DEPARTMENT OF ECONOMICS

Working Paper

ON THE ROBUSTNESS OF UNIT ROOT TESTS
IN THE PRESENCE OF DOUBLE UNIT ROOTS

Niels Haldrup
Peter Lildholdt

Working Paper No. 2000-1
Centre for Dynamic Modelling in Economics

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

CENTRE FOR DYNAMIC MODELLING IN ECONOMICS
DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

WORKING PAPER

ON THE ROBUSTNESS OF UNIT ROOT TESTS IN THE PRESENCE OF DOUBLE UNIT ROOTS

Niels Haldrup
Peter Lildholdt

Working Paper No. 2000-1

DEPARTMENT OF ECONOMICS
SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

On the robustness of unit root tests in the presence of double unit roots

NIELS HALDRUP & PETER LILDHOLDT
DEPARTMENT OF ECONOMICS, UNIVERSITY OF AARHUS, AND
CENTRE FOR DYNAMIC MODELLING IN ECONOMICS*

28. February, 2000.

ABSTRACT. We examine some of the consequences on commonly used unit root tests when the underlying series is integrated of order two. It turns out that standard augmented Dickey-Fuller type of tests for a single unit root have excessive density in the explosive region of the distribution. The lower (stationary) tail, however, will be virtually unaffected in the presence of double unit roots. On the other hand, the Phillips-Perron test is shown to diverge to plus infinity asymptotically and thus will favor the explosive alternative. Numerical simulations are used to demonstrate the analytical results and some of the implications in finite samples.

KEYWORDS: Unit root tests, Phillips-Perron test, I(1) versus I(2).
JEL CLASSIFICATION: C12, C14, C22.

1. INTRODUCTION

It seems to be well recognized that most economic time series have properties that mimic those characterizing unit root (integrated) processes. For the majority of time series a characterization in terms of integration of order one, I(1), seems appropriate. However, some variables like prices, wages, money balances, stock-variables etc., appear to be smoother than normally observed for variables integrated of order one; such series are potentially integrated of order 2 whereby double differencing is needed to render the series stationary. The differenced series are therefore I(1); for instance, if the series are log-transformed, the growth rates will be integrated of order one. By now, there is a growing literature focusing on the complications implied by double unit roots. This literature is not only concerned with univariate testing for I(2), (Hasza and Fuller (1979), Dickey and Pantula (1987), Sen and Dickey (1987), Shin and Kim (1999), and Haldrup (1994a)), but it also focuses on the rather complex dynamic interactions occurring in I(2) cointegrated models (compare Johansen (1995,

*This paper was completed while the first author was visiting University of California, San Diego, in the spring of 2000. The Economics Department is gratefully acknowledged for its hospitality. We also want to thank Michael Jansson for helpful comments. The usual disclaimer applies.

1997), Kitamura (1995), Choi, Park, and Yu (1997), and Haldrup (1994b)). In Haldrup (1998) recent advances in the theoretical and empirical literature on I(2) are reviewed.

In the present paper, our attention is directed towards univariate testing for the order of integration, and there is mainly one particular problem we want to study. This concerns the behavior of standard univariate tests for a single unit root when, in fact, double unit roots are present. There are some problems in connection with unit root tests and the possibility of additional unit roots. The null hypothesis in conventional unit root testing typically looks like $H_0 : \alpha_1 = 1$ in the basic model $y_t = \alpha_1 y_{t-1} + v_t$ where α_1 is the autoregressive root at frequency zero and v_t is a general possibly autocorrelated process. The distribution of test statistics are typically derived under the assumption that there are no unit roots in v_t. However, it is known from the work of Dickey and Pantula (1987), among others, that the null-distribution of traditional augmented Dickey-Fuller tests will be affected in the presence of two unit roots, and hence their suggestion is to test for I(2) against I(1) rather than following the opposite procedure of testing I(1) prior to testing for I(2). This recommended sequence of testing will lead to similar tests with a size that can be controlled by choice of the significance level. Notwithstanding, in many applied papers researchers take the opposite route or examine only the level of the series, simply ignoring that I(2)-ness might be a possibility. Our paper examines in more detail analytically as well as numerically, the likely consequences of following this (reverse) route of testing. Since at least one unit root will be present when the series is either I(1) or I(2), the potential problem is that of similarity with respect to a nuisance parameter, that is, whether an additional unit root is present or absent under the null. It turns out that augmented Dickey-Fuller tests will tend to reject the null at a fraction very close to the nominal significance level when the I(1) critical values are used and the test is one-sided in the direction of the stationary alternative. However, the distribution mass is concentrated much more heavily in the upper tail compared to the Dickey-Fuller distribution. As a consequence, when testing against the explosive alternative, the Dickey-Fuller test will tend to reject a unit root too often in favor of explosiveness. The implications for the Phillips-Perron class of semi-parametric tests appear to be even more dramatic. We show that the semi-parametric test based on the Dickey-Fuller t-statistic asymptotically will have positive support and, in fact, will tend to plus infinity as the sample size grows. Hence, by allowing an explosive alternative, asymptotically the test will always reject the unit root hypothesis in favor of explosive behavior.

The paper proceeds as follows. In sections 2 and 3 the data generating mechanism is described and some summary results on the behavior of test statistics in the presence of a single unit root are provided. Next, in section 4, the properties of the test statistics are derived under the maintained assumption of double unit roots. Finally, section 5 concludes. All proofs can be found in the appendix.

2. The Data Generating Mechanism

As a starting point, consider the data generating mechanism

$$(1-\alpha_1 L)(1-\alpha_2 L)y_t = u_t \qquad t = 1, 2, ..., T \tag{1}$$

where initially the sequence $\{u_t\}$ is assumed to be i.i.d.$(0, \sigma_u^2)$. Slackening of this simplifying requirement will be made where appropriate. In particular, we may require u_t to follow the general regularity conditions of Phillips (1987), Assumption 2.1. The above autoregressive model has been analyzed under a number of different settings and in particular the presence of unit roots has attracted much attention. When $\alpha_1 = 1$ and $\mid \alpha_2 \mid < 1$, y_t is integrated of order one whilst $\alpha_1 = \alpha_2 = 1$ implies the presence of double unit roots. We want to focus on the behavior of test statistics which are designed to test for a single unit root to see how these statistics behave (under the null) by the presence of an additional unit root.

