Generalizations of the KPSS-test for Stationarity

Bart Hobijn

Department of Economics, New York University

Philip Hans Franses* and Marius Ooms

Econometric Institute, Erasmus University Rotterdam

January 1998

Abstract

We propose automatic generalizations of the KPSS-test for the null hypothesis of stationarity of a univariate time series. We can use the tests for the null hypotheses of trend stationarity, level stationarity and zero mean stationarity. We derive the asymptotic null distributions and we determine the rates of consistency against relevant nonstationary alternatives. We compare the properties of the tests with those of other recently proposed tests for stationarity. Theoretical results and Monte Carlo simulations support the relevance of the tests when an autoregressive process with large positive autocorrelations is likely under the null hypothesis.

Econometric Institute Report, no. 9802/A
`http://www.eur.nl/few/ei/papers`

Keywords I(0) test, rate of consistency, long run variance, heteroskedasticity and autocorrelation consistent covariance estimation, Choi's test, Leybourne and McCabe's test

JEL Code C22, C12

Acknowledgements
We thank In Choi and Benedikt Pötscher for providing us with useful references.

*Corresponding author: Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL- 3000 DR Rotterdam, The Netherlands. e-mail: franses@few.eur.nl. GAUSS programs used to obtain the results in this paper are available from *http://pages.nyu.edu/~bqh8457*.

1 Introduction

Tests for the null hypothesis of stationarity have not yet become part of the standard tools of empirical time series analysts. In many cases, however, the hypothesis of stationarity is more likely than the more frequently used hypothesis of (autoregressive) unit root nonstationarity. If one only uses autoregressive unit root (Dickey-Fuller) type tests the hypothesis of stationarity is only chosen if one rejects the null hypothesis of a unit root. Most unit root tests have low power against stationary, and highly autoregressive alternatives. This standard approach therefore entails that stationarity is not often found.

An important argument against the use of tests for the null hypothesis of stationarity is the difficulty to control their size when the process is stationary, but highly autoregressive. Probably the best known test for stationarity in econometrics, the so-called KPSS test introduced by Kwiatkowski, Phillips, Schmidt and Shin (1992) is oversized in that case: it rejects the true hypothesis of stationarity too often, again leading to undue preference for the hypothesis if unit root nonstationarity.

In this paper we develop a simple and automatic test for stationarity against random walk alternatives. This test should have reasonable size properties for the stationary but highly autoregressive case, since this is often a plausible model in practice. At the same time we also want good consistency properties against unit root alternatives. We use basically the same test statistic as KPSS, that is the ratio of the properly scaled sample variance of the partial sums of a series and an estimate of the long run variance of a series. Stock (1994) discusses related tests for I(0), and their possible interpretations in particular. Choi and Yu (1997) show how to apply the statistics to test for arbitrary integer orders of integration.

The standard KPSS test is oversized for highly autoregressive processes because it employs a semiparametric heteroskedasticity and autocorrelation consistent covariance estimator (HAC) of the long run variance of the process with an important positive finite sample bias. However, for the HAC estimator one can choose other bandwidths than the ones suggested by KPSS. In finite samples, the choice of bandwidth implies the following trade-off. Choosing too large a bandwidth implies that the long run variance is overestimated: the test statistic becomes too small and the test will have little or no power in finite samples if one employs common nominal significance levels. On the other hand, if one chooses the bandwidth too small and the process is highly autoregressive, then the long run variance is underestimated, the test statistic becomes too large and the test is oversized. The introduction of a more convenient estimator of the long run variance under the null hypothesis does not automatically repair the KPSS-type test. Some long run variance estimators which work well under the null, lead to inconsistency of the KPSS-type test under random walk alternatives, i.e., the power of the test for some relevant alternatives does not approach 1 as the sample size increases. In this paper we suggest an automatic form of the KPSS-test that reduces this size distortion without suffering from inconsistency.

Similar to the analysis in Stock (1994), we generalize the KPSS-test in two directions. First we show that the KPSS-test retains its conventional asymptotic properties when the long run variance is calculated using the automatic data dependent bandwidth selection procedure introduced by Newey and West (1994). This implies that test results become less sensitive to the choice of an a priori bandwidth, because the properties of the process are taken into account. We show that automatic bandwidth selection leads to a significant reduction in the size distortion of the test in the relevant case of a highly autoregressive process.

Second, we show how one can extend the KPSS procedure to test the joint hypotheses of zero mean stationarity or level stationarity. These joint hypotheses differ from the marginal null hypothesis of pure I(0)-ness where the deterministic part of the null hypothesis is partly left unspecified. We derive the asymptotic distributions under these joint hypotheses. We show that KPSS-type tests (designed for random walk alternatives) are also consistent in the following cases: (i), the null hypothesis of zero mean stationarity is tested when the process is actually stationary around a non-zero mean or a deterministic trend, and (ii) the null hypothesis of level stationary is tested while the process is actually trend stationary. Hobijn and Franses (1997) give a practical example in which both the joint hypotheses of zero mean stationarity and level stationarity are economically relevant. They show how to relate these with the concepts of unconditional and conditional convergence of per capita income levels. The consistency against deterministic I(0) alternatives is in line with results of Choi (1994, page 727). He mentions that his tests for the null of stationarity asymptotically reject in these two cases too. However, where Choi considers this to be a defect, we actually consider this a useful result when interpreting zero mean stationarity or level stationarity as *joint* hypotheses on the mean of the process *and* its memory characteristics.

We compare the properties of the generalized KPSS tests with those of related stationarity tests from the econometric literature, namely the residual-based test of Choi (1994), and the (parametric) test of Leybourne and McCabe (1994), henceforth (LBM). These tests were not analyzed in Stock (1994). Two results are especially worth mentioning. The first is that the Newey and West (1994) automatic bandwidth selection procedure is not applicable to Choi's test, since it does not yield a consistent test statistic. The second is that the LBM-test is not consistent against the alternative of a pure random walk.

We employ Monte Carlo analysis throughout to see whether our asymptotic results are useful for relevant sample sizes. Our results serve as an update and extension of the Monte Carlo evidence in Stock (1994).

The paper consists of two main sections. In Section 2 we introduce our generalizations of the KPSS test and prove their asymptotic properties. We derive its asymptotic distributions for three null hypotheses: trend stationarity, level stationarity and zero mean stationarity. Furthermore, we determine the rates of consistency if one uses the automatic bandwidth selection procedures in the estimation of the long run variance. We do this both for the Bartlett and for the Quadratic Spectral (QS) window. We also prove consistency against trend stationarity and level stationarity, when level stationarity and zero mean stationarity are tested, respectively.

In Section 3, we compare the small sample properties of our generalizations with those of the conventional KPSS test and with other related tests introduced by Choi (1994) and Leybourne and McCabe (1994, henceforth LBM). In the first part of Section 3 we review these two alternative tests and show how they differ from the KPSS test. In the second part of Section 3 we present results of Monte Carlo simulations that illustrate our main points. Section 4 concludes.

For the practitioner, our results seem to suggest that the KPSS test with automatic bandwidth selection is preferable to the Choi- and LBM-tests, since it has better small sample properties than the test by Choi (1994) and, contrary to the LBM-test, it is consistent against all relevant non-stationary models. As for the choice of spectral window, our simulations suggest that the main difference between using the Bartlett and the QS window is the difference in the rate of consistency. The QS window yields a higher (asymptotic) rate of consistency which seems also to lead to higher power in small samples.

2 KPSS-test and generalizations

In section 2.1 we first introduce the basic model and notation for the KPSS test. Then we discuss so-called HAC covariance estimation of the long run variance. In section 2.3 we generalize the asymptotic results for the KPSS test in two directions. First, we show that the test retains the asymptotic properties derived by KPSS when it is calculated using either the Bartlett or QS kernel in combination with the automatic bandwidth selection procedure proposed by Newey and West (1994). Second, we show that the test is also consistent in case the process contains an intercept or deterministic trend whenever the hypotheses of zero mean stationarity and level stationarity are tested, respectively.

