Corradi, Valentina; Swanson, Norman R.

Working Paper

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

Working Paper, No. 2003-22

Provided in Cooperation with:
Department of Economics, Rutgers University

Suggested Citation: Corradi, Valentina; Swanson, Norman R. (2003) : The Effect of Data Transformation on Common Cycle, Cointegration and Unit Root Tests : Monte Carlo Results and a Simple Test, Working Paper, No. 2003-22, Rutgers University, Department of Economics, New Brunswick, NJ

This Version is available at:
https://hdl.handle.net/10419/23178

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]Queen Mary, University of London and Rutgers University, respectively

September 2002
revised June 2003

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: common cycles, common trends, nonlinear transformation, nonstationarity, randomized procedure.

---

* Valentina Corradi, Department of Economics, Queen Mary, University of London, Mile End Road, London E1 4NS, UK, v.corradi@qmul.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 putting together an excellent conference. In addition, the many useful comments received from conference participants were also much appreciated. Finally, the authors wish to thank the editor, Heather Anderson, two anonymous referees, Marcelo Fernandes and Hashem Pesaran for helpful comments on this version of the paper; and Graham Elliot, Liudas Giraitis, Clive Granger, Vassilis Hajivassiliou, Javier Hidalgo, Nick Keifer, Yongcheol Shin, Andy Snell, and Farshid Vahid for comments on an earlier version of the paper.

1 Introduction

Engle and Kozicki (1993) have introduced the concept of common feature among time series. In their terminology, if the individual series exhibit a given feature, say serial correlation, heteroskedasticity, seasonality, etc., but there exist linear combination(s) which do not display that feature, then there exist a common feature among these series. They also suggest a testing procedure for the null hypothesis of no common feature, in the context of stationary series. In the context of nonstationary series, the most known example of a common feature is that of cointegration or common trends; in fact the individual series display a unit root, but there exist linear combination(s) which do not display unit root (see e.g. Engle and Granger (1987), Stock and Watson (1988) and King, Plosser, Stock and Watson (1991)). Vahid and Engle (1993) merge the literature on cointegration and that on common features: first they show that the existence of a serial correlation common feature among the first differences of cointegrated series imply the existence of common cycles in terms of Beveridge-Nelson-Stock-Watson decomposition, second they suggest a test for the number of common cycles, given the number of common stochastic trends. Extension to the case non perfectly synchronized cycles (co-dependent cycles) are considered by Vahid and Engle (1997) and Cubadda and Hecq (2001); common cycles in the presence of seasonally cointegrated series are considered by Cubadda (1999).[1]

As it is customary in most of the empirical macroeconomic literature, tests for cointegration, for 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

[1]Proietti (1997) and Hecq, Palm and Urbain (2001) analyze the common cycles-common trends decomposition via state-space models.

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 $I(0)$.[2] 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 (see below for further discussion). Furthermore, correct choice of data transformation is crucial when specifying 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 different but related issues that arise in the context of data transformation. First, we study the impact that incorrect data transformation has on frequently used tests in empirical macroeconomics, such as common trends and/or common cycles tests. This is mainly accomplished via a series of Monte Carlo experiments illustrating the rather substantive effect that incorrect transformation can have on the behavior of common cycles and/or 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. The second issue that we address is the circular testing problem that arises when choosing data transformation and the order of integratedness.

[2]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).

To overcome this problem, 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.

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), Kobayashi and McAleer (1999a,b) and Kramer and Davies (2002). 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, and it may be unclear which diverges at a faster rate. As a consequence, it is not clear whether many tests have nonzero asymptotic power against alternatives of interest (see Section 3.1 for examples and discussion). 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.[3] Our procedure, which distinguishes between the null hypothesis of a loglinear DGP and the alternative of a (level) linear DGP, does not rely on small sigma asymptotics. 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, for

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

example. 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. In this way, 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 $(1 - \alpha)$ 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 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 statistic are quite good for samples with more than 250 observations, when DGPs are calibrated using the U.S. output, consumption, and money variables examined by King, Plosser, Stock and Watson (KPSW: 1991). In addition, an empirical illustration is provided in which three datasets examined in previous papers are subjected to our data transformation test. First, we examine the KPSW dataset, and find mixed evidence, although their use of logged data is

generally supported, with the possible exception of their money variable. Second, the extended Nelson and Plosser (1982) dataset of Schotman and van Dijk (1991) is examined, and it is found, again, that the logged variables examined by those authors correspond to our test findings, with the possible exception of employment and wages. Finally, the term structure dataset of Hall, Anderson and Granger (1992) is examined, and strong support is found for their use of the actual unlogged (level) data.

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), and the Johansen (1988, 1991) cointegration test. 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. Section 3 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 are contained in Sections 2 and 3 respectively, while our empirical illustration is given in Section 4. Concluding remarks are gathered in Section 5. All proofs are collected in an appendix. Hereafter, $\xrightarrow{d^*} a.s. - \omega$ denotes convergence in distribution conditional on the sample, ω, $\forall\ \omega$ (i.e. for all samples 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 $I(1)$, and in the presence of common trends, the number of common cycles can be ascertained via the approach suggested

by Vahid and Engle (VE: 1993), or using the information-criteria approach recently proposed by Issler and Vahid (2002). In this section, our objective is to examine the effect of (incorrect) data transformation on the common cycle test proposed by VE.

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, and so there is (exactly) one common trend. On the other hand,

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

is a I(1) process with drift. Also note that, $\Delta Y_{2,t} - \Delta Y_{1,t} = (\beta - 1)\epsilon_t + (\beta - 1)\delta$. Thus, while $\Delta Y_{2,t}$ and $\Delta Y_{2,t}$ are serially correlated, as u_t is a AR(1) process, $\Delta Y_{2,t} - \Delta Y_{1,t}$ is not serially correlated and cannot be predicted using the lags of the series. This means that there is a serial correlation common feature among the first difference of two cointegrated series, and so, according to the definition of VE, there is (exactly) one common cycle. More precisely, $\begin{pmatrix} 1 & -\beta \end{pmatrix}$ defines the cointegrating vector and $\begin{pmatrix} 1 & -1 \end{pmatrix}$ defines the serial correlation common feature (common cycle) vector.

