NONLINEAR TRANSFORMATIONS OF INTEGRATED TIME SERIES

By C. W. J. Granger

University of California, San Diego

Jeff Hallman

Federal Reserve Board

First version received March 1989

Abstract. In this paper we consider the effects of nonlinear transformations on integrated processes and unit root tests performed on such series. A test that is invariant to monotone data transformations is proposed. It is shown that series are generally not cointegrated with nonlinear transformations of themselves, but the same transformation applied to a pair of cointegrated series can result in cointegration between the transformed series.

Keywords. Nonlinear transformations; integrated processes; unit root tests; cointegrated series; monotone data transformations; autocorrelations; Dickey–Fuller statistics.

1. INTRODUCTION

In this paper we are concerned with the effects of nonlinear transformations on integrated, particularly I(1), processes. Three questions are considered.

(i) If x_t is integrated and $z_t = f(x_t)$, will z_t appear to be integrated as well?
(ii) Are x_t and z_t cointegrated?
(iii) If x_t, y_t are I(1) and cointegrated, will $g(x_t)$, $g(y_t)$ also be cointegrated?

These questions arise naturally when considering regressions of the form

$$w_t = a + bx_t + cz_t \text{ (or } y_t) + \text{residuals}$$

where w_t is stationary. The terms x_t, z_t or y_t can only occur on the right-hand side if they are either I(0) or cointegrated. For example, a researcher may try to explain the unemployment rate in terms of r_t and $\log r_t$, where r_t is an interest rate.

The outline of the paper is as follows. Following this introduction, in Section 2 we address question (i) by contrasting the effects of several nonlinear transformations on the empirical autocorrelations and Dickey–Fuller (DF) statistics of a random walk. The DF test appears to be much more sensitive to nonlinear transformation than is the empirical autocorrelation function. A simple modification of the DF is proposed which works

0143-9782/91/03 207–18 \$02.50/0

correctly for a large class of transformations. In Section 3 we consider questions (ii) and (iii); the answers are generally no and yes, respectively, although the DF test is again somewhat misleading.

The topics considered are relevant because of the current interest in integrated, or 'unit root' series in econometrics and macroeconomics, and in nonlinear time series models. Properties of nonlinearly transformed series are examined in greater detail by Granger and Hallman (1988), while nonlinear theoretical relationships between integrated series are considered by Granger (1988) and Hallman (1989).

2. UNIT ROOT TESTS ON TRANSFORMED SERIES

There is now a substantial literature on the topic of testing for unit roots in linear time series models. The result obtained by Phillips (1987) forms the basis for the distributional theory of the various tests. Phillips assumes that a series y_t is generated by

$$y_t = y_{t-1} + u_t$$

where $y_0 = 0$ and u_t is assumed to satisfy the following assumptions.

ASSUMPTION 2.1 (PHILLIPS)

(a) $Eu_t = 0$

(b) $\sup_t E|u_t|^\beta < \infty$ for some $\beta > 2$.

(c) $\sigma^2 = \lim_{T\to\infty} E\left\{\frac{1}{T}\left(\sum u_t\right)^2\right\}$

exists and is greater than zero.

(d) $\{u_t\}_1^\infty$ is strong mixing with coefficients α_m that satisfy $\Sigma_1^\infty \alpha_m^{1-2/\beta} < \infty$.

Given these assumptions, Phillips shows that

$$T(\hat{\alpha} - 1) \equiv T\frac{\Sigma\,\Delta y_t y_{t-1}}{\Sigma\, y_{t-1}^2}$$

$$\to \frac{W(1)^2 - \sigma_u^2/\sigma^2}{2\int_0^1 W^2(r)dr}$$

and

$$t_\alpha \equiv \frac{\hat{\alpha} - 1}{\tilde{s}(\Sigma\, y_{t-1}^2)^{1/2}} \to \frac{(\sigma/\sigma_u)W(1)^2 - 1}{2(\int_0^1 W^2(r)dr)^{1/2}}$$

where $W(r)$ is a standard Brownian motion and $\tilde{s}^2$ is the usual estimate of the variance of the residuals from the regression. The statistic $-t_\alpha$ is called the Dickey–Fuller (DF) test statistic and its distribution is known by the same name. If $y_0 \neq 0$ it is subtracted from the other terms in the series. Models with more complicated serial correlation but still only a single unit root can be handled by including lags of Δy_t in the regression; this is the augmented Dickey–Fuller (ADF) test. For example, the ADF test using four lags is minus the t statistic of the coefficient α in the regression

$$\Delta y_t = \hat{\alpha} y_{t-1} + \sum_{i=1}^{4} \hat{b}_i \Delta y_{t-i}. \tag{2.1}$$

Both the simple and augmented versions of the test have the same limiting distribution.

The test given by (2.1) is designed to have power against the alternative hypothesis that y_t is generated by a stationary AR model with zero mean. A test for the more general alternative where the mean of the series may be nonzero is constructed by performing the same regression with the addition of a constant term, i.e.

$$\Delta y_t = c + \hat{\alpha} y_{t-1} + \sum_{i=1}^{4} \hat{b}_i \Delta y_{t-i}. \tag{2.2}$$

It is interesting to ask how the DF and ADF tests work with nonlinearly transformed series. Suppose that

$$x_t = x_{t-1} + \varepsilon_t,$$

where ε_t meets the requirements of Assumption 2.1, and let

$$y_t = f(x_t).$$

A mean value expansion has

$$y_t = y_{t-1} + \eta_t$$

with

$$\eta_t = f'(x_{t-1} + r_t)\varepsilon_t$$

where r_t lies in the interval $[y_{t-1}, y_t]$. There is no reason to expect η_t to meet the requirements of Phillips' assumption unless $f(\cdot)$ is affine. As examples, consider the following transformations, noting that the term $(\Sigma u_t)^2$ in (c) is just y_t^2.

$y_t = x_t^2$: here $\eta_t = e_t^2 + 2x_{t-1}\varepsilon_t$ and this violates all four parts of Assumption 2.1.

$y_t = x_t^3$: here $\eta_t = 3x_{t-1}\varepsilon_t + 3x_{t-1}\varepsilon_t^2 + \varepsilon_t^3$ also violates all four parts of Assumption 2.1.

$y_t = \operatorname{sgn}(x_t)$: this violates (c) since $(\Sigma_1^T \eta_t)^2 = y_T^2 = 1$, so that $\sigma^2 = 0$.