2.1 Model specification and test statistic

The model that we use to test for stationarity of the time series, y_t, is similar to the one used by KPSS and reads

$$y_t = \alpha + \beta t + d \sum_{i=1}^{t} u_i + \varepsilon_t \qquad t = 1, \ldots, T \tag{1}$$

where u_i and ε_t are both covariance stationary and short memory with mean zero and $d \in \{0, 1\}$. u_t and ϵ_t satisfy the usual assumptions as defined in Newey and West (1994) or Park and Phillips (1998). In particular the long run variance, defined below, exists and is strictly positive. Contrary to KPSS and LBM we will not only consider the null hypothesis of trend stationarity, but also those of level stationarity and zero mean stationarity. The corresponding parameter restrictions are tabulated in Table 1. Our statistic differs slightly for the various null hypotheses and so does its asymptotic distribution. Under the alternative ($d = 1$), the stochastic part of y_t consists of a random-walk component Σu_t and a noise component ϵ_t.

Our test statistic is a generalization of the KPSS test in the sense that it also considers the partial sum process of the residuals, e_t, of each of the following regressions, depending on the null hypothesis, i.e.,

$$H_\tau : \; y_t = a + bt + e_t \tag{2}$$

Table 1: Null Hypotheses

	Hypothesis	Restriction
H_τ	Trend stationarity	$d = 0$
H_μ	Level stationarity	$d = \beta = 0$
H_0	Zero mean stationarity	$d = \beta = \alpha = 0$

$$H_\mu : \; y_t = a + e_t \tag{3}$$

$$H_0 : \; y_t = e_t \tag{4}$$

We use $S_t = \sum_{i=1}^{t} e_i$, to obtain our respective test statistics, denoted by w_τ, w_μ or w_0, which are all based on

$$w = T^{-2} \sum_{t=1}^{T} S_t^2 / \widehat{\sigma}^2 \tag{5}$$

depending on the null hypothesis. Under the null hypothesis, $\widehat{\sigma}^2$ is a consistent estimator of the long run variance of ε_t, σ_ε^2, which is defined as

$$\sigma_\varepsilon^2 = \lim_{T\to\infty} T^{-1} E\left[\left(\sum_{t=1}^{T} \varepsilon_t\right)^2\right]$$

Note that S_t is I(1) under the null and I(2) under the alternative ($d = 1$).

We use the statistic in a one-sided test. The test rejects the I(0) hypothesis against I(1) alternatives for large values of w. The test is consistent for appropriate choices of $\widehat{\sigma}^2$. In this paper we do not consider a two-sided test of I(0)-ness of e_t against fractionally integrated alternatives I(δ), $\delta < 0$, or $\delta > 0$, see Lee and Schmidt (1996). In practice low values of the statistic may be associated with overdifferenced series. Note that $\sigma_\varepsilon^2 = 0$ if $\delta < 0$, and $\sigma_\varepsilon^2 \to \infty$ if $\delta > 0$. Both situations are ruled out by assumption.

2.2 Estimation of the long run variance

We first discuss the denominator of our test statistic in (5). One can also view this as an estimator of the spectral density of ε_t at frequency zero or as an estimator of T times the variance of the sample mean of y_t (in cases where $\beta = 0$). Hamilton (1994, section 10.5) presents a useful introduction to this topic.

So far, we only assumed that ε_t is covariance stationary with a finite and positive long run variance σ_ε^2. However, we need a consistent estimator of σ_ε^2 which is efficient under general assumptions on the serial correlation properties of ε_t. We employ estimators which are frequently used in so-called heteroskedasticity and autocorrelation consistent (HAC) estimation. Recent studies, as for example Den Haan and Levin (1997), suggest that the accuracy of inference obtained using w crucially depends on the actual choice of estimator for σ_ε^2. As an alternative to the HAC estimation one could employ a parametric filtering approach, by specifying a (ARMA-) model for ϵ_t and estimating σ_ϵ^2 implicitly using that model. LBM use such an approach,

Table 2: Kernels

Kernel	
Bartlett	$k_m(j) = \begin{cases} \left(1 - \frac{j}{m+1}\right) & \text{for } j \leq m \\ 0 & \text{otherwise} \end{cases}$
Quadratic Spectral	$k_m(j) = \frac{25}{12\pi^2 (j/m)^2} \left[\frac{\sin(6\pi(j/m)/5)}{6\pi(j/m)/5} - \cos(6\pi(j/m)/5) \right]$

see section 3 below. Here we employ the nonparametric approach that is based on estimators of the form

$$\widehat{\sigma}^2 = \widehat{\gamma}_0 + 2 \sum_{j=1}^{T-1} k_m(j)\, \widehat{\gamma}_j \tag{6}$$

where $\widehat{\gamma}_j = T^{-1} \sum_{t=j+1}^{T} e_t e_{t-j}$ is used as the estimate of the j-th order autocovariance of ε_t, and $k_m(.)$ is a kernel function depending on a bandwith parameter m.

Several kernels have been proposed to weight the estimated autocovariances in (6). We consider two of them: the Bartlett kernel and the Quadratic Spectral kernel (QS), as tabulated in table 2. We consider only these two for the following reasons: the Bartlett kernel is used by KPSS, while the Quadratic Spectral kernel has been shown by Andrews (1991) to be more efficient. It has optimal asymptotic mean squared error properties and Monte Carlo results for small samples, documented by Andrews (1991) and Newey and West (1994), indicate that it yields more accurate estimates of σ_ε^2 than other kernels in finite samples. Note that the QS kernel gives a nonzero weight to all computable sample autocorrelations.

Next, we have to choose the bandwidth, m. We compare two approaches. The first, as used by KPSS, is simply to choose m_T as a deterministic function of the sample size T. Using the Bartlett kernel, KPSS assure consistency of $\widehat{\sigma}^2$ by choosing $m_T = O\left(T^{1/4}\right)$. [1] In particular, they choose $m_T(x) = \text{integer}\left[x\,(T/100)^{1/4}\right]$. However, if y_t is highly autoregressive with positive serial correlation, it turns out that $\widehat{\sigma}^2$ strongly depends on the choice of x: In case x is chosen large, then $\widehat{\sigma}^2$ is relatively large, w relatively small, and consequently the KPSS test turns out to have low power in small samples if one uses conventional nominal significance levels. In case x is chosen small, then, in small samples, $\widehat{\sigma}^2$ underestimates σ_ε^2 such that the KPSS test statistic is oversized. Hence, it can be preferable to use a data dependent procedure to estimate the optimal bandwidth parameter m_T, which thereby becomes a stochastic variable, $\widehat{m}_T$. This second approach was first explored by Andrews (1991) and later refined by Newey and West (1994). Newey and West's (1994) procedure is summarized in table 3.

In order to apply Newey and West's procedure, we again have to choose an a priori nonstochastic bandwidth parameter, in this case n_T. However, since outcome of the final bandwidth, $\widehat{m}_T$, only indirectly depends on n_T, it turns out that the value of $\widehat{\sigma}^2$ depends much less on n_T than on the deterministic choice of m_T in the previous approach. Hence, when the automatic bandwidth selection is used, the final

[1] Andrews (1991) showed that in case the QS kernel is applied, a choice of m such that $m_T = o\left(T^{1/2}\right)$ will yield consistency of $\widehat{\sigma}^2$.