In the subsequent sections we will first summarize (for the matter of reference) some well-known properties of Dickey-Fuller and Phillips-Perron tests when a single unit root is present; next we will examine the two classes of tests in the presence of double unit roots.

3. Behavior of test-statistics when a single unit root exists.

As a benchmark, assume that $\alpha_2 = 0$ and $\alpha_1 = 1$, that is, y_t follows a random walk. For this situation a number of authors (White(1958), Fuller (1976), Dickey and Fuller (1979), and Phillips (1987)) have reported the limiting distributions of the normalized least squares estimator, $T(\widehat{\alpha}_1 - 1)$ and the $t-$statistic of $\mathrm{H}_0 : \alpha_1 = 1$ based on the regression

$$\Delta y_t = (\widehat{\alpha}_1 - 1)\, y_{t-1} + \widehat{u}_t. \tag{2}$$

The $t-$statistic for $\widehat{\alpha}_1$ is defined as $t_{\alpha_1} = (\widehat{\alpha}_1 - 1)/[s_u(\sum_{t=1}^T y_{t-1}^2)^{-1/2}]$ where $s_u^2 = T^{-1}\sum_{t=1}^T \widehat{u}_t^2$. Under the above conditions the following distribution results will apply:

$$T(\widehat{\alpha}_1 - 1) \;\Rightarrow\; \left(\int_0^1 W(r)\mathrm{d}W(r)\right)\left(\int_0^1 W(r)^2\mathrm{d}(r)\right)^{-1} \tag{3}$$

$$t_{\alpha_1} \;\Rightarrow\; \left(\int_0^1 W(r)\mathrm{d}W(r)\right)\left(\int_0^1 W(r)^2\mathrm{d}(r)\right)^{-1/2} \tag{4}$$

where $W(r)$ is a standard Brownian motion on $C[0,1]$, i.e. the space of continuous functions on the unit interval, and "$\Rightarrow$ " signifies weak convergence (in distribution). The distributions (3) and (4) are known as the Dickey-Fuller distributions.

When slackening the i.i.d. assumption about u_t, Phillips (1987) showed that under rather weak regularity conditions (see Phillips' paper for details), the relevant distributions become

$$T(\widehat{\alpha}_1 - 1) \;\Rightarrow\; \left(\int_0^1 W(r)\mathrm{d}W(r) + \lambda\right)\left(\int_0^1 W(r)^2\mathrm{d}(r)\right)^{-1} \tag{5}$$

$$t_{\alpha_1} \;\Rightarrow\; \frac{\sigma}{\sigma_u}\left(\int_0^1 W(r)\mathrm{d}W(r) + \lambda\right)\left(\int_0^1 W(r)^2\mathrm{d}(r)\right)^{-1/2} \tag{6}$$

with $\lambda = (\sigma^2 - \sigma_u^2)/2\sigma^2, \sigma_u^2 = \lim_{T\to\infty} T^{-1}E\left[\sum_{t=1}^T u_t^2\right], \sigma^2 = \lim_{T\to\infty} T^{-1}E\left[S_T^2\right]$ (the long-run variance), and $S_T = \sum_{t=1}^T u_t$. It can be easily seen that when $\sigma^2 = \sigma_u^2$, which applies for martingale difference sequences for instance, the resulting distributions are (3) and (4).

The situation with $\sigma^2 \neq \sigma_u^2$ is naturally of interest in practice because the limiting distributions will depend upon nuisance parameters. However, in the case where y_t follows an AR(p) process, estimation of a p'th order autoregression will remove the influence of the nuisance parameters such that the distribution results (3) and (4) will hold, and even in the case where MA terms are present, it is sufficient to let the order of the autoregression, k, grow with the sample size according to $k = o_p(T^{1/3})$, see Said and Dickey (1984). Hence, by this approach the nuisance parameter problem is solved in a fully parametric way.

Phillips (1987) and Phillips and Perron (1988) have suggested a semiparametric way of adjusting the above statistics[1]. In particular, Phillips (1987) has suggested the statistics Z_α and Z_t, which are constructed from the regression (2):

$$Z_\alpha = T(\widehat{\alpha}_1 - 1) - \frac{1}{2}(s^2 - s_u^2)(T^{-2}\sum_{t=1}^T y_{t-1}^2)^{-1} \tag{7}$$

$$Z_t = (s_u/s)t_{\widehat{\alpha}_1} - \frac{1}{2s}(s^2 - s_u^2)(T^{-2}\sum_{t=1}^T y_{t-1}^2)^{-1/2} \tag{8}$$

where s_u^2 and s^2 are consistent estimates of the population equivalents σ_u^2 and σ^2. The asymptotic distributions of the above statistics are given in (3) and (4).

Estimates of the long-run variance σ^2 and σ_u^2 can be obtained in various ways. It is commonplace to use the variance estimator

$$s_u^2 = \widehat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t^2 \tag{9}$$

[1]The Phillips-Perron tests generalize the Phillips Z tests by allowing deterministics in the auxiliary regressions. This will affect the limiting distributions in a very simple way since the Brownian motion processes should be replaced by appropriately detrended processes. Here we abstract from this generalization to make the exposition clear.

whereas a number of choices exists with respect to the estimator of σ^2, see e.g. Andrews (1991) and Newey and West (1994). The one we will be using here is a kernel estimator based on the sample autocovariances and can be written as

$$s^2 = \widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^2 + \frac{2}{T}\sum_{\tau=1}^{l}\omega_{\tau l}\sum_{t=\tau+1}^{T}\widehat{u}_t\widehat{u}_{t-\tau}. \tag{10}$$

In the present context we will use the Bartlett kernel which is the one Phillips (1987) used in his original paper defining the Z_α and Z_t statistics. This is given as $\omega_{\tau l} = 1-\tau/(l+1)$ where l defines the bandwidth parameter which should increase with T at an appropriate rate to ensure consistency. Of course, other kernels could be equally interesting to examine, see e.g. Andrews (1991), and Perron and Ng (1996). However, as argued by Newey and West (1994) the choice of bandwidth parameter appears to be more important than the actual choice of kernel.

