Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19103-6297
pier@ssc.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 97-026

"Choosing between Levels and Logs in the Presence of Deterministic and Stochastic Trends"

by

Valentina Corradi and Norman R. Swanson

Choosing Between Levels and Logs in the Presence of Deterministic and Stochastic Trends

Valentina Corradi(+) and Norman R. Swanson(++)*

(+) University of Pennsylvania
Department of Economics
3718 Locust Walk
Philadelphia, PA 19104-6297
corradi@econ.sas.upenn.edu, tel. (215) 898-1505

(++) Pennsylvania State University
Department of Economics
Kern Graduate Building
University Park, PA 16802
nswanson@psu.edu, tel. (814) 865-2234

May 1997

JEL classification: C22, C51.

Keywords: completely consistent procedure, deterministic trend, integratedness, Nelson-Plosser data, nonlinear transformation.

*The authors wish to thank John Chao, Frank Diebold, Philip Franses, Clive Granger, Atsushi Inoue, Yasu Murasawa, Phillip Howrey, Lutz Kilian, Shinichi Sakata, and seminar participants at the University of Michigan for helpful comments. Swanson thanks the Research and Graduate Studies Office at Pennsylvania State University for research support.

Abstract

In macroeconometrics, unit root tests are typically performed using logs. While this is sensible from a theoretical macroeconomic perspective, there is no clear reason, particularly from an empirical perspective, why logs should be used rather than levels. Further, standard unit root tests assume linearity under both the null and the alternative hypothesis. Violation of this linearity assumption can result in severe size and power distortion, both in finite and large samples. Finally, casual inspection of gnp per capita, for example, plotted against its fitted linear deterministic trend (see Figure 1) gives no clear indication that a loglinear specification should be preferred to a linear specification. Thus, it is reasonable to address the problem of data transformation before running a unit root test. In this paper we propose a simple completely consistent procedure for choosing between levels and log-levels specifications in the presence of deterministic and/or stochastic trends. Once we have chosen the proper data transformation, we remain with the standard problem of choosing between $I(1)$ or $I(0)$, either in levels or in logs. Based on a series of Monte Carlo experiments, we show that the frequency of selecting the correct data transformation is close to one, even for very small sample sizes, when our procedure is implemented. Empirical evidence is also presented, and suggests that 11 of the 14 variables in the Nelson and Plosser (1982) dataset are appropriately modeled in levels rather than in logs, when performing unit root tests.

1 Introduction

In macroeconometrics, unit root tests are typically performed using logs. This is consistent with much of the real business cycle literature (see e.g. Long and Plosser (1993) and King, Plosser, Stock, and Watson (1991)) where it is suggested, for example, that gnp should be modeled in logs, given an assumption that output is generated according to a Cobb-Douglas production function. While this is sensible from a theoretical macroeconomic perspective, there is no clear empirical reason why logs should be used rather than levels, when performing unit root tests, particularly given that standard unit root tests assume linearity under both the null and the alternative, and violation of this linearity assumption can result in severe size and power distortion, both in finite and large samples (e.g. see Granger and Hallman (1991) and Figure 2 below). Also, casual inspection of gnp per capita, for example, plotted against its fitted linear deterministic trend (see Figure 1) gives no clear indication that a loglinear specification should be preferred to a linear specification. Thus, it is reasonable to address the problem of data transformation before running a unit root test.

The current convention is to define an integrated process of order d (I(d)) as one which has the property that the partial sum of the dth difference, scaled by $T^{-1/2}$, satisfies a functional central limit theorem (FCLT). In this case, integratedness in logs does not imply integratedness in levels, and $vice-versa$. Thus, any *a priori* assumption concerning whether to model data in levels or logs has important implications for the outcome of unit root and related tests. For example, Granger and Hallman (1991) show that the percentiles of the empirical distribution of the Dickey-Fuller (1979) statistic constructed using $exp(X_t)$ are much higher, in absolute value, than the corresponding percentiles constructed using the original time series X_t, when X_t is a random walk process. Thus, inference based on the Dickey-Fuller statistic using the exponential transformation leads to an overrejection of the unit root null hypothesis, when standard critical values are used. More recently it has been shown in Corradi (1995) that if X_t is a random walk, then any convex transformation (such as exponentiation) is a submartingale, and any concave transformation (such as taking logs) is a supermartingale. However, while submartingales and supermartingales have a unit root component, their first differences do not generally satisfy typical FCLTs. Thus, Dickey-Fuller type tests no longer have well defined limiting distributions. Given all of the above considerations, it is of some interest to use a statistical procedure for selecting between linear and loglinear specifications, rather than simply assuming from the outset that a series is best modeled as linear or loglinear. Further, while Cox-type tests are available for the I(0) case, few results are available for the I(1) case.

In this paper we propose and examine a simple testing strategy for choosing among I(0) in levels, I(0) in logs around a nonzero linear deterministic trend, I(1) in levels around a positive linear deterministic trend, and I(1) in logs. Our approach consists of two steps. In the first step, two statistics are constructed, say V_{1T} and V_{2T}. If the underlying data generating process (DGP) is either I(0) in levels (possibly around a linear deterministic trend) or is I(1) in levels, then we show that both of the statistics approach zero in probability. On the other hand, if the underlying DGP is either I(0) in logs (around a nonzero linear deterministic trend) or is I(1) in logs, then one of the statistics diverges at a geometric rate, while the other approaches zero in probability. This enables us to distinguish between linear and loglinear model specifications. The procedure outlined above is completely consistent, as both type I and type II errors approach zero asymptotically (i.e. the asymptotic size is zero and the asymptotic power is one). Given that we now know whether the series is best characterized as linear or loglinear, we are left with the standard problem of selecting between I(1) and I(0). The second step of our testing strategy consists of using any procedure for choosing between I(0) and I(1) (e.g. Dickey and Fuller (1979), Kwiatkowski, Phillips, Schmidt, and Shin (1992), Phillips and Ploberger (1994), and Stock (1994)). As Step 1 is a completely consistent procedure, we ensure that the probability of accepting (rejecting) the true (false) DGP is asymptotically one, if a completely consistent procedure is used in Step 2; and that our approach does not does not suffer from the size distortion problem typical of sequential testing procedures, if a classical unit root test is used in Step 2.

In a stimulating recent paper, Kobayashi & McAleer (1996) propose a related procedure for distinguishing between integratedness in logs and integratedness in levels. However, their approach differs from ours in a number of respects. First, integratedness is a maintained assumption. Second, they derive the limiting distribution of their test statistic under the assumption that the variance of the innovations in both DGPs converges to zero as the sample size approaches infinity, at a sufficiently fast rate.

We examine the finite sample behavior of our procedure via a series of Monte Carlo experiments, and show that type I and type II errors associated with the use of V_{1T} and V_{2T} are very close to zero, even for samples as small as 50 observations. Also, the overall ability of our approach to select the "correct" DGP is shown to be quite reasonable when augmented Dickey-Fuller or Kwiatkowski, Phillips, Schmidt, and Shin (1992, hereafter KPSS) unit root tests are used in the second step of the approach.

Empirical evidence based on Nelson and Plosser (1982) data is also presented. Our main findings can be summarized as follows. First, for 11 out of 14 variables, our procedure suggests constructing unit root tests using levels data. The 3 variables for which we choose logs are: employment, wages, and nominal gnp. Second, we find that unemployment is I(0) using both levels and logs. Third, we find that two other variables are I(0) based on our choice of data transformation. These two variables, employment and real wages, also constitute two of the three variables for which we choose log specifications. As these two variables were modeled in logs by Nelson and Plosser (1982), our findings disagree with their finding that both variables are I(1). However, it should be noted that the t-statistic reported by Nelson and Plosser for real wages is -3.04, which is borderline I(0). Also, the contradiction between our finding and theirs for employment is due to our different approach for selecting the number of lag augmentations to use in our application of the ADF test. In particular, they choose two lags while we choose one lag. Indeed, any choice of lags other than 1 (i.e. 0 ,2, ..., 10) leads to a test statistic value of at most -3.19, in agreement with their findings. Third, for employment, gnp per capita, money, real gnp, and wages, using the inappropriate data transformation leads to I(1) being found when the data are I(0) using the appropriate transformation, and *vice-versa*. This suggests that the use of the correct data transformation is important when constructing unit root tests, thus providing evidence as to the usefulness of our procedure.

The rest of the paper is organized as follows. Section 2.1 gives a heuristic discussion of the approach and the statistics which we use. In Section 2.2 we outline the main results and describe the sequential approach used. Section 3 summarizes the findings of several Monte Carlo experiments. In Section 4, the procedure is applied to a number of U.S. macroeconomic series. Section 5 gives concluding remarks.

2 Distinguishing Between I(0) and I(1) Processes in Logs and Levels

2.1 Pitfalls With Classical Hypothesis Testing Approaches

Given a series of observations on an underlying strictly positive process X_t, $t = 1, 2, \ldots$, our objective is to decide whether: (1) X_t is an I(0) process (possibly around a linear deterministic trend), (2) $\log X_t$ is an I(0) process around a nonzero linear deterministic trend, (3) X_t is an I(1) process (around a positive linear deterministic trend), and (4) $\log X_t$ is an I(1) process, (possibly around a linear deterministic trend). A natural approach to this problem is to construct a test that has a well defined limiting distribution under a particular DGP, and diverges to infinity under all of the other above DGPs.

While it is easy to define a test having a well defined distribution under one of (1)-(4), it not clear how to ensure that the test has power against all of the remaining DGPs. To illustrate the problem, consider the sequence $\hat{\epsilon}_t$, given as the residuals from a regression of X_t on a constant and a time trend. Now, construct the test statistic proposed by Kwiatkowski, Phillips, Schmidt, and Shin (1992, hereafter KPSS):

$$S_T = \frac{1}{\hat{\sigma}_T^2} T^{-2} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} \hat{\epsilon}_t^2 \right)^2 ,$$

where $\hat{\sigma}_T^2$ is a heteroskedasticity and autocorrelation (HAC) robust estimator of $var\left(T^{-1/2} \sum_{j=1}^{t} \epsilon_t\right)$. It is known from KPSS that if X_t is I(0) (possibly around a linear deterministic trend), then S_T has a well defined limiting distribution under the null hypothesis, while S_T diverges at rate T/l_T under the alternative that X_t is an integrated process, where l_T is the lag truncation parameter used in the estimation of the variance term in S_T. However, if the underlying DGP is $\log X_t = \alpha_1 + \delta_1 t + \sum_{j=1}^{t} \epsilon_j$, $\delta_1 > 0$ (i.e. $logX_t$ is a unit root process) then both $\hat{\sigma}_T^2$ and $T^{-2} \sum_{t=1}^{T} \left(\sum_{j=1}^{t} \hat{\epsilon}_j \right)^2$ will tend to diverge at a geometric rate, given that $X_t = \exp(\alpha_1 + \delta_1 t + \sum_{j=1}^{t} \epsilon_j)$. In this case it is not clear whether the numerator or the denominator is exploding at a faster rate. This problem is typical of all tests which are based on functionals of partial sums and variance estimators, and arises because certain *nonlinear* alternatives are not treatable using standard FCLTs.

So far we have analyzed the case in which we perform a test with $I(0)$ as the null hypothesis and $I(1)$ as the alternative. In this case, the statistic is typically constructed in terms of functionals of partial sums scaled by a variance estimator. Another common procedure is to test for the null of $I(1)$ versus the alternative of $I(0)$ using Dickey-Fuller type tests. To illustrate the problems associated with this approach, consider the following simple example. Assume that $\log X_t = \log X_{t-1} + \epsilon_t$, $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$. However, we perform a Dickey-Fuller test using levels. For example, we compute $T(\hat{\alpha}_T - 1)$, where

$$\hat{\alpha}_T = \frac{\sum_{t=2}^{T} X_t X_{t-1}}{\sum_{t=2}^{T} X_{t-1}^2}.$$

Now, $X_t = \exp(\log X_{t-1} + \epsilon_t) = X_{t-1}\exp(\epsilon_t)$, so that we can write:

$$T(\hat{\alpha}_T - 1) = \frac{T\sum_{t=2}^{T} X_{t-1}^2 (e^{\epsilon_t} - 1)}{\sum_{t=2}^{T} X_{t-1}^2}.$$

Note that as $X_t = X_0 \exp(\sum_{j=1}^{t} \epsilon_j)$, standard unit root asymptotics no longer apply. However, by confining our attention to the case where $\epsilon \sim N(0, \sigma_\epsilon^2)$, we can examine the properties of $T(\hat{\alpha}_T - 1)$, thus gaining insight into the performance of a Dickey-Fuller test using an incorrect transformation of the data. Notice that $Ee^{\epsilon_t} = e^{\frac{1}{2}\sigma_\epsilon^2} > 1$. Thus, we might expect that $T(\hat{\alpha}_T - 1)$ tends to diverge to $+\infty$. However, Granger and Hallman (1991) find that this statistic tends to overreject the null of a unit root. One possible explanation for the difference between their finding and our intuition is that the distribution of $e^{\epsilon_t} - 1$ is highly skewed to the left, and has a lower bound of negative one. Thus, even though the mean of $e^{\epsilon_t} - 1$ is positive, this is due to the very long right-tail of the distribution. When ϵ_t is drawn from a standard normal distribution, however, most observations are rather close to zero (e.g. 95% are between 2 and -2). These data, when transformed using $e^{\epsilon_t} - 1$, are mainly between -0.86 and 6.4, say. Further, the median of the distribution of $e^{\epsilon_t} - 1$ is zero. Now, in the context of finite samples, this suggests that if we truncate the distribution of $e^{\epsilon_t} - 1$ to be, say, between -0.8 and 1, then the mean of this truncated distribution will actually be negative (as we draw relatively fewer observations close to the upper bound than negative observations close to the lower bound). In the context of generating data in finite samples, as Granger and Hallman did, this situation indeed seems to have occurred, resulting in mostly large negative values being calculated for the expression $T(\hat{\alpha}_T - 1)$. Put another way, the negative elements of $T\sum_{t=1}^{T} X_{t-1}^2(e^{\epsilon_t} - 1)$ are usually quite large in magnitude, relative to most of the positive elements of the same sum. Of course, in large samples, and with large σ_ϵ^2 we should expect that this result will not hold, as the effect of large positive draws from the distribution of $e^{\epsilon_t} - 1$ begins to dominate the overall sum $T\sum_{t=1}^{T} X_{t-1}^2(e^{\epsilon_t} - 1)$. This intuition suggests that Granger and Hallman's results, while holding for the usual sample sizes and the usual error variances observed in economic time series, should not hold generally. It further suggests that indeed using levels data when the true process is I(1) in logs will produce either overrejection of the unit root null (as Hallman and Granger show), or underrejection of the null. Interestingly, these arguments also suggest that for very special cases (i.e. appropriately chosen σ_ϵ^2 and sample size), the empirical size of the Dickey-Fuller test may actually match the nominal size, even when the wrong data transformation is used! To examine this intuition further, we carried out a Monte Carlo experiment. Data were generated according to a loglinear random walk, with error variance equal to σ_ϵ^2, and the $T(\hat{\alpha}_T - 1)$ test statistic was used to check for a unit root at the 5% level. Samples of 100, 250, and 500 observations were used, 2000 replications were performed, and σ_ϵ was varied between 0.01 and 1000. The results are presented in Figure 2, and support our intuition, suggesting that the empirical size is close to and/or below the nominal size for extremely large samples and/or extremely large error variances. Further, for extremely small error variance, the statistic tends to zero, as the numerator of the of the test statistic tends to zero, again resulting in underrejection of the null. For standard deviations between 0.5 and 100, and for all of the sample sizes which we examine, Granger and Hallman's result holds, however, and the test severely overrejects the null hypothesis. In summary, there appears to be a need to carefully consider which transformation is used when constructing unit root tests, as the wrong transformation may yield entirely misleading results.