Table 3: Automatic bandwidth selection procedure

	Step	Kernel	
		Bartlett	*Quadratic Spectral*
1.	choose bandwidth parameter n_T, where $n \to \infty$ as $T \to \infty$ and	$n = o\left(T^{2/9}\right)$	$n = o\left(T^{2/25}\right)$
2.	calculate $\widehat{s}^{(0)} = \widehat{\gamma}_0 + 2\sum_{i=1}^{n} \widehat{\gamma}_i$		
3.	calculate $\widehat{s}^{(j)} = 2\sum_{i=1}^{n} i^j \widehat{\gamma}_i$		
4.	calculate	$\widehat{\gamma} = 1.1447\left(\left[\frac{\widehat{s}^{(1)}}{\widehat{s}^{(0)}}\right]^2\right)^{1/3}$	$\widehat{\gamma} = 1.3221\left(\left[\frac{\widehat{s}^{(2)}}{\widehat{s}^{(0)}}\right]^2\right)^{1/5}$
5.	choose bandwidth parameter	$\widehat{m}_T = \min\left\{T, \text{integer}\left[\widehat{\gamma}T^{1/3}\right]\right\}$	$\widehat{m}_T = \min\left\{T, \widehat{\gamma}T^{1/5}\right\}$

realization of the test statistic, and thus the conclusion of the test, depends less on this a priori choice. As Monte Carlo results in section 3 show, in the most relevant cases the best small sample results for w are obtained using an automatic bandwidth selection for the Quadratic Spectral kernel.

Two other approaches for the estimation of σ_ϵ^2, which could be more efficient if e_t closely resembles an autoregressive process, turn out to have serious defects in our context. The first tries to improve the estimator $\widehat{m}_T$. Andrews (1991) suggests the use of a least squares estimate of an AR model in order to modify m_T. However, Choi (1994, section 4) shows that Andrews' choice for $\widehat{m}_T$ grows too fast under the stochastic trend alternative with $d = 1$: $\widehat{m}_T = O_p(T)$. This results in a large denominator in (5) and therefore in an inconsistent test. Of course, one can truncate the stochastic $\widehat{m}_T$ using another deterministic upper bound, see e.g. Stock (1994) or Choi and Yu (1997), but this introduces a second arbitrary choice in the testing procedure, which we seek to avoid.

A second approach which leads to inconsistency is to directly "prefilter" e_t using a least squares estimate of an AR model, as suggested by Andrews and Monahan (1992). Under the null, this procedure can be viewed as a helpful model based correction for serial correlation, but under the I(1) alternative it leads to differencing of e_t which prevents divergence of test statistic (5) as $T \to \infty$, and therefore to inconsistency.

2.3 Asymptotic null distribution and consistency

In this subsection we present asymptotic properties of the test statistics under the null hypothesis as well as under the alternative. The proofs do not require any novel techniques and are given in the appendix. It turns out that the asymptotic properties differ slightly under the three null hypotheses. Our tests are consistent under all relevant alternatives. The rate of consistency depends on the type of kernel and on the corresponding bandwith selection procedure applied when estimating σ_ε^2.

We follow the notation of KPSS, which we summarize in Table 4. We derive the asymptotic distributions for the test statistic under the various null hypotheses stated in the following proposition that we prove in appendix A.

Proposition 1 Asymptotic distributions: If y_t can be represented by equation (1), then under the various null hypotheses H_τ, H_μ and H_0 the test statistic w

Table 4: Notation

$W(r)$	Standard Wiener process
$V(r)$	First level Brownian bridge
$V_2(r)$	Second level Brownian bridge
$\underline{W}(r)$	Demeaned Wiener process
$W^*(r)$	Demeaned and detrended Wiener process
$\rightarrow$	Weak convergence
σ_ε^2	Long run variance of the noise component, ε_t
σ_u^2	Variance of changes in the random walk component, u_t

Table 5: Upper tail critical values

Level	**10%**	**5%**	**2.5%**	**1%**
H_τ	0.119	0.148	0.178	0.219
H_μ	0.348	0.460	0.580	0.754
H_0	1.195	1.656	2.114	2.759

Based on 50000 Monte Carlo replications with sample size 5000

in (5) has the following asymptotic distributions

$$H_\tau: \quad w_\tau \rightarrow \int_0^1 V_2^2(r)\,dr$$

$$H_\mu: \quad w_\mu \rightarrow \int_0^1 V^2(r)\,dr$$

$$H_0: \quad w_0 \rightarrow \int_0^1 W^2(r)\,dr$$

Table 5 contains the critical values for the one-sided tests at various significance levels under the three null hypotheses. We reject against random walk alternatives for large values of w. It is also crucially important to consider the behavior of the statistic under relevant alternatives. We show in the appendix that the test based on w is consistent for all relevant alternatives. Our result, stated in the following proposition, generalizes that of KPSS in two ways. First, we show that the test for level stationarity and zero mean stationarity are not only consistent against the alternative where y_t contains a stochastic trend ($d = 1$) in (1). The tests are also consistent against alternatives where y_t contains a deterministic trend ($\beta \neq 0$) or an intercept ($\alpha \neq 0$), respectively. Second, we show that even when the automatic bandwidth selection procedure is used, which can sometimes improve the small sample size properties, the test is still consistent.

Proposition 2 Consistency: If y_t can be represented as in equation (1), then[2]

[2]More specifically, the rate of consistency is at least $T^{3/4}$in case the Bartlett kernel is used without automatic lag selection, $T^{14/25}$ in case the Bartlett kernel is used with automatic lag selection and $T^{92/125}$ in case the QS kernel is used with automatic lag selection.

1. *If* $d = 1$, *then* $w_\tau \to \infty$ as $T \to \infty$.

2. *If either* $d = 1$ *or* $\beta \neq 0$, then $w_\mu \to \infty$ as $T \to \infty$

3. *If either* $d = 1$, $\beta \neq 0$ or $\alpha \neq 0$, then $w_0 \to \infty$ as $T \to \infty$

   More specifically, the rate of consistency against fixed alternatives with $d = 1$ is at least $T^{3/4}$ for the Bartlett kernel without automatic bandwidth selection, $T^{14/25}$ for the Bartlett kernel with automatic bandwidth selection, and $T^{92/125}$ for the QS kernel with automatic bandwidth selection. This means $w = O_p(T^v)$, with $v = 3/4, 14/25$, and $92/125$ respectively.

This result implies that our test will asymptotically have a power of unity. That is, as the sample size, T, goes to infinity we will reject a false null hypothesis with certainty. In practice, however, our sample size is finite and the test will not have such power. In the following section we compare the small sample properties of the test statistic for the different estimation methods of the long run variance of ε_t and for different null hypotheses. We also compare the properties of the tests with the stationarity tests proposed by Choi (1994) and LBM.

3 Alternative tests and a comparison of small sample properties

The KPSS test is not the only relevant test for stationarity. In particular, Choi (1994) and LBM (1994) have proposed related tests in the econometric literature. In this section we briefly review their methods and illustrate how they differ from the KPSS test. Finally, we compare the small sample properties of the KPSS test for different HAC estimators of σ_ε^2 with the small sample properties of the tests by Choi (1994) and LBM (1994). Although one can adapt both these tests for the null hypotheses of zero mean and level stationarity, we confine our analysis to the trend stationarity case to save space.

3.1 Alternative tests for stationarity

Instead of considering the residuals, e_t, of equation (2), Choi (1994) proposes to consider the slightly modified regression

$$C_t = \widehat{\alpha} t + \widehat{\beta} \left(\frac{t\,(t+1)}{2} \right) + \widehat{S}_t$$

where $C_t = \sum_{i=1}^{t} y_i$, i.e., the partial sum process of y_t. His test statistic, denoted by ω_2, equals

$$\omega_2 = \left\{ \frac{T^{-1} \sum_{t=2}^{T} \widehat{S}_{t-1} \Delta \widehat{S}_t}{\widehat{\sigma}^2} - \frac{1}{2} \left(1 - \frac{\widehat{\sigma}^2_{\Delta S}}{\widehat{\sigma}^2} \right) \right\}$$

where $\widehat{\sigma}^2$ is a HAC estimator of σ_ε^2 and $\widehat{\sigma}^2_{\Delta S} = T^{-1} \sum_{t=1}^{T} \Delta S_t^2$. Note that there is no intercept in the regression to construct $\widehat{S}_t$, since Choi assumes $y_0 = 0$, a relevant assumption in his derivation. Under the null hypothesis of trend stationarity, Choi shows that

$$\omega_2 \to \left[\int_0^1 W^*(s)\, dW^*(s)\right]$$

with exact percentiles as tabulated in Choi (1994, page 731: Table 2). Similar to the KPSS test, Choi's test depends on the HAC estimator $\widehat{\sigma}^2$. However, contrary to the KPSS test, for Choi's test we obtain

Proposition 3 ω_2 is not consistent when Newey and West's automatic bandwidth selection procedure is applied.

