ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Co-Integration and Error Correction: Representation, Estimation, and Testing
Author(s): Robert F. Engle and C. W. J. Granger
Source: *Econometrica*, Vol. 55, No. 2 (Mar., 1987), pp. 251-276
Published by: The Econometric Society
Stable URL: https://www.jstor.org/stable/1913236
Accessed: 25-03-2020 12:31 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/1913236?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

The Econometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Econometrica*

JSTOR

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

Econometrica, Vol. 55, No. 2 (March, 1987), 251-276

CO-INTEGRATION AND ERROR CORRECTION: REPRESENTATION, ESTIMATION, AND TESTING

BY ROBERT F. ENGLE AND C. W. J. GRANGER[1]

The relationship between co-integration and error correction models, first suggested in Granger (1981), is here extended and used to develop estimation procedures, tests, and empirical examples.

If each element of a vector of time series x_t first achieves stationarity after differencing, but a linear combination $\alpha' x_t$ is already stationary, the time series x_t are said to be co-integrated with co-integrating vector α. There may be several such co-integrating vectors so that α becomes a matrix. Interpreting $\alpha' x_t = 0$ as a long run equilibrium, co-integration implies that deviations from equilibrium are stationary, with finite variance, even though the series themselves are nonstationary and have infinite variance.

The paper presents a representation theorem based on Granger (1983), which connects the moving average, autoregressive, and error correction representations for co-integrated systems. A vector autoregression in differenced variables is incompatible with these representations. Estimation of these models is discussed and a simple but asymptotically efficient two-step estimator is proposed. Testing for co-integration combines the problems of unit root tests and tests with parameters unidentified under the null. Seven statistics are formulated and analyzed. The critical values of these statistics are calculated based on a Monte Carlo simulation. Using these critical values, the power properties of the tests are examined and one test procedure is recommended for application.

In a series of examples it is found that consumption and income are co-integrated, wages and prices are not, short and long interest rates are, and nominal GNP is co-integrated with M2, but not M1, M3, or aggregate liquid assets.

KEYWORDS: Co-integration, vector autoregression, unit roots, error correction, multivariate time series, Dickey-Fuller tests.

1. INTRODUCTION

AN INDIVIDUAL ECONOMIC VARIABLE, viewed as a time series, can wander extensively and yet some pairs of series may be expected to move so that they do not drift too far apart. Typically economic theory will propose forces which tend to keep such series together. Examples might be short and long term interest rates, capital appropriations and expenditures, household income and expenditures, and prices of the same commodity in different markets or close substitutes in the same market. A similar idea arises from considering equilibrium relationships, where equilibrium is a stationary point characterized by forces which tend to push the economy back toward equilibrium whenever it moves away. If x_t is a vector of economic variables, then they may be said to be in equilibrium when the specific linear constraint

$$\alpha' x_t = 0$$

[1] The authors are indebted to David Hendry and Sam Yoo for many useful conversations and suggestions as well as to Gene Savin, David Dickey, Alok Bhargava, and Marco Lippi. Two referees provided detailed constructive criticism, and thanks go to Yoshi Baba, Sam Yoo, and Alvaro Ecribano who creatively carried out the simulations and examples. Financial support was provided by NSF SES-80-08580 and SES-82-08626. A previous version of this paper was entitled "Dynamic Model Specification with Equilibrium Constraints: Co-integration and Error Correction."

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

occurs. In most time periods, x_t will not be in equilibrium and the univariate quantity

$$z_t = \alpha' x_t$$

may be called the equilibrium error. If the equilibrium concept is to have any relevance for the specification of econometric models, the economy should appear to prefer a small value of z_t rather than a large value.

In this paper, these ideas are put onto a firm basis and it is shown that a class of models, known as error-correcting, allows long-run components of variables to obey equilibrium constraints while short-run components have a flexible dynamic specification. A condition for this to be true, called co-integration, was introduced by Granger (1981) and Granger and Weiss (1983) and is precisely defined in the next section. Section 3 discusses several representations of co-integrated systems, Section 4 develops estimation procedures, and Section 5 develops tests. Several applications are presented in Section 6 and conclusions are offered in Section 7. A particularly simple example of this class of models is shown in Section 4, and it might be useful to examine it for motivating the analysis of such systems.

2. INTEGRATION, CO-INTEGRATION, AND ERROR CORRECTION

It is well known from Wold's theorem that a single stationary time series with no deterministic components has an infinite moving average representation which is generally approximated by a finite autoregressive moving average process. See, for example, Box and Jenkins (1970) or Granger and Newbold (1977). Commonly however, economic series must be differenced before the assumption of stationarity can be presumed to hold. This motivates the following familiar definition of integration:

DEFINITION: A series with no deterministic component which has a stationary, invertible, ARMA representation after differencing d times, is said to be integrated of order d, denoted $x_t \sim I(d)$.

For ease of exposition, only the values $d = 0$ and $d = 1$ will be considered in much of the paper, but many of the results can be generalized to other cases including the fractional difference model. Thus, for $d = 0$ x_t will be stationary and for $d = 1$ the change is stationary.

There are substantial differences in appearance between a series that is $I(0)$ and another that is $I(1)$. For more discussion see, for example, Feller (1968) or Granger and Newbold (1977).

(a) If $x_t \sim I(0)$ with zero mean then (i) the variance of x_t is finite; (ii) an innovation has only a temporary effect on the value of x_t; (iii) the spectrum of x_t, $f(\omega)$, has the property $0 < f(0) < \infty$; (iv) the expected length of times between crossings of $x = 0$ is finite; (v) the autocorrelations, ρ_k, decrease steadily in magnitude for large enough k, so that their sum is finite.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

(b) If $x_t \sim I(1)$ with $x_0 = 0$, then (i) variance x_t goes to infinity as t goes to infinity; (ii) an innovation has a permanent effect on the value of x_t, as x_t is the sum of all previous changes; (iii) the spectrum of x_t has the approximate shape $f(\omega) \sim A\omega^{-2d}$ for small ω so that in particular $f(0) = \infty$; (iv) the expected time between crossings of $x = 0$ is infinite; (v) the theoretical autocorrelations, $\rho_k \to 1$ for all k as $t \to \infty$.

The theoretical infinite variance for an $I(1)$ series comes completely from the contribution of the low frequencies, or long run part of the series. Thus an $I(1)$ series is rather smooth, having dominant long swings, compared to an $I(0)$ series. Because of the relative sizes of the variances, it is always true that the sum of an $I(0)$ and an $I(1)$ will be $I(1)$. Further, if a and b are constants, $b \neq 0$, and if $x_t \sim I(d)$, then $a + bx_t$ is also $I(d)$.

If x_t and y_t are both $I(d)$, then it is generally true that the linear combination

$$z_t = x_t - ay_t$$

will also be $I(d)$. However, it is possible that $z_t \sim I(d-b)$, $b > 0$. When this occurs, a very special constraint operates on the long-run components of the series. Consider the case $d = b = 1$, so that x_t, y_t are both $I(1)$ with dominant long run components, but z_t is $I(0)$ without especially strong low frequencies. The constant a is therefore such that the bulk of the long run components of x_t and y_t cancel out. For $a = 1$, the vague idea that x_t and y_t cannot drift too far apart has been translated into the more precise statement that "their difference will be $I(0)$." The use of the constant a merely suggests that some scaling needs to be used before the $I(0)$ difference can be achieved. It should be noted that it will not generally be true that there is an a which makes $z_t \sim I(0)$.

An analogous case, considering a different important frequency, is when x_t and y_t are a pair of series, each having important seasonal component, yet there is an a so that the derived series z_t has no seasonal. Clearly this could occur, but might be considered to be unlikely.

To formalize these ideas, the following definition adapted from Granger (1981) and Granger and Weiss (1983) is introduced:

DEFINITION: The components of the vector x_t are said to be *co-integrated of order d, b*, denoted $x_t \sim CI(d, b)$, if (i) all components of x_t are $I(d)$; (ii) there exists a vector $\alpha (\neq 0)$ so that $z_t = \alpha' x_t \sim I(d-b)$, $b > 0$. The vector α is called the *co-integrating vector.*

Continuing to concentrate on the $d = 1$, $b = 1$ case, co-integration would mean that if the components of x_t were all $I(1)$, then the equilibrium error would be $I(0)$, and z_t will rarely drift far from zero if it has zero mean and z_t will often cross the zero line. Putting this another way, it means that equilibrium will occasionally occur, at least to a close approximation, whereas if x_t was not co-integrated, then z_t can wander widely and zero-crossings would be very rare, suggesting that in this case the equilibrium concept has no practical implications.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

The reduction in the order of integration implies a special kind of relationship with interpretable and testable consequences. If however all the elements of x_t are already stationary so that they are $I(0)$, then the equilibrium error z_t has no distinctive property if it is $I(0)$. It could be that $z_t \sim I(-1)$, so that its spectrum is zero at zero frequency, but if any of the variables have measurement error, this property in general cannot be observed and so this case is of little realistic interest. When interpreting the co-integration concept it might be noted that in the $N=2$, $d=b=1$ case, Granger and Weiss (1983) show that a necessary and sufficient condition for co-integration is that the coherence between the two series is one at zero frequency.

