The Effect of Data Transformation on Common Cycle, Cointegration, and Unit Root Tests: Monte Carlo Results and a Simple Test*

Valentina Corradi and Norman R. Swanson[1]

[1]University of Exeter and Rutgers University, respectively

September 2002

Abstract

In the conduct of empirical macroeconomic research, unit root, cointegration, common cycle, and related tests statistics are often constructed using logged data, even though there is often no clear reason, at least from an empirical perspective, why logs should be used rather than levels. Unfortunately, it is also the case that standard data transformation tests, such as those based on the Box-Cox transformation, cannot be shown to be consistent unless an assumption is made concerning whether the series being examined is $I(0)$ or $I(1)$, so that a sort of circular testing problem exists. In this paper, we address two quite different but related issues that arise in the context of data transformation. First, we address the circular testing problem that arises when choosing data transformation and the order of integratedness. In particular, we propose a simple randomized procedure, coupled with sample conditioning, for choosing between levels and log-levels specifications in the presence of deterministic and/or stochastic trends. Second, we note that even if pre-testing is not undertaken to determine data transformation, it is important to be aware of the impact that incorrect data transformation has on tests frequently used in empirical works. For this reason, we carry out a series of Monte Carlo experiments illustrating the rather substantive effect that incorrect transformation can have on the finite sample performance of common feature and cointegration tests. These Monte Carlo findings underscore the importance of either using economic theory as a guide to data transformation and/or using econometric tests such as the one discussed in this paper as aids when choosing data transformation.

JEL classification: C12, C22.
Keywords: deterministic trend, nonlinear transformation, nonstationarity, randomized procedure.

---

* Valentina Corradi, Department of Economics, University of Exeter, Streatham Court, Exeter EX4 4PU, U.K., v.corradi@exeter.ac.uk. Norman R. Swanson, Department of Economics, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA, nswanson@econ.rutgers.edu. This paper was prepared for the conference on common features in Rio de Jenario in 2002, and the authors would like to thank the coordinators, Heather Anderson, João Issler, and Farshid Vahid for their organizational efforts. In addition, the many useful comments received from the organizers and conference participants were also much appreciated. Finally, the authors wish to thank Marcelo Fernandez and Hashem Pesaran for helpful comments on this version, as well as Graham Elliot, Liudas Giraitis, Clive Granger, Vassilis Hajivassiliou, Javier Hidalgo, Nick Keifer, Yongcheol Shin, and Andy Snell, for comments on an earlier version of this paper.

1 Introduction

In the conduct of empirical macroeconomic research, unit root, cointegration, common cycle, and related tests are often constructed using logged data. This is consistent with much of the real business cycle literature (see e.g. Long and Plosser (1983), King, Plosser and Rebelo (1988(a)(b)) and King, Plosser, Stock, and Watson (1991)), where it is suggested, for example, that GDP should be modeled in logs, given an assumption that output is generated according to a Cobb-Douglas production function. More recently, Engle and Issler (1995), Issler and Vahid (2001), and Vahid and Issler (2002) also show that the real business cycles models cited above can generate both common trends and common cycles for the log variables. Therefore, from an economic theory point of view, there is a clear justification for running unit root, cointegration and common cycle tests using loglinear models. Nevertheless, it is not always obvious by simply inspecting the data, for example, which transformation is 'appropriate', when modeling economic data (see e.g. Figure 1). Put another way, from a purely empirical point of view (e.g. if one were constructing prediction models using data-mining techniques) it is not obvious which data transformation should be used. This distinction is important because there is an implicit assumption in much of the current literature that transformation done prior to application of tests (such as cointegration and common cycle tests) will not affect the limiting distribution of the test under the null hypothesis. Consider as a case in point the common cycle test of Vahid and Engle (1993). The asymptotic behavior of their test is based on the fact that the (smallest) canonical correlations are $I(0)$ processes. However, this is the case if and only if the first differences of the series are also $I(0)$.[1] Now, it is well known (see e.g. Granger and Hallman (1991) and Corradi (1995)) that nonlinear transformations of $I(1)$ processes are no longer $I(1)$, and so the first difference of nonlinear transformations of $I(1)$ processes are no longer $I(0)$. It is in this sense that data transformation can pose a problem. In particular, under incorrect data transformation we do not know what the asymptotic behavior of our tests is. Unfortunately, it is also the case that standard data transformation tests, such as those based on the Box-Cox transformation, cannot be shown to be consistent unless an assumption is made concerning whether the series being examined is $I(0)$ or $I(1)$, so that a sort of circular testing problem exists. Furthermore, correct choice of data transformation is crucial when specifying

[1]The current convention is to define an integrated process of order d (say I(d), using the terminology of Engle and Granger (1987)) as one which has the property that the partial sum of the d^{th} difference, scaled by $T^{-1/2}$, satisfies a functional central limit theorem (FCLT).

forecasting models using integrated and/or cointegrated variables, as documented in Arino and Franses (2000) and in Chao, Corradi and Swanson (2001). Of course, this problem is not unique to unit root, cointegration, and common cycle tests; rather, we focus on these tests as they are so widely used in the current practice on macroeconometrics.

In this paper, we address two quite different but related issues that arise in the context of data transformation. First, we address the circular testing problem that arises when choosing data transformation and the order of integratedness. In particular, we propose a simple randomized procedure, coupled with sample conditioning, for choosing between levels and log-levels specifications in the presence of deterministic and/or stochastic trends. Second, we note that even if pre-testing is not undertaken to determine data transformation, it is important to be aware of the impact that incorrect data transformation has on frequently used tests in empirical macroecomics. For this reason, we carry out a series of Monte Carlo experiments illustrating the rather substantive effect that incorrect transformation can have on the finite sample performance of common feature and cointegration tests. These Monte Carlo findings underscore the importance of either using economic theory as a guide to data transformation and/or using econometric tests such as the one discussed in this paper as aids when choosing data transformation.

In recent years, the choice of data transformation for nonstationary series (henceforth, by nonstationary we mean I(1)) has received considerable attention in the literature. Important contributions in the area include De Bruin and Franses (1999), Franses and Koop (1998), Franses and McAleer (1998), Kobayashi (1994) and Kobayashi and McAleer (1999a,b). One line of research (see e.g. Franses and Koop (1998) and Franses and McAleer (1998)) analyzes the joint problem of choosing the Box-Cox transformation (with levels and logs being special cases) and choosing between stationarity and nonstationarity. These tests, for example, should be useful for addressing data transformation when the order of integratedness is unknown, and Monte Carlo results reported in the papers are rather encouraging. However, we believe that work still remains to be done before a complete picture of the asymptotic behavior of such tests based on Box-Cox transformations can be obtained. Broadly speaking, the main issue that arises when studying the limiting behavior of these and related tests (e.g. tests constructed under both nonstationarity and nonlinearity) can be summarized as follows. Often, test statistics can be written in "ratio" form, where the denominator of the test is an estimator of a (long run) variance. In such cases, a well defined limiting distribution can be derived under the null hypothesis. However, under the alternative hypothesis, it is often the case that both the numerator and the denominator approach infinity, with the latter diverging at a faster (or at least not slower) rate than the former. As a consequence, some tests have zero asymptotic power against alternatives of interest. This problem is solved in a rather ingenious way in a recent paper by Kobayashi and McAleer (KM: 1999a), who propose a test for distinguishing between levels and logs in models with a unit root. In particular, by assuming that the variance of the innovation process approaches zero at a sufficiently fast rate as the sample increases, KM derive the limiting distribution of their test under the null hypothesis, and show that the probability of type II error approaches zero asymptotically. While the assumption of a variance approaching zero is suitable for some financial series, it is in general unsuitable for macroeconomic time series. [2]

Our first objective in this paper is to propose an alternative procedure for distinguishing between the null hypothesis of a loglinear DGP and a (level) linear DGP. Once we have chosen the correct data transformation, we can proceed by running standard unit root or stationarity tests, as well as cointegration or common cycle tests. Two points are worth making at this juncture. First, when defining the relevant models from among which to choose, we allow for rather general, dependent error processes. Thus, the test is robust to a rich variety of dynamics. Second, we overcome the test consistency problem discussed above by basing our test on the combined use of a randomization procedure coupled with sample conditioning. In particular, we add randomness to our basic statistic, proceed by conditioning on the sample, and show that for all samples except a set of measure zero, the statistic has a chi-squared limiting distribution under the null hypothesis, while it diverges under the alternative hypothesis. The asymptotic behavior of the statistic is driven by the probability measure governing the added randomness. Nevertheless, conditional on the sample and for all samples except a set of measure zero, we choose the null hypothesis with probability approaching α whenever is true, and we reject the null hypothesis with probability approaching one whenever is false. We see randomization as a useful device when we cannot rely on standard asymptotic theory. A "common sense" drawback of randomization is that the outcome of an experiment can depend on the added randomness and so researchers sharing the same data set may arrive at different conclusions. Yet, randomization is a well known device in the statistical literature, tracking back at least at Pearson (1950), who uses randomization for dealing with inference for random variables with a discontinuous distribution. Sample conditioning instead occurs when performing bootstrap tests. Given the same sample, each researcher obtains the same

[2] The device that KM use is called small sigma asymptotics (see e.g. Bickel and Doksum (1981)).

numerical value for the actual statistic, but such a statistic is then compared with bootstrap quantiles which differ across researchers, even if based by resampling from the same sample. However, there is a substantial difference between bootstrap tests and the randomized procedure suggested in this paper. In fact, in the case of bootstrap tests as the sample size gets large, all researchers will eventually reach the same conclusion: all of them always reject the null when is false, all of them reject the null in $\alpha\%$ of the cases (samples), when is true. In our context instead, as the sample size gets large, all researchers always reject the null when false, while $\alpha\%$ of researchers always reject the null when is true.

In a series of Monte Carlo experiments, we establish that the finite sample properties of the suggested statistics are quite good for samples of at least 300 observations, for DGPs calibrated using U.S. monetary data. In addition, an empirical illustration is provided in which the King, Plosser, Stock and Watson (1991) data set is examined. Results suggest that output, consumption and investment are "best" modelled as loglinear, as predicted by standard real business cycles models, while money is better modeled as linear, at least when considering the full sample.

As mentioned above, our second objective is somewhat unrelated to the first. In particular, we wish to assess the importance, from a finite sample perspective, of wrongly imposing a particular data transformation. This is done in the context of common cycle and cointegration tests. Our findings suggest that the impact of incorrect data transformation is rather important in finite samples, as the test statistics tend to behave quite poorly in our experimental settings. This part of our paper, thus serves to highlight the importance of either relying on economic theory for choice of data selection, or undertaking the test for data transformation. In particular, if economic theory does not clearly dictate data transformation, then one should examine the data carefully, making as much use of available data transformation tests as possible.