Proof: see appendix A.4.

Proposition 3 implies that when applying the Choi-test we still have to choose a deterministic bandwith m_T, which only depends on the sample size, in order to obtain a consistent test statistic.

If one is prepared to make more specific assumptions about the error process ϵ_t one can incorporate this extra information in the testing procedure and therefore hope to get a more efficient test. Leybourne and McCabe (1994), LBM, followed this approach, which turns out to be quite promising. They show that their test achieves the optimal consistency rate of T. Given their assumptions of the existence of a finite dimensional model, they do not need an infinite number of autocorrelations to consistently estimate the long run variance.

LBM assume a Gaussian stationary purely autoregressive process for y_t around a deterministic trend as their null hypothesis, i.e. $y_t - \alpha - \beta t \sim ARIMA\,(p,0,0)$. Since p is seldomly known a priori, one should check for sensitivity with respect to p in practice.

Their test consists of two stages. First they filter out all *non unit roots* of the autoregressive part of the process. Then they test for stationarity of the filtered process without a further correction for serial correlation. In order to obtain an estimate of these stable roots they estimate the following reduced form ARIMA$(p,1,1)$ model

$$\Delta y_t = \beta + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \eta_t - \theta \eta_{t-1} \tag{7}$$

assuming that $\eta_t \sim IIN\left(0,\sigma_\eta^2\right)$ using the now widely implemented method of exact maximum likelihood. The moving average parameter θ is included, because under the null hypothesis of stationarity, this model is overdifferenced and will contain a unit root in the MA-part, i.e., $\theta = 1$. Pötscher (1991) shows that, even though (7) is non-invertible in case of stationarity, the estimators of the autoregressive parameters, $\widehat{\phi}_i$, are consistent. Hence, by filtering out the autoregressive part of the process y_t through

$$y_t^* = y_t - \sum_{i=1}^{p} \widehat{\phi}_i y_{t-i} \tag{8}$$

y_t^* will, asymptotically, resemble an intercept, deterministic trend plus white noise under the null hypothesis. This means that we can calculate w using y_t^* instead of

e_t in (5) using the sample variance of y_t^* in the denominator. Unfortunately, this approach has a disadvantage, which we summarize in the following proposition.

Proposition 4 The LBM-test is not consistent against the alternative that $y_t \sim ARIMA(0,1,0)$, i.e. the alternative that y_t follows a pure random walk (with or without drift).

(*proof:*) In case $y_t \sim ARIMA(0,1,0)$, such that

$$y_t = \beta + y_{t-1} + \varepsilon_t$$

where ε_t is white noise, the estimated $ARIMA(1,1,1)$ model is

$$\Delta y_t = \phi \Delta y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}.$$

Clearly, the parameters ϕ and θ are not identified, leading to a nonvanishing positive probability that both $\widehat{\phi}$ and $\widehat{\theta}$ are arbitrarily close to unity as $T \to \infty$. Hannan and Deistler (1988, p. 183-184 and 204) formally show that, in this case, the likelihood function maximized by estimators $\widehat{\phi}$ and $\widehat{\theta}$, where in practice $\widehat{\phi}$ has to be confined to a stable and invertible region $[-1/(1+\delta), 1/(1+\delta)]$, will always have many local maxima for $\widehat{\phi}$ (and therefore here also for $\widehat{\theta}$) near the borders -1 and 1 as $\delta \to 0$ and $T \to \infty$. As $\delta \to 0$ and $T \to \infty$ and $\widehat{\phi}$ and $\widehat{\theta}$ remain unidentified, the filter (8) reduces with a nonnegligible probability to

$$y_t^* = \Delta y_t.$$

This shows that the unit root will be filtered out, resulting in a positive probability of the test statistic having the same asymptotic distribution under the alternative as under the null, which precludes consistency of the test.

The generalization of the LBM-test, introduced by McCabe, Leybourne and Shin (1997) to test for the null hypothesis of stationarity of the error term of a cointegrating regression and therefore for the null of cointegration, suffers from a similar problem. Note that the same problem does not occur for the generalized KPSS tests. There the consistency does not depend on the assumption that $\sigma_\epsilon^2/\sigma_u^2 > 0$ in (1) under the alternative with $d = 1$, see the proof of Proposition 2.

It thus turns out that the (often overlooked) identification assumption of the parameters in the auxiliary ARMA model is of crucial importance for the efficiency of the LBM test. For model (1) this means that the variance of random walk component $d\sum_{i=1}^{t} u_i$ must *not* be too large compared to the variance of noise component ϵ_t in order to reject often in large samples! The reduction in the power of the LBM test as the size of the random walk component *increases* is one the interesting results that we illustrate in the Monte Carlo simulations of the next section.

3.2 Monte Carlo simulations

The Monte Carlo results in this section illustrate four points. First, they show that the automatic data dependent bandwith selection procedure, in particular when used

Table 6: Bandwidth parameter sequences

Kernel	Bandwidth choice
Bartlett	$m_b(x) = \text{integer}\left[x(T/100)^{1/4}\right]$
	$n_b(x) = \text{integer}\left[x(T/100)^{1/4}\right]$
Quadratic Spectral	$m_{qs}(x) = \text{integer}\left[\frac{2}{3}x(T/100)^{2/9}\right]$
	$n_{qs}(x) = \text{integer}\left[x(T/100)^{2/25}\right]$

in combination with the QS kernel, yields better small sample results than when the bandwidth is chosen arbitrarily, as by KPSS. Second, we illustrate the validity of our extension of KPSS' consistency results by considering the size and power of the test in the presence of deterministic parts, i.e. an intercept and time trend. Thirdly, we verify Newey and West's (1994) claim that the outcome of the test is less sensitive to the bandwidth choice in case the automatic bandwidth selection procedure is used. Finally, we illustrate the inconsistency of the LBM test in case of a pure random walk alternative.

Tables 7 through 10, collected at the end of the paper, all contain the probability of rejection of the various null hypotheses at a nominal 5% significance level. The results presented are the outcomes of experiments for sample sizes 30, 100 and 500, each consisting of 1000 replications. $m_b(x)$, $n_b(x)$, $m_{qs}(x)$ and $n_{qs}(x)$ refer to the specific bandwidth choices, where m indicates a deterministic bandwidth choice a and n signifies that the automatic data dependent bandwidth selection procedure has been applied. The subscripts b and qs refer to the Bartlett and QS kernel respectively. The specific bandwidth choices are listed in table 6. $m_b(x)$ is similar to the bandwidth sequence used in KPSS, while, in order to make the results for the Bartlett and QS kernels comparable, $m_{qs}(x)$ is chosen such that $\widehat{\sigma}^2$ has the same asymptotic variance as when $m_b(x)$ is used. Both $n_b(x)$ and $n_{qs}(x)$ are the same as in Newey and West (1994). We follow Choi (1994) and have used the QS kernel for the calculation of his test, where we have chosen the bandwidth according to $m_{qs}(4)$. For the LBM-test we have chosen the AR-order $p = 1$ when the data generating process is an AR(1)-process and $p = 3$ whenever the process is MA(1).