$y_t = \sin x_t$: Granger and Hallman (1988) show this to be a stationary AR(1) process with variance $\frac{1}{2} + c\alpha^{4t}$, implying that the limit in (c) is

$$\lim_{T\to\infty} \frac{1 + 2c\alpha^{4T}}{2T} = \lim_{T\to\infty} \frac{1}{2T} = 0$$

$y_t = \exp(x_t)$: in this case $\eta_t = \{\exp(\varepsilon_t) - E\exp(\varepsilon_t)\}y_{t-1}$ has a variance exploding faster than t, thus violating (b) and (c). It is also clear from the expression for η_t that it is not mixing.

$y_t = 1/x_t$: to avoid problems associated with x_t taking nonpositive values, assume that x_0 is large. Then y_t will be bounded and $\lim_{T\to\infty}(y_t^2/T)$ will be zero, violating (c). (d) also fails as

$$\eta_t = -\varepsilon_t y_{t-1}^2 + (\varepsilon_t - \sigma^2) y_{t-1}^3.$$

As a simple example of what can happen to the DF test when the series tested is a transformation of a random walk, let x_t be the simplest type of random walk given by

$$x_t = x_{t-1} + \varepsilon_t \tag{2.3}$$

with $x_0 = 0$ and where ε_t $(t = 1, 2, \ldots, T)$ is an independent identically distributed (i.i.d.) series with $\text{prob}(\varepsilon_t = 1) = \text{prob}(\varepsilon_t = -1) = \frac{1}{2}$. x_t meets the conditions of Assumption 2.1, and so the DF test when performed on it will have the DF distribution. Considering the transformed series $y_t = \text{sgn}(x_t)$, it is seen that the change series Δy_t is just

$$\Delta y_t = \begin{cases} 2 & \text{if } x_t > 0 \text{ and } x_{t-1} < 0 \\ -2 & \text{if } x_t < 0 \text{ and } x_{t-1} > 0 \\ 0 & \text{otherwise} \end{cases}$$

so that $\Delta y_t y_{t-1}$ is -2 if x_t crosses zero between time $t-1$ and t, and is zero otherwise. $y_{t-1}^2 = 1$ for all t, of course, and so

$$\text{DF}(y_t) \equiv -\sum_{t=1}^{T} \Delta y_t y_{t-1} \bigg/ \hat{\sigma}\left(\sum_{t=1}^{T} y_{t-1}^2\right)^{1/2} = -\frac{-2 \times \text{no. of zero crossings of } x_t}{(T\hat{\sigma}^2)^{1/2}}.$$

Since $\hat{\sigma}^2$ is just the mean square error (MSE) of the regression of Δy_t on y_{t-1},

$$\hat{\sigma}^2 \leq \frac{1}{T}\sum_{t=1}^{T} \Delta y_t^2 = \frac{4}{T} \times \text{no. of zero crossings of } x_t$$

so that

$$\begin{aligned} \text{DF}(y_t) &\geq \frac{2 \times \text{no. of crossings}}{(4 \times \text{no. of crossings})^{1/2}} \\ &= (\text{no. of zero crossings of } x_t)^{1/2}. \end{aligned}$$

Feller (1968) shows that for the simple random walk given by (3.3) the number of returns to the origin divided by $T^{1/2}$ is asymptotically distributed as a truncated normal random variable. As the probability of a zero crossing is just half the probability of a return to the origin, it follows that twice the number of crossings divided by $T^{1/2}$ has the same distribution. The DF test statistic for y_t is at least $\text{O}(T^{1/4})$ and will become infinitely large as the sample size grows.

Tables I and II show the empirical distributions of the DF and ADF tests

TABLE I

DICKEY–FULLER EMPIRICAL DISTRIBUTION

Transformation	1%	5%	10%	25%	50%	75%	90%	95%	99%
x	−0.68	0.06	0.48	1.05	1.59	2.15	2.62	2.90	3.54
x^2	−2.61	−0.87	0.02	1.15	1.84	2.46	3.21	3.74	4.86
x^3	−4.06	−1.58	−0.11	1.23	1.99	2.65	3.33	3.78	4.78
$\|x\|$	−0.48	0.34	0.80	1.40	2.01	2.60	3.24	3.70	4.76
$\operatorname{sgn}(x)$	1.45	2.16	2.67	3.58	4.58	6.05	8.37	11.31	14.25
$\sin x$	5.75	6.17	6.34	6.70	7.07	7.46	7.82	8.00	8.50
$\exp(x)$	−11.6	2.99	4.04	5.05	6.03	7.22	8.68	10.13	36.06
$\ln(x+75)$	−0.74	0.03	0.50	1.06	1.59	2.14	2.63	2.94	3.50
$1/(x+75)$	−0.96	0.01	0.49	1.07	1.59	2.14	2.66	2.97	3.56

TABLE II

AUGMENTED DICKEY–FULLER EMPIRICAL DISTRIBUTION

Transformation	1%	5%	10%	25%	50%	75%	90%	95%	99%
x	−0.83	0.03	0.40	1.03	1.57	2.14	2.64	2.95	3.58
x^2	−2.63	−1.20	−0.25	1.07	1.82	2.45	3.06	3.44	4.23
x^3	−4.0	−2.03	−0.61	1.04	1.87	2.48	3.04	3.39	3.95
$\|x\|$	−0.78	0.12	0.59	1.29	1.89	2.47	3.05	3.35	4.23
$\operatorname{sgn}(x)$	0.53	1.24	1.56	2.08	2.82	4.01	6.08	6.90	10.92
$\sin x$	3.71	4.08	4.27	4.63	5.06	5.49	5.89	6.15	6.70
$\exp(x)$	−10.5	−2.29	1.88	3.25	4.07	4.72	5.39	7.76	39.1
$\ln(x+75)$	−0.85	−0.03	0.42	1.03	1.58	2.14	2.64	2.94	3.60
$1/(x+75)$	−1.02	−0.07	0.43	1.04	1.59	2.15	2.65	2.97	3.64

on several transformations of a Gaussian random walk. These were found by creating 2000 random walks of length 200, making the indicated transformations and recording the values of the test statistics. Four lags of the dependent variable were used for the ADF statistic, and constants were included in both the ADF and DF regressions.

