NBER TECHNICAL WORKING PAPER SERIES

A SIMPLE MLE OF COINTEGRATING VECTORS IN HIGHER ORDER INTEGRATED SYSTEMS

James H. Stock

Mark W. Watson

Technical Working Paper No. 83

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 1989

The authors thank Rob Engle, Danny Quah, Ken West, and the participants in the NBER/FMME Summer Institute Workshop on New Econometric Methods in Financial Time Series, July 17-20, 1989 for helpful comments on an early version of this paper. Stock thanks the Sloan Foundation for financial support. This research was supported in part by National Science Foundation grants no. SES-86-18984 and SES-89-10601. This paper is part of NBER's research programs in Economic Fluctuations and Financial Markets and Monetary Economics.

NBER Technical Working Paper #83
December 1989

A SIMPLE MLE OF COINTEGRATING VECTORS IN HIGHER ORDER INTEGRATED SYSTEMS

ABSTRACTS

An MLE of the unknown parameters of cointegrating vectors is presented for systems in which some variables exhibit higher orders of integration, in which there might be deterministic components, and in which the cointegrating vector itself might involve variables of differing orders of integration. The estimator is simple to compute: it can be calculated by running GLS for standard regression equations with serially correlated errors. Alternatively, an asymptotically equivalent estimator can be computed using OLS. Usual Wald test statistics based on these MLE's (constructed using an autocorrelation-robust covariance matrix in the case of the OLS estimator) have asymptotic χ^2 distributions.

James H. Stock
Kennedy School of Government
Harvard University
Cambridge, MA 02138

Mark W. Watson
Department of Economics
Northwestern University
Evanston, IL 60208

1. Introduction

Let y_t, $t=1,\ldots,T$ be a n-dimensional cointegrated stochastic process. In theory, the Gaussian maximum likelihood estimator (MLE) of the unknown coefficients of the cointegrating vectors can be found by parameterizing the covariance matrix of $(y_1,\ldots,y_T)$ and directly computing the Gaussian likelihood. In practice, however, this entails inverting the nT×nT covariance matrix and so is computationally impractical. This has led various researchers to compute the MLE using factorizations of the likelihood that reduce computational demands, typically to the level of nonlinear simultaneous equations regression. Research so far has focused on the case that each element of y_t individually is integrated of order 1 (is I(1)), typically with no drift term. Johansen (1988) and Ahn and Reinsel (1987) independently derived the asymptotic distribution of the MLE when the cointegrated system is parameterized as a vector error correction model, and Johansen (1989) extended this result to the case of nonzero drifts. Phillips (1988a) derived asymptotic representations for MLE's in a cointegrated ARMA model. Phillips and Hansen (1989) considered a two-step zero frequency seemingly unrelated regression estimator, and Phillips (1988b) used spectral methods to factor the likelihood and to compute the MLE in the frequency domain.

This paper adopts an alternative factorization of the likelihood that permits the derivation of a computationally simple MLE that readily extends to systems with deterministic components and with higher orders of integration and cointegration. The empirical problem motivating this research is the analysis of a standard four variable macroeconomic system involving the logarithms of money, prices, real output and the level of interest rates (respectively m, p, Q, and r). For the postwar U.S., a plausible empirical description of these series is that money and prices are doubly integrated (I(2)) processes with no drift in inflation or money growth, output is I(1) around a linear trend, and the interest rate is I(1) with no drift (see for example King, Plosser, Stock and Watson [1987] or Hoffman and Rasche [1989]). There are two possible cointegrating relations among the data: first, real balances, m-p -- or perhaps $m-\theta_p p$ -- are possibly I(1), and second, there could be a stable money demand relation for which $m-\theta_p p-\theta_Q Q-\theta_r r$ would be I(0). This system involves variables that are integrated of different orders, have different deterministic components, and are related by a system of cointegrating vectors.

The factorization of the likelihood is discussed for I(1) variates in Section 2 and for I(d) variates in Section 3. The properties of the MLE's and test statistics are examined in Section 4. Section 5 presents a brief example and Section 6 summarizes the results of a Monte Carlo experiment. Section 7 concludes.

2. Gaussian Estimation: The I(1) Case

Suppose that each element of y_t is I(1), that $E\Delta y_t=0$, and that the $n\times r$ matrix of r cointegrating vectors α is $\alpha = (-\theta \quad I_r)'$, where θ is a $r\times(n-r)$ submatrix of unknown parameters. The task is to obtain the Gaussian MLE of θ. Our starting point is the triangular representation, (2.1a) $$\Delta y_t^1 = u_t^1$$

(2.1b) $$y_t^2 = \theta y_t^1 + u_t^2$$

where y_t is partitioned as (y_t^1, y_t^2), where y_t^1 is $(n-r)\times 1$ and y_t^2 is $r\times 1$ and where $u_t = (u_t^{1\prime}\ u_t^{2\prime})'$ is a stationary stochastic process with full rank spectral density matrix. The key feature in (2.1) is that the levels of y_t appear in only the final r equations. Bewley (1979) derived a representation with this feature for an error correction system under the assumption that y_t^1 is strictly exogenous in (2.1b) and not necessarily I(1). Hylleberg and Mizon (1989) assumed that y_t^1 is I(1) and generalized Bewley's formulation to the case where y_t^1 is not strictly exogenous in (2.1b), a generalization which they termed the Bewley representation. Campbell and Shiller (1987,1989) and Campbell (1987) used the form (2.1) in applications where they parameterized u_t as an unrestricted VAR. This representation has been used extensively by Phillips (1988a,1988b), typically without parametric structure on the I(0) process u_t.

For the development of the MLE it is assumed that u_t is Gaussian. In general, u_t^1 and u_t^2 will be cross-correlated at leads and lags; only when this cross-correlation is zero is the GLS estimator of θ in (2.1b) the MLE. The factorization adopted here addresses this cross-correlation by making the disturbances in the y_t^2 equation independent of the entire sequence $\{y_t^1\}$. Let $\tilde{u}_t^2 = u_t^2 - E[u_t^2|\{u_t^1\}] = u_t^2 - d(L)u_t^1 = u_t^2 - d(L)\Delta y_t^1$, where d(L) is in general two-sided and where the conditional expectation is linear in $\{u_t^1\}$ because u_t is Gaussian. By construction $\tilde{u}_t^2$ is independent of

$\{\Delta y_t^1\}$. Because u_t^1 and $\bar{u}_t^2$ are stationary and Gaussian with finite second moments, they have one-sided Wold representations $u_t^1 = c_{11}(L)\epsilon_t^1$ and $\bar{u}_t^2 = c_{22}(L)\epsilon_t^2$, where ϵ_t^1 and ϵ_t^2 are independent and normalized so that $E\epsilon_t\epsilon_t' = I$ where $\epsilon_t = (\epsilon_t^{1\prime} \ \ \epsilon_t^{2\prime})'$. Thus

$$u_t^1 = c_{11}(L)\epsilon_t^1 \tag{2.2}$$
$$u_t^2 = c_{21}(L)\epsilon_t^1 + c_{22}(L)\epsilon_t^2$$

where $c_{11}(L)$ and $c_{22}(L)$ are one-sided and in general $c_{21}(L) = d(L)c_{11}(L)^{-1}$ is two-sided. Thus (2.1) can be written,

$$\Delta y_t^1 = c_{11}(L)\epsilon_t^1 \tag{2.3a}$$

$$y_t^2 = \theta y_t^1 + d(L)\Delta y_t^1 + c_{22}(L)\epsilon_t^2 \tag{2.3b}$$

where ϵ_t is NIID(0,I) and $\{\epsilon_t^2\}$ is independent of $\{y_t^1\}$.