Table 7 illustrates the empirical size properties of the various test statistics in case the data are generated by an AR(1) process $y_t = \phi\, y_{t-1} + \epsilon_t$. We use the header KPSS for the test statistics proposed in Section 2. Note first that the size properties are not much affected by the level of detrending. The most interesting case is that for an AR-parameter of .9. In this case the test is less oversized when the automatic bandwidth selection procedure is used compared to the regular KPSS test. In particular, for sample sizes 100 and 500 the size of the test statistic improves significantly when the bandwidth is selected automatically. In these cases the size does not differ much for both kernels. However, when we consider the power of the test statistic, as tabulated in Table 9, we find that in the case of automatic bandwidth selection the QS kernel yields a significantly higher power than the Bartlett kernel. This compares well with the higher rate of consistency of the test statistic when the QS kernel is used instead of the Bartlett kernel.

When compared to Choi's test we find that the KPSS test is more strongly oversized in case $\phi > 0$. However, Choi's test seems to have very low power; as can be seen in Table 9, even at a sample size of $T = 500$ in the most extreme case considered the false null hypothesis is only rejected with a probability of 60%, while for the KPSS test this is virtually 100%. As for the LBM test we find, as expected, that, because of the application of an autoregressive filter, it has superior size properties in small samples compared to both the KPSS- and Choi-test, see Table 7. Unfortunately, Table 9 illustrates the result of Proposition 4: the more the process resembles a random walk under the alternative, the less often the false null hypothesis of stationarity is rejected.

In completing our discussion of Table 7 we mention that Stock (1994) obtained an empirical size of 0.10 for $\phi = 0.9$ and $T = 100$, when he applied a KPSS test combined with a truncated version of an automatic AR(1) based bandwidth selection for the QS-kernel as suggested by Andrews(1991). This number is precisely between the results for the Choi test and the LBM test.

Table 8 shows the empirical size of the test statistics when the data are generated by an MA(1)-process. In general, the size of the test does not seem to depend on the bandwith selection procedure and kernel applied. This is to be expected because, if the process is MA(1), the truncation of the sample autocovariances matters less since the autocovariances of the order higher than 1 are zero. However, in small samples, it seems that the automatic bandwidth selection procedure selects $m = 0$ when $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$ where ε_t white noise and θ close to -1. This leads to an underestimation of the long run variance of y_t and, consequently, to an oversized test statistic in the columns $n_b(4)$ and $n_{qs}(4)$. This effect seems to dominate the downweighting (in absolute terms) of the negative first order covariance in (6) if $m > 0$. In itself, that effect leads to overestimation of the long run variance, and consequently to underrejection for the statistics that employ the fixed bandwidths, see the columns $m_b(4)$ and $n_{qs}(4)$.

The power results of Table 9 clearly show the deterioration of power as the order of detrending increases from demeaning to detrending. Stock (1994) provides analytical asymptotic local power results describing this phenomenon. Most tests are asymptotically locally most powerful in a set-up as presented in Table 9. Stock also provides empirical size-adjusted power results for generalizations of the KPSS test in an ARMA context.

From Table 10 one can easily see that the power of the test statistic against the alternative of a deterministic component is high. Zero mean stationarity is rejected in 100% of the cases for almost all the alternatives listed, apart from the one where an intercept equals only 1/10 of the variance of the noise. Level stationarity is rejected at a rate higher than 95% for all magnitudes of the deterministic trend studied, which should not come as a surprise given the superconsistency of $\widehat{\beta}$ in this normal regression case. Hence, even in a relatively small sample, like $T = 30$ listed in the table, the test manages to reliably detect deterministic parts and reject the null hypothesis if necessary. Notice that this property is particularly useful for the empirical analysis in Hobijn and Franses (1997).

Table 11 illustrates a main point of the paper. It presents the empirical sizes and powers of the test statistic in case the data are generated by an AR(1) process for all four estimation methods of $\widehat{\sigma}^2$ and for several bandwidth sequence parameters, i.e. $x \in \{2, 4, 8\}$. This table illustrates Newey and West's claim that the outcome of the test will be less sensitive for the choice of the bandwidth parameter n_T in case the final bandwidth, $\widehat{m}_T$, is selected automatically than when m_T is chosen a priori. The table also shows that the procedure that we suggest does not work for stationary AR processes with an alternating autocorrelation function, see the results for $\phi = -0.9$ in the right hand side columns. So, strong seasonal movements have to be removed to avoid rejecting stationarity too often in small samples.

Finally, table 12 illustrates another important point of our paper: the inconsistency of the LBM test statistic against the alternative of a random walk. It presents the power at a nominal 5% significance level of all three tests. The LBM statistic is calculated for the null hypothesis that y_t is an AR(1)-process around a deterministic trend[3], while w_τ is calculated using the QS kernel and the automatic bandwidth selection procedure with $n_{qs}(4)$. As can be seen, the power of w_T converges to 1 as T goes to infinity, as proved in proposition 2. However, the power of the LBM-test does not increase as the sample size increases. That is, as $T = 5000$ the false null hypothesis is only rejected in 33% of the cases. Though Choi's test, contrary to that of LBM, is consistent against a pure random walk it seems to have an inferior rate of consistency compared to our version of the KPSS test.

4 Concluding remarks

We have extended the KPSS test for stationarity in two directions. First, we have shown that a combination of the KPSS test and the automatic bandwidth selection procedure introduced by Newey and West (1994) yields a consistent test statistic that has better small sample properties than the original KPSS test. Secondly, we have derived the asymptotic distribution of the KPSS test for the additional null hypotheses of zero mean stationarity and level stationarity and we have shown the consistency of the test against the alternative of a deterministic component, like an intercept or a deterministic trend.

Our Monte Carlo simulations show that the best small sample results of the test in case the process exhibits a high degree of persistence are obtained using both the automatic bandwidth selection procedure and the Quadratic Spectral kernel. The resulting small sample properties are still open to improvement. This might possibly be achieved by estimating a specific model for the time series studied, as proposed by Leybourne and McCabe (1994). However, such a parameterization of the test should not go together with the loss of consistency against a relevant class of alternatives.

[3]Because y_t is a random walk, the model (7) with $p = 1$ is overidentified. This means that the ML estimate $\widehat{\phi}_1$ is sensitive to the starting values used in the numerical maximization procedure applied to the loglikelihood. The starting values that we have used are the first order sample autocorrelation of y_t for ϕ_1 and 1 for θ, because these would be consistent estimators under the null hypothesis.

References

[1] Andrews, D.W.K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, 59, 817-858.

[2] Andrews, D.W.K., and Monahan, J.C., (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica*, 60, 953-966.

[3] Brockwell, P.J., and Davis, R.A. (1991), *Time Series: Theory and Methods (2nd ed.)*, New York, Springer Verlag.

[4] Choi, I. (1994), "Residual Based Tests for the Null of Stationarity with Applications to U.S. Macroeconomic Time Series", *Econometric Theory*, 10, 720-746.

[5] Choi, I., and Yu, B. (1997), "A General Framework for Testing I(m) against I($m + k$)", *Journal of Economic Theory and Econometrics*, 3, 103-138.

[6] Den Haan, W.J. and A. Levin (1997), "A Practitioner's Guide to Robust Covariance Matrix Estimation", IN: *Handbook of Statistics 15: Robust Inference*, Editors: Maddala, G.S., and Rao, C.R., Chapter 12, 291-341, Amsterdam, North-Holland, Elsevier.

[7] Fuller, W.A. (1976), *Introduction to Statistical Time Series Analysis*, New York, Wiley.

[8] Hamilton, J.D. (1994), *Time Series Analysis*, Princeton, NJ, Princeton University Press.

[9] Hannan, E.J. and Deistler, M., *The Statistical Theory of Linear Systems*, New York, Wiley.

[10] Hobijn, B. and P.H.B.F. Franses (1997), "Asymptotically Perfect and Relative Convergence of Productivity", *Econometric Institute Report no. 9725/A,* Erasmus University Rotterdam.

[11] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root", *Journal of Econometrics*, 54, 159-178.

[12] Lee, D., and Schmidt, P. (1996), "On the power of the KPSS test of stationarity against fractionally-integrated alternatives", *Journal of Econometrics,* 73, 285-302.

[13] Leybourne, S.J. and B.P.M. McCabe (1994), "A Consistent Test for a Unit Root", *Journal of Business and Economic Statistics*, 12, 157-166.