The rest of the paper is organized as follows. Section 2 discusses the issue of data transformation in the context of the common cycle test of Vahid and Engle (1993), while Section 3 does the same for cointegration tests. Section 4 discusses problems involved with testing for unit roots in the presence of incorrect data transformation, and introduces the randomized statistic for selecting data transformation when the order of integratedness of the series is unknown. Monte Carlo findings pertaining to the different parts of the paper and contained in Sections 2, 3, and 4, and a small empirical illustration is also given in Section 4. Concluding remarks are gathered in Section 5. All proofs are collected in an appendix. Hereafter, $\overset{d^*}{\rightarrow}\ a.s. - \omega$ denotes convergence in distribution conditional on the sample, ω, $\forall\ \omega$ (i.e. for all sample except a set of measure zero).

2 Common Cycle Tests Under Incorrect Data Transformation

Assume that the objective is to carry out a test for the number of common cycles. If the investigator knows that all variables are $I(0)$, then the correct data transformation can be chosen via a Box-Cox transformation approach, and once the appropriate transformation is chosen a common feature test for serial correlation can be carried out along the lines suggested by Engle and Kozicki (1993). If the investigator knows that the series are loglinear or (level) linear, then (s)he can find the number of cointegrating vectors after deciding whether the series are all $I(1)$, for example, using unit root and Johansen (1988, 1991) cointegration tests. In addition, if all series are $I(1)$, a common feature test can be carried out using first differenced data, either along the lines of Engle and Kozicki (1993) or using the information-criteria approach recently proposed by Issler and Vahid (2002). Finally, in the presence of common trends, the number of common cycles can be ascertained via the approach suggested by Vahid and Engle (VE: 1993). In this section, our objective is to examine the effect of (incorrect) data transformation on the common cycle test proposed by VE.[3]

For sake of simplicity, we posit a simple DGP characterized by one common trend and one common cycle, as in Example 1 in VE. Let $Y_{i,t}$, $i = 1, 2$, denote the variable in levels and $y_{i,t}$ denote the natural logarithm of $Y_{i,t}$. Consider the following data generating process (DGP):

$$Y_{1,t} = Y_{1,0} + \sum_{j=1}^{t} \epsilon_j + \delta t + u_t \tag{1}$$

and

$$Y_{2,t} = Y_{2,0} + \beta \sum_{j=1}^{t} \epsilon_j + \beta\delta t + u_t, \tag{2}$$

where ϵ_j is $iid(0, \sigma_\epsilon^2)$ and $u_t = \rho u_{t-1} + \eta_t$ with $|\rho| < 1$ and η_t $iid(0, \sigma_\eta^2)$. It is immediate to see that

$$Y_{2,t} - \beta Y_{1,t} = Y_{2,0} - \beta Y_{1,0} + (1-\beta)u_t,$$

so that $Y_{2t} - \beta Y_{1t}$ is a stationary process, while

$$Y_{2,t} - Y_{1,t} = Y_{2,0} - Y_{1,0} + (\beta - 1)\sum_{j=1}^{t} \epsilon_j + (\beta - 1)\delta t.$$

[3]Such a test can be seen as an extension of the serial correlation common feature test of Engle and Kozicki (1993) to the case of $I(1)-$cointegrated variables.

Thus, $\Delta Y_{2,t} - \Delta Y_{1,t} = (\beta-1)\epsilon_t + (\beta-1)\delta$ cannot be predicted using the lags of the series. Therefore, according to the definition of VE, $\begin{pmatrix} 1 & -\beta \end{pmatrix}$ defines the cointegrating vector and $\begin{pmatrix} 1 & -1 \end{pmatrix}$ defines the cofeature vector, so that in this example we have exactly one common trend and one common cycle. Suppose we want to test the null of zero common cycles (i.e. the cofeature space, is of dimension s¿1) versus the alternative of one common cycle, conditional on the fact that there is one cointegrating vector. Following VE (pp. 349-350), the test statistic is given by $C(p,2) = -(T-p-1)\sum_{i=1}^{2}\log(1-\lambda_i^2)$, where λ_i^2 $i = 1,2$ are the two (smallest) canonical correlations between ΔY_t and $(\Delta Y_{t-1}, ..., \Delta Y_{t-p}, \widehat{Z}_{t-1})$, where $\Delta Y_t = \begin{pmatrix} \Delta Y_{2t} & \Delta Y_{1,t} \end{pmatrix}$, $\widehat{Z}_t = Y_{2t} - \widehat{\beta}_T Y_{1t}$, $\widehat{\beta}_T$ is an estimator of the cointegrating vector, and the number of lags, p, is chosen using a model selection procedure, say. If ΔY_t is a stationary process (and so the canonical correlations are stationary processes), and if $\widehat{\beta}_T$ is $T-$consistent for β, then $C(p,2)$ is χ^2_{4p+2} under the null.[4]

Now suppose that the DGP is as in equations (1) and (2), but we run the test using logged data. From (1) and (2) we see that $y_{1t} = \log\left(Y_{1,0} + \sum_{j=1}^{t}\epsilon_t + \delta t + u_t\right)$ and $y_{2t} = \log\left(Y_{2,0} + \beta\sum_{j=1}^{t}\epsilon_t + \beta\delta t + u_t\right)$. First, note that the existence of a cofeature vector in levels does not imply the existence of a cofeature vector in the logs; and the existence of a cointegrating vector in levels does not imply the existence of a cointegrating vector in logs. Second, note that the vector Δy_t is no longer stationary, and the limiting distribution of the cointegrating vector is no longer well defined.[5] Thus, in general, we cannot ascertain the limiting behavior of the statistic above, when we implement the test using logs instead of levels. In fact, the canonical correlations are linear combination of scaled sums of $\Delta y_{i,t}$, and so do not satisfy the central limit theorem unless $\Delta y_{i,t}$ is a short memory series.

Now assume that the DGP is:

$$y_{1,t} = y_{1,0} + \sum_{j=1}^{t}\epsilon_j + \delta t + u_t \tag{3}$$

and

$$y_{2,t} = y_{2,0} + \beta\sum_{j=1}^{t}\epsilon_j + \beta\delta t + u_t, \tag{4}$$

so that we have exactly one common trend and one common cycle in this loglinear DGP. Suppose that we perform the common feature test using levels instead of logs. Note that $Y_{1,t} = \exp y_{1,t} = \exp\left(y_{1,0} + \sum_{j=1}^{t}\epsilon_j + \delta t + u_t\right)$ and $Y_{2,t} = \exp y_{2,t} = \exp\left(y_{2,0} + \beta\sum_{j=1}^{t}\epsilon_j + \beta\delta t + u_t\right)$, so that we

[4] The number of degree of freedom (see VE (1993), pp. 349) is $s^2 + snp + sr - sn$. In the present context, $n = s = 2$ and $r = 1$.

[5] Granger and Hallman (1991) point out that cointegration between Y and X does not imply cointegration between $g(Y)$ and $g(X)$, for any nonlinear function g.

have neither a common trend nor a common cycle in the levels series. It also immediate to see that ΔY_t is not a stationary vector and also that the partial sums of ΔY_t will tend to diverge at an exponential rate. Therefore, the common cycles statistic does not have a well defined limiting distribution.

The above results are meant to serve as a reminder of the importance of data transformation. Of course, this does not mean that if theory suggests a particular data transformation, then we should not use it. Rather that we should consider constructing tests of data transformation that are robust to the order of integration of the data (as in Section 4) whenever we do not have priors concerning the appropriate data transformation. To illustrate the empirical relevance of the issue discussed above, the results of a small series of Monte Carlo experiments based on the above DGPs are reported in Table 1. In particular, we began by calibrating some simple random walk models using quarterly real and seasonally adjusted U.S. M2 for the period 1970:1-1994:1.[6] This particular series used was the money variable examined in King, Plosser, Stock and Watson (1991) and later updated by Corradi, Swanson and White (2000). The fitted models were: $\widehat{X}_t = 30.78 + X_{t-1}$, $\hat{\sigma} = 17.27$ and $\widehat{\log X}_t = 0.0188 + \log X_{t-1}$, $\hat{\sigma} = 0.0099$. Using these parameters as a guide, we specified levels DGPs according to (1) and (2) with $\sigma_\epsilon = 17.27$, $\delta = 30$, $Y_{1,0} = 100$, $Y_{2,0} = 200$, and $\sigma_u = \{0.1\sigma_\epsilon,\ 0.1\sigma_\epsilon,\ 10\sigma_\epsilon\}$. Data were also generated according to (3) and (4) with $\sigma_\epsilon = 0.01$, $\delta = 0.015$, $Y_{1,0} = 0.5$, $Y_{2,0} = 1$, and $\sigma_u = \{0.1\sigma_\epsilon,\ 0.1\sigma_\epsilon,\ 10\sigma_\epsilon\}$. In all experiments, we set $\rho = \{0.3,\ 0.6,\ 0.9\}$ and sample of $T = \{100,\ 250,\ 500\}$ observations were used. Also, the cointegrating rank was assumed to be unity (even when data were "incorrectly" transformed), lags were estimated using each of the Akaike and the Schwarz Information Criteria (AIC and SIC), and the cointegrating vector was estimated using the maximum likelihood procedure of Johansen (1988,1991). All entries in the table are rejection frequencies based on a sequential procedure. First, we tested the null hypothesis that $s > 0$. When this hypothesis failed to reject based on a 5% test, a second test of the hypothesis $s > 1$ was carried out. If this test resulted in rejection at the 5% level, coupled with failure to reject $s > 0$ in the first step of the procedure, then a rejection was recorded (i.e. evidence of a single common cycle was found). Under correct data transformation, then, we expect the rejection frequency to be high, while under incorrect data transformation the test is invalid and there is no common cycle, so that it is unclear what the findings of the test will be. It is worth noting that in Table 1, Panels A and

[6] Results based on models parameterized using U.S. GDP as a benchmark were also tabulated, and are available upon request.

B, the empirical rejection frequency of the common cycles test is very high, even for samples of only 100 observations. However, when $\rho = 0.9$, so that the persistence driving the cointegration is very high, then the VE test does not perform well for samples of 100 observations, although performance is dramatically improved when the sample is increased to 250 observations. Interestingly, the VE test often finds evidence of common cycles even when the data are incorrectly transformed (see the last 6 rows of entries in each panel), although this finding is far from robust as there are various σ_ϵ, σ_u combinations for which little evidence of common cycles is found. Thus, as expected, the VE test becomes suspect under incorrect data transformation. While this finding should come as no surprise, it does serve to underscore the importance of data transformation and testing for data transformation when the correct transformation is not suggested by economic theory.[7] Finally, our simple setup does not allow us to easily distinguish the trade-offs between using the AIC versus the SIC for lag selection. For an in-depth discussion of lag selection and related issues in the context of common cycle tests, the reader is referred to Vahid and Issler (2002).