4. Behavior of test-statistics when double unit roots exist. Notice that we can rearrange the equation (1) to yield

$$\Delta y_t = (\alpha_1+\alpha_2-\alpha_1\alpha_2-1)y_{t-1} + \alpha_1\alpha_2\Delta y_{t-1} + u_t. \tag{11}$$

Inspired by this representation we can focus on the augmented Dickey-Fuller regression

$$\Delta y_t = (\widehat{\alpha}-1)y_{t-1} + \widehat{\gamma}\Delta y_{t-1} + \widehat{u}_t. \tag{12}$$

Observe that both when a single and double unit roots are present, $(\alpha-1) = \alpha_1+\alpha_2-\alpha_1\alpha_2-1 = 0$. Assume now that $\alpha_1=\alpha_2=1$ and u_t is a non i.i.d. sequence which is stationary. In this situation one of the authors, Haldrup (1994a), shows that based upon the regression (12), the statistics will have the following distributions:

$$\begin{aligned} T^2(\widehat{\alpha}-1) \;&\Rightarrow\; D^{-1}\Big\{\Big(\int_0^1 \overline{W}(r)dW(r)\Big)\Big(\int_0^1 W(r)^2dr\Big) \\ &\quad -\frac{1}{2}\Big(\int_0^1 W(r)dW(r)+\lambda\Big)\overline{W}(1)^2\Big\} \end{aligned} \tag{13}$$

$$\begin{aligned} t_\alpha \;&\Rightarrow\; \frac{\sigma}{\sigma_u}D^{-1/2}\Big\{\Big(\int_0^1 \overline{W}(r)dW(r)\Big)\Big(\int_0^1 W(r)^2dr\Big)^{1/2} \\ &\quad -\frac{1}{2}\Big(\int_0^1 W(r)dW(r)+\lambda\Big)\Big(\int_0^1 W(r)^2dr\Big)^{-1/2}\overline{W}(1)^2\Big\} \end{aligned} \tag{14}$$

with $D = \left(\int_0^1 \overline{W}(r)^2dr\right)\left(\int_0^1 W(r)^2dr\right) - \frac{1}{4}\overline{W}(1)^4$. In the above expressions we use the notation $\overline{W}(r) = \int_0^r W(s)ds$. Again the influence from nuisance parameters can be removed by increasing, (at an appropriate rate in T), the number of lags of the differenced series in the auxiliary augmented Dickey-Fuller regression.

Obviously, the distribution of the usual Dickey-Fuller t_α−statistic as displayed in (14) will be different from the I(1) Dickey-Fuller distribution (6). Hence the rejection probability of a test for a single unit root, which is based on the augmented Dickey-Fuller regression (using I(1) critical values), is likely to be different from the significance level in the presence of an additional unit root. Dickey and Pantula (1987) conducted a small scale Monte Carlo experiment to study these potential problems. For a sample of 50 observations they found that the unit root null was rejected in favor of the stationary alternative in slightly *more* than 5 % of the cases at a nominal 5 % level. The finding that the rejection probability exceeds the nominal size for this one-sided alternative is, of course, a very surprising result because one would require that a test for a single unit root would clearly indicate non-stationarity rather than stationarity when indeed two unit roots are present[2]. In Table 1, we have extended Dickey and Pantula's study and examine for various sample sizes the rejection frequencies against both a stationary one-sided and an explosive one-sided alternative. The experimental design is chosen such that it corresponds to (11) and (12) with $\alpha_1 = \alpha_2 = 1$ and $u_t \sim$ n.i.d.$(0, 1)$; the statistic of interest is the Dickey-Fuller t−ratio, t_α.

Table 1 about here

The simulations demonstrate that, in fact, the size distortions[3] from using the Dickey-Fuller lower tail critical values are very minor. There is indeed a minor excessive rejection rate of the Dickey-Fuller test at the 5% level, but this is nothing seeming to be of any practical relevance. Table 1 shows that the lower tail of the augmented Dickey-Fuller test statistic is virtually the same regardless of whether one or two unit roots exist. However, the upper tail is somewhat different with a larger concentration of density leading to fairly big size distortions when testing against

[2]The apparent over rejections of the augmented Dickey-Fuller test, caused by the presence of double unit roots, made Dickey and Pantula (1987) and Pantula (1989) suggest the alternative testing procedure, which is implemented by reversing the order of testing. Hence I(2) is tested prior to testing for I(1), and in so doing a consistent and similar test with a well-known distribution under the null (the Dickey-Fuller), and hence a controllable size, can be obtained.

[3]Coining the over rejection of the null hypothesis "*size-distortion*" might appear slightly misleading given that standard unit root tests are designed to test for a single unit root and do *not* encompass double unit roots in the maintained hypothesis. Hence the proper notion is that of *robustness* with respect to an additional unit root. Since a unit root will always be present under the null, even when the series is I(2), we will nevertheless in some cases refer to the size of the tests (with the reservations just given).

the explosive alternative. This reflects the fact that in finite samples at least, I(2) processes have properties that mimic those of explosive processes. In Figure 1 the above results are visualized for a sample size of $T = 50$ observations. The density function of t_α is displayed for the case of both a single and double unit roots. As seen, the upper tail is definitely affected by the presence of double unit roots, whereas the lower tail hardly changes. Hence, as long as one tests against a one-sided stationary alternative, the risky consequences of testing I(1) prior to testing for I(2) are rather limited which is opposed to the general conception.