In these tests the null hypothesis is that the series is I(1) and the alternative is that it is I(0). The first row of Table I shows that the test statistic is less than 2.90 95% of the time when H_0 is true. The results show that not only is H_0 always (correctly) rejected for $\sin x_t$, but it is also usually rejected for the long-memory processes $\operatorname{sgn}(x_t)$ and $\exp(x_t)$. It would certainly be incorrect to accept the latter two series as being I(0). The other transformations are also rejected too often, except for the last two. This is misleading, however, because the effect of adding 75 to the random walk before transforming it is to reduce the curvature of the transformation greatly, making both $\ln(x_t+75)$ and $1/(x_t+75)$ nearly linear transformations of x_t over this range. Adding a smaller constant than 75 would undoubtedly move the DF and ADF distributions to the right. Only those realizations in which x_t crossed the zero axis between observations 5 and 195 were used in obtaining the statistics

TABLE III

AUTOCORRELATIONS

Transformation	Lag 1	2	3	4	5	6	7	8	9	10
x	0.96	0.92	0.87	0.83	0.79	0.75	0.72	0.69	0.65	0.62
x^2	0.93	0.87	0.80	0.74	0.68	0.63	0.58	0.54	0.49	0.45
x^3	0.92	0.85	0.79	0.71	0.64	0.60	0.55	0.50	0.46	0.42
$\|x\|$	0.93	0.87	0.81	0.75	0.70	0.66	0.62	0.58	0.54	0.50
$\operatorname{sgn}(x)$	0.94	0.81	0.76	0.73	0.68	0.65	0.62	0.60	0.58	0.54
$\sin x$	0.59	0.33	0.20	0.12	0.09	0.06	0.01	−0.03	−0.06	−0.05
$\exp(x)$	0.60	0.42	0.30	0.23	0.19	0.18	0.17	0.15	0.12	0.12
$\ln(x+50)$	0.96	0.91	0.87	0.83	0.79	0.75	0.72	0.68	0.65	0.61
$1/(x+50)$	0.96	0.91	0.87	0.83	0.78	0.75	0.71	0.68	0.65	0.61

for the transformation $\operatorname{sgn}(x_t)$. About 85% of the realizations in the simulation had at least one such crossing.

In the Box–Jenkins modeling strategy, the shape of the correlogram is used to decide whether a series seems to be I(0) or I(1). If the autocorrelations decline slowly with lag length, an I(1) model is chosen. Table III presents the means across replications of the first ten autocorrelations obtained in an experiment similar to the one generating Tables I and II. For most of the transformed series, the correlogram closely resembles that of a random walk, even for the bounded series $\operatorname{sgn}(x)$. The exceptions are the stationary series $\sin x_t$ and the explosive series $\exp(x_t)$.

The DF and ADF unit root tests appear to be more sensitive than the autocorrelations to series transformations. Economists often transform their variables by taking logarithms, using Box–Cox transformations etc., before building models and making inferences. It can easily happen that a unit root exists in the original series, but the usual tests reject a unit root in the transformed series despite a high degree of autocorrelation in the latter. A test for unit roots that is invariant to a broad class of transformations would avoid this outcome.

It is not possible to construct a test which is invariant to every possible transformation of the series being tested. As a trivial example, consider the transformation $T(x_t) = 568.3$. No test on the transformed series can possibly yield any information about the original series. More interestingly, Granger and Hallman (1988) show that the sine (or cosine) of a random walk yields a stationary AR(1). This suggests that any periodic transformation will result in a stationary series, since periodic transformations can be arbitrarily well approximated by Fourier transforms. Given such a series with no long memory properties, it will not be possible to detect that it resulted from transforming a random walk.

Many nonparametric tests are based on notions of rank and ordering. Since the ordering of a series is unaffected by strictly monotone transformations, tests based on these notions have distributions that are unaffected by monotone transformations of the data. The ranks R_t of a time series x_t are defined by

$$R_t = \text{the rank of } x_t \text{ among } x_1, x_2, \ldots, x_T.$$

A simple test for unit roots in a (possibly) transformed series is to calculate the DF or ADF statistic of the ranks of the series rather than of the original series itself. Here the null hypothesis is that there exists a strictly monotone transformation of the time series being tested which has a unit root.

The question immediately arises: what is the distribution of what will be called the rank Dickey–Fuller (RDF) statistic and its augmented cousin (RADF)? Unfortunately we have not obtained an analytical answer to this question. Phillips' distributional results can be extended via the continuous mapping theorem to find the distribution of a test for a *given* transformation, but this will not solve the problem since the rank transformation is different for every sample. Rank statistics are usually applied in situations where it is known that the normalized sample ranks $R(x_i)/N$ converge to the population distribution function $F(x_i)$. For the null hypothesis here, there is no well-defined distribution function for the ranks to converge to, as x_t is a nonstationary random walk.

Despite the fact that 'nice' analytic representations of their distribution functions are not available, the RDF and RADF statistics are random variables which are easily computed for any given time series. The approach taken here is to investigate their usefulness as tests in some specific cases by means of computer simulation. Figure 1 shows estimated densities for the RDF statistic (with constant) for sample sizes 25, 50, 100, 200, 400 and 800. The plots were constructed by generating 5000 independent random walks of the indicated sample sizes, calculating and recording the RDF test statistics and finally estimating the density with a kernel estimator. As the figure indicates, the density does not change much as the sample size changes, except for the smallest sample size of 25. Figure 2 shows the corresponding densities for the augmented version of the test, RADF. Here there is marked variation in the density as the sample size varies, but this is also true of the ADF test as seen in Figure 3. The fact that an elegant asymptotic theory is available for the ADF but not for the RADF does not seem to make much difference to their small-sample behavior. Tables IV and V give percentiles of the RDF and RADF tests under the null hypothesis that x_t is a monotone transformation of a pure random walk.

Figures 4 and 5 compare the power of the RDF test with that of the DF. The upper left panel of Figure 4, for example, shows the fraction of rejections of the hypothesis $H_0 : \rho = 0$ in the model

$$\Delta x_t = -\rho x_{t-1} + \varepsilon_t$$

for several values of ρ when the sample size is 50. The four lines on the plot show the rejection percentages for the DF and RDF at the 5% and 10% significance levels. Figure 4 compares the tests when there is no constant