Now, suppose we want to test the null of zero common cycles 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 for the null $s = 2$, i.e. no common cycles, 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.[45] The intuition behind the VE statistic is the following. If $s = 2$, then the cofeature space is of dimension two and there are two independent cofeature vectors, i.e. no cycles. In this case the two canonical correlations approach zero as T gets large and the statistic has a well defined chi-square limiting distribution. Under the alternative, $s < 2$, at least one of the (squared) canonical correlation is strictly positive and thus the statistic diverges to infinity, as the sample size gets large.

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 cointegrating vector in levels does not imply the existence of a cointegrating vector in logs. Analogously, the existence of a common serial correlation feature among the first difference of the levels does not imply a common serial correlation feature among the first difference of the logged variables. Second, note that the vector Δy_t is no longer stationary, and the limiting distribution of the cointegrating vector is no longer well defined.[6] 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 common trends and common cycles tests 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 have neither a common trend nor a common cycle in the levels (differenced level) series.

[4]Note that in the present example, $n = 2$ and so the cofeature space cannot have dimension larger than 2.

[5]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 $r = 1$.

[6]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.

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. Note that the VE statistic hinges on the correct detection of the number of common trends and consistent estimation of the cointegrating vectors. So, incorrect data transformation has both a direct and an indirect effect on common cycles test.

The above discussion is 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 (see Section 3) whenever we do not have strong a priori about 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 DGP given as equations (1) - (4) above are reported in Tables 1 and 2. In particular, the money, consumption and output variables examined by King, Plosser, Stock and Watson (1991) and updated and examined by Corradi, Swanson and White (2000), which are depicted in Figure 1, and which are later reported on in Section 4 were used to calibrate some simple random walk type models for the period 1970:1-1994:1. The calibrated models are given in Table 3b and are referred to in that table as DGPS1-DGPS6 and DGPP1-DGPP6, so that there are 12 models in total. Of these, 6 models correspond to stationary processes, and 6 to I(1) processes. In this section, we use the 6 I(1) models (i.e. DGPS4-DGPS6 (log DGPs) and DGPP4-DGPP6 (levels DGPs) to guide in the calibration of our Monte Carlo. In particular, values for δ and σ_ϵ in (1) and (2) are taken from the Table.[7] Thereafter, it remains to set σ_u, the variance of the error term that links (1) and (2), to set ρ, and to set β. As the univariate DGPs in Table 3b do not provide direct guidance for setting σ_u and ρ, various values are tried. Additionally, we fix $\beta = 0.5$. More specifically, we specify levels DGPs according to (1) and (2) with $\sigma_\epsilon = 150$, $Y_{1,0} = 100$, $Y_{2,0} = 200$, and $\sigma_u = \{0.1\sigma_\epsilon, 1\sigma_\epsilon, 10\sigma_\epsilon\}$, where $\delta = 100$ when $\sigma_u = 0.1\sigma_\epsilon$, $\delta = 150$ when $\sigma_u = 1\sigma_\epsilon$, and $\delta = 200$ when $\sigma_u = 10\sigma_\epsilon$. Data are also generated according to (3) and (4) with $\sigma_\epsilon = 0.005$, $Y_{1,0} = 0.5$, $Y_{2,0} = 1$, and $\sigma_u = \{0.1\sigma_\epsilon, 1\sigma_\epsilon, 10\sigma_\epsilon\}$, where $\delta = 0.010$ when $\sigma_u = 0.1\sigma_\epsilon$,

[7]The Monte Carlo results reported in Section 3.3 also rely on the calibarted models given in Table 3b, although experiments reported in that section are based on univariate analyses, so that no error linking multiple equations (i.e. the u_t) is specified. While the lack of a u_t term in these later Monte Carlo results differentiates the DGPs used in that section of the paper with those used in this section, the models are still related, at least in the sense that the models fitted and reported on in Table 3b are the "benchmark" around which all Monte Carlo experiments in this paper are designed.

$\delta = 0.015$ when $\sigma_u = 1\sigma_\epsilon$, and $\delta = 0.020$ when $\sigma_u = 10\sigma_\epsilon$. In all experiments, we set $\rho = \{0.3, 0.6, 0.9\}$, $\beta = 0.5$, used samples of $T = \{100, 250, 500\}$ observations, and carried out 5000 Monte Carlo simulations.

Consider application of the VE test to these data. 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)[8], and the cointegrating vector was estimated using the maximum likelihood procedure of Johansen (1988, 1991). The null hypthesis is $H_0 : s = 2$ (i.e. no common cycles) versus $H_A : s < 2$ (i.e. one common cycle). All entries in the table report the rejection rates, based on a test at 5% level. 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. 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. Turning now to our findings which are reported in Table 1, it is worth noting that in empirical rejection frequencies for the common cycles test are very high under correct data transformation, 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 as well for samples of 100 observations, with rejection rates as low as 22%. These results are as might be expected. Interestingly, though, the VE test often finds evidence of common cycles even when the data are incorrectly transformed, although this finding is far from robust as there are some 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.

Using exactly the same setup as that discussed above, cointegration tests were also run. The null hypothesis tested in this experiment (using the Johansen trace test, which is here equivalent to the maximum eigenvalue test) is that the rank of the cointegrating space is 0. Thus, as above, we expect the rejection frequency getting closer 1 under correct data transformation, while under incorrect data transformation the test is invalid and there is no cointegration, so that it is unclear what the findings of the test will be. From Table 2, we see that the rejection rates are very high, regardless of data transformation, suggesting a finding of cointegration even when the data are incorrectly transformed. While the above findings should come as no surprise, it does serve to

[8]The maximum lag length considered was 12.

underscore the importance of data transformation and testing for data transformation when the correct transformation is not suggested by economic theory.[9] Put simply, the common trend model of KPSW and the common-trend/common-cycle model of VE are designed under the assumption that the empirical investigator knows the correct data transformation, and for this reason, one needs to be wary of common cycle, common trend, and related tests performed within the framework of these models if there is uncertainty concerning the correct data transformation. Finally, it is worth noting that 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).[10]

Before turning our attention to a test for data transformation, it is worth noting that 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 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.