Figure 1 about here

As we shall now see, the Phillips-Perron class of tests appears to behave much differently compared to the augmented Dickey-Fuller tests. Observe that these statistics, (defined in (7) and (8)) are based on the regression (2) and the associated least squares coefficient and its t-ratio for a zero coefficient null. It can be shown that when $\alpha_1 = \alpha_2 = 1$, the following limiting results will apply[4]:

Theorem 1. *For the regression model $\Delta y_t = (\widehat{\alpha}_1 - 1)y_{t-1} + \widehat{u}_t$ with the data generating mechanism $\Delta^2 y_t = u_t$ where u_t satisfies the general conditions of Phillips (1987), (Assumption 2.1), then for $T \to \infty$*

$$T(\widehat{\alpha}_1 - 1) \Rightarrow \frac{1}{2}(\int_0^1 \overline{W}(r)^2 dr)^{-1}\overline{W}(1)^2 \qquad (15)$$

$$T^{-1/2}t_{\alpha_1} \Rightarrow \frac{1}{2}\left(\int_0^1 V(r)^2 dr\right)^{-1/2}(\int_0^1 \overline{W}(r)^2 dr)^{-1/2}\overline{W}(1)^2 \qquad (16)$$

where $V(r) = \left\{W(r) - \frac{1}{2}(\int_0^1 \overline{W}(r)^2 dr)^{-1}\overline{W}(1)^2\overline{W}(r)\right\}$.

As opposed to the limiting results (5) and (6), the above statistics will have no nuisance parameters appearing in the limiting distributions[5]. This is of no practical relevance, however, because it is obvious that the regression model (2) is inadequate in the present situation as the residual $\widehat{u}_t$ will be highly serially correlated; in fact, $\widehat{u}_t$ will

[4]Some of the results reported below can be partially found in the work by Dickey and Pantula (1987) and Nabeya and Perron (1994). The result associated with $T(\widehat{\alpha}_1 - 1)$ is reported in both these references whilst the limiting result concerning $T^{-1/2}t_{\alpha_1}$ is reported in Dickey and Pantula using a somewhat different notation.

[5]Note that we cannot use the result (15) as a general way of solving nuisance parameter problems. One might think that by taking the cumulative sum of an I(1) series, hence becoming I(2), one could use (15) as the relevant distribution under the null hypothesis. However, such a 'test' is inconsistent since under the alternative $T(\widehat{\alpha}_1 - 1)$ will be bounded as well, c.f. (5).

be integrated of order one and hence the various statistics will suffer from a standard spurious regressions problem, see Phillips (1986). This is why $\widehat{\alpha}_1 - 1 = O_p(T^{-1})$, rather than $O_p(T^{-2})$ as in (13), and it is also the spurious regression phenomenon that makes t_{α_1} have a non-degenerate asymptotic distribution. Nevertheless, the Phillips-Perron tests are constructed from the quantities t_{α_1}, and $T(\widehat{\alpha}_1 - 1)$.

As a benchmark, assume that σ^2 and σ_u^2 are known figures that we need not estimate; hence the adjustment of the statistics given in (7) and (8) becomes rather trivial. Because $T^{-2}\sum_{t=1}^{T} y_{t-1}^2 = O_p(T^2)$ it can be seen that asymptotically the influence from the adjustment terms will vanish. Therefore, the limiting results in (15) and (16) will apply to (7) and (8) (apart from a scaling parameter of the last distribution). Observe that since the limiting distributions have only positive support, asymptotically the test statistics will never reject the unit root null in favor of the stationary alternative. In the limit, because $t_{\widehat{\alpha}_1} \to +\infty$, a single unit root will always be rejected in favor of the explosive alternative even though a unit root is known to exist under the null[6].

Let us now examine how the Phillips-Perron tests behave if nuisance parameters are estimated according to (9) and (10). We shall consider the two cases where either l is fixed or l increases with T in a particular way. First, assuming l to be a fixed number, the following can be shown.

Theorem 2. *Under the assumptions of Theorem 1, and considering the bandwidth parameter l to be fixed, then for $T \to \infty$*

$$T^{-1}s_u^2 \Rightarrow \sigma^2 \left(\int_0^1 V(r)^2 dr \right) \tag{17}$$

$$T^{-1}s^2 \Rightarrow \sigma^2 (l+1) \left(\int_0^1 V(r)^2 dr \right) \tag{18}$$

$$Z_\alpha = T(\widehat{\alpha}_1 - 1) + o_p(1) \tag{19}$$

$$T^{-1/2} Z_t \Rightarrow \frac{1}{2(l+1)^{1/2}} \left(\int_0^1 V(r)^2 dr \right)^{-1/2} \left(\int_0^1 \overline{W}(r)^2 dr \right)^{-1/2} \overline{W}(1)^2 \tag{20}$$

Hence, according to (19) and (20), the *qualitative* results obtained, compared to when σ^2 and σ_u^2 are known, will continue to hold. However, because it is required that $l \to \infty$ at a controlled rate as $T \to \infty$ for the estimators s^2 and s_u^2 to be consistent

[6]The evidence of the series being explosive is perhaps not too surprising given that I(2) and explosive processes have somewhat similar properties, see e.g. Haldrup (1998) for a discussion on this account.

(in the I(1) case) we need to focus on the behavior of the statistics for $l \to \infty$. In the original application of the above statistics to I(1) series, consistency requires that $l = o_p(T^{1/4})$, see Phillips (1987). But in the present situation weaker requirements are needed.

Theorem 3. *Under the assumptions of Theorem 1, and requiring that $l \to \infty$ such that $\frac{l}{T} \to 0$ for $T \to \infty$, that is, $l = O_p(T^{1-\varepsilon})$ for some $\varepsilon : 0 < \varepsilon < 1$*

$$(Tl)^{-1} s^2 \Rightarrow \sigma^2 \left(\int_0^1 V(r)^2 dr \right) \tag{21}$$

$$(l/T)^{1/2} Z_t \Rightarrow \frac{1}{2} \left(\int_0^1 V(r)^2 dr \right)^{-1/2} (\int_0^1 \overline{W}(r)^2 dr)^{-1/2} \overline{W}(1)^2 \tag{22}$$

As seen, asymptotically, $(l/T)^{1/2} Z_t$ will have the same limiting distribution as $T^{-1/2} t_\alpha$ depicted in (16). In finite samples we will thus expect that, by increasing l, the entire distribution is shifted less towards the right than would otherwise be the case. The explanation behind the limiting result (22) is that in the expression for Z_t, see (8), the second term is annihilated asymptotically because $(s^2 - s_u^2)/s$ will diverge at the rate l whilst $(T^{-2} \sum_{t=1}^{T} y_{t-1}^2)^{1/2}$ will diverge at rate T, the result follows by the assumption $l/T \to 0$ for $l, T \to \infty$.

