A POWERFUL, SIMPLE TEST FOR COINTEGRATION USING COCHRANE–ORCUTT

Bruce E. Hansen[0]

University of Rochester

April 1990

Revised: May 1990

[0]I am grateful to Adrian Pagan for helpful comments on an earlier draft. I have also benefited from discussions with Bill Brown, Robert Engle, Masao Ogaki, Changyong Rhee, Jim Stock and Mark Watson.

A POWERFUL, SIMPLE TEST FOR COINTEGRATION USING COCHRANE–ORCUTT

Bruce E. Hansen
Department of Economics
Harkness Hall
University of Rochester
Rochester, NY 14627
(716) 275–7307

ABSTRACT

It is well known that tests for cointegration have low power against reasonable alternatives in typical sample sizes. This paper analyzes existing tests and isolates a common problem. The distributional theory depends upon the dimensionality of the system (the number of variables, n), which is due to the fact that all n unit roots in the system are estimated. This does not make effective use of the structure of the alternative hypothesis, since typically the alternative of interest is a low dimensional cointegrating space.

This paper proposes a simple, powerful test using Cochrane–Orcutt. Under the null hypothesis of no cointegration, the asymptotic distribution of the AR(1) parameter estimate is identical to the Dickey–Fuller coefficient test for a univariate series. Standard unit root tests which allow for serial correlation may be employed on the Cochrane–Orcutt residuals under the distributional theory for the univariate case. This result stands in stark contrast to the results of Stock and Watson (1988), Johansen (1988) and Phillips and Ouliaris (1990), whose distributional theory depends upon n, the dimensionality of the system. Monte Carlo evidence is provided for the dramatic improvements in power achieved by this approach.

Keywords: Cointegration, unit roots, Cochrane–Orcutt, Power properties.

1. Introduction

Since Granger (1981) introduced the concept of cointegration, a number of authors have suggested various tests to determine the number of cointegrating vectors in a system of I(1) variables. Engle and Granger (1987) suggested using an Augmented Dickey–Fuller (ADF) test on the ordinary least squares (OLS) residuals from a regression of one variable upon the others. The asymptotic theory for this OLS "residual–based" approach has been developed by Engle and Yoo (1988), Phillips and Ouliaris (1990) and Hansen (1990). A likelihood ratio (LR) test using full information maximum likelihood estimation (MLE) was proposed and developed by Johansen (1988, 1989). Alternative tests have been proposed by Phillips and Durlauf (1986), Stock and Watson (1988), Phillips and Ouliaris (1988), Stock (1988), and Park, Ouliaris, and Choi (1988), among others.

Despite the large number of papers, the theory is still undeveloped. The large number of tests, each with their own non–standard distributions, is confusing to many applied researchers. The relationship between the Stock–Watson, Johansen, and residual–based tests has not been examined in the previous literature.

This paper discusses these tests in a unified framework with some promising results. The distributions of existing test statistics are shown to depend upon the number of variables in the system. One effect is that power decreases with the size of the system. The common cause of this dependence is that the tests estimate all the unit roots under the null. This ignores the structure of the alternative hypothesis, which most commonly is a low dimensional cointegrating space. For example, in a two variable system, the null hypothesis is no cointegration, or the presence of two unit roots. The alternative is that the variables are cointegrated, or that the two variables have one unit root between them. It therefore seems superfluous to estimate both roots.

An alternative GLS approach is proposed which avoids this problem. The procedure is quite simple to implement, reducing in a single equation setting to iterated

Cochrane–Orcutt regression. The asymptotic distribution of the AR(1) coefficient is identical to the distribution of a univariate Dickey–Fuller coefficient test. Some size distortion exists in finite samples, which a Monte Carlo analysis suggests can be corrected by an bias adjustment. Corrections for residual serial correlation can be made, as in Phillips (1987), Said and Dickey (1984), Stock (1988), or Park and Choi (1988).

The rest of the paper is organized as follows. Section 2 introduces the model and assumptions. Section 3 reviews the common cointegration tests: Stock–Watson, Johansen, and Phillips–Ouliaris, and demonstrates their curse of dimensionality. Section 5 studies Cochrane–Orcutt estimation in a single equation setting and derives the asymptotic null distribution. Section 5 presents an adjustment which reduces size distortion in small samples and presents evidence on power comparisons. Section 6 extends the analysis to allow for serial correlation. Section 7 extends the analysis to allow for fitted intercepts and trends. Section 8 extends the model to allow for tests for multiple cointegrating vectors.

Some notational conventions are given as follows. The symbol " $\Longrightarrow$ " denotes weak convergence of probability measures, " $\equiv$ " denotes equality in distribution, $BM(\Omega)$ denotes a process distributed as a vector Brownian motion with covariance matrix Ω , "[c]" denote the greatest integer less than or equal to c , and I(1) denotes a stochastic process integrated of order one. Arguments of functionals on the space [0,1] are frequently suppressed, and integrals on the space [0,1] such as $\int_0^1 B(s)$ are written as $\int_0^1 B$ and sometimes as $\int B$ to reduce notation. All limits are taken as the sample size T diverges to positive infinity. Proofs are left to an appendix.

2. Model

The question is whether or not the $n \times 1$ stochastic process $\{x_t : t = 1, \ldots\}$ is *cointegrated.* This requires a maintained assumption that each element of x_t is I(1), or "difference stationary". Under standard assumptions (finite second moments and asymptotically independent increments) this allows the application of the invariance principle:

(1) $$T^{-1/2} x_{[Tr]} \Rightarrow B(r) \equiv BM(\Omega)$$

where

(2) $$\Omega = \lim_{T\to\infty} T^{-1} E(x_T x_T') .$$

The null hypothesis is of *no cointegration,* which may be specified as

$$H_0 : \Omega > 0 .$$

The alternative is that Ω is of deficient rank. If $n > 2$, it is also of interest to uncover the *number* of distinct cointegrating vectors, which is equivalent to the *rank deficiency* of Ω. These issues are fairly well understood and have been discussed at length in the existing literature.

The asymptotic theory used in this paper will also require convergence to the matrix stochastic integral:

(3) $$T^{-1}\Sigma_1^T x_{t-1}\Delta x_t' \Rightarrow \int_0^1 B dB' + \Lambda , \qquad \Lambda = \lim_{T\to\infty} T^{-1}\Sigma_1^T E(x_{t-1}\Delta x_t') .$$

An assumption that is sufficient for (1) and (3) and the other results in this paper is

(4) $\{\Delta x_t\}$ is weakly stationary and strong mixing with mixing coefficients $\{\alpha_m\}$ satisfying $\Sigma_{m=1}^{\infty} m^2 \alpha_m^{1-1/r} < \infty$, and $\sup_{t\geq 1} E|\Delta x_t' \Delta x_t|^{2r} < \infty$, for some $r > 1$.

3. Cointegration Tests and the Curse of Dimensionality

There are two broad classes of tests which take the null of no cointegration. The first generalizes the Dickey–Fuller univariate unit root testing methodology directly to the multivariate case. This includes the Stock–Watson common trends test (Stock and Watson, 1988) and Johansen's likelihood ratio (LR) test (Johansen, 1988). The second class of tests apply a standard (univariate) unit root test to a least squares residual. This includes most of the tests discussed by Engle and Granger (1987), Engle and Yoo (1988), Phillips and Ouliaris (1990), Hansen (1990), and Park, Ouliaris and Choi (1988).

Stock and Watson (1988) fit by multivariate least squares

$$x_t = \hat{R}\, x_{t-1} + \hat{u}_t$$

and base their test upon $\hat{R}$, for under the null hypothesis, $\hat{R} \rightarrow_p I_n$. Serial correlation in Δx_t can be incorporated via prefiltering (as in Dickey and Fuller (1979)) or bias correction (as in Phillips (1987)). Stock–Watson suggest examining the latent roots of $\hat{R}$.

