sigma_{\eta t}^2\}$ is fixed throughout with $d = 4$ and $m = 0.1$ under Case (a), $d = 4$, $m = 0.1$, and $\gamma = 10$ under Case (b), and $m = 0$ and $d = 4$ under Case (c). In these cases, therefore, $\{\sigma_t^2\}$ and $\{\sigma_{\eta t}^2\}$ evolve according to the same function with the same parameters, whereas for the other entries in the table they evolve according to the same function but with different parameters. The local alternative considered is (5), except that we do not scale out the nuisance parameter $\bar{\omega}/\bar{\omega}_\eta$.[4] Results are reported for the linear trend case with $c = 10$. Results for the constant only case and for other values of c were qualitatively similar. Consequently, the results for the shift in $I(0)$ only pertain to the local alternative $H_c\colon \sigma_{\eta t}^2 = \sigma_\eta^2 = (10/T)^2$, $t = 1,\dots,T$, whereas all other results relate to $H_c\colon \sigma_{\eta t}^2 = (10/T)^2 \times \omega_\eta(t/T)^2$, $t = 1,\dots,T$, where $\omega_\eta(.)^2$ is as defined previously for each of Cases (a), (b) and (c).

Consider first Table 2, which reports benchmark results for the power of the $\mathcal{KPSS}$, $\mathcal{RS}$, and $\mathcal{KS}$ tests for the homoskedastic case, $\sigma_t^2 = 1$, $t = 1,\dots,T$, under the local alternative $H_c\colon \sigma_{\eta t}^2 = c^2/T^2$, $t = 1,\dots,T$, for $c = 1$, 5, 10, 15, 20, and 25. Observe that the $\mathcal{KS}$ test is dominated on local power by both the $\mathcal{KPSS}$ and $\mathcal{RS}$ tests. The $\mathcal{KPSS}$ test is the locally best invariant (LBI) test in this setting, so it is no surprise that it displays the highest power in most cases. However, the $\mathcal{RS}$ test is very competitive on power and, indeed, tends to display higher power than $\mathcal{KPSS}$ for $c \geq 20$.

TABLE 2. Empirical local power of stationarity tests: Homoskedastic errors

		$\mathcal{KPSS}$						$\mathcal{KS}$						$\mathcal{RS}$					
T	c	1	5	10	15	20	25	1	5	10	15	20	25	1	5	10	15	20	25
50		5.1	12.7	32.7	51.4	65.5	74.2	5.4	11.7	30.1	49.0	63.8	73.2	5.2	12.3	31.6	50.6	65.0	74.2
250		5.1	13.5	35.7	57.6	73.2	82.6	5.1	12.2	33.6	55.9	72.5	82.4	5.2	12.9	34.9	57.5	73.8	83.6

Turning to the results for the heteroskedastic cases in Table 3, a number of regularities are seen. First, in each of the cases of variance shifts in the $I(0)$, $I(1)$, and both $I(0)$ and $I(1)$ components, the $\mathcal{RS}$ and $\mathcal{KPSS}$ tests behave almost identically. Second, in the case of variance shifts in the $I(1)$ component only, all three tests behave almost identically. Third, in the case where variance shifts affect both the $I(0)$ and $I(1)$ components, for the entries in Cases (a) and (b) for $m = 0.1$, $d = 4$ and Case (c) for $m = 0.0$, $d = 4$ (i.e., instances where precisely the same variance process applies to both the $I(0)$ and $I(1)$ components) the results are very similar to those seen in Table 2 for $c = 10$. Fourth, and as predicted by the asymptotic distribution theory (cf. Remark 7), changing variances in the $I(0)$ and $I(1)$ components (but not both) effect very different outcomes: negative (positive) shifts in the variance of the $I(1)$ component result in increases (decreases) in power relative to the benchmark homoskedastic power in Table 2, whereas the converse is true for variance shifts in the $I(0)$ component. Fifth, and in contrast to the preceding point, shifts in both the $I(0)$ and $I(1)$ variances tend not to inflate power beyond the homoskedastic benchmark; indeed, for single and smooth transition breaks with early positive shifts the empirical rejection frequencies of all the tests are close to the nominal level. Finally, the effects on power (relative to the homoskedastic case) of heteroskedastic variances are most pronounced for the single break case and least pronounced in the trend case (cf. Table 1).