If x_t has N components, then there may be more than one cointegrating vector α. It is clearly possible for several equilibrium relations to govern the joint behavior of the variables. In what follows, it will be assumed that there are exactly r linearly independent co-integrating vectors, with $r \leq N-1$, which are gathered together into the $N \times r$ array α. By construction the rank of α will be r which will be called the "co-integrating rank" of x_t.

The close relationship between co-integration and error correcting models will be developed in the balance of the paper. Error correction mechanisms have been used widely in economics. Early versions are Sargan (1964) and Phillips (1957). The idea is simply that a proportion of the disequilibrium from one period is corrected in the next period. For example, the change in price in one period may depend upon the degree of excess demand in the previous period. Such schemes can be derived as optimal behavior with some types of adjustment costs or incomplete information. Recently, these models have seen great interest following the work of Davidson, Hendry, Srba, and Yeo (1978) (DHSY), Hendry and von Ungern Sternberg (1980), Currie (1981), Dawson (1981), and Salmon (1982) among others.

For a two variable system a typical error correction model would relate the change in one variable to past equilibrium errors, as well as to past changes in both variables. For a multivariate system we can define a general error correction representation in terms of B, the backshift operator, as follows.

DEFINITION: A vector time series x_t has an error correction representation if it can be expressed as:

$$A(B)(1-B)x_t = -\gamma z_{t-1} + u_t$$

where u_t is a stationary multivariate disturbance, with $A(0)=I$, $A(1)$ has all elements finite, $z_\tau = \alpha' x_\tau$, and $\gamma \neq 0$.

In this representation, only the disequilibrium in the previous period is an explanatory variable. However, by rearranging terms, any set of lags of the z can be written in this form, therefore it permits any type of gradual adjustment toward a new equilibrium. A notable difference between this definition and most of the applications which have occurred is that this is a multivariate definition which does not rest on exogeneity of a subset of the variables. The notion that one

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

variable may be weakly exogenous in the sense of Engle, Hendry, and Richard (1983) may be investigated in such a system as briefly discussed below. A second notable difference is that α is taken to be an unknown parameter vector rather than a set of constants given by economic theory.

3. PROPERTIES OF CO-INTEGRATED VARIABLES AND THEIR REPRESENTATIONS

Suppose that each component of x_t is $I(1)$ so that the change in each component is a zero mean purely nondeterministic stationary stochastic process. Any known deterministic components can be subtracted before the analysis is begun. It follows that there will always exist a multivariate Wold representation:

$$(3.1) \qquad (1-B)x_t = C(B)\varepsilon_t,$$

taken to mean that both sides will have the same spectral matrix. Further, $C(B)$ will be uniquely defined by the conditions that the function $\det[C(z)]$, $z = e^{i\omega}$, have all zeroes on or outside the unit circle, and that $C(0) = I_N$, the $N \times N$ identity matrix (see Hannan (1970, p. 66)). In this representation the ε_t are zero mean white noise vectors with

$$E[\varepsilon_t \varepsilon_s'] = 0, \quad t \neq s,$$
$$= G, \quad t = s,$$

so that only contemporaneous correlations can occur.

The moving average polynomial $C(B)$ can always be expressed as

$$(3.2) \qquad C(B) = C(1) + (1-B)C^*(B)$$

by simply rearranging the terms. If $C(B)$ is of finite order, then $C^*(B)$ will be of finite order. If $C^*(1)$ is identically zero, then a similar expression involving $(1-B)^2$ can be defined.

The relationship between error correction models and co-integration was first pointed out in Granger (1981). A theorem showing precisely that co-integrated series can be represented by error correction models was originally stated and proved in Granger (1983). The following version is therefore called the Granger Representation Theorem. Analysis of related but more complex cases is covered by Johansen (1985) and Yoo (1985).

GRANGER REPRESENTATION THEOREM: If the $N \times 1$ vector x_t given in (3.1) is co-integrated with $d = 1$, $b = 1$ and with co-integrating rank r, then:

(1) $C(1)$ *is of rank* $N - r$.

(2) *There exists a vector ARMA representation*

$$(3.3) \qquad A(B)x_t = d(B)\varepsilon_t$$

with the properties that $A(1)$ *has rank* r *and* $d(B)$ *is a scalar lag polynomial with* $d(1)$ *finite, and* $A(0) = I_N$. *When* $d(B) = 1$, *this is a vector autoregression.*

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

(3) *There exist* $N \times r$ *matrices,* α, γ, *of rank* r *such that*

$$\alpha' C(1) = 0,$$

$$C(1)\gamma = 0,$$

$$A(1) = \gamma\alpha'.$$

(4) *There exists an error correction representation with* $z_t = \alpha' x_t$, *an* $r \times 1$ *vector of stationary random variables*:

$$A^*(B)(1-B)x_t = -\gamma z_{t-1} + d(B)\varepsilon_t \tag{3.4}$$

with $A^*(0) = I_N$.

(5) *The vector* z_t *is given by*

$$z_t = K(B)\varepsilon_t, \tag{3.5}$$

$$(1-B)z_t = -\alpha'\gamma z_{t-1} + J(B)\varepsilon_t, \tag{3.6}$$

where $K(B)$ *is an* $r \times N$ *matrix of lag polynomials given by* $\alpha' C^*(B)$ *with all elements of* $K(1)$ *finite with rank* r, *and det* $(\alpha'\gamma) > 0$.

(6) *If a finite vector autoregressive representation is possible, it will have the form given by* (3.3) *and* (3.4) *above with* $d(B) = 1$ *and both* $A(B)$ *and* $A^*(B)$ *as matrices of finite polynomials.*

In order to prove the Theorem the following lemma on determinants and adjoints of singular matrix polynomials is needed.

LEMMA 1: *If* $G(\lambda)$ *is a finite valued* $N \times N$ *matrix polynomial on* $\lambda \in [0, 1]$, *with rank* $G(0) = N - r$ *for* $0 \leq r \leq N$, *and if* $G^*(0) \neq 0$ *in*

$$G(\lambda) = G(0) + \lambda G^*(\lambda),$$

then

(i) $\det(G(\lambda)) = \lambda^r g(\lambda) I_N$ *with* $g(0)$ *finite,*

(ii) $Adj(G(\lambda)) = \lambda^{r-1} H(\lambda)$,

where I_N *is the* $N \times N$ *identity matrix,* $1 \leq rank\ (H(0)) \leq r$, *and* $H(0)$ *is finite.*

PROOF: The determinant of G can be expressed in a power series in λ as

$$\det(G(\lambda)) = \sum_{i=0}^{\infty} \delta_i \lambda^i.$$

Each δ_i is a sum of a finite number of products of elements of $G(\lambda)$ and therefore is itself finite valued. Each has some terms from $G(0)$ and some from $\lambda G^*(\lambda)$. Any product with more than $N - r$ terms from $G(0)$ will be zero because this will be the determinant of a submatrix of larger order than the rank of $G(0)$. The only possible non-zero terms will have r or more terms from $\lambda G^*(\lambda)$ and

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

therefore will be associated with powers of λ of r or more. The first possible nonzero δ_i is δ_r.

Defining

$$g(\lambda) = \sum_{i=r}^{\infty} \delta_i \lambda^{i-r}$$

establishes the first part of the lemma since δ_r must be finite.

To establish the second statement, express the adjoint matrix of G in a power series in λ:

$$\text{Adj } G(\lambda) = \sum_{i=0}^{\infty} \lambda^i H_i,$$

Since the adjoint is a matrix composed of elements which are determinants of order $N-1$, the above argument establishes that the first $r-1$ terms must be identically zero. Thus

$$\begin{aligned}\text{Adj } G(\lambda) &= \lambda^{r-1} \sum_{r-1}^{\infty} \lambda^{i-r+1} H_i \\ &= \lambda^{r-1} H(\lambda).\end{aligned}$$

Because the elements of H_{r-1} are products of finitely many finite numbers, $H(0)$ must be finite.