[9]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).

[10]The Monte Carlo experiments reported in this paper are based on stylized DGPs, often with little dynamics, and are thus meant only to illustrate the potential pitfalls associated with application of common cycle tests to incorrectly transformed data. However, it seems reasonable to assume that the sorts of findings reported in this and later sections of the paper are applicable under more complicated dynamics, for example.

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

3.1 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 sort of problem is typical of all tests which are based on functionals of partial sums and variance estimators, including, for

example, the augmented Dickey-Fuller tests and Johansen cointegration tests, 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[11], 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)).[12] 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:

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

$H_A : H_1 \cup H_2$

[11] 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).

[12] 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:

$$S_{T,R}(\omega) = \int_U Z_{T,R}^2(u,\omega)\pi(u)du, \tag{5}$$

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

$$Z_{T,R}(u) = \frac{2}{\sqrt{R}} \sum_{i=1}^{R} \left(1\left\{ V_{\xi,T}^i(\omega) \leq u \right\} - \frac{1}{2} \right), \tag{6}$$

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

$$V_{\xi,T}^i(\omega) = \left(\frac{1}{T} \sum_{t=1}^{T} \left(\frac{\Delta X_t}{\Delta X_1} \right)^2 \right)^{1/2} \xi_i, \; i = 1, \cdots, R, \tag{7}$$

with ξ_i an $iidN(0,1)$ random variable, and $1\{\cdot\}$ denoting the indicator function. 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.

3.2 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, $S_{T,R}(\omega) \xrightarrow{d^*} \chi^2_1$, $a.s. - \omega$.

(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[13] 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 $V^i_{\xi,T}(\omega)$

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

diverges almost surely to $+\infty$ or to $-\infty$, $a.s.-\omega, \forall i$. In addition, because of symmetry we have 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.[14] 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,

$$\frac{1}{\sqrt{R}}\sum_{i=1}^{R}(1\left\{V^i_{\xi T}(\omega) \leq u\right\} - \frac{1}{2}) = \frac{1}{\sqrt{R}}\sum_{i=1}^{R}(1\left\{V^i_{\xi T}(\omega) \leq u\right\} - F(u)) + \sqrt{R}(F(u) - \frac{1}{2}). \qquad (8)$$

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 (8), 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, at a geometric rate. 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

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

given sample. A natural approach is thus to construct two statistics and then base inference on the smaller one. Without loss of generality, let U^+ be a compact set on the positive real line (including 0)[15]. 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 $P^* \left[\frac{1}{R} \min(S^a_{T,R}(\omega), S^b_{T,R}(\omega)) > \nu \right] \to 1$, $a.s. - \omega$.

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

[15] 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^-.

logged data. In the previous section, it was noted that a larger compact set, U, leads to higher 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.

3.3 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 3.1, which can be 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 3.1, 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 actual data, we again parameterize our DGPs using the models given in Table 3b. As mentioned above, the DGPs in Table 3b were calibrated using the KPSW dataset (see Section 4 for further discussion of this data). In fitting these models, we set $\rho = 0.75$ in all stationary cases. Samples of T= 100, 250, 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.[16] 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

[16]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.

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 the DGPs given in Table 3b and based on 5% nominal level tests are reported in Table 3a.

Turning to the results, notice that empirical rejection frequencies are close to the level of the test when data are generated under the null (DGPS1-DGPS6), for samples of 500 observations. For smaller samples of 250 observations, the evidence is mixed, with some parameterizations resulting in good empirical level, and others tending to over-reject the loglinear null hypothesis. Notice that these finding do not hinge on the value of R used, although power results suggest that any value of R above 0.75 works well. Empirical power is moderately good for samples of 250 observations, and improves somewhat with sample size, as expected. In addition, power improves as we move from DGPP1 to DGPP6, for example, because the trend parameter (i.e. δ_1 for H_1 or α_2 for H_2) increases. Correspondingly, increasing either δ_2 or α_4 results in improved empirical size, as evidenced by moving from DGP-S1 to DGP-S4, for example. Finally, 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 250 observations. However, empirical size/power trade-offs are very pronounced for smaller samples.[17]

4 Empirical Illustration

4.1 The King, Plosser, Stock and Watson (1991) Dataset

In keeping with the Monte Carlo experiments reported on in the previous sections, we now consider the quarterly U.S. data set examined by KPSW (1991), and updated in Corradi, Swanson and White (2000). In particular, the $S_{T,R}(\omega)$ test is carried out for four series: 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.[18]

[17]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, as is to be expected. These results are thus not included here.

[18]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.

Results for a variety of values of R, as well as for two different sub-samples, are given in Table 4. 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. Vahid and Engle (1993), and Diebold and Senhadji (1996)). One important implication of this finding is that the common cycle tests often applied to series like consumption and output have generally been correctly performed using logged data. 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.[19] For these reasons, we also looked at the sub-sample period beginning in 1984. For this sub-sample, the statistics for money are (2.043, 3.666 6.191, 7.419, 8.425), 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. However, it seems to be the case that if a shorter subsample of the data is used, then the KPSW approach of logging these variables is correct, and so the factor analysis carried out by KPSW is not subject to the criticisms outlined in Section 2 of this paper. Indeed, even if a longer sample is used, the mixed evidence concerning money and the evidence in favor of logs for the other variables still suggests that their factor analysis is not subject to our data transformation related criticisms. Thus, no definite choice among logs and levels is provided by the test when modelling money. Overall,

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

though, this illustration supports the common practice in empirical macroeconomics of logarithmic data transformation prior to unit root testing.[20]

