THE POWER OF COINTEGRATION TESTS

Jeroen J. M. Kremers, Neil R. Ericsson and Juan J. Dolado[1]

I. INTRODUCTION

Contrasting inferences about the presence of cointegration often appear in empirical investigations. For example, in applying the commonly used 'two-step' procedure proposed by Engle and Granger (1987), the Dickey–Fuller unit-root test may only marginally reject the null hypothesis of no cointegration, if it rejects at all. By contrast, the coefficient on the error-correction term in the corresponding dynamic model of the same data may be 'highly statistically significant', strongly supporting cointegration; cf. Kremers (1989), Hendry and Ericsson (1991a), and Campos and Ericsson (1988). Both procedures are tests of cointegration, so why should there be such a contrast? A plausible explanation centers on an implicit common factor restriction imposed when using the Dickey–Fuller statistic to test for cointegration. If that restriction is invalid, the Dickey–Fuller test remains consistent, but loses power relative to cointegration tests that do not impose a common factor restriction, such as those based upon the estimated error-correction coefficient.

This paper examines the asymptotic and finite sample properties of the two procedures for a simple, single-lag, bivariate process. Even with more lags and more variables, the reason for the low power of the Dickey–Fuller test remains. The error-correction-based test is preferable because it uses available information more efficiently than the Dickey–Fuller test.

Section II describes the process of interest and derives the relationship between the error-correction mechanism and the equation from which the Dickey–Fuller statistic is calculated. Section III presents the asymptotic distribution of each test statistic under the null hypothesis of no cointegra-

[1] This paper represents the views of the authors and should not be interpreted as reflecting those of the Dutch Ministry of Finance, the Board of Governors of the Federal Reserve System, the Bank of Spain, or other members of their staff. This paper was prepared in part while the second and third authors were visiting INSEE and CEPREMAP, who we thank for generous hospitality. We are grateful to Javier Andrés, Anindya Banerjee, Julia Campos, Christian Gourieroux, David Hendry, Søren Johansen, Augustín Maravall, Alain Monfort, Mark Salmon, Jim Stock, and Hong-Anh Tran for helpful discussions, and to Lisa Barrow and Rafael Domenech for research assistance. All numerical results were obtained using PC-NAIVE and PC-GIVE Version 6.01; cf. Hendry and Neale (1990) and Hendry (1989).

tion, while Section IV gives the corresponding asymptotic distributions under the alternative hypothesis of cointegration, using fixed and 'near non-cointegrated' alternatives. Section V generalizes the results for testing in multivariate, multiple-lag systems. Section VI interprets some Monte Carlo finite sample evidence in light of the asymptotic formulae. Section VII empirically illustrates the two testing procedures with Hendry and Ericsson's (1991b) quarterly data on UK narrow money demand. Derivations of all new results appear in the Appendix.

II. A SIMPLE BIVARIATE PROCESS

Using a simple dynamic bivariate process, this paper focuses on the relative merits of the two-step Engle–Granger and single-step dynamic-model procedures for testing for the existence of cointegration. See Engle and Granger (1987) on the former and Banerjee, Dolado, Hendry and Smith (1986) *inter alia* on the latter. The former is characterized by a Dickey–Fuller (DF) statistic used to test for the existence of a unit root in the residuals of a static cointegrating regression. The latter is based upon the t-ratio of the coefficient on the error-correction term in a dynamic model reparameterized as an error-correction mechanism (ECM), noting that cointegration implies and is implied by an ECM. This t-ratio is denoted the ECM statistic. This section describes the data generation process (DGP) and derives the analytical relationship between the ECM and the equation for the DF statistic.

The bivariate process considered is one of the simplest imaginable, and has been used elsewhere for expository purposes; cf. Davidson, Hendry, Srba and Yeo (1978) and Banerjee, Dolado, Hendry and Smith (1986). It is a linear first-order vector autoregression with normal disturbances, at least one unit root, and Granger-causality in only one direction. For expositional convenience, this DGP is written as a conditional ECM (1) and a marginal unit-root process (2):

$$\begin{aligned} \Delta y_t &= a\Delta z_t + b(y-z)_{t-1} + \varepsilon_t \\ \Delta z_t &= u_t \end{aligned} \qquad \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim IN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & \sigma_u^2 \end{bmatrix} \right) \qquad t=1,\ldots,T, \qquad \begin{matrix} (1) \\ (2) \end{matrix}$$

where Δ is the first-difference operator $1-L$, L is the lag operator, and T is the sample size. The variables y_t and z_t are integrated of order one [denoted $I(1)$] and are possibly cointegrated. For $y=\ln Y$ and $z=\ln Z$, a is the short-run elasticity of Y with respect to Z. The parameter b is the error-correction coefficient in the conditional model of y_t, given lagged y and current and lagged z; and ε_t and u_t are the disturbances in this conditional/marginal factorization. Without loss of generality, the cointegrating vector for $(y_t\ z_t)'$ is $(1, -1)$ if y_t and z_t are cointegrated.

For simplicity, the (hypothesized) cointegrating vector is assumed known. Such *a priori* knowledge of the cointegrating vector arises frequently in economic models of long-run behavior, as in modeling (logs of) consumers' expenditure and disposable income, wages and prices, money and income, or the exchange rate and foreign and domestic price levels.[2] Also, z_t is assumed weakly exogenous for the parameters in the conditional model (1); see Engle, Hendry and Richard (1983) and Johansen (1992a).

As Section V shows, the logical issues arising from common factor restrictions apply to processes more general than (1)–(2). Specifically, the cointegrating vector or vectors may be estimated and may enter more than one equation (e.g., no weak exogeneity); and a constant term, seasonal dummies, additional variables, and additional lags may be included. However, some statistics' distributions are more complicated with such generalizations, so we focus on this bivariate case.

The parameter space is restricted to $\{0 \le a \le 1,\ -1 < b \le 0\}$. In many empirical studies, $a \approx 0.5$ and $b \approx -0.1$, with $\sigma_u^2 > \sigma_\varepsilon^2$. That is, the short-run elasticity (a) is smaller than the long-run elasticity (unity), adjustment to remaining disequilibria is slow, and the innovation error variance for the regressor process is larger than that of the conditional ECM.

The variables y_t and z_t are cointegrated or not, depending upon whether $b<0$ or $b=0$. Thus, tests of cointegration rely upon some estimate of b. In the ECM approach, equation (1) itself is estimated by OLS (denoted by a circumflex ˆ):

$$\Delta y_t = \hat{a}\Delta z_t + \hat{b}w_{t-1} + \hat{\varepsilon}_t, \tag{3}$$

where the putative disequilibrium is:

$$w_t = y_t - z_t. \tag{4}$$

The t-ratio based upon $\hat{b}$ is the ECM statistic, denoted t_{ECM}. It is used to test the null hypothesis that $b=0$, i.e., that y and z are *not* cointegrated with a cointegrating vector $(1, -1)$.

The DF statistic derives from a different regression, so it is helpful to establish the relationship between the DF regression equation and the ECM in (1). Specifically, subtract Δz_t from both sides of (1) and re-arrange:

$$\Delta(y-z)_t = b(y-z)_{t-1} + [(a-1)\Delta z_t + \varepsilon_t]. \tag{5}$$

Noting (4), equation (5) may be rewritten as:

$$\Delta w_t = bw_{t-1} + e_t, \tag{6}$$

where the disturbance e_t is:

$$e_t = (a-1)\Delta z_t + \varepsilon_t. \tag{7}$$

[2] See Davidson, Hendry, Srba and Yeo (1978), Hendry, Muellbauer and Murphy (1990), Sargan (1964), Nymoen (1992), Hendry and Ericsson (1991a, 1991b), and Johansen and Juselius (1990a, 1990b) *inter alia*.

OLS estimation of (6) (denoted by a tilde ˜) generates:

$$\Delta w_t = \tilde{b} w_{t-1} + \tilde{e}_t. \quad (8)$$

The t-ratio based upon $\tilde{b}$ is the DF statistic, denoted t_{DF} here [$\hat{\tau}$ in Dickey and Fuller, 1979]. This t-ratio is also used to test whether or not y_t and z_t are cointegrated with cointegrating vector $(1, -1)$. See Dickey and Fuller (1979, 1981) and Engle and Granger (1987).