Johansen's test is quite similar in spirit. Serial correlation is directly handled by estimating

$$x_t = R\, x_{t-1} + A(L)\Delta x_{t-1} + e_t$$

by maximum likelihood. Under the null of no cointegration, $R = I_n$, which Johansen suggests testing by the likelihood ratio statistic. When the innovations are normal, this reduces to the Wald statistic for the test of $R = I$.[1] Both the Stock–Watson and Johansen tests can be seen as multivariate generalizations of the univariate unit root testing methodology: they take the null hypothesis that the n–variable system has n unit roots in the autoregressive representation, and test that all the leading roots are

[1]Johansen actually discusses a more general situation which allows for r cointegrating vectors under the null, where $r < n$. This is rarely used in practice, and not discussed here.

unity.

The second class of cointegration tests attempt to reduce the multivariate problem to a univariate test upon a least squares residual. Partition x_t as

$$x_t = (\underset{1}{x_{1t}} \ \underset{n-1}{x_{2t}'})'$$

and note that under the null hypothesis of no cointegration, for any $(n-1) \times 1$ vector α, if we define

$$u_t = x_{1t} - \alpha' x_{2t}$$

then $u_t \equiv I(1)$ and if we fit an AR(1) to u_t:

$$u_t = \hat{\rho}\, u_{t-1} + \hat{e}_t$$

then $\hat{\rho} \longrightarrow_p 1$. Under the alternative, some vector α exists such that for some $\rho < 1$, $u_t \equiv I(0)$ and thus $\hat{\rho} \longrightarrow_p \rho$. In practice, α is unknown and the common practice is to use a least squares regression of x_{1t} upon x_{2t}.

Unfortunately, the distributional theory of this procedure is not identical to the theory for the univariate case. Under the null, the OLS coefficient estimate $\hat{\alpha}$ does not converge to a constant, but stays random in the limit. This is part of the phenomenon of spurious regression; see Granger and Newbold (1974) and Phillips (1986). This inconsistency affects the behavior of the test statistics constructed from the residuals. It is not that difficult to see how this happens. The least squares residuals are

$$\hat{u}_t = x_{1t} - \Sigma_1^T x_{1t} x_{2t}' \left[\Sigma_1^T x_{2t} x_{2t}'\right]^{-1} x_{2t}$$

so standard invariance principle arguments give

$$(5) \quad T^{-1/2}\hat{u}_{[Tr]} = T^{-1/2}x_{1[Tr]} - \left[T^{-2}\Sigma_1^T x_{1t}x_{2t}'\right]\left[T^{-2}\Sigma_1^T x_{2t}x_{2t}'\right]^{-1}\left[T^{-1/2}x_{2[Tr]}\right]$$

$$\Rightarrow B_1(r) - \int_0^1 B_1 B_2' \left[\int_0^1 B_2 B_2'\right]^{-1} B_2(r) = B^*(r).$$

B^* is the residual from the continuous time regression of B_1 upon B_2. Unit root tests constructed from $\{\hat{u}_t\}$ can be shown to depend upon this random element, which depends upon n, the number of elements in the system.

This dependence upon dimensionality is important. Examination of the critical values in the tables of Stock and Watson (1988), Johansen (1988) and Phillips and Ouliaris (1990) reveals that the asymptotic distributions of the test statistics shift away from the origin as the dimensionality increases. Thus *larger* test statistics are needed for rejection, implying that *smaller* estimated AR(1) parameters are needed. This is expected to reduce power. To illustrate this fact I report a simple Monte Carlo experiment[2]. 5000 draws of samples of length 100 were made from the process

$$x_{1t} = \rho x_{1t-1} + e_{1t}$$

$$x_{2t} = x_{2t-1} + e_{2t}$$

$$e_t = (e_{1t}, e_{2t}')' \equiv N(0, I_n) .$$

No corrections for serial correlation were made. The (size–adjusted) power functions for 5% size tests are displayed in table 1 for the Stock–Watson, Johansen and Phillips–Ouliaris tests. Rejection frequencies are given as a function of ρ, for systems of size 2 through 5 variables. As expected, all three tests show dramatic reductions in power as the size of the system increases. It is also interesting to note that the residual–based tests (Phillips–Ouliaris) have much better power than the Johansen and Stock–Watson tests.

This dependence renders the tests fairly ineffective in moderate sample sizes with moderately large systems. We could term this problem a "curse of dimensionality" for cointegration testing. The reason why the test distributions depend upon the dimensionality of the system is because they involve estimation of all n unit roots in the system. The Stock–Watson and Johansen tests do this explicitly by estimating an n

[2]All calculations in this paper were made in GAUSS386 on a 386/33.

equation VAR. The residual–based procedure does this implicitly, as the first stage regression is spurious and implicitly estimates the $n-1$ unit roots of the regressors, while the test itself estimates the remaining unit root.

Since it rarely is the case that *all* unit roots need to be tested, it seems sensible to design tests which only estimate the number of unit roots which are under scrutiny. To be specific, in the common two–variable system, if it is agreed that each variable is roughly described as I(1), then the question of interest is whether the variables are *cointegrated*, which means that the system has *one* unit root, as opposed to *two* unit roots. The above procedures estimate both roots, and hence possess distributions which reflect this fact. An alternative procedure which circumvents this curse of dimensionality is developed in the next section.

Table 1. Finite Sample Power. T = 100

A. Stock–Watson

ρ	n = 2	n = 3	n = 4	n = 5
1.00	.05	.05	.05	.05
0.95	.10	.06	.06	.05
0.90	.28	.13	.09	.07
0.85	.56	.27	.16	.10
0.80	.82	.50	.29	.18
0.75	.96	.73	.49	.30

B. Johansen

ρ	n = 2	n = 3	n = 4	n = 5
1.00	.05	.05	.05	.05
0.95	.08	.05	.05	.05
0.90	.20	.10	.07	.06
0.85	.41	.19	.12	.09
0.80	.66	.32	.20	.12
0.75	.86	.51	.30	.17

C. Phillips–Ouliaris

ρ	n = 2	n = 3	n = 4	n = 5
1.00	.05	.05	.05	.05
0.95	.13	.09	.08	.07
0.90	.36	.23	.17	.13
0.85	.68	.46	.33	.26
0.80	.91	.72	.56	.44
0.75	.98	.92	.77	.65

4. Cochrane–Orcutt

The residual–based test estimates the two–equation system

(6) $$x_{1t} = x_{2t}'\alpha + u_t$$

(7) $$u_t = \rho u_{t-1} + \epsilon_t$$

sequentially by least squares. Denote these estimators by $(\hat{\alpha}, \hat{\rho})$.

In a classic paper, Cochrane and Orcutt (1949) suggested estimation of (α, ρ) as following. Quasi–difference the data:

(8) $$x_{1t}^* = x_{1t} - x_{1t-1}\hat{\rho}$$

$$x_{2t}^* = x_{2t} - x_{2t-1}\hat{\rho}$$

and then apply OLS to the transformed equation

(9) $$x_{1t}^* = x_{2t}^{*\prime}\tilde{\alpha} + \tilde{\epsilon}_t .$$

It is well known that if $\{x_t\}$ is covariance stationary and $E(\epsilon_t | \mathcal{F}_{t-1}) = 0$, then these estimates are consistent and asymptotically normal. It is also known that if $x_t \equiv$ I(1) and x_{1t} is cointegrated with x_{2t}, then $\tilde{\alpha}$ converges stochastically to the cointegrating vector and $\tilde{\rho} \rightarrow_p \rho < 1$. See, for example, Phillips and Park (1988).

This estimator has not been studied in the context of integrated variables which are not cointegrated. It is not, however, particularly difficult. As shown by Phillips and Ouliaris (1990), under no cointegration, $T(\hat{\rho} - 1)$ has a limiting distribution (implying that $\hat{\rho} \rightarrow_p 1$ at rate T). Thus x_{1t}^* and x_{2t}^* asymptotically are first differences of x_{1t} and x_{2t}, and $\tilde{\alpha}$ estimates the regression coefficient of Δx_{1t} upon Δx_{2t}.