Even if we decide to keep integratedness as a maintained assumption, and choose between I(1) in levels and I(1) in logs, or *vice versa*, we do not in general obtain a test which has unit asymptotic power. For example consider constructing a KPSS-type test using the first differences of the levels data (i.e. ΔX_t). Under the null of I(1) in levels the statistic has the usual well defined limiting distribution. However, under the alternative of I(1) in logs it does not necessarily diverge to infinity. Again the reason for this result is that both the numerator and the denominator tend to diverge to infinity if they have a positive linear deterministic trend, and in general we cannot determine whether the numerator or the denominator is diverging at a faster rate. Recently Kobayashi and McAleer (1996) develop a test for the null of I(1) in levels versus the alternative of I(1) in logs, and *vice versa*, under the maintained assumption of integratedness. However, in order to derive the limiting distribution of their statistic under the null and to show the power under the alternative, they assume that the variance of the innovation term approaches zero at a sufficiently fast rate.

In this paper, our aim is to develop a procedure for selecting among levels and log-levels linear specifications, while allowing for both deterministic and stochastic trends. In the short memory case, Cox-type tests for choosing between non-nested specifications, and in particular between levels and log-levels specifications are already available. These tests, however, are not applicable to $I(1)$ variables, in general. Our approach is somewhat different. In a first stage, we construct two simple statistics. If X_t is either $I(0)$ or $I(1)$ in levels, possibly around a linear deterministic trend component, then both of the statistics approach zero in probability. If instead X_t is $I(0)$ in logs around a non-zero linear trend or is $I(1)$ in logs with or without a deterministic trend component, then one of the two statistics diverges at a geometric rate, as the sample size gets large. (The case where $\log X_t$ is $I(0)$ without a deterministic trend component is discussed below.) Thus, we obtain a very simple completely consistent rule for selecting among levels and log-levels specifications. We are then left with the standard problem of choosing between $I(0)$ and $I(1)$, using data which are either in levels or logs. In a second stage, we can either employ a completely consistent procedure (e.g. Phillips and Ploberger (1994) or Stock (1994)), or we can rely on classical hypothesis testing procedures (e.g. Dickey and Fuller (1979) or Kwiatkowski, Phillips, Schmidt, and Shin (1992)).

2.2 Testing Strategies

We propose a simple completely consistent procedure for choosing between levels and log-levels specifications. The procedure avoids the pitfalls discussed above which are associated with standard hypothesis testing approaches. Consider the following four DGPs:

$H_1 : X_t = \alpha_0 + \delta_0 t + \epsilon_t,\ \delta_0 \geq 0$

$H_2 : X_t = \alpha_0 + \delta_0 t + \sum_{j=1}^{t} \epsilon_j,\ \delta_0 > 0$

$H_3(i) : \log X_t = \alpha_1 + \delta_1 t + \epsilon_t,\ \delta_1 = 0$

$H_3(ii) : \log X_t = \alpha_1 + \delta_1 t + \epsilon_t,\ \delta_1 > 0$

$H_3(iii) : \log X_t = \alpha_1 + \delta_1 t + \epsilon_t,\ \delta_1 < 0$

$H_4(i) : \log X_t = \alpha_1 + \delta_1 t + \sum_{j=1}^{t} \epsilon_j,\ \delta_1 = 0$

$H_4(ii) : \log X_t = \alpha_1 + \delta_1 t + \sum_{j=1}^{t} \epsilon_j,\ \delta_1 > 0$

$H_4(iii) : \log X_t = \alpha_1 + \delta_1 t + \sum_{j=1}^{t} \epsilon_j,\ \delta_1 < 0$

Also, assume the following:

Assumption 1: (i) $X_t \geq \beta > 0, \forall t > 0$. (ii) the partial sums of ϵ_t, scaled by $T^{-1/2}$ satisfy a functional central limit theorem.

Given Assumption 1 (ii) above, $\{\epsilon_t\}$ can display dependence and heterogeneity of various forms. For example, most strong mixing processes, mixingales, and near-epoch dependent functions of mixing sequences satisfy a FCLT under mild moment (or domination) conditions. Under H_1, X_t is an I(0) process (possibly around a constant and a positive linear deterministic trend), while under H_2, X_t is an I(1) process (around a constant and a positive linear deterministic trend). Under H_3, $\log X_t$ is an I(0) process with: (i) no deterministic trend (ii) a positive linear deterministic trend, and (iii) a negative linear deterministic trend. In these cases, the sign of the trend component plays a crucial role. For example, if $\delta_1 > 0$, X_t tends to explode at a geometric rate as $t \to \infty$, while if $\delta_1 < 0$, X_t converges to zero at a geometric rate, as $t \to \infty$. Finally, under H_4, $\log X_t$ is an I(1) process, possibly with a linear deterministic trend component.

It is worth noting that the above DGPs include the class of ARMA(p,q) and ARIMA(p,q) models. Let $\epsilon_t = \frac{A(L)}{B(L)} u_t$, where u_t is $iid(0, \sigma_u^2)$, and $A(L)$ and $B(L)$ are lag polynomials of order p and q respectively, with all roots outside the unit circle. Then, under H_2 and H_4, X_t and $\log X_t$ are ARIMA(p,q) processes. Under H_1 and H_3, X_t and $\log X_t$ are ARMA(p,q) when $\delta_i = 0, i = 0, 1$; and when $\delta_i \neq 0, i = 0, 1$, the deviations of X_t and $\log X_t$ from their trends are ARMA(p,q). Furthermore, in the $I(1)$ case (i.e. H_2 and H_4), the addition of an $I(0)$ error term, say v_t, is sufficient to ensure that the above DGPs include the class of DGPs considered by KPSS (1992). Needless to say, the inclusion of an additional $I(0)$ term in the DGPs in H_2 and H_4 does not in any way affect the asymptotic results stated below.

Our objective is to choose among $H_1, H_2, H_3(i) - (iii), H_4(i) - (iii)$. Hereafter, let $\hat{\eta}_t$ denote the residuals from the regression of X_t on a constant and a linear trend, and let $\hat{\xi}_t$ denote the residuals from the regression of $\log X_t$ on a constant and a linear time trend. Consider the following two statistics:

$$V_{1T} = \frac{\left(\frac{1}{T^{3/2}} \sum_{t=1}^{T} \hat{\xi}_t^2\right)^2}{\left(\frac{1}{T} \sum_{t=1}^{T} \hat{\eta}_t^2\right)^2}$$

$$V_{2T} = \left(\frac{1}{T^2} \sum_{t=1}^{T} \hat{\eta}_t^2\right) \left(\frac{1}{T^2} \sum_{t=1}^{T} \hat{\xi}_t^2\right)$$

Then, we have the following.

Proposition 2.2.1: Let Assumption 1 hold.

(a) Given the additional assumption under $H_3(i)$ that $\sup_{t\geq 0} E\left(e^{2\epsilon_t}\right) < \infty$, we have that under both H_1 and $H_3(i)$, $V_{1T} \xrightarrow{p} 0$, and $V_{2T} \xrightarrow{p} 0$, where V_{1T} converges at rate $T^{-(1-\gamma)}$ and V_{2T} converges at rate $T^{-2(1-1/2\gamma)}$, $\gamma > 0$ and arbitrarily small.

(b) Under H_2, $V_{1T} \xrightarrow{p} 0$ at rate $T^{-3+\gamma}$, and $V_{2T} \xrightarrow{p} 0$ at rate $T^{-1+\gamma}$, $\gamma > 0$ and arbitrarily small.

(c) Under $H_3(ii)$ and $H_4(ii)$, $V_{1T} \xrightarrow{p} 0$ at a geometric rate, and $V_{2T} \xrightarrow{p} \infty$ at a geometric rate.

(d) Under $H_3(iii)$ and $H_4(iii)$, $V_{1T} \xrightarrow{p} \infty$ at a geometric rate, and $V_{2T} \xrightarrow{p} 0$ at a geometric rate.

(e) Under $H_4(i)$, either

$$V_{1T} \xrightarrow{p} 0, V_{2T} \xrightarrow{p} \infty$$

or

$$V_{1T} \xrightarrow{p} \infty, V_{2T} \xrightarrow{p} 0,$$

where both the divergence to infinity and the convergence to zero occur at geometric rates.

Proof: See Appendix.

Proposition 2.2.1 can be interpreted as follows. Under H_1 and $H_3(i)$, $\frac{1}{T}\sum \hat{\eta}_t^2$ converges in probability to a non-deterministic limit (given the law of large numbers), while $\frac{1}{T}\sum \hat{\xi}_t = O_p(\log T)$. Thus, both V_{1T} and V_{2T} approach zero in probability. Under H_2, $\frac{1}{T^2}\sum \hat{\eta}_t^2$ converges in distribution to a well defined non-degenerate limit, and hence $\frac{1}{T}\sum \hat{\eta}_t^2$ diverges at rate T, On the other hand, $\frac{1}{T}\sum \hat{\xi}_t^2 = O_p(\log T)$. Thus, both V_{1T} and V_{2T} again approach zero in probability. Under $H_3(ii)$ and $H_4(ii)$, $\hat{\eta}_t = exp(\alpha_1 + \delta_1 t + \epsilon_t) - \hat{\alpha}_T - \hat{\delta}_T$ and $\hat{\eta}_t = exp(\alpha_1 + \delta_1 t + \sum_{j=1}^{t} \epsilon_j) - \hat{\alpha}_T - \hat{\delta}_T$, respectively. This implies that $\frac{1}{T}\sum \hat{\eta}_t^2$ explodes at a geometric rate. However, under $H_3(ii)$, $\frac{1}{T}\sum \hat{\xi}_t^2 = O_p(1)$, while under $H_4(ii)$, $\frac{1}{T^2}\sum \hat{\xi}_t^2 = O_p(1)$. It follows that $V_{1T} \xrightarrow{p} 0$ and $V_{2T} \xrightarrow{p} \infty$, and that convergence (divergence) occurs at a geometric rate in both cases. In the case of $\delta_1 < 0$, (i.e. $H_3(iii)$ and $H_4(iii)$), $\frac{1}{T}\sum \hat{\eta}_t^2$ converges to zero at a geometric rate. Thus, $V_{1T} \xrightarrow{p} \infty$ and $V_{2T} \xrightarrow{p} 0$. Finally, consider $H_4(i)$, where $X_t = exp(\alpha_1 + \sum_j^t \epsilon_j)$. Set $t = [Tr]$. Then, given that $\frac{1}{T^{1/2}}\sum_j^{[Tr]} \epsilon_j$ weakly converges to a nondegenerate limit, it follows that $|\sum_j^{[Tr]} \epsilon_j|$ diverges with probability approaching one, and so $\frac{1}{T^2}\sum \hat{\eta}_t^2$ either diverges to infinity or converges to zero, at a geometric rate. Thus, either $V_{1t} \xrightarrow{p} 0$ and $V_{2T} \xrightarrow{p} \infty$, or *vice versa*.

Let $H_A = H_1 \bigcup H_2 \bigcup H_3(i)$ and $H_B = H_3(ii) - (iii) \bigcup H_4(i) - (iii)$. The following summarizes our procedure.

STEP 1: If $V_{1T} \leq 1$ and $V_{2T} \leq 1$, choose H_A. Otherwise, choose H_B. If H_A is chosen, go to Step 2. Otherwise, go to Step 3.

STEP 2: Choose between $I(1)$ and $I(0)$ using data in levels. If the variable is found to be $I(1)$, then choose H_2. Otherwise, choose $H_1 \bigcup H_3(i)$.

STEP 3: Choose between $I(1)$ and $I(0)$ using logged data. If the variable is found to be $I(1)$, then choose H_4. Otherwise, choose H_3.

