A Test for Strict Stationarity

Luiz Renato Lima

and Department of Economics, University of Illinois at Urbana-Champaing*

and Graduate School of Economics, Getúlio Vargas Foundation†

http://www.econ.uiuc.edu/~oliverlm

Luiz.Renato@fgv.br

Breno Neri

Department of Economics, New York University‡

http://homepages.nyu.edu/~bpn207

breno.neri@nyu.edu

January 4, 2010

Abstract

We introduce a test for strict stationarity based on the fluctuations of the quantiles of the data, and we show that this test has power against the alternative hypothesis of unconditional heteroskedasticity while other tests for first order (level) stationarity as the KPSS test (Kwiatkowski et al., 1992) and, its robust version, the IKPSS test (de Jong et al., 2007) have low power against this alternative of time-varying variance. Moreover, our test has power against the alternative hypothesis of time-varying kurtosis, while the test for second order (covariance) stationarity introduced by Xiao and Lima (2007) has power close to size against this alternative.

keywords: strict stationarity testing, time-varying volatility, time-verying kurtosis.

JEL Classification: C12, C22.

1 Introduction

Several techniques employed in time-series econometrics rely on stationarity. So, the development of tests for stationarity is an active field of research.

*461 Wohlers Hall 461 S Sixth Street Champaing, IL 61820

†Praia de Botafogo 190, 11^{th} Floor Rio de Janeiro, RJ Brazil 22250-900

‡19 West 4^{th} Street, 6^{th} Floor New York, NY USA 10012

In 1992, Kwiatkowski, Phillips, Schmidt and Shim (KPSS) proposed a test for for first order (level) stationarity based on the following standardized empirical process:

$$S_T(r) := \frac{1}{\widehat{\omega}\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} (y_t - \overline{y}_T),$$

where $r \in [0, 1]$, $\overline{y}_T$ is the sample mean of $\{y_t\}_{t=1}^T$ and $\widehat{\omega}^2$ is a nonparametric consistent estimator of the long-run variance

$$\omega^2 = \lim_{T\to\infty} \mathbb{E}\left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (y_t - \overline{y}_T)\right)^2\right].$$

In order to measure the fluctuation of $S_T(r)$, they consider the functional $h\left(S_T(r)\right)$, where $h(\cdot)$ is the Cramér-von Mises metric. The KPSS test statistic is then given by

$$KPSS = \frac{1}{(\widehat{\omega}T)^2} \sum_{k=1}^{T} \left(\sum_{t=1}^{k} (y_t - \overline{y}_T)\right)^2,$$

and, under the null hypothesis of level stationarity,

$$KPSS \stackrel{d}{\Longrightarrow} \int_0^1 \kappa(\alpha)^2 d\alpha,$$

where $\kappa(\alpha) := W(\alpha) - \alpha W(1)$ is the standard Brownian bridge. The critical values can be found in KPSS (1992).

In a recent paper, de Jong et al. (2007) proposed a robust version of the KPSS test based on the following empirical process:

$$I_T(r) := \frac{1}{\widehat{\sigma}\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \operatorname{sign}\left(y_t - m_T\right),$$

where m_T is the sample median of $\{y_t\}_{t=1}^T$, $\widehat{\sigma}^2$ is a nonparametric consistent estimator of the long-run variance

$$\sigma^2 = \lim_{T\to\infty} \mathbb{E}\left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \operatorname{sign}\left(y_t - m_T\right)\right)^2\right],$$

and

$$\operatorname{sign}(x) := \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

And they applied the Cramér-von Mises metric to measure the fluctuation of the empirical process $I_T(r)$. This gives rise to the IKPSS test statistic

$$IKPSS = \frac{1}{(\widehat{\sigma}T)^2} \sum_{k=1}^{T} \left(\sum_{t=1}^{k} \operatorname{sign}(y_t - m_T) \right)^2 .$$

Under the null hypothesis of level stationarity, de Jong et al. (2007) show that $IKPSS \overset{d}{\Longrightarrow} \int_0^1 \kappa(\alpha)^2 d\alpha$, the same limiting distribution as the KPSS test statistic. Unlike the KPSS test, the IKPSS has correct size under the presence of fat-tailed errors. When the alternative hypothesis is unit root, the indicator test has lower power than the KPSS when tails are thin, but higher power when tails are fat.

However, when the aforementioned traditional stationarity tests are applied to test stationarity, it is difficult to detect alternatives with unconditional volatility (distribution scale) that changes over time.

In the same year, 2007, Xiao and Lima proposed a test for second order (covariance) stationarity based on the following standardized bivariate empirical process:

$$Z_T(r) := \frac{1}{\sqrt{T}} \widehat{\Omega}^{-\frac{1}{2}} \sum_{t=1}^{\lfloor Tr \rfloor} \begin{pmatrix} \widetilde{y}_t \\ v_t \end{pmatrix},$$

where $\widetilde{y}_t := y_t - \frac{1}{T} \sum_{j-1}^{T} y_j$ is the demeaned data, $v_t := \widetilde{y}_t^2 - \sigma_y^2$, $\sigma_y^2 := \frac{1}{T} \sum_{t=1}^{T} \widetilde{y}_t^2$ and $\widehat{\Omega}^{-\frac{1}{2}}$ is the inverse of the Choleski decomposition of $\widehat{\Omega}^2$, a nonparametric consistent estimator of the long-run variance

$$\Omega = \lim_{T \to \infty} \mathbb{E} \left[\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \widetilde{y}_t \\ v_t \end{pmatrix} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \widetilde{y}_t \\ v_t \end{pmatrix} \right)' \right].$$

Then, they applied the Kolmogorov metric to measure the fluctuation of the empirical process $Z_T(r)$. Their test statistic is then

$$XL = \max_{1 \le k \le T} \left\| \frac{1}{\sqrt{T}} \widehat{\Omega}^{-\frac{1}{2}} \sum_{t=1}^{k} \begin{pmatrix} \widetilde{y}_t \\ v_t \end{pmatrix} \right\|_1 .$$

Under the null hypothesis of covariance stationarity,

$$XL \overset{d}{\Longrightarrow} \sup_{0 \le r \le 1} \left\| \begin{pmatrix} W_1(r) - rW_1(1) \\ W_2(r) - rW_2(1) \end{pmatrix} \right\|_1 ,$$

where $\begin{pmatrix} W_1(r) - rW_1(1) & W_2(r) - rW_2(1) \end{pmatrix}'$ is the 2-dimensional standardized Brownian bridge. The critical values can be found in Xiao and Lima (2007).

Unlike the KPSS or the IKPSS, the XL test has power not only against the alternative hypothesis of distribution location varying on time but also against the alternative hypothesis of distribution scale (unconditional volatility) varying on time. However, all of the aforementioned tests have power close to size against the alternative hypothesis of time-varying kurtosis.

As Busetti and Harvey (2007) discuss, the distribution of a random variable may presents changes over time that does not impact the level or the variance. For instance, maybe the asymmetry or fatness of the tail is time-varying. This is particularly important in analyzing financial time-series. To exemplify this point, consider how changes in lower tail quantiles may impact decisions of a risk manager or a regulatory agency.[1]

In this paper, we propose a new test for the null hypothesis of strict stationarity as a useful complement to the previous procedures. This new test uses the sign of the data minus the sample quantiles. In this way, this new test can be seen as a generalization of the IKPSS test, since the latter uses the sign of the data minus the sample median only. Comparing to the KPSS, IKPSS and XL tests, the proposed test has power not only against unit root alternative, alternatives with structural changes in the mean and alternatives with unconditional heteroskedasticity, but also has good power in detecting changes in higher moments of the unconditional distribution.