The representation (2.3) leads to a factorization of the Gaussian likelihood that differs from the usual prediction error factorization in an important way. Let λ_1 denote the parameters of $c_{11}(L)$, let λ_2 denote the parameters of $d(L)$ and $c_{22}(L)$, and let Y^i denote $(y_1^i, \ldots, y_T^i)$, $i=1,2$. Then (2.3) implies that the likelihood can be factored as

$$f(Y^1, Y^2 | \theta, \lambda_1, \lambda_2) = f(Y^2 | Y^1, \theta, \lambda_2) f(Y^1 | \lambda_1) \ . \tag{2.4}$$

If the mapping from the original parameters to $(\lambda_1, [\theta, \lambda_2])$ is variation free, the factorization (2.4) is a sequential cut (see Engle, Hendry and Richard [1983]); that is, if there are no cross-restrictions between λ_1 and (θ, λ_2), Y^1 is weakly exogenous for (θ, λ_2). (This is a slight modification of Engle, Hendry, and Richard's definition of weak exogeneity since they write the likelihood in predicition error form, i.e. conditional on past data.) Thus maximum likelihood estimation of θ can be implemented by maximizing $f(Y^2|Y^1,\theta,\lambda_2)$, which reduces to estimating the parameters of the regression equation (2.3b) by GLS. In fact, because the regressor y_t^1 is I(1), an asymptotically equivalent estimator of θ can be obtained by estimating (2.3b) by OLS, a result discussed in Section 4.

This representation provides considerable insight into the large-sample properties of the GLS estimator of θ in (2.3b). Because $\{\epsilon_t^2\}$ is independent of the regressors, conditional on Y^1 the GLS estimator has a normal distribution and the Wald test statistic has a χ^2 distribution. Because y_t^1 is I(1), the conditional covariance matrix of the GLS estimator will differ across realizations of Y^1, even in large samples; thus unconditionally the GLS estimator of θ will have a large-sample distribution that is a random mixture of normals. Phillips (1988a) provides an insightful discussion of the intuition behind the property that the MLE of θ has a locally mixed asymptotically normal distribution.

We close this section by noting that, although the two-sided triangular representation (2.2) was developed for a Gaussian time series, it applies to general stationary stochastic processes with finite second moments. This result, which provides an alternative to the (one-sided) Wold representation theorem, is summarized in the following lemma which is proved in the Appendix.

<u>Lemma 2.1</u>. Let $u_t = (u_t^{1\prime}\ u_t^{2\prime})'$ be a n×1 stationary stochastic process with $E(u_{it})^2<\infty$, with full rank spectral density matrix, and with

no deterministic component, where u_t^1 is $(n-r)\times 1$ and u_t^2 is $r\times 1$. Then u_t has the representation (2.2) where $c_{11}(L)$ and $c_{22}(L)$ are one-sided, $c_{21}(L)$ is in general two-sided, $E\epsilon_t=0$, $E\epsilon_t\epsilon_t'=I_n$, and $E\epsilon_t\epsilon_s'=0$ for $t\neq s$. In addition, $C(L)$ is square summable.

3. Representation in I(d) Systems

This section presents an extension of the triangular representation (2.3) to systems in which variables may be integrated and cointegrated of different orders and in which there are arbitrary polynomial time trends. The I(d) generalization of (2.1) is

$$
\begin{aligned}
(3.1)\qquad \Delta^d y_t^1 &= \mu_{1,0} + u_t^1 \\
\Delta^{d-1} y_t^2 &= \mu_{2,0} + \mu_{2,1}t + \theta_{2,1}^{d-1}(\Delta^{d-1}y_t^1) + u_t^2 \\
\Delta^{d-2} y_t^3 &= \mu_{3,0} + \mu_{3,1}t + \mu_{3,2}t^2 + \\
&\qquad \theta_{3,1}^{d-1}(\Delta^{d-1}y_t^1) + \theta_{3,1}^{d-2}(\Delta^{d-2}y_t^1) + \theta_{3,2}^{d-2}(\Delta^{d-2}y_t^2) + u_t^3 \\
&\cdots \\
y_t^{d+1} &= \textstyle\sum_{j=0}^{d} \mu_{d+1,j}t^j + \sum_{j=1}^{d}\sum_{i=j}^{d}\theta_{d+1,j}^{d-i}(\Delta^{d-i}y_t^j) + u_t^{d+1}
\end{aligned}
$$

for $t=1,\ldots,T$, where the y_t^j are $k_j\times 1$ vectors which form a partition of y_t, i.e. $y_t=(y_t^{1\prime}\ y_t^{2\prime}\ \ldots\ y_t^{d+1\prime})'$. The stochastic process $u_t=(u_t^{1\prime}\ u_t^{2\prime}\ \ldots\ u_t^{d+1\prime})'$ is I(0) with a full rank spectral density matrix. Note that not every element of y_t need be I(d) for (3.1) to apply.

This representation partitions y_t into components corresponding to stochastic trends of different orders. Thus, abstracting from the deterministic components, y_t^1 is a $k_1 \times 1$ vector corresponding to the k_1 distinct I(d) stochastic trends in the system, $\Delta^{d-1} y_t^2 - \theta_{2,1}^{d-1}(\Delta^{d-1} y_t^1)$ is a $k_2 \times 1$ vector corresponding the k_2 distinct I(d-1) elements in the system, etc. It is straightforward to generalize the representation (3.1) to include higher order polynomials in t, although in applications many of the coefficients on t^j typically will be zero. A derivation of (3.1) from the Wold representation of an I(d) system with multiple cointegrating vectors and drifts is given in the Appendix.

As in the I(1) case, the likelihood function is parameterized so that the variables appearing on the right hand side of each equation are potentially weakly exogenous for the parameters in that equation. By repeated application of Lemma 2.1, $u_t = (u_t^{1\prime}\ u_t^{2\prime}\ \ldots\ u_t^{d+1\prime})'$ has the representation

$$(3.2) \qquad u_t = C(L)\epsilon_t$$

where $\epsilon_t = (\epsilon_t^{1\prime}\ \epsilon_t^{2\prime}\ \ldots\ \epsilon_t^{d+1\prime})'$, and where C(L) is a block lower triangular matrix partitioned conformably with u_t, with diagonal blocks $c_{ii}(L)$ that are one-sided polynomials in L and with lower off-diagonal blocks $c_{ij}(L)$ that are two-sided polynomials in L. The ℓ'th equation in (3.1) can then be written as:

$$(3.3) \quad \Delta^{d-\ell+1} y_t^\ell = \textstyle\sum_{j=0}^{\ell-1} \bar{\mu}_{\ell,j} t^j + \sum_{j=1}^{\ell-1}\sum_{i=j}^{\ell-1} \theta_{\ell,j}^{d-i}(\Delta^{d-i} y_t^j)$$
$$+ \textstyle\sum_{m=1}^{\ell-1} d_{\ell m}(L)[(\Delta^{d-m+1} y_t^m) - \sum_{j=1}^{m-1}\sum_{i=j}^{m-1} \theta_{m,j}^{d-i}(\Delta^{d-i} y_t^j)] + c_{\ell\ell}(L)\epsilon_t^\ell$$

where the $\bar{\mu}_{\ell,j}$ are functions of $\mu_{m,j}$ for $m=1,\ldots,\ell$. Generalizing d(L) in

(2.3), $\{d_{\ell m}(L)\}$ arises from the projection of u_t^ℓ onto $\{u_t^m\}$ for $m=1,\ldots,\ell-1$. When $\{y_t\}$ is Gaussian, ϵ_t is NIID$(0,I_n)$.

Equation (3.3) describes the cointegrating relation between the y_t^ℓ components in the system and higher order integrated components. The set of cointegrating vectors characterizing this relation are given by the matrices $\theta_{\ell,j}^{d-i}$ appearing in the second term on the right hand side of (3.3). Note that the equation contains all of the cointegrating vectors for $m<\ell$, which appear in the higher order "error correction" terms making up the third term on the right hand side of (3.3). For example, in a system with d=2 the equations describing cointegration in the levels contain any cointegrating relations between the first differences.

