Spurious logarithms and the KPSS statistic

Robert M. de Jong* and Peter Schmidt
Department of Economics
Michigan State University

December 4, 2001

Abstract

This paper analyzes the asymptotic behavior of two types of so-called KPSS tests when a logarithm transformation has been applied spuriously to data that are themselves an integrated time series. Although a different limit distribution is obtained, the asymptotic convergence behavior of the KPSS statistic is reminiscent of that of integrated time series, and it is shown that the KPSS test cannot distinguish consistently between an integrated time series and the logarithm of an integrated time series.

JEL numbers: C22; C32

Keywords: unit roots, unit root test, weak convergence, nonlinearity

*Address correspondence to: Robert M. de Jong, Department of Economics, Michigan State University, 217 Marshall Hall, East Lansing, MI 48824, USA, Email "dejongr@msu.edu", fax (517)-432-1068.

1 Introduction

Let w_t be an observable series. In this paper we ask whether we can distinguish the case that w_t is I(1) from the case that $\log(w_t)$, or more precisely $\log|w_t|$, is I(1). This is a matter of some importance in practice because the asymptotic theory for cointegrating regressions assumes that the variables in question are I(1), and it is not clear whether standard unit root tests can be trusted to distinguish whether w_t or $\log|w_t|$ is I(1).

Empirically, regressions are often run in logarithms and it is not clear whether standard theory applies. That is, if we regress $\log|y_t|$ on $\log|w_t|$, standard theory applies if these logarithmic variables are I(1) and cointegrated, but a *different* and as yet undeveloped theory would apply if y_t and w_t are I(1) in levels (but the regression is still run in logarithms).
In this paper we ask whether the KPSS tests of Kwiatkowski et al. (1992) can be used to make this distinction. The KPSS tests are most commonly used to test the null hypothesis of stationarity, but they are also standard tests of the unit root hypothesis (e.g., Shin and Schmidt (1992), Harvey (2001)).

For a sample $w_1, \ldots, w_n$, define $\bar{w}$ as the sample mean, and define the demeaned data as $w_t - \bar{w}$ $(t = 1, \ldots, n)$. There are three types of KPSS tests; the first one is based on the actual data w_t, the second is based on the demeaned data, and the third type considers demeaned and detrended data. The asymptotic distribution of all three statistics is a functional of Brownian motion. In this paper, we investigate the consequences for the first two types of KPSS statistics if the data w_t are generated as the logarithm of an integrated process. We find that the KPSS statistic cannot distinguish consistently between the case that w_t is I(1) and the case that $\log|w_t|$ is I(1).

2 The KPSS statistic, assumptions, and main results

The KPSS statistic "in levels" is defined as

$$\mathrm{KPSS}_1 = \frac{n^{-1}\sum_{t=1}^{n} S_t^2}{\sum_{t=1}^{n} w_t^2} \tag{1}$$

where $S_t = \sum_{j=1}^{t} w_j$, and the w_t are observed. It is well-known that for I(1) w_t, under regularity conditions,

$$n^{-1}\mathrm{KPSS}_1 = \frac{n^{-4}\sum_{t=1}^{n} S_t^2}{n^{-2}\sum_{t=1}^{n} w_t^2} \xrightarrow{d} \frac{\int_0^1 (\int_0^a W(r)dr)^2 da}{\int_0^1 W(r)^2 dr} \tag{2}$$

where "$\xrightarrow{d}$" denotes convergence in distribution, while for mean zero I(0) w_t, again under regularity conditions,

$$\text{KPSS}_1 = \frac{n^{-2}\sum_{t=1}^{n} S_t^2}{n^{-1}\sum_{t=1}^{n} w_t^2} \xrightarrow{d} \int_0^1 W(r)^2 dr. \tag{3}$$

