Unit root tests in the presence of uncertainty about the non-stochastic trend

Leila Ayat, Peter Burridge*

Department of Economics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received 1 November 1996; received in revised form 1 December 1998; accepted 1 April 1999

Abstract

A sequential procedure for determination of trend degree and testing for a unit root is introduced; its properties are investigated by Monte Carlo experiment. We implement the pseudo-GLS unit root tests of Elliott et al. (1996. Econometrica 64(4), 813–836), with lag length selected by the BIC criterion. Our procedure allows for quadratic trend, and we introduce a ‘GLS’-type test for this case. We compare the sequential procedure, in which trend degree is tested after a unit root pre-test, with a robust trend test recently developed by Vogelsang (1998. Econometrica 66(1), 123–149). The sequential procedure is advocated in preference both to informal use of the usual family of unit root tests and to alternative formal sequential methods that have been advanced in the literature. It is illustrated by application to the inventory data analysed in Hall (1994. Journal of Business and Economic Statistics 12(4), 461–470). © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C22

Keywords: Unit root tests; Trend degree; Sequential strategies

1. Introduction

In practice, testing for a unit root takes place in the presence of uncertainty about the appropriate degree of any deterministic trend that the data may

*Corresponding author. Tel.: + 44-121-414-6640, fax: + 44-121-414-7377.
E-mail address: burridgp@css.bham.ac.uk (P. Burridge)

0304-4076/00/$- see front matter © 2000 Elsevier Science S.A. All rights reserved.
PII: S 0 3 0 4 - 4 0 7 6 ( 9 9 ) 0 0 0 3 0 - 5
Consequently, unit root tests are often conducted after some kind of pre-test for the trend; such pre-tests may be very informal, such as inspection of time plots of the data, or may be implemented by testing the significance of the coefficient on the time trend in an equation fitted to the data. Whether such pre-tests are employed or not, it is very common for the various test statistics not to agree, as reported, for example, by Hall (1994) who applied the ADF $\tau_\mu$ and $\tau_\iota$ tests (see below) to 10 inventory series, finding that the tests were in agreement in roughly half the series.\(^1\) Commenting on the disagreements, Hall (1994, pp. 467–468) observes,

> Clearly there can be disagreement between the results of the tests due to the inherent variation in sampling. There are two alternative explanations that need to be considered, however. First, if $\tau_\mu$ is insignificant but $\tau_\iota$ is significant, then it may be due to misspecification of the trend term. West (1987) demonstrated that if $y_t$ is stationary about a linear time trend but the trend is omitted from the regression model, then $\tau_\mu$ converges in probability to 0, so asymptotically one never rejects a unit root. Therefore I interpret this type of conflicting result as evidence against a unit root. Similarly, if $\tau_\mu$ is significant but $\tau_\iota$ is insignificant, then this may be due to the inclusion of a redundant regressor. Dickey (1984) demonstrated that if $y_t$ is stationary about an intercept alone, then the inclusion of a linear time trend … leads to a considerable loss of power. Therefore I also interpret this type of conflict between the tests as evidence against a unit root.

The interpretive issue raised in the above comment arises in all applications of unit root tests, and the procedure we propose is designed to provide a systematic resolution of the problem via evaluation of the significance of the trend.\(^2\) This idea is not completely new, of course; the first proposal for a sequential procedure we can find appears in Perron (1988), and Dolado et al. (1990) advocated the sequential use of the Dickey–Fuller unit root tests and tests for the presence of a trend. So far as we are aware, however, there has been no subsequent study of the properties of such sequential tests, in spite of the no doubt very wide use of informal strategies of this type in empirical work. Given the attention devoted in the literature to modifying Dickey–Fuller tests (and their variants) in order to control test size, the failure to examine the properties of the sequential decision rules within which they are generally employed is a serious omission. The present paper begins to fill this gap.

In some settings, such as the estimation of average growth rates, the non-stochastic trend function may itself be the focus of interest, and we might hope to

---

\(^1\)Hall’s study compares lag length selection via information criteria, with selection by general to specific and specific to general testing, and he finds that unit root test outcomes are in some cases sensitive to the method used. See Section 5 for further discussion.

\(^2\)An alternative Bayesian approach has been investigated by Phillips and Ploberger (1994).
be able to draw inferences about the trend which are robust to the autocorrelation structure of the stochastic component. Vogelsang (1998) develops test statistics with size which is controllable irrespective of the stationarity or non-stationarity of the shocks, while Canjels and Watson (1997) explore the efficiency of various trend estimators in the same setting. While these latter papers focus on the construction of confidence intervals for trend estimates, we are primarily concerned with determining trend degree, so that size and power are both important.

Section 2 sets out the data-generation process assumed, and defines the unit root tests used. The sequential procedures are defined in Section 3. These procedures could be implemented using any of the many forms of unit root test statistics that are now available, but in view of their good power, we illustrate their performance using the ‘GLS’ forms of the ADF test introduced by Elliott et al. (1996) extended here to allow for a quadratic polynomial trend. Trend degree is tested either by conventional $t$-tests (with their standard critical values) applied to levels or to differences of the data depending on the outcome of a unit root pre-test, or using a robust test due to Vogelsang (1998) for which no pre-test is required. The quadratic trend is included as a default class: if the tests detect such a trend, then the series should not be modelled with a linear trend alone.$^3$

A feature of the fully sequential procedure we study is that tests for the degree of the trend are conducted after a unit root pre-test, with the highest trend degree maintained, has been used to determine whether or not to difference the data. This is only one possibility, however; in Section 3 we reject on grounds of redundancy the use of joint $F$-type tests for unit roots and trend, and in the Monte Carlo experiment described in Section 4 we illustrate the effect of replacing the ‘$t$’ tests for trend degree with a robust test. Section 5 applies the procedures to the inventory data studied by Hall (1994), and concludes.

2. Data-generation process, test definitions and invariance properties

2.1. The DGP

We write the data, $y_t$, as the sum of a deterministic (polynomial trend) component, $d_t$, and a purely stochastic component, $u_t$:

$$
y_t = d_t + u_t \quad (t = 1, 2, \ldots, T),
$$

$$
u_t = \rho u_{t-1} + v_t \quad (t = 2, 3, \ldots, T),
$$

$$
d_t = \sum_{i=0}^{k} \beta_i t^i, \quad v_t \sim \text{stationary.} \quad (1)
$$

$^3$As noted by a referee, the test for the presence of a quadratic trend also has power against some forms of trend breaks, and the associated unit root test may be less adversely affected by such breaks than existing tests. However, further analysis of this issue is beyond the scope of the present paper.
The model selection and hypothesis testing problem is to determine the degree of \( d_t \), and to test \( \rho = 1 \), against trend-stationary alternatives, \(|\rho| < 1\). In the leading case, \( v_t \) is IID(0, \( \sigma^2 \)); departures from this may be handled in a variety of ways, but following Hall (1994), we have used the ADF approach with lag length selected via an information criterion, as this is simple to implement and performs well in most cases – see Stock (1994) for a recent survey. Whatever method of individual test size-correction is employed, the contribution to overall size from the sequential procedure is likely to be similar, and this is our main concern. The initial observation is treated as random, and thus lies off the trend, in general.