Step 1 of the procedure follows directly from Proposition 2.2.1, as under H_A, both V_{1T} and V_{2T} approach zero in probability, while under H_B, either V_{1T} or V_{2T} diverges to infinity. We have chosen unity as the "cut-off" value in Step 1. Although at first glance, this may appear to be a somewhat arbitrary choice, our simulation results reported below suggest that the rule *works*, in the sense that the percentage of times that the wrong hypothesis is selected is arbitrarily close to zero, even when very small samples are used in the construction of the statistics. However, given that either V_{1T} or V_{2T} diverges at a geometric rate under H_B, the probability of choosing H_A, when H_B is true, approaches zero very rapidly. If we are concerned about the probability of choosing H_B, when instead H_A is true, we can choose a somewhat looser "cut-off" value. Once either H_A or H_B are chosen, it remains only to select between $I(0)$ and $I(1)$ in levels, or between $I(1)$ and $I(0)$ in logs, in which case we proceed either to Step 2 or Step 3, respectively. There are many procedures which can be used in Steps 2 and 3. For example, a completely consistent procedure for selecting among $I(1)$ and $I(0)$ may be implemented. This would ensure that the probability of selecting the true DGP using Steps 1-3 approaches unity as the sample size increases. Examples of completely consistent unit root procedures include Phillips and Ploberger (1994) and Stock (1994). Alternatively, many classical hypothesis testing procedures are available for implementation in Steps 2 and 3. If classical tests are used, then the entire procedure (Steps 1-3) is not affected by the usual size distortion problem associated with sequential testing procedures, as Step 1 is based on a completely consistent approach. Two widely used classical hypothesis tests which can be used in Steps 2 and 3 are: (i) Dickey-Fuller type tests for the null of $I(1)$ versus the alternative of $I(0)$ (using levels if H_A is chosen, and using logs if H_B is chosen); and (ii) KPSS tests for the null of $I(0)$ versus the alternative of $I(1)$ (again using levels if H_A is chosen, and using logs if H_B is chosen).

It should be noted that if the variable being examined using Steps 1-3 is found to be $I(0)$ in Step 2, then a choice remains between $H_3(i)$ and H_1. In particular, the series may be $I(0)$ in logs with no linear deterministic trend, or $I(0)$ in levels, possibly around a linear deterministic trend. Two comments are in order here. First, Step 2 uses levels data to choose between $I(1)$ and $I(0)$, even if the true DGP is $H_3(i)$. This, however, is of little concern, as $H_3(i)$ and H_1 are indistinguishable in the context of unit root tests and procedures. Second, it may be useful to examine the linear deterministic trend component of a fitted model of the data, for example, as a coefficient significantly different from zero suggests choosing H_1. When the same coefficient is not significantly different from zero, then a Cox-type test might be used to select among logs and levels in the context of $I(0)$ data (e.g. as suggested Pesaran and Pesaran (1993)).

So far we have only considered the problem of choosing between *levels* and *log − levels* specifications. One reason for this approach is that macroeconometricians typically model data either in levels or in logs. However, the above procedure can be easily modified to select among data in levels or data which has been transformed in some other manner, say $g(X_t)$. Define $\hat{\xi}_t$ as the residual from the regression of $g(X_t)$ on a constant and a linear deterministic time trend. Provided that g is unbounded and $\lim_{x\to\infty} g(x)/x = 0$, we can still rely on the ideas which lead to Proposition 2.2.1, and formulate a procedure that is analogous to the one outlined above. However, it should be noted that the rate of divergence of one of the two statistics is no longer geometric in general, under the g transformation, but depends on the particular functional form of g.

3 Finite Sample Performance

In this section the results of two related Monte Carlo experiments are presented. In the first experiment, the finite sample performance of using V_{1T} and V_{2T} to discriminate between H_A and H_B is examined (this corresponds to Step 1 of the above procedure). In the second experiment, the finite sample performance of the entire testing procedure is examined (by carrying out Steps 1-3). Thus, the first experiment addresses the usefulness of V_{1T} and V_{2T} for choosing between expressing a variable in levels or logs, prior to carrying out unit root tests. The second experiment uses the results of the first experiment in conjunction with standard strategies for selecting between I(0) and I(1) to fully classify the variable being examined as either: (1) I(0) in levels possibly around a positive (linear) deterministic trend or I(0) in logs with no deterministic trend (this is denoted by "H_1 or $H_3(i)$"); (2) I(0) in logs with a positive or negative deterministic trend (this is denoted by "$H_3(ii)$ or $H_3(iii)$"); (3) I(1) in levels around a positive deterministic trend (this is denoted by "H_2"); or (4) I(1) in logs with a positive, negative or zero deterministic trend (this is denoted by "$H_4 = H_4(i) - (iii)$"). Both experiments are based on the following DGP:

$$y_t = \beta_1 t + r_t + \eta_t,$$

$$r_t = r_{t-1} + \nu_t.$$

This is the same DGP as that used by KPSS (1992). One of the reasons why we use this DGP is that in the second experiment we construct KPSS (and also Augmented Dickey-Fuller) tests to select between I(0) and I(1), given prior knowledge (from Step 1 of the procedure) as to whether the data are better modeled in levels or in logs. Our experimental setup closely follows KPSS, although they consider ρ between -0.8 and 0.8, and we allow it to vary between -0.99 and 0.99. In particular, we model η_t as a $AR(1)$ process in I(0) cases (i.e. $\eta_t = \rho\eta_{t-1} + \mu_t$), allowing for different choices of $\rho \in (0,1)$. Clearly as ρ gets closer to one, it becomes somewhat more difficult to distinguish between $I(0)$ and $I(1)$. In $I(1)$ cases, η_t is generated as $iidN(0,1)$, and the finite sample properties of our procedure is examined by varying the parameter $\lambda = \sigma_\nu^2/\sigma_\eta^2$. In the $I(0)$ case, $\lambda = 0$ as $\sigma_\nu^2 = 0$, while in the $I(1)$ case, the lower the value of λ, the greater the likelihood of mistakenly identifying an $I(1)$ process as an $I(0)$ process, as the unit root component of y_t accounts for little of the stochastic variation of the series. Under H_3 and H_4, we set the initial value, say r_0, equal to zero, while under H_1 and H_2 we set $r_0 = 100$. In theory, under H_1 and H_2, we should impose some sort of "constraint" on the support of η_t and on the support of μ_t or ν_t, in order to ensure the positivity of $X_t, \forall t$. However, by using an initial value of r_0 which is large enough, it turns out that we never draw a negative value, and thus we avoid having to impose restrictions on the support of the distributions.

In summary, we generate data using the above DGP, and corresponding to the hypotheses $H_1 - H_4$. For example, note that for H_3 and H_4, data which are I(0) and I(1) in logs are generated. This is accomplished by setting $y_t = \log z_t$, where z_t is the underlying series in levels. Then, the series in levels is constructed by forming $exp(y_t)$. The parameterizations which we consider can be summarized as follows. For H_1 and $H_3(i) - (iii)$ data are generated using: $\beta_1 = 0$ (for H_1 and $H_3(i)$), $\beta_1 = 0.1$ (for H_1 and $H_3(ii)$), $\beta_1 = -0.1$ (for $H_3(iii)$),

$\rho = \{0.99, 0.9, 0.5, 0, -0.5, -0.9, -0.99\}$, and $\mu_t \sim iidN(0,1)$. In all of these cases, $\sigma_\nu^2 = 0$. For H_2 and $H_4(i) - (iii)$ we generate data using $\beta_1 = 0.1$ (for $H_4(i)$), $\beta_1 = 0.1$ (for H_2 and $H_4(ii)$), $\beta_1 = -0.1$ (for $H_4(iii)$), $\lambda = \{0.0001, 0.01, 1, 100, 100000\}$, and both η_t and $\nu_t \sim iidN(0,1)$. In all experiments, samples of 50, 100, 250, and 500 observations are used. However, for the sake of brevity, and because the results based on our first experiment change little, we only include tabulated results for samples of 50 and 100 observations. (Complete results are available upon request.) All experiments are based on 5000 replications, and in all cases where ADF and KPSS tests are run, critical values correspond to a 5% nominal size.

Tables 1A.1-1A.2 and 1B.1-1B.2 present results based on the first experiment. For DGPs generated according to H_1-H_4, the frequency of occurrence of all possible combinations of V_{1T} and V_{2T} are tabulated. Also, the expected magnitudes (from Proposition 2.1.1) of the two statistics are given in the last column of the tables. For example, in Table 1A.1 (sample size = 50) and Table 1A.2 (sample size = 100), Proposition 2.1.1 suggests that V_{1T} and V_{2T} should both be less than one for H_1 and $H_3(i)$. The same applies to H_2 in Tables 1B.1 and 1B.2. Together, these hypotheses form H_A (see above). It is immediately apparent, even upon cursory inspection, that the frequency of times that the expected combination of statistics is observed is always close to unity, even for samples of only 50 observations. In fact, there are only two notable exceptions to this finding. The first is the case of $H_3(i)$ when ρ is equal to 0.99 or -0.99 (see Tables 1A.1 and 1A.2). In this case, we do not find that V_{1T} and V_{2T} are both less than unity, as expected. This is due to the fact that the DGPs in these cases mimic unit root processes, so that our simulated data appear to have been generated according to log difference stationary processes with no linear deterministic trend (i.e. $H_4(i)$). The second exception occurs when data are generated according to DGP $H_4(i)$, and λ is very close to zero. A value for λ close to zero indicates that the variance of the random walk component is very small relative to the variance of y_t. Thus, the data tend to mimic data generated according to a short memory process (i.e. $H_3(i)$, as discussed in KPSS). In short, conditional on the two standard exceptions noted here, Step 1 of our procedure appears to perform surprisingly well, even for very small samples.

Tables 2A.1-2A.2 and 2B.1-2B.2 report results based on the second experiment, when the ADF test is used in Steps 2 and 3. The method used to select the number of lags in the ADF test regressions is discussed in the next section, and in the footnote to Table 4. Each ADF test reported on is carried out using a stepwise procedure, which begins with the formation of an ADF regression with both intercept and trend. If the series is found to be I(0) based on this regression, the procedure is stopped. Otherwise, the significance of the trend component is examined, and if the trend component is found to be insignificant, then a new regression is estimated with only an intercept. If the series is then found to be I(0), the procedure is stopped. Otherwise, we test whether or not to include the intercept term in the ADF regression, and continue by estimating a regression with no intercept or trend, if appropriate.

First, consider Tables 2A.1-2A.2, which report results based on data generated according to short memory processes. The entries of these tables can be interpreted as follows. Consider the first entry in the upper left corner of Table 2A.1, which is 0.334. This entry denotes the frequency of times that Steps 1-3 of our procedure result in the selection of "H_1 or $H_3(i)$", given that actual data are generated according to $H_1(\delta = 0)$. Put another way, the data are first subjected to Step 1 of our procedure, and either H_A or H_B is selected. Then, given the appropriate data transformation implied by Step 1, an ADF test is performed, and the data are further classified as either I(0) or I(1). For simplicity, when H_A is selected in Step 1, we use a levels transformation of the data. However, as H_A includes not only H_1 and H_2 (I(0) and I(1) in levels), but also $H_3(i)$ (I(0) in logs with no deterministic linear trend), this approach should be viewed as a simplification. In practice, though, this simplification does not affect our results, as H_1 and $H_3(i)$ are indistinguishable from an empirical perspective (i.e. when ADF tests are used). It should be noted that the empirical power of the ADF test (from Steps 2-3 of our procedure), as well as the "empirical power of the entire procedure" (from Steps 1-3) can also be inferred from the entries in Tables 2A.1-2A.2. Here, by the "empirical power of the entire procedure", we mean the frequency of rejection of the column entry "H_1 or $H_3(i)$" when the actual data are generated according to $H_1(\delta = 0)$, say.

To illustrate all of the above concepts, consider the row of entries in Table 2A.1 corresponding to data generated according to $H_3(ii)$ with $\rho = 0.99$. The entries are 0.008, 0.331, 0.012, and 0.647. The entry 0.331 is the frequency of times that the true DGP, $H_3(ii)$, falls within the class of DGPs ("$H_3(ii)$ or $H_3(iii)$") selected by the entire procedure. Also, the empirical power of the ADF test is 0.008+0.331=0.339, as these two entries amount to the probability of rejecting the I(1) null hypothesis when it is false. Finally, the empirical power of the entire procedure is 0.331, which is the intersection of the probability of finding H_B using Step 1 of our procedure, and the probability of finding that the variable is I(0) using the ADF test. Upon examination of Tables 2A.1-2A.2 it is clear that the empirical power of our procedure is very high for all values of ρ, except for ρ=0.99, -0.99. This is not surprising, as the ADF test has very little power for values of $|\rho|$ close to unity, and so our entire procedure likewise has low power. Further, even for as few as 100 observations, the overall power of the procedure improves appreciably for moderate values of ρ. For example, for $\rho = 0.5$ and data generated according to $H_3(ii)$, the empirical power of the procedure increases from 0.759 to 0.932 when the sample is increased from 50 to 100 observations.

Tables 2B.1-2B.2 are analogous to Tables 2A.1-2A.2, except that empirical size is presented, rather than empirical power, given that all data in these two tables are generated according to I(1) processes. Consider the row of entries corresponding to data generated according to $H_4(ii)(\delta > 0)$ with $\lambda = 1.0$. The entries are 0.040, 0.458, 0.031, and 0.470. In this example, the empirical size of the ADF test is 0.040+0.458=0.498. Also, the empirical size of our entire procedure is 0.040+0.458+0.031=0.529. Notice that the empirical size of the ADF test as well as of the entire procedure is always above 0.80 for in Table 2B.1 for $\lambda = 0.0001$. This is not surprising, as $\lambda = 0.0001$ is the case for which very little of the variation in y_t is due to the random walk component (r_t). Also, not surprisingly, as λ and the sample size increase, the empirical size of the ADF test as well as of the entire procedure improve dramatically.