The product of a matrix and its adjoint will always give the determinant so:

$$\begin{aligned}\lambda^r g(\lambda) I_N &= (G(0) + \lambda G^*(\lambda)) H(\lambda) \\ &= G(0) H(\lambda) \lambda^{r-1} + h(\lambda) G^*(\lambda) \lambda^r.\end{aligned}$$

Equating powers of λ we get

$$G(0)H(0) = 0.$$

Thus the rank of $H(0)$ must be less than or equal to r as it lies entirely in the column null space of the rank $N-r$ matrix $G(0)$. If $r=1$, the first term in the expression for the adjoint will simply be the adjoint of $G(0)$ which will have rank 1 since $G(0)$ has rank $N-1$. *Q.E.D.*

PROOF OF GRANGER REPRESENTATION THEOREM: The conditions of the Theorem suppose the existance of a Wold representation as in (3.1) for an N vector of random variables x_t which are co-integrated. Suppose the co-integrating vector is α so that

$$z_t = \alpha' x_t$$

is an r-dimensional stationary purely nondeterministic time series with invertible moving average representation. Multiplying α times the moving average representation in (3.1) gives

$$(1-B) z_t = (\alpha' C(1) + (1-B)\alpha' C^*(B)) \varepsilon_t.$$

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

For z_t to be $I(0)$, $\alpha' C(1)$ must equal 0. Any vector with this property will be a co-integrating vector; therefore $C(1)$ must have rank $N-r$ with a null space containing all co-integrating vectors. It also follows that $\alpha' C^*(B)$ must be an invertible moving average representation and in particular $\alpha' C^*(1) \neq 0$. Otherwise the co-integration would be with $b=2$ or higher.

Statement (2) is established using Lemma 1, letting $\lambda = (1-B)$, $G(\lambda) = C(B)$, $H(\lambda) = A(B)$, and $g(\lambda) = d(B)$. Since $C(B)$ has full rank and equals I_N at $B=0$, its inverse is $A(0)$ which is also I_N.

Statement (3) follows from recognition that A(1) has rank between 1 and r and lies in the null space of C(1). Since α spans this null space, $A(1)$ can be written as linear combinations of the co-integrating vectors

$$A(1) = \gamma\alpha'.$$

Statement (4) follows by manipulation of the autoregressive structure. Rearranging terms in (3.3) gives:

$$[\tilde{A}(B) + A(1)](1-B)x_t = -A(1)x_{t-1} + d(B)\varepsilon_t,$$

$$A^*(B)(1-B)x_t = -\gamma z_{t-1} + d(B)\varepsilon_t,$$

$A^*(0) = A(0) = I_N$.

The fifth condition follows from direct substitution in the Wold representation. The definition of co-integration implies that this moving average be stationary and invertible. Rewriting the error correction representation with $A^*(B) = I + A^{**}(B)$ where $A^{**}(0) = 0$, and premultiplying by α' gives:

$$(1-B)z_t = -\alpha'\gamma z_{t-1} + [\alpha' d(B) + \alpha' A^{**}(B)C(B)]\varepsilon_t$$

$$= -\alpha'\gamma z_{t-1} + J(B)\varepsilon_t.$$

For this to be equivalent to the stationary moving average representation the autoregression must be invertible. This requires that $\det(\alpha'\gamma) > 0$. If the determinant were zero then there would be at least one unit root, and if the determinant were negative, then for some value of ω between zero and one,

$$\det(I_r - (I_r - \alpha'\gamma)\omega) = 0,$$

implying a root inside the unit circle.

Condition six follows by repeating the previous steps, setting $d(B) = 1$.

Q.E.D.

Stronger results can be obtained by further restrictions on the multiplicity of roots in the moving average representations. For example, Yoo (1985), using Smith Macmillan forms, finds conditions which establish that $d(1) \neq 0$, that $A^*(1)$ is of full rank, and that facilitate the transformation from error correction models to co-integrated models. However, the results given above are sufficient for the estimation and testing problems addressed in this paper.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

The autoregressive and error correction representations given by (3.3) and (3.4) are closely related to the vector autoregressive models so commonly used in econometrics, particularly in the case when $d(B)$ can reasonably be taken to be 1. However, each differs in an important fashion from typical VAR applications. In the autoregressive representation

$$A(B)x_t = \varepsilon_t,$$

the co-integration of the variables x_t generates a restriction which makes $A(1)$ singular. For $r = 1$, this matrix will only have rank 1. The analysis of such systems from an innovation accounting point of view is treacherous as some numerical approaches to calculating the moving average representation are highly unstable.

The error correction representation

$$A^*(B)(1-B)x_t = -\gamma\alpha' x_{t-1} + \varepsilon_t$$

looks more like a standard vector autoregression in the differences of the data. Here the co-integration is implied by the presence of the levels of the variables so a pure VAR in differences will be misspecified if the variables are co-integrated.

Thus vector autoregressions estimated with co-integrated data will be misspecified if the data are differenced, and will have omitted important constraints if the data are used in levels. Of course, these constraints will be satisfied asymptotically but efficiency gains and improved multistep forecasts may be achieved by imposing them.

As $x_t \sim I(1)$, $z_t \sim I(0)$, it should be noted that all terms in the error correction models are $I(0)$. The converse also holds; if $x_t \sim I(1)$ are generated by an error correction model, then x_t is necessarily co-integrated. It may also be noted that if $x_t \sim I(0)$, the generation process can always be written in the error correction form and so, in this case, the equilibrium concept has no impact.

As mentioned above, typical empirical examples of error correcting behavior are formulated as the response of one variable, the dependent variable, to shocks of another, the independent variable. In this paper all the variables are treated as jointly endogenous; nevertheless the structure of the model may imply various Granger causal orderings and weak and strong exogeneity conditions as in Engle, Hendry, and Richard (1983). For example, a bivariate co-integrated system must have a causal ordering in at least one direction. Because the z's must include both variables and γ cannot be identically zero, they must enter into one or both of the equations. If the error correction term enters into both equations, neither variable can be weakly exogenous for the parameters of the other equation because of the cross equation restriction.

The notion of co-integration can in principle be extended to series with trends or explosive autoregressive roots. In these cases the co-integrating vector would still be required to reduce the series to stationarity. Hence the trends would have to be proportional and any explosive roots would have to be identical for all the series. We do not consider these cases in this paper and recognize that they may complicate the estimation and testing problems.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

4. ESTIMATING CO-INTEGRATED SYSTEMS

In defining different forms for co-integrated systems, several estimation procedures have been implicitly discussed. Most convenient is the error correction form (particularly if it can be assumed that there is no moving average term). There remain cross-equation restrictions involving the parameters of the co-integrating vectors; and therefore the maximum likelihood estimator, under Gaussian assumptions, requires an iterative procedure.

In this section, we will propose another estimator which is a two step estimator. In the first step the parameters of the co-integrating vector are estimated and in the second these are used in the error correction form. Both steps require only single equation least squares and it will be shown that the result is consistent for all the parameters. The procedure is far more convenient because the dynamics do not need to be specified until the error correction structure has been estimated. As a byproduct we obtain some test statistics useful for testing for co-integration.

From (3.5) the sample moment matrix of the data can be directly expressed. Let the moment matrix divided by T be denoted by:

$$M_T = 1/T^2 \sum_t x_t x_t'.$$

Recalling that $z_t = \alpha' x_t$, (3.5) implies that

$$\alpha' M_T = \sum_t [K(B)\varepsilon_t] x_t' / T^2.$$

Following the argument of Dickey and Fuller (1979) or Stock (1984), it can be shown that for processes satisfying (3.1),

$$\lim_{T \to \infty} E(M_T) = M \quad \text{a finite nonzero matrix,} \tag{4.1}$$

and

$$\alpha' M = 0, \quad \text{or} \quad (\text{vec}\,\alpha)'(I \otimes M) = 0. \tag{4.2}$$

Although the moment matrix of data from a co-integrated process will be nonsingular for any sample, in the limit, it will have rank $N - r$. This accords well with the common observation that economic time series data are highly collinear so that moment matrices may be nearly singular even when samples are large. Co-integration appears to be a plausible hypothesis from a data analytic point of view.

Equations (4.2) do not uniquely define the co-integrating vectors unless arbitrary normalizations are imposed. Let q and Q be arrays which incorporate these normalizations by reparametrizing α into θ, a $j \times 1$ matrix of unknown parameters which lie in a compact subset of R^j:

$$\text{vec}\,\alpha = q + Q\theta. \tag{4.3}$$

Typically q and Q will be all zeros and ones, thereby defining one coefficient in each column of α to be unity and defining rotations if $r > 1$. The parameters θ

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

are said to be "identified" if there is a unique solution to (4.2), (4.3). This solution is given by

$$(4.4) \qquad (I \otimes M)Q\theta = -(I \otimes M)q$$

where by the assumption of identification, $(I \otimes M)Q$ has a left inverse even though M does not.