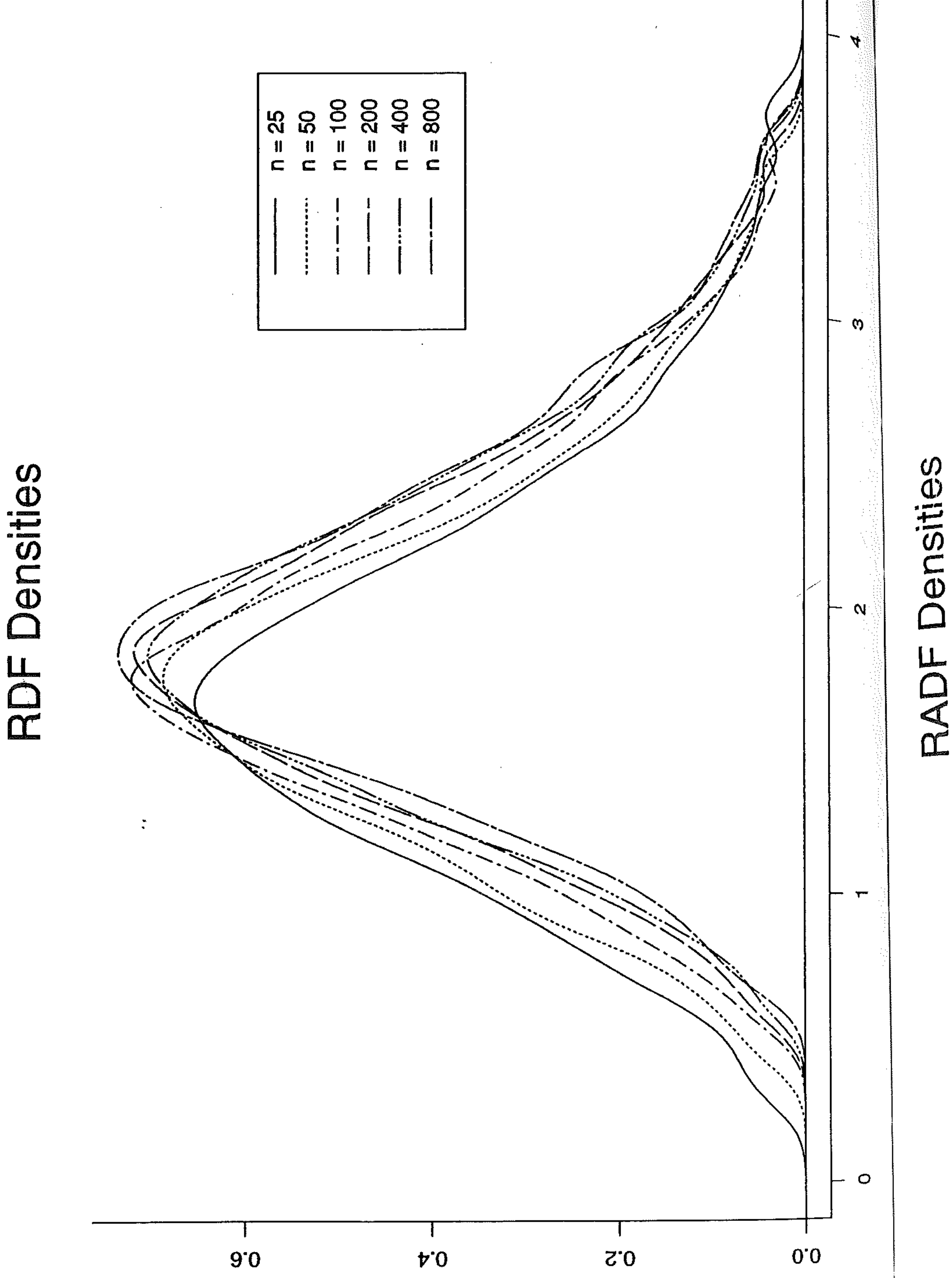

FIGURE 1. Rank Dickey–Fuller densities.

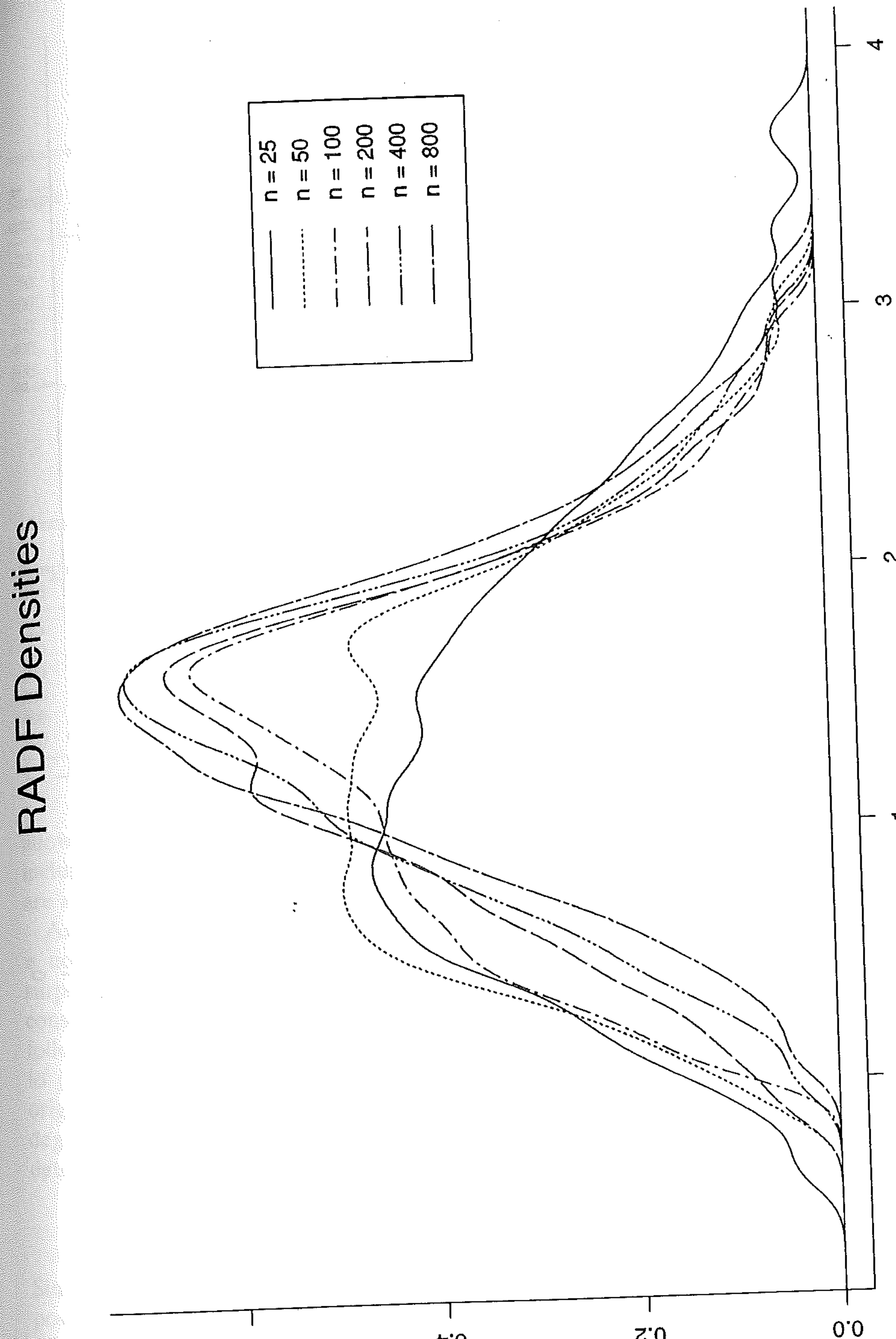

FIGURE 2. Rank augmented Dickey–Fuller densities.