Tables 3A.1-3A.2 and 3B.1-3B.2 are based on the same experiment reported on in Tables 2A.1-2A.2 and 2B.1-2B.2. However, there are a number of differences in the test procedure used, and in the layout of the tables. First, 6 different KPSS unit root test statistics - $\hat{\eta}_\eta$ and $\hat{\eta}_\tau$, for $l0$, $l4$, and $l12$, rather than one ADF unit root test, are used in Steps 2 and 3 of the procedure. Second, no results are reported for data generated according to $H_3(i)$. This is because 6 times as many columns of results would need to be presented in order to adequately partition the entries in the table, thus allowing for a reasonable interpretation of data generated according to $H_3(i)$. (These results, however, are available for the authors.) Third, entries in the tables correspond to the empirical size of the entire procedure (Tables 3A.1-3A.2) and to the empirical power of the entire procedure (Tables 3B.1-3B.2). For example, the upper left entry in Table 3A.1, which is 0.957, suggests that when data are generated according to $H_1(\delta = 0)$, the probability of finding that the actual data are generated by either H_2, $H_3(ii) - (iii)$, or H_4 is 0.957. This empirical size is very poor, and arises because the data are generated with $\rho = 0.99$, in which case the KPSS test used in Steps 2-3 of our procedure tends to severely overreject the null of I(0). Notice that the empirical size of the entire procedure improves as the lag truncation parameter is increased from $l0$ to $l12$, as $|\rho|$ decreases, and as the sample increases from 50 to 100 observations. This finding is in accord with the findings of KPSS (1992, Table 3). Now, consider the upper left entry in Table 3B.1, which is 1.000, suggesting that when the data are generated according to $H_2(\delta > 0)$ with $\lambda = 0.0001$, the probability of finding that the actual data are generated by H_2 is 1.000. In this case, the empirical power of our procedure is extremely good. However, this result is an artifact of the fact that the true data are generated with a positive deterministic trend, but the test statistic used is $\hat{\eta}_\eta$ with $l0$. This KPSS statistic invariably rejects the I(0) null because the deterministic trend is ignored, thus leading to the selection of H_2 by the entire procedure. Indeed, when data are generated according to $H_2(\delta > 0)$, the appropriate columns in Table 3B.1 are the last three, in which results based on $\hat{\eta}_\tau$ KPSS test statistics are reported. Here, entries are close to 0.05, which is actually the expected empirical size of the procedure. However, this result is completely expected, as $\lambda = 0.0001$, in which case the data mimic those generated according to an I(0) process (as discussed above). Of final note is that the empirical power of the procedure increases as λ and the sample size increase, but decreases as we move from $l0$ to $l12$. Again, this result is in accord with KPSS (1992, Table 4).

In summary, our procedure appears to perform as well in finite samples as the unit root test used in the second and third steps of the procedure. This is due to the impressive finite sample performance of the completely consistent approach used in the first step of the procedure, and suggests that practitioners may use our procedure with the knowledge that it will essentially perform as well as the classical unit root test used in Steps 2 and 3.

4 Empirical Application: U.S. Macroeconomic Series

In this section, an updated version of the Nelson and Plosser (1982) dataset is examined. V_{1T}, V_{2T}, ADF, and KPSS statistics are calculated for all of the variables. More precisely, ADF test statistics based on regressions with a constant and trend, $\hat{\tau}_\tau$, as used in Nelson and Plosser (1982), and KPSS statistics using two of the three lag truncation parameters used in our simulations - $l4$ and $l12$ (see above), and with deterministic trend, $\hat{\eta}_\tau$, are reported on. Complete results for ADF and KPSS test statistics, including $\hat{\tau}$, $\hat{\tau}_\mu$, $\hat{\tau}_\tau$, $\hat{\eta}_\mu$, and $\hat{\eta}_\tau$, for $l0$, $l4$, and $l12$, are available upon request from the authors. For the ADF test statistics, the number of lagged differences used in the test regressions was selected by examining their t-statistics, starting with 10 lags, and stopping when the last lagged difference had a coefficient significantly different from zero at a 95 percent level of confidence.

Although the Nelson-Plosser data have been examined many times, it has usually been assumed *a priori* that the data are best modeled in logs. Our approach is to first use Step 1 of the above procedure to ascertain whether the variables are better modeled as linear in logs or in levels. Thereafter, we apply ADF and KPSS tests to the appropriately transformed data, as well as to the inappropriately transformed data. Although the construction of ADF and KPSS test statistics to these data may seem a bit repetitive given the large number of previous studies in which the same data have been examined, we nevertheless include these results, in order to illustrate how our entire procedure (Steps 1-3) can be applied empirically, and because it turns out that using what we find to be the wrong data transformation can lead to misleading inference.

Table 4 contains our findings. The data are the same as those used by Schotman and van Dijk (1991), who updated the original Nelson-Plosser (1982) dataset. The sample size considered for each of the variables is equal to the length of the shortest series in the dataset, and is 80 observations (1909-1988). For 11 out of 14 variables, our procedure suggests constructing unit root tests using levels data. The 3 variables for which we choose logs are: employment, wages, and nominal gnp. It is somewhat surprising that we choose logs for nominal gnp, but choose levels for real gnp and per capita gnp. This is particularly surprising, as much of the real business cycle literature (e.g. Long and Plosser (1993) and King, Plosser, Stock, and Watson (1991)) suggests that gnp should be modeled in logs, given an assumption that output is generated according to a Cobb-Douglas production function, say. To examine this finding further, we plotted per capita gnp and nominal gnp against their fitted linear deterministic trends (see Figure 1). While there is clear evidence that nominal gnp is well modeled as loglinear around a deterministic trend, even cursory inspection of the plots indicates that choosing between levels and logs for per capita gnp not straightforward.

Table 4 also reports our results based on ADF and KPSS unit root tests, using both the correct and incorrect data transformation, where by "correct" we mean the transformation chosen by our procedure. A number of findings emerge. First, we find that unemployment is I(0) using both levels and logs. Further, even though we choose levels, our above definition of "levels" includes I(0) in logs with no deterministic trend. As pointed out above, I(0) in logs with no deterministic trend is the only loglinear specification which is included in the definition of "levels" which we use in Table

4. As unemployment is the only series which is found to be I(0) in levels, it is also the only series for which our procedure is unable to clearly distinguish between linear and loglinear specifications.

Second, we find that two other variables are I(0) based on the correct transformation. These two variables, employment and real wages also constitute two of the three variables for which we choose log specifications. As these two variables were modeled in logs by Nelson and Plosser (1982), our findings disagree with their findings that both variables are I(1). However, it should be noted that the t-statistic reported by Nelson and Plosser for real wages is -3.04, which is borderline I(0). Also, the contradiction between our finding and theirs for employment is due to our different approaches for selecting the number of lag augmentations to use in our application of the ADF test. In particular, they choose two lags while we choose one lag. Indeed, any choice of lags other than 1 (i.e. 0 ,2, ..., 10) leads to a test statistic value of at most -3.19, in agreement with their findings.

Third, for employment, gnp per capita, money, real gnp, and wages, using the incorrect data transformation leads to I(0) being found when the data are I(1) using the correct transformation, and *vice-versa*. This finding is the same regardless of whether ADF or KPSS test statistics are examined. This suggests that the use of the correct data transformation is important when constructing unit root tests, thus providing evidence of the usefulness of our procedure.

5 Conclusions

In this paper we propose a simple consistent procedure for choosing between levels and log-levels specifications in the presence of stochastic and/or deterministic trends. The procedure is designed to help in the selection of the appropriate data transformation when carrying out unit root tests, for example. Our procedure is carried out in two steps. First, the appropriate data transformation is chosen. Second, a standard unit root test is performed in order to select between $I(1)$ and $I(0)$, given the appropriate data transformation. The second step can be carried out by using a wide variety of classical hypothesis tests, or alternatively, using a completely consistent approach. If a completely consistent approach is used, then the entire procedure is completely consistent. However, even if a classical hypothesis test is used, the size distortion problem typical of sequential testing procedures is avoided, as our procedure for selecting the appropriate data transformation is completely consistent.

In order to assess the usefulness of the proposed procedure, a series of Monte Carlo experiments were carried out. Results based on these experiments are encouraging. For example, the frequency at which the correct data transformation is selected is very close to unity in most cases, even with samples as small as 50 observations. Empirical evidence is also presented which suggests that modeling macroeconomic variables in levels may in some cases be preferable to modeling them in logs. This evidence is based on an examination of the dataset used in Nelson and Plosser (1982). Perhaps most importantly, we show that for this dataset, using the inappropriate data transformation leads to I(1) being found when the data are I(0) using the appropriate transformation, and *vice-versa*, for 5 of 14 variables. This suggests that the use of the correct data transformation is important when constructing unit root tests, thus providing evidence of the usefulness of our procedure.

6 Appendix

Proof of Proposition 2.2.1: Unless otherwise stated, all summations run from 1 to T. Let $\hat{\alpha}_T$ and $\hat{\delta}_T$ denote the estimated coefficients from the regression of X_t on a constant and a linear time trend, and define

$$D_T = \frac{T^2(T+1)(2T+1)}{6} - \frac{T^2(T+1)^2}{4} = \frac{T^4}{12}$$

we have

$$\hat{\alpha}_T = \frac{T(T+1)(2T+1)}{6D_T}\sum X_t - \frac{T(T+1)}{2D_T}\sum tX_t \tag{1}$$

$$\hat{\delta}_T = -\frac{T(T+1)}{2D_T}\sum X_t + \frac{T}{D_T}\sum tX_t \tag{2}$$

Analogously, let $\tilde{\alpha}_T$ and $\tilde{\delta}_T$ be the coefficients from the regression of $\log X_t$ on a constant and a linear time trend, we have

$$\tilde{\alpha}_T = \frac{T(T+1)(2T+1)}{6D_T}\sum \log X_t - \frac{T(T+1)}{2D_T}\sum t\log X_t \tag{3}$$

$$\tilde{\delta}_T = -\frac{T(T+1)}{2D_T}\sum \log X_t + \frac{T}{D_T}\sum t\log X_t \tag{4}$$

(a) Under H_1,

$$\sqrt{T}(\hat{\alpha}_T - \alpha_0) = \frac{T(T+1)(2T+1)}{6D_T/T}\frac{1}{\sqrt{T}}\sum \epsilon_t - \frac{T(T+1)}{2D_T/T^2}\frac{1}{T^{3/2}}\sum t\epsilon_t$$

and

$$T^{3/2}(\hat{\delta}_T - \delta_0) = -\frac{(T(T+1)}{2D_T/T^2}\frac{1}{\sqrt{T}}\sum \epsilon_t + \frac{T}{D_T/T^3}\frac{1}{T^{3/2}}\sum t\epsilon_t$$

Thus

$$\frac{1}{T}\sum \hat{\epsilon}_t^2 = \frac{1}{T}\sum(\epsilon_t - (\hat{\alpha}_T - \alpha_0) - (\hat{\delta}_T - \delta_0)t)^2 = \frac{1}{T}\sum \epsilon_t^2 + o_p(1) \xrightarrow{p} \sigma_\epsilon^2$$

and so $\frac{1}{T^2}\sum \hat{\eta}_t^2 \xrightarrow{p} 0$ at rate T. Now $\log X_t = \log(\alpha_0 + \delta_0 t + \epsilon_t)$, so that $\tilde{\alpha}_T = O_p(\log T)$ and $\tilde{\delta}_T = O_p\left(\frac{\log T}{T}\right)$, when the trend component is nonzero. When $\delta_0 = 0$, then $\tilde{\alpha}_T = O_p(1)$ $\tilde{\delta}_T = O_p(\frac{1}{T})$. It follows that

$$\frac{1}{T^{3/2}}\sum(\log(\alpha_0 + \delta_0 t + \epsilon_t) - \tilde{\alpha}_T - \tilde{\delta}_T t)^2 \xrightarrow{p} 0 \qquad .$$

at rate $T^{-1/2+\gamma/2}$, and $\frac{1}{T^2}\sum \hat{\xi}_t^2$ approaches zero at rate $T^{-(1-\gamma)}$, $\gamma > 0$, arbitrarily small. So $V_{1T} \xrightarrow{p} 0$ at rate $T^{-1+\gamma}$, and $V_{2T} \xrightarrow{p} 0$ at rate $T^{-2(1-\gamma)}$, $\eta > 0$, arbitrary small.
Under $H_3(i)$, $\sqrt{T}(\tilde{\alpha}_T - \alpha_1) = O_p(1)$ and $T^{3/2}(\tilde{\delta}_T - \delta_1) = O_p(1)$, so that

$$\frac{1}{T^{3/2}}\sum \hat{\xi}_t^2 \xrightarrow{p} 0$$

at rate $T^{-1/2}$. Now $X_t = \exp(\alpha_1 + \epsilon_t)$, so that $E(X_t^2) = exp(2\alpha_1)E(exp(2\epsilon_t)) < \infty$, by assumption. Thus $\frac{1}{T}\sum X_t^2 = O_p(1)$, so $\hat{\alpha}_T = O_p(1)$ and $\hat{\delta}_T = O_p(T^{-1})$. So

$$\frac{1}{T}\sum \hat{\eta}_t^2 = \frac{1}{T}\sum(exp(\alpha_1 + \epsilon_t) - \hat{\alpha}_T - \hat{\delta}_T t)^2 = O_p(1)$$

also note that the expression above is strictly positive except for a set of measure zero. It follows that $V_{1T} \xrightarrow{p} 0, V_{2T} \xrightarrow{p} 0$, at rate $T^{-1+\gamma}$ and $T^{-2(1-\gamma)}$ respectively..