To describe finite sample distributions the descriptions above are inadequate as lower order terms can play a role, especially when l increases with the sample size without stating precisely the exact value of l to be chosen. As a reference case, consider the situation where l is fixed. In this situation the expression for $T^{-1/2} Z_t$ consists of two components with off-setting effects as can be seen from

$$T^{-1/2} Z_t = (\frac{T^{-1/2} s_u}{T^{-1/2} s})(T^{-1/2} t_{\widehat{\alpha}_1}) - \frac{1}{2T} \frac{(T^{-1} s^2 - T^{-1} s_u^2)}{T^{-1/2} s} (T^{-4} \sum_{t=1}^{T} y_{t-1}^2)^{-1/2}$$

The first term appears to be $O_p(l^{-1/2})$ and will always have positive support. The second term is $O_p(l^{1/2}/T)$ and will have negative support. The question is which term is likely to dominate when l increases for a fixed sample. This is also a question of practical relevance because there is only little guidance in the literature concerning the actual choice of the truncation parameter in finite samples[7], although theoretical rules for the rate in T follows from the asymptotics.

A small scale Monte Carlo experiment has been conducted in order to examine the quantitative implications of the above theoretical results. In Table 2 the finite sample distributions of the Z_t test for a range of sample sizes and choices of the

[7]However, consult Ng and Perron (1995) for instance for a comparison of various data dependent rules as well as rules of thumb in determining the truncation parameter l.

truncation parameter have been calculated with a data generating mechanism given by a double unit root process. The implications that follow from the analytical results are confirmed in the simulations. That is, the divergence of Z_t towards infinity with T but at a reduced rate when l increases as well. Although the distributions have some concentration of density in the negative region for increasing values of l, the concentration is only of moderate size and the practical implications are that it is very unlikely that the Phillips-Perron test will lead to acceptance of stationarity. Rather, the test will indicate explosiveness when double unit roots exist which is the asymptotic prediction.

Table 2 about here

5. Conclusion

In this paper we have examined the robustness of Dickey-Fuller and Phillips-Perron tests for a unit root. In particular, we have analyzed the implications of testing for I(1) when the series is really I(2). This is frequently seen in empirical studies where I(2)-ness is ignored as a likely alternative to the I(1) process. The results also have implications for following a route of testing where I(1) is tested prior to testing for I(2). It was found that when the underlying series is doubly integrated, it is likely to give rise to excessive rejection of the unit root null in favor of the explosive alternative because the test statistic will have a non-similar distribution caused by the extra unit root. However, the lower tail of the Dickey-Fuller test remains almost identical regardless of whether a single or two unit roots are present in the series. Therefore, as long as one tests against a one-sided stationary alternative, the risky consequences of testing I(1) prior to testing for I(2) are rather limited which is opposed to the general conception in the literature, and hence the order in which the number of unit roots should be tested appears practically irrelevant. It remains necessary, though, to consider also the possibility of an extra unit root; both when testing against one-sided and two sided alternatives. Otherwise one might get the wrong impression that the series is I(1) or explosive when in fact it is I(2).

6. Technical Appendix

Proof of results reported in section 4.

The following two lemmas will show useful throughout:

Lemma 4. *Suppose that* $\{y_t\}$ *is a random sequence generated according to (1) with* $\alpha_1 = \alpha_2 = 1$ *and with* $\{u_t\}_1^\infty$ *satisfying the regularity conditions of Phillips (1987), (Assumption 2.1), then as* $T \to \infty$

a) $T^{-3/2}y_{[Tr]} \Rightarrow \sigma \int_0^1 W(s)ds \equiv \sigma \overline{W}(r)$
b) $T^{-1/2}\Delta y_{[Tr]} \Rightarrow \sigma W(r)$
c) $T^{-4}\sum_{t=1}^{T} y_t^2 \Rightarrow \sigma^2 \int_0^1 \overline{W}(r)^2 dr$
d) $T^{-2}\sum_{t=1}^{T} \Delta y_t^2 \Rightarrow \sigma^2 \int_0^1 W(r)^2 dr$
e) $T^{-3}\sum_{t=1}^{T} y_{t-1}\Delta y_t \Rightarrow \sigma^2 \int_0^1 \overline{W}(r)W(r)dr = \frac{\sigma^2}{2}\overline{W}(1)^2$

Proof. The results (*a*), (*b*), (c), (*d*), and the first limiting result in (*e*) can be directly deduced from Lemma 2.1 of Park and Phillips (1989). The last equality sign associated with (*e*) can be proven along the following lines:

The general conditions to ensure that $T^{-1/2}S_{[Tr]} = T^{-1/2}\sum_{j=1}^{t} u_j \Rightarrow \sigma W(r)$ are assumed to hold, see Herrndorf (1984) and Phillips (1987). Note that since $y_t = y_{t-1} + \Delta y_t$, squaring and summing over T will yield

$$\begin{aligned} T^{-3}\sum_{t=1}^{T} y_{t-1}\Delta y_t &= T^{-3}\sum_{t=1}^{T}(y_t^2 - y_{t-1}y_t - \Delta y_t^2) = T^{-3}\sum_{t=1}^{T}(y_t^2 - y_{t-1}\Delta y_t + y_{t-1}^2 - \Delta y_t^2) \Leftrightarrow \\ 2T^{-3}\sum_{t=1}^{T} y_{t-1}\Delta y_t &= T^{-3}\sum_{t=1}^{T}(y_t^2 + y_{t-1}^2 - \Delta y_t^2) \Leftrightarrow \\ T^{-3}\sum_{t=1}^{T} y_{t-1}\Delta y_t &= \frac{1}{2}\left\{T^{-3}y_T^2 - T^{-3}\sum_{t=1}^{T}\Delta y_t^2\right\} = \frac{1}{2}T^{-3}y_T^2 + o_p(1) \Rightarrow \frac{\sigma}{2}\overline{W}(1)^2 \end{aligned}$$

by (*a*) and (*d*) of Lemma 4.