[14] MacNeill, I.B. (1978), "Properties of Sequences of Partial Sums of Polynomial Regression Residuals with Applications to Tests for Change of Regression at Unknown Times", *The Annals of Statistics*, 6, 422-433.

[15] McCabe, B.P.M., S.J. Leybourne and Y. Shin (1997), "A Parametric Approach to Testing the Null of Cointegration", *Journal of Time Series Analysis*, 18, 396-413.

[16] Newey, W.K. and K.D. West (1994), "Automatic Lag Selection in Covariance Matrix Estimation", *Review of Economic Studies*, 61, 631-653.

[17] Park, J.Y. and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Processes: Part 1", *Econometric Theory*, 4, 468-497.

[18] Pőtscher, B.M. (1991), "Noninvertibility and Pseudo-Maximum Likelihood Estimation of Misspecified ARMA-models", *Econometric Theory*, 7, 435-449.

[19] Stock, J.H. (1994), "Unit Roots, Structural Breaks and Trends", IN: *Handbook of Econometrics, Volume IV*, Editors: Engle, R.F., and McFadden, D.L., 2739-2841, Amsterdam, Elsevier Science.

A Proofs of propositions

In this paper we define the order in probability, ν, of a sequence of random variables $\{m_T\}_{T=1}^{\infty}$ as follows. $m_T = O_p(T^{\nu})$ if for every $\epsilon > 0$, there exists an $M > 0$ such that $P\{|m_T| > M/(T^{\nu})\} < \epsilon$, see e.g. Brockwell and Davis (1991, section 6.1). We say that m_T/T^{ν} converges in probability to zero, $m_T = o_p(T^{\nu})$, if and only if $P\{|m_T|/(T^{\nu}) > \epsilon\} \to 0$ as $T \to \infty$.

A.1 Lemmas used to prove propositions 1 and 2

In order to derive the asymptotic distributions and prove the consistency of w for the various null hypotheses and their alternatives, we introduce the following two lemmas

Lemma 1 *Under the various null hypotheses, H_τ, H_μ and H_0, the estimator of the long run variance $\widehat{\sigma}_{\varepsilon}^2$ is a consistent estimator of σ_{ε}^2.*

Proof: Newey and West (1994) and Andrews (1991).

Lemma 2 *If the Bartlett kernel is used with automatic bandwidth selection, then $m = \widehat{\gamma} T^{1/3} = o\left(T^{13/27}\right)$*

Proof: We simply have to prove that $\widehat{\gamma} = o_p\left(T^{4/27}\right)$. In the case that the estimated covariances do not decrease, we obtain that

$$\widehat{s}^{(1)}/\widehat{s}^{(0)} = O_p(n) = o_p\left(T^{2/9}\right) \tag{9}$$

since this is the most extreme case possible, this result must hold in general. But then (9) implies that

$$\widehat{\gamma} = 1.1447\left(\left[\widehat{s}^{(1)}/\widehat{s}^{(0)}\right]^2\right)^{1/3} = o_p\left(T^{4/27}\right)$$

This again yields that

$$m = \widehat{\gamma} T^{1/3} = o_p\left(T^{13/27}\right)$$

which completes the proof.

Lemma 3 *If the QS kernel is used with automatic bandwidth selection, then $m = \widehat{\gamma} T^{1/5} = o_p\left(T^{33/125}\right)$*

Proof: The proof is similar to that of the previous lemma, but now we obtain that

$$\widehat{s}^{(2)}/\widehat{s}^{(0)} = O_p\left(n^2\right) = o_p\left(T^{4/25}\right)$$

which implies that

$$\widehat{\gamma} = 1.3221\left(\left[\widehat{s}^{(2)}/\widehat{s}^{(0)}\right]^2\right)^{1/5} = o_p\left(T^{8/125}\right)$$

and thus

$$m = \widehat{\gamma} T^{1/5} = o_p\left(T^{33/125}\right)$$

which, again, completes the proof.

Lemma 4 *Let y_t have a representation as in (1), then the partial sum process of the residuals, e_t, of the regression of y_t on an intercept and a deterministic trend has the following asymptotic distribution for $d = 0$ and $d = 1$, respectively:*

$$d = 0: \quad \forall r \in (0,1] \quad T^{-1/2}\sigma_\varepsilon^{-1} \sum_{t=1}^{\langle Tr \rangle} e_t \to V_2(r)$$

$$d = 1: \quad \forall r \in (0,1] \quad T^{-1/2}\sigma_u^{-1} \sum_{t=1}^{\langle Tr \rangle} e_t \to \int_0^r W^*(r)\, dr$$

where $\langle Tr \rangle$ is the integer part of Tr.

If $\beta = 0$ in (1), then the partial sum process of the residuals, e_t, of the regression of y_t on an intercept has the following asymptotic distribution for $d = 0$ and $d = 1$, respectively:

$$d = 0: \quad \forall r \in (0,1] \quad T^{-1/2}\sigma_\varepsilon^{-1} \sum_{t=1}^{\langle Tr \rangle} e_t \to V(r)$$

$$d = 1: \quad \forall r \in (0,1] \quad T^{-1/2}\sigma_u^{-1} \sum_{t=1}^{\langle Tr \rangle} e_t \to \int_0^r \underline{W}(r)\, dr$$

If $\alpha = \beta = 0$, then the partial sum process of y_t has the following asymptotic distribution for $d = 0$ and $d = 1$ respectively

$$d = 0: \quad \forall r \in (0,1] \quad T^{-1/2}\sigma_\varepsilon^{-1} \sum_{t=1}^{\langle Tr \rangle} e_t \to W(r)$$

$$d = 1: \quad \forall r \in (0,1] \quad T^{-1/2}\sigma_u^{-1} \sum_{t=1}^{\langle Tr \rangle} e_t \to \int_0^r W(r)\, dr$$

Proof: Follows from Park and Phillips (1988) and MacNeill (1978)

A.2 Proof of proposition 1

We will only prove the result for the null hypothesis of trend stationarity. The other cases follow in a similar way from the lemmas above. The result is obtained by using the consistency of $\widehat{\sigma}_\varepsilon^2$ and applying the continuous mapping theorem to the results of lemma 4. According to lemma 4, under H_τ,

$$T^{-1/2}\sigma_\varepsilon^{-1} S_{\langle Tr \rangle} \to V_2(r)$$

Applying the continuous mapping theorem, we now obtain that

$$T^{-1}\sigma_\varepsilon^{-2} S_{\langle Tr \rangle}^2 \to V_2^2(r),$$

from which it follows that

$$T^{-2}\sigma_\varepsilon^{-2} \sum_{t=1}^{T} S_t^2 \to \int_0^1 V_2^2(r)\, dr.$$

From lemma 1 we know that $\widehat{\sigma}_\varepsilon^2$ is a consistent estimator of σ_ε^2, and we can therefore replace σ_ε^2 by $\widehat{\sigma}_\varepsilon^2$ to find that

$$w_\tau = T^{-2}\widehat{\sigma}_\varepsilon^{-2}\sum_{t=1}^{T} S_t^2 \to \int_0^1 V_2^2(r)\,dr$$

which completes the proof.

A.3 Proof of proposition 2

We will only prove the proposition for the QS kernel with automatic bandwidth selection and for the following cases

1. If $d = 1$, then $W_\tau \to \infty$ as $T \to \infty$
2. If $\beta \neq 0$, then $W_\mu \to \infty$ as $T \to \infty$
3. If $\alpha \neq 0$, then $W_0 \to \infty$ as $T \to \infty$

The other results can be proved in a similar way as the proof below using the lemmas introduced above.