Additionally, if the objective of the researcher is the construction of the "best" prediction model, where by best we have in mind the use of some loss function for comparing models such as the mean square forecast error (MSFE), then the researcher may want to rely just as much on empirical evidence (e.g. statistical tests) as on economic theory when choosing data transformation. The simple reason for this is that there are many competing theories (in macroeconomic for example), and certain theories may be supportive of different data transformations, while other theories may have nothing to say at all about data transformation. This issue is discussed further in Section 4 below, and is explored via Monte Carlo experimentation in Chao, Corradi and Swanson (2001), who find that incorrect data transformation can play havoc on the estimation and formulation of prediction models, in the sense that incorrect transformation can lead to models that are grossly misbehaved in the context of MSFE.

[7] Suppose we have a common cycle in logs (levels) and we run the test for the null of no common cycles, using levels (logs). Whichever conclusion we draw, it is bound to be incorrect. For example, if we reject we may conclude that there is no common cycle between the two series (although there really may be, between the "correctly" transformed series), while if we do not reject we may conclude that there is common cycle between the logged series, when instead there is no linear relation between the logged series at all (assuming that the correct data transformation is levels).

3 Cointegration Tests Under Incorrect Data Transformation

The above results concerning the validity of common cycle tests under incorrect data transformation also directly apply to the case of cointegration tests. For this reason, it may be useful to examine the finite sample performance of Johansen CI tests under incorrect data transformation. Using a setup similar to that of Chao, Corradi and Swanson (2001) our objective in this paper is to examine the finite sample behavior of the Johansen (1988,1991) cointegration test using data generated according to

$$\Delta Q_{1,t} = a + b(L)\Delta Q_{1,t-1} + cZ_{t-1} + \epsilon_t. \tag{5}$$

Data are generated according to the above vector error correction model, where $Q_{1,t} = (X_t, W_t')'$ is a vector if I(1) variables, W_t is an $n \times 1$ vector for $n = 2$, $Z_{t-1} = dQ_{1,t-1}$ is an $r \times 1$ vector of I(0) variables, r is the rank of the cointegrating space, d is an $r \times (n+1)$ matrix of cointegrating vectors, a is an $(n+1) \times 1$ vector, $b(L)$ is a matrix polynomial in the lag operator L, with p terms, each of which is an $(n+1) \times (n+1)$ matrix, p is the order of the VEC model, c is an $(n+1) \times r$ matrix, and ϵ_t is a vector error term. For DGPs generated as linear in levels, we report rejection frequencies for $a = (a_1, a_2)\prime$, $a_1 = a_2 = \{0.0, 0.1, 0.2\}$, $b = 0$, $c = (c_1, c_2)\prime$, $c_1 = -0.2, c_2 = \{0.0, 0.2, 0.4, 0.6\}$, and $\sigma^2_{\epsilon_i} = 1.0$, $i = 1, 2$. For loglinear DGPs, b and c are as above, $a_1 = a_2 = \{0.0, 0.01, 0.02\}$, and $\sigma^2_{\epsilon_i} = 0.09$, $i = 1, 2$. Results for other parameterizations examined are qualitatively similar, and are available upon request from the authors. The results of the experiment are quite straightforward, and are summarized in Table 3. First, the empirical size of the trace test statistic is severely upward biased, with bias increasing as T increases. Further, and as expected, the finite sample power (all cases where $c_2 \neq 0$) increases rapidly to unity as T increases. Thus, the null of no cointegration is over-rejected. Also, we know that estimators of cointegrating vectors are inconsistent under the wrong data transformation, even if the *true* cointegrating rank is known. Thus, we again have evidence that it should be useful to jointly test for integratedness and data transformation when both are uncertain either from a theoretical and/or from an empirical perspective. This is the subject to which we next turn our attention.

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

4.1 Set Up

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 around a linear deterministic trend, (2) $\log X_t$ is an I(0) process possibly 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. More precisely we want to choose among the following DGPs:

$H_1 : X_t = \alpha_0 + \delta_0 t + \rho X_{t-1} + \varepsilon_{1,t}$, $|\rho| < 1$ and $\delta_0 > 0$,

$H_2 : X_t = \delta_0 + X_{t-1} + \varepsilon_{1,t}$, $\delta_0 > 0$.

$H_3 : \log X_t = \alpha_1 + \delta_1 t + \rho \log X_{t-1} + \varepsilon_{2,t}$, $|\rho| < 1$ and $\delta_1 \geq 0$ and

$H_4 : \log X_t = \delta_1 + \log X_{t-1} + \varepsilon_{2,t}$, $\delta_1 \geq 0$.

Note that in order to ensure positivity we assume that the DGPs in levels have a positive trend component.

While it is easy to define a test that has a well defined distribution under one of $H_1 - H_4$, it is 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. In particular, construct the statistic for the null of stationarity proposed by Kwiatkowski, Phillips, Schmidt, and Shin (KPSS: 1992):

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

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 I(1) 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 $\log X_t = \delta_1 + \log X_{t-1} + \epsilon_t$, $\delta_1 > 0$ (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 $X_t = \exp(\log X_0 + \delta_1 t + \sum_{j=1}^{t} \epsilon_j)$. 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.

Recently Park and Phillips (PP: 1999, 2001) have developed an asymptotic theory for partial sums and for moments of nonlinear functions of integrated processes. The novel and important approach of Park and Phillips is based on the idea of replacing sample sums by spatial sums and then analyzing the average time spent by the process in the vicinity of given points. A key ingredient is the notion of local time of a Brownian motion. In our setup, we need to take into account the presence of a positive deterministic trend, at least for levels DGPs, however, and we are currently unable to generalize the PP results to the case of processes with deterministic drift components. The intuition behind the difficulty in providing such a generalization stems from the fact that we cannot embed an integrated process with deterministic drift into a continuous semimartingale[8], and to the best of our knowledge a local time theory is available only for continuous semimartingale processes. Broadly speaking, an integrated process with positive drift is dominated by the deterministic component and so it is transient. Thus, compact sets in the state space will be visited only a finite number of times, as the process will spend almost all time in the "proximity of infinity". Therefore, we shall follow a different approach, based on the combination of randomization and sample conditioning. In the sequel, in order to distinguish between H_1, H_2, H_3 and H_4 above, we rely on the following assumption:

Assumption A1: (i) $X_t > 0, \forall t \geq 0$, (ii) $\varepsilon_{i,t}, i = 1, 2$, is a zero-mean strictly stationary strong mixing process with mixing coefficient α_m satisfying $\sum_{m=0}^{\infty} \alpha_m^{\frac{\gamma}{4+2\gamma}} < \infty$, for any $\gamma > 0$, and (iii) $0 < E(\varepsilon_{i,1}^2) = \sigma_i^2 < \infty$ and $E(|\varepsilon_{i,1}|^{2(2+\gamma)}) < \infty$, $i = 1, 2$, for the same γ as in (ii).

Note that Assumption A1 suffices for the partial sums of $\{\varepsilon_{j,t}\}$ to satisfy a strong (and so a weak) invariance principle (see e.g. Corollary 4.1 and Theorem 3.1 in Berger (1990)).[9] As mentioned above, our main objective is to distinguish between levels and logs. This is because once we have chosen the correct data transformation, we can choose between $I(0)$ and $I(1)$ via standard tests. Now, group the above hypotheses as follows:

$H_0 : H_3 \cup H_4$, $\delta_1 > 0$

$H_A : H_1 \cup H_2$

[8] A semimartingale is a process given by the sum of a martingale plus an adaptive process of finite variation (see e.g. Revuz and Yor (1990), pp.121).

[9] The strict stationarity assumption can be relaxed at the price of strengthening the mixing condition. In fact, a strong invariance principle for strong mixing, non-stationary processes could be used (see e.g. Theorem 2 in Eberlain (1986)).

Thus, the null hypothesis is logs and the alternative is levels. The case of $\delta_1 = 0$, (i.e. no deterministic drift in the log DGPs) is somewhat more complex and will be treated subsequently. The proposed test statistic is:

$$S_{T,R}(\omega) = \int_U Z^2_{T,R}(u,\omega)\pi(u)du, \tag{6}$$

where U is a compact set on the real line, ω denotes the dependence of $S_{T,R}(\cdot)$ on the data, $\int_U \pi(u)du = 1$,

$$Z_{T,R}(u) = \frac{2}{\sqrt{R}}\sum_{i=1}^{R}\left(1\left\{V^i_{\xi,T}(\omega) \le u\right\} - \frac{1}{2}\right), \tag{7}$$

with $R = o(T)$, and $V^i_{\xi,T}(\omega)$ is defined as:

$$V^i_{\xi,T}(\omega) = \left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t}{\Delta X_1}\right)^2\right)^{1/2}\xi_i,\ \ i = 1,\cdots,R, \tag{8}$$

where ξ_i is an $iidN(0,1)$ random variable. Note that we divide all data by the initial value in order to make the statistic invariant to scalar multiplication of the observations. It turns out that for any sample, ω, which is a realization of a DGP under the null hypothesis (i.e. a log DGP), $S_{T,R}(\omega)$ converges in distribution to a χ^2 random variable, while for any ω which is a realization of a DGP under the alternative hypothesis (i.e. a level DGP), $S_{T,R}(\omega)$ diverges. Note that, as we proceed conditionally on the sample, the asymptotic behavior of the statistic is driven by the probability law governing the artificial randomness (i.e. the probability law governing ξ_i). Randomized procedures have previously been used in the literature, tracking back to Pearson (1950). For example, Dufour and Kiviet (1996) use a randomized test to obtain finite sample confidence intervals for structural changes in dynamic models; although in finite samples the level of the actual and of the randomized test may differ, they are equivalent in large samples. In a different context, Lütkepohl and Burda (1997) use a randomized approach for constructing Wald tests under non regular conditions - namely when the matrix of partial derivatives has reduced rank. They essentially overcome a certain singularity problem by adding randomness, and convergence to the limiting distribution is driven by both the probability law governing the sample and the probability law governing the added randomness. What differentiates our approach from the randomized procedures cited above is the joint use of randomization and sample conditioning. Our asymptotic result only holds conditionally on the sample, and for all samples except a set of measure zero. It is also worth noting, however, that randomization coupled with sample conditioning is used elsewhere to obtain conditional p-values and conditional percentiles, for example, when the limiting distribution of the actual statistic is data dependent (see e.g. Hansen (1996), Inoue (2001) and Corradi and Swanson (2002)). In these cases, though, inference is based on comparison of the actual statistic (which depends only on the sample) with conditional percentiles. In the present context, inference is based on the randomized statistic, conditional on the sample.