As the moment matrix M_T will have full rank for finite samples, a reasonable approach to estimation is to minimize the sum of squared deviations from equilibrium. In the case of a single co-integrating vector, $\hat{\alpha}$ will minimize $\alpha' M_T \alpha$ subject to any restrictions such as (4.3) and the result will be simply ordinary least squares. For multiple co-integrating vectors, define $\hat{\alpha}$ as the minimizer of the trace $(\alpha' M_T \alpha)$. The estimation problem becomes:

$$\begin{aligned} \underset{\alpha s.t.(4.3)}{\text{Min}} \operatorname{tr}(\alpha' M_T \alpha) &= \underset{\alpha s.t.\ (4.3)}{\text{Min}} \operatorname{vec} \alpha' (I \otimes M_T) \operatorname{vec} \alpha \\ &= \underset{\theta}{\text{Min}} (q + Q\theta)'(I \otimes M_T)(q + Q\theta), \end{aligned}$$

which implies the solution

$$(4.5) \qquad \hat{\theta} = -(Q'(I \otimes M_T)Q)^{-1}(Q'(I \otimes M_T)q), \quad \operatorname{vec} \hat{\alpha} = q + Q\hat{\theta}.$$

This approach to estimation should provide a very good approximation to the true co-integrating vector because it is seeking vectors with minimal residual variance and asymptotically all linear combinations of x will have infinite variance except those which are co-integrating vectors.

When $r = 1$ this estimate is obtained simply by regressing the variable normalized to have a unit coefficient upon the other variables. This regression will be called the "co-integrating regression" as it attempts to fit the long run or equilibrium relationship without worrying about the dynamics. It will be shown to provide an estimate of the elements of the co-integrating vector. Such a regression has been pejoratively called a "spurious" regression by Granger and Newbold (1974) primarily because the standard errors are highly misleading. They were particularly concerned about the non-co-integrated case where there was no relationship but the unit root in the error process led to a low Durbin Watson, a high R^2, and apparently high significance of the coefficients. Here we only seek coefficient estimates to use in the second stage and for tests of the equilibrium relationship. The distribution of the estimated coefficients is investigated in Stock (1984).

When $N = 2$, there are two possible regressions depending on the normalization chosen. The nonuniqueness of the estimate derives from the well known fact that the least squares fit of a reverse regression will not give the reciprocal of the coefficient in the forward regression. In this case, however, the normalization matters very little. As the moment matrix approaches singularity, the R^2 approaches 1 which is the product of the forward and reverse regression coefficients. This would be exactly true if there were only two data points which,

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

of course, defines a singular matrix. For variables which are trending together, the correlation approaches one as each variance approaches infinity. The regression line passes nearly through the extreme points almost as if there were just two observations.

Stock (1984) in Theorem 3 proves the following proposition:

PROPOSITION 1: *Suppose that x_t satisfies* (3.1) *with $C^*(B)$ absolutely summable, that the disturbances have finite fourth absolute moments, and that x_t is co-integrated* $(1, 1)$ *with r co-integrating vectors satisfying* (4.3) *which identify θ. Then, defining $\hat{\theta}$ by* (4.5),

$$(4.6) \qquad T^{1-\delta}(\hat{\theta}-\theta) \xrightarrow{p} 0 \quad \text{for} \quad \delta > 0.$$

The proposition establishes that the estimated parameters converge very rapidly to their probability limits. It also establishes that the estimates are consistent with a finite sample bias of order $1/T$. Stock presents some Monte Carlo examples to show that these biases may be important for small samples and gives expressions for calculating the limiting distribution of such estimates.

The two step estimator proposed for this co-integrated system uses the estimate of α from (4.5) as a known parameter in estimating the error correction form of the system of equations. This substantially simplifies the estimation procedure by imposing the cross-equation restrictions and allows specification of the individual equation dynamic patterns separately. Notice that the dynamics did not have to be specified in order to estimate α. Surprisingly, this two-step estimator has excellent properties; as shown in the Theorem below, it is just as efficient as the maximum likelihood estimator based on the known value of α.

THEOREM 2: *The two-step estimator of a single equation of an error correction system, obtained by taking $\hat{\alpha}$ from* (4.5) *as the true value, will have the same limiting distribution as the maximum likelihood estimator using the true value of α. Least squares standard errors will be consistent estimates of the true standard errors.*

PROOF: Rewrite the first equation of the error correction system (3.4) as

$$y_t = \gamma \hat{z}_{t-1} + W_t\beta + \varepsilon_t + \gamma(z_{t-1} - \hat{z}_{t-1}),$$

$$z_t = X_t\alpha,$$

$$\hat{z}_t = X_t\hat{\alpha},$$

where $X_t = x_t'$, W is an array with selected elements of Δx_{t-i} and y is an element of Δx_t so that all regressors are $I(0)$. Then letting the same variables without subscripts denote data arrays,

$$\sqrt{T}\begin{bmatrix} \gamma & -\gamma \\ \beta & -\beta \end{bmatrix} = [(\hat{z}, W)'(\hat{z}, W)/T]^{-1}[(\hat{z}, W)'(\varepsilon+\gamma)(z-\hat{z})]/\sqrt{T}.$$

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

This expression simplifies because $\hat{z}'(z-\hat{z})=0$. From Fuller (1976) or Stock (1984), $X'X/T^2$ and $X'W/T$ are both of order 1. Rewriting,

$$W'(z-\hat{z})/\sqrt{T}=[W'X/T][T(\alpha-\hat{\alpha})][1/\sqrt{T}],$$

and therefore the first and second factors to the right of the equal sign are of order 1 and the third goes to zero so that the entire expression vanishes asymptotically. Because the terms in $(z-\hat{z})/\sqrt{T}$ vanish asymptotically, least squares standard errors will be consistent.

Letting $S=\text{plim}\,[(\hat{z}, W)'(\hat{z}, W)/T]$,

$$\sqrt{T}\begin{bmatrix}\gamma & -\gamma\\ \beta & -\beta\end{bmatrix} \xrightarrow{A} D(0, \sigma^2 S^{-1})$$

where D represents the limiting distribution. Under additional but standard assumptions, this could be guaranteed to be normal.

To establish that the estimator using the true value of α has the same limiting distribution it is sufficient to show that the probability limit of $[(z, W)'(z, W)/T]$ is also S and that $z'\varepsilon/\sqrt{T}$ has the same limiting distribution as $\hat{z}'\varepsilon/\sqrt{T}$. Examining the off diagonal terms of S first,

$$\hat{z}'W/T-z'W/T=T(\hat{\alpha}-\alpha)'[W'X/T](1/T).$$

The first and second factors are of order 1 and the third is $1/T$ so the entire expression vanishes asymptotically:

$$\begin{aligned}(\hat{z}-z)'(\hat{z}-z)/T &= z'z/T-\hat{z}'\hat{z}/T\\ &= T(\hat{\alpha}-\alpha)'[X'X/T^2]T(\hat{\alpha}-\alpha)(1/T).\end{aligned}$$

Again, the first three factors are of order 1 and the last is $1/T$ so even though the difference between these covariance matrices is positive definite, it will vanish asymptotically. Finally,

$$(\hat{z}-z)'\varepsilon/\sqrt{T}=T(\hat{\alpha}-\alpha)'[X'\varepsilon/T]1/\sqrt{T},$$

which again vanishes asymptotically.

Under standard conditions the estimator using knowledge of α will be asymptotically normal and therefore the two-step estimator will also be asymptotically normal under these conditions. This completes the proof. *Q.E.D.*

A simple example will illustrate many of these points and motivate the approach to testing described in the next section. Suppose there are two series, x_{1t} and x_{2t}, which are jointly generated as a function of possibly correlated white noise disturbances ε_{1t} and ε_{2t} according to the following model:

$$x_{1t}+\beta x_{2t}=u_{1t}, \quad u_{1t}=u_{1t-1}+\varepsilon_{1t}, \tag{4.7}$$

$$x_{1t}+\alpha x_{2t}=u_{2t}, \quad u_{2t}=\rho u_{2t-1}+\varepsilon_{2t}, \quad |\rho|<1. \tag{4.8}$$

Clearly the parameters α and β are unidentified in the usual sense as there are no exogenous variables and the errors are contemporaneously correlated. The

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

reduced form for this system will make x_{1t} and x_{2t} linear combinations of u_{1t} and u_{2t} and therefore both will be $I(1)$. The second equation describes a particular linear combination of the random variables which is stationary. Hence x_{1t} and x_{2t} are $CI(1, 1)$ and the question is whether it would be possible to detect this and estimate the parameters from a data set.

Surprisingly, this is easy to do. A linear least squares regression of x_{1t} on x_{2t} produces an excellent estimate of α. This is the "co-integrating regression." All linear combinations of x_{1t} and x_{2t} except that defined in equation (4.8) will have infinite variance and, therefore, least squares is easily able to estimate α. The correlation between x_{2t} and u_{2t} which causes the simultaneous equations bias is of a lower order in T than the variance of x_{2t}. In fact the reverse regression of x_{2t} on x_{1t} has exactly the same property and thus gives a consistent estimate of $1/\alpha$. These estimators converge even faster to the true value than standard econometric estimates.