7. BOOTSTRAP PROCEDURES

In this section we show that the size biases caused by time-varying second moments can be corrected by properly adapting the heteroskedastic fixed regressor bootstrap of Hansen (2000) to the present framework. Interestingly, the heteroskedastic bootstrap allows us to retrieve asymptotically correct p-values even in the presence of autocorrelated errors. The rationale behind this result is that whereas the asymptotic null distribution of the $\mathcal{KPSS}$ statistic is affected by the heteroskedasticity function $\omega(\cdot)$ it is not affected by the short memory properties of the $I(0)$ component $\{\varepsilon_t\}$ (see Theorem 1). We outline the bootstrap procedure for the $\mathcal{KPSS}$-based procedure, although the $\mathcal{KS}$- and $\mathcal{RS}$-based procedures may be bootstrapped in an entirely analogous fashion.

Let $\mathcal{D}_0$ and $G(\cdot)$ denote the limiting null distribution of $\mathcal{KPSS}$ (Theorem 1) and its cumulative distribution function (c.d.f.), respectively. Let $\{\hat{u}_t\}$ denote the residuals obtained by regressing y_t on x_t and let $\{z_t\}_{t=1}^{T}$ denote an independent $N(0,1)$ sequence. The bootstrap sample is defined as $y_t^b := u_t^b := \hat{u}_t z_t$, $t = 1,\ldots,T$, and the bootstrap statistic is given by $\mathcal{KPSS}^b := s_b^{-2} T^{-2} \sum_{t=1}^{T} (\hat{S}_t^b)^2$ with $\hat{S}_t^b := \sum_{i=1}^{t} \hat{u}_i^b$, $s_b^2 := T^{-1} \sum_{t=1}^{T} (\hat{u}_t^b)^2$, $\{\hat{u}_t^b\}$ denoting the residuals obtained from the regression of y_t^b on x_t, $t = 1,\ldots,T$. The bootstrap p-value is $p_T^b := 1 - G_T^b(\mathcal{KPSS})$, where $G_T^b(\cdot)$ denotes the c.d.f. of $\mathcal{KPSS}^b$.

The usefulness of the heteroskedastic bootstrap in the present framework is given in Theorem 4, which shows (i) that the bootstrap allows us to retrieve the

TABLE 3. Empirical local power of stationarity tests under heteroskedastic errors: $x_t = (1, t)'$

		$\mathcal{KPSS}$						$\mathcal{KS}$						$\mathcal{RS}$					
		$m = m^*$		$m = 0.5$		$m = 0.9$		$m = m^*$		$m = 0.5$		$m = 0.9$		$m = m^*$		$m = 0.5$		$m = 0.9$	
T	d	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4	0.25	4
Case (a): Shift in $I(0)$ only																			
50		6.0	78.9	10.2	48.2	23.6	34.7	7.1	76.4	18.0	49.8	32.3	33.8	7.1	77.2	10.3	46.9	23.9	35.3
250		6.1	83.8	10.1	52.8	24.3	37.3	7.3	83.2	19.9	56.2	35.4	36.9	7.4	83.6	11.5	52.4	26.1	38.5
Case (a): Shift in $I(1)$ only																			
50		87.5	7.0	79.1	20.3	45.4	31.5	86.9	6.9	80.6	19.1	44.8	29.4	87.7	7.0	80.3	19.7	45.1	30.7
250		94.6	7.7	86.8	22.6	47.3	34.6	95.0	7.2	87.9	20.9	45.9	32.5	95.4	7.3	88.1	21.8	47.0	33.8
Case (a): Shift in both $I(0)$ and $I(1)$																			
50		4.3	29.5	7.1	10.5	14.7	6.6	5.6	34.3	15.0	19.9	28.7	7.9	5.3	28.6	7.9	10.6	16.6	7.9
250		4.3	29.2	6.8	11.3	14.1	7.1	5.6	40.7	16.7	21.9	31.8	8.2	5.4	31.4	9.2	12.8	18.1	8.2
Case (b): Shift in $I(0)$ only																			
50		7.1	72.9	9.5	48.1	17.6	35.3	7.6	70.8	14.6	47.8	24.4	33.3	7.5	71.7	10.2	46.9	17.3	34.6
250		6.8	79.9	9.4	52.3	17.9	38.5	7.5	79.6	16.2	54.1	26.2	37.3	7.4	79.8	10.6	52.3	18.5	38.6