4.2 Asymptotic Results

Hereafter let d^* denote convergence in distribution according to P^*, the probability law governing ξ_i, $i = 1, ..., R$, conditional on the sample. Also, E^* and Var^* denote the mean and the variance operators with respect to the probability law P^*. Finally, the notation $a.s. - \omega$ means conditional on the sample, and for all samples except a set of measure zero.

Theorem 1: Let A1 hold. If $R = T^a$, $0 < a < 1$, then as $T \to \infty$:

(i) Under H_0, $S_{T,R}(\omega) \overset{d^*}{\to} \chi^2_1$, $a.s. - \omega$.

(ii) Under H_A, there exists a $\nu > 0$ such that $P^* \left[\frac{1}{R} S_{T,R}(\omega) > \nu \right] \to 1$, $a.s. - \omega$.

Thus, the test statistic has a well defined limiting distribution for each sample which is a realization of a DGP under H_0 and diverges for each sample which is a realization of a DGP under H_A.

It is worth noting that the interpretation of *test size* in the current context differs from the interpretation associated with inference which is not sample conditioned. To see this difference, consider the following example. Suppose we draw 10000 samples from a DGP generated under H_0. In addition, there are 10000 people performing the same test. According to the usual definition, the size is 5% if all 10000 people decide in favor of H_0 based on examination of 9500 samples, while they all decide in favor of H_A based on the remaining 500 samples. On the other hand, for the sample conditioned statistic, some group[10] of 9500 people decide in favor of H_0 for each of the 10000 samples, while the remaining 500 people decide in favor of the alternative for each sample.

Although a detailed proof of the theorem above is given in the appendix, it is perhaps worthwhile to give an intuitive explanation of the result. Note first that conditional on the sample, $V^i_{\xi,T}(\omega) \sim N\left(0, \frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t}{\Delta X_1}\right)^2\right)$. Now, note that under the null hypothesis of a log DGP, $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t}{\Delta X_1}\right)^2$ diverges to infinity at a geometric rate as T gets large. It then follows that $V^i_{\xi,T}(\omega)$ diverges almost surely to $+\infty$ or to $-\infty$, $a.s. - \omega, \forall i$. In addition, because of symmetry we have

[10]The members of the group may change from sample to sample.

that $V^i_{\xi,T}(\omega)$ diverges to either plus or minus infinity with probability approaching $1/2$, $a.s. - \omega$. Thus, $E^*(1\{V^i_{\xi,T}(\omega) \leq u\}) = P^*(V^i_{\xi,T}(\omega) \leq u) = \frac{1}{2} + o(1)$, uniformly in u for U compact, and $Var^*\left(\frac{1}{\sqrt{R}}\sum_{i=1}^R \left(1\{V^i_{\xi,T}(\omega) \leq u\}\right)\right) = \frac{1}{4} + o(1)$, uniformly in u, $a.s. - \omega$. The desired result then follows directly from the central limit theorem for independent triangular arrays.[11] Under the alternative hypothesis of a level DGP, by the strong law of large numbers, $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2$ converges almost surely to a constant, say M. Let $F(u)$ be the CDF of a $N(0, M)$ random variable, evaluated at u. Now,

$$\frac{1}{\sqrt{R}}\sum_{i=1}^R (1\left\{V^i_{\xi T}(\omega) \leq u\right\} - \frac{1}{2}) = \frac{1}{\sqrt{R}}\sum_{i=1}^R (1\left\{V^i_{\xi T}(\omega) \leq u\right\} - F(u)) + \sqrt{R}(F(u) - \frac{1}{2}). \qquad (9)$$

The first term on the right hand side above is bounded in probability, because of the central limit theorem for empirical processes for independent triangular arrays, while the second term diverges at rate $\sqrt{R}$ whenever $F(u) \neq 1/2$ (i.e. whenever $u \neq 0$).

In practice, the interval over which u is integrated must be determined. For increasing width intervals which are centered at zero and for $\pi(u)$ uniform over U, finite sample power improves, while finite sample size deteriorates. The dependence of finite sample power on U in this case can be seen immediately from equation (9), as the second term on the right hand increases the further is $|u|$ from zero. On the other hand, finite sample size tends to gets worse the larger is $|u|$. Hence, there is a trade-off between finite sample size and power associated with the choice of the interval U. In practice, we also have to choose R. It is easy to see that the higher is the rate at which R grows, provided it grows at a slower rate than T, the higher is the finite sample power. The choice of U and R is analyzed in the Monte Carlo section below.

We now turn to the case where $\delta_1 = 0$ (i.e. the case of log DGPs without a deterministic trend component). For example, under H_4, $\Delta X_t = X_{t-1}\exp(\varepsilon_{2,t} - 1)$, where $X_{t-1} = \exp(\log X_0 + \sum_{j=1}^{t-1}\varepsilon_{2,j})$. As $\sum_{j=1}^{t-1}\varepsilon_{2,j}$ diverges either to plus or minus infinity, it follows that $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2$ either diverges to infinity or converges to zero. Thus, $V^i_{\xi,T}(\omega)$ either diverges to $\pm\infty$ or converges to zero, depending on ω. Intuitively, if $V^i_{\xi,T}(\omega)$ converges to zero, $1\{V^i_{\xi,T}(\omega) \leq u\} \to 1$, for all $u > 0$, and $1\{V^i_{\xi,T}(\omega) \leq u\} \to 0$, for all $u < 0$. On the other hand, when $V^i_{\xi,T}(\omega)$ diverges to $\pm\infty, 1\{V^i_{\xi T}(\omega) \leq u\} \to 1$ (resp. 0) with probability $\frac{1}{2}$, $a.s. - \omega$, for all $u \in U, U$ compact. Needless to say, it is unknown whether $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2$ converges to zero or diverges, for any given sample. A natural approach is thus to construct two statistics and then base inference on the smaller one.

[11] Note that conditionally on sample, $V^i_{\xi,T}$, $i = 1, ..., R$ is an independent triangular array.

Without loss of generality, let U^+ be a compact set on the positive real line (including 0)[12]. Define:

$$S^a_{T,R}(\omega) = \int_{U^+} \left(\frac{2}{\sqrt{R}} \sum_{i=1}^{R} \left(1\left\{ V^i_{\xi,T}(\omega) \leq u \right\} - \frac{1}{2} \right) \right)^2 \pi(u) du$$

and

$$S^b_{T,R}(\omega) = \int_{U^+} \left(\frac{2}{\sqrt{R}} \sum_{i=1}^{R} \left(1\left\{ V^i_{\xi,T}(\omega) \leq u \right\} - p \right) \right)^2 \pi(u) du, p = 1,$$

where $\int_{U^+} \pi(u)du = 1$. Note that $S^a_{T,R}(\omega)$ is the same as $S_{T,R}(\omega)$ above, with the additional requirement that it is computed over U^+. The choice between logs and levels in this context is facilitated by using $\min(S^a_{T,R}(\omega), S^b_{T,R}(\omega))$. The intuition for this test is as follows. Conditioning on a sample for which $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t}{\Delta X_1}\right)^2 \to \infty$ implies that $S^a_{T,R}(\omega)$ is asymptotically χ^2_1, while $S^b_{T,R}(\omega)$ diverges. On the other hand, conditioning on a sample for which $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t}{\Delta X_1}\right)^2 \to 0$, implies that $S^a_{T,R}(\omega)$ diverges, while $S^b_{T,R}(\omega)$ converges in probability to zero. This suggests using the above test within the context of the following hypotheses,

$H'_0 : H_4$ with $\delta_1 = 0$ and

$H'_A : H_1 \cup H_2 \cup H_3$ with $\delta_1 = 0$.

Theorem 2: Let Assumption A1 hold. If $R = T^a$, $0 < a < 1$, then as $T, R \to \infty$:

(i) Under H'_0, $\lim_{T,R\to\infty} P^* \left[\min(S^a_{T,R}(\omega), S^b_{T,R}(\omega)) > c_\beta\right] \leq \beta$, $a.s. - \omega$, where c_β is the $(1-\beta)th$ percentile of a χ^2_1 random variable.

(ii) If in addition, $E\left(\exp\left(2(2+\phi)\sum_{j=0}^{t-1}\rho^j\epsilon_{1,t-j}\right)\right) < \infty$, for some $\phi > 0$ and $|\rho| < 1$, then under H'_A there exists $\nu > 0$, such that $\forall\gamma < 1$, $P^*\left[\frac{1}{R^\gamma}\min(S^a_{T,R}(\omega), S^b_{T,R}(\omega)) > \nu\right] \to 1$, $a.s. - \omega$.

Thus, the asymptotic type I error is less than or equal to β, while the asymptotic type II error is zero, conditional on ω, and for all ω except a set of measure zero. Theorem 2 also holds for $\delta_1 > 0$. In this case, the smaller statistic is $S^a_{T,R}(\omega)$, $\forall\omega$. In the case where the test selects H'_A one cannot distinguish between DGPs in levels and DGPs in logs with short memory. In this case, it remains only to test the significance of the coefficient on a linear deterministic trend in a levels regression. Even if the process is actually short memory in logs, the test is well defined, as the exponential of a short memory process is short memory. Thus, a finding that the coefficient on the trend component is significant implies the consequent choice of levels data, otherwise use logged data. In the previous section, it was noted that a larger compact set, U, leads to higher

[12] Analogously, for U^- a compact set on the negative real line, set p in $S^b_{T,R}(\omega)$ equal to zero. Theorem 2 then holds for the min statistic defined on U^-.

finite sample power as well as higher finite sample size, for U centered around zero. In the current context, finite test performance trade-offs are not as straightforward. Consider $U^+ = [0,\ u_{\max}]$. For all samples in which $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t}{\Delta X_1}\right)^2$ converges to zero, larger $u_{\max}$ implies better finite sample size and worse finite sample power. On the other hand, the opposite holds for all samples in which $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t}{\Delta X_1}\right)^2$ diverges. For this reason, we recommend use of the statistic which is defined for U (see Theorem 1). If the null hypothesis is rejected, but there is ancillary evidence that the true DGP may be a unit root process in logs with no drift, continue by using the statistic described in Theorem 2.