While there are other consistent estimates of α, several apparently obvious choices are not. For example, regression of the first differences of x_1 on the differences of x_2 will not be consistent, and the use of Cochrane Orcutt or other serial correlation correction in the co-integrating regression will produce inconsistent estimates. Once the parameter α has been estimated, the others can be estimated in a variety of ways conditional on the estimate of α.

The model in (4.7) and (4.8) can be expressed in the autoregressive representation (after subtracting the lagged values from both sides and letting $\delta = (1-\rho)/(\alpha-\beta)$) as:

$$\Delta x_{1t} = \beta\delta x_{1t-1} + \alpha\beta\delta x_{2t-1} + \eta_{1t}, \tag{4.9}$$

$$\Delta x_{2t} = -\delta x_{1t-1} - \alpha\delta x_{2t-1} + \eta_{2t}, \tag{4.10}$$

where the η's are linear combinations of the ε's. The error correction representation becomes:

$$\Delta x_{1t} = \beta\delta z_{t-1} + \eta_{1t}, \tag{4.11}$$

$$\Delta x_{2t} = -\delta z_{t-1} + \eta_{2t}, \tag{4.12}$$

where $z_t = x_{1t} + \alpha x_{2t}$. There are three unknown parameters but the autoregressive form apparently has four unknown coefficients while the error correction form has two. Once α is known there are no longer constraints in the error correction form which motivates the two-step estimator. Notice that if $\rho \to 1$, the series are correlated random walks but are no longer co-integrated.

5. TESTING FOR CO-INTEGRATION

It is frequently of interest to test whether a set of variables are co-integrated. This may be desired because of the economic implications such as whether some system is in equilibrium in the long run, or it may be sensible to test such hypotheses before estimating a multivariate dynamic model.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

Unfortunately the set-up is nonstandard and cannot simply be viewed as an application of Wald, likelihood ratio, or Lagrange multiplier tests. The testing problem is closely related to tests for unit roots in observed series as initially formulated by Fuller (1976) and Dickey and Fuller (1979, 1981) and more recently by Evans and Savin (1981), Sargan and Bhargava (1983), and Bhargava (1984), and applied by Nelson and Plosser (1983). It also is related to the problem of testing when some parameters are unidentified under the null as discussed by Davies (1977) and Watson and Engle (1982).

To illustrate the problems in testing such an hypothesis, consider the simple model in (4.7) and (4.8). The null hypothesis is taken to be no co-integration or $\rho = 1$. If α were known, then a test for the null hypothesis could be constructed along the lines of Dickey and Fuller taking z_t as the series which has a unit root under the null. The distribution in this case is already nonstandard and was computed through a simulation by Dickey (1976). However, when α is not known, it must be estimated from the data. But if the null hypothesis that $\rho = 1$ is true, α is not identified. Thus only if the series are co-integrated can α be simply estimated by the "co-integrating regression," but a test must be based upon the distribution of a statistic when the null is true. OLS seeks the α which minimizes the residual variance and therefore is most likely to be stationary, so the distribution of the Dickey-Fuller test will reject the null too often if α must be estimated.

In this paper a set of seven test statistics is proposed for testing the null of non-co-integration against the alternative of co-integration. It is maintained that the true system is a bivariate linear vector autoregression with Gaussian errors where each of the series is individually $I(1)$. As the null hypothesis is composite, similar tests will be sought so that the probability of rejection will be constant over the parameter set included in the null. See, for example, Cox and Hinkley (1974, p. 134–136).

Two cases may be distinguished. In the first, the system is known to be of first order and therefore the null is defined by

$$\Delta y_t = \varepsilon_{1t}, \quad \Delta x_t = \varepsilon_{2t}, \quad \begin{bmatrix} (\varepsilon_{1t}) \\ (\varepsilon_{2t}) \end{bmatrix} \sim N(0, \Omega). \tag{5.1}$$

This is clearly the model implied by (4.11) and (4.12) when $\rho = 1$ which implies that $\delta = 0$. The composite null thus includes all positive definite covariance matrices Ω. It will be shown below that all the test statistics are similar with respect to the matrix Ω so without loss of generality, we take $\Omega = I$.

In the second case, the system is assumed merely to be a stationary linear system in the changes. Consequently, the null is defined over a full set of stationary autoregressive and moving average coefficients as well as Ω. The "augmented" tests described below are designed to be asymptotically similar for this case just as established by Dickey and Fuller for their univariate tests.

The seven test statistics proposed are all calculable by least squares. The critical values are estimated for each of these statistics by simulation using 10,000 replications. Using these critical values, the powers of the test statistics are

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

computed by simulations under various alternatives. A brief motivation of each test is useful.

1. CRDW. After running the co-integrating regression, the Durbin Watson statistic is tested to see if the residuals appear stationary. If they are nonstationary, the Durbin Watson will approach zero and thus the test rejects non-co-integration (finds co-integration) if DW is too big. This was proposed recently by Bhargava (1984) for the case where the series is observed and the null and alternative are first order models.

2. DF. This tests the residuals from the co-integrating regression by running an auxiliary regression as described by Dickey and Fuller and outlined in Table I. It also assumes that the first order model is correct.

3. ADF. The augmented Dickey-Fuller test allows for more dynamics in the DF regression and consequently is over-parametrized in the first order case but correctly specified in the higher order cases.

4. RVAR. The restricted vector autoregression test is similar to the two step estimator. Conditional on the estimate of the co-integrating vector from the co-integrating regression, the error correction representation is estimated. The test is whether the error correction term is significant. This test requires specification of the full system dynamics. In this case a first order system is assumed. By making the system triangular, the disturbances are uncorrelated, and under normality the t statistics are independent. The test is based on the sum of the squared t statistics.

5. ARVAR. The augmented RVAR test is the same as RVAR except that a higher order system is postulated.

6. UVAR. The unrestricted VAR test is based on a vector autoregression in the levels which is not restricted to satisfy the co-integration constraints. Under the null, these are not present anyway so the test is simply whether the levels would appear at all, or whether the model can be adequately expressed entirely in changes. Again by triangularizing the coefficient matrix, the F tests from the two regressions can be made independent and the overall test is the sum of the two F's times their degrees of freedom, 2. This assumes a first order system again.

7. AUVAR. This is an augmented or higher order version of the above test.

To establish the similarity of these tests for the first order case for all positive definite symmetric matrices Ω, it is sufficient to show that the residuals from the regression of y on x for general Ω will be a scalar multiple of the residuals for $\Omega = I$. To show this, let ε_{1t} and ε_{2t} be drawn as independent standard normals. Then

$$\begin{aligned} y_t &= \sum_{i=1,t} \varepsilon_{1i}, \\ x_t &= \sum_{i=1,t} \varepsilon_{2i}, \end{aligned} \tag{5.2}$$

and

$$u_t = y_t - x_t \sum x_t y_t / \sum x_t^2. \tag{5.3}$$

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

To generate y^* and x^* from Ω, let

$$\begin{aligned} \varepsilon_{2t}^* &= c\varepsilon_{2t}, \\ \varepsilon_{1t}^* &= a\varepsilon_{2t} + b\varepsilon_{1t}, \end{aligned} \tag{5.4}$$

where

$$c = \sqrt{\omega}_{xx}, \quad a = \omega_{yx}/c, \quad b^2 = \omega_{yy} - \omega_{yx}^2/\omega_{xx}.$$

Then substituting (5.4) in (5.2)

$$\begin{aligned} x^* &= cx, \quad y^* = ay + bx, \\ u^* &= y^* - x^* \sum y_t^* x_t^* / \sum x_t^{*2} \\ &= ay + bx - cx \sum (ay_t + bx_t) cx_t / \sum c^2 x_t^2 \\ &= au, \end{aligned}$$

thus showing the exact similarity of the tests. If the same random numbers are used, the same test statistics will be obtained regardless of Ω.

In the more complicated but realistic case that the system is of infinite order but can be approximated by a p order autoregression, the statistics will only be asymptotically similar. Although exact similarity is achieved in the Gaussian fixed regressor model, this is not possible in time series models where one cannot condition on the regressors; similarity results are only asymptotic. Tests 5 and 7 are therefore asymptotically similar if the p order model is true but tests 1, 2, 4, and 6 definitely are not even asymptotically similar as these tests omit the lagged regressors. (This is analogous to the biased standard errors resulting from serially correlated errors.) It is on this basis that we prefer not to suggest the latter tests except in the first order case. Test 3 will also be asymptotically similar under the assumption that u, the residual from the co-integration regression, follows a p order process. This result is proven in Dickey and Fuller (1981, pp. 1065-1066). While the assumption that the system is p order allows the residuals to be of infinite order, there is presumably a finite autoregressive model, possibly of order less than p, which will be a good approximation. One might therefore suggest some experimentation to find the appropriate value of p in either case. An alternative strategy would be to let p be a slowly increasing nonstochastic function of T, which is closely related to the test proposed by Phillips (1985) and Phillips and Durlauf (1985). Only substantial simulation experimentation will determine whether it is preferable to use a data based selection of p for this testing procedure although the evidence presented below shows that estimation of extraneous parameters will decrease the power of the tests.