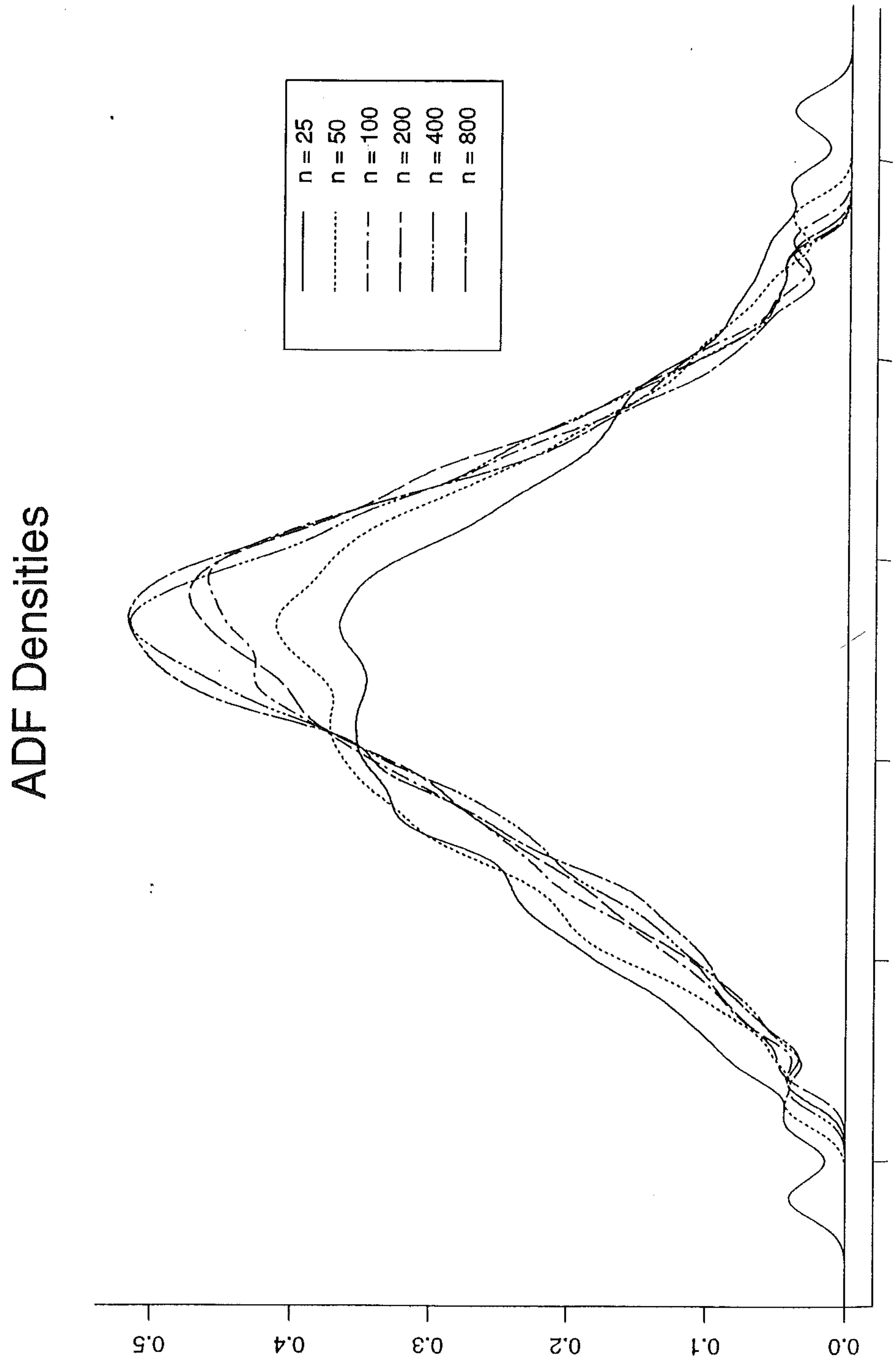

FIGURE 3. Augmented Dickey–Fuller densities.

TABLE IV

RANK DICKEY–FULLER PERCENTILES

Sample size	Without constant 10%	5%	1%	Constant included 10%	5%	1%
25	1.70	2.03	2.71	2.63	2.98	3.70
50	1.77	2.13	2.79	2.63	2.93	3.49
100	1.82	2.14	2.76	2.68	2.95	3.60
200	1.87	2.18	2.80	2.71	3.00	3.53
400	1.88	2.18	2.82	2.75	3.01	3.57
800	1.97	2.28	2.83	2.78	3.06	3.59

TABLE V

RANK AUGMENTED DICKEY–FULLER PERCENTILES

Sample size	Without constant 10%	5%	1%	Constant included 10%	5%	1%
25	1.67	2.05	2.87	2.39	2.72	3.48
50	1.57	1.91	2.56	2.37	2.66	3.25
100	1.61	1.92	2.52	2.41	2.68	3.24
200	1.66	1.95	2.57	2.48	2.75	3.27
400	1.70	2.04	2.61	2.55	2.82	3.42
800	1.79	2.08	2.73	2.65	2.92	3.51

allowed in the regression, while Figure 5 compares the tests with a constant included. Similar power comparisons between the ADF and RADF statistics are a topic for further research.

As the DF test without a constant is equivalent to a likelihood ratio test, it is not surprising that it is more powerful than its rank counterpart. What is surprising is that the RDF is apparently more powerful than the DF when constants are allowed into the regression. A possible explanation is as follows. When there are no lags of Δy_t involved, regression (2.2) is equivalent to (2.1) using the mean-corrected $\tilde{y}_t = y_t - \bar{y}$. When this is done with the original data, the mean is a parameter that has to be estimated, using up a degree of freedom. For ranks, however, the mean is almost completely determined by the number of observations—it is just

$$\frac{N(N+1)}{2} - \frac{\text{rank}(y_N)}{N}.$$

Since it is nearly deterministic, having to estimate it has little effect on the power of the test.