The likelihood function follows directly from (3.3). Let λ_ℓ denote the parameters of $c_{\ell\ell}(L)$ and $d_{\ell m}(L)$, $m=1,\ldots m_{\ell-1}$, let θ_ℓ denote $\{\theta_{\ell,j}^{d-i}\}$, let μ_ℓ denote $\{\bar{\mu}_{\ell,j}\}$, and let λ, θ, and μ represent the collection of λ_ℓ, θ_ℓ, and μ_ℓ. The likelihood function can be written as:

$$
\begin{aligned}
(3.4)\quad f(Y_T,\theta,\mu,\lambda) = {} & f(Y_T^{d+1}|Y^1{}_T,\ldots,Y_T^d,\theta_2,\theta_3,\ldots,\theta_{d+1},\bar{\mu}_{d+1},\lambda_{d+1}) \\
& \times f(Y_T^d|Y^1{}_T,\ldots,Y_T^{d-1},\theta_2,\theta_3,\ldots,\theta_d,\bar{\mu}_d,\lambda_d) \\
& \quad \cdots \\
& \times f(Y_T^2|Y_T^1,\theta_2,\mu_2,\lambda_2)f(Y_T^1,\mu_1,\lambda_1)
\end{aligned}
$$

where $Y_T=(y_1', y_2', \ldots, y_T')'$ and $Y_T^i=(y_1^{i\prime}, y_2^{i\prime}, \ldots, y_T^{i\prime})'$ for $i=1,\ldots,d+1$.

The factorization in (3.4) shows that, for $\ell>2$, in general Y^m for $m<\ell$ will not be weakly exogneous for θ_ℓ because the likelihood of Y^ℓ depends on

$\theta_2, \ldots, \theta_{\ell-1}$. In this case it will not be possible to condition on Y^m for $m<\ell$ when constructing these MLE's for the parameters in the ℓ'th equation. Thus the estimators of parameters in the ℓ'th block of equations will not in general be conditionally normally distributed and the Wald test statistics will not have an asymptotic χ^2 distribution.

An important exception to this situation is when all the cointegrating vectors making up θ_m for $m<\ell$ are known, for example when there are no such cointegrating vectors. Thus we consider estimation of the ℓ'th equation when θ_m for $m<\ell$ are known. The analysis is facilitated by first transforming (3.3) to isolate the regressors of different orders in probability:

$$(3.5) \qquad \Delta^{d-\ell+1} y_t^\ell = (z_t' \otimes I)\delta + e_t$$

where $e_t = c_{\ell\ell}(L)\epsilon_t^\ell$, where $z_t = (z_t^{1\prime}\ z_t^{2\prime}\ \ldots\ z_t^{2\ell\prime})'$ and where $\delta = (\delta_1'\ \delta_2'\ \ldots\ \delta_{2\ell}')'$, with $\delta_i = (\delta_{i1}'\ \delta_{i2}'\ \ldots\ \delta_{in}')'$, where δ_{ij} is the vector of coefficients on z_t^i in the j'th equation in the block of equations (3.3). By construction, $E[e_t | \{z_t\}] = 0$. The transformed regressors z_t in (3.5) are the canonical regressors discussed in detail in Sims, Stock and Watson (1990). They are constructed so that z_t^1 is a zero mean I(0) vector, z_t^2 is a constant, z_t^3 is dominated by a martingale, z_t^4 is a linear time trend, and so forth. In general $\sum_{t=1}^T z_t^i z_t^{i\prime}$ is $O_p(T^{i-1})$ for $i \geq 2$. Thus, z_t^1 contains the requisite number of leads and lags of u_t^m for $m<\ell$ dictated by the polynomial $d_{\ell m}(L)$, $z_t^2 = 1$, z_t^3 is composed of the singly integrated elements of u_t^m for $m=1,2,\ldots,\ell-1$, $z_t^4 = t$, z_t^5 is composed of the doubly integrated elements of u_t^m for $m=1,2,\ldots,\ell-2$, etc.

4. Estimation and Testing

This section considers the Gaussian estimation of the parameters δ in (3.5). It is assumed that y_t has the extended triangular representation (3.1) with u_t given by (3.2). It is also assumed that the conditions hold under which (3.5) obtains from (3.3). We consider estimation in the case that z_t and δ are finite dimensional, i.e. in which $\{d_{\ell m}(L)\}$ have finite orders. Although the motivation for the representation (3.5) is to provide a convenient framework for computing the Gaussian MLE, as usual the asymptotic distribution theory is valid under weaker assumptions than Gaussianity. It is therefore assumed that $\{\epsilon_t\}$ in (3.2) is a martingale difference sequence with $E[\epsilon_t\epsilon_t'|\epsilon_{t-1},\epsilon_{t-2},\ldots]=I_n$ and $\max_i \sup_t E[(\epsilon_{it})^4|\epsilon_{t-1},\epsilon_{t-2},\ldots]<\infty$.

There are two natural estimators of the parameters in (3.5), the GLS estimator based on an estimator of $c_{\ell\ell}(L)$ and the OLS estimator. These estimators are

$$\hat{\delta}_{gls} = [\sum\tilde{z}_t\tilde{z}_t']^{-1}[\sum\tilde{z}_t(\Delta^{d-\ell+1}\tilde{y}_t^{\ell})]$$
$$\hat{\delta}_{ols} = [(\sum z_t z_t') \otimes I_{k\ell}]^{-1}[\sum (z_t \otimes I_{k\ell})(\Delta^{d-\ell+1}y_t^{\ell})]$$

where $\tilde{z}_t=[z_t \otimes \hat{\Phi}(L)']$ and $\tilde{y}_t^{\ell} = \hat{\Phi}(L)y_t^{\ell}$, where $\hat{\Phi}(L)$ is a consistent estimator of $\Phi(L)=c_{\ell\ell}(L)^{-1}$.

Associated with the GLS estimator is the Wald statistic testing $R\delta=r$,

$$W_{gls} = (R\hat{\delta}_{gls}-r)'[R(\sum\tilde{z}_t\tilde{z}_t')^{-1}R']^{-1}(R\hat{\delta}_{gls}-r) \ .$$

Because the disturbance in (3.5) is serially correlated, the Wald statistic for $\hat{\delta}_{ols}$ must be constructed using a modified covariance matrix. When the hypotheses of interest do not involve the coefficients on the mean-zero stationary regressors in (3.5), this modification is the serial correlation robust estimator of the covariance matrix using the spectral density matrix of e_t at frequency zero, $\hat{c}_{\ell\ell}(1)\hat{c}_{\ell\ell}(1)'$. That is,

$$W_{ols} = [R\hat{\delta}_{ols}-r]'(R[(\sum z_t z_t') \otimes \hat{c}_{\ell\ell}(1)\hat{c}_{\ell\ell}(1)']^{-1}R')^{-1}[R\hat{\delta}_{ols}-r]$$

Define the scaling matrix T_T to be a block diagonal matrix partitioned conformably with z_t, with diagonal blocks $T_{1T}=T^{1/2}I$ and $T_{iT} = T^{(i-1)/2}I$ for $i>2$. From Sims, Stock, and Watson (1990, Section 2) z_t can be written as $z_t = G(L)\upsilon_t$, where $G(L)$ is a block lower triangular matrix and $\upsilon_t=(\xi_t^{0\prime}\ 1\ \xi_t^{1\prime}\ t\ \xi_t^{2\prime}\ \ldots\ \xi_t^{\ell-1\prime}\ t^{\ell-1})'$, where $\xi_t^0=(\epsilon_t^{1\prime}\ \epsilon_t^{2\prime}\ \ldots\ \epsilon_t^{\ell-1\prime})'$ and where ξ_t^g is defined recursively by $\xi_t^g=\sum_{s=1}^t\xi_s^{g-1}$ for $g\geq1$. Let $\Gamma_w(j) = E[w_t-E(w_t)][w_{t-j}-E(w_t)]'$ for any variable w_t.

The next four theorems, proven in the Appendix, summarize the asymptotic distributions of these statistics.