4.2 The Nelson and Plosser (1982) Dataset

These data have been extensively examined by numerous authors, including Nelson and Plosser (NP: 1982), Schotman and van Dijk (1991) and Andrews and Chen (1994), for example. In Table 5 we report results based on application of the $S_{T,R}(\omega)$ test to the NP dataset updated by Schotman and van Dijk (1991). A clear pattern emerges upon inspection of the results. Namely, for all series, with the possible exception of employment and real wages, loglinear models are preferred. This finding is largely in accord with the common practice of modelling these series in logs. However, it is interesting to note that for both sub-samples reported on in the table, employment and real wages exhibit evidence that the data might be better modeled in levels. Additionally, notice that the evidence on unemployment and velocity is mixed, suggesting that there is little to choose between logs and levels for these series.[21]

4.3 The Hall, Anderson and Granger (1992) Dataset

These data have been examined by Hall, Anderson and Granger (1992) and Anderson (1997), for example, and constitute U.S. treasury bill yields on bills with one to 11 months maturity (denoted as R1 to R11 in Table 6). A detailed discussion of these data is given in both of the aforementioned papers. Interestingly, and in contrast to our other empirical findings, all series strongly support using unlogged data, a finding which is in accord with the authors' use of unlogged data. Note that our findings on interest rates should be taken with caution, as interest rates do not exhibit a postive deterministic trend, thus violating assumption A1(i).

5 Concluding Remarks

Unit root and stationarity tests are severely biased, both in small and in large samples, when data have been incorrectly transformed. Additionally, common feature and cointegration tests do not

[20] 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.

[21] The interest rate variable in the dataset is not examined as various interest rates are examined and discussed in Section 4.3 below. Additionally, note that our findings on unemployment and velocity should be taken with caution, as interest rates do not exhibit a postive deterministic trend, thus violating assumption A1(i).

have well defined limiting distributions under incorrect data transformation. In this paper, we carry out a series of Monte Carlo experiments that suggest these tests are indeed incorrectly sized and may suffer power problems in such circumstances. Put another way, common trend models, such as that of King, Plosser, Stock and Watson (1991) and the common-trend/common-factor models, such as that of Vahid and Engle (1993) are designed under the assumption that the empirical investigator knows the correct data transformation, and for this reason, one needs to be wary of common cycle, common trend, and related tests performed within the framework of these models if there is uncertainty concerning the correct data transformation. Given these problems, we suggest that if theory does not unambiguously suggest a particular data transformation, then statistical tests for data transformation when the order of integratedness is unknown 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 roots, cointegration, and common cycles, for example. 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 250 observations. In addition, an empirical illustration based on the King, Plosser, Stock and Watson (1991) data set is given, and evidence of preference for loglinear models is provided, lending credence to their findings based on their factor model. Further empirical evidence is provided suggesting that the Nelson and Plosser (1982) data largely support the specification of loglinear models, and that Hall, Anderson and Granger (1992) were correct to use levels term structure data in their analysis.

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 $\forall u \in U$,

$$P^*(V^i_{\xi,T}(\omega) \le u) = \frac{1}{2} + O(T^{-1/2}), \tag{9}$$

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

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

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

$$\begin{aligned} P^*(0 \le V^i_{\xi,T}(\omega) \le u) &= \frac{1}{\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)^{1/2}\pi^{1/2}}\int_0^u \exp\left(-x^2/\frac{2}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)dx \\ &= O(T^{-1/2}), \end{aligned} \tag{10}$$

uniformly in i, $a.s.-\omega$, as $\sup_{u\in U}\int_0^u \exp\left(-x^2/\frac{2}{T}\sum_{t=1}^{T}\left(\frac{\Delta X_t(\omega)}{\Delta X_1(\omega)}\right)^2\right)dx \le \overline{u}$, $a.s.-\omega$, 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 $u \in U$,

$$\begin{aligned} \frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{V^i_{\xi,T}(\omega) \le u\} - \frac{1}{2}\right) &= \frac{1}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{V^i_{\xi,T}(\omega) \le u\} - E^*(1\{V^i_{\xi,T}(\omega) \le u\})\right) \\ &\quad + \sqrt{R}\left(E^*(1\{V^i_{\xi,T}(\omega) \le u\}) - \frac{1}{2}\right). \end{aligned} \tag{11}$$

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 (11) 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 (9),

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

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,

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

$$= \frac{1}{\epsilon^2}\frac{1}{R}\sum_{i=1}^{R}E^*\left(1\{\underline{u} \leq V^i_{\xi,T}(\omega) \leq \overline{u}\}\right) \leq \frac{1}{\epsilon^2}\sup_i P^*(\underline{u} \leq V^i_{\xi,T} \leq \overline{u}) = O(T^{-1/2}),$$

$a.s. - \omega$, because of (10). 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 $V^i_{\xi,T}(\omega) \xrightarrow{d^*} N(0, M)$, $a.s. - \omega$, 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,

$$\frac{2}{\sqrt{R}}\sum_{i=1}^{R}\left(1\{V^i_{\xi,T}(\omega) \leq u\} - \frac{1}{2}\right) = \frac{2}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}(\omega) \leq u\} - F(u)) + 2\sqrt{R}\left(F(u) - \frac{1}{2}\right). \tag{12}$$

As $F(u) = \frac{1}{2}$ when $u = 0$, the second term on the RHS of (12) 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 (12) 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 $R, T \to \infty$, $R/T \to 0$, $\frac{2}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}(\omega) \leq u\} - F_T(u,\omega)) \xrightarrow{d^*} N(0, 4F(u)(1-F(u))$, $a.s - \omega$, 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 (12) 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,

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

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 $t \to \infty$, $\left|\sum_{j=1}^{t} \varepsilon_{2,j}\right| \xrightarrow{a.s.} \infty$, 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, $\frac{1}{T^{2/7}} \Delta X_1 \xrightarrow{a.s.} 0$. 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) $\min(S_{T,R}^a(\omega), S_{T,R}^b(\omega)) = S_{T,R}^a(\omega)$ and $S_{T,R}^a(\omega) \xrightarrow{d^*} \chi_1^2$ $a.s. - \omega$, $\forall \omega \in \Omega_1$, and (b) $\min(S_{T,R}^a(\omega), S_{T,R}^b(\omega)) = S_{T,R}^b(\omega)$ and $S_{T,R}^b(\omega) \xrightarrow{pr^*} 0$, $a.s. - \omega$, $\forall \omega \in \Omega_2$. (with the notation $\xrightarrow{pr^*}$ we mean convergence in probability according to P^*, conditionally on the sample).