Finally, Tables VI and VII show the empirical distributions of the RDF and RADF tests from a simulation in which the statistics were computed for 200 observations of the indicated transformations of a pure random walk. 500

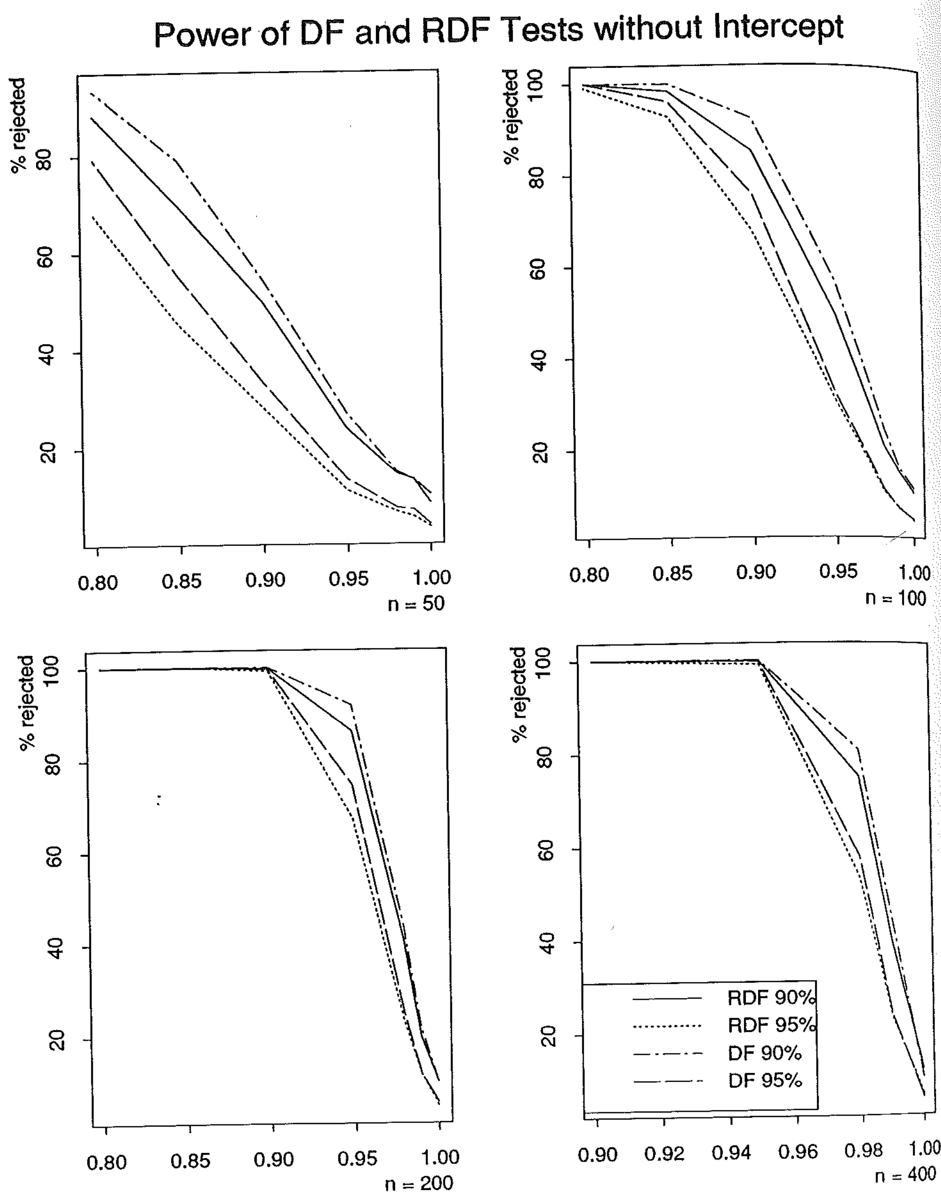

FIGURE 4. Power of Dickey–Fuller and rank Dickey–Fuller tests without intercept.

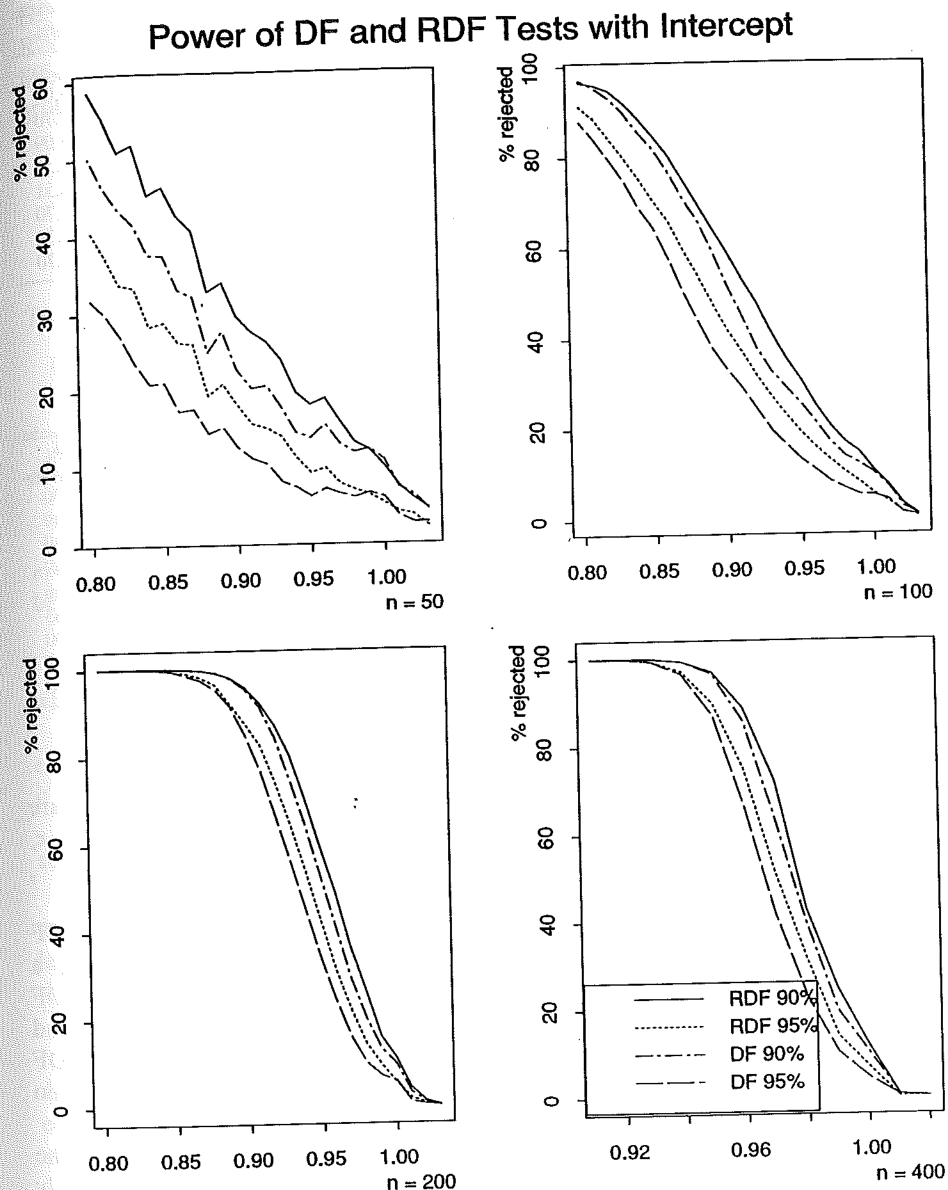

FIGURE 5. Power of Dickey–Fuller and rank Dickey–Fuller tests with intercept.