Part 1: If $d = 1$ then lemma 4 and the continuous mapping theorem can be applied to obtain that

$$T^{-1}\sigma_u^{-1}\sum_{t=1}^{T} S_t^2 = O_p\left(T^2\right) \tag{10}$$

Since, in this case, $e_t \sim I(1)$, that is e_t is integrated of the order 1, we obtain that

$$\forall k = 0, 1, \ldots : \ \widehat{\gamma}_k = O_p(T)$$

Since lemma 3 implies that $m = O_p\left(T^{33/125}\right)$, it follows that, because $\int_0^\infty k_m(x)\,dx = O_p(1)$,

$$\widehat{\sigma}_\varepsilon^2 = o_p\left(T^{33/125}\right)O_p(T) = o_p\left(T^{158/125}\right) \tag{11}$$

Combining (10) and (11) we obtain that if $d = 1$ then

$$w_\tau = T^{-2}\widehat{\sigma}_\varepsilon^{-2}\sum_{t=1}^{T} S_t^2 = \frac{O_p(T^2)}{o_p\left(T^{158/125}\right)}$$

Hence, when $d = 1$ then $W_\tau \to \infty$ as $T \to \infty$ and its rate of consistency is at least $T^{92/125}$.

Part 2: If $\beta \neq 0$, then it can be shown, as in Leybourne and McCabe (1994), that in case e_t is obtained from regression (3), then

$$\forall r \in (0, 1] : \ S_{\langle Tr\rangle} = \sum_{t=1}^{\langle Tr\rangle} e_t = O_p\left(T^2\right)$$

and

$$\forall k=0,1,\ldots:\ \widehat{\gamma}_k=O_p\left(T^2\right)$$

This implies that

$$T^{-2}\sigma_\varepsilon^{-1}\sum_{t=1}^{T}S_t^2=O_p\left(T^3\right)$$

and, since, according to lemma 3, $m=o_p\left(T^{33/125}\right)$, we obtain that

$$\widehat{\sigma}_\varepsilon^2=o_p\left(T^{33/125}\right)O_p\left(T^2\right)=o_p\left(T^{283/125}\right)$$

Similar to the proof of part 1 we now obtain that if $\beta\neq 0$, then $W_\mu\to\infty$ as $T\to\infty$ with a minimum rate of consistency that is again equal to $T^{92/125}$.

Part 3: This follows in a similar way as part 2 from the fact that if $\alpha\neq 0$ and e_t is obtained from regression (4), then

$$\forall r\in(0,1]:\ S_{\langle Tr\rangle}=\sum_{t=1}^{\langle Tr\rangle}e_t=O_p\left(T\right)$$

and

$$\forall k=0,1,\ldots:\ \widehat{\gamma}_k=O_p\left(1\right)$$

Combining these two results we now obtain that $W_0\to\infty$ as $T\to\infty$ with again a minimum rate of consistency that equals $T^{92/125}$.

A.4 Proof of proposition 3

We will prove this proposition only for the case in which the automatic bandwidth selection procedure is applied in combination with the QS window.

From Choi (1994, p. 744 eq. (A.4)) we obtain that, under the alternative

$$\sum_{t=2}^{T}\widehat{S}_{t-1}\Delta\widehat{S}_t=O_p\left(T^3\right)$$

and

$$\widehat{\sigma}_{\Delta S}^2=O_p\left(T\right).$$

We have already shown that under the alternative, $d=1$, $\widehat{\sigma}_\varepsilon^2=o_p\left(T^{283/125}\right)$. Hence, we obtain that

$$\frac{T^{-1}\sum_{t=2}^{T}\widehat{S}_{t-1}\Delta\widehat{S}_t}{\widehat{\sigma}_\varepsilon^2}=\frac{O_p\left(T^2\right)}{o_p\left(T^{283/125}\right)}$$

and thus

$$\omega_2=\left\{\frac{T^{-1}\sum_{t=2}^{T}\widehat{S}_{t-1}\Delta\widehat{S}_t}{\widehat{\sigma}^2}-\frac{1}{2}\left(1-\frac{\widehat{\sigma}_{\Delta S}^2}{\widehat{\sigma}^2}\right)\right\}$$

is not always consistent. That is, we can not exclude the possibility that $\omega_2=O_p\left(1\right)$ or even $\omega_2=o_p\left(1\right)$. The latter would imply that asymptotically the false null hypotheses would not be rejected with certainty.

Appendix B to: Econometric Institute Report 9802/A by Hobijn, Franses and Ooms, Tables 7-12

Table 7: Empirical size of test statistics in case of an AR(1) process

DGP: $y_t = \phi y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim IN(0,1)$

		H_τ						H_μ				H_0			
		KPSS				Choi	LBM	KPSS				KPSS			
T	ϕ	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_{qs}(4)$	$p=1$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$
30	*.9*	.44	.28	.80	.18	.17	.19	.49	.32	.74	.26	.54	.48	.69	.47
	.5	.13	.16	.36	.15	.25	.07	.12	.09	.27	.08	.21	.21	.30	.20
	0	.03	.31	.05	.33	.36	.04	.03	.09	.04	.08	.07	.14	.05	.14
	-.5	.02	.56	.00	.63	.50	.04	.01	.15	.01	.20	.04	.13	.00	.18
	-.9	.00	.88	.00	.98	.68	.04	.00	.06	.00	.49	.00	.06	.00	.20
100	*.9*	.58	.32	.85	.30	.08	.14	.47	.28	.71	.27	.43	.33	.57	.33
	.5	.09	.09	.20	.08	.13	.05	.10	.09	.16	.08	.11	.11	.16	.11
	0	.04	.06	.05	.06	.20	.05	.05	.06	.06	.06	.07	.09	.06	.09
	-.5	.01	.10	.02	.11	.31	.04	.02	.04	.04	.05	.03	.07	.04	.09
	-.9	.00	.01	.10	.50	.55	.05	.00	.00	.04	.10	.00	.01	.03	.09
500	*.9*	.68	.23	.87	.32	.07	.05	.47	.19	.65	.24	.39	.19	.51	.23
	.5	.10	.07	.14	.07	.07	.05	.08	.05	.11	.05	.09	.08	.11	.08
	0	.04	.04	.04	.04	.09	.05	.04	.04	.05	.05	.05	.06	.05	.05
	-.5	.03	.04	.05	.06	.17	.06	.03	.03	.04	.05	.04	.04	.05	.06
	-.9	.01	.02	.05	.07	.33	.06	.01	.01	.04	.04	.01	.04	.04	.06

NOTE: Probability of rejection at a nominal 5% level. Results based on 1000 replications

Table 8: Empirical size of test statistics in case of an MA(1) process

DGP: $y_t=\varepsilon_t+\theta\varepsilon_{t-1}$ $\varepsilon_t \sim IN(0,1)$

		H_τ						H_μ				H_0			
		KPSS				Choi	LBM	KPSS				KPSS			
T	θ	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_{qs}(4)$	$p=3$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$
30	*.9*	.06	.26	.21	.24	.27	.05	.06	.07	.15	.07	.12	.17	.18	.17
	.5	.07	.29	.18	.29	.29	.03	.06	.08	.14	.08	.13	.20	.18	.19
	0	.04	.34	.06	.35	.35	.02	.04	.09	.06	.10	.08	.15	.07	.17
	-*.5*	.00	.56	.00	.67	.50	.01	.00	.16	.00	.22	.01	.11	.00	.16
	-*.9*	.00	.68	.00	.89	.63	.00	.00	.18	.00	.56	.00	.00	.00	.10
100	*.9*	.05	.07	.07	.06	.14	.01	.06	.07	.08	.06	.08	.09	.08	.08
	.5	.05	.07	.07	.06	.17	.03	.06	.07	.08	.06	.09	.10	.09	.09
	0	.04	.06	.05	.07	.19	.04	.04	.05	.05	.05	.07	.08	.06	.08
	-*.5*	.01	.26	.00	.23	.38	.03	.01	.04	.01	.06	.01	.08	.00	.09
	-*.9*	.00	.47	.00	.85	.59	.00	.00	.09	.00	.36	.00	.00	.00	.06
500	*.9*	.06	.06	.06	.05	.09	.02	.07	.06	.06	.06	.07	.07	.07	.07
	.5	.06	.07	.06	.06	.09	.04	.06	.06	.06	.06	.07	.08	.07	.07
	0	.06	.06	.06	.06	.10	.05	.04	.04	.05	.05	.04	.04	.04	.05
	-*.5*	.01	.03	.01	.04	.18	.02	.01	.03	.01	.04	.01	.05	.01	.05
	-*.9*	.00	.21	.00	.45	.55	.00	.00	.05	.00	.07	.00	.02	.00	.03