Theorem 1. *Under the null hypothesis of no cointegration*

(10) $$\tilde{\alpha} \rightarrow_p \alpha \equiv \left[E(\Delta x_{2t}\Delta x_{2t}')\right]^{-1}E(\Delta x_{2t}\Delta x_{1t}) .$$

Theorem 1 demonstrates that under the null hypothesis of no cointegration, the Cochrane–Orcutt estimate of α converges to a constant, not to a random variable as does the least squares estimator. This suggests that a test for cointegration constructed using residuals from the Cochrane–Orcutt estimates will not display the curse of dimensionality.

Using the definition of α given in (10), define the error u_t by (6), and $\epsilon_t = \Delta u_t$. The second–stage residuals are given by

$$\tilde{u}_t = x_{1t} - x_{2t}'\tilde{\alpha} \ . \tag{11}$$

Consider the regression of $\tilde{u}_t$ upon $\tilde{u}_{t-1}$:

$$\tilde{u}_t = \tilde{\rho}\, \tilde{u}_{t-1} + \tilde{\epsilon}_t \, , \tag{12}$$

so that

$$\tilde{\rho} = \frac{\Sigma_2^T \tilde{u}_{t-1}\tilde{u}_t}{\Sigma_2^T \tilde{u}_{t-1}^2} \ . \tag{13}$$

Theorem 2. *Under the null hypothesis of no cointegration*

(a) $$T^{-1/2}\tilde{u}_{[Tr]} \Rightarrow \sigma\, W(r) \, ,$$

(b) $$T(\tilde{\rho}-1) \Rightarrow \frac{\frac{1}{2}\left[W(1)^2 - \sigma_\epsilon^2/\sigma^2\right]}{\int_0^1 W^2} = \frac{\int_0^1 W dW + \lambda/\sigma^2}{\int_0^1 W^2} \, ,$$

where $\sigma^2 = \eta'\Omega\eta > 0$, $\eta = (1 \ -\alpha')'$, $\lambda = \eta'\Lambda\eta$, $\sigma_\epsilon^2 = E(\epsilon_0^2)$, *and* $W(r)$ *is standard Brownian motion.*

Corollary 1. *If in addition to the null hypothesis of theorem 2,* u_t *is a martingale difference sequence* ,

$$T(\tilde{\rho}-1) \Rightarrow \frac{\int_0^1 W dW}{\int_0^1 W^2} \ .$$

The distributions in theorem 2 and corollary 1 do not depend upon the dimension of the system, which is unique among existing tests of the null of no cointegration. The distributions are identical in form to the results given in Phillips (1987) for the univariate Dickey–Fuller unit root test. These results are quite promising, for they suggest that the curse of dimensionality is not an inherent property of tests for cointegration, and may be circumvented by appropriate techniques.

5. Adjusting for Size

Continuing to abstract from residual serial correlation, it is of interest to address how well the asymptotic theory approximates the finite sample behavior. Theorem 1 shows that the second–round Cochrane–Orcutt estimator asymptotically achieves the Dickey–Fuller distribution, but it seems reasonable that iteration will improve finite sample performance. Figures 1 and 2 display non–parametric estimates[3] of the finite sample $(T = 100)$ density functions of the statistic $T(\tilde{\rho} - 1)$ after various iterations. The innovations are iid normals with no serial correlation. The densities in Figure 1 are calculated for a system with two variables. The Cochrane–Orcutt densities are substantially shifted away from the OLS residual density at all iterations. The right–most density is calculated with fixed regression coefficients, and is thus the finite–sample Dickey–Fuller distribution. After multiple iterations, the Cochrane–Orcutt estimator comes quite close to the Dickey–Fuller distribution. In figure 2 the experiment is repeated for a system of five variables. In this case the iterated estimator does not come as close to

[3]A normal kernel was applied to 10,000 draws.

the Dickey–Fuller distribution, indicating that the curse of dimensionality has not been completely avoided. The fatter tails indicate that the null rejection frequency will be too high if asymptotic critical values are used.

The source of this distortion lies in the fact that the serial correlation coefficient used to quasi–difference the data is *estimated*, and thus tends to take on values less than unity. Since the data is not fully differenced in finite samples, an I(1) component persists, contaminating the finite sample distribution. This conjecture is verified by the following Monte Carlo experiment under the null: difference the data, estimate the OLS coefficient, then calculate the Dickey–Fuller test on the levels residuals. These results are not reported here, for they merely reveal that the size distortion disappears. Unfortunately, this procedure is not a valid test as it is not generally powerful, since the OLS estimate on differenced data will not converge to the cointegrating vector under the alternative, except in special cases.

This experiment suggests a possible remedy. Since the estimated serial correlation coefficient is downwardly biased under the null, and hence under–differences the data, we can use a bias adjustment to achieve an improved rate of convergence. Consider the following procedure. Denote by $\hat{\alpha}$ and $\hat{\rho}$ the OLS estimates from (6) – (7). Define

$$\rho^{+} = \hat{\rho} + c/T \tag{14}$$

where $c \geq 0$ is a fixed constant. Quasi–difference the data as in (8) using this bias–adjusted estimator ρ^{+}. Now estimate α using this quasi–difference data as in (9) ; denote this estimator by α^{+}. Iterate if desired. At the final stage, estimate ρ *without* the bias adjustment, in order to use the standard tables. (Alternatively, subtract c/T from ρ^{+}.) Since the adjustment term vanishes at rate T, the asymptotic theory of the previous section is unaltered.

The obvious question is how to select the value c. Standard Cochrane–Orcutt implicitly uses $c = 0$. Consider the adjustment term in the leading case of normally distributed, iid errors. This allows the use of Monte Carlo integration methods. Figures 3 and 4 display the rejection frequencies of the Dickey–Fuller test applied to the bias–adjusted Cochrane–Orcutt residuals (4 iterations) using the asymptotic 5% critical value for a variety of adjustment parameters along the horizontal axis. Figure 3 sets $T = 50$ and figure 4 sets $T = 150$. Figures for $T = 100$ and $T = 200$ were calculated but not reported here for the results were qualitatively similar. Note that the intersection points on the vertical axis correspond to the rejection frequency of unadjusted Cochrane–Orcutt ($c = 0$).

The graphs yield a fair amount of information. First, unadjusted Cochrane–Orcutt displays considerable size distortion for n large and T small. Second, bias adjustment can reduce the magnitude of this problem even in small samples. Third, setting $c = 10$ appears to minimize the rejection frequency for all sample and system sizes. This minimum sometimes lies below the nominal size, but not excessively so. Fourth, the choice of bias adjustment parameter is not critical for small n or large T.

The recommendation that emerges from this analysis is to set $c = 10$. That is, iterate on

$$\rho^{+} = \tilde{\rho} + 10/T \ .$$

Although admittedly *ad hoc*, this procedure appears to work quite well. To assess the practical impact of these procedures, table 2 reports (size adjusted) power of a 5% size test from a Monte Carlo experiment with 3000 replications, sample size of 100, and iid normal errors. The AR parameter of the dependent variable was varied from 0.85 to 1 in steps of 0.05. Three tests were applied: Phillips–Ouliaris (Dickey–Fuller applied to the OLS residuals), unadjusted Cochrane–Orcutt, and bias–adjusted ($c = 10$) Cochrane–Orcutt. The Cochrane–Orcutt tests are uniformly more powerful than the Phillips–Ouliaris (OLS) test. For example, in a four variable system, a 5% size OLS test rejects the null 19% of the time when $\rho = 0.9$, while the bias–adjusted Cochrane–Orcutt test rejects 59% of the time. It is not an overstatement to call this improvement dramatic. Note as well that the power of the two Cochrane–Orcutt procedures are close, with the adjusted procedure having moderately better local power.

Table 2. Finite Sample Power. T = 100.