In contrast to the estimated ECM in (3), the estimated DF equation (8) ignores potential information contained in Δz_t. Equivalently, (6) imposes the restriction that a equals unity. That is, the short-run elasticity (a) equals the long-run elasticity (unity). More generally, (6) imposes a common factor, as follows from rewriting (4) and (6):

$$y_t = z_t + w_t \qquad w_t = (1+b)\, w_{t-1} + e_t \quad (9)$$

or

$$[1-(1+b)\,L]\, y_t = [1-(1+b)\,L]\, z_t + e_t, \quad (10)$$

where $[1-(1+b)\,L]$ is the factor common to y_t and z_t in (10).[3]

The transformation of (1) to (6), (9), and (10) provides several insights. First, (1), (6), (9), and (10) are equivalent representations, given the relationship between the errors ε_t and e_t in (7); but the two errors are not equal unless $a=1$ or $\Delta z_t = 0$. Second, and relatedly, the common factor restriction in (10) [and so in (6) and (9)] is invalid unless $a=1$, noting that:

$$[1-(1+b)\,L]\, y_t = [a-(a+b)\,L]\, z_t + \varepsilon_t, \quad (11)$$

from (1). Interestingly, even if the common factor restriction is invalid, e_t remains white noise for this DGP. Nonetheless, e_t is *not* an innovation with respect to current and lagged z and lagged y; cf. Granger (1983) and Hendry and Richard (1982) on the distinction between white noise and innovations. Since empirically estimated short- and long-run elasticities often differ markedly (as noted above), imposing their equality in the DF statistic is rather arbitrary. Third, (9) motivates the use of unit-root statistics in testing for cointegration. If w_t has a unit root, then w_t is non-stationary, $b=0$, and y_t and z_t are not cointegrated with the cointegrating vector $(1, -1)$. Conversely, if w_t has its root inside the unit circle, then w_t is stationary, $b<0$, and y_t and z_t are cointegrated.

III. DISTRIBUTION OF THE STATISTICS UNDER THE NULL HYPOTHESIS (NO COINTEGRATION)

The null hypothesis is *no* cointegration: that is, $b=0$ in (1)–(2). Because w_{t-1} [in (3) and (8)] is not stationary under this hypothesis, distributional results from 'standard' asymptotic theory do not apply. This section describes the asymptotic distributions of the DF and ECM statistics under that null hypothesis, and obtains a *normal* approximation to the distribution of the ECM t-ratio when $a \neq 1$.

For expositional convenience, we adopt certain notational conventions concerning Brownian motion (or Wiener) processes. Consider a normal, independently and identically distributed variable η_t, $t=1,\ldots,T$: that is, $\eta_t \sim IN(0, \sigma_\eta^2)$. In this paper, η_t is usually either e_t, ε_t, or u_t. Define $B_{T,\eta}(r)$ as the partial sum $\Sigma_1^{[Tr]}\eta_t/\surd(T\sigma_\eta^2)$, where r lies in $[0,1]$, and $[Tr]$ is the integer part of Tr. As discussed in Phillips (1987b), $B_{T,\eta}(r)$ converges weakly to a standardized Wiener process, denoted $B_\eta(r)$. Frequently, the argument r is suppressed, as is the range of integration over r, when that range is $[0,1]$. Thus, integrals such as $\int_0^1 B_\eta(r)^2\, dr$ are written as $\int B_\eta^2$. The symbol '$\Rightarrow$' denotes weak convergence of the associated probability measures as the sample size $T \to \infty$. See Banerjee, Dolado, Galbraith and Hendry (1992) for a detailed discussion of Wiener processes.

The DF statistic [from (8)] is:

$$\begin{aligned} t_{DF} &= \tilde{b}/ese(\tilde{b}) \\ &= [(\Sigma w_{t-1}^2)^{-1}(\Sigma w_{t-1}\Delta w_t)]/\surd[\tilde{\sigma}_e^2 \cdot (\Sigma w_{t-1}^2)^{-1}] \\ &= (\Sigma w_{t-1}^2)^{-1/2}(\Sigma w_{t-1} e_t)/\tilde{\sigma}_e, \end{aligned} \quad (12)$$

where $ese(\cdot)$ is the estimated standard error of its argument, $\tilde{\sigma}_e^2$ is the estimated residual variance in (8), and all summations Σ are from 1 to T unless otherwise noted. Dickey and Fuller (1979) show that:

$$t_{DF} \Rightarrow (\int B_e\, dB_e)/\surd(\int B_e^2) \quad (13)$$

under the null hypothesis. Dickey [in Fuller, 1976, p. 373] tabulates by Monte Carlo the finite sample distribution for t_{DF}, from which critical values may be taken for constructing a unit-root test.

The DF statistic has several important properties. First, its distribution is skewed to the left, and it has a negative median. In part because of these characteristics, the use of (negative) one-sided normal critical values may result in over-rejection under the null hypothesis. Second, the distribution of the DF statistic is invariant to σ_u, σ_ε, and a, even in finite samples; cf. (12).

Banerjee, Dolado, Hendry and Smith (1986, Theorem 4) derive the asymptotic distribution of the t-ratio on $\hat{b}$ in the ECM (3). Our Appendix corrects their formula and obtains a simpler *normal* approximation for $a \neq 1$. Since $\Sigma\Delta z_t w_{t-1}$ is $O_p(T)$ and $E(u_t\varepsilon_t)=0$, the ECM t-ratio is:

$$\begin{aligned} t_{ECM} &= \hat{b}/ese(\hat{b}) \\ &= (\Sigma w_{t-1}^2)^{-1/2}(\Sigma w_{t-1}\varepsilon_t/\hat{\sigma}_\varepsilon) + O_p(T^{-1/2}), \end{aligned} \quad (14)$$

where $\hat{\sigma}_\varepsilon^2$ is the estimated residual variance in (3), and Mann and Wald's (1943) order notation is used. Ignoring the term of $O_p(T^{-1/2})$, (14) is identical

[3] See Hendry and Mizon (1978) and Sargan (1964, 1980) on common factors.

to the DF statistic in (12), except that ε_t appears rather than e_t. Using properties of independent Brownian motion, the limiting distribution of t_{ECM} is:

$$t_{\mathrm{ECM}} \Rightarrow (\int B_e \, \mathrm{d}B_\varepsilon)/\sqrt{(\int B_e^2)}$$

$$\Rightarrow \frac{(a-1)\int B_u \, \mathrm{d}B_\varepsilon + s^{-1}\int B_\varepsilon \, \mathrm{d}B_\varepsilon}{\sqrt{[(a-1)^2\int B_u^2 + 2(a-1)s^{-1}\int B_u B_\varepsilon + s^{-2}\int B_\varepsilon^2]}}, \tag{15}$$

where s is the ratio $\sigma_u/\sigma_\varepsilon$ (assumed strictly positive).

As will be discussed below, the distribution of t_{ECM} depends on the relative importance of the two terms comprising e_t in (7), which are $(a-1)\Delta z_t$ and ε_t. Specifically, it is useful to define a 'signal-to-noise' ratio:

$$q = -(a-1)\, s, \tag{16}$$

where q^2 is the variance of $(a-1)\Delta z_t$ relative to that of ε_t. Equally, q^2 is $\mathscr{R}^2/(1-\mathscr{R}^2)$, where $\mathscr{R}^2$ is the population R^2 with $b=0$ for Δw_t regressed on w_{t-1} and Δz_t, as in (28) below.

The asymptotic distribution of the ECM statistic has several unusual properties. First, because Δz_t is observed and is conditioned upon in estimating (3), q measures the amount of information present on the invalidity of the common factor restriction (for a given T). Second, and relatedly, when $a=1$ (and so $q=0$), (15) simplifies to the DF distribution (13), noting that $e_t = \varepsilon_t$ (and hence $B_e = B_\varepsilon$) for $a=1$. Third, for $a \neq 1$, (15) can be reparameterized in terms of q exclusively, rather than a and s separately:

$$t_{\mathrm{ECM}} \Rightarrow \frac{\int B_u \, \mathrm{d}B_\varepsilon - q^{-1}\int B_\varepsilon \, \mathrm{d}B_\varepsilon}{\sqrt{[\int B_u^2 - 2q^{-1}\int B_u B_\varepsilon + q^{-2}\int B_\varepsilon^2]}}. \tag{17}$$

The asymptotic distribution of t_{ECM} is sensitive to a and s only insofar as they enter q.

Fourth, for large q, (17) is approximately a standardized normal distribution:

$$t_{\mathrm{ECM}} \Rightarrow N(0,1) + O_p(q^{-1}). \tag{18}$$

This second approximation is 'small-σ' in nature or, equivalently, assumes the signal-to-noise ratio for (3) to be large; cf. Kadane (1970, 1971).[4] As q varies from small to large, the asymptotic distribution of t_{ECM} shifts from the DF distribution to the normal distribution. To obtain (18), note that (17) is:

$$t_{\mathrm{ECM}} \Rightarrow (\int B_u \, \mathrm{d}B_\varepsilon)/\sqrt{(\int B_u^2)} + O_p(q^{-1}). \tag{19}$$

[4] Complementary interpretations exist. From (1) and (2) with $b=0$ and $a \neq 0$, y_t and z_t are virtually identical series for large q (a constant term and factor of proportionality aside) because the variance of $a\Delta z_t$ is large relative to that of ε_t. Thus, y_t and z_t appear cointegrated, giving rise to 'standard' inferential procedures for b. This reasoning does not apply to the DF statistic because it is invariant to the variance of e_t.

Since B_u and B_ε are independent Brownian motions, the ratio in (19) is normally distributed; see Phillips and Park (1988).