In the above equations, the KPSS_1 statistic is written in such a way as to make clear what the weighting with respect to n needs to be for the numerator and the denominator of the KPSS statistic. The KPSS statistic that considers data in deviations from mean is

$$\text{KPSS}_2 = \frac{n^{-1}\sum_{t=1}^{n} \bar{S}_t^2}{\sum_{t=1}^{n}(w_t - \bar{w})^2} \tag{4}$$

where $\bar{S}_t = \sum_{j=1}^{t}(w_t - \bar{w})$. For KPSS_2 we have, for I(1) w_t, under regularity conditions,

$$n^{-1}\text{KPSS}_2 \xrightarrow{d} \frac{\int_0^1 (\int_0^a \bar{W}(r)dr)^2 da}{\int_0^1 \bar{W}(r)^2 dr}, \tag{5}$$

while for I(0) w_t with arbitrary constant mean,

$$\text{KPSS}_2 \xrightarrow{d} \int_0^1 \bar{W}(r)^2 dr, \tag{6}$$

where $\bar{W}(r) = W(r) - \int_0^1 W(s)ds$.

Below, it is assumed that instead of being a stationary or an integrated process, the w_t that we observe are instead generated as $w_t = \log|x_t|$, where x_t satisfies assumptions implying that x_t is I(1). This situation would occur if the original data were I(1), but a logarithm transformation has been (incorrectly) applied to these data. Define $z_t = \log|n^{-1/2}x_t|$, implying that $w_t = z_t - (1/2)\log(n)$ and $w_t - \bar{w} = z_t - \bar{z}$. It will be assumed that

$$x_t = x_{t-1} + v_t, \tag{7}$$

where v_t is generated according to

$$v_t = \sum_{k=0}^{\infty} \phi_k \varepsilon_{t-k}, \tag{8}$$

where ε_t is assumed to be a sequence of i.i.d. random variables with mean zero, and it is assumed that $\sum_{k=0}^{\infty} \phi_k \neq 0$. In addition, we will assume that x_0 is an arbitrary random variable that is independent of all v_t. The following regularity conditions will be assumed to hold for ε_t:

Assumption 1

(a) $\sum_{k=0}^{\infty} k|\phi_k| < \infty$ *and* $E|\varepsilon_t|^p < \infty$ *for some* $p > 2$.

(b) *The distribution of* ε_t *is absolutely continuous with respect to the Lebesgue measure and has characteristic function* $\psi(s)$ *for which* $\lim_{s\to\infty} s^\eta \psi(s) = 0$ *for some* $\eta > 0$.

The following theorem is the key to determining the behavior of the KPSS statistic if a logarithm transformation has been applied to an I(1) process. In the theorem below, "$\Rightarrow$" denotes weak convergence; the theorem below provides an extension of results in de Jong (2001), where a result similar to the one below is shown, but only pointwise for $a = 1$.

Theorem 1 *Under Assumption 1,*

$$n^{-1} \sum_{t=1}^{[an]} \log |n^{-1/2} x_t| \Rightarrow \int_0^a \log |\sigma W(r)| dr \tag{9}$$

and

$$n^{-1} \sum_{t=1}^{[an]} (\log |n^{-1/2} x_t|)^2 \Rightarrow \int_0^a (\log |\sigma W(r)|)^2 dr, \tag{10}$$

where

$$\sigma^2 = \lim_{n\to\infty} E(n^{-1/2} x_n)^2 \in (0, \infty). \tag{11}$$

Note that the results of Theorem 1 are certainly not as straightforward as they may look at first sight. Because the logarithm function has a pole at 0, we cannot simply apply the continuous mapping theorem in order to arrive at even the pointwise result for $a = 1$, and the assertion that $\log |n^{-1/2} x_{[nr]}| \Rightarrow \log |\sigma W(r)|$ is not correct. To prove the above result, a separate proof is required, using results of de Jong (2001). In de Jong (2001), it is shown that

$$n^{-1} \sum_{t=1}^{n} T(n^{-1/2} x_t) \xrightarrow{d} \int_0^1 T(\sigma W(r)) dr \tag{12}$$

will typically hold as long as for all $K > 0$, $\int_{-K}^{K} |T(x)| dx < \infty$, in spite of possible poles in the function $T(.)$ and in spite of the fact that in the presence of poles in $T(.)$, the assertion $T(n^{-1/2} x_{[nr]}) \Rightarrow T(\sigma W(r))$ is not true in general. In addition, it can be shown that for functions with poles that are non-integrable (such as $T(x) = |x|^{-\phi}$ for $\phi > 1$), the result of Equation (12) is incorrect in general.