(b) Under H_2,

$$\frac{1}{\sqrt{T}}(\hat{\alpha}_T - \alpha_0) = \frac{T(T+1)(2T+1)}{6D_T/T}\frac{1}{T^{3/2}}\sum_{t=1}^{T}(\sum_{j=1}^{t}\epsilon_j) - \frac{T(T+1)}{2D_T/T^2}\frac{1}{T^{5/2}}\sum_{t=1}^{T} t\sum_{j=1}^{t}\epsilon_j,$$

so that

$$\frac{1}{\sqrt{T}}(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} 4\sigma_\epsilon\int_0^1 W_s ds - 6\sigma_\epsilon\int_0^1 sW_s ds$$

and

$$\sqrt{T}(\hat{\delta}_T - \delta_0) = -\frac{T(T+1)}{2D_T/T^2}\frac{1}{T^{3/2}}\sum_{t=1}^{T}(\sum_{j-1}^{t}\epsilon_j) + \frac{T}{D_T/T^3}\frac{1}{T^{5/2}}\sum_{t=1}^{T} t\sum_{j=1}^{t}\epsilon_j$$

so that

$$\sqrt{T}(\hat{\delta}_T - \delta_0) \xrightarrow{d} -6\sigma_\epsilon\int_0^1 W_s ds + 12\sigma_\epsilon\int_0^1 sW_s ds$$

Consequently,

$$\frac{1}{T^2}\sum\hat{\eta}_t^2 = \frac{1}{T^2}\sum(\epsilon_t - (\hat{\alpha}_T - \alpha_0) - (\hat{\delta}_T - \delta_0)t)^2 \xrightarrow{d} \sigma_\epsilon^2 \int_0^1 (W_s^\tau)^2 ds \quad .$$

where

$$W_s^\tau = W_s + (6s-4)\int_0^1 W_r dr + (-12s+6)\int_0^1 rW_r dr$$

Thus

$$\frac{1}{T}\sum\hat{\eta}_t^2 \xrightarrow{p} \infty$$

at rate T.

On the other hand, $\tilde{\alpha}_T = O_p(\log T)$ and $\tilde{\delta}_T = O_p\left(\frac{\log T}{T}\right)$, so that

$$\frac{1}{T^{3/2}}\sum(\log(\alpha_0 + \delta_0 t + \sum_{j=1}^{t}\epsilon_j) - \tilde{\alpha}_T - \tilde{\delta}_T t)^2 \xrightarrow{p} 0$$

at rate $T^{-1/2+\gamma}$, $\gamma > 0$, arbitrary small. Thus $V_{1T} \xrightarrow{p} 0$ at rate $T^{-3+\gamma}$, and $V_{2T} \xrightarrow{p} 0$ at rate $T^{-1+\gamma}$.

(c) Under $H_4(ii)$, $X_t = exp(\alpha_1 + \delta_1 t + \sum_{j=1}^{t}\epsilon_j)$, thus $|\hat{\alpha}_T|$ and $|\hat{\delta}_T|$, explode at a geometric rate as $T \to \infty$. Thus

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t = \frac{1}{T}\sum(exp(\alpha_1 + \delta_1 t + \sum_{j=1}^{t}\epsilon_j) - \hat{\delta}_T t - \hat{\alpha}_T)^2 \xrightarrow{p} \infty \quad .$$

at a geometric rate.

Note that, because of the law of the iterated logarithm for weakly dependent processes (Eberlain, theorem 1, 1986),

$$\limsup_{T\to\infty}\frac{1}{\sqrt{2T\log\log T}}|\sum_{j=1}^{t}\epsilon_j| = \sigma_\epsilon,$$

so that the asymptotic behavior of X_t is driven by the deterministic trend component; thus also under $H_3(ii)$,

$$\frac{1}{T}\sum\hat{\eta}_t^2 \xrightarrow{p} \infty$$

at a geometric rate. Under $H_3(ii)$,

$$\frac{1}{T^{3/2}}\sum\tilde{\xi}_t^2 = \frac{1}{T^{3/2}}\sum(\epsilon_t - (\tilde{\alpha}_T - \alpha_1) - (\tilde{\delta}_T - \delta_1)t)^2 \xrightarrow{p} 0$$

at rate $T^{-1/2}$, $V_{1T} \xrightarrow{p} 0, V_{2T} \xrightarrow{p} \infty$, and the convergence (divergence) occurs at a geometric rate. Under $H_4(ii)$, by the same argument used in part (b), we know that

$$\frac{1}{T^2}\sum\tilde{\xi}_t^2 \xrightarrow{d} \sigma_\epsilon^2\int_0^1 (W_s^\tau)^2 ds$$

where W_s^T is a detrended Brownian motion. Thus

$$\frac{1}{T^{3/2}} \sum \hat{\xi}_t^2 \xrightarrow{p} \infty$$

at rate $T^{1/2}$. As $\frac{1}{T} \sum \hat{\eta}_t^2, \frac{1}{T^2} \sum \hat{\eta}_t^2$ diverge at a geometric rate, the same outcome as in $H3(ii)$ occurs.

(d) Essentially by a similar argument as in (c), noting that for $\delta_1 < 0$, under $H_3(ii)$,

$$\frac{1}{T} \sum \hat{\eta}_t^2 = \frac{1}{T} \sum (exp(\delta_1 t + \alpha_1 + \epsilon_t) - \hat{\alpha}_T - \hat{\delta}_T t)^2 \xrightarrow{p} 0$$

at a geometric rate, and similarly under $H_4(ii)$. The behavior of $T^{-3/2} \sum \hat{\xi}_t^2$ is as in part (c) under both $H_3(iii)$ and $H_4(iii)$. Thus simply the asymptotic behavior of the two statistics is switched, and $V_{1T} \xrightarrow{p} \infty$ and $V_{2T} \xrightarrow{p} 0$, and the convergence (divergence) occurs at a geometric rate.

(e) The limiting behavior of $T^{-3/2} \sum \hat{\xi}_t^2$ is as in (b) and (c). Now as $T^{-1/2} \sum_{t=1}^{[Tr]} \epsilon_j \xrightarrow{d} N(0, r\sigma_\epsilon^2)$, for any $r \in [0, 1]$, it follows that, for $r \in [0, 1]$,

$$\lim_{T \to \infty} P\left(|\sum_{j=1}^{[Tr]} \epsilon_j| > C_T \right) = 1$$

where $\frac{C_T}{T^{1/2-\gamma/2}} \to \infty$, as $T \to \infty$, for $\gamma > 0$, arbitrarily small.
Given (3)-(4), we see that $|\hat{\alpha}_T|$ and $|\hat{\delta}_T|$ diverge to infinity, or approach zero, at a geometric rate, depending whether $\sum_{j=1}^T \epsilon_j \xrightarrow{p} \infty$ or $\sum_{j=1}^T \epsilon_j \xrightarrow{p} -\infty$, but we know that $\sum_{j=1}^T \epsilon_j$, in absolute values, diverges to infinity, with probability approaching one. Thus depending on the divergence to ∞ or to $-\infty$ of $\sum \epsilon_j$, we know that either $V_{1T} \xrightarrow{p} 0$ and $V_{2T} \xrightarrow{p} \infty$, or *vice versa*. In all cases the divergence (convergence) occurs at a geometric rate.

7 References

Corradi, V., (1995), Nonlinear Transformation of Integrated Time Series, *Journal of Time Series Analysis*, 16, 539-550.

Dickey, D.A. and W.A. Fuller, (1979), Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association*, 74, 427-431.

Eberlain, E., (1986), On Strong Invariance Principles under Dependence Assumptions, *Annals of Probability*, 14, 260-270.

Granger, C.W.J. and J. Hallman, (1991), Nonlinear Transformations of Integrated Time Series, *Journal of Time Series Analysis*, 12, 207-224.

Granger, C.W.J. and N.R. Swanson, (1996), Further Developments in the Study of Cointegrated Variables, *Oxford Bulletin of Economics and Statistics*, 58, 537-553.

Granger, C.W.J. and N.R. Swanson, (1997), An Introduction to Stochastic Unit Root Processes, *Journal of Econometrics*, forthcoming.

King, R.G., C.I. Plosser, J.H. Stock and M.M. Watson, (1991), Stochastic Trends and Economic Fluctuations, *American Economic Review*, 81, 819-840.

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

Kobayashi, M. and M. McAleer, (1996), Tests of Linear and Logarithmic Transformations for Integrated Processes, Yokohama National University and University of Western Australia, manuscript.

Long, J.B. and C.I. Plosser, (1983), Real Business Cycles, *Journal of Political Economy*, 91, 39-69.

Nelson, C.R. and C.I. Plosser, (1982), Trends and Random Walks in Macro-economic Time Series, *Journal of Monetary Economics*, 10, 139-162.

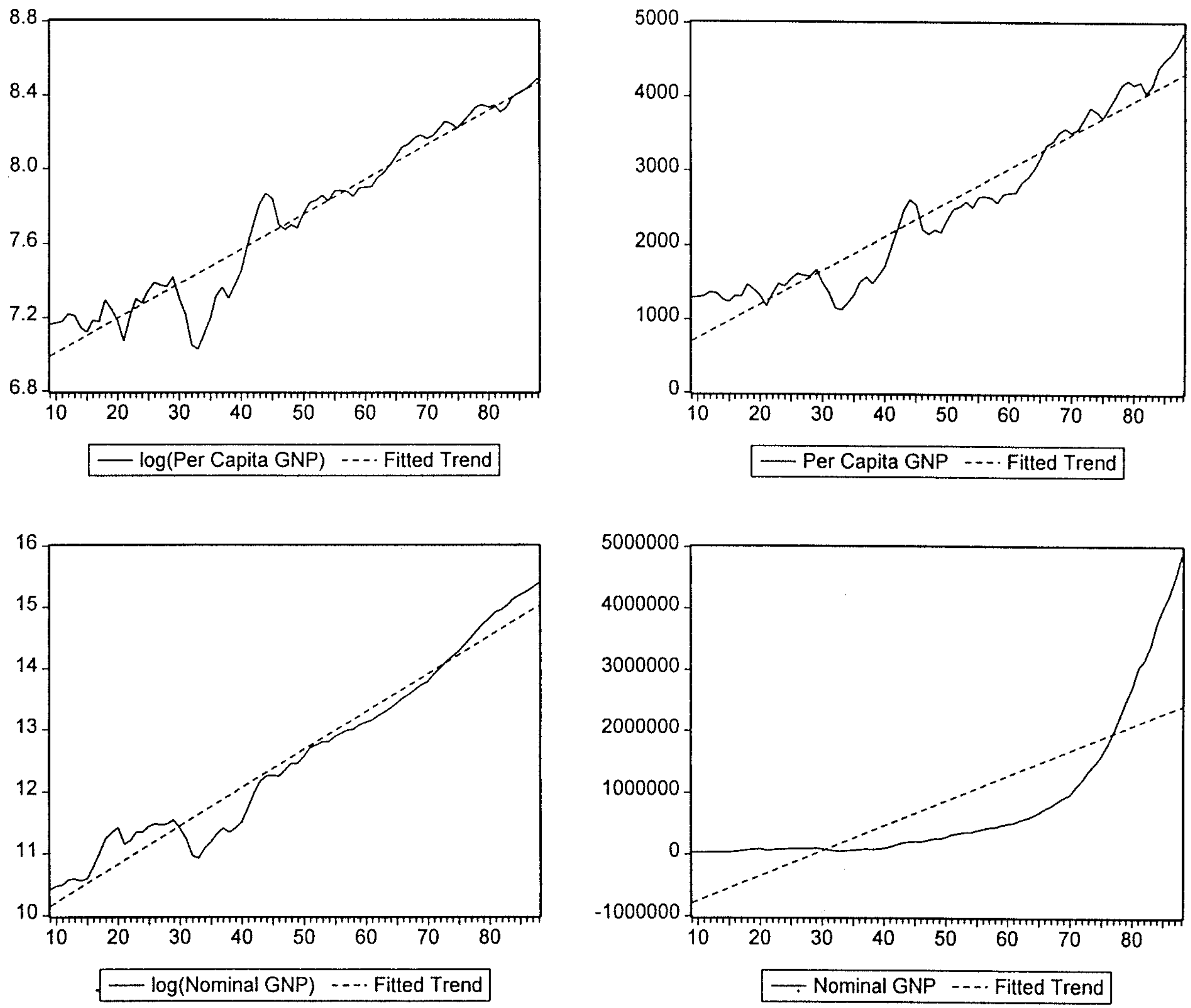
Pesaran B. and M.H. Pesaran, (1993), A Simulation Approach to the Problem of Computing Cox's Statistics for Testing Non-nested Models, *Journal of Econometrics*, 57, 377-392.

Phillips, P.C.B. and W. Ploberger, (1994), Posterior Odds Testing for a Unit Root with Data-Based Model Selection, *Econometric Theory*, 10, 774-808.

Schotman, P.C. and H.K. van Dijk, (1991), On Bayesian Routes to Unit Roots, *Journal of Applied Econometrics*, 6, 387-401.

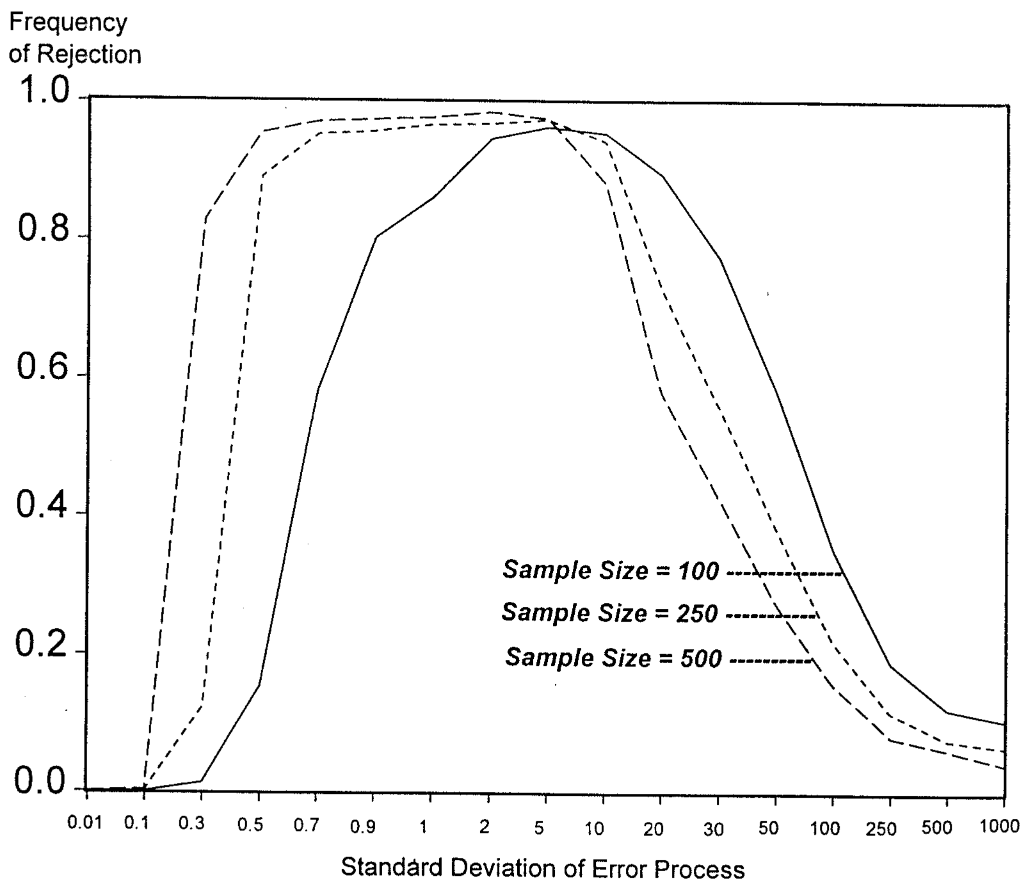
Stock, J.H., (1994), Deciding Between I(1) and I(0), *Journal of Econometrics*, 63, 105-131.