This paper is organized as follows: Section 2 describes our testing procedure; Section 3 brings the Monte Carlo; an empirical exercise is done in Section 4; and Section 5 concludes.

2 A Test for Strict Stationarity

Let $\{y_t\}_{t=1}^T$ be the data and, for $\tau \in [0, 1]$, define

$$b(\tau) := \arg\max_{b \in \mathbb{R}} \sum_{t=1}^{T} \rho_\tau \left(y_t - b \right),$$

where

$$\rho_\tau(u) = (1_{u<0} - \tau)u.$$

Therefore, $b(\tau)$ is simply the τ^{th} sample unconditional quantile of $\{y_t\}_{t=1}^T$.

Notice that ρ_τ is not everywhere differentiable but, since it is convex, we can still compute the subgradient. The subgradient plays the same role in quantile estimation as the score function in maximum likelihood estimation. The subgradient of ρ_τ is given by[2]

$$\psi_\tau(u) = 1_{u<0} - \tau.$$

We now define the empirical process

$$S_T(r, \tau) := \frac{1}{\widehat{\pi}(\tau)\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \psi_\tau \left(y_t - b(\tau) \right),$$

[1] Value-at-Risk (VaR), a measure of risk based on a lower tail quantile, is of considerable importance in financial regulation (Lima and Neri, 2007).

[2] In fact, the subgradient of ρ_τ at zero is not unique; it can be any element of the closed interval $[-\tau, 1 - \tau]$.

where $r \in [0,1]$ and $\widehat{\pi}(\tau)^2$ is a nonparametric consistent estimator of

$$\pi(\tau)^2 := \lim_{T\to\infty} \mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_\tau\left(y_t - b_0(\tau)\right)\right)^2\right],$$

where $b_0(\tau)$ is the population τ^{th} unconditional quantile of the $\{y_t\}_{t=1}^{T}$.

This paper proposes to test for strict stationarity by using the Kolmogorv-Smirnoff metric to measure the fluctuation of $S_T(r,\tau)$ across various quantiles $\tau \in \Gamma_w = [w, 1-w]$, for some $w \in \left(0, \frac{1}{2}\right)$, which gives rise to the following test statistic:

$$SS = \max_{\tau\in\Gamma_w}\max_{1\le k\le T}\frac{1}{\widehat{\pi}(\tau)\sqrt{T}}\left|\sum_{t=1}^{k}\psi_\tau\left(y_t - b(\tau)\right) - \frac{k}{T}\sum_{t=1}^{T}\psi_\tau\left(y_t - b(\tau)\right)\right|.$$

$\widehat{\pi}(\tau)^2$ can be computed as the HAC estimator,

$$\widehat{\pi}(\tau)^2 := \frac{1}{T}\sum_{i=1}^{T}\sum_{j=1}^{T}K\left(\frac{i-j}{q_T}\right)\psi_\tau\left(y_i - b(\tau)\right)\psi_\tau\left(y_j - b(\tau)\right),$$

where K is a kernel function.

Assumption 1. *The kernel function K satisfies:*

1. $\int_{-\infty}^{\infty}|\omega(\xi)|\,d\xi < \infty$*, where*

$$\omega(\xi) := \frac{1}{2\pi}\int_{-\infty}^{\infty}K(x)e^{-ix\xi}dx;$$

2. *K is continuous at all but a finite number of points, $K(x) = K(-x)$, $|K(x)| \le l(x)$, where $l(x)$ is non-increasing and $\int_0^\infty |l(x)|\,dx < \infty$, and $K(0) = 1$;*

3. $\lim_{T\uparrow\infty} q_T = \infty$ *and* $\lim_{T\uparrow\infty}\frac{q_T}{T} = 0$.

Assumption 1 is equal to Assumption 2 used in de Jong et al. (2007) and although it excludes the use of the uniform kernel function, it allows choices such as the Bartlett, Quadratic Spectral, and Parzen kernels.

Assumption 2. *(Null Hypothesis H_0)*

1. *$\{y_t\}_{t=1}^{\infty}$ is a strictly stationary stochastic process and $b_0(\tau)$ is the unique population τ^{th} unconditional quantile of y_t;*

2. *$\{y_t\}_{t=1}^{\infty}$ is strong $(\alpha-)$ mixing and, for some finite $\kappa > 2$, $C > 0$ and $\eta > 0$, $\alpha(m) \le Cm^{-\frac{\kappa}{\kappa-2}-\eta}$;*

3. *$y_t - b_0(\tau)$ have a continuous density f in a neighborhood $[-\eta,\eta]$ of 0 for some $\eta > 0$, and $\inf_{y\in[-\eta,\eta]} f(y) > 0$;*

4. $\sigma^2 \in (0, \infty)$.

Theorem 1. *Under Assumption 1 and Assumption 2,*

$$SS \overset{d}{\Longrightarrow} \sup_{\tau \in \Gamma_w} \sup_{0 \leq r \leq 1} |B(r, \tau)|,$$

where $B(r, \tau)$ is the Brownian Pillow (or the tucked Brownian Sheet).

A proof for Theorem 1 for the case in which the innovations are i.i.d. was done by Qu (2005). The critical values of our test are computed through the simulation of 10^5 time series with 1000 observations $e_t \sim i.i.d.U[0,1]$.[3] $B(r, \tau)$ is then approximated by

$$\frac{1}{\widehat{\gamma}(\tau)\sqrt{T}} \left(\sum_{t=1}^{k} 1_{e_t \leq \tau} - \frac{k}{T} \sum_{t=1}^{T} 1_{e_t \leq \tau} \right),$$

where $k = \lfloor Tr \rfloor$ and $\widehat{\gamma}(\tau)^2$ is the sample variance (over k) of $\sum_{t=1}^{k} 1_{e_t \leq \tau} - \frac{k}{T} \sum_{t=1}^{T} 1_{e_t \leq \tau}$. The supremum of the absolute value of this approximating process is obtained by maximizing over k and τ. We considered $\tau \in [0.10, 0.90]$ with increments of 0.01. Figure 1 displays (1) the histogram of the 10^5 realizations of the maximum of the absolute value of the approximating process, (2) the probability density estimated nonparametrically using a Gaussian kernel and bandwidth given by Silverman's rule-of-thumb (Silverman, 1986), and (3) the quantiles 90%, 95% and 99%. So, the critical values for the significance levels of 10%, 5% and 1% are 1.65, 1.77 and 2.01, respectively.