(a) That $S_{T,R}^a(\omega) \xrightarrow{d^*} \chi_1^2$ $a.s. - \omega$,$\forall \omega \in \Omega_1$ follows directly by the same arguments used in the proof of Theorem 1(i). Now,

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

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

so that $S_{T,R}^b(\omega)$ 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 $V^i_{\xi,T}(\omega)\xrightarrow{pr^*}0$, $a.s.-\omega$, $\forall\omega\in\Omega_2$, $\forall i$. 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 $V^i_{\xi,T}\xrightarrow{pr^*}0$, $\forall\omega\in\Omega_2$, at the same rate. Thus, $\forall u\in U^+$, $1\{V^i_{\xi,T}(\omega)\le u\}\xrightarrow{pr^*}1$, $a.s.-\omega$, $\forall\omega\in\Omega_2$, at an exponential rate as $T\to\infty$. It follows that $\forall u\in U^+$, as $T,R\to\infty, R/T\to 0$, $\frac{1}{\sqrt{R}}\sum_{i=1}^{R}(1\{V^i_{\xi,T}\le u\}-1)\xrightarrow{pr^*}0$, and so $S^b_{T,R}(\omega)\xrightarrow{pr^*}0$, $a.s.-\omega$, $\forall\omega\in\Omega_2$. Also, note that

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

so that $S^a_{T,R}(\omega)$ 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

Anderson, H.M. (1997), Transaction Costs and Nonlinear Adjustment Towards Equilibrium in the US Treasury Bill Market, *Oxford Bulletin of Economics and Statistics,* 59, 465-480.

Anderson, H.M. and F. Vahid (1998), Testing Multiple Equation Systems for Common Nonlinear Components, *Journal of Econometrics,* 84, 1-36.

Andrews, D.W.K. and H.-Y. Chen, (1994), Approximately Median-Unbiased Estimation of Autoregressive Models, *Journal of Business and Economic Statistics,* 12, 187-204.

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.

Cubadda, G., (1999), Common Cycles in Seasonal Nonstationary Time Series, *Journal of Applied Econometrics*, 14, 273-291.

Cubadda, G. and A. Hecq, (2001), On Non-Contemporaneous Short-Run Comovements, *Economic Letters*, 73, 389-397.

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), 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.

Hall, A.D., H.M. Anderson and C.W.J. Granger, (1992), A Cointegration Analysis of Treasury Bill Yields, *The Review of Economics and Statistics,* 74, 116-126.

Hamilton, J., (1994), *Time Series Analysis*, Princeton University Press, Princeton.

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

Hecq, A., F.C. Palm and JP Urbain, (2000), Premanent-Transitory Decomposition in VAR Models with Cointegration and Common Cycles, *Oxford Bullettin of Economics and Statistics*, 62, 511-532.

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.

Kramer, W. and L. Davies, (2002), Testing for Unit Roots in the Context of Misspecified Logarithmic Random Walks, *Economic Letters*, 74, 313-319.

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.

Nelson, C.R. and C.I. Plosser, (1982), Trends and Random Walks in Macroeconomic Time Series, *Journal of Monetary Economics,* 10, 129-162.

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.

Proietti, T., (1997), Short-Run Dynamics in Cointegrated Systems, *Oxford Bulletin of Economics and Statistics*, 59, 405-422.

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

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

Stock, J.H. and M.W. Watson, (1988), Testing for Common trends, *Journal of the American Statistical Association*, 83, 1097-1107.

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.

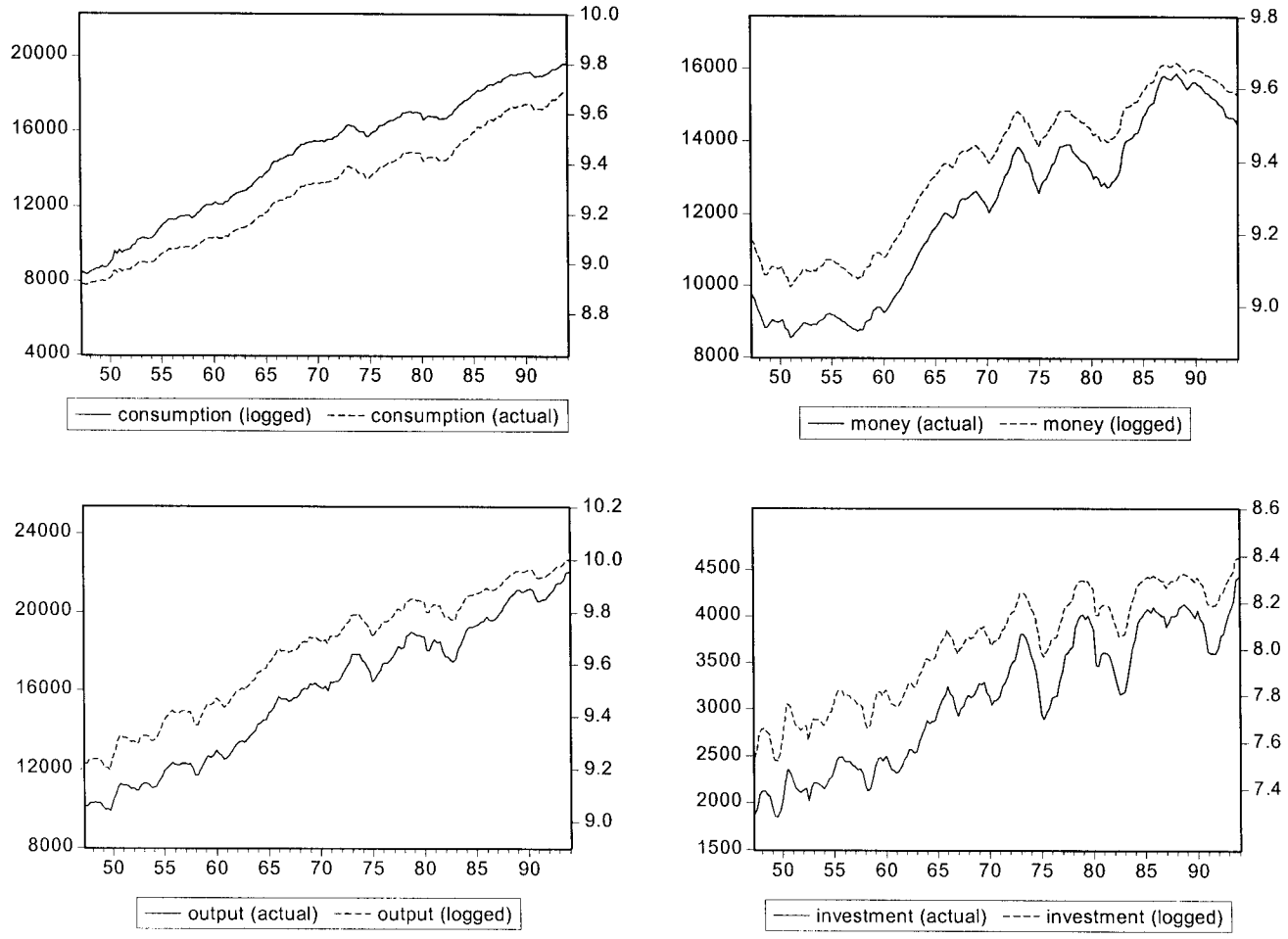
Van Dijk, D. and P.H. Franses (1999), Modelling Multiple Regimes in the Business Cycle, *Macroeconomic Dynamics* 3, 311-340.