Thus, when the common factor restriction in (9) is invalid and Δz_t contributes substantively to the determination of Δy_t, the t-ratio on the error-correction term in (3) is approximately normal, even when the error-correction coefficient is zero and so y_t and z_t are not cointegrated. That simplifies conducting inference with t_{ECM} when q is large.[5] The distribution of t_{DF} is independent of a, σ_u, and σ_ε (and thus of s and q), even in finite samples, so no parallel approximation exists for t_{DF}.

To summarize, in so far as distributions under the null are concerned, t_{ECM} has a distinct advantage over t_{DF} when q is known to be large because of the former's approximate normality under that condition. The next section considers distributions under the alternative hypothesis of cointegration, and so the issue of power.

IV. DISTRIBUTION OF THE STATISTICS UNDER THE ALTERNATIVE HYPOTHESIS (COINTEGRATION)

The alternative hypothesis is cointegration: namely, $b<0$ in (1)–(2). This section examines the asymptotic distributions of the DF and ECM statistics under both fixed and local alternatives. *A priori*, the distributions derived under either alternative could approximate the underlying finite sample distributions well, so both alternatives are of interest. Under a fixed alternative, w_{t-1} in (3) and (8) *is* stationary, so distributional results follow from conventional central limit theorems. Under a local alternative, the non-conventional asymptotic theory developed by Phillips (1988) for near-integrated series can be applied.

Section 4.1 compares the asymptotic distributions of the DF and ECM statistics under a fixed alternative; Section 4.2 compares them under a local alternative. When $a=1$, the two statistics are asymptotically equivalent. When $a \neq 1$, the ECM test can be arbitrarily more powerful than the DF test.

4.1. Distributions under a Fixed Alternative

Under a fixed alternative, this subsection analyzes the components of the DF and ECM statistics, from which the properties of the statistics themselves can be compared.

For the DF statistic, the numerator is:

$$\begin{aligned}\tilde{b} &= (\Sigma w_{t-1}^2)^{-1}(\Sigma w_{t-1}\Delta w_t) \\ &= b + (\Sigma w_{t-1}^2)^{-1}(\Sigma w_{t-1} e_t),\end{aligned} \tag{20}$$

[5] If no information is available on the magnitude of q, then it appears advisable to use the DF critical values for the ECM statistic because they are larger in absolute value than the critical values for the normal. This choice follows from the definition of statistical size involving the supremum over the appropriate parameter space, here, being over the range of a and s.

from which it follows that:

$$T^{1/2}\cdot(\tilde{b}-b)\Rightarrow N(0,\ \sigma_e^2/\sigma_w^2), \tag{21}$$

where $\sigma_w^2=\sigma_e^2/[1-(1+b)^2]$. The denominator of the DF statistic is:

$$ese(\tilde{b})=T^{-1/2}\sigma_e/\sigma_w+O_p(T^{-1}). \tag{22}$$

For the ECM statistic, the numerator is:

$$\hat{b}=b+(\Sigma w_{t-1}^2)^{-1}(\Sigma w_{t-1}\varepsilon_t)+O_p(T^{-1}), \tag{23}$$

which implies:

$$T^{1/2}\cdot(\hat{b}-b)\Rightarrow N(0,\ \sigma_\varepsilon^2/\sigma_w^2). \tag{24}$$

The denominator of the ECM statistic is:

$$ese(\hat{b})=T^{-1/2}\sigma_\varepsilon/\sigma_w+O_p(T^{-1}). \tag{25}$$

Combining these results obtains a relationship between the two statistics:

$$\begin{aligned}\frac{t_{\mathrm{ECM}}}{t_{\mathrm{DF}}}&=\frac{\hat{b}/ese(\hat{b})}{\tilde{b}/ese(\tilde{b})}\\ &=\sigma_e/\sigma_\varepsilon+O_p(T^{-1/2}).\end{aligned} \tag{26}$$

That is, the ECM statistic is approximately $\sigma_e/\sigma_\varepsilon$ times the DF statistic. That factor of proportionality is at least unity, and in general is greater than unity, noting that:

$$\begin{aligned}\sigma_e^2/\sigma_\varepsilon^2&=[(a-1)^2\sigma_u^2+\sigma_\varepsilon^2]/\sigma_\varepsilon^2\\ &=(1+q^2)\geq 1\end{aligned} \tag{27}$$

from (7). The degree of inequality depends upon q. Relative power is likewise affected, as illustrated in Section VI via Monte Carlo.

Intuition for the differences between the statistics is as follows. The ECM regression conditions on both Δz_t and w_{t-1}, whereas the DF regression conditions on only w_{t-1}, thereby losing potentially valuable information from Δz_t. Rewriting (5) helps clarify:

$$\Delta w_t=bw_{t-1}+(a-1)\Delta z_t+\varepsilon_t, \tag{28}$$

where, as an extreme example, $\varepsilon_t\approx 0$, $a\neq 1$, and Var(Δz_t) is 'substantial' (and so q is large). The ECM (28) has a near perfect fit, a and b are estimated with near exact precision, and the ECM t-ratio for b is (arbitrarily) large. However, the DF statistic is invariant to the variance of e_t (and so to the values of a and s), and the distribution of the DF statistic depends upon only b and T. For a suitably small (but nonzero) value of b and a given T, the DF statistic has little power (e.g., approximating its size) while the ECM statistic has power close to unity. This arises because the DF statistic ignores valuable information about Δz_t that is present in e_t. Nevertheless, both statistics are $O_p(T^{1/2})$ under a fixed alternative, so motivating a local alternative to obtain statistics of $O_p(1)$.

4.2. Distributions under a Local Alternative

To formalize the previous intuition, we apply Phillips's (1988) noncentral distribution theory to analyze the local asymptotic properties of the test statistics. The DGP is (1)–(2) with the local alternative:

$$b=e^{c/T}-1\approx c/T, \tag{29}$$

where c is a negative fixed scalar. The local alternative (29) parallels the usual Pitman-type local alternative, except that, in order to obtain statistics of $O_p(1)$, (29) differs from the null by $O_p(T^{-1})$, rather than by $O_p(T^{-1/2})$.

To proceed, we follow Phillips (1987b) and use the diffusion process:

$$\begin{aligned}K_\eta(r)&=\int_0^r e^{(r-j)c}\,\mathrm{d}B_\eta(j)\\ &=B_\eta(r)+c\int_0^r e^{(r-j)c}B_\eta(j)\,\mathrm{d}j,\end{aligned} \tag{30}$$

where $K_\eta(r)$ is an implicit function of c. If $c=0$, then $K_\eta(r)$ is $B_\eta(r)$. As with B_η, the argument r and the limits of integration are dropped if no ambiguity arises from doing so.

Under the local alternative (29), the DF statistic is distributed as:

$$t_{\mathrm{DF}}\Rightarrow c(\textstyle\int K_e^2)^{1/2}+(\int K_e\,\mathrm{d}B_e)/\surd(\int K_e^2); \tag{31}$$

see Phillips (1987b, p. 541; 1988, (26)). As shown in the Appendix, the ECM statistic is distributed as:

$$t_{\mathrm{ECM}}\Rightarrow c(1+q^2)^{1/2}(\textstyle\int K_e^2)^{1/2}+\dfrac{(a-1)\int K_u\,\mathrm{d}B_\varepsilon+s^{-1}\int K_\varepsilon\,\mathrm{d}B_\varepsilon}{\surd[(a-1)^2\int K_u^2+2(a-1)s^{-1}\int K_uK_\varepsilon+s^{-2}\int K_\varepsilon^2]}. \tag{32}$$

Properties of the asymptotic distributions in (31) and (32) are closely related to results under the null hypothesis. First, when $c=0$, (32) simplifies to the distribution under the null, (17). Likewise, the asymptotic distribution (31) for the DF statistic reduces to (13) under the null. Second, when $a=1$, (32) simplifies to the DF distribution (31). Third, for $a\neq 1$, (32) can be reparameterized in terms of c and q exclusively:

$$t_{\mathrm{ECM}}\Rightarrow c(1+q^2)^{1/2}(\textstyle\int K_e^2)^{1/2}+\dfrac{\int K_u\,\mathrm{d}B_\varepsilon-q^{-1}\int K_\varepsilon\,\mathrm{d}B_\varepsilon}{\surd[\int K_u^2-2q^{-1}\int K_uK_\varepsilon+q^{-2}\int K_\varepsilon^2]}. \tag{33}$$

Fourth, for large q, (33) is approximately a standardized normal distribution:

$$t_{ECM} \Rightarrow N(c(1+q^2)^{1/2}(\int K_u^2)^{1/2}, 1) + O_p(q^{-1}), \quad (34)$$

conditional on the process for u_t. Fifth, the unconditional mean of t_{ECM} can be approximated as:

$$E(t_{ECM}) \approx \gamma/\sqrt{2}, \quad (35)$$

where $\gamma = c(1+q^2)^{1/2}$.