3 Monte Carlo Experiment

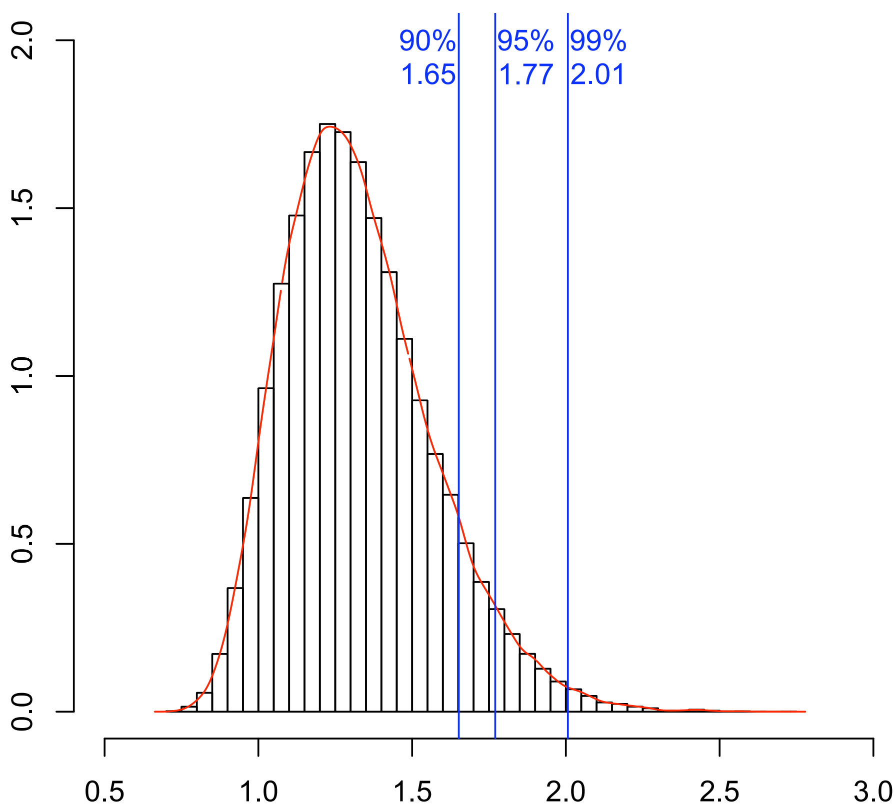
In this section we report the results of our Monte Carlo experiment that investigate the size and power of the KPSS, IKPSS, XL and our test for strict stationarity (SS). Our experiment is coded in R and it is run in one of the Linux HPCCs (High Performance Computation Clusters) at New York University (NYU). We follow de Jong et al. (2007) and vary tail thickness by considering t distributions with different degrees of freedom. In particular, we consider t_∞ (normal), t_5, t_3, t_2, and t_1 (Cauchy). We consider sample sizes $T = 100$, $T = 500$ and $T = 1000$. The significance level of the tests is 5%. For the SS test, we set $\tau \in [0.10, 0.90]$ with increments of 0.01. Our results are based on $N = 10^5$ replications.

3.1 Serially Independent Innovations

We begin our experiment with the case in which the errors ε_t are i.i.d. and are distributed as t_∞ (normal), t_5, t_3, t_2 or t_1 (Cauchy). In Subsection 3.2,

[3] All numerical procedures used in this paper are implemented in R, and can be downloaded from http://homepages.nyu.edu/~bpn207. R is a free computer programming language very suited to statisticians and econometricians, and can be downloaded from http://www.r-project.org.

Figure 1: Histogram of 10^5 realizations of the maximum of the absolute value of the approximating process, probability density estimated nonparametrically using a Gaussian kernel and bandwidth given by Silverman's rule-of-thumb (Silverman, 1986); and the quantiles 90%, 95% and 99%, showing that the critical values for the significance levels of 10%, 5% and 1% are 1.65, 1.77 and 2.01, respectively.



we investigate the effect of short memory via a bootstrap experiment, with the errors being $\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t$, $\varepsilon_0 = 0$ and ξ_t are i.i.d. innovations that can be distributed as t_∞, t_5, ... or t_1.

In this Subsection, since the errors are i.i.d., we use $q_T = 0$ lags to compute the long-run variance for all the four tests.

3.1.1 Size

Table 1: Size of the tests at 5% significance level.

	t_∞	t_5	t_3	t_2	t_1
			$T = 100$		
KPSS	0.049	0.050	0.048	0.044	0.029
IKPSS	0.049	0.050	0.050	0.049	0.050
XL	0.030	0.023	0.019	0.015	0.006
SS	0.039	0.040	0.040	0.040	0.039
			$T = 500$		
KPSS	0.050	0.049	0.049	0.045	0.028
IKPSS	0.051	0.050	0.050	0.050	0.049
XL	0.043	0.035	0.027	0.020	0.007
SS	0.050	0.049	0.049	0.049	0.049
			$T = 1000$		
KPSS	0.050	0.049	0.049	0.046	0.028
IKPSS	0.050	0.050	0.051	0.049	0.049
XL	0.047	0.039	0.032	0.022	0.007
SS	0.050	0.051	0.051	0.051	0.051

We first consider the size of the tests, so our Data Generating Process (DGP) is

$$y_t = \varepsilon_t,$$

with ε_t i.i.d. t_∞, ... or t_1. Our results are displayed in Table 1 and are easy to summarize. For the KPSS and IKPSS, our results are, as expected, very close to the ones obtained by de Jong et al. (2007). The KPSS test is undersized for infinite variance (and mean) fat tail distributions, t_2 and t_1. The size distortion of the KPSS test for the Cauchy distribution does not come as a surprise, since this test requires the existence of the first moment. The KPSS has correct size for the normal and finite variance fat tailed data, such as t_5 and t_3. The empirical size of the KPSS test does not seem to depend on the sample size. The IKPSS has empirical size very close to nominal size in all the cases.

The XL test is too conservative. It is even more undersized the smaller the sample size is and the fatter the tail of the distribution is. Under normality, it has empirical size close to nominal size for moderate sample sizes ($T = 500$ or $T = 1000$). Note that the XL test requires the existence of the first two moments, so this test is supposed to have a large size distortion for the t_2 distribution, and even larger for the Cauchy distribution.

Notice that the IKPSS and SS tests are robust to distributions without finite mean and/or variance. The robustness of the SS test does not come as a surprise since it is well known that quantile estimation does not depend on distributional assumptions. However, for very small samples ($T = 100$), the SS test is more conservative than the IKPSS. This happens because we estimate 81 unconditional quantiles in order to compute the SS test. Since the precision of such estimates depends on the density of observations around the quantiles, the performance of the SS test tends to deteriorate in very small samples. Indeed, for sample sizes equal to 500 or 1000 the empirical size of the SS test is very close to the nominal size, 0.05.

3.1.2 Power against Alternatives with Unit Root

We parameterize the unit root alternative in a fashion similar to de Jong et al. (2007),

$$y_t = \lambda r_t + \varepsilon_t,$$

where

$$r_t = \sum_{j=1}^{t} \mu_j$$

is a random walk, and μ_t and ε_t are i.i.d. and independent from each other, and follow the same distribution (normal, ... or Cauchy). The scale factor λ measures the relative importance of the random walk component. We considered $\lambda = 0.01$ and $\lambda = 0.1$.

First of all, the results, summarized in Table 2, indicate that the power of all the four tests is increasing on λ and on T, as one would expect.

As noted by de Jong et al. (2007), the IKPSS test has more power than the KPSS test for fat tail distributions, but it has less power for normal and t_5 distributions. Actually, the power of the IKPSS test is increasing on the fatness of the tail, which also happens with the SS test (and with the XL test, except for a few cases). Under normality, the KPSS has more power than the other three tests.

Both the SS and the IKPSS tests have more power than the XL test in all cases. Moreover, for this alternaive hypothesis of unit root, the XL test has less power than the KPSS test, except for the infinite mean distribution (Cauchy).