A. N = 2

ρ	OLS	Coch–Orc	Coch–Orc (Biased Adjusted)
1.00	.05	.05	.05
0.95	.15	.26	.31
0.90	.38	.65	.70
0.85	.70	.91	.91

B. N = 4

ρ	OLS	Coch–Orc	Coch–Orc (Biased Adjusted)
1.00	.05	.05	.05
0.95	.09	.19	.27
0.90	.19	.49	.59
0.85	.36	.80	.82

6. Incorporating Serial Correlation

This section extends the previous analysis to allow for serial correlation in the residual $\{u_t\}$. Several popular techniques are considered: (a) Phillips' Z_α and Z_t; (b) Augmented Dickey–Fuller (ADF); (c) Stock quadratic variation; and (d) Park–Choi variables addition. In each of the theorems, W denotes standard Brownian motion.

(a) Phillips' Z_α and Z_t

Phillips (1987) suggested correcting for serial correlation in unit root testing by using a bias correction. This technique extends directly to the Cochrane–Orcutt cointegration test. If $\{\tilde{u}_t\}$ are the levels residuals from (possibly bias adjusted) Cochrane–Orcutt iteration, and $\tilde{\rho}$ and $\tilde{\epsilon}_t$ are given in (12) – (13), then define the covariance estimates

$$\tilde{\gamma}_m = T^{-1} \sum_{t=m+1}^{T} \tilde{\epsilon}_{t-m} \tilde{\epsilon}_t$$

the bias and variance estimates

$$\tilde{\lambda} = \sum_{m=1}^{M} w_m \tilde{\gamma}_m \qquad \tilde{\sigma}^2 = \tilde{\gamma}_0 + 2\tilde{\lambda} ,$$

and the bias–corrected coefficient estimate

$$\tilde{\rho}^* = \frac{\Sigma_2^T \tilde{u}_{t-1} \tilde{u}_t - T\tilde{\lambda}}{\Sigma_2^T \tilde{u}_{t-1}^2} .$$

Note that this bias correction is distinct from the bias adjustment of the previous section. The weights w_m are usually selected so that for each m, $\lim_{M\uparrow\infty} w_m = 1$, while the truncation parameter M is selected to grow to infinity slowly with sample size. Some proofs, such as Newey and West (1987), require $M = o(T^{1/4})$. Andrews obtains consistency if $M = o(T)$ under our conditions. One simple choice for the weights is the Bartlett window $w_m = 1 - |m|/(M+1)$.

Phillips (1987) suggested the following test statistics to test for a unit root:

$$Z_\alpha = T(\tilde{\rho}^* - 1)$$

$$Z_t = \frac{\tilde{\rho}^* - 1}{s}, \qquad s^2 = \left[\Sigma_2^T \tilde{u}_{t-1}^2\right]^{-1} \tilde{\sigma}^2,$$

We have the following theory.

Theorem 3. *Under the null of no cointegration:*

(i) $$Z_\alpha \Rightarrow \frac{\int_0^1 W dW}{\int_0^1 W^2},$$

(ii) $$Z_t \Rightarrow \frac{\int_0^1 W dW}{\left[\int_0^1 W^2\right]^{1/2}}.$$

The distributions in theorem 2 are the standard Dickey–Fuller coefficient and t distribution, which are tabulated in Fuller (1976).

(b) Augmented Dickey–Fuller (ADF)

Said and Dickey (1984) suggested using an autoregressive approximation to capture the serial correlation properties of the residuals in tests for unit roots. Engle and Granger (1987) suggested the ADF test on the OLS residuals as a cointegration test. Phillips and Ouliaris (1990) demonstrated that this test statistic possesses an asymptotic distribution which is free of nuisance parameters other than the dimensionality of the system. We can consider two combinations of this approach with the Cochrane–Orcutt procedure.

First, consider taking the Cochrane–Orcutt residuals $\{\tilde{u}_t\}$ from the procedure discussed in the previous sections. Then fit by OLS

$$\tilde{u}_t = \rho\, \tilde{u}_{t-1} + \theta_1 \Delta\tilde{u}_{t-1} + \theta_2 \Delta\tilde{u}_{t-2} + \cdots + \theta_p \Delta\tilde{u}_{t-p} + \epsilon_t .$$

The ADF statistic is the t–statistic for the hypothesis that $\rho = 1$. Denote this by ADF_t. It is usually assumed that $p \uparrow \infty$ at some controlled rate such as $p = o(T^{1/3})$. Under these conditions we find.

Theorem 4. *Under the null of no cointegration and the assumption that* $\{\Delta x_t\}$ *is generated by a finite–order ARMA process,*

$$\mathrm{ADF}_t \Rightarrow \frac{\int_0^1 W dW}{\left[\int_0^1 W^2\right]^{1/2}} .$$

This shows that the ADF statistics possesses the standard Dickey–Fuller distribution if the Cochrane–Orcutt residuals are used, in contrast to the case when the OLS residuals are used.

Alternatively, we could estimate the two equation system

$$x_{1t} = x_{2t}'\alpha + u_t$$

$$u_t = \rho u_{t-1} + \theta_1 \Delta u_{t-1} + \cdots + \theta_p \Delta u_{t-p} + \epsilon_t$$

by Cochrane–Orcutt using an AR(p+1) correction. (GLS or MLE could alternatively be used with no change in the asymptotic theory.) The t–statistic for the hypothesis that $\rho = 1$ will again possess an asymptotic Dickey–Fuller distribution. This procedure may be preferred in finite samples to the first ADF procedure for it takes the serial correlation into account at all iterations. Note that the test that $\rho = 1$ is equivalent to the test that $\Sigma_{i=1}^{p+1} \theta_i = 1$ in the equation

$$u_t = \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_{p+1} u_{t-p-1} + \epsilon_t$$

which is the parameterization appearing in many statistics packages. This procedure has the advantage that it is quite simple to implement.

(c) **Stock Quadratic variation**

Stock (1988) proposed a class of tests for unit roots, a leading candidate being

$$g_1(2) = T^{-2}\Sigma_1^T u_t^2 \;/\; \hat{\sigma}^2 \quad ,$$

where $\hat{\sigma}^2$ is the long–run variance of $\{\Delta u_t\}$, as described in the section on the Phillips' tests. Under the null hypothesis of a unit root, this statistic has the asymptotic distribution

$$g_1(2) \Rightarrow \int_0^1 W^2 \, . \tag{16}$$

Under the alternative, $g_1(2)$ vanishes, so the test is to reject for significantly small values. Stock also discusses tests for cointegration, by replacing u_t by the OLS residual from a candidate cointegrating regression. The limiting distribution of this test statistic depends upon the number of variables and thus displays the curse of dimensionality.

If, however, Cochrane–Orcutt residuals are used, then result (16) still holds.

Theorem 5. *Under the null hypothesis of no cointegration*

$$\tilde{g}_1(2) = T^{-2}\Sigma_1^T \tilde{u}_t^2 \;/\; \tilde{\sigma}^2 \Rightarrow \int_0^1 W^2 \, .$$

(d) **Park–Choi Variables addition**

Park, Ouliaris, and Choi (POC) (1988) proposed a test for cointegration based upon variable addition. If $\{\hat{u}_t\}$ is the OLS residual from a non–cointegrating regression POC suggest regressing $\hat{u}_t$ against "spurious" trends. Abstracting from their inclusion of time trends under the null, they effectively suggest the linear regression

$$\hat{u}_t = \hat{\alpha} + \hat{\beta}' k_t + \hat{\epsilon}_t \; , \qquad k_t = \begin{bmatrix} t \\ t^2 \end{bmatrix} \text{, say .} \tag{17}$$

The trend function k_t may contain a variety of powers of time trends. Using $\hat{R}^2$ to denote the standard coefficient of determination, the F–statistic for testing $\beta = 0$ is

$$\hat{F} = T\left[\frac{\hat{R}^2}{1-\hat{R}^2}\right].$$

Now, for $\delta_T = \begin{bmatrix} T & 0 \\ 0 & T^2 \end{bmatrix}$ and $\bar{k} = T^{-1}\Sigma_1^T k_t$, we have

$$\text{(18)} \qquad \delta_T^{-1}(k_{[Tr]} - \bar{k}) \Rightarrow \begin{bmatrix} r - 1/2 \\ r^2 - 1/3 \end{bmatrix} = k(r), \text{ say}.$$

Thus using (5) and (18)

$$\hat{R}^2 \Rightarrow \frac{\int B^* k'(\int kk')^{-1} \int kB^*}{\int (B^* - \int B^*)^2} = R^2(B^*, k) \text{ , say.}$$

Hence

$$\hat{J} = T^{-1}F \Rightarrow \frac{R^2(B^*, k)}{1 - R^2(B^*, k)}$$

Under the alternative hypothesis of no unit root, the F–statistic has a limiting chi–square distribution so the $\hat{J}$ statistic vanishes. Thus POC suggest rejecting the null hypothesis of no cointegration when J is sufficiently small. As in the other tests based upon OLS residuals, this distributional theory depends upon n (as B^* depends upon n).