The powers of the DF and ECM statistics can be summarized, as follows. For a given pair of values for c and T, the DF statistic has an associated asymptotic power, derivable from (31) and its critical value. For the same (c, T) pair and some comparable critical value, q can be arbitrarily large, in which case the ECM statistic is conditionally approximately normally distributed with unit variance. Further, its unconditional mean is negative and arbitrarily large, so its power can be arbitrarily close to unity. Thus, the ECM test has greater power than the DF test when q is sufficiently large, and the two tests have the same power when $q = 0$.

V. GENERALIZATIONS

The common factor 'problem' of the DF statistic remains when (1) includes additional variables, additional lags of variables, a constant term, seasonal dummies, and/or a more complicated cointegrating vector. Furthermore, augmented versions of the DF statistic [such as Dickey and Fuller's, 1981 ADF statistic] and non-parametric corrections [such as in Phillips, 1987a, and Phillips and Perron, 1988] do *not* resolve this problem. This section examines the common factor problem for a more general structure. It then shows how common factors can appear in systems procedures, as illustrated by Stock and Watson's (1988) test for common trends and avoided by Johansen's (1988) procedure.

Consider three generalizations of (1): lagged as well as current values of Δy_t and Δz_t may appear, z_t is a vector rather than a scalar, and the cointegrating vector is $(1, -\lambda')'$, being normalized on y but being otherwise unrestricted. Letting $d(L)$ and $a(L)$ be suitable scalar and vector polynomials in the lag operator L, (1) becomes:

$$d(L)\,\Delta y_t = a(L)'\Delta z_t + b(y - \lambda' z)_{t-1} + \varepsilon_t. \quad (36)$$

Subtracting $d(L)\lambda'\Delta z_t$ from both sides (rather than Δz_t as in Section II) obtains:

$$d(L)\,\Delta(y - \lambda' z)_t = b(y - \lambda' z)_{t-1} + \{[a(L)' - d(L)\,\lambda']\,\Delta z_t + \varepsilon_t\} \quad (37)$$

or

$$d(L)\,\Delta w_t = b w_{t-1} + e_t, \quad (38)$$

where

$$w_t = y_t - \lambda' z_t \quad (39)$$

and

$$e_t = [a(L)' - d(L)\,\lambda']\,\Delta z_t + \varepsilon_t. \quad (40)$$

Equations (38), (39), and (40) generalize (6), (4), and (7). When e_t is not white noise, (38) is not a regression equation, and below we comment on that case.

The ADF statistic is based upon (38), and so imposes the common factor restriction:

$$a(L) = d(L)\,\lambda. \quad (41)$$

If invalid, that restriction implies a loss of information (and so a loss of power) for the ADF test relative to the ECM test from (36). The caveat about common factors applies to other single-equation unit-root-type cointegration tests constructed from a static relationship between y_t and z_t, including Phillips's (1987a) Z_α and Z_t statistics, Phillips and Perron's (1988) generalizations thereon, and Sargan and Bhargava's (1983) statistic. The problem is not with the unit root tests *per se*: they may be quite useful for determining an individual series's order of integration. Rather, the difficulty arises from testing for *co*integration via testing for a unit root (or the lack thereof) in the purported disequilibrium measure $y_t - \lambda' z_t$.

The ADF tests applied to (38) may encounter an additional difficulty. Whereas e_t is white noise in the simple example (6), it need not be in (38); cf. (7) and (40). If not, then, in order to generate white noise errors, the ADF regression would need a lag length longer than that required in the ECM. Conversely, choosing too short a lag length for the ADF statistic can create misleading inferences; cf. Kremers (1988).

System analysis of cointegration faces similar problems. In a system notation following Johansen (1988), let x_t denote the entire vector of $I(1)$ variables under study, of dimension $p \times 1$. One interesting and commonly used representation for x_t is the Gaussian, finite-order vector autoregressive process:

$$\pi(L)\,x_t = \nu_t \qquad \nu_t \sim IN(0, \Omega_\nu) \quad (42)$$

or

$$\Delta x_t = \pi x_{t-1} + \Gamma(L)\,\Delta x_{t-1} + \nu_t, \quad (43)$$

where $\pi(L)$ is the lth order, $p \times p$ matrix polynomial $\Sigma_{i=0}^{l} \pi_i L^i$, $\Gamma(L)$ is a related $p \times p$ matrix polynomial, and $\pi = \pi(1)$. But for the normalization $\pi_0 = I_p$, $\pi(L)$ is unrestricted; so π and $\Gamma(L)$ are also unrestricted. Cointegration of variables in x_t implies that π is of reduced rank (r, say), so π can be factorized as:

$$\pi = \alpha\beta', \quad (44)$$

where α and β are full-rank $p \times r$ matrices. The rows of β' are cointegrating vectors, and the coefficients in α are the weights on the cointegrating vectors in each equation.

Some 'systems' procedures focus on the roots of $\beta' x_t$ rather than on the properties of x_t itself. Such procedures impose 'system common factors', as can be seen by pre-multiplying (43) by β':

$$\beta' \Delta x_t = (\beta' \alpha)\, \beta' x_{t-1} + \beta' \Gamma(L)\, \Delta x_{t-1} + \beta' v_t \qquad (45)$$

or

$$[I_r - G(L)\, L]\, \Delta w_t = (\beta' \alpha)\, w_{t-1} + \psi_t, \qquad (46)$$

where w_t is now the vector $\beta' x_t$, $G(L)$ is an $r \times r$ matrix polynomial in L, and ψ_t is:

$$\psi_t = [\beta' \Gamma(L) - G(L)\, \beta']\, \Delta x_{t-1} + \beta' v_t. \qquad (47)$$

Equations (46)–(47) parallel (38) and (40) for a single equation.

The disturbance ψ_t may contain valuable, predictable information for two reasons. First, unless the restriction $G(L)\, \beta' = \beta' \Gamma(L)$ holds, lags of Δx_t enter ψ_t. Second, if z_t is weakly exogenous, then $\beta' v_t$ may be explained in part by current z [as in (1)]. Both reasons imply a loss of information from analyzing w_t rather than x_t when testing for cointegration.

As an example, Stock and Watson's (1988) test for common trends imposes common factors, except when the maintained hypothesis is p common trends (i.e., no cointegration). Stock and Watson's statistic is derived from a vector autoregression in the hypothesized common trends $\beta'_{\perp} x_t$ [their equation (3.1)], which is an autoregression 'complementing' (46). Unless $\beta_{\perp}$ is square, their autoregression omits lags in $\beta' x_t$, and so ignores potentially valuable information.

Johansen (1988, 1991) and Johansen and Juselius (1990a) derive a likelihood-based method for testing the rank of π and, conditional upon a given rank, conducting inference about α and β. Because (43) is the basis for inference, this method avoids common factor problems. All short-run dynamics in $\Gamma(L)$ are unrestricted, and so are 'structural' rather than 'error' dynamics: the Johansen procedure parallels the ECM procedure, but with the system complete. Conversely, the ECM procedure is a special case of Johansen's for a system in which the cointegrating vectors appear in only the equation of interest. Under that condition, it is valid to analyze only the equation of interest, as a conditional equation; cf. Dolado, Ericsson and Kremers (1989) and Johansen (1992a).

VI. FINITE SAMPLE EVIDENCE

To analyze the size and power of the DF and ECM tests, a set of Monte Carlo experiments were conducted with (1) and (2) as the DGP. Without loss of generality, $\sigma_{\varepsilon}^2 = 1$. That leaves the parameters (s, a, b) and the sample size T as experimental design variables, noting that s now is σ_u. This Monte Carlo study is solely meant to illustrate the common factor issue, so we chose a full factorial design of:

$$(a, s) = [(1.0, 1), (0.5, 6), (0.5, 16)]$$

$$b = (0.0\ [\text{no cointegration}],\ -0.05\ [\text{cointegration}])$$

$$T = 20, \qquad (48)$$

resulting in six experiments. The number of replications per experiment was $N = 10{,}000$, the first 20 observations of each replication were discarded in order to attenuate the effect of initial values, and new z's were generated for each replication.

The parameter values were chosen with the following in mind. For $a = 1.0$ (and so $q = 0$), only $s = 1$ is considered, since the analytical results in Sections III and IV imply exact or asymptotic invariance of the statistics to s when the common factor restriction is valid. For $a = 0.5$, the values $s = 6$ and $s = 16$ imply $q = 3$ and $q = 8$ respectively, with the latter very 'strongly' violating the common factor restriction. The two values of b, 0.0 and -0.05, imply lack of and existence of cointegration respectively, although, in the latter case, the stationary root of the system is still large: 0.95. Finally, the sample size is small by most econometric standards, and implies a low power of the DF statistic for the nonzero value of b.

Table 1 lists rejection frequencies of the DF and ECM statistics under the hypotheses of no cointegration and cointegration. These rejection frequencies correspond to size and power, provided the correct critical values are used. Panels A and B of the table report rejection frequencies for one-sided tests at two nominal sizes, 5 percent and 1 percent. For each, three critical values are examined: those from Dickey in Fuller (1976, Table 8.5.2, p. 373) for $T = 25$, those of the normal distribution, and (for power) those estimated from our Monte Carlo with $b = 0$. The values of b and q appear at the top of the table: they define the experiments, and q in particular is important for the ECM statistic.