The SS test has performance very similar to the IKPSS test. In all the cases, the SS test has power very close to the winner, when it is not the winner itself. For the infinite mean cases (Cauchy distribution), the SS test is the most powerful test among all the four tests analyzed, except for one case ($T = 100$ and $\lambda = 0.01$).

Table 2: Power of the tests, at 5% significance level, against the alternative hypothesis of unit root.

	$\lambda = 0.01$					$\lambda = 0.1$				
	t_∞	t_5	t_3	t_2	t_1	t_∞	t_5	t_3	t_2	t_1
					$T = 100$					
KPSS	0.061	0.060	0.064	0.068	0.141	0.588	0.590	0.590	0.587	0.564
IKPSS	0.057	0.060	0.069	0.100	0.477	0.488	0.561	0.633	0.735	0.921
XL	0.035	0.028	0.026	0.029	0.146	0.442	0.468	0.502	0.563	0.679
SS	0.047	0.047	0.055	0.079	0.453	0.500	0.558	0.627	0.739	0.951
					$T = 500$					
KPSS	0.307	0.308	0.315	0.337	0.413	0.988	0.987	0.986	0.974	0.873
IKPSS	0.230	0.299	0.394	0.593	0.980	0.972	0.983	0.991	0.997	1.000
XL	0.213	0.211	0.229	0.293	0.513	0.980	0.980	0.982	0.984	0.963
SS	0.241	0.290	0.375	0.571	0.983	0.982	0.989	0.995	0.999	1.000
					$T = 1000$					
KPSS	0.606	0.607	0.608	0.608	0.582	1.000	0.999	0.999	0.996	0.934
IKPSS	0.507	0.595	0.697	0.858	0.999	0.998	0.999	1.000	1.000	1.000
XL	0.509	0.512	0.533	0.596	0.713	0.999	0.999	0.999	0.999	0.990
SS	0.539	0.605	0.694	0.853	1.000	0.999	1.000	1.000	1.000	1.000

3.1.3 Power against Alternatives with Unconditional Heteroskedasticity

Recall that the driving force of the KPSS (IKPSS) test is the fluctuation of the data around the sample mean (median). So they should have low power to detect processes with a constant distribution location, but with a distribution scale that changes over time. Such processes are not strict stationarity.

To investigate this possibility, we consider the following DGP:

$$y_t = \sqrt{1 + st}\varepsilon_t.$$

Notice that now the scale factor is varying over time! We considered $s = 0.01$ and $s = 0.05$. In this model, there is no unit root, the mean (when it exists) and median are constant over time, but the distribution scale is changing. More precisely, the variance (when it exists) is changing linearly over time at rate s.

Table 3 exhibits our results. Basically, the KPSS test has power equal to size even for large sample sizes ($T = 1000$). So, since it is undersized for fat tail distributions (t_2 and Cauchy), it has power less than significance level, 5%. In fact, it is a biased test (power less than size) in several instances.

The IKPSS test has power close to size. Even for large samples ($T = 1000$), the maximum power offered by the IKPSS is never more than 0.072.

The XL test has power against this alternative of time-varying scale for thin tail distributions. For the t_2 distribution, its power is low. For the Cauchy distribution, its power is very close to its size and, actually, it is never greater

Table 3: Power of the tests, at 5% significance level, against the alternative hypothesis of time-varying volatility.

	$s = 0.01$					$s = 0.05$				
	t_∞	t_5	t_3	t_2	t_1	t_∞	t_5	t_3	t_2	t_1
					$T = 100$					
KPSS	0.049	0.049	0.049	0.043	0.028	0.052	0.051	0.050	0.044	0.026
IKPSS	0.051	0.051	0.050	0.051	0.051	0.055	0.056	0.056	0.056	0.056
XL	0.098	0.050	0.032	0.019	0.006	0.388	0.171	0.086	0.039	0.008
SS	0.072	0.064	0.061	0.058	0.051	0.239	0.190	0.167	0.146	0.102
					$T = 500$					
KPSS	0.051	0.053	0.050	0.046	0.027	0.053	0.053	0.052	0.046	0.025
IKPSS	0.057	0.057	0.057	0.056	0.057	0.066	0.065	0.065	0.066	0.066
XL	1.000	0.823	0.399	0.110	0.010	1.000	0.943	0.600	0.188	0.012
SS	0.974	0.913	0.851	0.758	0.505	1.000	0.999	0.996	0.984	0.860
					$T = 1000$					
KPSS	0.054	0.053	0.051	0.047	0.026	0.056	0.052	0.051	0.046	0.026
IKPSS	0.060	0.060	0.060	0.061	0.060	0.072	0.068	0.070	0.070	0.070
XL	1.000	0.982	0.705	0.210	0.010	1.000	0.989	0.786	0.280	0.013
SS	1.000	1.000	1.000	1.000	0.973	1.000	1.000	1.000	1.000	0.999

than 0.013, even when $T = 1000$. This low power of the XL test for both the t_2 and the t_1 distributions are not so surprising, since this test requires the existence of the first two moments. These distortions of the XL test for the infinite variance and/or infinite mean cases can be seen throughout the paper, in several tables.

The SS test has very good power against this alternative of time-varying scale. Even for moderate sample sizes ($T = 500$), it offers power 1, or very close to 1, for almost all distributions and, when $s = 0.05$, it offers power above 98% for four (out of five) distributions. The SS test has more power than all the other tests in almost all cases, against this alternative hypothesis of time-varying scale.

3.1.4 Power against Alternative with Time-Varying Kurtosis

The results above show that both the XL and the SS tests can reveal lack of stationarity in the data even when it has constant mean (or median). In this sense, these tests can be used to test the null hypothesis of covariance stationarity.

However, if a process is strict stationary then the data must also have no excess fluctuation around other sample quantiles. Recall that the driving force of the new test is the fluctuation of the data around sample quantiles $\tau \in [0.10, 0.90]$. If the data exhibit excessive fluctuation around sample quantiles then the null hypothesis of strict stationarity will be rejected.

To investigate this, consider a family of real-valued discrete random variables $X(\nu)$ parametrized by $\nu \in \left[\sqrt{2}, \infty\right)$ and defined by the following probability mass distribution:

$$P\left(X(\nu) = x\right) = \begin{cases} \frac{1}{\nu^2} & \text{if } x = -\frac{\nu}{\sqrt{2}}, \\ 1 - \frac{2}{\nu^2} & \text{if } x = 0, \\ \frac{1}{\nu^2} & \text{if } x = \frac{\nu}{\sqrt{2}}, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\mathbb{E}\left[X(\nu)\right] = 0$, $\mathbb{E}\left[X(\nu)^2\right] = 1$, $\mathbb{E}\left[X(\nu)^3\right] = 0$ and $\mathbb{E}\left[X(\nu)^4\right] = \frac{\nu^2}{2}$, so the expectation, variance and skewness do not vary with ν, but the kurtosis depends on ν. Now, define

$$\eta_t := X\left(\sqrt{2} + 8\frac{t}{T}\right) \tag{3.1}$$

and consider the DGP

$$y_t = \eta_t + \varepsilon_t,$$

that is, the process is now the error ε_t, that can be distributed as normal, ..., or Cauchy, plus a discrete random variable η_t that has zero mean (and median) and skewness, and unit variance, but has time-varying kurtosis.

It is worthwhile to notice that Kapetanios (2007) says that stationarity tests applied to such processes with changes only in higher unconditional moments have not been analyzed in the literature, and Xiao and Lima (2007) say that many widely used stationarity tests cannot even capture changes in the unconditional variance.