2.2. The Dickey–Fuller family of unit root tests

The currently most widely employed tests for a unit root, the so-called ‘augmented’ versions of those developed by Dickey and Fuller (1979, 1981), are based on the \( t \)-statistic for \( \rho = 1 \) in the OLS regressions:

\[
\Delta y_t = (\hat{\rho}_a - 1)y_{t-1} + \sum_{j=1}^{p} a_j \Delta y_{t-j} + e_{at},
\]

(2a)

\[
\Delta y_t = (\hat{\rho}_b - 1)y_{t-1} + \hat{\kappa}_{00} + \sum_{j=1}^{p} a_j \Delta y_{t-j} + e_{bt},
\]

(2b)

or

\[
\Delta y_t = (\hat{\rho}_c - 1)y_{t-1} + \hat{\kappa}_{01} + \hat{\kappa}_{11}t + \sum_{j=1}^{p} a_j \Delta y_{t-j} + e_{ct}.
\]

(2c)

To these widely used tests we may add the test introduced by Ouliaris et al. (1989), based on

\[
\Delta y_t = (\hat{\rho}_d - 1)y_{t-1} + \hat{\kappa}_{02} + \hat{\kappa}_{12}t + \hat{\kappa}_{22}t^2 + \sum_{j=1}^{p} a_j \Delta y_{t-j} + e_{dt}.
\]

(2d)

The test equations, being augmented with \( p \) lags of \( \Delta y_t \) on the right-hand side, thus approximate \( v_t \) by a stationary AR(\( p \)). There are a number of equivalent ways of calculating the test regressions; for ease of comparison with the ‘GLS’ tests discussed below and implemented in our experiments, we can think of partitioning the regression into a first stage removal of the trend by OLS, followed by OLS estimation of (2a) applied to the detrended series (augmented by \( p \) lagged differences).
The DF–GLS tests of Elliott et al. (1996) (ERS), differ from (2b)–(2d) in that the trend is estimated by pseudo-generalised least squares. The GLS test statistics are thus defined as the ‘t’ statistic on the coefficient of $y_t^*$ in the OLS regression

$$\Delta y_t^* = (\hat{\rho}^* - 1)y_{t-1}^* + \sum_{j=1}^{p} a_j \Delta y_{t-j}^* + \text{error},$$  \hspace{1cm} (3)$$

in which $y_t^* = y_t$ (no detrending, corresponding to (2a)), or

$$y_t^* = y_t - \hat{\beta}_{00,\text{GLS}}$$  \hspace{1cm} (4)$$
(de-meaned only, corresponding to (2b)), or

$$y_t^* = y_t - \hat{\beta}_{01,\text{GLS}} - \hat{\beta}_{11,\text{GLST}}$$  \hspace{1cm} (5)$$
(de-meaned and de-trended, corresponding to (2c)), or

$$y_t^* = y_t - \hat{\beta}_{02,\text{GLS}} - \hat{\beta}_{12,\text{GLST}} - \hat{\beta}_{22,\text{GLST}^2}$$  \hspace{1cm} (6)$$
(de-meaned and quadratic de-trended, corresponding to (2d)).

The ‘GLS’ regressions differ in the three cases, since in each case the quasi-differencing operator is chosen by setting the test asymptotic power function (against a sequence of alternatives, $\rho = 1 + c/T$) tangent to its power envelope at a power of 50% when size is set at 5%. ERS obtained $\tilde{c}_0 = -7.0$ (de-meaned) and $\tilde{c}_1 = -13.5$ (de-meaned and de-trended), but did not investigate higher degree tests.

To obtain $\tilde{c}$ for the quadratic trend test we followed ERS and set size at 5%, and sought that value of $\tilde{c}$ which would yield asymptotic power of 50%. To do this, we took $T = 1000$ and estimated the power at various $c$ values using 20,000 Monte Carlo replications. The value of $\tilde{c}$ obtained was $-18.5$; we then checked the power for this value of $c$ at $T = 500$ (i.e. $\rho = 0.963$) and found it to be very close to 50%, thus confirming that $\tilde{c}_2 = -18.5$ is appropriate, since the power is unchanging as $T$ increases from 500 to 1000. Critical values for this new test are given in Appendix B.

Writing $\tilde{\rho}_j = 1 + \tilde{c}_j/T$, ($j = 0, 1, 2$) we can define the $\hat{\beta}_{ij,\text{GLS}}$ as follows. $\hat{\beta}_{00,\text{GLS}}$ is the OLS regression coefficient obtained by regressing the vector,

$$[y_1, y_2 - \tilde{\rho}_0 y_1, \ldots, y_T - \tilde{\rho}_0 y_{T-1}],$$
on the vector,
\[
[1, 1 - \bar{\rho}_0, \ldots, 1 - \bar{\rho}_0]',
\]
similarly, \([\hat{\beta}_{01}, \hat{\beta}_{11}]_{\text{GLS}}\) results from the OLS regression of the vector,
\[
[y_1, y_2 - \bar{\rho}_1 y_1, \ldots, y_T - \bar{\rho}_1 y_{T-1}]',
\]
on the matrix,
\[
\begin{bmatrix}
1, & 1 - \bar{\rho}_1, \ldots, 1 - \bar{\rho}_1 \\
1, & 2 - \bar{\rho}_1, \ldots, T - (T - 1)\bar{\rho}_1
\end{bmatrix}
\]
and finally, \([\hat{\beta}_{02}, \hat{\beta}_{12}, \hat{\beta}_{22}]_{\text{GLS}}\) results from the OLS regression of the vector,
\[
[y_1, y_2 - \bar{\rho}_2 y_1, \ldots, y_T - \bar{\rho}_2 y_{T-1}]',
\]
on the matrix,
\[
\begin{bmatrix}
1, & 1 - \bar{\rho}_2, \ldots, 1 - \bar{\rho}_2 \\
1, & 2 - \bar{\rho}_2, \ldots, T - (T - 1)\bar{\rho}_2 \\
1, & 4 - \bar{\rho}_2, \ldots, T^2 - (T - 1)^2\bar{\rho}_2
\end{bmatrix}
\]
ERS find that in practical sample sizes, the GLS-detrended unit root tests enjoy a power advantage over their OLS-detrended counterparts. The asymptotic efficiency gain in trend estimation is substantial when the data are near-integrated, and has been quantified by Lee and Phillips (1994). In finite samples, the efficiency of the trend estimates may be calculated exactly when \(v_t\) in (1) is a stationary ARMA process, and the pseudo-GLS trend estimates are not, in general, more efficient than OLS (see Canjels and Watson, 1997); however, the power advantage of the ERS tests remains because the null distribution of the tests is shifted to the right relative to the corresponding Dickey–Fuller distributions (some evidence on this point is contained in Burridge and Taylor, 1999).

As is well known, the three regressions (2a)–(2c) give unit root test statistics, \(\tau\), \(\tau_{\mu}\) and \(\tau_{\nu}\), respectively, with different invariance properties and limit distributions. The same is true of the GLS tests. Because the sequential procedure we examine relies on these invariance properties, we collect these with proof and discussion, in Appendix A.
3. Sequential testing strategies

Since we are assuming that the degree of any polynomial trend that may be present in the data is unknown, the objective of the testing strategy should be to identify the class of model, that is, to test the unit root and determine the trend degree. To this end we consider a sequence of pre-tests.

If the data are generated by (1), with $\rho = \rho_0$, known, $v_t \sim \text{IID}(0, \sigma^2)$, but the trend parameters unknown, then fitting equation (2b), (2c), or (2d) without using our knowledge of $\rho$ produces inefficient trend estimates in finite samples; efficient estimation, by GLS, requires the dependent variable to be transformed to $(1 - \rho L)y_t = y_t^\dagger$, say. If $\rho = 1$ this leads to

$$y_t^\dagger = \Delta y_t = (\beta_1 - \beta_2) + 2\beta_2 t + v_t$$

if $k = 2$, or

$$y_t^\dagger = \Delta y_t = \beta_1 + v_t$$

if $k = 1$; in either case, the non-stochastic component may be efficiently estimated and its significance tested by standard $t$-tests. Furthermore, as shown by Dickey (1984) and West (1987), unit root tests lose power if the trend specified is of higher degree than necessary (via inefficient estimation of $\rho$), and are inconsistent (biased in favour of the null) if the trend fitted is of lower degree than is present in the DGP. To minimise such power losses, we seek reliable inference about the trend in the presence of uncertainty about $\rho$. At the same time, we want reliable inference about $\rho$ in the presence of uncertainty about the trend. The invariance results set out in Appendix A combined with more efficient trend estimation using (7) or (8) if unit root pretests applied to (2d) or (2c) fail to reject, suggest a strategy which we now describe in detail.