Case (b): Shift in $I(1)$ only																		
50	86.7	9.4	79.2	20.8	55.2	29.9	85.9	8.9	78.9	18.8	54.3	27.4	87.0	9.1	79.6	20.0	54.7	28.9
250	94.2	9.9	87.2	22.4	59.6	32.7	94.5	9.3	87.7	20.7	59.1	30.4	94.9	9.6	88.2	21.7	60.0	32.1
Case (b): Shift in both $I(0)$ and $I(1)$																		
50	5.4	29.9	6.6	14.6	9.3	9.9	6.0	36.8	11.9	20.8	18.9	10.1	5.8	29.4	7.7	14.8	10.0	10.2
250	5.0	32.1	6.3	15.5	9.2	10.8	5.7	40.6	13.0	23.0	20.5	11.3	5.6	32.5	7.8	16.6	10.7	11.3
Case (c): Shift in $I(0)$ only																		
50	8.9	48.1	13.3	38.3	28.6	33.4	11.5	46.7	23.8	37.4	30.3	31.9	9.6	47.0	13.3	38.2	26.9	33.3
250	8.9	52.3	13.2	42.5	29.7	36.2	11.9	52.2	25.9	42.1	32.4	34.9	9.7	52.6	14.4	43.0	29.0	36.4
Case (c): Shift in $I(1)$ only																		
50	79.6	20.5	66.3	26.8	36.1	32.4	79.3	18.9	66.7	24.8	33.2	30.0	79.9	19.7	66.3	25.8	34.4	31.4
250	87.7	22.8	73.1	29.9	39.6	35.2	87.8	20.8	73.8	27.6	37.1	33.1	88.5	21.9	74.2	29.0	38.6	34.4
Case (c): Shift in both $I(0)$ and $I(1)$																		
50	7.3	31.9	10.1	25.3	21.6	20.6	10.1	33.0	20.4	25.7	25.4	19.6	8.1	31.8	10.4	25.6	21.1	20.8
250	7.6	35.6	10.2	27.6	22.1	23.1	10.7	37.8	22.9	28.5	27.5	22.4	8.7	36.2	11.7	28.6	22.5	23.5

Note: (i) In the first block of results $\sigma_{\eta t} = 10/T$, $t = 1, \ldots, T$, whereas in the second and third blocks $\sigma_{\eta t} = (10/T)\omega_\eta(t/T)$, $t = 1, \ldots, T$; (ii) for Cases (a) and (b), $m^* = 0.1$, whereas for Case (c), $m^* = 0.0$.

correct asymptotic null distribution and hence that the p-values based on $G_T^b(\mathcal{KPSS})$ are asymptotically pivotal and (ii) that a test based on the bootstrap p-values is consistent.

THEOREM 4. *(i) Under the conditions of Theorem 1, $\mathcal{KPSS}^b \xrightarrow{w}_p \mathcal{D}_0$ and $p_T^b \xrightarrow{w} U[0,1]$, where $\xrightarrow{w}_p$ denotes weak convergence in probability (see Giné and Zinn, 1990). (ii) Under the conditions of Theorem 2, $p_T^b \xrightarrow{p} 0$.*

In practice, $G_T^b(\cdot)$ is not known but can be approximated in the usual way through numerical simulation by generating N (conditionally) independent bootstrap statistics, $\mathcal{KPSS}_n^b$, $n = 1,\ldots,N$, computed as before but from $y_{n,t}^b := u_{n,t}^b := \hat{u}_t z_{n,t}$, $t = 1,\ldots,T$, with $\{\{z_{n,t}\}_{t=1}^T\}_{n=1}^N$ a doubly independent $N(0,1)$ sequence. The simulated bootstrap p-value is then computed as $\tilde{p}_T^b := N^{-1}\sum_{n=1}^N \mathbb{I}(\mathcal{KPSS}_n^b \geq \mathcal{KPSS})$ and is such that $\tilde{p}_T^b \xrightarrow{a.s.} p_T^b$ as $N \to \infty$.