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.

Figure 1: Actual and Logged Data

Variables from the King-Plosser-Stock-Watson Dataset (1947-1994)



Notes: See discussion in Section 4 for further details regarding construction of the variables depicted above.

Table 1: 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						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.993	0.999	0.992	1.000	0.995	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.993	0.999	0.992	1.000	0.995	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.993	0.999	0.992	1.000	0.995	1.000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.993	0.999	0.992	1.000	0.995	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.993	0.999	0.992	1.000	0.995	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.993	0.999	0.992	1.000	0.995	1.000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.970	0.993	0.973	0.995	0.492	0.919
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.218	0.773	0.002	0.114	0.002	0.080
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.126	0.526	0.000	0.003	0.000	0.000
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.981	0.999	0.451	0.861	0.003	0.024
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.984	0.999	0.928	0.997	0.583	0.984
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.804	0.954	0.697	0.884	0.313	0.776
Panel B: $\rho = 0.6$						
Data Generated in Logs, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.991	0.997	0.993	1.000	0.991	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.991	0.997	0.993	1.000	0.991	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.991	0.997	0.993	1.000	0.991	1.000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.991	0.997	0.993	1.000	0.991	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.991	0.997	0.993	1.000	0.991	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.991	0.997	0.993	1.000	0.991	1.000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.983	0.997	0.975	0.994	0.540	0.929
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.828	0.982	0.588	0.899	0.009	0.428
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.670	0.911	0.052	0.446	0.000	0.059
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.965	0.974	0.706	0.972	0.018	0.605
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.985	0.998	0.920	0.998	0.574	0.978
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.970	0.995	0.856	0.991	0.476	0.954
Panel C: $\rho = 0.9$						
Data Generated in Logs, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.235	0.221	0.844	0.842	0.993	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.234	0.220	0.844	0.843	0.993	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.235	0.221	0.844	0.843	0.993	1.000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.236	0.220	0.843	0.842	0.993	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.236	0.220	0.843	0.841	0.993	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.236	0.221	0.843	0.842	0.993	1.000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.980	0.996	0.978	0.996	0.541	0.930
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.982	0.998	0.719	0.970	0.221	0.806
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.954	0.990	0.734	0.941	0.123	0.644
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.397	0.379	0.878	0.901	0.760	0.984
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.484	0.468	0.948	0.993	0.694	0.988
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.587	0.579	0.941	0.997	0.695	0.987

(*) Notes: Entries are based on the application of Vahid-Engle common cycle tests and denote the frequency of rejection of $H_0 : s = 2$ (i.e. no common cycles), where the alternative is $H_A : s < 2$ (i.e. one common cycle), based on 5% nominal level tests. Details of the models used to simulate data for these experiments are given above. In all experiments, 5000 Monte Carlo simulations were run (see above for further details).

Table 2: Johansen Cointegration 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						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.999	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	0.995	1.000	1.000	1.000
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.999	1.000	1.000	0.999	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Panel B: $\rho = 0.6$						
Data Generated in Logs, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.999	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.997	0.999	0.997	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Panel C: $\rho = 0.9$						
Data Generated in Logs, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.685	0.682	0.993	0.994	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.683	0.680	0.994	0.995	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.683	0.680	0.994	0.995	1.000	1.000
Data Generated in Levels, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.686	0.683	0.994	0.995	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.687	0.684	0.994	0.995	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.687	0.685	0.994	0.995	1.000	1.000
Data Generated in Logs, Test Done on Levels Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	1.000	1.000	1.000	1.000	1.000	1.000
Data Generated in Levels, Test Done on Logged Data						
$\sigma_u^2 = 0.1\sigma_\epsilon^2$	0.724	0.721	0.978	0.979	1.000	1.000
$\sigma_u^2 = 1.0\sigma_\epsilon^2$	0.847	0.845	1.000	1.000	1.000	1.000
$\sigma_u^2 = 10.0\sigma_\epsilon^2$	0.909	0.910	1.000	1.000	1.000	1.000

(*) Notes: See notes to Table 1. Entries are Johansen trace test statistic rejection frequencies, for the null hypothesis that the cointegrating rank is 0. As our models have two variables, and the alternative of interest is a finding of a cointegrating space of rank 1 or 2, the trace test is the same as the maximum eigenvalue test in our experiments. All results are based on 5% nominal level tests. In all experiments, 5000 Monte Carlo simulations were run.