Theorem 4.1 Suppose that y_t satisfies (3.3) and (3.5) where $c_{jj}(L)$ is $d+1-j$ summable, $j=1,\ldots,d+1$, that $\Phi(L)=c_{\ell\ell}(L)^{-1}$ has known order $q<\infty$, and $d_{\ell m}(L)$ has a known finite order. Then $(T_T \otimes I_n)(\hat{\delta}_{gls}-\delta) \Rightarrow Q^{-1}\phi$, where after partioning Q and ϕ conformably with δ:

$Q_{11} = E[(z_t^1 \otimes \Phi(L)')(z_t^1 \otimes \Phi(L)')']$, $Q_{1j} = 0$, $j>2$, and

$Q_{ij} = [V_{ij} \otimes \Phi(1)'\Phi(1)]$ for $i,j > 2$, where $V_{22} = 1$,

$$V_{mp} = G_{mm}(1)[\textstyle\int_0^1 W_1^{(m-1)/2}(s)W_1^{(p-1)/2}(s)'ds]G_{pp}(1)' ,$$

$$m=3,5,7,\ldots,2\ell-1,\ p=3,5,7,\ldots,2\ell-1$$

$$V_{mp} = G_{mm}(1)[\textstyle\int_0^1 t^{(m-2)/2}W_1^{(p-1)/2}(s)'ds]G_{pp}(1)' = V'_{pm} ,$$

$$m=2,4,6,\ldots,2\ell,\ p=3,5,6,\ldots,2\ell-1$$

$V_{mp} = 2/(p+m-2)G_{mm}(1)G_{pp}(1)'$, $m=2,4,6,\ldots,2\ell$, $p=2,4,6,\ldots,2\ell$,

$\phi_1 \sim N(0,Q_{11})$

$\phi_m = \mathrm{Vec}(\Phi(1)'[G_{mm}(1)(\int_0^1 s^{(m-2)/2}dW_2(s)')]')$, $m=2,6,\ldots,2\ell$

$\phi_m = \mathrm{Vec}(\Phi(1)'[G_{mm}(1)(\int_0^1 W_1^{(m-1)/2}(s)dW_2(s)')]')$, $m=3,5,7,\ldots,2\ell-1$

where W_1 and W_2 are independent standard Weiner processes of dimension $\sum_{m=1}^{\ell-1}k_m$ and k_ℓ respectively and where ϕ_1 is independent of ϕ_m, $m>1$.

Theorem 4.2. Under the assumptions of Theorem 1,

(a) $(\Upsilon_T \otimes I_n)(\hat{\delta}_{ols}-\delta) \Rightarrow [V^{-1}\otimes I_r]\omega$, where after partioning V and ω conformably with δ:

$\omega_1 \sim N(0,\Sigma_{\omega_1})$, where $\Sigma_{\omega_1} = \sum_{j=-\infty}^{\infty}[\Gamma_{z1}(j) \otimes \Gamma_e(j)]$.

$\omega_m = \mathrm{Vec}(c_{\ell\ell}(1)'[G_{mm}(1)(\int_0^1 s^{(m-2)/2}dW_2(s)')]')$, $m=2,6,\ldots,2\ell$

$\omega_m = \mathrm{Vec}(c_{\ell\ell}(1)'[G_{mm}(1)(\int_0^1 W_1^{(m-1)/2}(s)dW_2(s)')]')$, $m=3,5,7,\ldots,2\ell-1$

where ω_1 is independent of ω_m, $m>1$, and where $V = [V_{ij}]$, $i,j = 1, 2,\ldots,2\ell$, where $V_{11} = Ez_t^1 z_t^{1\prime}$, $V_{1j} = 0$, $j\geq 2$, and V_{ij}, $i,j > 2$ are given in Theorem 4.1. This holds even if $c_{\ell\ell}(L)^{-1}$ has infinite order but is 1-summable.

(b) Partition $\delta = (\delta_1{}'\ \delta_*{}')'$ so that δ_1 denotes the elements of δ corresponding to z_t^1 and δ_* corresponds to the remaining parameters.

Similarly partition $\hat{\delta}_{ols}$, $\hat{\delta}_{gls}$, $z_t=(z_t^{1\prime}\ z_t^{*\prime})'$, and $T_T=\text{diag}(T_{1T},T_{*T})$. Then $(T_{*T} \otimes I_n)(\hat{\delta}_{*ols}-\hat{\delta}_{*gls}) \xrightarrow{p} 0$.

Theorem 4.3. Under the Assumptions of Theorem 4.1, $W_{gls} \Rightarrow \chi^2_m$.

Theorem 4.4. Suppose that the first $\dim(z_t^1)$ columns of R equal zero and that $\hat{c}_{\ell\ell}(1) \xrightarrow{p} c_{\ell\ell}(1)$. Then under the assumptions of Theorem 4.1, $W_{ols} - W_{gls} \xrightarrow{p} 0$ and $W_{ols} \Rightarrow \chi^2_m$.

Note that $c_{\ell\ell}(L)^{-1}$ need not be finitely parameterized to implement the OLS estimator but $c_{\ell\ell}(1)$ needs to be consistently estimated to construct W_{ols}.

The asymptotic equivalence of the dynamic OLS and the feasible GLS estmators (Theorem 4.2(b)) for the coefficients on the integrated regressors is a consequence of the trending properties of these regressors. That is, for m>2 the GLS-transformed regressors are asymptotically colinear with their untransformed counterparts. This result extends the familiar result for the case of a constant and polynomial time trend (Grenander and Rosenblatt [1957]) and extends the results of Phillips and Park (1986) to the general integrated regression model with regressors of various orders of integration.

The result concerning the asymptotic χ^2 distribution of the Wald test statistics applies whether or not the integrated regressors have components that are polynomials in time. However, the limiting distribution of the estimator itself will differ depending on whether time (say) is included as a regressor and whether some of the regressors have a time trend component. For example, suppose that y_t is bivariate I(1), y_t^1 has nonzero drift, and time is excluded from the GLS or dynamic OLS regression. West (1988) showed that the static OLS estimator from the regression of y_t^2 onto y_t^1 has a large sample normal distribution with a nonrandom variance, a result that extends to the MLE computed by either GLS or dynamic OLS. Moreover, the test statistics are asymptotically χ^2. Although the asymptotic distribution of θ changes depending on whether Δy_t^1 has a nonzero mean, the distribtuon of the test statistic does not. Precisely which elements of y_t contain deterministic components and which polynomials of t are included in (3.3) determine the transformation to the canonical regressors, z_t.

It is useful to identify two circumstances in which the assumptions underlying the results in this section are violated and the asymptotic χ^2 result does not obtain. To simplify, consider the d=1 case. The first circumstance is when constraints are imposed on $\mu_{2,0}$ in (3.1). The constant term in (3.3) is $\bar{\mu}_{2,0} = \mu_{2,0}-d_{21}(1)\mu_{1,0}$, so that restrictions on $\mu_{2,0}$ impose cross-equation restrictions between the coefficients in (3.3) and the first block of equations in (3.1), implying that y_t^1 is not weakly exogenous for θ. The second is noted by Phillips (1988a) who points out that if the unit root in the y_t^1 process is estimated rather than imposed *a priori*, the asymptotic χ^2 distribution for the Wald statistic will not obtain. This follows from (2.3) since (y_t^1) fails to be weakly exogenous for θ because Δy_t^1 is replaced by $(1-\rho L)y_t^1$, imposing a cross-equation constraint.

These theorems apply to the case that there are a fixed number of regressors. Conceptually, one could view this estimator as semiparametric by embedding this parametric regression in a sequence of regressions where the number of regressors increase as a function of the sample size. A formal treatment of this extension would entail generalizing the univariate I(0) results of Berk (1974) and the univariate I(1) results of Said and Dickey (1984) to the I(d), vector-valued case, an extension beyond the scope of this paper.