Consider using Cochrane–Orcutt residuals in place of the OLS residuals in regression (17). Denote the coefficient of determination from this regression by $\tilde{R}^2$. This statistic has the same asymptotic behavior as that obtained by Park and Choi (1988) in the context of unit root testing:

Theorem 6. *Under the null hypothesis of no cointegration*

$$\text{(a)} \qquad \tilde{R}^2 \Rightarrow \frac{\int Wk'(\int kk')^{-1} \int kW}{\int (W - \int W)^2} \equiv R^2(W, k) \quad ,$$

$$\text{(b)} \qquad \tilde{J} = T^{-1}\tilde{F} \Rightarrow \frac{R^2(W, k)}{1 - R^2(W, k)} .$$

7. Allowing for Intercepts and Trends

If desired, the data may be demeaned, or demeaned and detrended, before applying the test for cointegration. Detrending is desirable if the data are thought of as "I(1) with drift". Demeaning is desirable if the cointegrating relationship may contain a constant. If the data is demeaned or detrended, then it is important to recognize that the distributions of the test statistics are different. It is not difficult to see that the appropriate distributions for the Z_α, Z_t and ADF_t statistics are given by the unit root test distributions in Fuller (1976) for models with intercepts and trends. Similarly, the Stock and Park–Choi statistics should be compared with the appropriate tables which take detrending into account.

An alternative method of detrending has been proposed in Schmidt (1990) and analyzed in Schmidt and Phillips (1989). This method could be applied to cointegration tests as well. Note that if

$$x_t = x_t^0 + \pi t$$

where $x_t^0 \equiv I(1)$, then the "detrended" variable

$$x_t' = x_t - \frac{t}{T} x_T$$

is free of the deterministic trend. This is equivalent to resumming the demeaned first differences of x_t. Schmidt and Phillips propose demeaning x_t' and using this variable in standard unit root tests. Since

$$T^{-1/2} x'_{[Tr]} \Rightarrow B(r) - rB(1) ,$$

a Brownian bridge, the limiting distributions are somewhat different then those tabulated in Fuller (1976) and therefore have been tabulated in Schmidt and Phillips (1989). If demeaned x_t' is used in a Cochrane–Orcutt cointegration test then these critical values could be used.

8. Testing for Multiple Cointegrating Relationships

Suppose $n > 2$ and the question is *how many* cointegrating vectors exist. Most of the papers written on testing for cointegration do not attempt to address this issue (other than Stock and Watson (1988)). I sketch here an outline of how the Cochrane–Orcutt approach can be extended to handle this testing situation.

Partition x_t as

$$x_t = (\underset{n_1}{x_{1t}'} \ \underset{n_2}{x_{2t}'})'$$

where n_1 contains the maximum number of cointegrating relationships to be discovered. Consider the parameterization

$$x_{1t} = Ax_{2t} + u_t \tag{18}$$

$$u_t = R\, u_{t-1} + \epsilon_t \tag{19}$$

$$H_o : R = I .$$

This parameterization reduces the number of unit roots which will be examined to n_1 , when (18) and (19) are *jointly* estimated. A vector generalization of Cochrane–Orcutt, for example, may be used. Under the null hypothesis, no cointegrating vectors exist, and thus $\tilde{R} \rightarrow_p I$. The Stock–Watson approach may now be applied to the matrix $\tilde{R}$ (adjusted for serial correlation as in their paper, yielding $\tilde{R}^*$), for the eigenvalues of the adjusted matrix $T(\tilde{R}^* - I)$ will converge to the limiting distributions given in their paper, where critical values are tabulated. One cointegrating relationship is found if the *smallest* root of $T(\tilde{R}^* - I)$ is significantly less than zero; two cointegrating relationships are found if the *second smallest* root of $T(\tilde{R}^* - I)$ is significantly less than zero ; etc.

A practical problem emerges in this approach in that no unique partition of x_t into x_{1t} and x_{2t} exists. No general selection method appears to exist, and choices may have to be made on a case–by–case basis.

References

Andrews, D. W. K. (1989). "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," Cowles Foundation Discussion Paper No. 877–R.

Campbell, J. Y. (1987). "Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis," *Econometrica*, 55, 1249–1273.

Campbell, J. Y. and R. J. Shiller (1987). "Cointegration and tests of present value models," *Journal of Political Economy*, 95, 1062–1088.

Cochrane, D. and G. H. Orcutt (1949). "Application of least squares regression to relationships containing autocorrelated error terms," *Journal of the American Statistical Association*, 44, 32–61.

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.

Engle, Robert F. and C.W.J. Granger (1987). "Co–Integration and error correction: representation, estimation and testing," *Econometrica*, 55, 251–276.

Engle, R. F. and B. S. Yoo (1987). "Forecasting and testing in co–integrated systems," *Journal of Econometrics*, 35, 143–159.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, New York: John Wiley & Sons.

Granger, C.W.J. (1981). "Some properties of time series data and their use in econometric model specification," *Journal of Econometrics*, 121–130.

Granger, C.W.J. and P. Newbold (1974). "Spurious regressions in econometrics," *Journal of Econometrics*, 2, 111–120.

Hansen, B. E. (1990). "Efficient estimation and testing of cointegrating vectors in the presence of deterministic trends," mimeo, University of Rochester.

Johansen, S. (1988). "Statistical analysis of cointegrating vectors," *Journal of Economic Dynamics and Control*, 12.

________________ (1988b). "Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models," mimeo.

Newey, W., and K. D. West (1987). "A simple positive semi–definite heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Park, J. Y. and B. Choi (1988). "A new approach to testing for a unit root," mimeo, Cornell University.

Park, J. Y., S. Ouliaris, and B. Choi (1988) "Spurious regressions and tests for cointegration," mimeo, Cornell University.

Phillips, P. C. B. (1986). "Understanding spurious regressions in econometrics," *Journal of Econometrics*, 33, 311–340.

________________ (1987). "Time series regression with a unit root," *Econometrica*, 55, 277–301.

Phillips, P.C.B. and S.N. Durlauf (1986). "Multiple time series with integrated variables," *Review of Economic Studies*, 53, 473–496.

Phillips, P.C.B. and B. E. Hansen (1990). "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, 57, 99–125.

Phillips, P.C.B. and S. Ouliaris (1988). "Testing for cointegration using principal components methods," *Journal of Economic Dynamics and Control*, 12, 205–230.

________________ (1990). "Asymptotic properties of residual based tests for cointegration," *Econometrica*, 58, 165–193.

Phillips, P.C.B. and J.Y. Park (1988). "Asymptotic equivalence of OLS and GLS in regressions with integrated regressors," *Journal of the American Statistical Association*, 83, 111–115.

Said, S. E. and D. A. Dickey (1984). "Testing for unit roots in autoregressive–moving average models of unknown order," *Biometrika*, 71, 599–607.

Schmidt, P. (1990). "Dickey–Fuller tests with drift," *Advances in Econometrics*, 8, 161–200.

Schmidt, P., and P.C.B. Phillips (1989). "Testing for a unit root in the presence of deterministic trends," Cowles Foundation Discussion Paper No. 933.