4.3 Finite Sample Evidence

In this section the results of a small set of Monte Carlo experiments are reported. Data are generated according to $H_1 - H_4$ in Section 2.1, which are here written as,

$H_1 : X_t = \alpha_1 + \delta_1 t + \rho X_{t-1} + \varepsilon_{1,t}$,

$H_2 : X_t = \alpha_2 + X_{t-1} + \varepsilon_{2,t}$,

$H_3 : \log X_t = \alpha_3 + \delta_2 t + \rho \log X_{t-1} + \varepsilon_{3,t}$,

$H_4 : \log X_t = \alpha_4 + \log X_{t-1} + \varepsilon_{4,t}$,

where all errors are assumed to be *iid* $\mathrm{N}(0, \sigma_i)$ random variables, $i = 1, 2, 3, 4$. Notice that α_4 in H_4 corresponds to δ_1 in the version of H_4 given in Section 2.1, for example. In general, then, we are assuming that there is a deterministic trend in the time series under investigation, so that $S^a_{T,R}(\omega)$ and $S^b_{T,R}(\omega)$ do not need to be calculated, and $S_{T,R}(\omega)$ is thus used throughout. In order to consider parameterizations which are illustrative of the types of DGPs observed in reality, we again calibrate our models using quarterly real and seasonally adjusted U.S. M2 for the period 1970:1-1994:1 (see above discussion).[13] After fixing $\rho = 0.75$, the following estimates were obtained,

$\widehat{X}_t = 73.31 + 8.84t + 0.75X_{t-1}$, $\hat{\sigma}_1 = 32.56$,

$\widehat{X}_t = 30.78 + X_{t-1}$, $\hat{\sigma}_2 = 17.27$

$\widehat{\log X}_t = 1.621 + 0.0050t + 0.75 \log X_{t-1}$, $\hat{\sigma}_3 = 0.0205$, and

$\widehat{\log X}_t = 0.0188 + \log X_{t-1}$, $\hat{\sigma}_4 = 0.0099$.[14]

[13] Results based on models parameterized using U.S. GDP as a benchmark were also tabulated, and are available upon request.

[14] The above models are meant to be used only as benchmark DGPs, and are not necessarily indicative of what might be viewed as the "best" univariate linear time series model for money. In addition, if one of these models were assumed to be "true", then the others would necessarily be misspecified. Further, there is likely a structural break in the money data being examined (see e.g. Swanson (1998)), so that all of the models may be rather inaccurate. These last two issues, while important for empirical analysis of the money data have no impact on our Monte Carlo

Using these estimated models as our "benchmark" models, data are generated according to DGPs with: $\alpha_1 = 75$, $\alpha_2 = (20, 30, 40, 50)$, $\alpha_3 = 2$, $\alpha_4 = (0.010, 0.015, 0.020, 0.025)$, $\delta_1 = (5, 10, 15, 20)$, $\delta_2 = (0.003, 0.004, 0.005, 0.006)$, and σ_i $i, 1, 2, 3, 4$ is set equal to its estimated value. Samples of $T=$ 100, 200, 300, 400, and 500 observations were simulated. Also, we set $R = (T^{0.50}, T^{0.75}, T^{0.90}, T^{0.95})$. The range of u is $-1.0 \leq u \leq 1.0$, and 100 statistics for 100 increments within this range were calculated.[15] All simulations are based on 500x500 Monte Carlo trials, where the first 500 corresponds to the number of Monte Carlo iterations, and the second 500 corresponds to the number of different ξ_i , $i = 1, ..., R$ vectors that are drawn. (Put another way, for each new ξ vector, a new statistic is calculated and inference based on that statistic is carried out. For each draw of the DGP, this is repeated 500 times.) Rejection frequencies based on these DGPs and a 5% nominal level are graphically depicted in Figures 1 and 2. The different plots in the figures correspond to different parameterizations, and are labelled as follows:

Rejection Frequencies When Data Are Generated According to Log DGPs (Empirical Size)

DGP-S1: $\alpha_4 = 0.010$, DGP-S2: $\alpha_4 = 0.015$, DGP-S3: $\alpha_4 = 0.020$, DGP-S4: $\alpha_4 = 0.025$, DGP-S5: $\alpha_3 = 2$, $\delta_2 = 0.003$, DGP-S6: $\alpha_3 = 2$, $\delta_2 = 0.004$, DGP-S7: $\alpha_3 = 2$, $\delta_2 = 0.005$, DGP-S8: $\alpha_3 = 2$, $\delta_2 = 0.006$.

Rejection Frequencies When Data Are Generated According to Levels DGPs (Empirical Power)

DGP-P1: $\alpha_2 = 20$, DGP-P2: $\alpha_2 = 30$, DGP-P3: $\alpha_2 = 40$, DGP-P4: $\alpha_2 = 50$, DGP-P5: $\alpha_1 = 75$, $\delta_1 = 5$, DGP-P6: $\alpha_1 = 75$, $\delta_1 = 10$, DGP-P7: $\alpha_1 = 75$, $\delta_1 = 15$, DGP-P8: $\alpha_1 = 75$, $\delta_1 = 20$.

Turning to the results, recall first that the graphs denoted by DGP-S3 (DGP-S8) and DGP-P2 (DGP-P6) correspond most closely to the estimated models (i.e. to what we have termed our "benchmark" models), and in these cases, empirical rejection frequencies are close to the level of the test when data are generated under the null (DGP-S3 and DGP-S8), while rejection frequencies are above 0.60 for all values of R except $R = T^{0.5}$, when data are generated under the alternative (DGP-P2 and DGP-P6). As expected, empirical power improves as we move from DGP-P1 to DGP-P4, for example, because the trend parameter (i.e. δ_1 for H_1 or α_2 for H_2) increases. The same argument can be made when viewing DGP-P5 - DGP-P8 in Figure 2. Correspondingly,

analysis, however, as we are simply using the above parameterizations as given baseline models in our experiments (e.g. whether the parameters are consistent or not has no implications for the Monte Carlo experiments, per se). Please see the subsequent section for a more detailed discussion of the empirical properties of the money data used here.

[15] Various ranges and increments for u were examined, including ranges for u between -100 and 100. Results were found to be robust to the choice of U and the number of increments.

increasing either δ_2 or α_4 results in improved empirical size, as evidenced by moving from DGP-S1 to DGP-S4, for example. Interestingly, empirical size is close to nominal in all cases, as long as samples of around 300 or more observations are used. The same can be said of empirical power, as rejection frequencies are generally above 0.80 in all cases except $R = T^{0.5}$, when samples of around 300 or more observations are used. The trade-off between smaller and bigger R is also as expected - increasing R results in worse empirical size and better empirical power. In summary, while our experiments are rather limited in scope, we have some evidence that the proposed test may be useful, even for samples of as few as 300 observations. However, empirical size/power trade-offs are very pronounced for smaller samples.[16]

4.4 Empirical Illustration

In keeping with the Monte Carlo experiments reported on in the previous sections, we now consider the quarterly U.S. data set examined by King, Plosser, Stock and Watson (KPSW: 1991), and updated in Corradi, Swanson and White (2000). In particular, the $S_{T,R}(\omega)$ test is carried out for four series, including: consumption, investment, money, and output. Note that variables of the type examined here are all clearly upward trending, as documented in Stock and Watson (1989), for example, thus supporting our use of this particular version of the data transformation test. Also, note that the variables are constructed as in KPSW.[17] Results for a variety of values of R, as well as for two different sub-samples, are given in Table 1.

[16]Monte Carlo results based on the sequential application of our data transformation test, unit root tests, cointegration tests, and common cycles tests suggest that the finite sample performance of the latter tests (after first carrying out our data transformation test) is similar to cases where the correct data transformation is known. These results are available upon request from the authors.

[17]Using citibase mnemonics, the series are constructed as follows: consumption=gcq/p; investment=gifq/p; money=fm2/p; output=(gdpq-ggeq)/p, with p=p16*1000000, p16=U.S. population, gcq=real consumption expenditures, fm2=nominal seasonally adjusted M2 stock, gdpq=real GDP, and ggeq=real government expenditures on goods and services. Thus, all series are per capita.

Table 1: Empirical Illustration: The King-Plosser-Stock-Watson Data Set *

Series	*Data Transformation Statistics*				
	$R = T^{0.50}$	$R = T^{0.75}$	$R = T^{0.90}$	$R = T^{0.95}$	$R = T^{0.99}$
Panel A: Sample: 1947 quarter 1 - 1994 quarter 1					
Consumption	1.21	1.65	2.79	3.58	4.24
Investment	2.14	5.82	11.9	15.8	19.5
Money	2.59	7.73	16.2	21.4	26.4
Output	1.07	0.90	0.95	1.07	1.03
Panel B: Sample: 1970 quarter 1 - 1994 quarter 1					
Consumption	1.27	1.05	1.14	1.32	1.35
Investment	1.32	1.41	1.90	2.35	2.58
Money	1.80	3.65	6.35	7.96	9.41
Output	1.41	1.87	2.82	3.52	4.00

* Entries in the table are $S_{T,R}(\omega)$ statistics calculated as discussed above, and are distributed as χ_1^2 random variables so that 1% and 5% critical values are 6.63 and 3.84, respectively. Data are quarterly and correspond to those series constructed and examined by King, Plosser, Stock and Watson (1991), except that the data have been updated through 1994, as discussed in Corradi, Swanson and White (2000).

A number of conclusions can be made based on these results. First, consumption and output are best modelled in logs, a result that agrees in large part with previous empirical practice (see e.g. Engle and Granger (1987), Vahid and Engle (1993), and Diebold and Senhadji (1996)). Second, the evidence on investment is mixed. For the longer sub-sample, the statistics based on $R = T^{0.5}$ supports logs, while the statistics based on different choices of R support levels. However, we know that the power of these tests is rather low for $R = T^{0.5}$. For this reason, and given that there is always a possibility of structural breaks (and hence poor test performance) among economic variables, we also constructed test statistics for the smaller sub-sample reported on in Panel B of the table. Notice that in this case, the null hypothesis of a loglinear DGP for investment is never rejected, regardless of the value of R (the maximum value of the statistic is 2.58 and the 5% critical value is 3.84). Thus, although the evidence is somewhat mixed, it appears that investment is better modelled in logs, particularly if more recent data are being modelled. Third, the evidence on money is mixed. In both sub-samples, findings depend upon the choice of R. Again, one reason for this may be the presence of a structural break. Indeed, in the early 1980's (prior to 1984) the federal reserve bank experimented with policy aimed at targeting the money stock. In addition, at about the same time, there was an apparent structural break in the money stock due to the introduction of interest bearing checking accounts and due to a surge in credit card usage, for example.[18] For

[18] See Clements and Hendry (1999a,b) for a detailed discussion of forecasting failure in the presence of structural breaks in economic series.

these reasons, we also looked at the sub-sample period beginning in 1984. For this sub-sample, the statistics for money are (2.04, 3.67, 6.19, 7.49, 8.42), for the various values of R reported on in the table. Note that although there is now stronger evidence than before for modelling money in logs, the evidence is still mixed. Thus, no definite choice among logs and levels is provided by the test when modelling money. Overall, though, this illustration supports the common practice in empirical macroeconomics of logarithmic data transformation prior to unit root testing.[19]

5 Concluding Remarks

Unit root and stationarity tests are severely biased, both in small and in large sample, in the presence of incorrect data transformation. Additionally, common feature and cointegration tests do not have well defined limiting distributions under incorrect data transformation, and our Monte Carlo experiments suggest that these tests are indeed incorrectly sized and suffer power problems in such circumstances. Given these problems, we suggest that if theory does not unambiguously suggest a particular data transformation, then statistical tests for data transformation under unknown order of integratedness may be useful. Along these lines, we propose a simple test, based on the combined use of a randomization procedure and sample conditioning, for choosing between linearity in logs and linearity in levels, in the presence of deterministic and/or stochastic trends. For any sample which is a realization of a DGP under the null hypothesis (i.e. a log DGP), the statistic has a χ^2 limiting distribution, while for any sample which is a realization of a DGP under the alternative (i.e. a level DGP) the statistic diverges. Once we have chosen the correct the data transformation, we remain with the standard problem of testing for a unit root, for cointegration or for common cycles. A Monte Carlo exercise is used to examine the finite sample behavior of the suggested testing procedure, and our findings are rather encouraging for samples of at least 300 observations. In addition, an empirical illustration based on the King, Plosser, Stock and Watson (1991) data set is given, and evidence of preference for the loglinear models for output and consumptions is provided.