In Table 4 we report results for the bootstrapped KPSS testing procedure, outlined before, applied to data generated according to Case (a) of Section 6.1. The results are therefore directly comparable with those given for Case (a) in Table 1 for the $\mathcal{KPSS}$ test. Results are reported only for this case because this was the form of heteroskedasticity that effected the most significant size distortions in the original tests. The reported results are for experiments run over $N = 1{,}000$ bootstrap replications. Benchmark entries for the case where the errors are homoskedastic are also reported in the column labeled "IID."

A comparison of the results in Tables 1 and 4 shows that the bootstrap performs very well in practice with empirical sizes much closer to the nominal level than for the standard $\mathcal{KPSS}$ test. Some oversizing, associated with early negative and late positive variance breaks, is still seen for $T = 50$ but is much reduced relative to that seen for the standard $\mathcal{KPSS}$ test and is largely eliminated for $T = 250$. The undersizing seen in the standard $\mathcal{KPSS}$ test for early

TABLE 4. Empirical size of bootstrap KPSS tests: Heteroskedastic errors, Case (a)

			$m = 0.1$		$m = 0.5$		$m = 0.9$		
x_t	T	d	0.25	4	0.25	4	0.25	4	IID
1	50		5.2	7.8	5.4	5.7	8.0	5.1	5.2
	250		5.1	6.1	5.0	5.4	6.0	5.1	5.0
$(1, t)'$	50		5.3	9.7	5.7	6.0	8.8	5.5	5.5
	250		5.1	6.1	5.1	5.3	5.8	5.2	5.0

Note: (i) In the column headed x_t, 1 and $(1, t)'$ indicate that $\hat{u}_t$, $t = 1,\ldots,T$, are the OLS residuals from the regression of y_t on a constant, and a constant and linear time trend respectively; (ii) entries in the column headed IID relate to the case of homoskedastic errors.

positive and late negative breaks is eliminated by the bootstrap. Although not reported here, qualitatively similar improvements (available on request) were seen for bootstrapped implementations of the $\mathcal{KS}$ and $\mathcal{RS}$ tests and for data generated under Cases (b) and (c).

8. CONCLUSIONS

In this paper we have analyzed the effects that time-varying second moments of a very general form have on the stationarity tests of Kwiatkowski et al. (1992), Lo (1991), and Xiao (2001). We have demonstrated that, in general, heteroskedasticity changes the limiting distributions of these stationarity test statistics under both the null and local alternatives and (for appropriately rescaled statistics) global alternatives. We have presented Monte Carlo simulation results to quantify the finite-sample effects of heteroskedasticity on the size and power properties of the three tests. Results were presented for variances displaying either a single break, a smooth transition break, or a linear/piecewise-linear trend. Bootstrap versions of the tests, adapted from the heteroskedastic bootstrap principle of Hansen (2000), were developed and shown to greatly improve the finite-sample size properties of the tests. Although not considered here, it would be interesting and reasonably straightforward to extend the results presented in this paper to the corresponding tests for the null hypothesis of cointegration of Shin (1994), inter alia.

NOTES

1. Busetti and Taylor (2003) consider the model discussed here under the constraint that $\sigma_t = \sigma_{\eta t}$. In our framework we do not require this constraint to hold.
2. Indeed, for Case (c) we also considered the generalized trend function, $f(s)^r = f_0^r + (f_1^r - f_0^r)(s-m)(1-m)^{-1}\mathbb{I}(s \geq m)$, for a range of values of r but found very little dependence on r.
3. All simulation experiments were conducted using the RNDN function of Gauss 3.1 over 40,000 Monte Carlo replications.
4. Recall that this was done in Theorem 3 purely to simplify the right member of (6).

REFERENCES

Busetti, F. & A.C. Harvey (2001) Testing for the presence of a random walk in series with structural breaks. *Journal of Time Series Analysis* 22, 127–150.

Busetti, F. & A.M.R. Taylor (2003) Variance shifts, structural breaks and stationarity tests. *Journal of Business & Economic Statistics* 21, 510–531.

Cavaliere, G. (2004a) Testing stationarity under a permanent variance shift. *Economics Letters* 82, 403–408.

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

Cavaliere, G. & A.M.R. Taylor (2004) Stationarity Tests under Time-Varying Second Moments. Discussion Papers in Economics 04-12, University of Birmingham.