The most general maintained model allows for a quadratic trend:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + u_t.$$  

With (9) maintained, we can obtain a unit root test statistic invariant to $\beta$ from regression, (2d) or the ‘GLS’ variant, and this test forms the first step of our sequence. At every step of the sequence the lag order, $p$, must be chosen; in our experiments, this was done by minimising the Schwarz information criterion (BIC), $\ln(\hat{\sigma}^2) + p\ln(T)/T$, with $3 \leq p \leq 8$ (the lower bound being used for

---

$^4$ As pointed out by a referee, the OLS trend estimates in regressions (7) and (8) are asymptotically equivalent to GLS even if $v_t$ is serially correlated; see Grenander and Rosenblatt (1957, Section 7.3).
comparability with the results of ERS, who impose this). We comment further on this after setting out the procedure; the complete algorithm is:

3.1. Strategy S1

1. Perform a preliminary unit root test invariant to quadratic trend under the null.
   2(a). If the unit root is not rejected at step 1, provisionally maintain this hypothesis and estimate \( \Delta y_t = \beta_{01}^* + \beta_{11}^* t + \sum_{j=1}^T a_j \Delta y_{t-j} + \epsilon_t \), testing for the null that \( k = 1 \), (that is, \( \beta_2 = 0 \), in (9)) using the \( t \)-statistic on \( \beta_{11}^* \) referred to standard tables.
   2(b). If the unit root is rejected at step 1, test for \( k = 1 \) using the \( t \)-statistic on \( \hat{a}_{11} \) in Eq. (2d), again referred to standard tables.
3(a). If \( k = 1 \) was rejected at step 2, we stop, since the unit root test already conducted is the only one available which is invariant to the maintained quadratic trend.
3(b). If \( k = 1 \) was not rejected, and the unit root was not rejected, perform a second provisional unit root test invariant to linear trend under the null.
3(c). If \( k = 1 \) was not rejected, but the unit root was rejected at step 2, test for \( k = 0 \) using the \( t \)-statistic on \( \hat{a}_{11} \) in (2c) referred to standard tables and stop.
4. If the unit root was not rejected at 3(b), estimate \( \Delta y_t = \beta_{00}^* + \sum_{j=1}^T a_j \Delta y_{t-j} + \epsilon_t \), testing the null that \( k = 0 \) using the \( t \)-statistic on \( \beta_{00}^* \).
5(a). If \( k = 0 \) is rejected at step 4, stop.
5(b). If \( k = 0 \) is accepted at step 4, conduct a further provisional unit root test invariant to the mean under the null.
6(a). If the unit root is not rejected at 5(b), test the magnitude of the initial observation, \( y_1 \), relative to the increments in \( y \) using \( y_1/\sqrt{(T^{-1} \sum (\Delta y_i)^2)} \) referred to \( N(0, 1) \).
6(b). If the unit root is rejected at 5(b), stop.
7(a). If \( y_1 \) differs significantly from zero, stop.
7(b). If \( y_1 \) does not differ significantly from zero, perform a unit root test which is not invariant to the mean under the null.

In many applications, steps 1 and 2, and steps 6 and 7, will be redundant, so we have also simulated a shorter sequential strategy:

3.2. Strategy S2

Start from step 3(b), finish at step 5.
To see if use of a robust trend test can be recommended in this setting, we have also simulated:
3.3. Strategy S3

Test for the presence of linear trend using Vogelsang’s (1998) t-PS1 statistic. If no trend is detected, perform a unit root test invariant to the mean under the null; if trend is detected, perform a unit root test invariant to linear trend under the null.

3.4. Strategy S3*

Test for linear trend using both t-PS1 and the standard t-statistic normalised by $T^{-1/2}$ (referred to critical values in Table II(ii) of Vogelsang, 1998), rejecting the no-trend null if either test rejects and proceeding as in S3.

Before commenting on other possible sequences of tests, we make some general observations.

(a) Our treatment of the trend and extra lags presupposes that not more than one unit root may be present, as we now illustrate for the case, $p = 1$. Suppose $u_t$ is the non-stationary AR(2):

$$\Delta u_t = \phi \Delta u_{t-1} + \varepsilon_t,$$

(10)

while $y_t$ is given by Eq. (9), with $\beta_2$ possibly zero, to be tested. Assume for the moment that $p$ is known. The first stage unit root test will be based on estimates from the equation

$$y_t = \rho y_{t-1} + \psi \Delta y_{t-1} + \alpha_{02} + \alpha_{12} t + \alpha_{22} t^2 + \text{error},$$

(11)

while the DGP implies that

$$\Delta y_t = (\beta_1 - \beta_2) + 2\beta_2 t + \phi \Delta u_{t-1} + \varepsilon_t.$$ 

(12)

Substituting for $\Delta u_{t-1}$ using (9) and rearranging, we find

$$y_t = y_{t-1} + \psi \Delta y_{t-1} + \{\beta_1(1 - \psi) + \beta_2(3\psi - 1)\} + 2\beta_2(1 - \psi).t + \varepsilon_t.$$ 

(13)

If the unit root test using (11) fails to reject the null, our procedure, S1, then tests the coefficient on $t$ in (14) estimated in differences

$$\Delta y_t = \psi \Delta y_{t-1} + \{\beta_1(1 - \psi) + \beta_2(3\psi - 1)\} + 2\beta_2(1 - \psi).t + \varepsilon_t,$$

(14)

which is efficient provided $\psi \neq 1$, that is, provided there is only a single unit root in $u_t$. This result generalises readily to $p > 1$. 
(b) Choosing lag length via the BIC criterion applied at each step produced the same results in our experiments as choosing lag length only at the odd-numbered steps and retaining this length, but we cannot rule out the possibility that the former could be advantageous in some situations, and so that is what we recommend.

The S1 algorithm has eight possible outcomes: (i) quadratic trend + unit root, (ii) quadratic trend + stationary, (iii) linear trend + unit root, (iv) linear trend + stationary, (v) non-zero mean + unit root, (vi) non-zero mean + stationary, (vii) zero mean + unit root, (viii) zero mean + stationary. The process chosen is the same for outcomes (v) and (vii) (unit root, no trend), and also for (vi) and (viii) (stationary, no trend). For a pure unit root test, the size (or power) of the algorithm is thus the sum of the proportions of even outcomes, while the probability that a linear trend is identified is the sum of the proportions in outcomes (iii) and (iv), and so on. In reporting the experimental results in the next section we concentrate on three issues: overall size/power for the null, \( \rho = 1 \); size/power for linear trend; proportion of correct model identifications.

A feature of the S1 and S2 algorithms is that non-rejection of a unit root at an earlier step can be overturned later if the data allow trend degree to be reduced, but not vice-versa.\(^5\) This accords with what most researchers would do in practice. As sample size increases, the overall size of the unit root test in the sequential algorithms will reduce in the presence of a linear trend because of the power of the trend degree tests. The non-zero size of the latter will result in the trend in some series being misclassified as of too high degree, however.