In Table I, the seven test statistics are formally stated. In Table II, the critical values and powers of the tests are considered when the system is first order. Here the augmented tests would be expected to be less powerful because they estimate parameters which are truly zero under both the null and alternative. The other four tests estimate no extraneous parameters and are correctly specified for this experiment.

From Table II one can perform a 5 per cent test of the hypothesis of non-co-integration with the co-integrating regression Durbin Watson test, by simply

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

TABLE I

THE TEST STATISTICS: REJECT FOR LARGE VALUES

1. The Co-integrating Regression Durbin Watson: $y_t = \alpha x_t + c + u_t$

   $\xi_1 = DW$. The null is $DW = 0$.

2. Dickey Fuller Regression: $\Delta u_t = -\phi u_{t-1} + \varepsilon_t$.

   $\xi_2 = \tau_\phi$: the t statistic for ϕ.

3. Augmented DF Regression: $\Delta u_t = -\phi u_{t-1} + b_1 \Delta u_{t-1} + \cdots + b_t \Delta u_t - p + \varepsilon_t$.

   $\xi_3 = \tau_\phi$.

4. Restricted VAR: $\Delta y_t = \beta_1 u_{t-1} + \varepsilon_{1t}$, $\Delta x_t = \beta_2 u_{t-1} + \gamma \Delta y_t + \varepsilon_{2t}$.

   $\xi_4 = \tau_{\beta 1}^2 + \tau_{\beta 2}^2$.

5. Augmented Restricted VAR: Same as (4) but with p lags of Δy_t and Δx_t in each equation.

   $\xi_5 = \tau_{\beta 1}^2 + \tau_{\beta 2}^2$.

6. Unrestricted VAR: $\Delta y_t = \beta_1 y_{t-1} + \beta_2 x_{t-1} + c_1 + \varepsilon_{1t}$, $\Delta x_t = \beta_3 y_{t-1} + \beta_4 x_{t-1} + \gamma \Delta y_t + c_2 + \varepsilon 2_t$.

   $\xi_6 = 2[F_1 + F_2]$ where F_1 is the F statistic for testing β_1 and β_2 both equal to zero in the first equation, and F_2 is the comparable statistic in the second.

7. Augmented Unrestricted VAR: The same as (6) except for p lags of Δx_t and Δy_t in each equation.

   $\gamma_7 = 2[F_1 + F_2]$.

NOTES: y_t and x_t are the original data sets and u_t are the residuals from the co-integrating regression.

checking DW from this regression and, if it exceeds 0.386, rejecting the null and finding co-integration. If the true model is Model II with $\rho = .9$ rather than 1, this will only be detected 20 per cent of the time; however if the true $\rho = .8$ this rises to 66 per cent. Clearly, test 1 is the best in each of the power calculations and should be preferred for this set-up, while test 2 is second in almost every case. Notice also that the augmented tests have practically the same critical values as the basic tests; however, as expected, they have slightly lower power. Therefore, if it is known that the system is first order, the extra lags should not be introduced. Whether a pre-test of the order would be useful remains to be established.

In Table III both the null and alternative hypotheses have fourth order autoregressions. Therefore the basic unaugmented tests now are misspecified while the augmented ones are correctly specified (although some of the intervening lags could be set to zero if this were known). Notice now the drop in the critical values of tests 1, 4, and 6 caused by their nonsimilarity. Using these new critical values, test 3 is the most powerful for the local alternative while at $\rho = .8$, test 1 is the best closely followed by 2 and 3. The misspecified or unaugmented tests 4 and 6 perform very badly in this situation. Even though they were moderately powerful in Table II, the performance here dismisses them from consideration.

Although test 1 has the best performance overall, it is not the recommended choice from this experiment because the critical value is so sensitive to the particular parameters within the null. For most types of economic data the differences are not white noise and, therefore, one could not in practice know

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

TABLE II

CRITICAL VALUES AND POWER

I MODEL: Δy, Δx independent standard normal, 100 observations, 10,000 replications, $p = 4$.

Statistic	Name	Critical Values 1%	5%	10%
1	CRDW	.511	.386	.322
2	DF	4.07	3.37	3.03
3	ADF	3.77	3.17	2.84
4	RVAR	18.3	13.6	11.0
5	ARVAR	15.8	11.8	9.7
6	UVAR	23.4	18.6	16.0
7	AUVAR	22.6	17.9	15.5

II MODEL: $y_t + 2x_t = u_t$, $\Delta u_t = (\rho - 1)u_{t-1} + \varepsilon_t$, $x_t + y_t = v_t$, $\Delta v_t = \eta_t$; $\rho = .8, .9, 100$ observations, 1000 replications, $p = 4$.

Statistic	Name	Rejections per 100: $\rho = .9$ 1%	5%	10%
1	CRDW	4.8	19.9	33.6
2	DF	2.2	15.4	29.0
3	ADF	1.5	11.0	22.7
4	RVAR	2.3	11.4	25.3
5	ARVAR	1.0	9.2	17.9
6	UVAR	4.3	13.3	26.1
7	AUVAR	1.6	8.3	16.3
		Rejections per 100: $\rho = .8$		
Statistic	Name	1%	5%	10%
1	CRDW	34.0	66.4	82.1
2	DF	20.5	59.2	76.1
3	ADF	7.8	30.9	51.6
4	RVAR	15.8	46.2	67.4
5	ARVAR	4.6	22.4	39.0
6	UVAR	19.0	45.9	63.7
7	AUVAR	4.8	18.3	33.4

what critical value to use. Test 3, the augmented Dickey-Fuller test, has essentially the same critical value for both finite sample experiments, has theoretically the same large sample critical value for both cases, and has nearly as good observed power properties in most comparisons, and is therefore the recommended approach.

Because of its simplicity, the CRDW might be used for a quick approximate result. Fortunately, none of the best procedures require the estimation of the full system, merely the co-integrating regression and then perhaps an auxiliary time series regression.

This analysis leaves many questions unanswered. The critical values have only been constructed for one sample size and only for the bivariate case, although recently, Engle and Yoo (1986) have calculated critical values for more variables

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

TABLE III

CRITICAL VALUES AND POWER WITH LAGS

MODEL I: $\Delta y_t = .8\Delta y_{t-4} + \varepsilon_t$, $\Delta x_t = .8\Delta x_{t-4} + \eta_t$; 100 observations, 10,000 replications, $p = 4$, ε_t, η_t independent standard normal.

		Critical Values		
Statistic	Name	1%	5%	10%
1	CRDW	.455	.282	.209
2	DF	3.90	3.05	2.71
3	ADF	3.73	3.17	2.91
4	RVAR	37.2	22.4	17.2
5	ARVAR	16.2	12.3	10.5
6	UVAR	59.0	40.3	31.4
7	AUVAR	28.0	22.0	19.2

MODEL II: $y_t + 2x_t = u_t$, $\Delta u_t = (\rho - 1)u_{t-1} + .8\Delta u_{t-4} + \varepsilon_t$, $y_t + x_t = v_t$, $\Delta v_t = .8\Delta v_{t-4} + \eta_t$; $\rho = .9, .8$. 100 observations, 1000 replications, $p = 4$.

		Rejections per 100: $\rho = .9$		
Statistic	Name	1%	5%	10%
1	CRDW	15.6	39.9	65.6
2	DF	9.4	25.5	37.8
3	ADF	36.0	61.2	72.2
4	RVAR	.3	4.4	10.9
5	ARVAR	26.4	48.5	62.8
6	UVAR	.0	.5	3.5
7	AUVAR	9.4	26.8	40.3
		Rejections per 100: $\rho = .8$		
Statistic	Name	1%	5%	10%
1	CRDW	77.5	96.4	98.6
2	DF	66.8	89.7	96.0
3	ADF	68.9	90.3	94.4
4	RVAR	7.0	42.4	62.5
5	ARVAR	57.2	80.5	89.3
6	UVAR	2.5	10.8	25.9
7	AUVAR	32.2	53.0	67.7

and sample sizes using the same general approach. There is still no optimality theory for such tests and alternative approaches may prove superior. Research on the limiting distribution theory by Phillips (1985) and Phillips and Durlauf (1985) may lead to improvements in test performance.

Nevertheless, it appears that the critical values for ADF given in Table II can be used as a rough guide in applied studies at this point. The next section will provide a variety of illustrations.