5. An Example

The motivating empirical problem stated in the introduction was estimating the parameters (θ_p, θ_Q, and θ_r) of a cointegrating money demand relation. In Engle and Granger's (1987) terminology, money (m_t) and prices (p_t) are cointegrated of order (2,1), i.e. $m_t-\theta_p p_t$ is I(1). Were θ_p known to be 1, θ_Q and θ_r could be estimated in the I(1) framework of Section 2, with $y_t^1 = (Q_t, r_t)$ and $y_t^2 = m_t-\theta_p p_t$. If θ_p is unknown, or if one wishes to test $\theta_p=1$, the MLE of (θ_p, θ_Q, θ_r) can be obtained using the framework of Section 3. The equations corresponding to (3.1) are

$$\ell=1: \quad \Delta^2 p_t = u_t^1$$

$$(5.1) \qquad \ell=2: \quad \begin{bmatrix}\Delta Q_t \\ \Delta r_t\end{bmatrix} = \mu_{2,0} + u_t^2$$

$$\ell=3: \quad m_t = \mu_{3,0} + \theta_p p_t + \theta_Q Q_t + \theta_r r_t + u_t^3$$

so that d=2, $y_t^1=p_t$, $y_t^2=(Q_t, r_t)$, and $y_t^3=m_t$. Thus inflation has zero drift (this could be relaxed), Q_t has nonzero drift, r_t could have zero or nonzero drift, and the money demand cointegrating vector implies that m_t is I(2) and inherets any deterministic components of p_t, Q_t, and r_t.

The error triangularization (3.2) yields the regression,

$$(5.2)\quad m_t = \bar{\mu}_{3,0} + \theta_p p_t + \theta_Q Q_t + \theta_r r_t + d_{31}(L)\Delta^2 p_t + d_{32;Q}(L)\Delta Q_t + d_{32;r}(L)\Delta r_t + c_{33}(L)\epsilon_t^3$$

where $d_{31}(L)$ and $d_{32}(L)$ are two-sided and finitely parameterized. Because there are no cointegrating vectors with unknown coefficients in the ℓ=2 equation and because there are no restrictions on $\mu_{3,0}$ (so that there are no cross-restrictions on the drifts), (p_t, Q_t, r_t) are weakly exogenous for (θ_p, θ_Q, θ_r). Thus GLS or dynamic OLS on (5.2) asymptotically yields the MLE.

6. Monte Carlo Results

This section summarizes a comparative study of the sampling properties of six estimators of cointegrating vectors in two different probability models. The six estimators are: the static OLS estimator (Engle and Granger [1987], Stock [1987]), the dynamic OLS estimator $\hat{\delta}_{ols}$ and the GLS estimator $\hat{\delta}_{gls}$ introduced in Sections 2-4, the zero frequency band spectrum estimator of Phillips (1988b), the fully modified estimator of Phillips and Hansen (1989) (essentially this is a zero frequency SUR estimator), and Johansen's (1988) VAR maximum likelihood estimator. All of the estimators except static OLS are asymptotically equivalent for the data generation processes considered, at least when they are interpreted as semi-parametric estimators.

The Monte Carlo experiments study bivariate models in which x_t and y_t are each I(1) with no drift and y_t-θx_t is I(0). The two models and the results are summarized in Tables 1 and 2. Two sample sizes were used in the simulations, T=160 (40 years of quarterly data) and T=360 (30 years of monthly data).

Model 1. The two equivalent representations for this model, presented in Table 1, correspond to the usual prediction error decomposition (A in Table 1) and to the two-sided triangular representation of Sections 2 and 3 (B in Table 1). None of the estimators correspond to exact maximum likelihood for this model. For example, the GLS estimator is constructed assuming that w_t follows an AR rather than an MA process and the Johansen estimator uses a VAR for the x_t, y_t process.

The first column shows that the dynamic OLS and GLS estimators have no significant bias. The other MLE's have small biases, approximately one-fifth the bias of the static OLS estimator. The distribution of the t-statistics is shifted to the right for the OLS estimator. The exact 5% critical values for the Wald statistics differ somewhat from 3.84, the value appropriate for the χ_1^2 distribution, less so in the larger sample. The largest descrepancy is for the dynamic GLS estimator, where the asymptotic 5% critical value leads to tests with sizes of 1.6% (T=160) and 2.6% (T=360).

Model 2. The second model that we consider has been used by Engle and Granger (1987), Banerjee et al (1988) and Phillips and Hansen (1988). Here x_t and y_t follow a cointegrated VAR(1) process so the Johansen estimator corresponds to exact maximum likelihood. The other estimators provide alternative approximations to the MLE.

The results are shown in Table 2. Not surprisingly the Johansen estimator has the best performance, the static OLS estimator the worst. It is interesting to contrast the dynamic OLS, GLS, Phillips and Phillips/Hansen estimators to the Johansen estimator on the one hand and the static OLS estimator on the other hand. The Phillips and Phillips/Hansen estimators have biases on the order of 30%-50% of the the magnitude of the bias in static OLS; the bias in the dynamic OLS and GLS estimators is somewhat less. The distribution of the Wald statistics for these (approximate) MLE's are sharply shifted to the right, albeit much less so than the static OLS estimator. The rightward shift in the distribution of the t and Wald statistics is more severe for the Phillips and Phillips/Hansen estimators than for the dynamic OLS and GLS estimators. Comparing the results for T=160 and T=360 suggests the convergence of the estimators implied by the asymptotic theory. The reason for the relatively poor performance of these estimators can be traced to the relatively poor estimates of the relevant spectra at frequency zero constructed using relatively short lag windows.

7. Conclusions

The two new asymptotic MLE's are easy to implement in practice and can be applied to a wide range of problems. The Monte Carlo simulations indicated that the performance of the various MLE's can vary substantially in finite samples. For the first design, all the MLE's exhibited reasonable performance, perhaps with the exception of the GLS estimator. For the second design, all the MLE's (except the exact MLE for this design) behaved relatively poorly. When performance is poor, it is linked to poor performance of estimators of the spectral density matrix of the errors at frequency zero.

Appendix

Proof of Lemma 2.1.

The proof is a modification of Anderson's (1971, Theorem 7.6.7) proof of the Wold decomposition. Let Ψ_t^i denote the Hilbert space spanned by $\{u_t^i, u_{t-1}^i, u_{t-2}^i, \ldots\}$, let $P(u_t^i|\Psi_s^j)$ denote the linear projection of u_t^i onto Ψ_s^j, and let $\epsilon_t^1 = u_t^1 - P(u_t^1|\Psi_{t-1}^1)$. Then $u_t^1 = c_{11}(L)\epsilon_t^1$ is the Wold representation of u_t^1. The assumption that u_t has a nonsingular spectral density matrix implies that $c_{11}(L)$ is invertible (Anderson [1971], Theorem 7.6.9). Let $\Psi_\infty^1 = \bigcup_{t=-\infty}^{\infty}\Psi_t^1$ so $P(u_t^2|\Psi_\infty^1) = \tilde{c}_{21}(L)u_t^1 = c_{21}(L)\epsilon_t^1$, where $c_{21}(L) = \tilde{c}_{21}(L)c_{11}(L)^{-1}$. Now $u_t^2 - P(u_t^2|\Psi_\infty^1)$ has the Wold representation $u_t^2 - P(u_t^2|\Psi_\infty^1) = c_{22}(L)\epsilon_t^2$, where $\epsilon_t^2 = u_t^2 - P(u_t^2|\Psi_{t-1}^2 \oplus \Psi_\infty^1)$. By construction, $E\epsilon_t^2\epsilon_t^{1\prime} = 0$, $E\epsilon_t^1\epsilon_t^{1\prime} = I_{n-r}$, $E\epsilon_t^2\epsilon_t^{2\prime} = I_r$ by appropriate normalization of $c_{11}(0)$ and $c_{22}(0)$, and $E\epsilon_t\epsilon_s' = 0$ for $t \neq s$. Finally, $c(L)$ is square summable because $Eu_tu_t' < \infty$ by assumption. $\square$

Derivation of (3.1).