The behaviour of the S1 and S2 algorithms reflects the interplay between test power at any given step and the quality of the approximation to the sampling distribution of the \( t \)-statistic delivered at the succeeding step. For example, in S1, if the preliminary unit root test has low power (as when \( \rho \) is moderately large), then trend stationary series will often go to Step 2A, in which the degree of the trend is tested in a misspecified equation. This has the effect of reducing the probability that a stationary process with a linear trend is correctly identified if it also has a large AR root. Similarly, in situations in which unit root tests remain over-sized even after augmentation by lagged differences, trend degree will be tested by applying the \( t \)-statistic to inappropriate critical values. In fact, as shown by our Monte Carlo experiments, such effects do not appear to seriously undermine the performance of the S2 procedure, when compared with use of Vogelsang’s robust trend test in S3.

\(^5\) In an earlier version of the paper we allowed unit root rejections to be overturned also, but, following up a suggestion by James MacKinnon, we found after further experiment that the size reduction achieved by this was in most cases more than offset by loss of power.
3.5. Alternative sequential procedures

Perron (1988, pp. 316–317) proposes a strategy (in each case using the Phillips–Perron modified DF tests) which seems to amount to the following. Estimate Eq. (2c) and test

\[ H_{01}: [(\rho - 1), \, z_{0c}, \, z_{1c}] = [0, \, z_0, \, 0], \]

using the \( F \)-test, \( \Phi_3 \), of Dickey and Fuller (1981). If this null is not rejected, test

\[ H_{02}: (\rho - 1) = 0 \]

with the unit root test, and if the null is rejected here, stop. If neither \( H_{01} \) nor \( H_{02} \) is rejected, test

\[ H_{03}: [(\rho - 1), \, z_{0c}, \, z_{1c}] = [0, \, 0, \, 0], \]

using the \( F \)-test, \( \Phi_2 \). If \( H_{03} \) is rejected, stop. Otherwise, estimate Eq. (2b) and test

\[ H_{04}: [(\rho - 1), \, z_{0b}] = [0, \, 0], \]

using the \( F \)-test, \( \Phi_1 \). If this null is not rejected, test

\[ H_{05}: (\rho - 1) = 0, \]

using the unit root test, and if the null is rejected here, stop. If neither \( H_{04} \) nor \( H_{05} \) is rejected and the series has a zero mean then test

\[ H_{06}: (\rho - 1) = 0. \]

As with our strategy, the aim is to use the unit root test with most power; however, the outcomes of the three \( F \)-tests may be ambiguous: what should we do if, say, \( H_{03} \) is rejected and \( H_{02} \) is not? In Perron’s strategy we implicitly treat rejection of \( H_{03} \) as evidence of the presence of drift, not as evidence against the unit root, which clearly it could be. It seems to us that these joint \( F \)-tests will generally beg the question of whether it is the unit root or non-stochastic part of the null which is to be rejected; to interpret the \( \Phi \)-type tests, therefore, one would need to conduct a separate test for trend degree. The S1–S3 procedures do this.

A much simpler decision rule that has been advocated, according to folklore, (we have not found it in print) is

Test \( H_{06} \), and if the test rejects, stop.

Test \( H_{05} \), and if the test rejects, stop.
<table>
<thead>
<tr>
<th>Trend</th>
<th>$\rho$</th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.0</td>
<td>0.11</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.27</td>
<td>0.20</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>0.52</td>
<td>0.41</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>0.70</td>
<td>0.92</td>
<td>0.41</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.10</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td>0.1</td>
<td>1.0</td>
<td>0.12</td>
<td>0.15</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.12</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.90</td>
<td>0.18</td>
<td>0.11</td>
<td>0.06</td>
</tr>
</tbody>
</table>

Table 1
Size/power of sequential rules, all tests nominal level 5%

- $S_1$: Allow quadratic trend and use $t$ tests
- $S_2$: Allow linear trend and use $t$ tests
- $S_3$: Allow linear trend and use Vogelsang tests

For each DGP, $C_1$: Size/power against $t$ tests
$C_2$: Size/power against Trend $\neq 0$
$C_3$: Proportion correct model identified

$T = 100, 5000$ replications.
<table>
<thead>
<tr>
<th>Trend = 0.2^t</th>
<th>1.0</th>
<th>0.95</th>
<th>0.90</th>
<th>0.70</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>0.71 0.73 0.65 0.57 0.57 0.50 0.74 0.85 0.66 0.97 1.0 0.93 0.86 0.95 0.78 0.70 0.70 0.63 0.72 0.72 0.61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.68 0.70 0.68 0.54 0.53 0.53 0.70 0.83 0.70 0.93 1.0 0.93 0.82 0.94 0.82 0.66 0.67 0.66 0.66 0.66 0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.67 0.93 0.67 0.37 0.45 0.36 0.70 1.0 0.70 0.93 1.0 0.93 0.82 1.0 0.82 0.58 0.73 0.58 0.51 0.62 0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>0.08 0.51 0.39 0.09 0.26 0.13 0.06 0.80 0.69 0.65 1.0 0.34 0.12 0.95 0.80 0.11 0.32 0.18 0.16 0.31 0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.05 0.47 0.42 0.07 0.18 0.14 0.04 0.79 0.75 0.53 1.0 0.47 0.08 0.95 0.87 0.07 0.25 0.20 0.10 0.21 0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.04 0.13 0.10 0.04 0.04 0.02 0.04 0.31 0.27 0.53 1.0 0.47 0.07 0.54 0.46 0.05 0.07 0.04 0.07 0.06 0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>0.11 0.53 0.08 0.12 0.18 0.07 0.12 0.95 0.09 0.76 1.0 0.58 0.18 1.0 0.13 0.14 0.26 0.09 0.20 0.26 0.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.08 0.51 0.08 0.09 0.13 0.07 0.09 0.94 0.09 0.66 1.0 0.66 0.13 1.0 0.13 0.10 0.20 0.10 0.14 0.19 0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.08 0.25 0.08 0.05 0.06 0.03 0.09 0.51 0.09 0.66 1.0 0.66 0.13 0.74 0.13 0.07 0.12 0.06 0.07 0.10 0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>0.21 0.67 0.18 0.19 0.24 0.15 0.20 0.98 0.16 0.86 1.0 0.73 0.33 1.0 0.27 0.25 0.34 0.20 0.31 0.36 0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.18 0.65 0.18 0.16 0.20 0.15 0.17 0.98 0.17 0.80 1.0 0.80 0.28 1.0 0.28 0.21 0.30 0.21 0.25 0.30 0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.17 0.47 0.18 0.08 0.14 0.08 0.17 0.78 0.17 0.80 1.0 0.80 0.28 0.94 0.28 0.15 0.26 0.15 0.14 0.22 0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>0.71 0.99 0.65 0.57 0.66 0.51 0.74 1.0 0.66 0.97 1.0 0.93 0.86 1.0 0.78 0.70 0.84 0.63 0.72 0.82 0.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>0.68 0.99 0.68 0.54 0.64 0.54 0.70 1.0 0.70 0.93 1.0 0.93 0.82 1.0 0.82 0.66 0.82 0.66 0.66 0.78 0.66</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>0.68 0.99 0.68 0.50 0.75 0.50 0.70 1.0 0.70 0.93 1.0 0.93 0.82 1.0 0.82 0.66 0.93 0.66 0.64 0.88 0.64</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Test $H_{02}$, and if the test rejects, stop.

Test $\rho = 1$ using (2d).

(The final step being added by us.)

This rule has been justified informally on the grounds that the first two tests are under-sized and lose power when an intercept or non-stochastic trend are present. However, the rule is not truly sequential, since it is equivalent to 'reject the unit root when at least one of the tests rejects'. Viewed solely as a unit root test (as it must be), this procedure has slightly greater size and power than S1, but leaves the trend degree uncertain.