TABLE VI

RANK DICKEY–FULLER (WITH CONSTANT) EMPIRICAL DISTRIBUTION

Transformation	1%	5%	10%	25%	50%	75%	90%	95%	99%
x	0.49	0.91	1.16	1.48	1.90	2.34	2.81	3.06	3.70
x^2	0.97	1.36	1.57	1.95	2.43	3.11	3.77	4.22	5.09
x^3	0.49	0.91	1.16	1.48	1.90	2.34	2.81	3.06	3.70
$\|x\|$	0.97	1.36	1.57	1.95	2.43	3.11	3.77	4.22	5.09
$\mathrm{sgn}(x)$	−0.99	−0.09	0.21	1.57	2.84	4.11	5.15	5.53	6.53
$\sin x$	6.14	6.45	6.58	6.92	7.25	7.63	7.96	8.14	8.51
$\exp(x)$	0.49	0.91	1.16	1.48	1.90	2.34	2.81	3.06	3.70
$\ln(x+75)$	0.49	0.91	1.16	1.48	1.90	2.34	2.81	3.06	3.70
$1/(x+75)$	0.49	0.91	1.16	1.48	1.90	2.34	2.81	3.06	3.70

TABLE VII

RANK AUGMENTED DICKEY–FULLER (WITH CONSTANT) EMPIRICAL DISTRIBUTION

Transformation	1%	5%	10%	25%	50%	75%	90%	95%	99%
x	0.22	0.71	0.91	1.28	1.74	2.14	2.59	2.89	3.30
x^2	0.67	0.98	1.21	1.59	2.09	2.61	3.17	3.49	4.49
x^3	0.22	0.71	0.91	1.28	1.74	2.14	2.59	2.89	3.30
$\|x\|$	0.67	0.98	1.21	1.59	2.09	2.61	3.17	3.49	4.49
$\mathrm{sgn}(x)$	−1.24	−0.42	0.01	0.53	1.59	2.33	2.92	3.53	4.16
$\sin x$	3.71	4.09	4.27	4.67	5.06	5.54	5.93	6.22	6.63
$\exp(x)$	0.22	0.71	0.91	1.28	1.74	2.14	2.59	2.89	3.30
$\ln(x+75)$	0.22	0.71	0.91	1.28	1.74	2.14	2.59	2.89	3.30
$1/(x+75)$	0.22	0.71	0.91	1.28	1.74	2.14	2.59	2.89	3.30

trials were performed to obtain the percentiles shown. Rank statistics are invariant to monotone transformations, and so the computed statistics for x, x^3, $\exp(x)$, $\ln(x+75)$, and $1/(x+75)$ are all identical. Since $|x| = (x^2)^{1/2}$, their statistics are also identical.

For the strictly monotone transformations in the tables, RDF and RADF have the correct size by construction. For the other transformations, a comparison of Tables VI and VII with Tables I and II indicates that the RDF and RADF distributions appear considerably more robust than the DF and ADF distributions. Only for the $\sin x$ transformation do the RDF and RADF tests consistently reject the null hypothesis, but this is the correct thing to do as $\sin x_t$ is a stationary AR(1).

A reasonable strategy for unit root testing is to compute both the conventional and the rank versions of the DF and ADF tests, since it is rarely known with certainty that the underlying data-generating process (DGP) is linear. If it is, both kinds of tests have the correct size and similar power. Otherwise the rank versions of the tests are more applicable. If the ADF test rejects its null while the RADF does not, for example, we might look at a plot of the rank transformation to see if it is suggestive of a parametric transformation yielding a series that could reasonably be modeled as a linear I(1) process. The fact that all the transformed series in Tables I and II have DF and ADF distributions shifted to the right indicates that the case where RADF rejects and ADF does not is unlikely unless the process really is linear. In this case we might find the ADF test more believable on the grounds that its asymptotic distribution has been worked out.

3. COINTEGRATED VARIABLES

Two questions will be considered.

(i) If x_t is I(1), can x_t and $g(x_t)$ be cointegrated for some function $g(\cdot)$?
(ii) If x_t, y_t are I(1) and cointegrated, will $g(x_t)$, $g(y_t)$ also be cointegrated?

It will be assumed that x_t is a pure Gaussian random walk, possibly with drift, generated by

$$x_t = m + x_{t-1} + e_t$$

$e_t \sim$ i.i.d. $\mathrm{N}(0, \sigma^2)$, so that $x_t \sim \mathrm{N}(mt, \sigma^2 t)$. For the second question, y_t will be assumed to be given by

$$y_t = \alpha x_t + \varepsilon_t$$

where ε is i.i.d. Gaussian, mean zero and independent of e_t.

Denote $E\{g(x_t)\} \equiv \mu_t$. In general, this will be a function of time. For example, if $g(x) = x^2$, then $\mu_t = m^2 t^2 + \sigma^2 t$ which is a function of time even if x_t has no drift.

If $g(x_t)$, x_t are cointegrated with a constant cointegrating parameter α, then

$$g(x_t) - \mu_t = \alpha x_t + a_t$$

where a_t is I(0). $\hat{a}_t$ will be uncorrelated with x_t if α is estimated by ordinary least squares (OLS). Is there a constant α such that $g(x_t) - \mu_t - \alpha x_t$ is I(0)? A simple form of Stein's lemma says that if x is Gaussian then

$$\operatorname{cov}\{g(x)x\} = E\{g'(x)\} \operatorname{var}(x)$$

and so the OLS estimate of α tends asymptotically to $E\{g'(x_t)\}$. There are essentially three cases.

(i) $\lim_{t\to\infty} E\{g'(x_t)\} = c$, a constant, in which case cointegration will occur.
(ii) $\lim_{t\to\infty} E\{g'(x_t)\} = 0$ and there is no cointegration.
(iii) $\lim_{t\to\infty} E\{g'(x_t)\} = G_t$, a function of time. In this case there is no constant-parameter cointegration. There may or may not be time-varying parameter cointegration, but this will not be considered in this paper.

It is easily seen that if $g(x) = \alpha x^k$ for some integer k, then x_t and $g(x_t)$ can only be (constant-parameter) cointegrated if $k = 1$. Similarly, if $g(x) = \exp(\lambda x)$, there cannot be cointegration.