Assume that the $n\times1$ vector y_t has Wold representation $\Delta^d y_t = \mu + F^d(L)a_t$, where (i) a_t is a martingale difference sequence with $E(a_ta_t'|a_{t-1}, a_{t-2}, \ldots) = \Sigma_a$ and $\max_i \sup_t E(a_{it}^4) < \infty$, (ii) $a_s = 0$ for $s \leq 0$, (iii) $F^d(L) = \sum_{j=0}^{\infty} F_j^d L^j$, with $\sum_{j=0}^{\infty} j^d |F_j^d| < \infty$, (iv) $F^d(e^{-i\omega})$ is nonsingular for $\omega \neq 0$, and (v) $\text{rank}[F^d(1)] = k_1 \leq n$. The triangular representation (3.1) is constructed by repeated application of the following Lemma:

Lemma A.1. Assume that the $n\times1$ vector x_t is generated by $\Delta x_t = \sum_{j=0}^{m}\mu_j t^j + F(L)a_t$, where a_t satisfies (i) and (ii), $F(L)$ is ℓ-summable, satisfies (iv), and rank[F(1)]=k≤n. Without loss of generality assume that x_t is ordered so that the upper k×n block of F(1) has full row rank. Then x_t can be represented as:

$$\Delta x_t^1 = \sum_{j=0}^{m}\bar{\mu}_{1,j}t^j + D_1(L)a_t$$
$$x_t^2 = \sum_{j=0}^{m+1}\bar{\mu}_{2,j}t^j + \theta x_t^1 + D_2(L)a_t$$

where $x_t=(x_t^{1\prime}\ x_t^{2\prime})'$, where x_t^1 is k×1, x_t^2 is (n-k)×1, and $D(L)=[D_1(L)'\ D_2(L)']'$ is $(\ell-1)$ summable. When μ_m lies in the column space of F(1), $\bar{\mu}_{2,m+1}=0$.

Proof. The result holds trivially for k=n, so consider k<n. Order x_t so that F(L) can be partitioned as $F(L)=[F_1(L)'\ F_2(L)']'$ where $F_1(L)$ is k×n, $F_2(L)$ is (n-k)×n, and $F_1(1)$ has full row rank. By definition $\alpha' F(1)=0$. Because $F_1(1)$ has full row rank, $F_2(1)=\tilde{\alpha}' F_1(1)$ for some k×r matrix $\tilde{\alpha}$. Now partition μ_i as $(\mu_{1,i}'\ \ \mu_{2,i}')'$, so that

$$\text{(A.1)} \qquad \Delta x_t^2 - \tilde{\alpha}'\Delta x_t^1 = \sum_{i=0}^{m}(\mu_{2,i}-\tilde{\alpha}'\mu_{1,i})t^i + [F_2(L)-\tilde{\alpha}'F_1(L)]a_t.$$

Accumulating (A.1) yields $x_t^2 - \tilde{\alpha}' x_t^1 = \sum_{i=0}^{m+1}\bar{\mu}_{2,i}t^i + D_2(L)a_t$, where $D_2(L) = F_2^*(L) - \tilde{\alpha}'F_1^*(L)$, where $F_i^*(L) = (1-L)^{-1}(F_i(L)-F_i(1))$, i=1,2. Because F(L) is ℓ summable, $F^*(L)$ is $(\ell-1)$ summable. If μ_m lies in the column space of F(1), then $\mu_{2,m}-\tilde{\alpha}'\mu_{1,m}=0$ so $\bar{\mu}_{2,m+1}=0$. The Lemma follows by setting $\bar{\mu}_{1,i} = \mu_{1,i}$ (i=0,...,m) and $D_1(L)=F_1(L)$. □

To construct the triangular representation (3.1), apply Lemma A.1 to $x_t = \Delta^{d-1}y_t$ to yield the decompostion:

$$
\begin{aligned}
\Delta^{d}\bar{y}^1_t &= \tilde{\mu}_{1,0} + F_1^{d-1}(L)a_t \\
\Delta^{d-1}\bar{y}^2_t &= \tilde{\mu}_{2,0} + \tilde{\mu}_{2,1}t + \theta^{d-1}_{2,1}(\Delta^{d-1}\bar{y}^1_t) + F_2^{d-1}(L)a_t
\end{aligned}
$$

where y_t has been partitioned into $k_1\times 1$ and $(n-k_1)\times 1$ components $\bar{y}^1_t$ and $\bar{y}^2_t$. Now assume that $F^{d-1}(1)=[F_1^{d-1}(1)'\ \ F_2^{d-1}(1)']'$ has rank $k_1+k_2\leq n$, and apply the lemma to $x_t = [\Delta^{d-1}\bar{y}^1_t{}',\ \ (\Delta^{d-2}\bar{y}^2_t-\theta^{d-1}_{2,1}\Delta^{d-2}\bar{y}^1_t)]$. Continuing this process until Δx_t has full rank spectral density matrix at frequency zero yields the triangular representation (3.1), with $u^j_t=D_j(L)a_t$, $j=1, \ldots, d+1$, where rank $[D_j(1)]=k_j$. □

Proof of Theorem 4.1.

First consider the infeasible GLS estimator $\bar{\delta}_{gls}$ constructed using $\Phi(L) = c_{\ell\ell}(L)^{-1}$. Note that $(T_T\otimes I)(\bar{\delta}_{gls}-\delta) = Q_T^{-1}\phi_T$, where $Q_T = (T_T^{-1}\otimes I)\sum_t\bar{z}_t\bar{z}_t'(T_T^{-1}\otimes I)$ and $\phi_T = (T_T^{-1}\otimes I)\sum_t\bar{z}_t\bar{y}^\ell_t$, with $\bar{z}_t = [z_t\otimes\Phi(L)']$ and $\bar{y}^\ell_t = \Phi(L)y^\ell_t$. The convergence of Q_{11T} to Q_{11} follows from a standard application of the weak law of large numbers. For Q_{ijT} with i or j $\geq$ 2:

$$
\begin{aligned}
Q_{ijT} &= (T_{iT}^{-1}\otimes I)\textstyle\sum_t[\sum_{m=0}^q(z^i_{t-m}\otimes\Phi_m')][\sum_{h=0}^q(z^j_{t-h}\otimes\Phi_h')]'(T_{jT}^{-1}\otimes I) \\
&= (T_{iT}^{-1}\otimes I)\textstyle\sum_t[\sum_{m=0}^q\sum_{h=0}^q(z^i_{t-m}z^{j}_{t-h}{}'\otimes\Phi_m'\Phi_h)](T_{jT}^{-1}\otimes I) \\
&= (T_{iT}^{-1}\otimes I)\textstyle\sum_t[\sum_{m=0}^q\sum_{h=0}^q(z^i_tz^{j}_t{}'\otimes\Phi_m'\Phi_h)](T_{jT}^{-1}\otimes I) + o_p(1) \\
&\Rightarrow (V_{ij}\otimes\Phi(1)'\Phi(1))
\end{aligned}
$$

where the last two lines follow from Lemma 1 of Sims, Stock and Watson (1990) (SSW). For ϕ_{iT}, i $\geq$ 2:

$$\phi_{iT} = (T_{iT}^{-1}\otimes I)\sum_t[\sum_{m=0}^{q}(z_{t-m}^{i} \otimes \Phi_m')]\epsilon_t^{\ell}$$
$$= (T_{iT}^{-1}\otimes I)\sum_t(z_t^i \otimes \Phi(1)')]\epsilon_t^{\ell} + o_p(1)$$
$$= \mathrm{vec}(\Phi(1)'[T_{iT}^{-1}\sum_t z_t^i\epsilon_t^{\ell}{}']')$$

$$\Rightarrow \begin{cases} \mathrm{Vec}(\Phi(1)'[G_{mm}(1)(\int_0^1 s^{(m-2)/2}dW_2(s)')]'), & m=2,6,\ldots,2\ell \\ \mathrm{Vec}(\Phi(1)'[G_{mm}(1)(\int_0^1 W_1^{(m-1)/2}(s)dW_2(s)')]'), & m=3,5,7,\ldots,2\ell-1 \end{cases}$$