Figure 1: Logged and Levels Nonimal GNP and Per Capita GNP



Footnote: The variables are updated Nelson-Plosser (1982) data taken from Schotman and van Dijk (1991), and are all from 1909-1988. fitted trends are based on a linear regression with constrant and linear deterministic trend.

Figure 2: Empirical Size of the Dickey-Fuller Test Using Levels Instead of Logs



Footnote: Graphs report percentage of rejections of the I(1) null when performing a Dickey-Fuller test at a 5% level, using levels data, given that the true DGP is I(1) in logs. Figures plotted are based on 2000 Monte Carlo replications, with the standard deviation of the error process used to generate the data reported along the horizontal axis.

Table 1A.1: Performance of V_{1T} and V_{2T} for Selecting Among Levels and Log Transformations[1]

Hypotheses Examined: H_1 and H_3. Sample Size = 50 Observations

Data Generated According To:	Parameter Values	V_{1T} and V_{2T} $(V_{1T}\leq1, V_{2T}\leq1)$	$(V_{1T}\leq1, V_{2T}>1)$	$(V_{1T}>1, V_{2T}\leq1)$	$(V_{1T}>1, V_{2T}>1)$	Expected Signs of (V_{1T}, V_{2T})
H_1 (δ=0)	ρ=0.99	1.000	0.000	0.000	0.000	$(V_{1T}\leq1, V_{2T}\leq1)$
	ρ=0.90	1.000	0.000	0.000	0.000	
	ρ=0.50	1.000	0.000	0.000	0.000	
	ρ=0.00	1.000	0.000	0.000	0.000	
	ρ=−0.50	1.000	0.000	0.000	0.000	
	ρ=−0.90	1.000	0.000	0.000	0.000	
	ρ=−0.99	1.000	0.000	0.000	0.000	
H_1 (δ>0)	ρ=0.99	1.000	0.000	0.000	0.000	$(V_{1T}\leq1, V_{2T}\leq1)$
	ρ=0.90	1.000	0.000	0.000	0.000	
	ρ=0.50	1.000	0.000	0.000	0.000	
	ρ=0.00	1.000	0.000	0.000	0.000	
	ρ=−0.50	1.000	0.000	0.000	0.000	
	ρ=−0.90	1.000	0.000	0.000	0.000	
	ρ=−0.99	1.000	0.000	0.000	0.000	
$H_3(i)$	ρ=0.99	0.292	0.480	0.226	0.000	$(V_{1T}\leq1, V_{2T}\leq1)$
	ρ=0.90	0.801	0.192	0.006	0.000	
	ρ=0.50	1.000	0.000	0.000	0.000	
	ρ=0.00	1.000	0.000	0.000	0.000	
	ρ=−0.50	0.999	0.000	0.000	0.000	
	ρ=−0.90	0.617	0.382	0.000	0.000	
	ρ=−0.99	0.088	0.910	0.000	0.000	
$H_3(ii)$	ρ=0.99	0.021	0.970	0.002	0.000	$(V_{1T}\leq1, V_{2T}>1)$
	ρ=0.90	0.000	1.000	0.000	0.000	
	ρ=0.50	0.000	1.000	0.000	0.000	
	ρ=0.00	0.000	1.000	0.000	0.000	
	ρ=−0.50	0.000	1.000	0.000	0.000	
	ρ=−0.90	0.000	1.000	0.000	0.000	
	ρ=−0.99	0.000	1.000	0.000	0.000	
$H_3(iii)$	ρ=0.99	0.085	0.024	0.890	0.000	$(V_{1T}>1, V_{2T}\leq1)$
	ρ=0.90	0.000	0.000	1.000	0.000	
	ρ=0.50	0.000	0.000	1.000	0.000	
	ρ=0.00	0.000	0.000	1.000	0.000	
	ρ=−0.50	0.000	0.000	1.000	0.000	
	ρ=−0.90	0.000	0.000	1.000	0.000	
	ρ=−0.99	0.042	0.087	0.870	0.000	

[1] Entries correspond to the frequency of times that the test statistics are: (1) $(V_{1T}\leq1, V_{2T}\leq1)$, (2) $(V_{1T}\leq1, V_{2T}>1)$, (3) $(V_{1T}>1, V_{2T}\leq1)$, and (4) $(V_{1T}>1, V_{2T}>1)$. Data are generated according to the DGPs outlined in the first two columns of the table. The expected signs given in the final column of the table are based on Proposition 2.2.1 and correspond to Step 1 of the procedure outlined in Section 2 of the paper. All entries are based on 5000 Monte Carlo simulations.

Table 1A.2: Performance of V_{1T} and V_{2T} for Selecting Among Levels and Log Transformations[1]

Hypotheses Examined: H_1 and H_3. Sample Size = 100 Observations

Data Generated According To:	Parameter Values	V_{1T} and V_{2T} $(V_{1T}\leq 1, V_{2T}\leq 1)$	$(V_{1T}\leq 1, V_{2T}>1)$	$(V_{1T}>1, V_{2T}\leq 1)$	$(V_{1T}>1, V_{2T}>1)$	Expected Signs of (V_{1T}, V_{2T})
H_1 (δ=0)	ρ=0.99	1.000	0.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}\leq 1)$
	ρ=0.90	1.000	0.000	0.000	0.000	
	ρ=0.50	1.000	0.000	0.000	0.000	
	ρ=0.00	1.000	0.000	0.000	0.000	
	ρ=−0.50	1.000	0.000	0.000	0.000	
	ρ=−0.90	1.000	0.000	0.000	0.000	
	ρ=−0.99	1.000	0.000	0.000	0.000	
H_1 (δ>0)	ρ=0.99	1.000	0.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}\leq 1)$
	ρ=0.90	1.000	0.000	0.000	0.000	
	ρ=0.50	1.000	0.000	0.000	0.000	
	ρ=0.00	1.000	0.000	0.000	0.000	
	ρ=−0.50	1.000	0.000	0.000	0.000	
	ρ=−0.90	1.000	0.000	0.000	0.000	
	ρ=−0.99	1.000	0.000	0.000	0.000	
$H_3(i)$	ρ=0.99	0.287	0.573	0.138	0.000	$(V_{1T}\leq 1, V_{2T}\leq 1)$
	ρ=0.90	0.820	0.179	0.000	0.000	
	ρ=0.50	1.000	0.000	0.000	0.000	
	ρ=0.00	1.000	0.000	0.000	0.000	
	ρ=−0.50	1.000	0.000	0.000	0.000	
	ρ=−0.90	0.709	0.290	0.000	0.000	
	ρ=−0.99	0.047	0.953	0.000	0.000	
$H_3(ii)$	ρ=0.99	0.002	0.997	0.000	0.000	$(V_{1T}\leq 1, V_{2T}>1)$
	ρ=0.90	0.000	1.000	0.000	0.000	
	ρ=0.50	0.000	1.000	0.000	0.000	
	ρ=0.00	0.000	1.000	0.000	0.000	
	ρ=−0.50	0.000	1.000	0.000	0.000	
	ρ=−0.90	0.000	1.000	0.000	0.000	
	ρ=−0.99	0.000	1.000	0.000	0.000	
$H_3(iii)$	ρ=0.99	0.089	0.021	0.889	0.000	$(V_{1T}>1, V_{2T}\leq 1)$
	ρ=0.90	0.000	0.000	1.000	0.000	
	ρ=0.50	0.000	0.000	1.000	0.000	
	ρ=0.00	0.000	0.000	1.000	0.000	
	ρ=−0.50	0.000	0.000	1.000	0.000	
	ρ=−0.90	0.000	0.000	1.000	0.000	
	ρ=−0.99	0.073	0.068	0.858	0.000	

[1] See notes to Table 1A.1.

Table 1B.1: Performance of V_{1T} and V_{2T} for Selecting Among Levels and Log Transformations[1]

Hypotheses Examined: H_2 and H_4. Sample Size = 50 Observations.

Data Generated According To:	Parameter Values	V_{1T} and V_{2T} $(V_{1T}\leq 1, V_{2T}\leq 1)$	$(V_{1T}\leq 1, V_{2T}>1)$	$(V_{1T}>1, V_{2T}\leq 1)$	$(V_{1T}>1, V_{2T}>1)$	Expected Signs of (V_{1T}, V_{2T})
H_2 ($\delta>0$)	$\lambda=0.0001$	1.000	0.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}\leq 1)$
	$\lambda=0.01$	1.000	0.000	0.000	0.000	
	$\lambda=1.0$	1.000	0.000	0.000	0.000	
	$\lambda=100$	1.000	0.000	0.000	0.000	
	$\lambda=10000$	1.000	0.000	0.000	0.000	
$H_4(i)$ ($\delta=0$)	$\lambda=0.0001$	1.000	0.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}>1)$ or $(V_{1T}>1, V_{2T}\leq 1)$
	$\lambda=0.01$	0.999	0.000	0.001	0.000	
	$\lambda=1.0$	0.191	0.454	0.354	0.000	
	$\lambda=100$	0.001	0.628	0.370	0.000	
	$\lambda=10000$	0.001	0.628	0.369	0.000	
$H_4(ii)$ ($\delta>0$)	$\lambda=0.0001$	0.000	1.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}>1)$
	$\lambda=0.01$	0.000	1.000	0.000	0.000	
	$\lambda=1.0$	0.071	0.897	0.031	0.000	
	$\lambda=100$	0.001	0.672	0.325	0.001	
	$\lambda=10000$	0.001	0.672	0.325	0.001	
$H_4(iii)$ ($\delta<0$)	$\lambda=0.0001$	0.000	0.000	1.000	0.000	$(V_{1T}>1, V_{2T}\leq 1)$
	$\lambda=0.01$	0.000	0.000	1.000	0.000	
	$\lambda=1.0$	0.116	0.014	0.869	0.000	
	$\lambda=100$	0.001	0.573	0.425	0.000	
	$\lambda=10000$	0.001	0.573	0.425	0.000	

[1] See notes to Table 1A.1.

Table 1B.2: Performance of V_{1T} and V_{2T} for Selecting Among Levels and Log Transformations[1]

Hypotheses Examined: H_2 and H_4. Sample Size = 100 Observations.

Data Generated According To:	Parameter Values	V_{1T} and V_{2T} $(V_{1T}\leq 1, V_{2T}\leq 1)$	$(V_{1T}\leq 1, V_{2T}>1)$	$(V_{1T}>1, V_{2T}\leq 1)$	$(V_{1T}>1, V_{2T}>1)$	Expected Signs of (V_{1T}, V_{2T})
H_2 ($\delta>0$)	λ=0.0001	1.000	0.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}\leq 1)$
	λ=0.01	1.000	0.000	0.000	0.000	
	λ=1.0	1.000	0.000	0.000	0.000	
	λ=100	1.000	0.000	0.000	0.000	
	λ=10000	1.000	0.000	0.000	0.000	
$H_4(i)$ ($\delta=0$)	λ=0.0001	1.000	0.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}>1)$
	λ=0.01	0.999	0.000	0.000	0.000	or
	λ=1.0	0.214	0.456	0.328	0.000	$(V_{1T}>1, V_{2T}\leq 1)$
	λ=100	0.002	0.657	0.340	0.000	
	λ=10000	0.001	0.657	0.341	0.001	
$H_4(ii)$ ($\delta>0$)	λ=0.0001	0.000	1.000	0.000	0.000	$(V_{1T}\leq 1, V_{2T}>1)$
	λ=0.01	0.000	1.000	0.000	0.000	
	λ=1.0	0.049	0.937	0.013	0.000	
	λ=100	0.002	0.707	0.290	0.000	
	λ=10000	0.001	0.708	0.290	0.000	
$H_4(iii)$ ($\delta<0$)	λ=0.0001	0.000	0.000	1.000	0.000	$(V_{1T}>1, V_{2T}\leq 1)$
	λ=0.01	0.000	0.000	1.000	0.000	
	λ=1.0	0.083	0.005	0.911	0.000	
	λ=100	0.001	0.598	0.400	0.000	
	λ=10000	0.002	0.598	0.400	0.000	

[1] See notes to Table 1A.1.