As implicit in the definition of η_t, we choose the equation

$$\nu(t) := \sqrt{2} + 8\frac{t}{T}. \tag{3.2}$$

to relate the time t to the parameter $\nu(t)$. But why do we choose this equation? First, note that

$$\begin{aligned} P\left(X(\nu) \neq 0\right) &= P\left(X(\nu) = \frac{\nu}{\sqrt{2}}\right) + P\left(X(\nu) = -\frac{\nu}{\sqrt{2}}\right) \\ &= \frac{2}{\nu^2}, \end{aligned}$$

so we have to restrict ν to be at least $\sqrt{2}$. Also, note that $P\left(X(\nu) = 0\right) > 0.98$ if $\nu > 10$, so ν should not be much larger than 10 in our simulations. In summa, $\nu(t)$ cannot be less than $\sqrt{2}$, and it should not be larger than 10, hence Eq. (3.2) seems to be a reasonable choice.[4]

[4]The results of our simulation are sensitive to the choice of Eq. (3.2). More precisely, the SS test loses some power if $\eta_t \neq 0$ too seldom or too often, as one could expect. However, the other tests have never power against the alternative of time-varying kurtosis, no matter the choice for Eq. (3.2) is.

Table 4: Power of the tests, at 5% significance level, against the alternative hypothesis of time-varying kurtosis.

	t_∞	t_5	t_3	t_2	t_1
$T = 100$					
KPSS	0.049	0.048	0.049	0.045	0.027
IKPSS	0.052	0.051	0.051	0.051	0.052
XL	0.050	0.032	0.024	0.017	0.006
SS	0.085	0.070	0.065	0.060	0.051
$T = 500$					
KPSS	0.050	0.050	0.048	0.046	0.027
IKPSS	0.051	0.050	0.050	0.052	0.052
XL	0.060	0.044	0.032	0.022	0.007
SS	0.386	0.273	0.221	0.178	0.112
$T = 1000$					
KPSS	0.049	0.048	0.050	0.046	0.028
IKPSS	0.051	0.050	0.051	0.050	0.051
XL	0.058	0.046	0.035	0.023	0.007
SS	0.723	0.536	0.438	0.342	0.199

Since the KPSS and the IKPSS tests are not able to detect time-varying variance when the mean (when it exists) and median are constant over time, we expect they are not able to detect time-varying kurtosis when both the distribution location and the distribution scale are constant over time. This is exactly what we see in Table 4. Their power and size are about the same.

The XL test presents very low power (never greater than 0.06). Except for the normal distribution, its power is less than the significance level, 5%.

The new test has good power when the sample size is moderate ($T = 500$ and, specially, $T = 1000$). When the sample size is very small ($T = 100$), its power is small, but we have to considerate that the SS test is a bit too conservative when the sample size is very small. Our test performs well when the kurtosis exists (normal and t_5 distributions), as one would expect; its power decreases with the fatness of the tail.

These results show that the SS test can reveal lack of stationarity in the data even when they have constant mean (or median), variance and skewness (if they exist). The new test is actually testing the null hypothesis of strict stationarity.

3.2 Serially Dependent Innovations

In this Subsection, the errors ε_t are serially correlated. More specifically,

$$\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t, \tag{3.3}$$

with $\varepsilon_0 = 0$, $\rho \in (0, 1)$ and $\xi_t \sim i.i.d.t_\iota$, $\iota \in \{\infty, 5, 3, 2, 1\}$.

Then, we need a sampling scheme to compute size and power of the tests. Following Psaradakis (2006), we use the so-called stationary bootstrap method introduced by Politis and Romano (1994) that, like the regular block bootstrap proposed by Künsch (1989) and Liu and Singh (1992), does not depend on any parametric assumptions about the data-generating mechanism. However, unlike the block bootstrap, the stationary bootstrap generates resampled data that are stationary.

In the stationary bootstrap, (overlapping) blocks are draw, with replacement, from the original time series, until the resampled time series has the same length as the original time series. Each observation in the original time series has the same probability of being the beginning of a block. The length of each block is given by a geometric distribution with parameter p, so the average block length is $\frac{1}{p}$.

To considerably reduce the time required to run the simulations, we use the technique suggested by Davidson and MacKinnon (1999), in which they build only one resampled time series for each of the N replications of the Monte Carlo, instead of resampling many time series for each of the N replications. The idea is that each resampled time series come from the same DGP; they have the same distribution. So the test statistics extracted from them can be pooled together to form the empirical distribution of the test statistic under the null hypothesis of stationarity. In other words, $\mathbb{F}_i$, the frequency of rejection (size, under the null, or power, under the alternative hypothesis) of the test i, is given by

$$\mathbb{F}_i = \frac{1}{N} \sum_{j=1}^{N} 1_{\mathbb{T}_i^j > \mathbb{Q}_i^{(1-\alpha)}},$$

where $\mathbb{T}_i^j$ is the test statistic of the test i applied to the original time series generated by the Monte Carlo in the replication j, $\mathbb{Q}_i^{(1-\alpha)}$ is the $1-\alpha$ quantile of the empirical distribution of the test statistics of the i^{th} test calculated from the N resampled time series, $\left\{ \mathbb{T}_i^{j*} \right\}_{j=1}^{N}$, and $\alpha = 0.05$ is the significance level.

To estimate the long-run variance we use the Bartlett Kernel. We use the same number of lags used by both Kwiatkowski et al. (1992) and de Jong et al. (2007):

$$q_T = round \left[4 \left(\frac{T}{100} \right)^{\frac{1}{4}} \right]. \tag{3.4}$$

However, if the errors are highly serial correlated, using the number of lags given by Eq. (3.4) may generate underestimated long-run variances, which would lead to oversized tests; the tests would reject too often. On the other hand, if the errors present very small serial correlation, the use of Eq. (3.4) may overestimate the long-run variance, which would lead to a loss in power. To avoid these distortions, Xiao and Lima (2007) use, in their paper, a data-dependent bandwidth selection:

$$q_T^{(XL)} = round \left[\min \left\{ \left(\frac{3T}{2} \right)^{\frac{1}{3}} \left(\frac{2\left|\widehat{\rho}\right|}{1-\widehat{\rho}^2} \right)^{\frac{2}{3}}, 8 \left(\frac{T}{100} \right)^{\frac{1}{3}} \right\} \right], \tag{3.5}$$

where $\widehat{\rho}$ is an estimate of the first-order autoregression coefficient of $\widehat{\varepsilon}_t$. They also experiment with

$$q_T^{(XL)} = round\left[\min\left\{\left(\frac{3T}{2}\right)^{\frac{1}{3}}\left(\frac{2\left|\widehat{\rho}\right|}{1-\widehat{\rho}^2}\right)^{\frac{2}{3}}, 12\left(\frac{T}{100}\right)^{\frac{1}{3}}\right\}\right], \tag{3.6}$$

but this latter option seems to overestimate the long-run variance: the tests they analyze are undersized and with a lower power than the same tests but with the use of the Eq. (3.5).