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

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

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

Hamori, S. & A. Tokihisa (1997) Testing for a unit root in the presence of a variance shift. *Economics Letters* 57, 245–253.

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

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

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

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

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, & Y. Shin (1992) Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54, 159–178.

Lo, A.W. (1991) Long-term memory in stock market prices. *Econometrica* 59, 1279–1314.

Nyblom, J. (1989) Testing the constancy of parameters over time. *Journal of the American Statistical Association* 84, 223–230.

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

Phillips, P.C.B. & Y. Sun (2001) Nonorthogonal Hilbert projections in trend regression. *Econometric Theory*, Problem, 17, 854; Solution, 18, 1011–1015.

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

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

Shin, Y. (1994) A residual-based test of the null of cointegration against the alternative of no cointegration. *Econometric Theory* 10, 91–115.

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

Xiao, Z. (2001) Testing the null hypothesis of stationarity against an autoregressive unit root alternative. *Journal of Time Series Analysis* 22, 87–105.

APPENDIX

Proof of Theorem 1. Define the partial sum $S_{[sT]} := \sum_{t=1}^{[sT]} u_t$. Under Assumptions $\mathcal{V}$ and $\mathcal{E}$, $T^{-1/2}S_{[sT]} \xrightarrow{w} \lambda_\varepsilon \int_0^s \omega(r)\,dB(r) = \lambda_\varepsilon \bar{\omega} B_\omega(s)$ (see also Cavaliere, 2004b). Similarly, $T^{-1/2}\sum_{t=1}^{[sT]} u_t x_t' \delta_T \xrightarrow{w} \lambda_\varepsilon \bar{\omega} \int_0^s dB_\omega(r) F(r)'$. After some algebra, the preceding results taken together with the convergence result $T^{-1}\sum_{t=1}^{T} \delta_T x_t x_t' \delta_T \to \int_0^1 F(s)F(s)'\,ds$ (Assumption $\mathcal{X}$) allow us to conclude that $T^{-1/2}\hat{S}_{[sT]} \xrightarrow{w} \lambda_\varepsilon \bar{\omega} V_\omega(s)$ and, hence, by the continuous mapping theorem (CMT), that $T^{-2}\sum_{t=1}^{T} \hat{S}_t^2 \xrightarrow{w} \lambda_\varepsilon^2 \bar{\omega}^2 \int_0^1 V_\omega(s)^2\,ds$. Note that the CMT also allows us to prove that $T^{-1/2}\max_{t\le T}|\hat{S}_t| \xrightarrow{w} \lambda_\varepsilon \bar{\omega} \sup_{s\in[0,1]}|V_\omega(s)|$ and that $T^{-1/2}(\max_{t\le T}\hat{S}_t - \min_{t\le T}\hat{S}_t) \xrightarrow{w} \lambda_\varepsilon \bar{\omega}(\sup_{s\in[0,1]} V_\omega(s) - \inf_{s\in[0,1]} V_\omega(s)) = \lambda_\varepsilon \bar{\omega} \sup_{s,s'\in[0,1]}|V_\omega(s) - V_\omega(s')|$ (see Remarks 3 and 5). The proof is completed by showing that $\hat{\lambda}^2 \xrightarrow{p} \lambda_\varepsilon^2 \bar{\omega}^2$, which follows from Cavaliere (2004b, Thm. 4). ■