Table 2A.1: Complete Procedure Performance Using ADF for Steps 2-3[1]

Hypotheses Examined: H_1 and H_3. Sample Size = 50 Observations

Data Generated According To:	Parameter Values	Model Selected			
		Short Memory		Integrated (I(1))	
		H_1 or $H_3(i)$	$H_3(ii)$ or $H_3(iii)$	H_2	$H_4(i)$–(iii)
H_1 (δ=0)	ρ=0.99	0.334	0.000	0.665	0.000
	ρ=0.90	0.381	0.000	0.619	0.000
	ρ=0.50	0.830	0.000	0.170	0.000
	ρ=0.00	0.887	0.000	0.112	0.000
	ρ=–0.50	0.896	0.000	0.103	0.000
	ρ=–0.90	0.894	0.000	0.105	0.000
	ρ=–0.99	0.890	0.000	0.109	0.000
H_1 (δ>0)	ρ=0.99	0.340	0.000	0.659	0.000
	ρ=0.90	0.376	0.000	0.623	0.000
	ρ=0.50	0.759	0.000	0.241	0.000
	ρ=0.00	0.881	0.000	0.118	0.000
	ρ=–0.50	0.894	0.000	0.105	0.000
	ρ=–0.90	0.895	0.000	0.104	0.000
	ρ=–0.99	0.893	0.000	0.106	0.000
$H_3(i)$	ρ=0.99	0.155	0.234	0.137	0.473
	ρ=0.90	0.476	0.060	0.325	0.137
	ρ=0.50	0.812	0.000	0.187	0.000
	ρ=0.00	0.855	0.000	0.144	0.000
	ρ=–0.50	0.848	0.000	0.151	0.000
	ρ=–0.90	0.408	0.347	0.208	0.035
	ρ=–0.99	0.051	0.817	0.037	0.094
$H_3(ii)$	ρ=0.99	0.008	0.331	0.012	0.647
	ρ=0.90	0.000	0.376	0.000	0.623
	ρ=0.50	0.000	0.759	0.000	0.241
	ρ=0.00	0.000	0.881	0.000	0.118
	ρ=–0.50	0.000	0.894	0.000	0.105
	ρ=–0.90	0.000	0.895	0.000	0.104
	ρ=–0.99	0.000	0.893	0.000	0.106
$H_3(iii)$	ρ=0.99	0.058	0.318	0.026	0.596
	ρ=0.90	0.000	0.380	0.000	0.619
	ρ=0.50	0.000	0.760	0.000	0.239
	ρ=0.00	0.000	0.882	0.000	0.117
	ρ=–0.50	0.000	0.894	0.000	0.105
	ρ=–0.90	0.000	0.897	0.000	0.102
	ρ=–0.99	0.029	0.855	0.012	0.102

[1] See notes to Table 1A.1. The entries in this table can be interepreted as follows. Consider the first entry in the upper left corner of Table 2A.1, which is 0.334. This entry denotes the frequency of times that Steps 1-3 of our procedure result in the selection of H_1 or $H_3(i)$, given that actual data are generated according to $H_1(\delta=0)$, and we interpret it as the "empirical power" of our entire procedure. Analogously, the empirical power of the ADF test is equal to the sum of the 3rd and 4th columns of the table (see above discussion). Put another way, the data are first subjected to Step 1 of our procedure, and either H_A or H_B is selected. Then, given the appropriate data transformation implied by Step 1, an ADF test is performed, and the data are further classified as either I(0) or I(1). ADF tests are performed as discussed above. Each entry is based on 5000 Monte Carlo replications.

Table 2A.2: Complete Procedure Performance Using ADF for Steps 2-3[1]

Hypotheses Examined: H_1 and H_3. Sample Size = 100 Observations

Data Generated According To:	Parameter Values	Model Selected: Short Memory H_1 or $H_3(i)$	Short Memory $H_3(ii)$ or $H_3(iii)$	Integrated (I(1)) H_2	Integrated (I(1)) $H_4(i)$–(iii)
H_1 (δ=0)	ρ=0.99	0.205	0.000	0.795	0.000
	ρ=0.90	0.409	0.000	0.590	0.000
	ρ=0.50	0.925	0.000	0.074	0.000
	ρ=0.00	0.949	0.000	0.050	0.000
	ρ=–0.50	0.955	0.000	0.044	0.000
	ρ=–0.90	0.961	0.000	0.038	0.000
	ρ=–0.99	0.961	0.000	0.039	0.000
H_1 (δ>0)	ρ=0.99	0.213	0.000	0.786	0.000
	ρ=0.90	0.370	0.000	0.629	0.000
	ρ=0.50	0.932	0.000	0.068	0.000
	ρ=0.00	0.957	0.000	0.042	0.000
	ρ=–0.50	0.963	0.000	0.036	0.000
	ρ=–0.90	0.968	0.000	0.031	0.000
	ρ=–0.99	0.966	0.000	0.033	0.000
$H_3(i)$	ρ=0.99	0.179	0.136	0.108	0.575
	ρ=0.90	0.590	0.044	0.229	0.135
	ρ=0.50	0.891	0.000	0.108	0.000
	ρ=0.00	0.922	0.000	0.078	0.000
	ρ=–0.50	0.921	0.000	0.078	0.000
	ρ=–0.90	0.574	0.279	0.134	0.011
	ρ=–0.99	0.035	0.915	0.011	0.037
$H_3(ii)$	ρ=0.99	0.001	0.213	0.000	0.785
	ρ=0.90	0.000	0.370	0.000	0.629
	ρ=0.50	0.000	0.932	0.000	0.068
	ρ=0.00	0.000	0.957	0.000	0.042
	ρ=–0.50	0.000	0.963	0.000	0.036
	ρ=–0.90	0.000	0.968	0.000	0.031
	ρ=–0.99	0.000	0.966	0.000	0.033
$H_3(iii)$	ρ=0.99	0.073	0.196	0.016	0.714
	ρ=0.90	0.000	0.365	0.000	0.634
	ρ=0.50	0.000	0.931	0.000	0.068
	ρ=0.00	0.000	0.957	0.000	0.042
	ρ=–0.50	0.000	0.963	0.000	0.036
	ρ=–0.90	0.000	0.968	0.000	0.031
	ρ=–0.99	0.057	0.894	0.016	0.031

[1] See notes to Table 2A.1.

Table 2B.1: Complete Procedure Performance Using ADF for Steps 2-3[1]

Hypotheses Examined: H_2 and H_4. Sample Size = 50 Observations.

Data Generated According To:	Parameter Values	Model Selected: Short Memory H_1 or $H_3(i)$	Short Memory $H_3(ii)$ or $H_3(iii)$	Integrated (I(1)) H_2	Integrated (I(1)) $H_4(i)$–(iii)
H_2 (δ>0)	λ=0.0001	0.887	0.000	0.113	0.000
	λ=0.01	0.855	0.000	0.144	0.000
	λ=1.0	0.576	0.000	0.423	0.000
	λ=100	0.408	0.000	0.591	0.000
	λ=10000	0.406	0.000	0.593	0.000
$H_4(i)$ (δ=0)	λ=0.0001	0.855	0.000	0.144	0.000
	λ=0.01	0.813	0.000	0.185	0.0000
	λ=1.0	0.119	0.397	0.071	0.411
	λ=100	0.001	0.346	0.000	0.652
	λ=10000	0.001	0.346	0.000	0.652
$H_4(ii)$ (δ>0)	λ=0.0001	0.000	0.878	0.000	0.121
	λ=0.01	0.000	0.845	0.000	0.154
	λ=1.0	0.040	0.458	0.031	0.470
	λ=100	0.001	0.343	0.000	0.655
	λ=10000	0.000	0.344	0.000	0.654
$H_4(iii)$ (δ<0)	λ=0.0001	0.000	0.880	0.000	0.119
	λ=0.01	0.000	0.844	0.000	0.155
	λ=1.0	0.085	0.445	0.030	0.438
	λ=100	0.001	0.349	0.000	0.649
	λ=10000	0.001	0.344	0.000	0.654

[1] See notes to Table 2A.1. This table is analogous to Table 2A.1, except that empirical size is presented, rather than empirical power, given that all data used in the experiments reported on in this two table are generated according to I(1) processes. For example, consider the row of entries corresponding to data generated according to $H_4(ii)(\delta>0)$, and $\lambda=1.0$. The entries are 0.040, 0.458, 0.031, and 0.470. In this example, the empirical size of the ADF test is 0.040+0.458=0.498. Also, the empirical size of our entire procedure is 0.040+0.458+0.031=0.529. Put another way, the data are first subjected to Step 1 of our procedure, and either H_A or H_B is selected. Then, given the appropriate data transformation implied by Step 1, an ADF test is performed, and the data are further classified as either I(0) or I(1). ADF tests are performed as discussed above. Each entry is based on 5000 Monte Carlo replications.

Table 2B.2: Complete Procedure Performance Using ADF for Steps 2-3[1]

Hypotheses Examined: H_2 and H_4. Sample Size = 100 Observations.

Data Generated According To:	Parameter Values	Model Selected: Short Memory		Integrated (I(1))	
		H_1 or $H_3(i)$	$H_3(ii)$ or $H_3(iii)$	H_2	$H_4(i)$–(iii)
H_2 (δ>0)	λ=0.0001	0.956	0.000	0.043	0.000
	λ=0.01	0.879	0.000	0.120	0.000
	λ=1.0	0.397	0.000	0.602	0.000
	λ=100	0.282	0.000	0.717	0.000
	λ=10000	0.285	0.000	0.715	0.000
$H_4(i)$ (δ=0)	λ=0.0001	0.917	0.000	0.083	0.000
	λ=0.01	0.810	0.000	0.189	0.000
	λ=1.0	0.149	0.250	0.064	0.535
	λ=100	0.002	0.216	0.000	0.781
	λ=10000	0.001	0.212	0.000	0.786
$H_4(ii)$ (δ>0)	λ=0.0001	0.000	0.956	0.000	0.043
	λ=0.01	0.000	0.878	0.000	0.121
	λ=1.0	0.031	0.317	0.018	0.633
	λ=100	0.002	0.214	0.000	0.783
	λ=10000	0.001	0.211	0.000	0.787
$H_4(iii)$ (δ<0)	λ=0.0001	0.000	0.956	0.000	0.043
	λ=0.01	0.000	0.877	0.000	0.122
	λ=1.0	0.068	0.302	0.015	0.614
	λ=100	0.001	0.212	0.000	0.786
	λ=10000	0.001	0.208	0.000	0.790

[1] See notes to Table 2B.1

Table 3A.1: Complete Procedure Performance Using KPSS for Steps 2-3[1]

Hypotheses Examined: H_1 and H_3. Sample Size = 50 Observations

Data Generated According To:	Parameter Values	$\hat{\eta}_\mu$ $l0$	$\hat{\eta}_\mu$ $l4$	$\hat{\eta}_\mu$ $l12$	$\hat{\eta}_\tau$ $l0$	$\hat{\eta}_\tau$ $l4$	$\hat{\eta}_\tau$ $l12$
H_1 (δ=0)	ρ=0.99	0.957	0.685	0.290	0.973	0.635	0.179
	ρ=0.90	0.888	0.454	0.109	0.957	0.506	0.112
	ρ=0.50	0.347	0.103	0.029	0.500	0.110	0.054
	ρ=0.00	0.052	0.040	0.060	0.360	0.040	0.048
	ρ=−0.50	0.001	0.000	0.880	0.060	0.018	0.037
	ρ=−0.90	0.000	0.011	0.000	0.000	0.520	0.024
	ρ=−0.99	0.000	0.115	0.000	0.000	0.660	0.004
H_1 (δ>0)	ρ=0.99	0.966	0.732	0.397	0.973	0.635	0.179
	ρ=0.90	0.954	0.742	0.369	0.957	0.506	0.112
	ρ=0.50	1.000	0.999	0.952	0.500	0.118	0.054
	ρ=0.00	1.000	1.000	1.000	0.053	0.040	0.048
	ρ=−0.50	1.000	1.000	1.000	0.000	0.880	0.037
	ρ=−0.90	0.996	1.000	1.000	0.000	0.075	0.024
	ρ=−0.99	0.436	0.999	0.763	0.000	0.646	0.004
$H_3(ii)$	ρ=0.99	0.967	0.742	0.413	0.974	0.645	0.196
	ρ=0.90	0.954	0.742	0.369	0.957	0.506	0.112
	ρ=0.50	1.000	0.999	0.952	0.500	0.118	0.054
	ρ=0.00	1.000	1.000	1.000	0.053	0.040	0.048
	ρ=−0.50	1.000	1.000	1.000	0.000	0.880	0.037
	ρ=−0.90	0.996	1.000	1.000	0.000	0.075	0.024
	ρ=−0.99	0.436	0.999	0.763	0.000	0.646	0.004
$H_3(iii)$	ρ=0.99	0.963	0.764	0.441	0.975	0.664	0.245
	ρ=0.90	0.951	0.748	0.375	0.957	0.506	0.112
	ρ=0.50	1.000	0.999	0.950	0.500	0.118	0.054
	ρ=0.00	1.000	1.000	1.000	0.053	0.040	0.048
	ρ=−0.50	1.000	1.000	1.000	0.000	0.880	0.037
	ρ=−0.90	0.995	1.000	1.000	0.000	0.075	0.024
	ρ=−0.99	0.478	0.999	0.782	0.042	0.640	0.046

[1] See notes to Table 2A.1. Reported frequencies are based on the same set of experiments reported on in Tables 2A.1-2A.2 and 2B.1-2B.2. However, there are a number of differences in the manner in which the findings are reported. First, entries are based on 6 different KPSS unit root test statistics - $\hat{\nu}_\nu$ and $\hat{\nu}_\tau$, for $l0$, $l4$, and $l12$, rather than on one ADF unit root test. Second, entries correspond to the empirical size of the entire procedure. For example, the upper left entry, which is 0.957, suggests that when data are generated according to $H_1(\delta=0)$, the probability of finding that the actual data are generated by either H_2, H_3, or H_4 is 0.957. This entry corresponds to the frequency of times that the correct DGP (i.e. the DGP according to which the experimental data were generated) was chosen based on the entire procedure (Steps 1-3). Put another way, the data are first subjected to Step 1 of our procedure, and either H_A or H_B is selected. Then, given the appropriate data transformation implied by Step 1, a KPSS test is performed, and the data are further classified as either I(0) or I(1). KPSS tests are performed as discussed above. Each entry is based on 5000 Monte Carlo replications.