[19] We leave the discussion of the implementation of our tests in nonlinear contexts, such as when fitting smooth transition and related models (see e.g. Van Dijk and Franses (1999)) and when there are outliers (see e.g. Van Dijk, Franses and Lucas (1999a) and Van Dijk, Franses and Lucas (1999b)) to future research.

6 Appendix

Proof of Theorem 1: (i) First note that conditional on the sample, $\forall i$, $V^i_{\xi,T} \sim N\left(0, \frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)$. Let $\Omega^+ = \{\omega : \lim_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^T\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)^{1/2} = \infty\}$. We begin by showing that $P(\Omega^+) = 1$. Under DGP H_3, $\Delta X_t = X_{t-1}\exp(\delta_1\sum_{j=0}^{t-1}\rho^j + \sum_{j=0}^{t-1}\rho^j(\varepsilon_{2,t-j} - \varepsilon_{2,t-j-1}))$, where $X_{t-1} = \exp(\log X_0 + \alpha_1\sum_{j=0}^{t-1}\rho^j + \delta_1\sum_{j=0}^{t-1}\rho^j(t-j) + \sum_{j=0}^{t-1}\rho^j\varepsilon_{2,t-j-1})$. Under DGP H_4, $\Delta X_t = X_{t-1}\exp(\delta_1 + \varepsilon_{2,t})$, where $X_{t-1} = \exp(\log X_0 + \delta_1(t-1) + \sum_{j=0}^{t-1}\varepsilon_{2,j})$. The functional law of the iterated logarithm for strong mixing processes (e.g. Berger Theorem 3.1, 1990) states that $\limsup_{t\to\infty}\frac{1}{\sqrt{2t\log\log t}}\sum_{j=1}^t \varepsilon_{2,j} = O_{a.s.}(1)$. The deterministic trend component is then the dominant term in both DGPs. It follows that $\frac{1}{T}\sum_{t=1}^T\left(\frac{\Delta X_t}{\Delta X_1}\right)^2 \overset{a.s}{\to} \infty$, at a geometric rate. Thus, $P(\Omega^+) = 1$. We now proceed conditionally on the sample, and with the notation $a.s. - \omega$, we mean conditionally on $\omega \in \Omega^+$. Hereafter, let $U = [\underline{u}, \overline{u}]$. We first need to show that for $\forall u \in U$,

$$P^*(V^i_{\xi,T}(\omega) \leq u) = \frac{1}{2} + O(T^{-1/2}), \tag{10}$$

where the $O(T^{-1/2})$ term holds uniformly in i and u, $a.s - \omega$. Suppose $u > 0$, then

$$P^*(V^i_{\xi,T}(\omega) \leq u) = P^*(V^i_{\xi,T}(\omega) \leq 0) + P^*(0 \leq V^i_{\xi,T}(\omega) \leq u) a.s. - \omega.$$

As ξ_i is a zero mean normal, $P^*(V^i_{\xi,T}(\omega) \leq 0) = \frac{1}{2}$. Therefore, it suffices to show that $P^*(0 \leq V^i_{\xi,T} \leq u) = O(T^{-1/2})$, uniformly in u and i, $a.s. - \omega$. Now,

$$\begin{aligned} P^*(0 \quad &\leq \quad V^i_{\xi,T}(\omega) \leq u) = \frac{1}{\left(\frac{1}{T}\sum_{t=1}^T\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)^{1/2}\pi^{1/2}}\int_0^u \exp\left(-x^2/\frac{2}{T}\sum_{t=1}^T\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)dx \\ &= \quad O(T^{-1/2}), uniformlyini, a.s. - \omega, \end{aligned} \tag{11}$$

as $\sup_{u\in U}\int_0^u \exp\left(-x^2/\frac{2}{T}\sum_{t=1}^T\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)dx \leq \overline{u}$, $a.s. - \omega$, and $\frac{1}{T}\sum_{t=1}^T\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2$ diverges at a faster rate than T. A similar argument applies to the case of $u < 0$. Hereafter, E^* denotes the expectation with respect to the probability measure P^*. Now, for any given $u \in U$,

$$\begin{aligned} \frac{1}{\sqrt{R}}\sum_{i=1}^R\left(1\{V^i_{\xi,T}(\omega) \leq u\} - \frac{1}{2}\right) \quad &= \quad \frac{1}{\sqrt{R}}\sum_{i=1}^R\left(1\{V^i_{\xi,T}(\omega) \leq u\} - E^*(1\{V^i_{\xi,T}(\omega) \leq u\})\right) \\ &\quad + \sqrt{R}\left(E^*(1\{V^i_{\xi,T}(\omega) \leq u\}) - \frac{1}{2}\right). \end{aligned} \tag{12}$$

Note that $E^*(1\{V^i_{\xi,T}(\omega) \leq u\}) = P^*(V^i_{\xi,T}(\omega) \leq u) = \frac{1}{2} + O(T^{-1/2})$, where the $O(T^{-1/2})$ term holds uniformly in i and u, $a.s - \omega$. As R grows at a rate slower than T, the last term on the RHS of

(12) approaches zero, $a.s. - \omega$. Recall that Var^* denotes the variance with respect the probability measure P^*. Now, as $V^i_{\xi,T}(\omega)$ is independent of $V^j_{\xi,T}(\omega)$, $\forall i \neq j$, $a.s. - \omega$, and recalling (10),

$$\begin{aligned}
Var^*\left(\frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{V^i_{\xi,T}(\omega)\leq u\}\right)\right) &= \frac{1}{R}\sum_{i=1}^{R}(E^*(1\{V^i_{\xi,T}(\omega)\leq u\}^2)-(E^*(1\{V^i_{\xi,T}(\omega)\leq u\}))^2) \\
&= \frac{1}{R}\sum_{i=1}^{R}E^*(1\{V^i_{\xi,T}(\omega)\leq u\})-\frac{1}{4}-O(T^{-1}) \\
&= \frac{1}{2}+O(T^{-1/2})-\frac{1}{4}-O(T^{-1})=\frac{1}{4}+O(T^{-1/2}),
\end{aligned}$$

uniformly in i and u, $a.s - \omega$. By noting that $1\{V^i_{\xi,T}(\omega) \leq u\}$, $i = 1, ..., R$ and $R = T^a$, $0 < a < 1$, is an independent triangular array, by the central limit for independent triangular arrays (see e.g. Davidson (2000, p.52)), for all $u \in U$,

$$\frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{V^i_{\xi,T}(\omega)\leq u\}-\frac{1}{2}\right)\xrightarrow{d}N(0,1/4).$$

We now need to show that the convergence above holds uniformly in u. That is, we need to show that

$$\frac{1}{\sqrt{R}}\sum_{i=1}^{R}1\{V^i_{\xi,T}(\omega)\leq u\}-\frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{u\leq V^i_{\xi,T}(\omega)\leq u'\}\right)=o_{P^*}(1),$$

with the $o_{P^*}(1)$ term independent of u and u'. Without loss of generality, let $u < u'$. Then,

$$\frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{V^i_{\xi,T}(\omega)\leq u\}-1\{V^i_{\xi,T}(\omega)\leq u'\}\right)=\frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{u\leq V^i_{\xi,T}(\omega)\leq u'\}\right).$$

Now,

$$\sup_{u,u'\in U}P^*\left(\left|\frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{u\leq V^i_{\xi,T}(\omega)\leq u'\}\right)\right|>\epsilon\right)\leq\frac{1}{\epsilon^2}\sup_{u,u'\in U}E^*\left(\frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{u\leq V^i_{\xi,T}(\omega)\leq u'\}\right)\right)^2$$

$$=\frac{1}{\epsilon^2}\frac{1}{R}\sum_{i=1}^{R}\sup_{u,u'\in U}E^*\left(1\{u\leq V^i_{\xi,T}(\omega)\leq u'\}\right)\leq\frac{1}{\epsilon^2}\sup_i P^*(\underline{u}\leq V^i_{\xi,T}\leq\overline{u})=O(T^{-1/2}),$$

$a.s. - \omega$, because of (11). The desired result then follows.