Table 3a: $S_{T,R}(\omega)$ Test Performance for Various Data Generating Processes (*)

Empirical Level Results								
I(0) Data Generating Processes					I(1) Data Generating Processes			
		T=100	T=250	T=500			T=100	T=250	T=500
$DGPS1$	$R = T^{0.5}$	0.026	0.003	0.030	$DGPS4$	$R = T^{0.5}$	0.018	0.000	0.030
	$R = T^{0.75}$	0.281	0.078	0.020		$R = T^{0.75}$	0.218	0.022	0.020
	$R = T^{0.90}$	0.454	0.280	0.074		$R = T^{0.90}$	0.427	0.114	0.070
	$R = T^{0.95}$	0.513	0.363	0.059		$R = T^{0.95}$	0.510	0.192	0.059
	$R = T^{0.99}$	0.577	0.420	0.091		$R = T^{0.99}$	0.596	0.251	0.087
$DGPS2$	$R = T^{0.5}$	0.017	0.000	0.030	$DGPS5$	$R = T^{0.5}$	0.009	0.000	0.030
	$R = T^{0.75}$	0.216	0.021	0.020		$R = T^{0.75}$	0.066	0.010	0.020
	$R = T^{0.90}$	0.422	0.106	0.070		$R = T^{0.90}$	0.221	0.022	0.070
	$R = T^{0.95}$	0.495	0.187	0.059		$R = T^{0.95}$	0.320	0.059	0.060
	$R = T^{0.99}$	0.569	0.246	0.087		$R = T^{0.99}$	0.416	0.059	0.090
$DGPS3$	$R = T^{0.5}$	0.013	0.000	0.030	$DGPS6$	$R = T^{0.5}$	0.003	0.000	0.030
	$R = T^{0.75}$	0.144	0.011	0.020		$R = T^{0.75}$	0.027	0.010	0.020
	$R = T^{0.90}$	0.356	0.034	0.070		$R = T^{0.90}$	0.054	0.020	0.070
	$R = T^{0.95}$	0.444	0.077	0.060		$R = T^{0.95}$	0.077	0.042	0.060
	$R = T^{0.99}$	0.530	0.094	0.090		$R = T^{0.99}$	0.150	0.055	0.090
Empirical Power Results									
$DGPP1$	$R = T^{0.5}$	0.133	0.251	0.378	$DGPP4$	$R = T^{0.5}$	0.137	0.250	0.367
	$R = T^{0.75}$	0.459	0.588	0.698		$R = T^{0.75}$	0.433	0.533	0.627
	$R = T^{0.90}$	0.600	0.746	0.825		$R = T^{0.90}$	0.543	0.682	0.793
	$R = T^{0.95}$	0.647	0.789	0.843		$R = T^{0.95}$	0.577	0.738	0.839
	$R = T^{0.99}$	0.692	0.814	0.865		$R = T^{0.99}$	0.623	0.776	0.876
$DGPP2$	$R = T^{0.5}$	0.131	0.245	0.362	$DGPP5$	$R = T^{0.5}$	0.132	0.251	0.362
	$R = T^{0.75}$	0.439	0.563	0.672		$R = T^{0.75}$	0.427	0.568	0.707
	$R = T^{0.90}$	0.567	0.725	0.810		$R = T^{0.90}$	0.577	0.771	0.840
	$R = T^{0.95}$	0.611	0.771	0.853		$R = T^{0.95}$	0.638	0.812	0.843
	$R = T^{0.99}$	0.659	0.796	0.882		$R = T^{0.99}$	0.697	0.831	0.854
$DGPP3$	$R = T^{0.5}$	0.132	0.245	0.352	$DGPP6$	$R = T^{0.5}$	0.126	0.257	0.381
	$R = T^{0.75}$	0.427	0.552	0.649		$R = T^{0.75}$	0.465	0.633	0.755
	$R = T^{0.90}$	0.558	0.706	0.809		$R = T^{0.90}$	0.647	0.787	0.866
	$R = T^{0.95}$	0.599	0.753	0.832		$R = T^{0.95}$	0.706	0.819	0.883
	$R = T^{0.99}$	0.647	0.789	0.850		$R = T^{0.99}$	0.750	0.847	0.899

(*) Notes: See discussion in Section 3.3 and list of DGPs used in Table 3b. Entries are rejection frequencies of the $S_{T,R}(\omega)$ test, where the null hypothesis is the data are generated according to a loglinear model that may be either I(0) or I(1). (As discussed above, note that if the null fails to reject, then standard unit root tests to the logged data can be used to assess whether the data are I(0) or I(1). (The finite sample properties of sequentially applied unit root tests of this sort are similar to the finite sample properties of Dickey-Fuller, Phillips-Perron, and other unit root tests examined by many authors in the unit root literature (see e.g. Hamilton (1994)). All results are based on 5% nominal level tests. In all experiments, 5000 Monte Carlo simulations were run.

Table 3b: Data Generating Processes Used in Table 3a (*)

Mnemonic	*DGP*
$DGPS1$	$\log X_t = 2.0 + 0.0020t + \rho \log X_{t-1} + \varepsilon_{1,t}$
$DGPS2$	$\log X_t = 2.0 + 0.0025t + \rho \log X_{t-1} + \varepsilon_{1,t}$
$DGPS3$	$\log X_t = 2.0 + 0.0030t + \rho \log X_{t-1} + \varepsilon_{1,t}$
$DGPS4$	$\log X_t = 0.010 + \log X_{t-1} + \varepsilon_{2,t}$
$DGPS5$	$\log X_t = 0.015 + \log X_{t-1} + \varepsilon_{2,t}$
$DGPS6$	$\log X_t = 0.020 + \log X_{t-1} + \varepsilon_{2,t}$
$DGPP1$	$X_t = 2500 + 15t + \rho X_{t-1} + \varepsilon_{3,t}$
$DGPP2$	$X_t = 2500 + 20t + \rho X_{t-1} + \varepsilon_{3,t}$
$DGPP3$	$X_t = 2500 + 25t + \rho X_{t-1} + \varepsilon_{3,t}$
$DGPP4$	$X_t = 100 + X_{t-1} + \varepsilon_{4,t}$
$DGPP5$	$X_t = 150 + X_{t-1} + \varepsilon_{4,t}$
$DGPP6$	$X_t = 200 + X_{t-1} + \varepsilon_{4,t}$

(*) Notes: See discussion in Section 3.3. In the above models: $\varepsilon_{1,t}$ is *iid* $N(0, 0.005^2)$; $\varepsilon_{2,t}$ is *iid* $N(0, 0.005^2)$; $\varepsilon_{3,t}$ is *iid* $N(0, 150^2)$; and $\varepsilon_{4,t}$ is *iid* $N(0, 150^2)$. All data generating processes where (roughly) calibrated by using parameters consistent with the money, consumption and output variables from the KPSW dataset used to calibrate the experiments reported in Tables 1 and 2 above.