In Panel A (5 percent critical values) under 'no cointegration', rejection frequencies for t_{DF} are virtually unchanged as q varies, in line with the invariance result. With the Dickey–Fuller critical value, the rejection frequency for t_{ECM} matches that of t_{DF} for $q = 0$, and shrinks to well below the nominal rejection frequency for large q (e.g., 3.5 percent for $q = 8$). With the Gaussian critical value, the rejection frequency for t_{ECM} is 9.5 percent for $q = 0$, approximately double the nominal value, and tends toward the nominal value for large q. Such over-rejection limits the use of Gaussian critical values in practice.

In Panel A under 'cointegration', the power of the DF statistic is approximately 10 percent, whether with Dickey–Fuller or estimated critical values. As expected, its power is insensitive to q and to the choice of critical value.

TABLE 1
Rejection Frequencies and Estimated Means of the Statistics

Critical value and statistic	No cointegration: b = 0.0			Cointegration: b = −0.05		
	q			q		
	0	3	8	0	3	8
A. Rejection frequency at the 5 percent critical value (in percent)						
Dickey–Fuller (−1.95)						
DF	5.4	5.6	5.4	9.6	10.3	10.1
ECM	5.4	4.1	3.5	9.9	50.2	91.6
Gaussian (−1.645)						
DF	9.4	9.5	9.7	17.3	18.1	17.4
ECM	9.5	7.2	6.4	17.3	60.6	94.3
Estimated[1]						
DF	[−2.01]	[−2.03]	[−2.02]	8.2	8.9	8.8
ECM	[−2.02]	[−1.88]	[−1.80]	8.6	52.4	92.9
B. Rejection frequency at the 1 percent critical value (in percent)						
Dickey–Fuller (−2.66)						
DF	1.1	1.3	1.2	2.1	2.1	2.3
ECM	1.3	1.2	0.9	2.3	30.2	82.8
Gaussian (−2.326)						
DF	2.5	2.7	2.4	4.5	4.7	4.6
ECM	2.6	2.1	1.7	4.5	39.2	87.3
Estimated[1]						
DF	[−2.76]	[−2.80]	[−2.77]	1.6	1.6	1.7
ECM	[−2.80]	[−2.76]	[−2.62]	1.7	27.9	83.4
C. Estimated means of the statistics[2]						
mean(t_{DF})	−0.34	−0.38	−0.37	−0.95	−0.96	−0.95
mean(t_{ECM})	−0.34	−0.13	−0.04	−0.93	−2.09	−5.08
$\gamma/\sqrt{2}$	0.0	0.0	0.0	−0.71	−2.24	−5.70

Notes:
[1] Under the null of no cointegration, Monte Carlo estimates of the critical values are reported, in square brackets. Under the alternative, rejection frequencies are reported. The estimated critical values used for the DF statistic are the averages of those obtained under the null: −2.02 for 5 percent and −2.78 for 1 percent. The estimated critical values used for the ECM statistic are those obtained under the null, and they vary with q.
[2] Monte Carlo standard errors on the estimated means are approximately 0.01.

The power of the ECM statistic for $q=0$ is virtually identical to that of the DF statistic. However, as q increases, so does the power of the ECM statistic. At $q=8$, its power is over 90 percent. The common factor restriction is disastrous for the Dickey–Fuller procedure in such instances. Conversely, the ECM procedure can gain markedly in power because it allows more flexible dynamics than the DF procedure. Panel B reports similar results at the 1 percent critical value.

Panel C lists the estimated means of t_{DF} and t_{ECM} across experiments, and the approximate asymptotic mean of t_{ECM}, which is $\gamma/\sqrt{2}$. The estimated mean of the DF statistic appears invariant to q, as implied by Sections III and IV. Its estimated mean is more negative with cointegration than without cointegration, reflecting *inter alia* the negative noncentrality $c(\int K_e^2)^{1/2}$ in (31). The estimated mean of t_{ECM} is *not* invariant to q. Under the null of no cointegration, it tends to zero as q increases. With cointegration, the estimated mean of t_{ECM} is approximately $\gamma/\sqrt{2}$, and becomes large and negative as q increases. In these experiments, $q=3$ and $q=8$ appear quite 'large' for the mean of t_{ECM}, but not for tail properties. That suggests using the Dickey–Fuller or related critical values for t_{ECM} rather than Gaussian critical values, in order to control size.

VII. EMPIRICAL EVIDENCE

This section tests for cointegration in Hendry and Ericsson's (1991b) quarterly data on UK money demand to show how the DF and ECM statistics can differ empirically. The data are nominal M_1 (M), 1985 price total final expenditure (Y), the corresponding deflator (P), the three-month local authority interest rate ($R3$), and the (learning-adjusted) retail sight deposit interest rate (Rra). Below, lower case denotes logarithms. Hendry and Ericsson (1991b) describe the data in their appendix. Johansen (1992b) finds that m and p appear $I(2)$, and are cointegrated as $m-p$, which is $I(1)$. Thus, to avoid possible inferential complexities with $I(2)$ variables, we consider whether or not $m-p$, y, Δp, $R3$, and Rra are cointegrated.

The static regression of these variables obtains:

$$\widetilde{(m-p)}_t = -0.07\, y_t + 0.94\, \Delta p_t - 2.1\, R3_t + 6.9\, Rra_t + 11.8 \qquad (49)$$

$$T = 100[1964(3)\text{–}1989(2)] \quad \tilde{\sigma} = 9.646\% \quad dw = 0.18.$$

While direct statistical inference on the estimated coefficients in (49) is difficult, note that the income elasticity is negative, not positive; and the inflation elasticity is positive, not negative. Neither property is 'economically sensible'. Additionally, the two interest rate semi-elasticities are numerically quite different in absolute magnitude, so an interest rate differential does not seem plausible as a measure of the opportunity cost.

The augmented Dickey–Fuller regression ADF(4) for the residuals w_t from (49) is:

$$\Delta\widetilde{w}_t = \underset{(0.053)}{-0.182}\, w_{t-1} + \sum_{i=1}^{4} \tilde{\phi}_i \Delta w_{t-i} \tag{50}$$

$$T=95\,[1965(4)\text{–}1989(2)] \quad \tilde{\sigma}=3.690\% \quad t_{ADF}=-3.41.$$

Here and in equations below, ϕ_i denotes a generic coefficient, and standard errors are in parentheses. MacKinnon's (1991) 10 percent critical value for the DF statistic is -4.25 for $T=95$, so the variables do not appear cointegrated by this measure. Even so, the coefficient on w_{t-1} is negative and large numerically, implying a root of approximately 0.8.

In the error-correction framework, the long-run relationship between the variables may be obtained by estimating an autoregressive distributed lag in the variables and solving numerically for that long-run solution. Estimating the fifth-order autoregressive distributed lag for $m-p$, y, Δp, $R3$, and Rra obtains this long-run solution:

$$(m-p)_t = \underset{(0.27)}{1.10}\, y_t - \underset{(1.8)}{7.4}\, \Delta p_t - \underset{(1.2)}{7.3}\, R3_t + \underset{(0.7)}{7.2}\, Rra_t - \underset{(2.9)}{0.8} \tag{51}$$

$$T=100\,[1964(3)\text{–}1989(2)].$$

The long-run income elasticity is near unity, and inflation has a strong negative long-run effect. Further, the interest-rate coefficients are nearly equal in magnitude, opposite in sign, so in the long run, interest rates appear to matter only through the net interest rate ($R3-Rra$, denoted R^*).

Re-estimating the autoregressive distributed lag as an error-correction model obtains:

$$\Delta\widehat{(m-p)}_t = \underset{(0.023)}{-0.149}\, w_{t-1} + \sum_{i=1}^{4} \hat{\phi}_i \Delta(m-p)_{t-i}$$

$$+ \sum_{i=0}^{4} \hat{\phi}_i^{*\prime}(\Delta y_{t-i}, \Delta^2 p_{t-i}, \Delta R3_{t-i}, \Delta Rra_{t-i}) \tag{52}$$

$$T=100\,[1964(3)\text{–}1989(2)] \quad \hat{\sigma}=1.320\% \quad t_{ECM}=-6.39,$$

where the lagged residual from (51) is now w_{t-1}, the error-correction term. Even in this highly over-parameterized model, the ECM statistic exceeds MacKinnon's (1991) DF 1 percent critical value of -5.18. The equation standard error in (52) is far smaller than that in (50), implying that the common factor restriction in (50) is invalid [COMFAC $\chi^2(20)=64.6$].