Table 3A.2: Complete Procedure Performance Using KPSS for Steps 2-3[1]

Hypotheses Examined: H_1 and H_3. Sample Size = 100 Observations

Data Generated According To:	Parameter Values	$\hat{\eta}_\mu$ $l0$	$\hat{\eta}_\mu$ $l4$	$\hat{\eta}_\mu$ $l12$	$\hat{\eta}_\tau$ $l0$	$\hat{\eta}_\tau$ $l4$	$\hat{\eta}_\tau$ $l12$
H_1 (δ=0)	ρ=0.99	0.993	0.793	0.505	0.998	0.815	0.393
	ρ=0.90	0.944	0.449	0.170	0.993	0.597	0.195
	ρ=0.50	0.356	0.096	0.043	0.543	0.110	0.051
	ρ=0.00	0.049	0.040	0.029	0.053	0.045	0.032
	ρ=–0.50	0.000	0.017	0.017	0.000	0.020	0.022
	ρ=–0.90	0.000	0.000	0.003	0.000	0.000	0.005
	ρ=–0.99	0.000	0.000	0.000	0.000	0.000	0.000
H_1 (δ>0)	ρ=0.99	0.995	0.879	0.700	0.998	0.815	0.393
	ρ=0.90	1.000	0.994	0.951	0.993	0.597	0.195
	ρ=0.50	1.000	1.000	1.000	0.543	0.110	0.051
	ρ=0.00	1.000	1.000	1.000	0.053	0.045	0.032
	ρ=–0.50	1.000	1.000	1.000	0.000	0.020	0.022
	ρ=–0.90	1.000	1.000	1.000	0.000	0.000	0.005
	ρ=–0.99	0.968	1.000	1.000	0.000	0.000	0.000
$H_3(ii)$	ρ=0.99	0.995	0.880	0.701	0.998	0.816	0.394
	ρ=0.90	1.000	0.994	0.951	0.993	0.597	0.195
	ρ=0.50	1.000	1.000	1.000	0.543	0.110	0.051
	ρ=0.00	1.000	1.000	1.000	0.053	0.045	0.032
	ρ=–0.50	1.000	1.000	1.000	0.000	0.020	0.022
	ρ=–0.90	1.000	1.000	1.000	0.000	0.000	0.005
	ρ=–0.99	0.968	1.000	1.000	0.000	0.000	0.000
$H_3(iii)$	ρ=0.99	0.997	0.886	0.702	0.998	0.831	0.448
	ρ=0.90	0.999	0.994	0.955	0.993	0.597	0.195
	ρ=0.50	1.000	1.000	1.000	0.543	0.110	0.051
	ρ=0.00	1.000	1.000	1.000	0.053	0.045	0.032
	ρ=–0.50	1.000	1.000	1.000	0.000	0.020	0.022
	ρ=–0.90	1.000	1.000	1.000	0.000	0.000	0.005
	ρ=–0.99	0.971	1.000	1.000	0.073	0.073	0.074

[1] See notes to Table 3A.1.

Table 3B.1: Complete Procedure Performance Using KPSS for Steps 2-3[1]

Hypotheses Examined: H_2 and H_4. Sample Size = 50 Observations.

Data Generated According To:	Parameter Values	$\hat{\eta}_\mu$ $l0$	$\hat{\eta}_\mu$ $l4$	$\hat{\eta}_\mu$ $l12$	$\hat{\eta}_\tau$ $l0$	$\hat{\eta}_\tau$ $l4$	$\hat{\eta}_\tau$ $l12$
H_2 (δ>0)	λ=0.0001	1.000	1.000	1.000	0.052	0.040	0.046
	λ=0.01	1.000	1.000	1.000	0.137	0.104	0.068
	λ=1.0	1.000	1.000	1.000	0.907	0.580	0.171
	λ=100	1.000	1.000	1.000	0.969	0.621	0.174
	λ=10000	1.000	1.000	1.000	0.970	0.621	0.173
$H_4(i)$ (δ=0)	λ=0.0001	0.000	0.000	0.000	0.000	0.000	0.000
	λ=0.01	0.000	0.060	0.000	0.000	0.000	0.000
	λ=1.0	0.744	0.542	0.246	0.736	0.472	0.140
	λ=100	0.954	0.689	0.319	0.968	0.620	0.174
	λ=10000	0.954	0.688	0.319	0.968	0.620	0.173
$H_4(ii)$ (δ>0)	λ=0.0001	1.000	1.000	1.000	0.052	0.040	0.046
	λ=0.01	0.999	0.999	0.998	0.137	0.104	0.068
	λ=1.0	0.870	0.671	0.368	0.841	0.541	0.160
	λ=100	0.955	0.690	0.318	0.967	0.620	0.174
	λ=10000	0.955	0.689	0.318	0.969	0.620	0.173
$H_4(iii)$ (δ<0)	λ=0.0001	1.000	1.000	1.000	0.052	0.040	0.046
	λ=0.01	1.000	1.000	0.999	0.137	0.104	0.068
	λ=1.0	0.826	0.655	0.383	0.803	0.514	0.151
	λ=100	0.952	0.686	0.320	0.968	0.620	0.174
	λ=10000	0.951	0.687	0.319	0.968	0.620	0.173

[1] See notes to Table 3A.1. Reported frequencies are based on the same set of experiments reported on in Tables 2A.1-2A.2 and 2B.1-2B.2. However, there are a number of differences in the manner in which the findings are reported. First, entries are based on 6 different KPSS unit root test statistics - $\hat{\nu}_\nu$ and $\hat{\nu}_\tau$, for $l0$, $l4$, and $l12$, rather than on one ADF unit root test. Second, entries correspond to the empirical power of the entire procedure. For example, the upper left entry, which is 1.000, suggests that when the data are generated according to $H_2(\delta>0)$, the probability of finding that the actual data are generated by H_2 is 1.000. This entry corresponds to the frequency of times that the correct DGP (i.e. the DGP according to which the experimental data were generated) was chosen based on the entire procedure (Steps 1-3). Put another way, the data are first subjected to Step 1 of our procedure, and either H_A or H_B is selected. Then, given the appropriate data transformation implied by Step 1, a KPSS test is performed, and the data are further classified as either I(0) or I(1). KPSS tests are performed as discussed above. Each entry is based on 5000 Monte Carlo replications.

Table 3B.2: Complete Procedure Performance Using KPSS for Steps 2-3[1]

Hypotheses Examined: H_2 and H_4. Sample Size = 100 Observations.

Data Generated According To:	Parameter Values	$\hat{\eta}_\mu$ $l0$	$\hat{\eta}_\mu$ $l4$	$\hat{\eta}_\mu$ $l12$	$\hat{\eta}_\tau$ $l0$	$\hat{\eta}_\tau$ $l4$	$\hat{\eta}_\tau$ $l12$
H_2 (δ>0)	λ=0.0001	1.000	1.000	1.000	0.058	0.048	0.037
	λ=0.01	1.000	1.000	1.000	0.354	0.273	0.167
	λ=1.0	1.000	1.000	1.000	0.990	0.798	0.398
	λ=100	1.000	1.000	1.000	0.998	0.816	0.405
	λ=10000	1.000	1.000	1.000	0.998	0.815	0.405
$H_4(i)$ (δ=0)	λ=0.0001	0.000	0.000	0.000	0.000	0.000	0.000
	λ=0.01	0.000	0.000	0.000	0.000	0.000	0.000
	λ=1.0	0.775	0.633	0.431	0.779	0.632	0.319
	λ=100	0.989	0.806	0.553	0.995	0.814	0.405
	λ=10000	0.991	0.808	0.554	0.997	0.814	0.404
$H_4(ii)$ (δ>0)	λ=0.0001	1.000	1.000	1.000	0.058	0.048	0.037
	λ=0.01	1.000	1.000	1.000	0.354	0.273	0.167
	λ=1.0	0.941	0.818	0.640	0.941	0.759	0.379
	λ=100	0.991	0.806	0.554	0.996	0.815	0.405
	λ=10000	0.992	0.807	0.554	0.997	0.815	0.405
$H_4(iii)$ (δ<0)	λ=0.0001	1.000	1.000	1.000	0.058	0.048	0.037
	λ=0.01	1.000	1.000	1.000	0.354	0.273	0.167
	λ=1.0	0.908	0.800	0.660	0.908	0.732	0.365
	λ=100	0.993	0.805	0.564	0.997	0.815	0.404
	λ=10000	0.993	0.806	0.565	0.996	0.814	0.405

[1] See notes to Table 3B.1.

Table 4: Levels Versus Logs and I(1) Versus I(0): An Re-examination of the Nelson-Plosser Dataset[1]

Variable	(V_{1T},V_{2T})	Correct Transformation	ADF Test Results Correct $(\hat{\tau}_\tau, lags, I(?))$	ADF Test Results Incorrect $(\hat{\tau}_\tau, lags, I(?))$	KPSS Test Results Correct $(\hat{\eta}_{\tau,l4}, I(?), \hat{\eta}_{\tau,l12}, I(?))$	KPSS Test Results Incorrect $(\hat{\eta}_{\tau,l4}, I(?), \hat{\eta}_{\tau,l12}, I(?))$
cpi	(8.080e-12, 0.0211)	levels	(1.050, 2, I(1))	(-2.539, 5, I(1))	(0.404, I(1), 0.178, I(1))	(0.350, I(1), 0.156, I(1))
employment	(1.874e-22, 29.43)	logs	(-3.802, 1, I(0))	(-0.274, 9, I(1))	(0.166, I(1), 0.111, I(0))	(0.393, I(1), 0.183, I(1))
gnp deflator	(4.670e-12, 0.0226)	levels	(0.212, 1, I(1))	(-2.079, 1, I(1))	(0.421, I(1), 0.182, I(1))	(0.359, I(1), 0.158, I(1))
industrial prod.	(1.147e-10, 0.0015)	levels	(-0.975, 2, I(1))	(-2.466, 9, I(1))	(0.507, I(1), 0.206, I(1))	(0.151, I(1), 0.094, I(0))
gnp per capita	(2.636e-16, 0.1989)	levels	(-3.011, 1, I(1))	(-3.821, 1, I(0))	(0.370, I(1), 0.183, I(1))	(0.144, I(0), 0.098, I(0))
interest rate	(1.239e-05, 7.362e-05)	levels	(0.485, 9, I(1))	(-1.594, 0, I(1))	(0.441, I(1), 0.185, I(1))	(0.466, I(1), 0.182, I(1))
money	(4.134e-15, 0.6707)	levels	(-0.753, 6, I(1))	(-3.692, 8, I(0))	(0.407, I(1), 0.183, I(1))	(0.219, I(1), 0.114, I(0))
nominal gnp	(3.287e-28, 7.523e+06)	logs	(-2.984, 6, I(1))	(1.244, 9, I(1))	(0.341, I(1), 0.154, I(1))	(0.423, I(1), 0.186, I(1))
real gnp	(2.598e-14, 0.0210)	levels	(0.013, 9, I(1))	(-3.733, 1, I(0))	(0.504, I(1), 0.208, I(1))	(0.168, I(1), 0.110, I(0))
real wages	(2.690e-09, 3.096e-05)	levels	(-1.749, 1, I(1))	(-1.367, 1, I(1))	(0.222, I(1), 0.105, I(0))	(0.263, I(1), 0.130, I(0))
s&p500	(1.930e-10, 0.0250)	levels	(0.413, 4, I(1))	(-2.838, 1, I(1))	(0.384, I(1), 0.201, I(1))	(0.301, I(1), 0.148, I(1))
unemployment	(3.038e-06, 0.0015)	levels	(-3.503, 6, I(0))	(-3.845, 10, I(0))	(0.117, I(0), 0.073, I(0))	(0.086, I(0), 0.062, I(0))
velocity	(0.0014, 2.630e-07)	levels	(-3.206, 1, I(1))	(-3.255, 1, I(1))	(0.437, I(1), 0.188, I(1))	(0.425, I(1), 0.187, I(1))
wages	(1.678e-19, 87.57)	logs	(-3.787, 10, I(0))	(1.420, 10, I(1))	(0.283, I(1), 0.134, I(0))	(0.441, I(1), 0.189, I(1))

[1] Variables reported on are the same as those used by Schotman and van Dijk (1991), who updated the original Nelson-Plosser (1982) dataset. The sample size considered is 80 observations (1909-1988). The V_{1T} and V_{2T} statistics reported are used as in Step 1 of the procedure outlined in Section 2 above. For example, when both statistics values are less than unity, we choose H_A - the series is either (i) I(0) or I(1) in levels, or (ii) I(0) in logs around a zero deterministic trend. Otherwise, H_B is chosen. We take the approach of denoting the transformation associated with the selection of H_A as "levels", although it should be noted that if the series turned out to be I(0), then a log I(0) with no linear deterministic trend representation may also be appropriate. This is the case for unemployment. Summarizing this argument: First, for all series that are found to be I(1), the "levels" label is appropriate Second, for series that are found to be I(0), a further a Cox-type test may indicate whether the series is better modeled using levels data, or using a logged data with no linear deterministic trend. ADF and KPSS statistics are also reported, both for the "correct" transformation based on our procedure, and for the "incorrect" transformation. All ADF test statistics are reported for ADF regressions with a constant and a trend ($\hat{\tau}_\tau$ - 5% CV is -3.465). For the ADF test statistics, the number of lagged differences used (given as "lags" was selected by examining the appropriate t-statistics, starting with 10 lags, and stopping when the last lagged difference had a coefficient significantly different from zero at a 95% level of confidence. KPSS statistics are reported for two of the three lag truncation parameters used in our simulations - $l4$ and $l12$ (see above), and for the case in which a deterministic trend is included in the test statistic construction ($\hat{\eta}_\tau$ - 5% CV is 0.146). For all of the reported test statistics, classification of the variables as either I(1) or I(0) is based on the above 5% critical values.