Table 4: Empirical Illustration I: 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	1.212	1.658	2.799	3.589	4.244
investment	2.147	5.828	11.98	15.86	19.52
money	2.590	7.730	16.22	21.42	26.40
output	1.072	0.905	0.950	1.071	1.039
interest rate	1.454	2.864	5.448	7.142	8.706
Panel B: Sample: 1970 quarter 1 - 1994 quarter 1					
consumption	1.270	1.052	1.145	1.328	1.357
investment	1.327	1.415	1.901	2.351	2.580
money	1.809	3.657	6.353	7.967	9.418
output	1.413	1.870	2.820	3.529	4.000
Panel C: Sample: 1984 quarter 1 - 1994 quarter 1					
consumption	2.616	5.312	9.256	11.20	12.77
investment	3.256	7.164	12.67	15.40	17.65
money	2.043	3.666	6.191	7.419	8.425
output	2.398	4.685	8.094	9.778	11.13

* Entries in the table are $S_{T,R}(\omega)$ statistics calculated as discussed above, and are distributed as χ^2_1 random variables so that 1%, 5%, and 10% critical values are 6.635, 3.842, and 2.706, 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). See above for further details.

Table 5: Empirical Illustration II: The Nelson-Plosser 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: 1909 - 1988				
cpi	1.033	0.823	1.039	0.979	1.057
employment	1.123	1.298	1.930	2.064	2.362
gnp deflator	1.015	0.777	0.960	0.881	0.937
industrial production	1.023	0.793	0.991	0.919	0.984
gnp per capita	1.014	0.759	0.917	0.822	0.867
money	1.014	0.734	0.861	0.753	0.782
nominal gnp	1.015	0.733	0.862	0.753	0.783
real gnp	1.036	0.835	1.059	1.003	1.087
real wages	1.418	2.630	4.738	5.457	6.473
s&p500	1.017	0.733	0.874	0.766	0.800
unemployment	1.124	1.307	1.949	2.087	2.390
velocity	1.050	0.917	1.203	1.176	1.296
wages	1.016	0.779	0.964	0.885	0.943
	Panel B: Sample: 1959 - 1988				
cpi	0.992	1.045	0.782	0.767	0.816
employment	1.017	1.030	0.762	0.756	0.800
gnp deflator	0.989	1.052	0.801	0.786	0.838
industrial production	0.999	1.037	0.770	0.758	0.805
gnp per capita	0.994	1.043	0.778	0.764	0.812
money	0.990	1.076	0.846	0.837	0.894
nominal gnp	0.989	1.046	0.788	0.774	0.823
real gnp	0.998	1.222	1.129	1.171	1.288
real wages	1.143	1.780	2.233	2.477	2.770
s&p500	1.101	1.615	1.895	2.074	2.309
unemployment	1.525	3.103	4.763	5.537	6.354
velocity	1.239	2.147	2.949	3.349	3.785
wages	0.984	1.117	0.924	0.929	0.999

* Entries in the table are $S_{T,R}(\omega)$ statistics calculated as discussed above, and are distributed as χ^2_1 random variables so that 1%, 5%, and 10% critical values are 6.635, 3.842, and 2.706, respectively. Data are annual and correspond to those series constructed and examined Schotman and van Dijk (1991) See above for further details.

Table 6: Empirical Illustration III: The Hall-Anderson-Granger 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: 1970 month 1 - 1988 month 12					
R1	4.464	18.43	42.59	56.77	71.35
R2	3.264	12.58	28.83	38.41	48.21
R3	4.984	21.00	48.68	64.91	81.57
R4	4.787	20.03	46.38	61.83	77.71
R5	4.158	16.92	39.00	52.00	65.34
R6	4.589	19.05	44.06	58.74	73.82
R7	5.240	22.25	51.65	68.89	86.57
R8	5.401	23.05	53.52	71.40	89.71
R9	5.779	24.92	57.99	77.38	97.22
R10	5.405	23.06	53.56	71.45	89.79
R11	4.480	18.51	42.77	57.02	71.65
Panel B: Sample: 1984 month 1 - 1988 month 12					
R1	2.551	7.429	13.88	17.18	19.87
R2	1.967	5.074	9.244	11.42	13.09
R3	2.522	7.312	13.65	16.88	19.53
R4	2.548	7.416	13.86	17.14	19.83
R5	2.330	6.532	12.10	14.96	17.27
R6	2.574	7.522	14.07	17.40	20.14
R7	2.894	8.777	16.54	20.48	23.77
R8	2.884	8.742	16.47	20.40	23.67
R9	3.069	9.442	17.81	22.07	25.64
R10	2.921	8.883	16.74	20.74	24.07
R11	2.530	7.341	13.71	16.95	19.61

* Entries in the table are $S_{T,R}(\omega)$ statistics calculated as discussed above, and are distributed as χ^2_1 random variables so that 1%, 5%, and 10% critical values are 6.635, 3.842, and 2.706, respectively. Data are monthly and correspond to those series constructed and examined Hall, Anderson and Granger (1992). In particular, the data are monthly treasury-bill nominal yield to maturity figures - R1 is the series for bills with one month to maturity, R2 is the series for bills with 2 months to maturity, etc. See above for further details.