An example where apparent cointegration might seem possible is when $g(x) = \ln(a + x)$, where a is large and positive throughout the sample period and it is assumed that x_t has no drift, so that $m = 0$. In this case,

$$g'(x) = \frac{1}{a + x}$$

$$\approx a^{-1}\left(1 - \frac{x}{a} + \frac{x^2}{a^2}\right) + O(a^{-4})$$

so that

$$E\{g'(x)\} = \frac{1}{a} + \frac{\sigma^2 t}{a^3} + O(a^{-4})$$

Provided that σ^2 times the number of observations included in a sample is small compared with a^3, $E\{g'(x)\}$ will approximate the (small) constant $1/a$ and apparent constant-parameter cointegration may occur.

In Tables VIII and IX the results of tests for cointegration (DF and ADF) between x and $g(x)$ are given for several functions $g(\cdot)$. Selected percentiles of the empirical distribution of the DF and ADF tests performed on the residuals of a regression of x_t on the indicated function $g(x_t)$, where x_t is a pure random walk of 200 observations, are shown. The tables are based on a simulation experiment with 500 trials for each function.

TABLE VIII

PERCENTILES OF DICKEY–FULLER COINTEGRATION TEST

Transformation	55%	60%	65%	70%	75%	80%	85%	90%	95%
x^2	2.81	2.95	3.13	3.33	3.52	3.77	4.04	4.34	4.75
x^3	3.16	3.30	3.45	3.62	3.82	4.07	4.27	4.67	5.24
$\sin x$	1.87	1.96	2.14	2.23	2.38	2.56	2.67	2.88	3.12
$\exp(x)$	2.96	3.10	3.21	3.30	3.44	3.57	3.76	4.06	4.49
$\ln(x + 75)$	3.13	3.26	3.42	3.58	3.74	4.01	4.27	4.53	5.06
$1/(x + 75)$	3.15	3.26	3.43	3.59	3.73	3.99	4.29	4.54	5.06

TABLE IX

PERCENTILES OF AUGMENTED DICKEY–FULLER COINTEGRATION TEST

Transformation	55%	60%	65%	70%	75%	80%	85%	90%	95%
x^2	2.69	2.84	2.97	3.10	3.36	3.58	3.75	4.19	4.87
x^3	2.83	2.96	3.08	3.23	3.42	3.62	3.82	4.07	4.33
$\sin x$	1.69	1.80	1.94	2.05	2.16	2.28	2.47	2.65	3.06
$\exp(x)$	2.00	2.10	2.18	2.30	2.39	2.52	2.74	2.88	3.30
$\ln(x + 75)$	2.98	3.09	3.26	3.39	3.54	3.74	3.95	4.25	4.59
$1/(x + 75)$	2.96	3.11	3.23	3.40	3.54	3.75	3.92	4.25	4.58

The values in the table can be compared with the 5% and 10% critical values of the DF(3.37, 3.02) and ADF(3.25, 2.98) tests for cointegration from Engle and Yoo (1987). It is seen that, except for the sine function, the cointegration tests can be somewhat misleading. For all the other transformations, the tests find cointegration a third or more of the time when it should not theoretically be there. It should be noted that the critical values for these tests were found using *independent* series x_t, y_t. Certainly x_t and $g(x_t)$ are not independent of each other.

Turning to the second question, a mean value expansion shows that

$$\begin{aligned} g(y_t) &= g(\alpha x_t + \varepsilon_t) \\ &\approx g(\alpha x_t) + \varepsilon_t g'(\alpha x_t + r_t) \end{aligned}$$

where r_t is some remainder term. As seen in Section 4, the second term will generally appear to be I(0) in mean with some heteroskedasticity, particularly if x_t, ε_t are independent. If it is assumed that this is correct, $g(y_t) - g(\alpha x_t)$ is I(0). It follows that $g(x_t)$, $g(y_t)$ are cointegrated if either (i) $\alpha = 1$ or (ii) $g(x)$ is homogeneous, so that $g(\alpha x) = \alpha^\lambda g(x)$, in which case the cointegrating parameter is α^λ. It should be pointed out that these results are only approximate.

The answer to the two questions posed at the beginning of the section are generally no and yes respectively. The second case requires $g(\cdot)$ to be homogeneous or the series to be scaled so that the cointegrating coefficient is 1. Granger and Hallman (1988) give an example where x_t, y_t are not cointegrated but x_t^2, y_t^2 are.

4. CONCLUSIONS

Nonlinear transformations of integrated series generally retain the long memory properties of traditional I(1) series, such as slowly declining autocorrelations. However, the DF and ADF unit root tests performed on such transformed series will often reject the null hypothesis that the series was generated by a linear process with a unit root. Since an investigator is rarely certain that the generating process for his data is in fact linear, a unit root test that is invariant to monotone data transformations is desirable. The test proposed here is to perform the DF or ADF test on the ranks of the series, rather than on the series itself. The power functions of the rank tests are very close to the power functions of the conventional tests, but the rank versions have the desired invariance property by construction.

In theory, a nonlinearly transformed series generally cannot be cointegrated with the original series. This emphasizes the importance of having the correct functional form when investigating a hypothesized long-run relationship $y_t = f(x_t)$. If the actual cointegrating relationship is $y_t = g(x_t)$, then y_t and $f(x_t)$ will be cointegrated only if g is an affine transformation of f. Hallman

(1989) addresses this issue. Testing for cointegration by performing unit root tests on the residuals from a regression of x_t on $f(x_t)$ can be misleading, often finding cointegration when it theoretically cannot be there.

Finally, if x_t, y_t are cointegrated series, then $g(x_t)$, $g(y_t)$ can also be cointegrated if either (i) $g(\cdot)$ is homogeneous or (ii) the data are scaled so that the cointegrating coefficient for x_t, y_t is 1.

ACKNOWLEDGEMENTS

This paper was prepared under National Science Foundation grant SES 8902950.

REFERENCES

ENGLE, R. F. and YOO, B. S. (1987) Forecasting and testing in cointegrated systems. *J. Economet.* 35, 143–59.

FELLER, W. (1968) *An Introduction to Probability Theory and Its Application,* Vol. 1. New York: Wiley.

GRANGER, C. W. J. (1988) Introduction to processes having equilibria as simple attractors: the Markov case. Discussion Paper, University of California, San Diego.

— and HALLMAN, J. J. (1988) The algebra of I(1). Finance and Economics Discussion Series 45, Board of Governors of the Federal Reserve System.

HALLMAN, J. J. (1989) Cointegration with transformed variables. Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System. In preparation.

PHILLIPS, P. C. B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.