Using Theorem 1, first, the denominators of both KPSS statistics are analyzed:

Lemma 1 *Under Assumption 1,*

$$n^{-1}((\log(n))^{-2}\sum_{t=1}^{n} w_t^2 \xrightarrow{p} 1/4, \tag{13}$$

and

$$n^{-1}\sum_{t=1}^{n}(w_t - \bar{w})^2 \xrightarrow{d} \int_0^1 (\log|W(r)| - \int_0^1 \log|W(s)|ds)^2 dr. \tag{14}$$

For the KPSS_1 and KPSS_2 statistics, the following results can now be shown:

Theorem 2 *Under Assumption 1,*

$$\log(n)(n^{-1}\text{KPSS}_1 - (1/3)) \xrightarrow{d} -(4/3)\log|\sigma| + 2\int_0^1 (r^2 - 1)\log|W(r)|dr \tag{15}$$

and

$$n^{-1}\text{KPSS}_2 \xrightarrow{d} \frac{\int_0^1(\int_0^a(\log|W(r)| - \int_0^1 \log|W(s)|ds)^2 dr da}{\int_0^1(\log|W(r)| - \int_0^1 \log|W(s)|ds)^2 dr}. \tag{16}$$

The above result implies that when a logarithm transformation has been applied incorrectly to I(1) data, the asymptotic convergence behavior of both KPSS tests will be identical to the situation where w_t is I(1), in the sense that the scaling factors are the same, although a different limit distribution results. That implies that the KPSS test cannot be a suitable means for distinguishing integrated processes from integrated processes to which a logarithm transformation was applied. Our results do not settle the question of whether some other statistic could be used to distinguish consistently between an integrated process and the logarithm of an integrated process. Work on this question continues.

References

de Jong, R. M. (2001), A continuous mapping theorem-type result without continuity, mimeo, Michigan State University, available at `http://www.msu.edu/user/dejongr`.

de Jong, R. M., Amsler, C. and P. Schmidt (2001), A robust version of the KPSS test, based on indicators, mimeo, Michigan State University, available at http://www.msu.edu/user/dejongr.

Harvey, A. (2001), A unified approach to testing for stationarity and unit roots, mimeo, University of Cambridge.

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: How sure are we that economic time series have a unit root?, Journal of Econometrics 54, 159-178.

Shin, Y. and P. Schmidt (1992), The KPSS stationarity test as a unit root test, Economics Letters 38, 387-392.

Appendix

Proof of Theorem 1:

Below, only the first result of Theorem 1 will be proven, since the second can be shown analogously. From the proof of de Jong (2001), it follows that for all $a \in [0,1]$, pointwise in a,

$$Z_n(a) = n^{-1} \sum_{t=1}^{[an]} \log |n^{-1/2} x_t| \xrightarrow{d} \int_0^a \log |\sigma W(r)| dr. \tag{17}$$

This is because by Theorem 1 of de Jong (2001),

$$n^{-1} \sum_{t=1}^{n} \log |c n^{-1/2} x_t| \xrightarrow{d} \int_0^1 \log |c W(r)| dr, \tag{18}$$

and therefore for each $a \in [0,1]$,

$$n^{-1} \sum_{t=1}^{[an]} \log |n^{-1/2} x_t| = a(an)^{-1} \sum_{t=1}^{[an]} \log |a^{1/2} (an)^{-1/2} x_t| \xrightarrow{d} a \int_0^1 \log |a^{1/2} W(r)| dr, \tag{19}$$

and because $a^{1/2} W(r)$ is distributed identically to $W(ar)$, the last expression can be rewritten as

$$a \int_0^1 \log |W(ar)| dr = \int_0^a \log |W(r)| dr. \tag{20}$$

Also, note that from de Jong (2001), it follows that for some large N not depending on ε,