The contrast between the DF and ECM statistics is robust to the choice of lag length and to whether or not long-run price homogeneity is imposed. Further, results from system analysis match the ECM results above. For a corresponding vector autoregression, Ericsson, Campos and Tran (1991) test and strongly reject the null of no cointegration in favor of one cointegrating vector, using Johansen's (1988, 1991) procedure. The system estimate of the first cointegrating vector is $(1, -0.77, 5.67, 5.82, -7.72)$, close to that in (51), noting that signs on unnormalized coefficients reverse. The first column in the estimated weighting matrix $\hat{\alpha}$ is $(-0.22, 0.00, 0.04, 0.07, 0.01)'$, consistent with weak exogeneity of Δp, y, $R3$, and Rra in the money equation for the cointegrating vector. That exogeneity permits valid conditional inference in the money equation, such as with the autoregressive distributed lag above.

The ECM statistic in (52) contains an estimated cointegrating vector, so the appropriateness of MacKinnon's tables for this t_{ECM} is as yet a conjecture, albeit a natural one. As an alternative, consider Hendry and Ericsson's (1991b) equation (6) — a constant, parsimonious simplification of an autoregressive distributed lag in the money demand variables:

$$\Delta\widehat{(m-p)}_t = \underset{(0.13)}{-0.69}\, \Delta p_t - \underset{(0.06)}{0.17}\, \Delta(m-p-y)_{t-1} - \underset{(0.060)}{0.630}\, R_t^* \tag{53}$$

$$- \underset{(0.009)}{0.093}\,(m-p-y)_{t-1} + \underset{(0.004)}{0.023}$$

$$T=100\,[1964(3)\text{–}1989(2)] \quad \hat{\sigma}=1.313\% \quad t_{ECM}=-10.87.$$

This equation *imposes* the long-run coefficients on prices and income, thus mirroring the analysis in Sections II–IV. While the error correction coefficient is somewhat smaller than before, the ECM statistic is even more highly significant than in (52). Prices and income have short-run elasticities of 0.31 and zero respectively, which contrast with their unit long-run elasticities and imply substantial violation of the common factor restriction in (50). Hendry and Ericsson (1991b, Section 4) further discuss the economic and statistical merits of (53).

VIII. SUMMARY

Over the last several years, testing for cointegration has become an important facet of the empirical analysis of economic time series, and various tests have been proposed and widely applied. This paper illustrates how a statistic based upon the estimation of an ECM can be approximately normally distributed when no cointegration is present, even though the equivalent DF statistic has a non-normal asymptotic distribution. With cointegration, the ECM statistic can generate more powerful tests than those based upon the DF statistic applied to the residuals of a static cointegrating relationship. These differences arise because the DF statistic ignores potentially valuable information — specifically, it imposes a possibly invalid common factor restriction. Phrased somewhat differently, a loss of information can occur from assuming error dynamics rather than structural dynamics. Both empirical and Monte Carlo finite sample evidence support these analytical results.

Ministry of Finance, The Hague, and Erasmus University, Rotterdam The Netherlands
International Finance Division, Federal Reserve Board, Washington DC
Research Department, Bank of Spain, Madrid

REFERENCES

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D. F. (1992). *Co-integration, Error-Correction and the Econometric Analysis of Non-stationary Data*, Oxford University Press, forthcoming.

Banerjee, A., Dolado, J. J., Hendry, D. F. and Smith, G. W. (1986). 'Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence', *BULLETIN*, Vol. 48, pp. 253–77.

Billingsley, P. (1968). *Convergence of Probability Measures*, John Wiley and Sons, New York.

Campos, J. and Ericsson, N. R. (1988). 'Econometric Modeling of Consumers' Expenditure in Venezuela', International Finance Discussion Paper No. 325, Board of Governors of the Federal Reserve System, Washington, DC.

Davidson, J. E. H., Hendry, D. F., Srba, F. and Yeo, S. (1978). 'Econometric Modelling of the Aggregate Time-series Relationship between Consumers' Expenditure and Income in the United Kingdom', *Economic Journal*, Vol. 88, pp. 661–92.

Dickey, D. A. and Fuller, W. A. (1979). 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root', *Journal of the American Statistical Association*, Vol. 74, pp. 427–31.

Dickey, D. A. and Fuller, W. A. (1981). 'Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root', *Econometrica*, Vol. 49, pp. 1057–72.

Dolado, J. J., Ericsson, N. R. and Kremers, J. J. M. (1989). 'Inference in Conditional Dynamic Models with Integrated Variables', paper presented at the European Meeting of the Econometric Society, Munich, Germany, September.

Engle, R. F. and Granger, C. W. J. (1987). 'Co-integration and Error Correction: Representation, Estimation, and Testing', *Econometrica*, Vol. 55, pp. 251–76.

Engle, R. F., Hendry, D. F. and Richard, J.-F. (1983). 'Exogeneity', *Econometrica*, Vol. 51, pp. 277–304.

Ericsson, N. R., Campos, J. and Tran, H.-A. (1991). 'PC-GIVE and David Hendry's Econometric Methodology', International Finance Discussion Paper No. 406, Board of Governors of the Federal Reserve System, Washington, DC; *Revista de Econometria*, forthcoming.

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

Granger, C. W. J. (1983). 'Forecasting White Noise' in Zellner, A. (ed.), *Applied Time Series Analysis of Economic Data*, Bureau of the Census, Washington, DC.

Hendry, D. F. (1989). *PC-GIVE: An Interactive Econometric Modelling System*, Version 6.0/6.01, Oxford, Institute of Economics and Statistics and Nuffield College, University of Oxford.

Hendry, D. F. and Ericsson, N. R. (1991a). 'An Econometric Analysis of U.K. Money Demand in *Monetary Trends in the United States and the United Kingdom* by Milton Friedman and Anna J. Schwartz', *American Economic Review*, Vol. 81, pp. 8–38.

Hendry, D. F. and Ericsson, N. R. (1991b). 'Modeling the Demand for Narrow Money in the United Kingdom and the United States', *European Economic Review*, Vol. 35, pp. 833–81.

Hendry, D. F. and Mizon, G. E. (1978). 'Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England', *Economic Journal*, Vol. 88, pp. 549–63.

Hendry, D. F., Muellbauer, J. N. J. and Murphy, A. (1990). 'The Econometrics of DHSY', Chapter 13 in Hey, J. D. and Winch, D. (eds.), *A Century of Economics: 100 Years of the Royal Economic Society and the Economic Journal*, Basil Blackwell, Oxford.

Hendry, D. F. and Neale, A. J. (1990). *PC-NAIVE: An Interactive Program for Monte Carlo Experimentation in Econometrics*, Version 6.01, Oxford, Institute of Economics and Statistics and Nuffield College, University of Oxford.

Hendry, D. F. and Richard, J.-F. (1982). 'On the Formulation of Empirical Models in Dynamic Econometrics', *Journal of Econometrics*, Vol. 20, pp. 3–33.

Johansen, S. (1988). 'Statistical Analysis of Cointegration Vectors', *Journal of Economic Dynamics and Control*, Vol. 12, pp. 231–54.

Johansen, S. (1989). 'The Power Function of the Likelihood Ratio Test for Cointegration', Preprint No. 1989:8, Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, Denmark.

Johansen, S. (1991). 'Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models', *Econometrica*, Vol. 59, No. 6, pp. 1551–80.

Johansen, S. (1992a). 'Cointegration in Partial Systems and the Efficiency of Single-equation Analysis', *Journal of Econometrics*, Vol. 52, pp. 389–402.

Johansen, S. (1992b). 'Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data', *Journal of Policy Modeling*, Vol. 14, pp. 313–34.

Johansen, S. and Juselius, K. (1990a). 'Maximum Likelihood Estimation and Inference on Cointegration — With Applications to the Demand for Money', *BULLETIN*, Vol. 52, pp. 169–210.

Johansen, S. and Juselius, K. (1990b). 'Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for UK', Preprint No. 1990:1, Institute of Mathematical Statistics, University of Copenhagen, Copenhagen, Denmark; *Journal of Econometrics*, forthcoming.

Kadane, J. B. (1970). 'Testing Overidentifying Restrictions When the Disturbances Are Small', *Journal of the American Statistical Association*, Vol. 65, pp. 182–85.

Kadane, J. B. (1971). 'Comparison of *k*-Class Estimators When the Disturbances Are Small', *Econometrica*, Vol. 39, pp. 723–37.

Kremers, J. J. M. (1988). 'Long-run Limits on the US Federal Debt', *Economics Letters*, Vol. 28, pp. 259–62.

Kremers, J. J. M. (1989). 'U.S. Federal Indebtedness and the Conduct of Fiscal Policy', *Journal of Monetary Economics*, Vol. 23, pp. 219–38.

Kremers, J. J. M., Ericsson, N. R. and Dolado, J. J. (1992). 'The Power of Cointegration Tests', International Finance Discussion Paper No. 431, Board of Governors of the Federal Reserve System, Washington, DC.

MacKinnon, J. G. (1991). 'Critical Values for Cointegration Tests', Chapter 13, in Engle, R. F. and Granger C. W. J. (eds.), *Long-Run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford.

Mann, H. B. and Wald, A. (1943). 'On Stochastic Limit and Order Relationships', *Annals of Mathematical Statistics*, Vol. 14, pp. 217–26.