4. Experimental results

Table 1 reports the results of 5000 replications of S1–S3 using the ERS ‘GLS’-type ADF tests in the following design:

(a) sample size, $T = 100$, all series initialised at zero at $t = -25$, so that when $|\rho| < 1$, the sample from $y_0$ to $y_T$ is effectively stationary, while when $\rho = 1$, $y_0$ is a draw from N(0, 25) (in the IID case);
(b) $\varepsilon_t \sim$ IID Gaussian $(0, 1)$, $v_t = \varepsilon_t$, $(1 - \phi L)v_t = \varepsilon_t$ with $\phi = \pm 0.5$, $v_t = (1 - \theta L)\varepsilon_t$ with $\theta = \pm 0.5$, $\pm 0.8$;
(c) $\rho = 1.0, 0.95, 0.90, \text{ and } 0.70$;
(d) $d_t = \beta_0 + \beta_1t$, with $\beta_0 = 0$ and $\beta_1 = 0.0, 0.10, 0.20$;
(e) nominal size of 5% for all tests

4.1. Behaviour under the unit root null

S1–S3 have different overall significance levels (against the null that $\rho = 1$), reflecting the different numbers of unit root tests performed. With no trend, and when the serial correlation correction works effectively (for IID or AR(1) shocks), S1 has actual size about 11%, S2 has size about 8%, and S3 size about 7%. Since only one unit root test is performed in S3, one might expect the size to be no greater than 5%; however, the data-dependent lag selection employed, together with the use of approximate critical values for the ERS $\tau_{\mu\text{GLS}}$ test (i.e. those for the DF $\tau$ test – see ERS, 1996 for details) results in some size inflation even for IID shocks. Vogelsang’s trend test is, except for the positive MA root case, better sized than the unit-root pre-test based ‘t’ test used in S2, while the unit-root pre-test-based $t$-test for quadratic trend used in S1 is also over-sized (compare C2 in the first two rows of Table 1). We thus find that in the absence of a trend, S3 identifies the correct model more often under the unit-root null than S1 or S2.

With a small trend ($0.1t$, equal to $\sigma^2/10$ per period), the unit root test sizes of all three algorithms are slightly reduced, while the low power of the robust trend
test used in S3 now reverses the ranking on C3, the proportion of correct model identifications. Only when the presence of a positive MA or negative AR root inflates the trend test power do any of the algorithms correctly identify the random walk with drift in more than one sample in four. When the trend increases to $\sigma_t^2/5$ per period, S2 is a clear winner because S1 too frequently identifies a spurious quadratic trend, while the S3 trend test still lacks power.

4.2. Behaviour under the (trend) stationary alternative

Except when there is a large MA root, the probability of correct model identification, C3, is U-shaped as $\rho$ varies from 1 to 0.70, reflecting the low power of unit root tests for alternatives close to the null. When no trend is present, S1 is the most successful algorithm, measured by C3, for all stationary alternatives, because the extra power against the unit root outweighs the probability of finding a spurious quadratic trend. We may, however, not be indifferent between the various types of errors that go to make up $(1 - C3)$, the probability of an incorrect model identification.

When the trend is small, S2 is always better than S1, and always as good as S3, and better in the case of positive AR or negative MA roots. For the larger trend, the ranking is the same, although the differences are smaller.

Following a referee's suggestion, we investigated the use of the normalised $t$-statistic, $T^{-1/2}t - W_T$ in Vogelsang's terminology, to test for the trend in $S3^*$. The idea here is that in stationary series this test will be very conservative, contributing neither to size nor to power, but that it has better power under the unit root null than does $t$-PS1. The results, not tabulated, show that for $\rho \leq 0.95$, adding this test to S3 makes no difference except when trend $= 0.2t$; in the latter case, power to detect trend is increased, but never exceeds that of S2. When $\rho = 1$, we find slight size-inflation of the trend test except when $\theta = 0.5$ or 0.8, so that for zero trend, S3 is better than $S3^*$, while if trend $= 0.1t$ or $0.2t$, $S3^*$ is effectively the same as S2.

4.3. Some general comments

We have reported results with the nominal size of every individual test held at 5%, which produces algorithms with actual size (against the unit root null) greater than this, even when the shocks are serially uncorrelated. Results for nominal size 1 and 10% were also produced, but have been omitted to save space – the patterns revealed were similar – the main notable feature being that overall size against a unit root is between 3 and 4% when every test has nominal size 1%. The S1 algorithm will apply four unit root tests to some series, yet its size in the IID case is only 11% even with data-based lag selection. However, the test for quadratic trend is quite badly behaved under the unit root null, which reduces the appeal of this algorithm unless detection of a non-linear trend is
Table 2

Summary results for $T = 250$, IID shocks, 1000 replications trend $= 0$ trend $= 0.1t$ trend $= 0.2t$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 1$</td>
<td>S1</td>
<td>0.11</td>
<td>0.15</td>
<td>0.81</td>
<td>0.09</td>
<td>0.40</td>
<td>0.29</td>
<td>0.08</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>0.08</td>
<td>0.09</td>
<td>0.87</td>
<td>0.06</td>
<td>0.35</td>
<td>0.31</td>
<td>0.05</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>0.07</td>
<td>0.06</td>
<td>0.90</td>
<td>0.05</td>
<td>0.10</td>
<td>0.06</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td>$\rho = 0.95$</td>
<td>S1</td>
<td>0.66</td>
<td>0.11</td>
<td>0.56</td>
<td>0.37</td>
<td>0.51</td>
<td>0.32</td>
<td>0.37</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>0.61</td>
<td>0.05</td>
<td>0.56</td>
<td>0.33</td>
<td>0.47</td>
<td>0.33</td>
<td>0.33</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>0.56</td>
<td>0.03</td>
<td>0.54</td>
<td>0.28</td>
<td>0.51</td>
<td>0.28</td>
<td>0.32</td>
<td>0.71</td>
</tr>
<tr>
<td>$\rho = 0.90$</td>
<td>S1</td>
<td>0.91</td>
<td>0.14</td>
<td>0.77</td>
<td>0.82</td>
<td>0.89</td>
<td>0.75</td>
<td>0.82</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>0.87</td>
<td>0.07</td>
<td>0.81</td>
<td>0.78</td>
<td>0.86</td>
<td>0.78</td>
<td>0.78</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>0.82</td>
<td>0.03</td>
<td>0.79</td>
<td>0.76</td>
<td>0.92</td>
<td>0.76</td>
<td>0.78</td>
<td>0.98</td>
</tr>
<tr>
<td>$\rho = 0.70$</td>
<td>S1</td>
<td>1.0</td>
<td>0.11</td>
<td>0.89</td>
<td>1.0</td>
<td>1.0</td>
<td>0.93</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>1.0</td>
<td>0.05</td>
<td>0.95</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>0.96</td>
<td>0.03</td>
<td>0.93</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

thought to be a priority. In this particular setting, there seems, on balance, to be no advantage in using the robust trend test to choose the unit root test, primarily because of its low power. However, this conclusion rests on the assumption that all modelling errors are equally costly. In a setting in which spurious trend detection in the presence of a unit root was particularly to be avoided, the S3 algorithm might be preferred, but not otherwise.

The experiments were repeated with a sample size of 250; results for IID shocks are given in Table 2, which provide further support for use of the S2 algorithm in this setting. Only under the unit root null, with no trend present does S3 do better.

5. Application to inventory series and conclusion

5.1. Inventory series

Table 3 reports the model classes identified by the S1–S3 procedures, together with the individual unit root test outcomes for the 10 US SIC inventory series studied by Hall (1994).