Stock, J. H. (1988). "A new class of tests for integration and cointegration," mimeo, Harvard University.

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

APPENDIX

Proof of Theorem 1. Note that

$$\tilde{\alpha} = \left[T^{-1}\Sigma_2^T x_{2t}^* x_{2t}^{*\prime}\right]^{-1} T^{-1}\Sigma_2^T x_{2t}^* x_{1t}^* .$$

Now

$$T^{-1}\Sigma_2^T x_{2t}^* x_{2t}^{*\prime} = T^{-1}\Sigma_2^T[\Delta x_{2t} - x_{2t-1}(\hat{\rho}-1)]\,[\Delta x_{2t}^{\prime} - x_{2t-1}^{\prime}(\hat{\rho}-1)]$$

$$= T^{-1}\Sigma_2^T \Delta x_{2t}\Delta x_{2t}^{\prime} - T^{-1}\Sigma_2^T x_{2t-1}\Delta x_{2t}^{\prime}(\hat{\rho}-1)$$

$$- T^{-1}\Sigma_2^T \Delta x_{2t} x_{2t-1}^{\prime}(\hat{\rho}-1) + T^{-2}\Sigma_2^T x_{2t-1}x_{2t-1}^{\prime}[T(\hat{\rho}-1)](\hat{\rho}-1)$$

$$= T^{-1}\Sigma_2^T \Delta x_{2t}\Delta x_{2t}^{\prime} + o_p(1) \;\rightarrow_p\; E\left[\Delta x_{2t}\Delta x_{2t}^{\prime}\right] .$$

Similarly,

$$T^{-1}\Sigma_2^T x_{2t}^* x_{1t}^* \rightarrow_p E\left[\Delta x_{2t}\Delta x_{1t}\right] ,$$

from which it follows that

$$\tilde{\alpha} \rightarrow_p \left[E(\Delta x_{2t}\Delta x_{2t}^{\prime})\right]^{-1} E(\Delta x_{2t}\Delta x_{1t}) . \qquad \square$$

Proof of Theorem 2.

(i)
$$T^{-1/2}\tilde{u}_{[Tr]} = T^{-1/2}x_{1[Tr]} - \tilde{\alpha}^{\prime}T^{-1/2}x_{2[Tr]}$$

$$\Rightarrow B_1(r) - \alpha^{\prime}B_2(r)$$

$$= \eta^{\prime}B(r) \equiv (\eta^{\prime}\Omega\eta)^{1/2}\,W(r) .$$

(ii)
$$T(\tilde{\rho}-1) = \frac{T^{-1}\Sigma_2^T \tilde{u}_{t-1}\Delta\tilde{u}_t}{T^{-2}\Sigma_2^T \tilde{u}_{t-1}^2} = \frac{\frac{1}{2T}\left[\tilde{u}_T^2 - u_1^2 - \Sigma_2^T \Delta\tilde{u}_t\right]}{T^{-2}\Sigma_2^T \tilde{u}_{t-1}^2}$$

$$\Rightarrow \frac{\frac{1}{2}\left[\sigma^2 W(1)^2 - \sigma_\epsilon^2\right]}{\sigma^2 \int_0^1 W^2} = \frac{\frac{1}{2}\left[W(1)^2 - \sigma_\epsilon^2/\sigma^2\right]}{\int_0^1 W^2}$$

where the convergence uses part (i), the continuous mapping theorem, and the fact that

$$\begin{aligned} T^{-1}\Sigma_2^T \Delta\tilde{u}_t^2 &= T^{-1}\Sigma_2^T[\Delta u_t - (\tilde{\alpha}-\alpha)'\Delta x_{2t}]^2 \\ &= T^{-1}\Sigma_2^T \epsilon_t^2 - 2(\tilde{\alpha}-\alpha)' T^{-1}\Sigma_2^T \epsilon_t \Delta x_{2t} \\ &\qquad + (\tilde{\alpha}-\alpha)' T^{-1}\Sigma_2^T \Delta x_{2t}\Delta x_{2t}'(\tilde{\alpha}-\alpha) \\ &\rightarrow_p \sigma_\epsilon^2 . \end{aligned}$$

The equality in part (ii) follows by Ito's lemma and the observation

$$\begin{aligned} \sigma^2 - \sigma_\epsilon^2 &= \eta'\Omega\eta - \eta' E(\Delta x_0 \Delta x_0')\eta \\ &= \eta'\left[\Omega - E(\Delta x_0 \Delta x_0')\right]\eta = \eta'\left[\Lambda + \Lambda'\right]\eta = 2\,\eta'\Lambda\eta . \quad \square \end{aligned}$$

Proof of Theorem 3.

$$\begin{aligned} \tilde{\epsilon}_t &= \tilde{u}_t - \tilde{\rho}\,\tilde{u}_{t-1} = \Delta\tilde{u}_t - (\tilde{\rho}-1)\,\tilde{u}_{t-1} \\ &= \tilde{\eta}'\Delta x_t - (\tilde{\rho}-1)\,\tilde{\eta}' x_{t-1} . \end{aligned}$$

Thus

(A1) $$\tilde{\lambda} = \Sigma_1^M w_m \tilde{\gamma}_m = \tilde{\eta}'\Sigma_1^M w_m T^{-1}\Sigma_t \Delta x_{t-m}\Delta x_t' \tilde{\eta} \rightarrow_p \eta'\Lambda\eta = \lambda ,$$

and

(A2) $$\tilde{\sigma}^2 \rightarrow_p \sigma^2 .$$

Hence

$$Z_\alpha = T(\tilde{\rho}^* - 1) = \frac{T^{-1}\Sigma_2^T \tilde{u}_{t-1}\Delta\tilde{u}_t - \tilde{\lambda}}{T^{-2}\Sigma_2^T \tilde{u}_{t-1}^2}$$

$$\Rightarrow \frac{\sigma^2 \int_0^1 W dW + \lambda - \lambda}{\sigma^2 \int_0^1 W^2} = \frac{\int_0^1 W dW}{\int_0^1 W^2} .$$

The result for Z_t follows similarly.

Proof of Theorem 4. The proof is identical to that of theorem 4.2 in Phillips and Ouliaris (1990), except that $\tilde{\eta} \rightarrow_p \eta$. This gives

$$ADF_t \Rightarrow \frac{\int_0^1 Q dQ}{\left[\int_0^1 Q^2\right]^{1/2}\left[\eta' \Omega \eta\right]^{1/2}} = \frac{\int_0^1 W dW}{\left[\int_0^1 W^2\right]^{1/2}}$$

where $Q = \eta' B = \sigma W$.

Proof of Theorem 5. Immediate from theorem 2(i) and (A2) .

Proof of Theorem 6.

$$\tilde{R}^2 = \frac{T^{-3/2}\Sigma_t \tilde{u}_t (k_t - \bar{k})' \delta_T^{-1} \left[T^{-1}\Sigma_t \delta_T^{-1}(k-\bar{k})(k-\bar{k})' \delta_T^{-1}\right]^{-1} T^{-3/2}\Sigma_t \delta_T^{-1}(k-\bar{k})\tilde{u}_t}{T^{-2}\Sigma_t [\tilde{u} - (T^{-1}\Sigma_t \tilde{u}_t)]^2}$$

$$\Rightarrow \frac{\sigma \int W k' \left(\int k k' \right)^{-1} \int k W \sigma}{\int (\sigma W - \int \sigma W)^2} = \frac{\int W k' (\int k k')^{-1} \int k W}{\int (W - \int W)^2} .$$