(ii) Let $\Omega_A = \{\omega : \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2 \to M, 0 < M < \infty\}$. We begin by showing that $P(\Omega_A) = 1$. Now, $\Delta X_t = \delta_0\sum_{j=0}^{t-1}\rho^j + \sum_{j=0}^{t-1}\rho^j(\varepsilon_{1,t-j} - \varepsilon_{1,t-j-1})$, under DGP H_1, and $\Delta X_t = \delta_0 + \varepsilon_{1,t}$, under H_2. Given A1, it follows by the strong law of large numbers that $\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t}{\Delta X_1}\right)^2 \xrightarrow{a.s.} M$, and so $P(\Omega_A) = 1$. (Hereafter, with the notation $a.s. - \omega$, we mean for all $\omega \in \Omega_A$.) From the previous statements, it follows that $V^i_{\xi,T}(\omega)$ is a zero mean normal random variable with variance equal to

$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2$, so that $V^i_{\xi,T}(\omega) \overset{d^*}{\to} N(0, M)$, $a.s. - \omega$, as $T \to \infty$, and $\forall i$. Let $F(u)$ be the cumulative distribution function (CDF) of a $N(0, M)$, evaluated at $u \in U$. Then,

$$\frac{2}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}(\omega) \leq u\} - \frac{1}{2}) = \frac{2}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}(\omega) \leq u\} - F(u)) + 2\sqrt{R}(F(u) - \frac{1}{2}). \tag{13}$$

As $F(u) = \frac{1}{2}$ when $u = 0$, the second term on the RHS of (13) diverges to + or $-\ \infty$ at rate $\sqrt{R}$, $a.s - \omega$, for all $u \neq 0$. In addition, the first term on the RHS of (13) is bounded in probability, as can be shown by noting that,

$$\frac{2}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}(\omega) \leq u\} - F(u)) = \frac{2}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}(\omega) \leq u\} - F_T(u)) - 2\sqrt{R}(F_T(u) - F(u)),$$

where $F_T(u) = P^*(V^i_{\xi,T}(\omega) \leq u)$. As $V^i_{\xi,T}$ has finite variance and is independent i, the Berry-Essen theorem (e.g. Davidson (1994) p.408) can be applied, yielding that,

$$\sup_{u\in U}(F_T(u) - F(u)) = O(T^{-1/2}), a.s. - \omega.$$

In addition, as $R/T \to 0$, $\sup_{u\in U}\sqrt{R}(F_T(u) - F(u)) \to 0$, $a.s. - \omega$. Now, $E^*(1\{V^i_{\xi,T}(\omega) \leq u\}) = \int_{-\infty}^{u} dF_T(s) = F_T(u)$, and $Var^*(1\{V^i_{\xi,T}(\omega) \leq u\}) = F_T(u)(1 - F_T(u))$. Thus, as $R, T \to \infty$, $R/T \to 0$, $\frac{2}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}(\omega) \leq u\} - F_T(u,\omega)) \overset{d^*}{\to} N(0, 4F(u)(1-F(u))$, $a.s - \omega$, as $F_T(u) \to F(u)$ when $T \to \infty$. Also, as $\frac{1}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T} \leq u\})$ is stochastic equicontinuous on U, it follows that $\sup_{u\in U}\frac{1}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T} \leq u\})$ weakly converges to the supremum of a Gaussian process. Thus, the LHS of (13) diverges in probability at rate $\sqrt{R}$, $\forall\omega \in \Omega_A$, with $\Pr(\Omega_A) = 1$.

Proof of Theorem 2: (i) Under DGP H_4, $\delta_1 = 0$ and $\Delta X_t = X_{t-1}\exp(\varepsilon_{2,t})$, where $X_{t-1} = \exp(\log X_0 + \sum_{j=1}^{t-1}\varepsilon_{2,j})$. Let

$$\Omega_1 : \{\omega : \left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)^{1/2} \to \infty\},$$

and

$$\Omega_2 : \{\omega : \left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)^{1/2} \to 0\}.$$

We begin by establishing that $P(\Omega_1 \cup \Omega_2) = 1$. This can be done by first showing that $\Pr(\omega : \lim_{t\to\infty}\left|\sum_{j=1}^{t}\varepsilon_{2,j}\right| = \infty) = 1$. Given A1, the strong invariance principle for stationary α-mixing processes (e.g. Eberlain (1986), Theorem 2) ensures that,

$$r \to \frac{1}{\sqrt{T}} \sum_{j=1}^{[Tr]} \varepsilon_{2,j} = \sigma W(r) + O_{a.s.}(T^{-\theta} \log\log T), for 0 < \theta < 1/2,$$

where $r \in (0,1]$, $\sigma^2 = E(\varepsilon_{2,t}^2)$, and W is a standard Brownian motion process. Now define,

$$\Psi = \{(t,\omega) \in [0,\infty) \times \Omega : W(t,\omega) = 0\},$$

and $\forall \omega \in \Omega$, define,

$$\Psi(\omega) = \{t \in [0,\infty) : W(t,\omega) = 0\}.$$

From Theorem 2.9.6 in Karatzas and Shreve (1991), it follows that $\Psi(\omega)$ has zero Lebesgue measure, $\forall \omega \in \Omega^*$, where $P(\Omega^*) = 1$. Thus, it also follows that as $t \to \infty$, $\left|\sum_{j=1}^t \varepsilon_{2,j}\right| \overset{a.s.}{\to} \infty$, at rate T^θ, $0 < \theta < 1/2$. This implies that $\forall \omega$ for which $\sum_{j=1}^t \varepsilon_{2,j}(\omega) \to \infty$, $\Delta X_t(\omega)^2 \to \infty$, and $\forall \omega$ for which $\sum_{j=1}^t \varepsilon_{2,j}(\omega) \to -\infty$, $\Delta X_t(\omega)^2 \to 0$. Now, $\Pr(\Delta X_1 = 0) = 0$, and given the moment conditions in A1, $\frac{1}{T^{2/7}} \Delta X_1 \overset{a.s.}{\to} 0$. Thus, $\forall \omega$ for which $\Delta X_t(\omega)^2 \to \infty$, we also have that $(\Delta X_t(\omega)/\Delta X_1(\omega))^2 \to \infty$, and $\forall \omega$ for which $\Delta X_t(\omega)^2 \to 0$, $(\Delta X_t(\omega)/\Delta X_1(\omega))^2 \to 0$, both at a geometric rate. It follows that $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2 \to \infty$ or $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2 \to 0$, $\forall \omega$, also at a geometric rate. Thus, $P(\Omega_1 \cup \Omega_2) = 1$. It remains to establish that as $T, R \to \infty$ and $R/T \to 0$, (a) $\min(S^a_{T,R}(\omega), S^b_{T,R}(\omega)) = S^a_{T,R}(\omega)$ and $S^a_{T,R}(\omega) \overset{d^*}{\to} \chi^2_1$ $a.s. - \omega$, $\forall \omega \in \Omega_1$, and (b) $\min(S^a_{T,R}(\omega), S^b_{T,R}(\omega)) = S^b_{T,R}(\omega)$ and $S^b_{T,R}(\omega) \overset{pr^*}{\to} 0$, $a.s. - \omega$, $\forall \omega \in \Omega_2$. (with the notation $\overset{pr^*}{\to}$ we mean convergence in probability according to P^*, conditionally on the sample).

(a) That $S^a_{T,R}(\omega) \overset{d^*}{\to} \chi^2_1$ $a.s. - \omega$,$\forall \omega \in \Omega_1$ follows directly by the same arguments used in the proof of Theorem 1(i). Now,

$$S^b_{T,R}(\omega) = \int_{U^+} \left(\frac{2}{\sqrt{R}} \sum_{i=1}^R \left(\left(1\left\{V^i_{\xi,T}(\omega) \le u\right\} - \frac{1}{2}\right) - \frac{1}{2}\right)\right)^2 \pi(u)du$$

$$= S^a_{T,R}(\omega) + \frac{1}{4}\sqrt{R} - \int_{U^+} \frac{1}{\sqrt{R}} \sum_{i=1}^R (1\left\{V^i_{\xi,T}(\omega) \le u\right\} - \frac{1}{2})\pi(u)du,$$

so that $S^b_{T,R}(\omega)$ diverges at rate $\sqrt{R}$.

(b) Recall that $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2 \to 0, \forall \omega \in \Omega_2$. As $V^i_{\xi,T}(\omega)$ is a zero mean normal with variance equal to $\frac{1}{T}\sum_{t=1}^T \left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2$, it follows that $V^i_{\xi,T}(\omega) \overset{pr^*}{\to} 0$, $a.s. - \omega$, $\forall \omega \in \Omega_2$, $\forall i$. Furthermore,

$\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2 \to 0$ at an exponential rate and $V_{\xi,T}^{i} \overset{pr^*}{\to} 0$, $\forall\omega \in \Omega_2$, at the same rate. Thus, $\forall u \in U^{+}$, $1\{V_{\xi,T}^{i}(\omega) \leq u\} \overset{pr^*}{\to} 1,\ a.s.-\omega$, $\forall\omega \in \Omega_2$, at an exponential rate as $T \to \infty$. It follows that $\forall u \in U^{+}$, as $T, R \to \infty, R/T \to 0$, $\frac{1}{\sqrt{R}}\sum_{i=1}^{R}(1\{V_{\xi,T}^{i} \leq u\} - 1) \overset{pr^*}{\to} 0$, and so $S_{T,R}^{b}(\omega) \overset{pr^*}{\to} 0$, $a.s. - \omega$, $\forall\omega \in \Omega_2$. Also, note that

$$S_{T,R}^{a}(\omega) = S_{T,R}^{b}(\omega) + \frac{1}{4}\sqrt{R} + \int_{U^{+}}\left(\frac{1}{\sqrt{R}}\sum_{i=1}^{R}(1\left\{V_{\xi,T}^{i}(\omega) \leq u\right\} - \frac{1}{2}\right)\pi(u)du,$$

so that $S_{T,R}^{a}(\omega)$ diverges at rate $\sqrt{R}$.

(ii) Note that, under DGP H_1, $\Delta X_t = \delta_0 + (\rho-1)X_{t-1} + \epsilon_{1,t} - \epsilon_{1,t-1}$; under DGP H_2, $\Delta X_t = \delta_0 + \epsilon_{1,t}$, and finally under DGP H_3, $\Delta X_t = \exp(\alpha_3 + \sum_{j=1}^{t}\rho^{j-1}\epsilon_{2,t-j+1}) - \exp(\alpha_3 + \sum_{j=1}^{t-1}\rho^{j-1}\epsilon_{2,t-j+1})$. The desired result then comes by the same argument followed in the proof of Theorem 1(ii).

7 References

Arino, M.A. and P.H. Franses, (2000), Forecasting the Level of Vector Autoregressions of Log Transformed Series, *International Journal of Forecasting*, 16, 111-116.

Berger, E., (1990), An Almost Sure Invariance Principle for Stationary Ergodic Sequences of Banach Space Valued Random Variables, *Probability Theory and Related Fields,* 84, 161-201.

Bickel, P.J. and K. Doksum, (1981), An Analysis of Transformations Revisited, *Journal of the American Statistical Association*, 76, 296-310.

Chao, J.C., V. Corradi and N.R. Swanson, (2001), Data Transformation and Forecasting in Models with Unit Roots and Cointegration, *Annals of Economics and Finance*, 2, 59-76.

Clements, M.P. and D.F. Hendry, (1999a), *Forecasting Economic Time Series*: *The Zeuthen Lectures on Economic Forecasting,* MIT Press, Cambridge.

Clements, M.P. and D.F. Hendry, (1999b), On Winning Forecasting Competitions in Economics, *Spanish Economic Review*, 1, 123-160.

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

Corradi, V. and N.R. Swanson, and H. White, (2000), Testing for Stationary Ergodicity and Comovements Between Nonlinear Discrete Time Markov Processes, *Journal of Econometrics,* 96, 39-73.