NOTE: Probabilities of rejection at a 5% nominal level. Results based on 1000 replications

Table 9: Power of test statistics

DGP: $y_t = y_0 + \sum_{i=1}^{t} u_i + \varepsilon_t$ $\varepsilon_t \sim IN(0,1)$, $u_t \sim IN(0,\sigma^2_u)$. *($y_0$ based on 50 startup observations)*

		H_τ						H_μ				H_0			
		KPSS				Choi	LBM	KPSS				KPSS			
T	σ^2_u	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_{qs}(4)$	$p=1$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$
30	0.01	.06	.31	.08	.08	.34	.06	.13	.15	.16	.12	.60	.63	.60	.60
	0.1	.20	.31	.28	.17	.24	.14	.40	.33	.51	.35	.85	.84	.87	.84
	1	.46	.35	.71	.34	.15	.26	.61	.47	.79	.54	.87	.84	.92	.85
	10	.49	.34	.83	.36	.16	.24	.62	.47	.84	.54	.87	.84	.92	.85
	100	.51	.35	.84	.37	.16	.28	.65	.50	.85	.56	.89	.85	.93	.87
100	0.01	.29	.26	.34	.28	.12	.32	.52	.49	.56	.52	.77	.75	.80	.77
	0.1	.69	.57	.81	.68	.10	.74	.76	.65	.88	.76	.87	.82	.93	.87
	1	.81	.57	.95	.79	.15	.45	.82	.69	.93	.81	.87	.82	.92	.87
	10	.82	.60	.96	.80	.17	.28	.81	.66	.94	.80	.86	.81	.93	.86
	100	.83	.60	.96	.82	.18	.29	.83	.69	.94	.83	.88	.83	.95	.88
500	0.01	.95	.89	.96	.95	.40	.97	.95	.88	.97	.96	.94	.85	.96	.95
	0.1	.99	.90	1.00	.99	.55	1.00	.99	.91	1.00	.99	.96	.85	.98	.97
	1	1.00	.91	1.00	1.00	.57	.59	.99	.89	1.00	1.00	.98	.86	.99	.98
	10	1.00	.91	1.00	1.00	.60	.30	1.00	.88	1.00	1.00	.98	.86	.98	.98
	100	1.00	.91	1.00	1.00	.61	.34	.99	.90	1.00	.99	.97	.86	.99	.98

NOTE: Empirical probabilities of rejection at nominal 5% significance level. Results based on 1000 replications

Table 10: Empirical power and size of test statistics in case of deterministic components

DGP: $y_t=\alpha+\beta t+\varepsilon_t \quad \varepsilon_t \sim IN(0,1)$

			H_τ				H_μ				H_0			
			Bartlett		*QS*		*Bartlett*		*QS*		*Bartlett*		*QS*	
T	α	β	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$	$m_b(4)$	$n_b(4)$	$m_{qs}(4)$	$n_{qs}(4)$
30	*0.1*	*0*	.04	.04	.04	.31	.03	.02	.04	.08	.12	.13	.10	.20
	1	*0*	.04	.04	.05	.36	.04	.03	.05	.11	1.00	1.00	1.00	1.00
	10	*0*	.04	.04	.06	.30	.03	.03	.04	.08	1.00	1.00	1.00	1.00
	0	*0.1*	.03	.03	.04	.32	.99	.98	.99	.96	.99	.98	1.00	.95
	0	*1*	.05	.03	.06	.31	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0	*10*	.03	.03	.04	.32	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
100	*0.1*	*0*	.04	.04	.05	.06	.04	.04	.04	.05	.16	.16	.15	.18
	1	*0*	.05	.05	.06	.07	.04	.04	.05	.04	1.00	1.00	1.00	1.00
	10	*0*	.04	.04	.05	.07	.04	.04	.05	.05	1.00	1.00	1.00	1.00
	0	*0.1*	.05	.05	.06	.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0	*1*	.03	.03	.04	.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0	*10*	.04	.04	.04	.06	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
500	*0.1*	*0*	.05	.05	.05	.05	.07	.07	.07	.06	.53	.53	.53	.53
	1	*0*	.04	.04	.04	.05	.06	.06	.06	.06	1.00	1.00	1.00	1.00
	10	*0*	.05	.05	.04	.05	.05	.05	.05	.04	1.00	1.00	1.00	1.00
	0	*0.1*	.05	.05	.06	.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0	*1*	.05	.05	.05	.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0	*10*	.05	.05	.05	.05	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

NOTE: Probabilities of rejection at a nominal significance level of 5%. Results based on 1000 replications

Table 11: Sensitivity of test statistics for bandwidth choice

DGP: $y_t=\phi y_{t-1}+\varepsilon_t \quad \varepsilon_t \sim IN(0,1)$

		H_τ											
		Bartlett						*Quadratic Spectral*					
T	ϕ	$m_b(2)$	$m_b(4)$	$m_b(8)$	$n_b(2)$	$n_b(4)$	$n_b(8)$	$m_{qs}(2)$	$m_{qs}(4)$	$m_{qs}(8)$	$n_{qs}(2)$	$n_{qs}(4)$	$n_{qs}(8)$
30	1.0	.69	.50	.18	.50	.32	.54	.89	.85	.34	.54	.22	.46
	.9	.62	.43	.15	.43	.27	.56	.85	.80	.30	.47	.19	.51
	.5	.21	.14	.07	.14	.20	.61	.40	.34	.09	.16	.18	.68
	0	.04	.03	.03	.06	.32	.59	.04	.04	.04	.06	.32	.75
	-.5	.01	.01	.02	.60	.55	.62	.00	.00	.16	.54	.61	.88
	-.9	.19	.00	.34	.27	.88	.60	.00	.00	.80	.90	.96	.99
100	1.0	.93	.82	.56	.76	.57	.40	.99	.96	.71	.78	.55	.31
	.9	.79	.58	.29	.50	.29	.20	.98	.84	.43	.53	.28	.13
	.5	.20	.11	.07	.12	.11	.15	.45	.22	.08	.11	.09	.14
	0	.05	.04	.04	.05	.05	.18	.05	.05	.04	.05	.06	.19
	-.5	.01	.02	.02	.02	.12	.28	.00	.04	.06	.06	.13	.30
	-.9	.00	.00	.00	.01	.01	.08	.00	.11	.30	.37	.54	.75
500	1.0	1.00	1.00	.97	.98	.91	.79	1.00	1.00	.99	1.00	.97	.85
	.9	.93	.66	.34	.42	.24	.14	1.00	.85	.45	.54	.33	.15
	.5	.21	.11	.07	.10	.08	.09	.51	.14	.07	.10	.07	.07
	0	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05	.05
	-.5	.01	.03	.03	.01	.03	.03	.00	.04	.04	.04	.04	.05
	-.9	.00	.00	.01	.01	.02	.05	.00	.05	.07	.07	.07	.05

NOTE: Probability of rejection at a nominal 5% significance level. Results based on 1000 replications.

Table 12: Power of the test statistics against a random walk

	DGP: $y_t=y_{t-1}+\varepsilon_t$ ε_t~$IN(0,1)$		
		H_τ	
	KPSS	Choi	LBM
T	$n_{qs}(4)$	$m_{qs}(4)$	$p=1$
30	.23	.15	.28
100	.56	.18	.30
500	.96	.58	.31
1000	.99	.67	.32
5000	1.00	.80	.33

NOTE: Probability of rejection at a 5% nominal significance level.
Results based on 1000 replications