The data are monthly, seasonally adjusted, real dollar inventory holdings running from 1958:12 to 1988:4. Six versions of the sequential procedure are

---

6 In general, unit root tests should not be conducted on seasonally adjusted data, but our intention here is simply to illustrate the behaviour of the sequential algorithm on series which have been studied before.
Table 3
Sequential procedure applied to inventory series; all tests at nominal 5%

<table>
<thead>
<tr>
<th>SIC</th>
<th>$S_{1\text{df}}$</th>
<th>$S_{2\text{df}}$</th>
<th>$S_{1\text{gl}}$</th>
<th>$S_{2\text{gl}}$</th>
<th>$S_3$</th>
<th>$S_3^*$</th>
<th>$\tau_{\mu}^a$</th>
<th>$\tau_{\mu}^b$</th>
<th>$\tau_{\mu}^c$</th>
<th>$\tau_{\mu}^d$</th>
<th>$\tau_{\mu}^{ql}$</th>
<th>$\tau_{\mu}^{qgl}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>R</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>23</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>A</td>
<td>A</td>
<td>R</td>
<td>A</td>
<td>A</td>
<td>R</td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>A</td>
<td>R</td>
<td>R</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>A</td>
<td>R</td>
<td>R</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>26</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>A</td>
<td>R</td>
<td>R</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>R</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>29</td>
<td>4*</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>A</td>
<td>R</td>
<td>R</td>
</tr>
</tbody>
</table>

*Note: a,b,c,d,e,f individual test outcomes, R = reject.

Key to cols 2–7:
1 = Quadratic trend + unit root.
2 = Quadratic trend + stationary.
3 = Linear trend + unit root.
4 = Linear trend + stationary.
5 = No trend + unit root.
6 = No trend + stationary.

reported: ADF test-based, or GLS test-based, in each case with (S1) and without (S2) the quadratic trend tests, and the GLS test-based variants of $S_3$ and $S_3^*$. We used nominal size of 5% for every test, and a maximum lag of 8.7

Note, firstly, that there is substantial agreement between the ADF and the GLS versions of the procedure; with the quadratic trend tests included, the only disagreement is for SIC 26, which the ADF classifies as stationary around a quadratic trend and the GLS as having no trend but a unit root; without the quadratic trend tests there are three disagreements, SIC 22 being either trendless and stationary (ADF) or trendless and non-stationary (GLS), SIC 25 being either non-stationary around a linear trend (ADF) or stationary around a linear trend (GLS), and SIC 26 being either stationary around a linear trend (ADF) or non-stationary and trendless (GLS). In each of the latter three cases at least one procedure identifies a quadratic trend when this is tested for, suggesting that the inferences based on a maintained linear trend may be fragile. Faced with the choice of GLS or ADF based inference, we would opt for the GLS tests here because of their greater power. It is notable also that the quadratic trend tests

7 Only for SIC 21 would a longer lag have been chosen by the BIC – see Hall's discussion for a demonstration of the importance of the method of lag length selection.
reject the linear trend null for half the series: this is quite strong evidence against reliance on the $\tau_r$ test result for these series; for example, SIC 23 is classified as stationary around a quadratic trend if our full procedure is used, but as non-stationary and trendless otherwise.

Strategy S3 gives the same model as S2, except on SICs 21 and 27 in which no trend is detected, while S3* detects the trend in SIC 27 but not in SIC 21. These results using robust trend tests are thus remarkably consistent with those we obtain via sequential pre-testing.

A second important feature of the results is that the various unit root tests, taken individually, do not always agree. Focusing on the GLS tests, we see in Table 3 that all three tests (at nominal 5%) agree in only half the series. For the series on which $\tau_{gls}$ rejects a unit root, however, so also does the quadratic trend test, and each of these series is classified as stationary around a trend, an outcome which echoes Hall’s interpretation of the disagreements between $\tau_\mu$ and $\tau_r$ which he found.

6. Conclusion

A sequential procedure for unit root testing and trend estimation has been explored which has a clear advantage (simultaneous identification of trend degree) over a naive strategy which rejects a unit root when at least one of the DF family of tests rejects. Overall size of the procedure can be reduced by conducting each component test at a smaller size than is required. The ability of the procedure to correctly classify series with and without linear trends is dependent on the magnitude of the trend, and on the auto-regressive root, deterministic trends being much easier to detect when stochastic trends are absent. By using unit root tests as pre-tests before testing trend degree (in a levels equation if the pre-test rejects, in a differenced equation otherwise) we are able to avoid the use of the non-standard sampling distributions of the $t$-statistics on the trend coefficients which arise when the unit root is estimated in the equation used to test the trend. The size distortions which result are not negligible, but neither are they disastrous, measured by the probability of identifying the correct model (C3 in Tables 1 and 2).

The accommodation of level shifts and trend breaks into unit root testing procedures has been the subject of a number of recent studies, and is discussed at length by Stock (1994). We have not sought to incorporate these developments into the strategy we propose, but extensions along these lines should be possible. The present study develops and illustrates a feasible model selection strategy incorporating tests for both trend degree and unit roots, which clarifies and formalises current practice and which can readily be adopted in applied work.
Acknowledgements

We thank Alastair Hall (who also supplied the inventory data), James Mac-Kinnon, John Nankervis, and other participants at the July 1996 ESRC Econometric Study Group meeting in Bristol, and two anonymous referees, whose comments have greatly improved the paper. All computations were performed in GAUSS. This research was supported by UK ESRC awards R00023 4797 and 6390.

Appendix A

We first present some convenient notation; as in the body of the paper, we write $\tilde{\rho}_1 = 1 + \tilde{c}/T$, and introduce the polynomial matrix, $D_k = [1t \cdots t^k]$, (in which $t^k = [1, 2^k, \ldots, T^k]$), and $(T - 1)x(T - 1)$ quasi-differencing operator,

$$H_j = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ -\tilde{\rho}_j & 1 & 0 & \cdots & 0 \\ 0 & -\tilde{\rho}_j & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\tilde{\rho}_j & 1 \end{bmatrix}$$

so that writing $y = [y_2, \ldots, y_T]'$, $y_0 = [y_1, \ldots, y_{T-1}]'$, the GLS trend estimates are as follows:

$\hat{\beta}_{00\text{GLS}}$ is the OLS regression coefficient obtained by regressing the vector, $H_0y$, on the vector, $H_0D_0$, while $[\hat{\beta}_{01\text{GLS}}, \hat{\beta}_{11\text{GLS}}]'$ results from the OLS regression of the vector, $H_1y$ on the matrix, $H_1D_1$, and finally, $[\hat{\beta}_{02\text{GLS}}, \hat{\beta}_{12\text{GLS}}, \hat{\beta}_{22\text{GLS}}]'$ results from the OLS regression of the vector, $H_2y$ on the matrix, $H_2D_2$.

A.1. Invariance results

We present the relevant results (for the case in which $v_t$ is IID) in Proposition 1, extension to the AR($p$) case being covered in a remark.

Let $u = [u_2, \ldots, u_T]'$, $u_0 = [u_1, \ldots, u_{T-1}]'$. Suppose the non-stochastic trend is of degree, $k$, with coefficients, $\beta(k) = [\beta_0, \ldots, \beta_k]'$, and introduce $\gamma(k)$, the vector of coefficients, $\gamma_i$, which solve the equation,

$$\sum_{i=0}^{k} \gamma_i t^i \equiv \beta_0 + \sum_{i=1}^{k} \beta_i (t - 1)^i. \quad (A.1)$$
We may now write the DGP as
\[ y = D_k \beta(k) + u, \quad (A.2) \]
with \( y_0 = D_k \gamma(k) + u_0 \), and \( u = u_0 \rho_0 + v, v \sim N(0, \sigma^2 I) \).