Corradi, V. and N.R. Swanson, (2002), A Consistent Test for Nonlinear Out of Sample Predictive Accuracy, *Journal of Econometrics,* 110, 353-381.

Davidson, J., (1994), *Stochastic Limit Theory*, Oxford University Press, Oxford.

Davidson, J., (2000), *Econometric Theory*, Blackwell Publishers, Oxford.

De Bruin, P. and P.H. Franses, (1999), Forecasting Power Transformed Time Series, *Journal of Applied Statistics,* 27, 807-815.

Diebold, F.X. and A.S. Senhadji, (1996), Deterministic versus Stochastic Trends in US GNP, Yet Again, *American Economic Review,* 86, 1291-1298.

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.

Dufour, J.M. and J.F. Kiviet, (1996), Exact Tests for Structural Change in First-Order Dynamic Models, *Journal of Econometrics,* 70, 39-68.

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

Engle, R.F. and C.W.J. Granger, (1987), Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica,* 55, 251-276.

Engle, R.F. and S. Kozicki, (1993), Testing for Common Features, *Journal of Business and Economic Statistics*, 11, 369-395.

Engle, R.F. and J.V. Issler, (1995), (1995), Estimating Common Sectoral Cycles, *Journal of Monetary Economics*, 35, 83-113.

Franses, P.H. and G. Koop, (1998), On the Sensitivity of Unit Root Inference to Nonlinear Data Transformations, *Economics Letters*, 59, 7-15.

Franses, P.H. and M. McAleer, (1998), Testing for Unit Roots and Nonlinear Transformations, *Journal of Time Series Analysis,* 19, 147-164.

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

Hansen, B.E., (1996), Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis, *Econometrica,* 64, 413-430.

Inoue, A., (2001), Testing for Distributional Change in Time Series, *Econometric Theory*, 17, 156-187.

Issler, J.V. and F. Vahid, (2001), Common Cycles and the Importance of Transitory Shocks to Macroeconomic Aggregates", *Journal of Monetary Economics,* 47, 449-475.

Karatzas, J. and S.E. Shreve, (1991), *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

King, R.G., C.I. Plosser and S. Rebelo, (1988(a)), Production and Growth and Business Cycles: I. The Basic Neoclassical Model, *Journal of Monetary Economics,* 21, 195-232.

King, R.G., C.I. Plosser and S. Rebelo, (1988(a)), Production and Growth and Business Cycles: II. New Directions, 21, 309-342.

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

Kobayashi, M., (1994), Power of Tests for Nonlinear Transformation in Regression Analysis, *Econometric Theory,* 10, 357-371.

Kobayashi, M. and M. McAleer, (1999a), Tests of Linear and Logarithmic Transformations for

Integrated Processes, *Journal of the American Statistical Association,* 94, 860-868.

Kobayashi, M. and M. McAleer, (1999b), Analytical Power Comparisons of Nested and Nonnested Tests for Linear and Loglinear Regression Models, *Econometric Theory,* 15, 99-113.

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.

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

Lütkepohl, H. and M.M. Burda, (1997), Modified Wald Tests under Nonregular Conditions, *Journal of Econometrics,* 78, 315-332.

Park, J.Y. and P.C.B. Phillips, (1999), Asymptotics for Nonlinear Transformations of Integrated Time Series, *Econometric Theory,* 15, 269-298.

Park, J.Y. and P.C.B. Phillips, (2001), Nonlinear Regression with Integrated Time Series, *Econometrica*, 69, 117-162.

Pearson, E.S. (1950), On Questions Raised by the Combination of Tests Based on Discontinuous Distributions, *Biometrika*, 37, 383-398.

Revuz, D. and M. Yor, (1990), *Continuous Martingales and Brownian Motion*, Springer and Verlag, Berlin.

Stock, J.H. and M.W. Watson, (1989), Interpreting the Evidence on Money-Income Causality, *Journal of Econometrics*, 40, 161-181.

Swanson, N.R., (1998), Money and Output Viewed Through a Rolling Window, *Journal of Monetary Economics,* 41, 455-474.

Vahid, F. and R.F. Engle, (1993), Common Trends and Common Cycles, *Journal of Applied Econometrics*, 8, 341-360.

Vahid, F. and J.V. Issler, (2002), The Importance of Common Cyclical Features in VAR Analysis: a Monte Carlo Study, *Journal of Econometrics*, 109, 341-63.

Van Dijk, D. and P.H. Franses (1999), Modelling Multiple Regimes in the Business Cycle, *Macroeconomic Dynamics* 3, 311-340.

Van Dijk, D., P.H. Franses and A. Lucas (1999a), Testing for Smooth Transition Nonlinearity in the Presence of Outliers, *Journal of Business and Economic Statistics* 17, 217-235.

Van Dijk, D., P.H. Franses and A. Lucas (1999b), Testing for ARCH in the Presence of Additive Outliers, *Journal of Applied Econometrics* 14, 539-562.

Table 2: VE Common Cycle Test Performance Under Various Data Transformations[(*)]

	T=100		T=250		T=500	
σ_u^2	AIC	SIC	AIC	SIC	AIC	SIC
Panel A: $rho = 0.3$						
Data Generated in Logs, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.99440	1.00000	0.99380	0.99980	0.99300	1.00000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.99420	1.00000	0.99380	0.99980	0.99300	1.00000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.99400	1.00000	0.99380	0.99980	0.99300	1.00000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.99420	1.00000	0.99380	0.99980	0.99300	1.00000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.99400	1.00000	0.99380	0.99980	0.99300	1.00000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.99400	1.00000	0.99380	0.99980	0.99300	1.00000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.97680	0.99580	0.96940	0.99500	0.71580	0.97900
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.31160	0.83400	0.00020	0.12180	0.00000	0.00020
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.33780	0.72980	0.00000	0.01240	0.00000	0.00000
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.66260	0.61120	0.04580	0.27920	0.00000	0.00500
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.98280	0.99980	0.32260	0.93620	0.00000	0.03020
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.97640	0.99860	0.80560	0.98960	0.40540	0.96420
Panel B: $rho = 0.6$						
Data Generated in Logs, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.99340	0.99920	0.99300	1.00000	0.99280	1.00000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.99340	0.99920	0.99300	1.00000	0.99280	1.00000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.99340	0.99920	0.99300	1.00000	0.99280	1.00000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.99340	0.99920	0.99300	1.00000	0.99280	1.00000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.99340	0.99920	0.99300	1.00000	0.99280	1.00000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.99340	0.99920	0.99300	1.00000	0.99280	1.00000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.98440	0.99680	0.97920	0.99680	0.70760	0.97000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.86460	0.98700	0.29480	0.86280	0.02600	0.50840
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.77020	0.92640	0.03360	0.46820	0.00000	0.04560
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.75440	0.74300	0.65960	0.96840	0.17500	0.76920
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.98560	0.99780	0.69820	0.98960	0.01120	0.67660
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.98840	0.99960	0.89760	0.99880	0.49820	0.98180
Panel C: $rho = 0.9$						
Data Generated in Logs, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.25300	0.24160	0.83540	0.83580	0.99560	1.00000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.25360	0.24260	0.83480	0.83520	0.99560	1.00000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.25400	0.24300	0.83500	0.83540	0.99560	1.00000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.25360	0.24280	0.83560	0.83600	0.99560	1.00000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.25380	0.24280	0.83500	0.83540	0.99560	1.00000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.25280	0.24200	0.83500	0.83540	0.99580	1.00000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.98420	0.99620	0.98120	0.99740	0.67100	0.96580
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.98500	0.99820	0.91320	0.99340	0.11900	0.79040
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.96340	0.99500	0.56860	0.93560	0.04780	0.61540
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.99180	0.99900	0.98580	0.99760	0.89660	0.98980
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.77560	0.77540	0.94240	0.99840	0.64280	0.98540
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.93320	0.93760	0.96140	0.99960	0.63720	0.98800

[(*)] Notes: Entries are based on the sequential application of Vahid-Engle common cycle tests and denote cases where a test of the null hypothesis that s¿0 resulted in failure to reject, followed by rejection of the null hypothesis that s¿1, based on 5experiments are given above. In all experiments, 5000 Monte Carlo simulations were run (see above for further details).

Table 3: Johansen Cointegration Test Performance under Various Data Transformations [(*)]

			T=100		T=250		T=500	
a_1	c_1	c_2	Trace1	Trace2	Trace1	Trace2	Trace1	Trace2
Panel A: DGP in logs								
0.0	0.0	0.0	0.339	0.426	0.596	0.733	0.94	0.95
0.01	0.0	0.0	0.298	0.383	0.534	0.671	0.914	0.924
0.02	0.0	0.0	0.310	0.376	0.560	0.644	0.914	0.906
0.00	-0.2	0.2	0.978	0.982	1.000	1.000	1.000	1.000
0.00	-0.2	0.4	0.998	0.996	1.000	1.000	1.000	1.000
0.00	-0.2	0.6	0.999	1.000	1.000	1.000	1.000	1.000
0.01	-0.2	0.2	0.973	0.965	1.000	0.999	1.000	1.000
0.01	-0.2	0.4	0.997	0.994	1.000	1.000	1.000	1.000
0.01	-0.2	0.6	1.000	0.998	1.000	1.000	1.000	1.000
0.02	-0.2	0.2	0.967	0.956	0.999	0.999	1.000	1.000
0.02	-0.2	0.4	0.997	0.990	1.000	1.000	1.000	1.000
0.02	-0.2	0.6	1.000	0.997	1.000	1.000	1.000	1.000
Panel B: DGP in levels								
0.1	0.0	0.0	0.152	0.163	0.354	0.235	0.657	0.395
0.2	0.0	0.0	0.498	0.208	0.922	0.434	0.999	0.763
0.1	-0.2	0.2	0.989	0.965	1.000	1.000	1.000	1.000
0.1	-0.2	0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.1	-0.2	0.6	1.000	1.000	1.000	1.000	1.000	1.000
0.2	-0.2	0.2	1.000	0.965	1.000	1.000	1.000	1.000
0.2	-0.2	0.4	1.000	1.000	1.000	1.000	1.000	1.000
0.2	-0.2	0.6	1.000	1.000	1.000	1.000	1.000	1.000

[(*)] Notes: Entries are Johansen trace test statistic rejection frequencies (Trace1 is based on models with an intercept and Trace2 is based on models with an intercept and trend in test regressions). Panel A presents results when the DGP is in logs, and data are exponentiated; while in Panel B the DGP is in levels, and data are logged. For each panel, entries below the horizontal rule correspond to power, while those above the horizontal rule correpond to empirical test level. All results are based on 5% nominal level tests. In all experiments, 5000 Monte Carlo simulations were run (see above for further details).