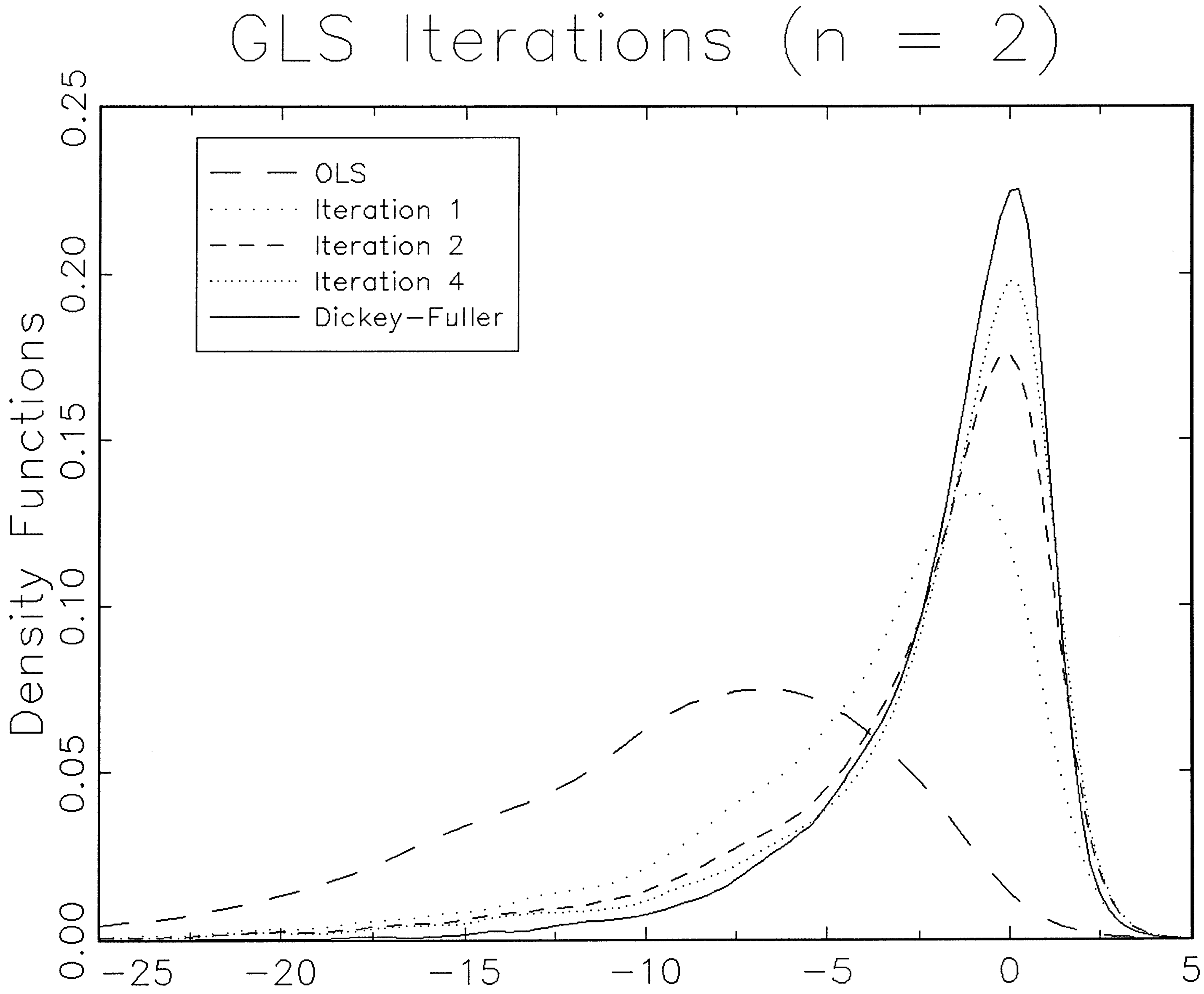
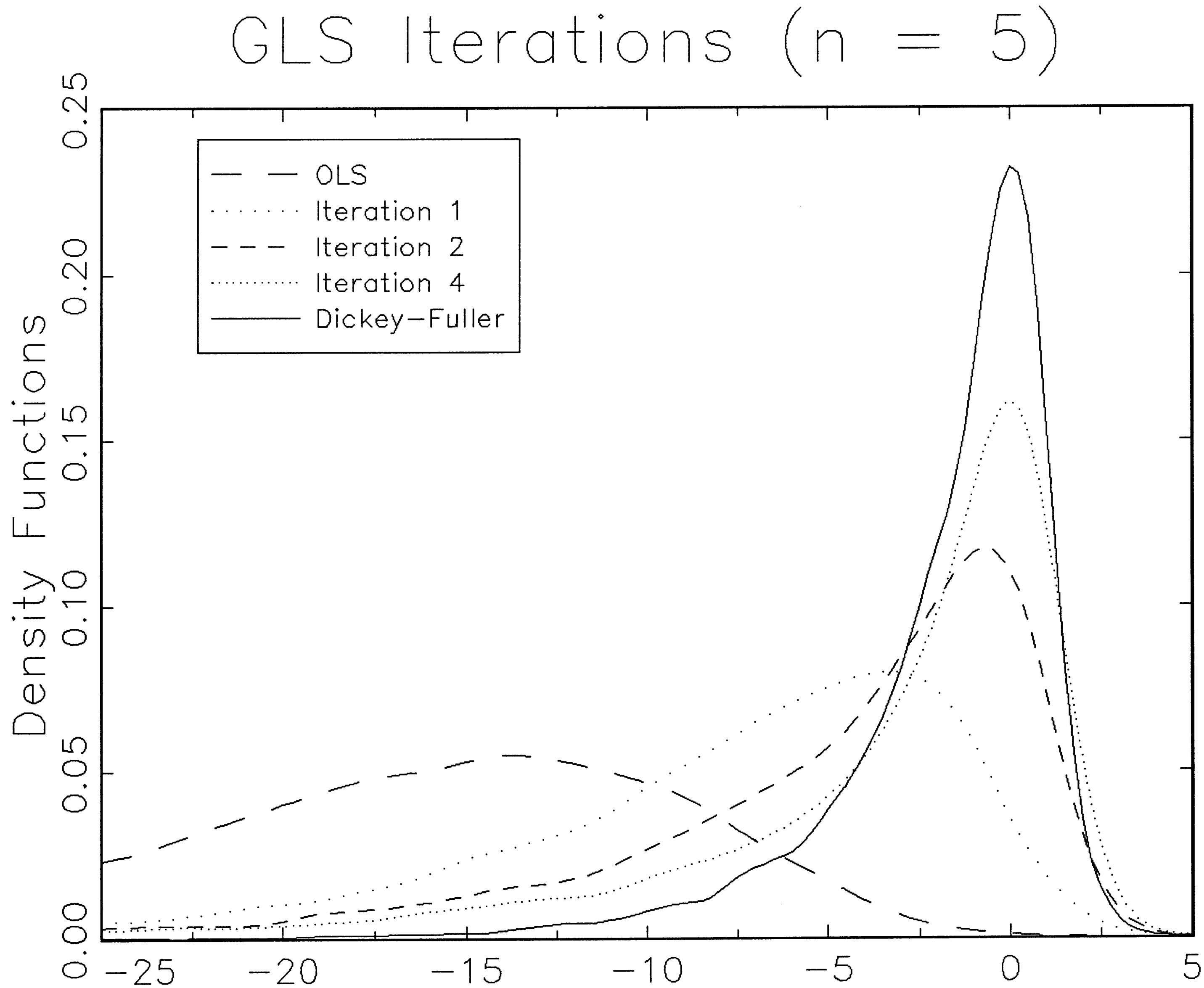
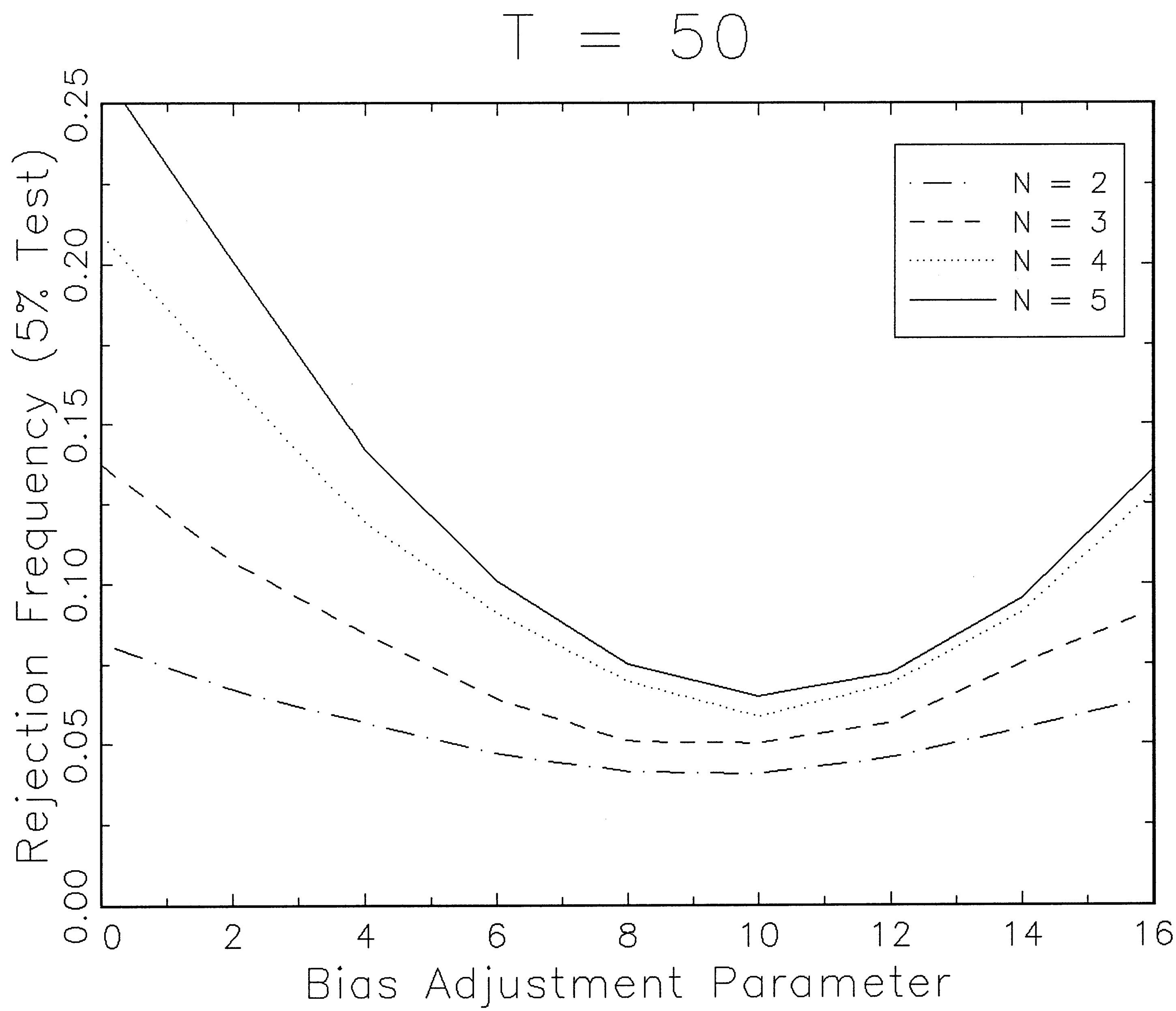