Lemma 5. *Under the conditions of Lemma 4*
a) $T^{-1/2}\widehat{u}_t \Rightarrow \sigma\left\{W(r) - \frac{1}{2}(\int_0^1 \overline{W}(r)^2 dr)^{-1}\overline{W}(1)^2\overline{W}(r)\right\} \equiv \sigma V(r)$
b) $T^{-1}s_u^2 = T^{-2}\sum_{t=1}^{T} \widehat{u}_t^2 \Rightarrow \sigma^2 \int_0^1 V(r)^2 dr$
c) $T^{-1}s^2 \Rightarrow \sigma^2(l+1)\left(\int_0^1 V(r)^2 dr\right)$ *for a fixed value of* l
d) $(Tl)^{-1}s^2 \Rightarrow \sigma^2\left(\int_0^1 V(r)^2 dr\right)$ *for* l=$O_p(T^{1-\varepsilon}), 0 < \varepsilon < 1.$

Proof. First we need to find the limiting distribution of $T(\widehat{\alpha}_1 - 1)$. It follows straightforwardly from (*c*) and (*e*) of Lemma 4, that

$$\begin{aligned} T(\widehat{\alpha}_1 - 1) &= (T^{-4}\sum_{t=1}^{T} y_t^2)^{-1}(T^{-3}\sum_{t=1}^{T} y_{t-1}\Delta y_t) \\ &\Rightarrow \frac{1}{2}(\int_0^1 \overline{W}(r)^2 dr)^{-1}\overline{W}(1)^2 \end{aligned} \tag{A1}$$

Now, turning to (a), the least squares residuals from (2) are given as $\widehat{u}_t = \Delta y_t - (\widehat{\alpha}_1 - 1)y_{t-1}$ which by appropriate scaling yields $T^{-1/2}\widehat{u}_t = T^{-1/2}\Delta y_t - T(\widehat{\alpha}_1 -$

$1)T^{-3/2}y_{t-1}$. The required result follows from the sub-results (a) and (b) of Lemma 4, and (A1).

Also, (b) follows immediately from (a).

Result (c) can be shown along the following lines:

$$\begin{aligned} T^{-1}s^2 &= T^{-1}s_u^2 + 2T^{-2}\sum_{\tau=1}^{l}\left(\left(1-\frac{\tau}{l+1}\right)\sum_{t=\tau+1}^{T}\hat{u}_t\hat{u}_{t-\tau}\right) \\ &= T^{-1}s_u^2 + 2T^{-2}\sum_{\tau=1}^{l}\left(1-\frac{\tau}{l+1}\right)\sum_{t=\tau+1}^{T}\left(\sum_{i=1}^{\tau}\Delta\widehat{u}_{t-i+1}\hat{u}_{t-\tau}\right) \\ &\quad +2T^{-2}\sum_{\tau=1}^{l}\left(1-\frac{\tau}{l+1}\right)\sum_{t=\tau+1}^{T}\hat{u}_{t-\tau}^2 \qquad \text{(A2)} \end{aligned}$$

where we have exploited that $\hat{u}_t = \sum_{i=1}^{\tau}\triangle\hat{u}_{t-i+1} + \hat{u}_{t-\tau}$. Now, since

$$\sum_{t=\tau+1}^{T}\sum_{i=1}^{\tau}\triangle\hat{u}_{t-i+1}\hat{u}_{t-\tau} = O_p(T) \qquad \text{(A3)}$$

it is seen that for fixed bandwidth l, the second term in (A2) is going to vanish asymptotically. Hence we have

$$\begin{aligned} T^{-1}s^2 &= T^{-1}s_u^2 + 2\sum_{\tau=1}^{l}\left(1-\frac{\tau}{l+1}\right)\left(T^{-2}\sum_{t=\tau+1}^{T}\hat{u}_{t-\tau}^2\right) + o_p(1) \\ &\Rightarrow \sigma^2\int_0^1 V(r)^2dr + 2\frac{l}{2}\sigma^2\int_0^1 V(r)^2dr \\ &= (l+1)\sigma^2\int_0^1 V(r)^2dr \end{aligned}$$

This proves (c).

Turning to (d), the proofs where l increases with T can be shown along the following lines. We note that because $\sum_{\tau=1}^{l}\left(1-\frac{\tau}{l+1}\right) = O(l)$ and $\frac{l}{T} = o(1)$, the second term in (A2) is $O_p(\frac{l}{T})$. Thus

$$\begin{aligned} (Tl)^{-1}s^2 &= (Tl)^{-1}s_u^2 + 2l^{-1}T^{-2}\sum_{\tau=1}^{l}\left(1-\frac{\tau}{l+1}\right)\sum_{t=\tau+1}^{T}\left(\sum_{i=1}^{\tau}\Delta\widehat{u}_{t-i+1}\hat{u}_{t-\tau}\right) \\ &\quad +2l^{-1}T^{-2}\sum_{\tau=1}^{l}\left(1-\frac{\tau}{l+1}\right)\sum_{t=\tau+1}^{T}\hat{u}_{t-\tau}^2 \\ &= O_p(l^{-1}) + O_p(T^{-1}) + \frac{2}{l}\sum_{\tau=1}^{l}\left(1-\frac{\tau}{l+1}\right)\left(T^{-2}\sum_{t=\tau+1}^{T}\hat{u}_{t-\tau}^2\right) \end{aligned}$$

$$\Rightarrow \sigma^2 \int_0^1 V(r)^2 dr, \qquad \text{for } l, T \to \infty \text{ and } l = O_p(T^{1-\varepsilon}), 0 < \varepsilon < 1$$

Proof of Theorem 1.

The result (15) has already been shown in (A1). The limit (16) can be seen by appropriate scaling of the $t-$statistic defined as $t_{\alpha_1} = (\widehat{\alpha}_1 - 1)/[s_u(\sum_{t=1}^T y_{t-1}^2)^{-1/2}]$, i.e. $T^{-1/2} t_{\alpha_1} = T(\widehat{\alpha}_1 - 1)/[T^{-1/2} s_u (T^{-4} \sum_{t=1}^T y_{t-1}^2)^{-1/2}]$. The result then follows from (c) of Lemma 4, (b) of Lemma 5, and (A1).