**Proposition 1.** If the DGP is \((A.2)\), and the equation fitted to the data is
\[ y = [D_k y_0][\hat{\beta}' + e], \quad (A.3) \]
in which \( k \geq k \), then

(a) If \((A.3)\) is estimated by OLS, the ‘t’ statistic for testing \( \rho = \rho_0 \) is invariant to \( \beta(k) \).

(b) If \((A.3)\) is estimated by first detrending \( y \) and \( y_0 \) by pseudo-GLS and then estimating \( \hat{y}^* = y_0 \hat{\rho} + e \), by OLS, the ‘t’ statistic for testing \( \rho = \rho_0 \) is invariant to \( \beta(k) \).

(c) In both (a) and (b), the ‘t’ statistics for testing \( x_{k+1} = 0, \ldots, x_k = 0 \), are invariant to \( \beta(k) \).

(d) The ‘t’ statistics in parts (a)–(c) are not invariant to \( \rho_0 \) or \( k \).

(e) If \((A.3)\) is estimated by OLS, partitioning \( \hat{z} \) into \( \hat{z}_k \) and \( \hat{z}_{k-k} \) we can write
\[
\begin{bmatrix}
\hat{z}_k \\
\hat{z}_{k-k} \\
\hat{\rho}
\end{bmatrix} =
\begin{bmatrix}
\beta(k) \\
0_{k-k} \\
\rho_0
\end{bmatrix} +
\begin{bmatrix}
\hat{\rho} \gamma(k) \\
0_{k-k} \\
0
\end{bmatrix},
\]
where the distribution of the random vector,
\[ w(k) = [[D_k [u_0]] [D_k [u_0]]^{-1} [D_k [u_0]]'] y \]
is invariant to \( \beta(k) \), but depends on \( k, \rho_0 \) and \( u_0 \).

(f) If \( |\rho_0| < 1 \), and \((A.3)\) is estimated by OLS, then the ‘t’ statistics for \( x_{k+1} = 0, \ldots, x_k = 0 \), have their usual asymptotic distribution.

(g) If \( \rho_0 = 1 \) and \((A.3)\) is replaced by OLS estimation of
\[ \Delta y = D_{k-1} \hat{\delta} + e, \quad (A.4) \]
then \( \hat{\delta} \) efficiently estimates the first \( k \) elements of \( [\beta(k) - \gamma(k)] \), and the ‘t’ statistic for \( \delta_k = 0 \) has its usual (Student t) distribution.

**Remark 1.** Parts (a), (c), (d), (f) and (g) are well known, at least for \( k = 1 \), (see for example DeJong et al., 1992, pp. 427–428). The requirement that \( k \geq k \) is essential for the unit root test to be consistent (see West, 1987, and Perron, 1988 for details). In their proposed sequential strategy, Dolado et al. (1990) advocate testing the coefficient on \( t \) in \((A.3)\) with \( k = 1 \), then referring the unit root test
statistic to Normal tables if the coefficient on $t$ is not zero: although the limit distribution of the unit root test statistic is Normal in such a case ($k < k$), this is not an appropriate thing to do, because the resulting test is inconsistent, as proved by Perron (1988), and our strategy differs at this point. Part (b), though trivial to prove, is included because it is required for the practical implementation of the GLS tests.

The formulation of Part (e) we believe to be new; its importance lies in the fact that when the trend degree, $k$, is unknown, we will want to be able to test the hypothesis that $k = k - 1$, and the question is whether estimates of (A.3) are suitable for this. Notice that when $\rho_0 = 1$, the random vector, $w(k)$, has, when suitably normalised, a non-normal limit distribution. This shows that the distribution of $\hat{z}_k$ is both non-normal and not invariant to $\beta$, in general, (via $\gamma$). Similarly, $\hat{z}_{k-k}$ is non-normal. As a result, 't' statistics formed from the elements of $\hat{z}_{k-k}$ have non-standard distributions. On the other hand, Part (f) shows that when the data are trend-stationary, estimates from (A.3) may be used for an approximate 't' test of $k = k - 1$. Finally, Part (g) tells us that when a unit root is present, the hypothesis that $k = k - 1$, may be tested by a standard 't' test applied to $\hat{\delta}_k$ in a regression of $\Delta y$ on $D_{k-1}$.

Since both $\rho$ and $k$ are in practice unknown, Parts (e–g) invite the use of a unit root test as a pre-test before the trend degree is tested either in levels (no unit root) or differences (unit root) – this is implemented in our sequential procedure.

**Remark 2.** If the test regression is augmented by additional lags, the invariance results for the 't' statistics on $\hat{\rho}$ are unaffected. For the 'GLS' tests, this follows from the invariance of the detrended series, $y^*$, to $\beta$, while for the 'OLS' tests, we can use a partitioned regression argument to obtain the same result.

To summarise, provided the trend polynomial in the estimated equation is of degree $k \geq k$, we may construct a $t$-test of $\rho = 1$ with critical values invariant to $\beta$, while if $k > k$, the $t$-statistic on the higher degree coefficients is invariant to $\beta$. The sampling distributions of these $t$-statistics depend on $\rho$ in all cases, however.

### A.2. Model reduction when $\rho = 1$

Suppose $k = k = 2$, and $\rho = 1$. In this situation, Part (g) of Proposition 1 tells us that the natural way to proceed is to estimate

$$\Delta y = D_1 \delta + e,$$  \hspace{1cm} (A.5)

which will be fully efficient for $\delta = [\beta_1 - \beta_2, 2\beta_2]'$, with the $t$ test of $\delta_2 = 0$ (which tests $k = 1$) having its standard sampling distribution. On the other hand, if $\rho \neq 1$, then (A.5) is misspecified; in particular, if $\rho \ll 1$, the $t$-test on $\delta_2$ is very conservative. However, a small experiment suggested that a unit root
pre-test was quite effective in correcting size, so this was incorporated in the strategy.

Similar considerations arise if \( k = k_1 = 1 \), so that an efficient way to test \( k = 0 \) with \( \rho = 1 \) maintained is to estimate

\[
\Delta y = D_0 \delta + e,
\]

and conduct a \( t \)-test of \( \delta = 0 \).

Finally, with \( k = 0 \) maintained, and motivated by the fact that the sampling distribution of \( \hat{\rho}_a \) in (2a) depends on \( y_1/\sigma \), (the DF test being conservative when \( y_1/\sigma \) is large). We might wish to test the plausibility of \( y_1 = v_1 \), i.e. of \( y_1 \) being drawn from \((0, \sigma^2)\). With \( \rho = 1 \) maintained, we can do this using the ratio,

\[
y_1 / (T^{-1} \Sigma(\Delta y_i)^2)^{1/2}.
\]

We include this step in our sequential procedure, but in most practical situations it will be redundant: for the majority of economic time series (A.7) will be large, and we did not fix \( y_1 \) at zero in any of our experiments.

**Proof of Proposition 1.** (1) ‘\( t \)’ statistics on \( \hat{\rho} \).

Since \( k \gg k_1 \), the non-stochastic trend lies in the space spanned by the columns of \( D_k \), so for any non-singular fixed \( T \times T \) matrix, \( H \), not involving \( \beta \), the detrended \( y \) vector using pseudo-GLS estimates of \( \beta \) from the regression of \( Hy \) on \( HD_k \) is invariant to \( \beta \): The detrended vector is

\[
y^H = \{ I - D_k ((HD_k)/(HD_k))^{-1}(HD_k)' H \} y
\]

\[
= \{ I - D_k ((HD_k)/(HD_k))^{-1}(HD_k)' H \} \left[ \begin{array}{c} D_k \beta \\ 0 \end{array} \right] + u,
\]

\[
= \{ I - D_k ((HD_k)/(HD_k))^{-1}(HD_k)' H \} u,
\]

which is invariant to \( \beta \). Obviously, \( \hat{\rho} \) and its associated ‘\( t \)’ statistic calculated from the regression of \([y_1^H, \ldots, y_T^H]'\) on \([y_1^H, \ldots, y_{T-1}^H]'\) will also be invariant to \( \beta \), which proves Part (b).