To deal with these issues, Hobijn et al. (2004) suggest the modification of the Eq. (3.4) to

$$q_T = round\left[\aleph\left(\frac{T}{100}\right)^{\frac{1}{4}}\right],$$

where $\aleph$ is chosen in other to minimize these size distortions, maybe in a two-step procedure, as in Andrews (1991) or Newey and West (1994). Anyway, in their simulations, Hobijn et al. (2004) use $\aleph = 4$, and they show that the results are not too sensitive to the choice of the bandwidth. Some experiments we did reached the same conclusion, so we also use $\aleph = 4$, that is, we stick to the usual Eq. (3.4).[5]

In this Subsection, we report results for the cases with first-order autoregressive coefficient $\rho = 0.3$ and with parameter of the geometric distribution that gives the length of the blocks in the bootstrap $p = 0.1$. We also run simulations with $\rho = 0.6$ and $p = 0.03$, and the results do not change too much. However, in fact, for the cases with $\rho = 0.6$, a larger number of lags than what is given by Eq. (3.4) would be more suited, since the size distortions start to be very noticeable when $\rho = 0.6$. But, for $\rho = 0.3$, Eq. (3.4) is the one offering the best results.

Our Monte Carlo has $N = 10^5$ replications.

3.2.1 Size

Let us begin with the size of the tests. The DGP is again

$$y_t = \varepsilon_t,$$

where ε_t follows the AR(1) in Eq. (3.3) with ρ=0.3.

Table 5 brings the results. When $T = 1000$, all tests have size close to level (5%) for almost all distributions. But for the infinity mean distribution (Cauchy), the KPSS is slightly undersized, while the XL test seems to be oversized.

However, when the sample size gets smaller, all the four tests get somewhat oversized.

[5]See Hobijn et al. (2004) for a further and interesting discussion about the influence of the bandwidth on the size and power of stationarity tests.

Table 5: Size of the tests at 5% significance level.

	t_∞	t_5	t_3	t_2	t_1
			$T = 100$		
KPSS	0.068	0.070	0.069	0.066	0.058
IKPSS	0.070	0.069	0.071	0.070	0.070
XL	0.063	0.061	0.059	0.056	0.060
SS	0.073	0.072	0.071	0.074	0.075
			$T = 500$		
KPSS	0.056	0.058	0.056	0.052	0.044
IKPSS	0.053	0.056	0.054	0.056	0.056
XL	0.055	0.054	0.049	0.047	0.059
SS	0.054	0.056	0.052	0.054	0.053
			$T = 1000$		
KPSS	0.053	0.052	0.054	0.051	0.043
IKPSS	0.052	0.053	0.054	0.053	0.052
XL	0.051	0.052	0.049	0.045	0.063
SS	0.051	0.053	0.054	0.052	0.052

3.2.2 Power against Alternatives with Unit Root

Our DGP under the alternative hypothesis with unit root is

$$y_t = \lambda r_t + \varepsilon_t,$$

where

$$r_t = \sum_{j=1}^{t} \mu_j$$

is a random walk and ε_t follows the AR(1) in Eq. (3.3) with $\rho = 0.3$ and innovations ξ_t, and μ_t and ξ_t are i.i.d. and independent from each other, but follow the same distribution (normal, ... or Cauchy).

We can see in Table 6 that, similar to the results from last Subsection (i.i.d. case), the KPSS test has more power than the IKPSS under normality, but the IKPSS test has more power than the KPSS for fat tail distributions (t_3, t_2 and t_1). For the t_5 distribution, the KPSS and the IKPSS have about the same power. Actually, the power of the IKPSS, XL and SS tests are increasing with the fatness of the tail. Under normality, the KPSS test offers the best power.

The SS and the IKPSS tests have very similar results, and they are particularly good for fat tails distributions (specially t_3, t_2 and t_1).

The tests are vaguely less powerful here, with serially correlated errors, than in the last Subsection, where the errors were i.i.d.. However, they are, in general, very powerful. For $\lambda = 0.1$ and $T = 1000$, the power of the four tests are close to 1 regardless of the distribution.

Table 6: Power of the tests, at 5% significance level, against the alternative hypothesis of unit root.

	$\lambda = 0.01$					$\lambda = 0.1$				
	t_∞	t_5	t_3	t_2	t_1	t_∞	t_5	t_3	t_2	t_1
$T = 100$										
KPSS	0.074	0.073	0.072	0.076	0.115	0.316	0.319	0.314	0.324	0.333
IKPSS	0.072	0.074	0.078	0.091	0.235	0.286	0.315	0.339	0.391	0.484
XL	0.065	0.065	0.064	0.067	0.132	0.230	0.242	0.256	0.296	0.363
SS	0.076	0.078	0.078	0.090	0.234	0.285	0.305	0.327	0.391	0.545
$T = 500$										
KPSS	0.190	0.191	0.199	0.216	0.292	0.783	0.782	0.778	0.766	0.675
IKPSS	0.158	0.193	0.246	0.366	0.747	0.758	0.772	0.786	0.811	0.847
XL	0.137	0.142	0.161	0.207	0.354	0.748	0.745	0.751	0.774	0.771
SS	0.156	0.178	0.220	0.332	0.785	0.774	0.789	0.807	0.849	0.950
$T = 1000$										
KPSS	0.419	0.420	0.424	0.429	0.432	0.922	0.922	0.923	0.912	0.815
IKPSS	0.366	0.424	0.501	0.640	0.921	0.909	0.915	0.923	0.933	0.948
XL	0.332	0.339	0.368	0.438	0.533	0.913	0.908	0.908	0.917	0.901
SS	0.369	0.410	0.481	0.623	0.951	0.926	0.932	0.943	0.963	0.994

3.2.3 Power against Alternatives with Unconditional Heteroskedasticity

Again, the DGP is

$$y_t = \sqrt{1 + st}\varepsilon_t,$$

where ε_t follows the AR(1) in Eq. (3.3) with $\rho = 0.3$.

Table 7 shows that the XL test is somewhat more powerful here with serial correlated errors than in the counterpart i.i.d. case (Table 3). On the other hand, the SS test seems to have a lower power here, with serial correlated errors. Anyway, even though the power of the SS test reduces with the fatness of the tail, the SS test is more powerful than the XL test for fat tail distributions (t_3, t_2 and t_1), while the XL test is more powerful under normality. This is because the power of the XL test declines faster than the power of the SS test with the fatness of the tail. For the t_5 distribution, we have mixed results.

The power of the KPSS test is about its size. But, surprisingly, the IKPSS presents power faintly above size when $T = 500$ and, specially, when $T = 1000$. Anyway, its power is very small: it is never above 0.075 even when $T = 1000$ and $s = 0.5$.

3.2.4 Power against Alternative with Time-Varying Kurtosis

Finally, consider the DGP

$$y_t = \eta_t + \varepsilon_t,$$

Table 7: Power of the tests, at 5% significance level, against the alternative hypothesis of time-varying volatility.