GLS Iterations (n = 2)
Density Functions
0.00
0.05
0.10
0.15
0.20
0.25
−25
−20
−15
−10
−5
0
5
OLS
Iteration 1
Iteration 2
Iteration 4
Dickey−Fuller

Figure 1



GLS Iterations (n = 5)
Density Functions
0.00
0.05
0.10
0.15
0.20
0.25
−25
−20
−15
−10
−5
0
5
OLS
Iteration 1
Iteration 2
Iteration 4
Dickey−Fuller

Figure 2



T = 50
Rejection Frequency (5% Test)
0.00
0.05
0.10
0.15
0.20
0.25
N = 2
N = 3
N = 4
N = 5
0
2
4
6
8
10
12
14
16
Bias Adjustment Parameter

Figure 3

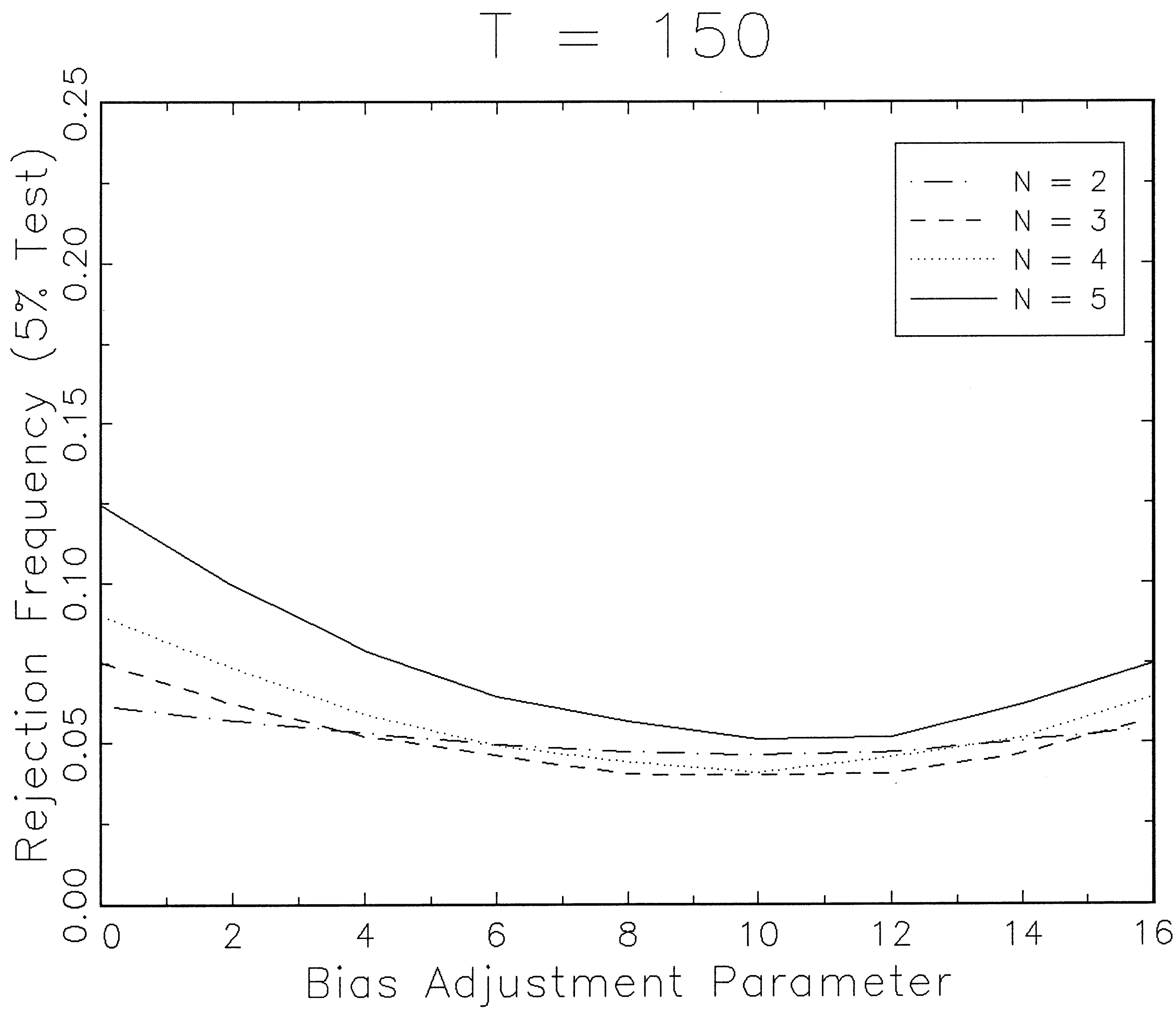

T = 150
Rejection Frequency (5% Test)
0.00
0.05
0.10
0.15
0.20
0.25
N = 2
N = 3
N = 4
N = 5
0
2
4
6
8
10
12
14
16
Bias Adjustment Parameter

Figure 4