where the last line follows from Lemma 1 of SSW. The joint convergence and distribution of ϕ follows from SSW Lemmas 1 and 2. To prove that the feasible GLS estimator has the same limit, as usual let

$$\hat{Q}_T = (T_T^{-1}\otimes I)\sum_t[\sum_{m=0}^{q}(z_{t-m} \otimes \hat{\Phi}_m')][\sum_{h=0}^{q}(z_{t-h} \otimes \hat{\Phi}_h')]'(T_T^{-1}\otimes I)$$
$$\hat{\phi}_T = (T_T^{-1}\otimes I)\sum_t[\sum_{m=0}^{q}(z_{t-m} \otimes \hat{\Phi}_m')]\epsilon_t^{\ell}$$

so $(T_T^{-1}\otimes I)(\hat{\delta}_{gls}-\overline{\delta}_{gls}) = \hat{Q}_T(\hat{\phi}_T-\phi_T) + (\hat{Q}_T-Q_T)\phi_T$. Assume that $\hat{\Phi}_j \xrightarrow{p} \Phi_j$ for $j=1,\ldots,q$. Evidently $\hat{Q}_T - Q_T \xrightarrow{p} 0$ and $\hat{\phi}_T - \phi_T \xrightarrow{p} 0$, from which asymptotic equivalence of GLS and feasible GLS follows. □

Proof of Theorem 4.2.

(a) By assumption, $c_{jj}(L)$ is d+1-j summable for j=1,2,...,d+1. This implies that the diagonal entries $G_{jj}(L)$ of G(L) corresponding to the stochastic elements, ξ_t^j, in υ_t from equation (3.7) are j summable. The theorem then follows from Lemma 1 of SSW. □

(b) Theorems 4.1 and 4.2 imply that $T_{1T}^{-1}\sum_t z_t^1 z_t^{*}{}'T_{*T}^{-1} \xrightarrow{p} 0$. First consider the infeasible GLS estimator $\overline{\delta}_{gls}$, defined in the proof of Theorem 4.1. Theorem 4.1 implies that

$$(T_{*T} \otimes I)(\bar{\delta}_{*gls}-\delta_*) = B_{*T}^{-1}[(T_{*T}^{-1} \otimes I)\sum_t(z_t \otimes \Phi(L)')\epsilon_t^{\ell}] + o_p(1)$$

where $B_{*T} = (T_{*T}^{-1}\otimes I)[\sum_t(z_t^*\otimes\Phi(1)')(z_t^{*\prime}\otimes\Phi(1))](T_{*T}^{-1}\otimes I)$. Now
$B_{*T} = (T_{*T}^{-1}\otimes I)(I\otimes\Phi(1)'\Phi(1))[\sum_t(z_t^*z_t^{*\prime}\otimes I)](T_{*T}^{-1}\otimes I)$
so $B_{*T}^{-1} = [(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes I)(z_t^{*\prime}\otimes I)(T_{*T}^{-1}\otimes I)]^{-1}[I\otimes\Phi(1)'\Phi(1)]^{-1}$. Also,

$$\begin{aligned}(T_{*T}\otimes I)(\hat{\delta}_{*ols}-\delta_*) &= [(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes I)(z_t^{*\prime}\otimes I)(T_{*T}^{-1}\otimes I)]^{-1} \\ &\quad \times (T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes I)c_{\ell\ell}(L)\epsilon_t^{\ell} + o_p(1) \\ &= B_{*T}^{-1}(I\otimes\Phi(1)'\Phi(1))(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes I)c_{\ell\ell}(L)\epsilon_t^{\ell} + o_p(1) \\ &= B_{*T}^{-1}(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes\Phi(1)'\Phi(1))c_{\ell\ell}(L)\epsilon_t^{\ell} + o_p(1) .\end{aligned}$$

Thus

$$\begin{aligned}(T_{*T}\otimes I)(\hat{\delta}_{*ols}-\bar{\delta}_{*gls}) &= B_{*T}^{-1}\{(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes\Phi(1)'\Phi(1))c_{\ell\ell}(L)\epsilon_t^{\ell} \\ &\quad - \sum_t(z_t^*\otimes\Phi(L)')\epsilon_t^{\ell}\} + o_p(1) \\ &= B_{*T}^{-1}(T_{*T}^{-1}\otimes I)\sum_t\{(z_t^*\otimes\Phi(1)')(I\otimes\Phi(1))[c_{\ell\ell}(1)\epsilon_t^{\ell} + c_{\ell\ell}^*(L)\Delta\epsilon_t^{\ell}] \\ &\quad - (z_t^*\otimes\Phi(1)')\epsilon_t^{\ell} + [z_t^*\otimes(\Phi(1)-\Phi(L))']\epsilon_t^{\ell}\} + o_p(1) \\ &= B_{*T}^{-1}(A_{1T} + A_{2T})\end{aligned}$$

where the final equality follows from $\Phi(1)c_{\ell\ell}(1)=I$, and where $\Phi^*(L)=(1-L)^{-1}(\Phi(L)-\Phi(1))$, $c_{\ell\ell}^*(L)=(1-L)^{-1}(c_{\ell\ell}(L)-c_{\ell\ell}(1))$, and

$$\begin{aligned}A_{1T} &= (T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes\Phi(1)'\Phi(1))c_{\ell\ell}^*(L)\Delta\epsilon_t^{\ell} \\ A_{2T} &= (T_{*T}^{-1}\otimes I)\sum_t[z_t^*\otimes(\Phi(1)-\Phi(L))']\epsilon_t^{\ell} \\ &= (T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes\Phi^*(L)')\Delta\epsilon_t^{\ell} .\end{aligned}$$

Because $B_{*T} \Rightarrow Q_*$ (the $(*,*)$ block of Q given in Theorem 4.1), the result follows if

$A_{1T} \xrightarrow{p} 0$ and $A_{2T} \xrightarrow{p} 0$. Because $\Phi^*(L)$ has a finite order by assumption and $E(\epsilon_t^\ell | (z_t^*))=0$, standard telescoping arguments imply that $A_{1T} \xrightarrow{p} 0$ and $A_{2T} \xrightarrow{p} 0$. □

Proof of Theorem 4.3.

The result follows from Theorem 4.1 above and Theorem 4 of Johansen (1988) or alternatively from section 4 of Phillips (1988a). □

Proof of Theorem 4.4.

This follows directly from Theorem 4.3 and the proof of Theorem 4.2. □

References

Ahn, and G. Reinsel (1987), "Estimation of Partially Nonstationary Multivariate Autoregressive Models," forthcoming, Journal of the American Statistical Association.

Anderson, T.W. (1971), The Statistical Analysis of Time Series. Wiley: New York.

Banerjee, A., J.J. Dolado, D.F. Hendry, and G.W. Smith (1986), "Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence," Oxford Bulletin of Economics and Statistics, 48, 253-277.

Berk, K.N. (1974), "Consistent Autoregressive Spectral Estimates," Annals of Statistics, 2, 489-502.

Bewley, R.A. (1979), "The Direct Estimation of the Equilibrium Response in a Linear Model," Economic Letters, 3, 357-361.

Campbell, J.Y. (1987), "Does Savings Anticipate Labor Income? An Alternative Test of the Permanent Income Hypothesis," Econometrica, 55, 1249-1274.

Campbell, J.Y. and R.J. Shiller (1987), "Cointegration and Tests of Present Value Models," Journal of Political Economy, 95, 1062-1088.

Campbell, J.Y. and R.J. Shiller (1989), "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," The Review of Financial Studies, 1, 195-228.

Engle, R.F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing," Econometrica, 55, 251-276.

Engle, R.F., D.F. Hendry, and J.F. Richard (1983), "Exogeneity," Econometrica 51, no. 2, 277-304.

Grenander, U. and M. Rosenblatt (1957), Statistical Analysis of Stationary Time Series, John Wiley and Sons: New York.