	$s = 0.01$					$s = 0.05$				
	t_∞	t_5	t_3	t_2	t_1	t_∞	t_5	t_3	t_2	t_1
$T = 100$										
KPSS	0.068	0.068	0.069	0.064	0.057	0.071	0.069	0.069	0.064	0.055
IKPSS	0.068	0.069	0.070	0.070	0.072	0.075	0.075	0.072	0.074	0.076
XL	0.130	0.097	0.081	0.064	0.058	0.305	0.202	0.141	0.096	0.059
SS	0.098	0.093	0.093	0.088	0.081	0.201	0.172	0.159	0.141	0.112
$T = 500$										
KPSS	0.058	0.054	0.058	0.053	0.044	0.059	0.058	0.057	0.055	0.042
IKPSS	0.062	0.059	0.062	0.061	0.063	0.070	0.071	0.070	0.070	0.073
XL	0.873	0.722	0.393	0.161	0.059	0.998	0.872	0.557	0.237	0.062
SS	0.823	0.699	0.612	0.494	0.249	0.984	0.946	0.895	0.801	0.449
$T = 1000$										
KPSS	0.055	0.057	0.054	0.053	0.042	0.059	0.059	0.057	0.052	0.040
IKPSS	0.064	0.063	0.063	0.064	0.066	0.074	0.075	0.073	0.071	0.072
XL	1.000	0.967	0.686	0.260	0.066	1.000	0.980	0.757	0.328	0.068
SS	1.000	0.999	0.994	0.971	0.660	1.000	1.000	1.000	0.999	0.864

where η_t is given by Eq. (3.1) and ε_t follows the AR(1) in Eq. (3.3) with $\rho = 0.3$.

As you can see in Table 8, the KPSS, IKPSS and XL tests have power really close to size (Table 5), but the SS test has power against this alternative hypothesis of time-varying kurtosis. Its power declines with the fatness of the tail. Under normality, its power is 0.417 when $T = 1000$. However, the power of the SS test is clearly smaller than its counterpart with i.i.d. innovations (Table 4).

4 An Empirical Illustration

In this section we present an empirical analysis in which the use of the SS test can lead to a significant different finding.

We use the log returns on the S&P 500 index, from 01/03/1991 to 08/11/2008, summing up to $T = 4438$ observations. A visual inspection in the first panel of Figure 2 leads to the belief that the returns r_t exhibit mean reversion, which suggests that the returns r_t do not have a unit root. However note, yet in the first panel, that the variance seems to change over time; about the central third of the plot seems to have a higher volatility than the rest of the time series. Then, we expect that both the KPSS and the IKPSS tests cannot reject the null hypothesis of stationarity, but that both the XL and the SS tests can.

Time series commonly present some serial correlation, so the practitioner should use some resampling scheme to compute the p-values of the tests. In this

Figure 2: (1) Plot of the log returns on S&P 500 from 01/03/1991 to 08/11/2008; (2) plot of the standardized returns; (3) plots of the variances of both the returns and the standardized returns; (4) plot of the kurtoses of the standardized returns.

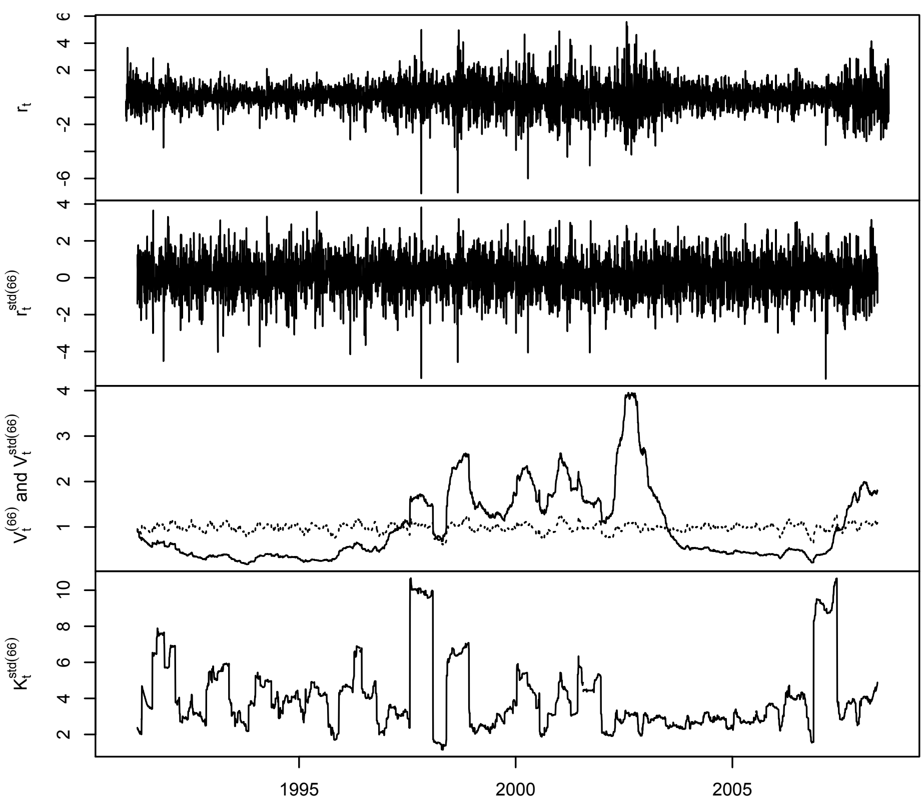


Table 8: Power of the tests, at 5% significance level, against the alternative hypothesis of time-varying kurtosis.

	t_∞	t_5	t_3	t_2	t_1
			$T = 100$		
KPSS	0.066	0.067	0.068	0.064	0.057
IKPSS	0.068	0.068	0.069	0.068	0.069
XL	0.086	0.069	0.063	0.059	0.059
SS	0.100	0.091	0.089	0.083	0.076
			$T = 500$		
KPSS	0.054	0.057	0.056	0.051	0.044
IKPSS	0.055	0.055	0.055	0.055	0.056
XL	0.071	0.059	0.051	0.047	0.058
SS	0.219	0.159	0.127	0.097	0.064
			$T = 1000$		
KPSS	0.052	0.053	0.054	0.051	0.042
IKPSS	0.052	0.051	0.054	0.052	0.054
XL	0.063	0.057	0.049	0.045	0.063
SS	0.417	0.271	0.209	0.145	0.072

Section we report results both by comparing the tests statistics to the critical values of the asymptotic distributions and by p-values computed via bootstrap. Since the returns r_t are reasonably uncorrelated,[6], the two approaches agree, that is, the critical values of the asymptotic distributions are valid.

Due to the lack of serial correlation in the returns, we use $q_T = 0$ lags to compute the long-run variance, *i.e.*, the long-run variance is simply the contemporaneous variance. Also, we employ a simple bootstrap (block length is one), which can be seen as a particular case of the stationary bootstrap used in the last Section with $p = 1$, the probability of the geometric distribution that gives that gives the length of the blocks. We tried some different number of lags q_T and some different probabilities p of the geometric distribution, but the results do not vary much. The bootstrap is based on $R = 10^5$ replications.[7]

The p-value $\mathbb{P}_i$ of the test i is then given by

$$\mathbb{P}_i = \frac{1}{R} \sum_{j=1}^{R} 1_{\mathbb{T}_i < \mathbb{T}_i^{j*}},$$

where $\mathbb{T}_i$ is the test statistic of the test i computed on the time series of returns and $\mathbb{T}_i^{j*}$ is the test statistic of the test i computed on the j^{th} bootstrapped sample.

[6]We cannot reject, at 5% of significance level, the null hypothesis that the first order autoregressive coefficient is zero.

[7]Note that, in this Section, on the contrary of the last Section, we are not using the Davidson and MacKinnon's (1999) fast bootstrap scheme, since there is no Monte Carlo. Rather, there is only one original time series, the returns time series.