Proof of Theorem 2.

The results (17) and (18) of Theorem 2 are already shown in Lemma 5. (19) follows as

$$\begin{aligned} Z_\alpha &= T(\widehat{\alpha}_1 - 1) - \frac{1}{2}(T^{-1}s^2 - T^{-1}s_u^2)(T^{-4}\sum_{t=1}^T y_{t-1}^2)^{-1} \cdot T^{-1} \\ &= T(\widehat{\alpha}_1 - 1) + o_p(1) \end{aligned}$$

by use of Lemma 4 and Lemma 5, whereas (20) is given as

$$\begin{aligned} T^{-1/2} Z_t &= (\frac{T^{-1/2}s_u}{T^{-1/2}s})(T^{-1/2}t_{\widehat{\alpha}_1}) - \frac{1}{2T}\frac{(T^{-1}s^2 - T^{-1}s_u^2)}{T^{-1/2}s}(T^{-4}\sum_{t=1}^T y_{t-1}^2)^{-1/2} \\ &= (\frac{T^{-1/2}s_u}{T^{-1/2}s})(T^{-1/2}t_{\widehat{\alpha}_1}) + o_p(1) \\ &\Rightarrow \frac{1}{2(l+1)^{1/2}}\left(\int_0^1 V(r)^2 dr\right)^{-1/2} (\int_0^1 \overline{W}(r)^2 dr)^{-1/2} \overline{W}(1)^2 \end{aligned}$$

Proof of Theorem 3

Appropriate normalization of (7) yields

$$(l/T)^{1/2} Z_t = \left(\frac{T^{-1/2}s_u}{(Tl)^{-1/2}\, s}\right) T^{-1/2} t_{\widehat{\alpha}_1} - \frac{1}{2\,(Tl)^{-1/2}\, s}\left(T^{-1}s^2 - T^{-1}s_u^2\right)(T^{-2}\sum_{t=1}^T y_{t-1}^2)^{-1/2}$$

From this expression it follows from Lemma 4 and Lemma 5 that

$$\begin{aligned} \frac{1}{2\,(Tl)^{-1/2}\, s}\left(T^{-1}s^2 - T^{-1}s_u^2\right) &= O_p(l) \\ (T^{-2}\sum_{t=1}^T y_{t-1}^2)^{1/2} &= O_p(T) \end{aligned}$$

Thus, given the assumption that l grows at a slower rate than T, see also Perron and Ng (1994), section 4,

$$\begin{aligned}(l/T)^{1/2}Z_t &= \left(\frac{T^{-1/2}s_u}{(Tl)^{-1/2}s}\right)T^{-1/2}t_{\widehat{\alpha}_1}+o_p(1)\\ &\Rightarrow \frac{1}{2}\left(\int_0^1 V(r)^2dr\right)^{-1/2}(\int_0^1 \overline{W}(r)^2dr)^{-1/2}\overline{W}(1)^2\end{aligned}$$

REFERENCES

[1] Andrews, D.W.K., (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59,** 817-858.

[2] Choi, I. J. Y. Park and B. Yu (1997). Canonical cointegration regression and testing for cointegration in the presence of I(1) and I(2) variables. *Econometric Theory* **13,** 850-876

[3] Dickey, D. A., and W. A. Fuller, (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association* **74**, 427-431.

[4] Dickey, D. A. and S. G. Pantula, (1987). Determining the order of differencing in autoregressive processes. *Journal of Business and Economic Statistics* **5**, 455-461.

[5] Fuller, W. A, (1976). Introduction to statistical time series, New York, John Wiley & Sons.

[6] Haldrup, N., (1994a). Semiparametric tests for double unit roots. *Journal of Business and Economic Statistics* **12**, 109-122.

[7] Haldrup, N., (1994b). The asymptotics of single-equation cointegration regressions with I(1) and I(2) variables. *Journal of Econometrics,* **63**, 153-81.

[8] Haldrup, N. (1998). An econometric analysis of I(2) variables. *Journal of Economic Surveys* **12**, 595-650.

[9] Hasza, D. P. and W. A. Fuller, (1979). Estimation of autoregressive processes with unit roots. *The Annals of Statistics* **7**, 1106-1120.

[10] Herrndorf, N., (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Annals of Probability* **12**, 141-53.

[11] Johansen, S., (1995) A statistical analysis of cointegration for I(2) variables. *Econometric Theory* **11,** 25-29.

[12] Johansen, S., (1997) A likelihood analysis of the I(2) model. *Scandinavian Journal of Statistics,* **24,** 433-462.

[13] Kitamura, Y., (1995) Estimation of cointegrated systems with I(2) processes. *Econometric Theory,* **11,** 1-24.

[14] Nabeya, S., and P. Perron, (1994). Local asymptotic distributions related to the AR(1) model with dependent errors. *Journal of Econometrics* **62**, 229-264.

[15] Newey, W.K., and K.D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies* **61,** 631-653.

[16] Ng, S., and P. Perron, (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* **90,** 268-281.

[17] Pantula, S. G., (1989). Testing for unit roots in time series data. *Econometric Theory* **5**, 256-271.

[18] Park, J. Y., and P. C. B. Phillips, (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory* **5**, 95-131.

[19] Perron, P., and S. Ng, (1996). Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* **63**, 435-463.

[20] Phillips, P. C. B., (1987). Time series regression with unit root. *Econometrica* **55**, 277-302.

[21] Phillips, P. C. B. and P. Perron, (1988). Testing for a unit root in time series regression. *Biometrika* **75**, 335-346.

[22] Said, E.S., and D.A. Dickey, (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599-607.

[23] Sen, D. L. and D. A. Dickey, (1987). Symmetric test for second differencing in univariate time series. *Journal of Business and Economic Statistics* **5**, 463-473.