Hansen, B.E. and P.C.B. Phillips (1988), "Estimation and Inference in Models of Cointegration: A Simulation Study," Cowles Foundation Discussion Paper No. 881.

Hoffman, D. and R.H. Rasche (1989), "Long-Run Income and Interest Elasticities of Money Demand in the United States," NBER Discussion Paper no. 2949.

Hylleberg, S. and G.E. Mizon (1989), "Cointegration and Error Correction Mechanisms," Economic Journal, March 1989.

Johansen, S. (1988), "Statistical Analysis of Cointegration Vectors," Journal of Economic Dynamics and Control, 12, 231-255.

Johansen, S. (1989), "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregression Models," manuscript, Institute for Mathematical Statistics, University of Copenhagen.

King, R., C. Plosser, J.H. Stock, and M.W. Watson (1987), "Stochastic Trends and Economic Fluctuations," NBER Discussion Paper No. 2229.

Phillips, P.C.B. (1988a), "Optimal Inference in Cointegrated Systems," Cowles Foundation Discussion Paper No. 866.

Phillips, P.C.B. (1988b), "Spectral Regression for Cointegrated Time Series," Cowles Foundation Discussion Paper No. 872.

Phillips, P.C.B. and B.E. Hansen (1989), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," forthcoming, Review of Economic Studies.

Phillips, P.C.B. and J.Y. Park (1986), "Asymptotic Equivalence of OLS and GLS in Regression with Integrated Regressors," Cowles Foundation Discussion Paper No. 802.

Said, S.E. and D.A. Dickey (1984), "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika, 71, 599-608.

Sims, C.A., J.H. Stock, and M.W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," forthcoming, Econometrica.

Stock, J.H. (1987), "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," Econometrica, 55, 1035-1056.

West, K. D. (1988), "Asymptotic Normality when Regressors have a Unit Root," Econometrica, 56, 1397-1418.

Table 1
Monte Carlo Results, Probability Model 1

Probability Model: $\Delta x_t = v_t$

$$y_t = \alpha + \theta x_t + u_t$$

where

A.
$$\begin{cases} v_t = e_t^1 + 0.5e_{t-1}^2 \\ u_t = e_t^2 + .7e_{t-1}^1 + .4e_{t-1}^2, \quad e_t^i \text{ NIID}(0,1) \end{cases}$$

equivalently:

B.
$$\begin{cases} v_t = \epsilon_t^1 \\ u_t = .4\epsilon_{t+1}^1 + .16\epsilon_t^1 + .56\epsilon_{t-1}^1 + w_t, \end{cases}$$

where $w_t = \epsilon_t^2 - .40\epsilon_{t-1}^2 - .35\epsilon_{t-2}^2$, where ϵ_t^1 NIID(0,1.25) and ϵ_t^2 NIID(0,1.03).

A. T=160

Estimator	Bias $E(\hat{\theta}-\theta)$	Std. Dev. $(\hat{\theta}-\theta)$	Empirical Critical Values $t_{.05}$	$t_{.95}$	$F_{.95}$	Frac. Rejections (5% Nominal Size)
Static OLS	.010	.024	-1.26	2.02	4.62	.031
Dynamic OLS	.000	.019	-1.67	1.66	4.16	.041
Dynamic GLS	.000	.019	-1.94	1.93	5.77	.016
Phillips BSR	.002	.019	-1.41	1.57	3.45	.063
Phillips/Hansen FM	.002	.019	-1.53	1.73	4.02	.045
Johansen VAR-MLE	-.003	.021	--	--	4.68	.031

B. T=360

Estimator	Bias $E(\hat{\theta}-\theta)$	Std. Dev. $(\hat{\theta}-\theta)$	Empirical Critical Values $t_{.05}$	$t_{.95}$	$F_{.95}$	Frac. Rejections (5% Nominal Size)
Static OLS	.005	.010	-1.30	1.98	4.45	.035
Dynamic OLS	.000	.008	-1.67	1.72	4.01	.045
Dynamic GLS	-.000	.008	-1.87	1.79	4.97	.026
Phillips BSR	.000	.008	-1.50	1.63	3.39	.066
Phillips/Hansen FM	.001	.008	-1.48	1.70	3.76	.053
Johansen VAR-MLE	.001	.009	--	--	4.69	.030

Notes to Table 1: The first column gives the bias, the second column the standard deviation. The third and fourth columns respectively present the 5% and 95% ordinates of the empirical distribution of the t-statistic for θ. The t-statistic for the static OLS regression was computed using the usual OLS formula; the t-statistics for the other estimators were computed by the appropriate method suggested by asymptotic theory. The fifth column presents the 5% critical value for the empirical distributions of the Wald test for the hypothesis $\theta=\theta_0$. The sixth column show the percent rejections from the empirical distribution of the Wald statistic computed using the usual $\chi^2_{1;.05}$ critical value of 3.84. The dynamic OLS estimators were constructed with 1 lead and lag of Δx_t in the regressions and the covariance matrix was calculated using a Bartlett lag window with 5 (T=160) and 8 (T=360) autocovariances. The GLS estimator was constructed with 1 lead and lag of Δx_t in the regressions and (estimated) AR(4) GLS corrections. Phillips BSR denotes the zero frequency band spectrum regression estimator described in Phillips (1988b) and the Phillips/Hansen FM estimator refers to the fully modified estimator described in Phillips and Hansen (1988). The estimated spectra for these estimators were computed using the same lag window and number of lags as the dynamic OLS covariance matrix. The Johansen VAR-MLE was computed using a VAR(5) (T=160) and a VAR(8) (T=360). The Johansen procedure yields an estimator and a likelihood ratio statistic. Thus, there are no entries in the third and fourth columns for the Johansen estimator, and the fifth column contains the 95th percentile of the empirical distribution of the likelihood ratio statistic. All results are based on 1000 Monte Carlo replications.

Table 2
Monte Carlo Results, Probability Model 2

Probability Model: $x_t = a_t^2 - a_t^1$

$y_t = 2a_t^1 - a_t^2$

where: $a_t^1 = a_{t-1}^1 + \epsilon_t^1$

$a_t^2 = 0.6a_{t-1}^2 + \epsilon_t^2$, ϵ_t^i NIID(0,1)

A. T=160

Estimator	Bias $E(\hat{\theta}-\theta)$	Std. Dev. $(\hat{\theta}-\theta)$	Empirical Critical Values $t_{.05}$	$t_{.95}$	$F_{.95}$	Frac. Rejections (5% Nominal Size)
Static OLS	.087	.070	-0.02	8.27	68.33	.000
Dynamic OLS	.018	.058	-2.16	3.03	9.82	.002
Dynamic GLS	.027	.061	-1.66	2.65	7.58	.006
Phillips BSR	.046	.060	-1.17	3.59	12.90	.000
Phillips/Hansen FM	.044	.061	-1.28	3.74	14.01	.000
Johansen VAR-MLE	-.000	.054	--	--	4.32	.041

B. T=360

Estimator	Bias $E(\hat{\theta}-\theta)$	Std. Dev. $(\hat{\theta}-\theta)$	Empirical Critical Values $t_{.05}$	$t_{.95}$	$F_{.95}$	Frac. Rejections (5% Nominal Size)
Static OLS	.041	.037	-0.42	8.06	64.88	.000
Dynamic OLS	.001	.025	-2.09	2.24	6.75	.009
Dynamic GLS	.003	.025	-1.70	1.87	4.88	.027
Phillips BSR	.016	.026	-1.29	2.79	7.80	.005
Phillips/Hansen FM	.014	.027	-1.43	2.83	8.28	.004
Johansen VAR-MLE	.002	.024	--	--	4.14	.042

Notes: For this model $y_t - \theta x_t$ is I(0) with $\theta=-2$. The dynamic OLS and GLS regression were computed using 5 (T=160) and 8 (T=360) lags in the regressions with y_t the left hand variable. The Johansen estimator was computed using a VAR(1). See the notes to Table 1.