Proof of Theorem 2. Because $\{\sigma_{\eta t}\}$ satisfies Assumption $\mathcal{V}$ then $T^{-1/2}\mu_{[sT]} \xrightarrow{w} \bar{\omega}_\eta B_{\omega_\eta}(s)$, $B_{\omega_\eta}(s) := (1/\bar{\omega}_\eta)\int_0^s \omega_\eta(r)\,dB(r)$. This result also implies that $T^{-1/2} \times (u_{[sT]} + \mu_{[sT]}) \xrightarrow{w} \bar{\omega}_\eta B_{\omega_\eta}(s)$ as $\{u_t + \mu_t\}$ is dominated by $\{\mu_t\}$. As in the proof of Theorem 1, it easily follows that the residuals $\{\hat{u}_t\}$ obey the functional central limit theorem $T^{-1/2}\hat{u}_{[sT]} \xrightarrow{w} \bar{\omega}_\eta W_{\omega_\eta}(s)$ and, hence, by the CMT, the numerator of $\mathcal{KPSS}$ satisfies $T^{-4}\sum_{t=1}^T \hat{S}_t^2 \xrightarrow{w} \bar{\omega}_\eta^2 \int_0^1 [\int_0^s W_{\omega_\eta}(r)\,dr]^2\,ds$. Finally, as in Kwiatkowski et al. (1992) one can show that for any $j = o(T^{1/2})$, $T^{-1}\hat{\gamma}(j) = T^{-2}\sum_{t=|j|+1}^T \hat{u}_t\hat{u}_{t-|j|} \xrightarrow{w} \bar{\omega}_\eta^2 \int_0^1 W_{\omega_\eta}(s)^2\,ds$ and hence that $q_T^{-1}T^{-1}\hat{\lambda}^2 \xrightarrow{w} \bar{k}\bar{\omega}_\eta^2 \int_0^1 W_{\omega_\eta}(s)^2\,ds$, $\bar{k} := \int_{-\infty}^{\infty} k(x)\,dx$. ■

Proof of Theorem 3. The proof follows directly from Theorems 1 and 2, using the CMT and the fact that, because $\mu_t = O_p(T^{-1/2})$, $\hat{\lambda}^2 \xrightarrow{p} \lambda_\varepsilon^2 \bar{\omega}^2$, as in Theorem 1. ■