6. EXAMPLES

Several empirical examples will be presented to show performance of the tests in practice. The relationship between consumption and income will be studied

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

in some detail as it was analyzed from an error correction point of view in DHSY and a time series viewpoint in Hall (1978) and others. Briefer analyses of wages and prices, short and long term interest rates, and the velocity of money will conclude this section.

DHSY have presented evidence for the error correction model of consumption behavior from both empirical and theoretical points of view. Consumers make plans which may be frustrated; they adjust next period's plans to recoup a portion of the error between income and consumption. Hall finds that U.S. consumption is a random walk and that past values of income have no explanatory power which implies that income and consumption are not co-integrated, at least if income does not depend on the error correction term. Neither of these studies models income itself and it is taken as exogenous in DHSY.

Using U.S. quarterly real per capita consumption on nondurables and real per capita disposable income from 1947-I to 1981-II, it was first checked that the series were $I(1)$. Regressing the change in consumption on its past level and two past changes gave a t statistic of $+.77$ which is even the wrong sign for consumption to be stationary in the levels. Running the same model with second differences on lagged first differences and two lags of second differences, the t statistic was -5.36 indicating that the first difference is stationary. For income, four past lags were used and the two t statistics were $-.01$ and -6.27 respectively, again establishing that income is $I(1)$.

The co-integrating regression of consumption (C) on income (Y) and a constant was run. The coefficient of Y was .23 (with a t statistic of 123 and an R^2 of .99). The DW was however .465 indicating that by either table of critical values one rejects the null of "non-co-integration" or accepts co-integration at least at the 5 per cent level. Regressing the change in the residuals on past levels and four lagged changes, the t statistic on the level is 3.1 which is essentially the critical value for the 5 per cent ADF test. Because the lags are not significant, the DF regression was run giving a test statistic of 4.3 which is significant at the 1 per cent level, illustrating that when it is appropriate, it is a more powerful test. In the reverse regression of Y on C, the coefficient is 4.3 which has reciprocal .23, the same as the coefficient in the forward regression. The DW is now .463 and the t statistic from the ADF test is 3.2. Again the first order DF appears appropriate and gives a test statistic of 4.4. Whichever way the regression is run, the data rejects the null of non-co-integration at any level above 5 per cent.

To establish that the joint distribution of C and Y is an error correction system, a series of models was estimated. An unrestricted vector autoregression of the change in consumption on four lags of consumption and income changes plus the lagged levels of consumption and income is given next in Table IV. The lagged levels are of the appropriate signs and sizes for an error correction term and are individually significant or nearly so. Of all the lagged changes, only the first lag of income change is significant. Thus the final model has the error correction term estimated from the co-integrating regression and one lagged change in income. The standard error of this model is even lower than the VAR suggesting the efficiency of the parameter restrictions. The final model passes a

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

TABLE IV

REGRESSIONS OF CONSUMPTION AND INCOME

Dep. Var.:	C	ΔEC	ΔEC	ΔC	ΔC
Y	.23 (123)				
$C(-1)$				−.19 (−2.5)	
$Y(-1)$				.046 (2.5)	
$EC(-1)$		−.22 (−3.1)	−.26 (−4.3)		−.14 (−2.2)
$\Delta C(-1)$				.092 (0.9)	
$\Delta C(-2)$				.017 (0.2)	
$\Delta C(-3)$				.16 (1.5)	
$\Delta C(-4)$				.009 (0.1)	
$\Delta Y(-1)$				.059 (1.8)	.068 (2.5)
$\Delta Y(-2)$				−.023 (−.7)	
$\Delta Y(-3)$				−.027 (−.8)	
$\Delta Y(-4)$				−.020 (−.7)	
$\Delta EC(-1)$		−.13 (−1.4)			
$\Delta EC(-2)$		.12 (1.4)			
$\Delta EC(-3)$		.03 (0.4)			
$\Delta EC(-4)$		−.13 (−1.6)			
CONST	.52 (85)			.10 (2.4)	.003 (2.6)
σ	.01628	.00999	0.01015	.01094	.01078
DW	.46	2.0	2.2	2.0	1.9
Dep. Var:	Y	ΔEY	ΔEY	ΔY	ΔY
C	4.29 (123)				
$C(-1)$				.15 (.67)	
$Y(-1)$				−.034 (.63)	
$EY(-1)$		−.23 (−3.2)	−.26 (−4.4)		−.053 (−1.1)
$\Delta C(-1)$				.79 (2.5)	.66 (2.4)
$\Delta C(-2)$				−.48 (−1.5)	
$\Delta C(-3)$				.68 (2.2)	
$\Delta C(-4)$				.56 (1.8)	.60 (2.1)
$\Delta Y(-1)$				−.027 (−.3)	
$\Delta Y(-2)$				−.051 (−.5)	
$\Delta Y(-3)$				.011 (.1)	
$\Delta Y(-4)$				−.23 (−2.5)	−.19 (2.1)
$\Delta EY(-1)$		−.13 (−1.5)			
$\Delta EY(-2)$		.12 (1.4)			
$\Delta EY(-3)$		.03 (0.4)			
$\Delta EY(-4)$		−.14 (−1.6)			
CONST	2.22 (−50)			−.071 (−.6)	.016 (4.6)
σ	.07012	.04279	.04350	.03255	.03321
DW	.46	2.0	2.2	2.1	2.2

NOTES: Data are from 1947-I to 1981-II. EC are the residuals from the first regression and EY are the residuals from the sixth regression. T ratios are in parentheses.

series of diagnostic tests for serial correlation, lagged dependent variables, nonlinearities, ARCH, and omitted variables such as a time trend and other lags.

One might notice that an easy model building strategy in this case would be to estimate the simplest error correction model first and then test for added lags of C and Y, proceeding in a "simple to general" specification search.

The model building process for Y produced a similar model. The same unrestricted VAR was estimated and distilled to a simple model with the error

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

correction term, first and fourth lagged changes in C and a fourth lagged change in Y. The error correction is not really significant with a t statistic of -1.1 suggesting that income may indeed be weakly exogenous even though the variables are co-integrated. In this case the standard error of the regression is slightly higher in the restricted model but the difference is not significant. The diagnostic tests are again generally good.

Campbell (1985) uses a similar structure to develop a test of the permanent income hypothesis which incorporates "saving for a rainy day" behavior. In this case the error correction term is approximately saving which should be high when income is expected to fall (such as when current income is above permanent income). Using a broader measure of consumption and narrower measure of income he finds the error correction term significant in the income equation.

The second example examines monthly wages and prices in the U.S. The data are logs of the consumer price index and production worker wage in manufacturing over the three decades of 50's, 60's and 70's. Again, the test is run both directions to show that there is little difference in the result. For each of the decades there are 120 observations so the critical values as tabulated should be appropriate.

For the full sample period the Durbin Watson from the co-integrating regression in either direction is a notable .0054. One suspects that this will be insignificantly different from zero even for samples much larger than this. Looking at the augmented Dickey Fuller test statistic, for p on w we find $-.6$ and for w on p we find $+.2$. Adding a twelfth lag in the ADF tests improves the fit substantially and raises the test statistics to .88 and 1.50 respectively. In neither case do these approach the critical values of 3.2. The evidence accepts the null of non-co-integration for wages and prices over the thirty year period.

For individual decades none of the ADF tests are significant at even the 10 per cent level. The largest of these six test statistics is for the 50's regressing p on w which reaches 2.4, which is still below the 10 per cent level of 2.8. Thus we find evidence that wages and prices in the U.S. are not co-integrated. Of course, if a third variable such as productivity were available (and were $I(1)$), the three might be co-integrated.

The next example tests for co-integration between short and long term interest rates. Using monthly yields to maturity of 20 year treasury bonds as the long term rate (R_t) and the one month treasury bill rate r_t as the short rate, co-integration was tested with data from February, 1952 to December, 1982. With the long rate as the dependent variable, the co-integrating regression gave:

$$R_t = 1.93 + .785\, r_t + ER_t, \quad DW = .126, \quad R^2 = .866,$$

with a t ratio of 46 on the short rate. The DW is not significantly different from zero, at least by Tables II and III; however, the correct critical value depends upon the dynamics of the errors (and of course the sample size is 340—much greater than for the tabulated values). The ADF test with four lags gives:

$$\Delta ER_t = \underset{(-3.27)}{-.06}\ ER_{t-1}$$

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

$$+ \underset{(4.55)}{.25}\ \Delta ER_{t-1} - \underset{(-4.15)}{.24}\ \Delta ER_{t-2} + \underset{(-4.15)}{.24}\ \Delta ER_{t-3} - \underset{(-1.48)}{.09}\ \Delta ER_{t-4}.$$

When the twelfth lag is added instead of the fourth, the test statistic rises to 3.49. Similar results were found with the reverse regression where the statistics were 3.61 and 3.89 respectively. Each of these test statistics exceeds the 5 per cent critical values from Table III. Thus these interest rates are apparently co-integrated.