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} En^{-1}\sum_{t=N+1}^{n} |\log|n^{-1/2}x_t||I(|n^{-1/2}x_t|<\varepsilon)=0. \tag{21}$$

Since we have pointwise convergence by the result of Equation (17), it suffices to show stochastic equicontinuity. To show this, note that, for all $\varepsilon > 0$ and $a \geq b$ and $a - b < \varepsilon$, defining "empty summations" as zero,

$$\sup_{a,b:|a-b|<\varepsilon} |Z_n(a)-Z_n(b)| = \sup_{a,b:|a-b|<\varepsilon} |n^{-1}\sum_{t=[bn]+1}^{[an]} \log|n^{-1/2}x_t||$$

$$\leq o_P(1) + n^{-1}\sum_{t=N+1}^{n} |\log|n^{-1/2}x_t||I(|n^{-1/2}x_t|\leq\varepsilon)$$

$$+ \sup_{a,b:|a-b|<\varepsilon} n^{-1}\sum_{t=[bn]+1}^{[an]} |\log|n^{-1/2}x_t||I(\varepsilon\leq|n^{-1/2}x_t|\leq 1)$$

$$+ \sup_{a,b:|a-b|<\varepsilon} n^{-1}\sum_{t=[bn]+1}^{[an]} |\log|n^{-1/2}x_t||I(|n^{-1/2}x_t|>1), \tag{22}$$

where the first term is uniform in a and b. By the result of Equation (21), the second term satisfies

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty} En^{-1}\sum_{t=N+1}^{n} |\log|n^{-1/2}x_t||I(|n^{-1/2}x_t|\leq\varepsilon)=0. \tag{23}$$

The third term can be bounded almost surely by

$$\sup_{a,b:|a-b|<\varepsilon} n^{-1}\sum_{t=[bn]+1}^{[an]} |\log(\varepsilon)| \leq \varepsilon|\log(\varepsilon)| \tag{24}$$

which also goes to 0 as $\varepsilon \to 0$. For the last term, we have the upper bound of

$$\varepsilon \sup_{1\leq t\leq n} |\log|n^{-1/2}x_t||I(|n^{-1/2}x_t|>1), \tag{25}$$

and since

$$\sup_{1\leq t\leq n} |\log|n^{-1/2}x_t||I(|n^{-1/2}x_t|>1) = O_P(1) \tag{26}$$

by the continuous mapping theorem, the result now follows. □

Proof of Lemma 1:

Note that, since $w_t = z_t - (1/2)\log(n)$,

$$n^{-1}(\log(n))^{-2}\sum_{t=1}^{n} w_t^2$$

$$= n^{-1}(\log(n))^{-2}\sum_{t=1}^{n} z_t^2 + (1/4) - n^{-1}((\log(n))^{-2}\log(n)\sum_{t=1}^{n} z_t. \tag{27}$$

Obviously the lemma is therefore complete if we can show that the first and last term in the above equation converge to 0 in probability. To show the first result, note that by Theorem 1,

$$n^{-1}\sum_{t=1}^{n} z_t^2 \xrightarrow{d} \int_0^1 (\log|\sigma W(r)|)^2 dr \tag{28}$$

implying that this term converges to 0 in probability at rate $(\log(n))^{-2}$. Also,

$$n^{-1}\sum_{t=1}^{n} z_t \xrightarrow{d} \int_0^1 \log|\sigma W(r)| dr, \tag{29}$$

which implies that the last term in Equation (27) converges to 0 at rate $(\log(n))^{-1}$. For the second part of the lemma, simply note that by Theorem 1,

$$n^{-1}\sum_{t=1}^{n}(w_t - \bar{w})^2 = n^{-1}\sum_{t=1}^{n}(z_t - \bar{z})^2 \xrightarrow{d} \int_0^1 (\log|W(r)| - \int_0^1 \log|W(s)|ds)^2 dr. \tag{30}$$

□

Proof of Theorem 2:

First note that given the result of Lemma 1,

$$\log(n)(n^{-1}\mathrm{KPSS}_1 - 1/3) = \log(n)(\frac{n^{-3}(\log(n))^{-2}\sum_{t=1}^{n} S_t^2}{n^{-1}(\log(n))^{-2}\sum_{t=1}^{n} w_t^2} - 1/3) \tag{31}$$

will have a limit distribution identical to that of

$$4\log(n)[n^{-3}(\log(n))^{-2}\sum_{t=1}^{n} S_t^2 - (1/12)] \tag{32}$$

if we can find a limit distribution for the last statistic. Next, note that

$$S_t = \sum_{j=1}^{t} z_j - (1/2)t\log(n) \equiv S_t^Z - (1/2)t\log(n) \tag{33}$$

and therefore

$$\sum_{t=1}^{n} S_t^2 = \sum_{t=1}^{n} (S_t^Z)^2 + (1/4)(\log(n))^2 \sum_{t=1}^{n} t^2 - \log(n)\sum_{t=1}^{n} tS_t^Z. \tag{34}$$

Noting that

$$(1/4)(\log(n))^2 \sum_{t=1}^{n} t^2 = (1/12)n^3(\log(n))^2 + O(n^2(\log(n))^2), \tag{35}$$

it follows that

$$\log(n)[n^{-3}(\log(n))^{-2}\sum_{t=1}^{n} S_t^2 - (1/12)]$$

$$= o_P(1) + n^{-3}(\log(n))^{-1}\sum_{t=1}^{n}(S_t^Z)^2 - n^{-3}\sum_{t=1}^{n} tS_t^Z. \tag{36}$$

The second term on the left-hand side in the last equation converges to 0 in probability because by Theorem 1, $n^{-1}S_{[an]}^Z \Rightarrow \int_0^a \log|\sigma W(r)|dr$, implying that by the continuous mapping theorem,

$$n^{-3}\sum_{t=1}^{n}(S_t^Z)^2 = n^{-1}\sum_{t=1}^{n}(n^{-1}S_t^Z)^2 \xrightarrow{d} \int_0^1 (\int_0^a \log|\sigma W(r)|dr)^2 da. \tag{37}$$

For the last term, again by the continuous mapping theorem,

$$-n^{-3}\sum_{t=1}^{n} tS_t^Z \xrightarrow{d} -\int_0^1 a(\int_0^a \log|\sigma W(r)|dr)da = \int_0^1 (1/2)(r^2-1)\log|\sigma W(r)|dr, \tag{38}$$

and the last expression equals

$$-(1/3)\log|\sigma| + \int_0^1 (1/2)(r^2-1)\log|W(r)|dr. \tag{39}$$

After multiplication by 4 as suggested by Equation (32), we now obtain the result for KPSS_1. To show the result for KPSS_2, note that

$$n^{-1}\text{KPSS}_2 = \frac{n^{-3}\sum_{t=1}^{n}(\sum_{j=1}^{t}(w_j - \bar{w}))^2}{n^{-2}\sum_{t=1}^{n}(w_t - \bar{w})^2} = \frac{n^{-3}\sum_{t=1}^{n}(\bar{S}_t^Z)^2}{n^{-2}\sum_{t=1}^{n}(z_t - \bar{z})^2}, \tag{40}$$

and because $n^{-1}\bar{S}_{[an]}^Z \Rightarrow \int_0^a (\log|\sigma W(r)| - \int_0^1 \log|\sigma W(s)|ds)dr$ by Theorem 1, using Lemma 1 and the continuous mapping theorem, it now follows that

$$n^{-1}\text{KPSS}_2 \xrightarrow{d} \frac{\int_0^1 (\int_0^a (\log|\sigma W(r)| - \int_0^1 \log|\sigma W(s)|ds)^2 dr da}{\int_0^1 (\log|\sigma W(r)| - \int_0^1 \log|\sigma W(s)|ds)^2 dr}, \tag{41}$$

and noting that the σ cancel out, we obtain the result of the theorem for KPSS_2. $\square$