Nymoen, R. (1992). 'Finnish Manufacturing Wages 1960–1987: Real-wage Flexibility and Hysteresis', *Journal of Policy Modeling*, Vol. 14.

Phillips, P. C. B. (1986). 'Understanding Spurious Regressions in Econometrics', *Journal of Econometrics*, Vol. 33, pp. 311–40.

Phillips, P. C. B. (1987a). 'Time Series Regression with a Unit Root', *Econometrica*, Vol. 55, pp. 277–301.

Phillips, P. C. B. (1987b). 'Towards a Unified Asymptotic Theory for Autoregression', *Biometrika*, Vol. 74, pp. 535–47.

Phillips, P. C. B. (1988). 'Regression Theory for Near-integrated Time Series', *Econometrica*, Vol. 56, pp. 1021–43.

Phillips, P. C. B. and Durlauf, S. N. (1986). 'Multiple Time Series Regression with Integrated Processes', *Review of Economic Studies*, Vol. 53, pp. 473–95.

Phillips, P. C. B. and Park, J. Y. (1988). 'Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions with Integrated Regressors', *Journal of the American Statistical Association*, Vol. 83, pp. 111–15.

Phillips, P. C. B. and Perron, P. (1988). 'Testing for a Unit Root in Time Series Regression', *Biometrika*, Vol. 75, pp. 335–46.

Sargan, J. D. (1964). 'Wages and Prices in the United Kingdom: A Study in Econometric Methodology', in Hart, P. E., Mills, G. and Whitaker, J. K. (eds.), *Econometric Analysis for National Economic Planning*, Colston Papers, Vol. 16, Butterworths, London, 25–63 (with discussion); reprinted in Hendry, D. F. and Wallis, K. F. (eds.) (1984) *Econometrics and Quantitative Economics*, Basil Blackwell, Oxford.

Sargan, J. D. (1980). 'Some Tests of Dynamic Specification for a Single Equation', *Econometrica*, Vol. 48, pp. 879–97.

Sargan, J. D. and Bhargava, A. (1983). 'Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk', *Econometrica*, Vol. 51, pp. 153–74.

Stock, J. H. and Watson, M. W. (1988). 'Testing for Common Trends', *Journal of the American Statistical Association*, Vol. 83, pp. 1097–1107.

White, H. (1984). *Asymptotic Theory for Econometricians*, Academic Press, Orlando, Florida.

APPENDIX: ASYMPTOTIC DISTRIBUTIONS

This Appendix derives asymptotic distributions under a local alternative of cointegration, following (e.g.) Phillips (1987b) and Johansen (1989). The DGP is (1)–(2) with $b = e^{c/T} - 1$. The proofs proceed by rescaling summations to be $O_p(1)$, applying the functional limit results in Table A.1, and dropping terms of $o_p(1)$. Setting $c = 0$ obtains the distributions under the null hypothesis of no cointegration. Distributions under the fixed alternative follow from Section 4.1. See Kremers, Ericsson and Dolado (1992, Appendix) for further details.

Section III's notation for Brownian motion is used throughout. As a reference for the building blocks of the proofs, Table A.1 lists correspondences between sample moments and limiting distributions under the null hypothesis. Correspondences under the local alternative follow from suitable replacement of 'B' by 'K'. See Billingsley (1968, Chapters 2 and 4), White (1984), Phillips (1986, Appendix; 1987a; 1987b; 1988), Phillips and Durlauf (1986), Phillips and Park (1988), Banerjee, Dolado, Hendry and

TABLE A.1

Asymptotic Distributions of Sample Moments Under the Null Hypothesis of No Cointegration

Sample moment	*Brownian motion representation*	*Alternative representation*
	Basic relationships	
$T^{-2}\Sigma(y_t^*)^2$	$\sigma_\varepsilon^2 \int B_\varepsilon^2$	—
$T^{-2}\Sigma z_t^2$	$\sigma_u^2 \int B_u^2$	—
$T^{-2}\Sigma z_t y_t^*$	$\sigma_\varepsilon \sigma_u \int B_\varepsilon B_u$	—
$T^{-1}\Sigma y_{t-1}^* \varepsilon_t$	$\sigma_\varepsilon^2 \int B_\varepsilon \, dB_\varepsilon$	$(\sigma_\varepsilon^2/2)[B_\varepsilon(1)^2 - 1]$
$T^{-1}\Sigma z_{t-1} u_t$	$\sigma_u^2 \int B_u \, dB_u$	$(\sigma_u^2/2)[B_u(1)^2 - 1]$
$T^{-1}\Sigma w_{t-1} e_t$	$\sigma_e^2 \int B_e \, dB_e$	$(\sigma_e^2/2)[B_e(1)^2 - 1]$
$T^{-1}\Sigma y_{t-1}^* u_t$	$\sigma_\varepsilon \sigma_u \int B_\varepsilon \, dB_u$	—
$T^{-1}\Sigma z_{t-1} \varepsilon_t$	$\sigma_\varepsilon \sigma_u \int B_u \, dB_\varepsilon$	—
$T^{-1/2}\Sigma \Delta z_{t-1} \varepsilon_t$	$\sigma_\varepsilon \sigma_u \int dB_u \, dB_\varepsilon$	$N(0, \sigma_\varepsilon^2 \sigma_u^2)$
	Implied auxiliary relationships	
$T^{-1}\Sigma w_{t-1} \varepsilon_t$	$\sigma_\varepsilon \sigma_e \int B_e \, dB_\varepsilon$	or $(a-1)\sigma_\varepsilon \sigma_u \int B_u \, dB_\varepsilon + \sigma_\varepsilon^2 \int B_\varepsilon \, dB_\varepsilon$
$T^{-2}\Sigma w_t^2$	$\sigma_e^2 \int B_e^2$	or $(a-1)^2 \sigma_u^2 \int B_u^2 + 2(a-1)\sigma_\varepsilon \sigma_u \int B_u B_\varepsilon + \sigma_\varepsilon^2 \int B_\varepsilon^2$

Notes:

1. The variable y_t^* is defined as: $y_t^* = \sum_{i=0}^{t} \varepsilon_i$.

2. Because u_t and ε_t are independent and $e_t = (a-1)u_t + \varepsilon_t$, it follows that $\sigma_e B_e = (a-1)\sigma_u B_u + \sigma_\varepsilon B_\varepsilon$ and $\sigma_e \, dB_e = (a-1)\sigma_u \, dB_u + \sigma_\varepsilon \, dB_\varepsilon$. Likewise, under the local alternative, $\sigma_e K_e = (a-1)\sigma_u K_u + \sigma_\varepsilon K_\varepsilon$ and $\sigma_e \, dK_e = (a-1)\sigma_u \, dK_u + \sigma_\varepsilon \, dK_\varepsilon$.

3. Under the local alternative, three of the formulae in the table change: $T^{-1}\Sigma w_{t-1} e_t \Rightarrow \sigma_e^2 \int K_e \, dB_e$, $T^{-1}\Sigma w_{t-1} \varepsilon_t \Rightarrow \sigma_\varepsilon \sigma_e \int K_e \, dB_\varepsilon$, and $T^{-2}\Sigma w_t^2 \Rightarrow \sigma_e^2 \int K_e^2$, with corresponding adjustments for their decompositions.

Smith (1986, Appendix), and Banerjee, Dolado, Galbraith and Hendry (1992) for derivation of the results in the table.

The DF Statistic. The DF statistic is:

$$t_{DF} = (\Sigma w_{t-1}^2)^{-1/2} \cdot (\Sigma w_{t-1} \Delta w_t / \tilde{\sigma}_e)$$

$$= c(T^{-2}\Sigma w_{t-1}^2/\sigma_e^2)^{1/2} + (T^{-2}\Sigma w_{t-1}^2/\sigma_e^2)^{-1/2} \cdot (T^{-1}\Sigma w_{t-1} e_t/\sigma_e^2) + O_p(T^{-1/2})$$

$$\Rightarrow c(\textstyle\int K_e^2)^{1/2} + (\int K_e \, dB_e)/\sqrt{(\int K_e^2)}. \qquad \text{(A1)}$$

See Dickey and Fuller (1979) and Phillips (1987a, 1987b) for details. From (A1), the (exact) distribution of t_{DF} is invariant to the scaling of w_t, and so to the choice of a, σ_u, and σ_ε. With no cointegration, the last line of (A1) simplifies to $(\int B_e \, dB_e)/\sqrt{(\int B_e^2)}$, the 'Dickey-Fuller' distribution.