Table 9: Tests statistics of the four analyzed tests applied to 4438 observations of returns on the S%P 500 index. To indicate statistical significance at 10%, 5% and 1%, we use *, ** and ***, respectively. The lower panel shows the p-values, computed via bootstrap.

	r_t	$r_t^{std(121)}$	$r_t^{std(66)}$	$r_t^{std(33)}$
Tests Statistics				
KPSS	0.282	0.184	0.178	0.214
IKPSS	0.264	0.262	0.256	0.281
XL	6.298***	1.389	1.083	1.138
SS	4.933***	2.129***	2.153***	2.042***
P-Values (via bootstrap)				
KPSS	0.153	0.303	0.310	0.243
IKPSS	0.170	0.174	0.179	0.152
XL	0.000	0.475	0.823	0.769
SS	0.000	0.005	0.003	0.009

The results can be seen in the first column of Table 9. As we expected, both the KPSS and the IKPSS tests fail to reject the null hypothesis of stationarity since the returns are mean reverting, but both the XL and the SS tests reject the null of stationarity at 1% significance level. The absence of unit root is not a sufficient condition for stationarity since the scale of the return distribution may be varying over time. To visualize how the variance is varying on time, we compute variances using a rolling window of length $2h+1$.[8] More specifically, given $h \in \mathbb{N}$, define $\left\{V_t^{(h)}\right\}_{t=h+1}^{T-h}$ by

$$V_t^{(h)} := \frac{1}{2h+1} \sum_{j=t-h}^{t+h} r_j^2.$$

We compute this for windows of one year ($h = 121$), one semester ($h = 66$) and one quarter ($h = 33$).[9] We plot $V_t^{(66)}$ in the third panel of Figure 2 with a continuous curve; we will explain the dotted curve shortly. The plots for both the annually and the quarterly windows are very similar. Indeed, the variance is time-varying and it is larger in the middle third of the time series, as we could visually inspect in the plot of the returns r_t in the first panel.

Now, let us define a standardized return,

$$r_t^{std(h)} := \frac{r_t}{\sqrt{V_t^{(h)}}},$$

[8] We use a rolling window instead of an ARCH type variance (Engle, 1982) because we are interested in the unconditional variance rather than the conditional one.

[9] We adopt the usual approximations that one month has about 22 observations and that one year has about 252 observations).

and compute variances using the rolling window,

$$V_t^{std(h)} := \frac{1}{2h+1} \sum_{j=t-h}^{t+h} r_j^{std(h)^2}.$$

The standardized return with semesterly window $r_t^{std(66)}$ is plotted in the second panel of Figure 2, while its variance $V_t^{std(66)}$ is plotted in the third panel, as a dotted curve. The plots for both the annually and the quarterly windows are very similar.

Notice that the variance of the standardized return (the dotted curve) is always about the unity, *i.e.*, the standardized return has unit variance, constant over time. So, the standardized returns are probably covariance stationary. Indeed, when applying the four stationarity tests to the standardized return, we find the the KPSS, the IKPSS and the XL tests fail to reject the null hypothesis of stationarity. However, the SS test rejects it at any usual significance level. What is going on? Probably, the standardized returns have the first two moments constant over time, but it has higher moments that are time-varying.

To investigate this, let us compute kurtoses of the standardized returns using the rolling window:

$$K_t^{std(h)} := \frac{1}{2h+1} \sum_{j=t-h}^{t+h} r_j^{std(h)^4}.$$

The fourth and last panel of Figure 2 brings the plot of $K_t^{std(66)}$. Indeed, the kurtosis is not constant over time. Actually, it is quite erratic. We see similar behavior for the skewness, when computed with the rolling window. The plots for both the annually and the quarterly windows are not too dissimilar.

In summa, the SS test can capture these fluctuations in higher moments of the returns, and even in higher moments of the standardized returns, so it can strongly reject the null hypothesis of strict stationarity. This empirical exercise casts doubts on results in the literature that are obtained from models that assume stationarity of returns.

5 Conclusion

In this paper we introduce a new test for strict stationarity. We show, through comprehensive Monte Carlo experiments, that this test is comparable to both classical and new tests for stationarity, in terms of power against alternative hypothesis with unit root or unconditional heteroskedasticity.

More importantly, we show that this test has good power against alternative hypothesis with higher moments varying on time, like a time-varying kurtosis, while the other tests fail to reject this hypothesis.

It is important to note that the new test requires larger sample sizes to have power, against this alternative hypothesis of time-varying kurtosis, approaching 1, when the first three moments (when they exist) are constant over time.

However, this is not an issue when considering sample sizes typically used in financial econometrics, as the empirical exercise shows.

Moreover, the new test is particularly more powerful than the other analyzed tests for fat tail distributions. This superiority of the SS test for fat tail distributions makes it very suitable when analyzing financial time series, which are know for presenting fat tails.

To reinforce this suitability of the SS test for financial econometrics, we finish the paper with an empirical exercise in which the SS test is the only analyzed test able to detect the non-stationarity of the standardized returns on the S&P 500 index. This result leads us to believe that many models that assume stationarity of stock returns may not be a good approximation of the reality.

References

[1] Andrews, D. W. K., 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," Econometrica 59 (3), 817-58.

[2] Busetti, F. and A. Harvey, 2007. "Tests of Time-Invariance," Mimeo, University of Cambridge.

[3] Davidson, R. and J. G. MacKinnon, 1999. "The Size Distortion of Bootstrap Tests," Econometric Theory 15 (3), 361-76.

[4] de Jong, R. M., C. Amsler and P. Schmidt, 2007. "A Robust Version of the KPSS Test Based on Indicators," Journal of Econometrics 137, 311-33.

[5] Engle, R. F., 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of The Variance of United Kingdom Inflation," Econometrica 50 (4), 987-1007.

[6] Hobijn, B., P. H. Franses and M. Ooms, (2004). "Generalizations of the KPSS-Test for Stationarity," Statistica Neerlandica 58 (4), 483-502.

[7] Kapetanios, G., 2007. "Testing for Strict Stationarity," Working Paper 602, Queen Mary University of London.

[8] Künsch, H. R., 1989. "The Jackknife and The Bootstrap for General Stationary Observations," The Annals of Statistics 17 (3), 1217-41.

[9] Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin, 1992. "Testing the Null Hypothesis of Stationarity Against the Alternative of Unit Root," Journal of Econometrics 54, 159-78.

[10] Lima, L. R. and B. Neri, 2007. "Comparing Value-at-Risk Methodologies," Brazilian Review of Econometrics 27 (1), 1-25.

[11] Liu, R. Y. and K. Singh, 1992. "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in: LePage, R. and L. Billard (Eds.), Exploring The Limits of Bootstrap. Wiley, New York, 225-48.

[12] Newey, W. K. and K. D. West, 1994. "Automatic Lag Selection in Covariance Matrix Estimation," Review of Economic Studies 61 (4), 631-53.

[13] Politis, D. N. and J. P. Romano, 1994. "The Stationary Bootstrap," Journal of American Statistical Association 89, 1303-13.

[14] Psaradakis, Z., 2006. "Blockwise Bootstrap Testing for Stationarity," Statistics and Probability Letters 76, 562-70.

[15] Qu, Z., 2005. "Test for Structural Change in Regression Quantiles," Working paper, University of Illinois.

[16] R Development Core Team, 2008. "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria.

[17] Silverman, B. W., 1986. "Density Estimation for Statistics and Data Analysis," Chapman and Hall, London.

[18] Xiao, Z. and L. R. Lima, 2007. "Testing Covariance Stationarity," Econometric Reviews 26 (6), 643-67.