Proof of Theorem 4. (i) Conditionally on $\{\hat{u}_t\}_{t=1}^T$, $M_T^b(s) := T^{-1/2}\sum_{t=1}^{[sT]} u_t^b \delta_T x_t = T^{-1/2}\sum_{t=1}^{[sT]} z_t \hat{u}_t \delta_T x_t$ is Gaussian with covariance kernel $\Lambda_T^M(r,s) = \Lambda_T^M(\min\{r,s\}) := T^{-1}\sum_{t=1}^{[\min\{r,s\}T]} \hat{u}_t^2 \delta_T x_t x_t' \delta_T$ (see, e.g., Hansen, 1996). Similarly, the process $S_T^b(s) := T^{-1/2}\sum_{t=1}^{[sT]} y_t^b = T^{-1/2}\sum_{t=1}^{[sT]} u_t^b$ is Gaussian with covariance kernel $\Lambda_T(r,s) = \Lambda_T(\min\{r,s\}) := T^{-1}\sum_{t=1}^{[\min\{r,s\}T]} \hat{u}_t^2$. To simplify notation, but without loss of generality, assume that x_t contains a constant, i.e., $x_t := (1, \tilde{x}_t')'$; then $S_T^b(s) = (1,0')M_T^b(s)$, so that the asymptotic distribution of $S_T^b(\cdot)$ easily follows from that of $M_T^b(\cdot)$. Now, $\Lambda_T^M(s) \xrightarrow{w} \Lambda_\omega^M(s) := \int_0^s \omega(r)^2 F(r)F(r)'\,dr$, which is a consequence of the fact that $\delta_T^{-1}\hat{\beta}$ (notice that the true value of β is zero here) is of $O_p(T^{-1/2})$ and the mixing properties of $\varepsilon_t^2 - E(\varepsilon_t^2)$. It therefore follows that $M_T^b(s) \xrightarrow{w}_p \int_0^s \omega(r)F(r)\,dB(r) = \bar{\omega}\int_0^s F(r)\,dB_\omega(r)$ and, by the CMT, that $S_T^b(s) \xrightarrow{w}_p \bar{\omega}B_\omega(s)$. Hence, $T^{-1/2}\hat{S}_{[sT]}^b = S_T^b(s) - M_T^b(1)'(T^{-1}\sum_{t=1}^T \delta_T x_t x_t' \delta_T)^{-1} M_T^b(s)$ weakly converges to $\bar{\omega}(B_\omega(s) - Q_{\mathbb{F}}B_\omega(s)) = \bar{\omega}V_\omega(s)$. Notice that $s_b^2 := T^{-1}\sum_{t=1}^T (\hat{u}_t^b)^2 = T^{-1}\sum_{t=1}^T \hat{u}_t^2(z_t^2 - 1) + T^{-1}\sum_{t=1}^T \hat{u}_t^2$. As in Cavaliere (2004b) it is straightforward to show that $T^{-1}\sum_{t=1}^T \hat{u}_t^2 \xrightarrow{p} \bar{\omega}^2$, so that $s_b^2 \xrightarrow{p} \bar{\omega}^2$ if $T^{-1}\sum_{t=1}^T \hat{u}_t^2(z_t^2 - 1) \xrightarrow{p} 0$, a result that follows from a standard application of the weak law of large numbers for martingale difference sequences. The preceding results imply that $\mathcal{KPSS}^b \xrightarrow{w}_p \mathcal{D}_0$ and hence that $G_T^b(\cdot) \to G(\cdot)$ uniformly in probability. The remainder of the proof is identical to the proof of Theorem 5 in Hansen (2000). (ii) Let $\hat{S}_T^b(s) := T^{-1}\hat{S}_{[sT]}^b$. Conditionally on $\{\hat{u}_t\}$, $\hat{S}_T^b(\cdot)$ is Gaussian with zero mean and covariance kernel $\Lambda_T^S(r,s) = \Lambda_T^S(\min\{r,s\}) := T^{-2}\sum_{t=1}^{[\min\{r,s\}T]} \hat{u}_t^2 \xrightarrow{w} \Lambda^S(\min\{r,s\})$, $\Lambda^S(s) := \bar{\omega}_\eta^2 \int_0^s W_{\omega_\eta}(r)^2\,dr$. Hence, $\hat{S}_T^b(s) \xrightarrow{w}_p \bar{\omega}_\eta^2 \int_0^s W_{\omega_\eta}(r)\,dB_{z_1}(r)$ and by the CMT $T^{-3}\sum_{t=1}^T (\hat{S}_t^b)^2 \xrightarrow{w}_p \bar{\omega}_\eta^2 \int_0^1 (\int_0^s W_{\omega_\eta}(r)\,dB_{z_1}(r))^2\,ds$, $B_{z_1}(\cdot)$ being a standard Brownian motion, independent of $W_{\omega_\eta}(r)$. Using similar arguments it can be shown that $T^{-2}\sum_{t=1}^T \hat{u}_t^2 \xrightarrow{w} \bar{\omega}_\eta^2 \int_0^1 W_{\omega_\eta}(s)^2\,ds$ and that $T^{-3/2}\sum_{t=1}^T \hat{u}_t^2(z_t^2 - 1) \xrightarrow{w}_p \bar{\omega}_\eta^2 \int_0^1 W_{\omega_\eta}(s)^2\,dB_{z_2}(s)$, $B_{z_2}(\cdot)$ being a standard Brownian motion, independent of $W_{\omega_\eta}(\cdot)$ and $B_{z_1}(\cdot)$. Consequently, the standardized variance estimator $T^{-1}s_b^2 = T^{-2}\sum_{t=1}^T \hat{u}_t^2 z_t^2 = T^{-2}\sum_{t=1}^T \hat{u}_t^2 + T^{-2}\sum_{t=1}^T \hat{u}_t^2(z_t^2 - 1)$ satisfies $T^{-1}s_b^2 = T^{-2}\sum_{t=1}^T \hat{u}_t^2 + O_p(T^{-1/2}) \xrightarrow{w}_p \bar{\omega}_\eta^2 \int_0^1 W_{\omega_\eta}(s)^2\,ds$. Taken together, the preceding results imply that $\mathcal{KPSS}^b$ weakly converges in probability to the random variable $(\int_0^1 W_{\omega_\eta}(s)^2\,ds)^{-1} \int_0^1 (\int_0^s W_{\omega_\eta}(r)\,dB_{z_1}(r))^2\,ds$, whose c.d.f. is denoted by $\tilde{G}^b(\cdot)$, and hence that $G_T^b(\cdot) \xrightarrow{p} \tilde{G}^b(\cdot)$, uniformly. Consequently, $p_T^b = 1 - \tilde{G}^b(\mathcal{KPSS}) + o_p(1)$. Because, by Theorem 2, $\mathcal{KPSS}$ diverges at the rate $q_T^{-1}T$, it follows that $p_T^b \xrightarrow{p} 0$. ■

https://doi.org/10.1017/S0266466605050553 Published online by Cambridge University Press