The ECM Statistic. The OLS estimator $(\hat{a} \;\; \hat{b})'$ in (3) is:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \Sigma(\Delta z_t)^2 & \Sigma \Delta z_t w_{t-1} \\ \Sigma w_{t-1} \Delta z_t & \Sigma w_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \Sigma \Delta z_t \Delta y_t \\ \Sigma w_{t-1} \Delta y_t \end{bmatrix}. \qquad \text{(A2)}$$

Substituting the definition of Δy_t into (A2) and pre-multiplying by the matrix diag$(T^{1/2}, T)$ obtains:

$$\begin{bmatrix} T^{1/2}(\hat{a}-a) \\ T(\hat{b}-b) \end{bmatrix} = \begin{bmatrix} T^{-1}\Sigma(\Delta z_t)^2 & T^{-3/2}\Sigma \Delta z_t w_{t-1} \\ T^{-3/2}\Sigma w_{t-1}\Delta z_t & T^{-2}\Sigma w_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2}\Sigma \Delta z_t \varepsilon_t \\ T^{-1}\Sigma w_{t-1}\varepsilon_t \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & \sigma_e^2 \int K_e^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_\varepsilon \sigma_u \int dB_\varepsilon \, dB_u \\ (a-1)\sigma_\varepsilon \sigma_u \int K_u \, dB_\varepsilon + \sigma_\varepsilon^2 \int K_\varepsilon \, dB_\varepsilon \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} (\sigma_\varepsilon/\sigma_u)\int dB_\varepsilon \, dB_u \\ \{(a-1)\sigma_\varepsilon\sigma_u \int K_u \, dB_\varepsilon + \sigma_\varepsilon^2 \int K_\varepsilon \, dB_\varepsilon\}/\{\sigma_e^2 \int K_e^2\} \end{bmatrix}. \qquad \text{(A3)}$$

The rates of convergence for $\hat{a}$ and $\hat{b}$ imply that:

$$\hat{\sigma}_\varepsilon^2 = \Sigma \hat{\varepsilon}_t^2/(T-2).$$

$$= \sigma_\varepsilon^2 + O_p(T^{-1/2}). \qquad \text{(A4)}$$

By partitioned inversion of the matrices involved in calculating t_{ECM}, and applying the limit results in Table A.1 under the local alternative, the ECM statistic is:

$$t_{ECM} = \hat{b}/ese(\hat{b})$$

$$= c(\sigma_e/\sigma_\varepsilon)(T^{-2}\Sigma w_{t-1}^2/\sigma_e^2)^{1/2} + (T^{-2}\Sigma w_{t-1}^2)^{-1/2}(T^{-1}\Sigma w_{t-1}\varepsilon_t/\sigma_\varepsilon)$$

$$+ O_p(T^{-1/2}) \qquad \text{(A5)}$$

$$\Rightarrow c(1+q^2)^{1/2}(\textstyle\int K_e^2)^{1/2} + (\int K_e \, dB_\varepsilon)/\sqrt{(\int K_e^2)}$$

$$\Rightarrow c(1+q^2)^{1/2}(\textstyle\int K_e^2)^{1/2} + \frac{(a-1)\int K_u \, dB_\varepsilon + s^{-1}\int K_\varepsilon \, dB_\varepsilon}{\sqrt{[(a-1)^2 \int K_u^2 + 2(a-1)s^{-1}\int K_u K_\varepsilon + s^{-2}\int K_\varepsilon^2]}},$$

noting (16) and the relation between e_t, u_t, and ε_t (and so between K_e, K_u, and K_ε).

Under the null hypothesis, (A5) simplifies to (15), which itself can be written as (17) when $a \neq 1$ and as the Dickey-Fuller distribution (13) when $a = 1$. Equations (A3) and (15) correct Banerjee, Dolado, Hendry and Smith (1986, Theorem 4).

Under the local alternative, (A5) simplifies to (A1) when $a = 1$. When $a \neq 1$, (A5) can be reparameterized in terms of c and q alone, rather than in terms of c, a, and s:

$$t_{ECM} \Rightarrow c(1+q^2)^{1/2}([q^2/(1+q^2)]\textstyle\int K_u^2 - 2[q/(1+q^2)]\int K_u K_\varepsilon + (1+q^2)^{-1}\int K_\varepsilon^2)^{1/2}$$

$$+ \frac{\int K_u \, dB_\varepsilon - q^{-1}\int K_\varepsilon \, dB_\varepsilon}{\sqrt{[\int K_u^2 - 2q^{-1}\int K_u K_\varepsilon + q^{-2}\int K_\varepsilon^2]}}, \qquad \text{(A6)}$$

noting that $(1+q^2)K_e^2 = q^2K_u^2 - 2qK_uK_\varepsilon + K_\varepsilon^2$. In order to obtain a 'large-q' approximation without having $t_{ECM} \rightarrow -\infty$, we hold $c(1+q^2)^{1/2}$ constant while expanding in q. Thus, we define a new parameter γ, which is:

$$\gamma = c(1+q^2)^{1/2}. \qquad \text{(A7)}$$

For large q and constant γ, (A6) simplifies to:

$$t_{ECM} \Rightarrow \gamma(\textstyle\int K_u^2)^{1/2} + (\int K_u \, dB_\varepsilon)/\sqrt{(\int K_u^2)} + O_p(q^{-1}). \qquad \text{(A8)}$$

Derivation of the distribution of (A8) parallels Phillips and Park (1988, p. 114, Proof of Theorem 2.3). The bivariate Brownian motion $(B_\varepsilon, K_u)'$ is defined on a probability space, denoted (Ω, F, P). Let F_u denote the sub σ-field of F generated by K_u. Then the second term on the right-hand side of (A8) is a standardized normal distribution, conditional on F_u (and also unconditionally). Thus, t_{ECM} is itself approximately conditionally distributed as a standardized normal variate:

$$t_{ECM}\Big|_{F_u} \Rightarrow N(\gamma(\textstyle\int K_u^2)^{1/2}, 1) + O_p(q^{-1}). \qquad \text{(A9)}$$

In essence, (A9) is conditional on $\{u_t\}$, and so on $\{z_t\}$.

Under the null hypothesis, $\gamma = 0$ so t_{ECM} is both conditionally and unconditionally asymptotic $N(0,1)$, to $O_p(q^{-1})$, from Phillips and Park (1988). Comparison of the unconditional distributions of t_{ECM} and t_{DF} under the local alternative requires several steps.

First, note that the distribution of t_{DF} in (A1) is invariant to q. Thus, for given values of T, c, and its critical value, t_{DF} has a given power, p^* (say). Second, $(\int K_u^2)^{1/2}$ in (A9) is non-negative; and, for any $\theta(1 \geq \theta > 0)$, there exists a $\kappa > 0$ such that:

$$\text{Prob}[(\int K_u^2)^{1/2} \geq \kappa] > 1 - \theta. \tag{A10}$$

Third, note that c is negative; and γ in (A9) is $c(1+q^2)^{1/2}$, which is $O(q)$. Now, consider a critical value for t_{ECM} equivalent to that for t_{DF}. For some q large enough, $\gamma(\int K_u^2)^{1/2}$ [and so t_{ECM} itself] is more negative than that critical value with probability arbitrarily close to unity. Thus, for large q, tests using t_{ECM} have greater power than those using t_{DF}.

An approximation to the unconditional mean of t_{ECM} helps in analyzing the Monte Carlo simulations:

$$E(t_{ECM}) \approx E[\gamma(\int K_u^2)^{1/2}] \approx \gamma[E(\int K_u^2)]^{1/2} \approx \gamma/\surd 2. \tag{A11}$$

The two approximations arriving at $\gamma[E(\int K_u^2)]^{1/2}$ are standard. The derivation of $E(\int K_u^2)$ proceeds as follows.

The integral $\int K_u^2$ can be generated as the large-T limit of $T^{-2}\Sigma\xi_t^2/\sigma_u^2$ for the process:

$$\xi_t = \rho\xi_{t-1} + u_t \quad u_t \sim IN(0, \sigma_u^2) \quad t = 1, \ldots, T, \tag{A12}$$

where $\rho = e^{c/T}$, $c < 0$, and $\xi_0 = 0$. Without loss of generality, $\sigma_u^2 = 1$. For any $t > 0$,

$$\begin{aligned} E(\xi_t^2) &= (1-\rho^{2t})/(1-\rho^2) \\ &= (1-e^{2ct/T})/(1-e^{2c/T}) \end{aligned} \tag{A13}$$

by repeated substitution of (A12). Thus, it follows that:

$$E(T^{-2}\Sigma\xi_t^2) = \frac{T^{-1}}{1-e^{2c/T}} - \left[\frac{T^{-1}}{1-e^{2c/T}}\right]^2 e^{2c/T}[1-e^{2c}]. \tag{A14}$$

Applying L'Hôpital's rule (as $T \rightarrow \infty$), the large-sample limit of (A14) is:

$$\lim_{T\rightarrow\infty} E(T^{-2}\Sigma\xi_t^2) = (e^{2c} - 1 - 2c)/(4c^2). \tag{A15}$$

Applying L'Hôpital's rule again (this time as $q \rightarrow \infty$ and so as $c \rightarrow 0$) obtains:

$$\lim_{c\rightarrow 0} \lim_{T\rightarrow\infty} E(T^{-2}\Sigma\xi_t^2) = \lim_{c\rightarrow 0} e^{2c}/2 = 1/2. \tag{A16}$$