To prove Part (a), take \( H = I \), and use partitioned regression, to obtain

\[
\hat{\rho} = (y_0'y_0)^{-1} y_0'y_1
\]

which is invariant to \( \beta \), as is its ‘\( t \)’ statistic, because both \( y_0' \) and \( y_1' \) are so.

(2) ‘\( t \)’ statistics on \( \hat{\rho}_a \).

The difficulty in working directly with OLS applied to (A.10), below, is that \( y_0 \) depends on \( \beta \) and \( \rho_0 \):

\[
y = [D_k | y_0] [\hat{z}, \hat{\rho}]' + e.
\]
However, this dependence may be isolated by application of the non-singular transformation,

\[ M = \begin{bmatrix} I_{k+1} & -\gamma' \\ 0' & 1 \end{bmatrix}, \]  

(A.11)

where \( \gamma' = [\gamma(k)', \theta(k - k)'] \). Using \( M \) we obtain

\[ [D_k | y_0]M = [D_k | u_0], \]  

(A.12)

which in turn yields the OLS estimates in (A.10) as

\[ \begin{bmatrix} \hat{x} \\ \hat{\rho} \end{bmatrix} = M[[D_k | u_0]'[D_k | u_0]]^{-1}[D_k | u_0]'[D_k \beta(k) + u]. \]  

(A.13)

Writing \( D_k \beta = D_k[\beta(k) | 0] \), and \( u = u_0 \cdot \rho_0 + v \), we obtain

\[ \begin{bmatrix} \hat{x} \\ \hat{\rho} \end{bmatrix} = M \begin{bmatrix} \beta(k) \\ 0 \\ \rho_0 \end{bmatrix} + [[D_k | u_0]'[D_k | u_0]]^{-1}[D_k | u_0]'v. \]  

(A.14)

Giving

\[ \begin{bmatrix} \hat{x} \\ \hat{\rho} \end{bmatrix} - M \begin{bmatrix} \beta(k) \\ 0 \\ \rho_0 \end{bmatrix} = Mw(k) \text{ say}, \]  

(A.15)

where \( w(k) = [w(k)_1, \ldots, w(k)_{k+2}]' \) is invariant to \( \beta \).

Substituting for \( M \) we find

\[ \begin{bmatrix} \hat{x} \\ \hat{\rho} \end{bmatrix} - \begin{bmatrix} \beta(k) - \rho_0 \gamma(k) \\ 0 \\ \rho_0 \end{bmatrix} = w(k) - \begin{bmatrix} w(k)_{k+2} \gamma(k) \\ 0 \\ 0 \end{bmatrix}, \]  

(A.16)

in which replacing \( w(k)_{k+2} \) in the final term by \( \hat{\rho} - \rho_0 \) establishes Part (e).

(3) Part (f).

This is a standard result for a trend-stationary process.

(4) Part (g).
We illustrate this for the case, \( k = k = 2 \). We have \( \gamma(k) = [\beta_0 - \beta_1 + \beta_2, \beta_1 - 2\beta_2, \beta_2]' \), and (A.16) specialises to

\[
\begin{bmatrix}
\hat{\alpha}_0 \\
\hat{\alpha}_1 \\
\hat{\alpha}_2 \\
\hat{\rho}
\end{bmatrix} =
\begin{bmatrix}
\beta_0 - \rho_0(\beta_0 - \beta_1 + \beta_2) \\
\beta_1 - \rho_0(\beta_1 - 2\beta_2) \\
\beta_2(1 - \rho_0) \\
\rho_0
\end{bmatrix} =
\begin{bmatrix}
w(2)_1 - w(2)_4(\beta_0 - \beta_1 + \beta_2) \\
w(2)_2 - w(2)_4.(\beta_1 - 2\beta_2) \\
w(2)_3 - w(2)_4.\beta_2 \\
w(2)_4
\end{bmatrix}.
\]

(A.17)

Setting \( \rho_0 = 1 \) yields

\[
\begin{bmatrix}
\hat{\alpha}_0 \\
\hat{\alpha}_1 \\
\hat{\alpha}_2 \\
\hat{\rho}
\end{bmatrix} =
\begin{bmatrix}
\beta_1 - \beta_2 \\
2\beta_2 \\
0 \\
1
\end{bmatrix} =
\begin{bmatrix}
w(2)_1 - w(2)_4(\beta_0 - \beta_1 + \beta_2) \\
w(2)_2 - w(2)_4.(\beta_1 - 2\beta_2) \\
w(2)_3 - w(2)_4.\beta_2 \\
w(2)_4
\end{bmatrix}
\]

(A.18)

and replacing \( w(2)_4 \) by \( (\hat{\rho} - 1) \) in the top three lines of the right-hand side and rearranging, we obtain

\[
\begin{bmatrix}
\hat{\alpha}_0 \\
\hat{\alpha}_1 \\
\hat{\alpha}_2 \\
\hat{\rho}
\end{bmatrix} =
\begin{bmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
1
\end{bmatrix} =
\begin{bmatrix}
w(2)_1 - \hat{\rho}(\beta_0 - \beta_1 + \beta_2) \\
w(2)_2 - \hat{\rho}(\beta_1 - 2\beta_2) \\
w(2)_3 - \hat{\rho}\beta_2 \\
w(2)_4
\end{bmatrix}.
\]

in which we see immediately that to test the hypothesis that \( \beta_2 = 0, \) (i.e. that \( k = 1 \), with \( k = 2 \) fitted, we could use the ‘\( i' \) statistic on \( \hat{\alpha}_2 \), which has a non-standard distribution.

However, given that \( \hat{\alpha}_2 \) is also an inefficient estimator in this situation, the natural way to proceed is to estimate

\[
\Delta y = D_1\delta + \epsilon,
\]

(A.19)

since when \( \rho_0 = 1 \), the DGP is

\[
\Delta y_t = \Delta d_t + v_t,
\]

(A.20)

that is,

\[
\Delta y = D_1\delta + v,
\]

(A.21)
which is a classical linear regression model, so that $\hat{\delta}$ is fully efficient for $\delta = [\beta_1 - \beta_2, 2\beta_2]$, with the $t$ test of $\delta_2 = 0$ having its standard sampling distribution.

**Appendix B**

Critical values for unit root test with quadratic trend

$$GLS \text{ test } \tilde{c} = -18.5.$$  

<table>
<thead>
<tr>
<th>$T$</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>-5.04</td>
<td>-4.22</td>
<td>-3.84</td>
</tr>
<tr>
<td>50</td>
<td>-4.45</td>
<td>-3.80</td>
<td>-3.48</td>
</tr>
<tr>
<td>100</td>
<td>-4.16</td>
<td>-3.61</td>
<td>-3.32</td>
</tr>
<tr>
<td>250</td>
<td>-4.05</td>
<td>-3.48</td>
<td>-3.20</td>
</tr>
<tr>
<td>500</td>
<td>-4.05</td>
<td>-3.47</td>
<td>-3.19</td>
</tr>
</tbody>
</table>

Generated by Monte Carlo with 20,000 replications.

**References**