[24] Shin, D. W., and H. J. Kim, (1999). Semiparametric tests for double unit roots based on symmetric estimators. *Journal of Business and Economic Statistics* **7**, 67-73.

[25] White, J.S., (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics* **29**, 1188-1197.

7. Tables and Figures

Table 1. Rejection frequencies of augmented Dickey-Fuller $t-$test for a single unit root when two unit roots exist. The tests are against one-sided stationary and explosive alternatives at 1, 5, and 10 % levels and for sample sizes $T = 25, 50, 100, 250$, and 500.

	Stationary alternative			Explosive alternative		
T	1%	5%	10%	1%	5%	10%
25	.017	.055	.090	.094	.289	.411
50	.015	.058	.099	.077	.231	.339
100	.013	.056	.102	.059	.182	.278
250	.012	.054	.100	.044	.144	.234
500	.012	.053	.099	.038	.135	.223

Note: The simulations are based on 10,000 Monte Carlo replications.

Table 2. Empirical fractiles for the Z_t test when the underlying process is generated according to $\Delta^2 y_t = \varepsilon_t$, with $\varepsilon_t \sim$n.i.d.$(0, 1)$, $t = 1, 2, .., T$.

		Empirical fractiles								
T	l	1%	2.5%	5%	10%	50%	90%	95%	97.5%	99%
50	0	-.39	-.28	-.10	.65	8.52	18.04	21.01	23.95	26.91
50	4	-.96	-.74	-.56	-.15	4.03	9.01	10.67	12.35	14.34
50	8	-1.15	-.95	-.75	-.42	3.18	7.41	8.95	10.48	12.29
50	12	-1.20	-1.04	-.87	-.56	2.83	6.83	8.42	10.00	11.91
100	0	-.24	-.14	.17	1.19	12.57	26.64	30.79	34.68	38.66
100	4	-.69	-.49	-.33	.17	5.75	12.62	14.59	16.63	18.78
100	8	-.92	-.69	-.53	-.11	4.39	9.95	11.42	13.24	15.03
100	12	-1.06	-.84	-.64	-.29	3.76	8.63	10.08	11.79	13.40
500	0	-.07	.07	.71	2.86	28.56	59.98	69.00	77.17	87.14
500	4	-.25	-.16	.14	1.11	12.83	27.02	31.32	34.99	39.60
500	8	-.37	-.25	-.04	.69	9.60	20.30	23.59	26.46	29.80
500	12	-.47	-.32	-.15	.47	8.04	17.03	19.84	22.34	25.16

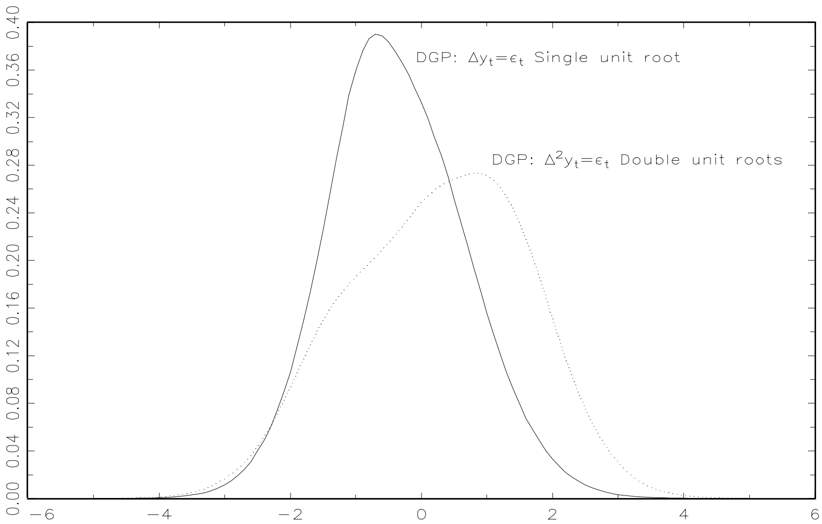

Figure 1: Density functions of t_α in the presence of a single and double unit roots for $T = 50$. The plots are drawn from 250.000 Monte Carlo replications and a normal density kernel estimator is used for smoothing.

Working Paper

1999-17	Eric Toulemonde: Deterring Entry through High Wages.
1999-18	Pierre M. Picard and Eric Toulemonde: On the Equivalence of Taxes Paid by Employers and Employees.
1999-19	Eric Toulemonde: Wage Bargaining: Reconciling Theory and Evidence.
1999-20	N.E. Savin and Allan H. Würtz: Empirically Relevant Power Comparisons for Limited Dependent Variable Models.
1999-21	Martin Paldam: Corruption and Religion. Adding to the Economic Model?
1999-22	Henrik Christoffersen, Martin Paldam and Allan Würtz: Public Versus Private Production. A Study of the Cost of School Cleaning in Denmark.
1999-23	Svend Jespersen: Economic Development without Fisher Separation: "Trickle-up" or "Trickle-down"?
1999-24	Mette Yde Skaksen and Jan Rose Sørensen: Should Trade Unions Appreciate Foreign Direct Investments?
1999-25	Palle Andersen: A Note on Alternative Measures of Real Bond Rates.
1999-26	Torben M. Andersen, Niels Haldrup and Jan Rose Sørensen: Product Market Integration and European Labour Markets.
1999-27	Licun Xue: Negotiation-Proof Nash Equilibrium.
1999-28	Torben M. Andersen and Niels C. Beier: Noisy Financial Signals and Persistent Effects of Nominal Shocks in Open Economies.
1999-29	Michael Jansson: Consistent Covariance Matrix Estimation for Linear Processes.
2000-1	Niels Haldrup and Peter Lildholdt: On the Robustness of Unit Root Tests in the Presence of Double Unit Roots.

CENTRE FOR DYNAMIC MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Working papers, issued by the Centre for Dynamic Modelling in Economics:

1999-11	Martin Paldam: The Big Pattern of Corruption. Economics, Culture and the Seesaw Dynamics.
1999-21	Martin Paldam: Corruption and Religion. Adding to the Economic Model?
2000-1	Niels Haldrup and Peter Lildholdt: On the Robustness of Unit Root Tests in the Presence of Double Unit Roots.