This finding is entirely consistent with the efficient market hypothesis. The one-period excess holding yield on long bonds as linearized by Shiller and Campbell (1984) is:

$$EHY = DR_{t-1} - (D-1)R_t - r_t$$

where D is the duration of the bond which is given by

$$D = ((1+c)^i - 1)/(c(1+c)^{i-1}$$

with c as the coupon rate and i the number of periods to maturity. The efficient market hypothesis implies that the expectation of the EHY is a constant representing a risk premium if agents are risk averse. Setting $EHY = k + \varepsilon$ and rearranging terms gives the error correction form:

$$\Delta R_t = (D-1)^{-1}(R_{t-1} - r_{t-1}) + k' + \varepsilon_t,$$

implying that R and r are co-integrated with a unit coefficient and that for long maturities, the coefficients of the error correction term is c, the coupon rate. If the risk premium is varying over time but is $I(0)$ already, then it need not be included in the test of co-integration.

The final example is based upon the quantity theory equation: $MV = PY$. Empirical implications stem from the assumption that velocity is constant or at least stationary. Under this condition, log M, log P, and log Y should be co-integrated with known unit parameters. Similarly, nominal money and nominal GNP should be co-integrated. A test of this hypothesis was constructed for four measures of money: $M1$, $M2$, and $M3$, and L, total liquid assets. In each case the sample period was 1959-I through 1981-II, quarterly. The ADF tests statistics were:

$M1$	1.81	1.90
$M2$	3.23	3.13
$M3$	2.65	2.55
L	2.15	2.13

where in the first column the log of the monetary aggregate was the dependent variable while in the second, it was log GNP. For only one of the $M2$ tests is the test statistic significant at the 5 per cent level, and none of the other aggregates are significant even at the 10 per cent level. (In several cases it appears that the DF test could be used and would therefore be more powerful.) Thus the most stable relationship is between $M2$ and nominal GNP but for the other aggregates, we reject co-integration and the stationarity of velocity.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

7. CONCLUSION

If each element of a vector of time series x_t is stationary only after differencing, but a linear combination $\alpha' x_t$ need not be differenced, the time series x_t have been defined to be co-integrated of order (1, 1) with co-integrating vector α. Interpreting $\alpha' x_t = 0$ as a long run equilibrium, co-integration implies that equilibrium holds except for a stationary, finite variance disturbance even though the series themselves are non-stationary and have infinite variance.

The paper presents several representations for co-integrated systems including an autoregressive representation and an error-correction representation. A vector autoregression in differenced variables is incompatible with these representations because it omits the error correction term. The vector autoregression in the levels of the series ignores cross equation constraints and will give a singular autoregressive operator. Consistent and efficient estimation of error correction models is discussed and a two step estimator proposed. To test for co-integration, seven statistics are formulated which are similar under various maintained hypotheses about the generating model. The critical values of these statistics are calculated based on a Monte Carlo simulation. Using these critical values, the power properties of the tests are examined, and one test procedure is recommended for application.

In a series of examples it is found that consumption and income are co-integrated, wages and prices are not, short and long interest rates are, and nominal *GNP* is not co-integrated with $M1$, $M3$, or total liquid assets, although it is possibly with $M2$.

Department of Economics, University of California—San Diego, La Jolla, CA 92093, *U.S.A.*

Manuscript received September, 1983; final revision received June, 1986.

REFERENCES

BHARGAVA, ALOK (1984): "On the Theory of Testing For Unit Roots in Observed Time Series," manuscript, ICERD, London School of Economics.

BOX, G. E. P., AND G. M. JENKINS (1970): *Time Series Analysis, Forecasting and Control.* San Francisco: Holden Day.

CAMPBELL, JOHN Y. (1985): "Does Saving Anticipate Declining Labor Income? An Alternative Test of the Permanent Income Hypothesis," manuscript, Princeton University.

COX, D. R., AND C. V. HINKLEY (1974): *Theoretical Statistics.* London: Chapman and Hall.

CURRIE, D. (1981): "Some Long-Run Features of Dynamic Time-Series Models," *The Economic Journal,* 91, 704-715.

DAVIDSON, J. E. H., DAVID F. HENDRY, FRANK SRBA, AND STEVEN YEO (1978): "Econometric Modelling of the Aggregate Time-series Relationship Between Consumer's Expenditure and Income in the United Kingdom," *Economic Journal,* 88, 661-692.

DAVIES, R. R. (1977): "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika,* 64, 247-254.

DAWSON, A. (1981): "Sargan's Wage Equation: A Theoretical and Empirical Reconstruction," *Applied Economics,* 13, 351-363.

DICKEY, DAVID A. (1976): "Estimation and Hypothesis Testing for Nonstationary Time Series," PhD. Thesis, Iowa State University, Ames.

DICKEY, DAVID A, AND WAYNE A. FULLER (1979): "Distribution of the Estimators for Autoregressive Time Series With a Unit Root," *Journal of the American Statistical Assoc.,* 74, 427-431.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms

——— (1981): "The Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," *Econometrica*, 49, 1057-1072.

ENGLE, ROBERT F., DAVID F. HENDRY, AND J. F. RICHARD (1983): "Exogeneity," *Econometrica*, 51, 277-304.

ENGLE, ROBERT F., AND BYUNG SAM YOO (1986): "Forecasting and Testing in Co-integrated Systems," U.C.S.D. Discussion Paper.

EVANS, G. B. A., AND N. E. SAVIN (1981): "Testing for Unit Roots: 1," *Econometrica*, 49, 753-779.

FELLER, WILLIAM (1968): *An Introduction to Probability Theory and Its Applications, Volume I.* New York: John Wiley.

FULLER, WAYNE A. (1976): *Introduction to Statistical Time Series.* New York: John Wiley.

GRANGER, C. W. J. (1981): "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 121-130.

——— (1983): "Co-Integrated Variables and Error-Correcting Models," unpublished UCSD Discussion Paper 83-13.

GRANGER, C. W. J., AND P. NEWBOLD (1977): *Forecasting Economic Time Series.* New York: Academic Press.

——— (1974): "Spurious Regressions in Econometrics," *Journal of Econometrics*, 26, 1045-1066.

GRANGER, C. W. J., AND A. A. WEISS (1983): "Time Series Analysis of Error-Correcting Models," in *Studies in Econometrics, Time Series, and Multivariate Statistics.* New York: Academic Press, 255-278.

HALL, ROBERT E. (1978): "A Stochastic Life Cycle Model of Aggregate Consumption," *Journal of Political Economy*, 971-987.

HANNAN, E. J. (1970): *Multiple Time Series.* New York: Wiley.

HENDRY, DAVID F., AND T. VON UNGERN-STERNBERG (1981): "Liquidity and Inflation Effects on Consumer's Expenditure," in *Essays in the Theory and Measurement of Consumer's Behavior*, ed. by A. S. Deaton. Cambridge: Cambridge University Press.

JOHANSEN, SOREN (1985): "The Mathematical Structure of Error Correction Models," manuscript, University of Copenhagen.

NELSON, C. R., AND CHARLES PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139-162.

PAGAN, A. R. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221-248.

PHILLIPS, A. W. (1957): "Stabilization Policy and the Time Forms of Lagged Responses," *Economic Journal*, 67, 265-277.

PHILLIPS, P. C. B. (1985): "Time Series Regression with Unit Roots," Cowles Foundation Discussion Paper No. 740, Yale University.

PHILLIPS, P. C. B., AND S. N. DURLAUF (1985): "Multiple Time Series Regression with Integrated Processes," Cowles Foundation Discussion Paper 768.

SALMON, M. (1982): "Error Correction Mechanisms," *The Economic Journal*, 92, 615-629.

SARGAN, J. D. (1964): "Wages and Prices in the United Kingdom: a Study in Econometric Methodology," in *Econometric Analysis for National Economic Planning*, ed. by P. E. Hart, G. Mills, and J. N. Whittaker. London: Butterworths.

SARGAN, J. D., AND A. BHARGAVA (1983): "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153-174.

SHILLER, R. J., AND J. Y. CAMPBELL (1984): "A Simple Account of the Behaviour of Long-Term Interest Rates," *American Economic Review*, 74, 44-48.

STOCK, JAMES H. (1984): "Asymptotic Properties of Least Squares Estimators of Co-Integrating Vectors," manuscript, Harvard University.

WATSON, MARK W., AND ROBERT ENGLE (1985): "A Test for Regression Coefficient Stability with a Stationary AR(1) Alternative," forthcoming in *Review of Economics and Statistics.*

YOO, SAM (1985): "Multi-co-integrated Time Series and Generalized Error Correction Models," manuscript in preparation, U.C.S.D.

This content downloaded from 71.126.161.212 on Wed, 25 Mar 2020 12:31:54 UTC
All use subject to https://about.jstor.org/terms