Econometrica, Vol. 61, No. 4 (July, 1993), 783–820

A SIMPLE ESTIMATOR OF COINTEGRATING VECTORS IN HIGHER ORDER INTEGRATED SYSTEMS

BY JAMES H. STOCK AND MARK W. WATSON[1]

Efficient estimators of cointegrating vectors are presented for systems involving deterministic components and variables of differing, higher orders of integration. The estimators are computed using GLS or OLS, and Wald Statistics constructed from these estimators have asymptotic χ^2 distributions. These and previously proposed estimators of cointegrating vectors are used to study long-run U.S. money (M1) demand. M1 demand is found to be stable over 1900–1989; the 95% confidence intervals for the income elasticity and interest rate semielasticity are (.88, 1.06) and (−.13, −.08), respectively. Estimates based on the postwar data alone, however, are unstable, with variances which indicate substantial sampling uncertainty.

KEYWORDS: Error correction models, unit roots, money demand.

1. INTRODUCTION

PARAMETERS DESCRIBING THE LONG-RUN RELATION between economic time series, such as the long-run income and interest elasticities of money demand, often play an important role in empirical macroeconomics. If these variables are cointegrated as defined by Engle and Granger (1987), then the task of describing these long-run relations reduces to the problem of estimating cointegrating vectors. Recent research on the estimation of cointegrating vectors has focused on the case in which each series is individually integrated of order 1 (is I(1)), typically with no drift term. Johansen (1988a) and Ahn and Reinsel (1990) derived the asymptotic distribution of the Gaussian MLE when the cointegrated system is parameterized as a vector error correction model (VECM), and Johansen (1991) extended this result to the case of nonzero drifts. A series of papers has considered other efficient estimators based on a different model for cointegrated systems, the triangular representation. Phillips (1991a) studied estimation in a cointegrated model with general I(0) errors; Phillips and Hansen (1990) and Park (1992) considered two-step, frequency-zero seemingly unrelated regression estimators; and Phillips (1991b) used spectral methods to compute efficient estimators in the frequency domain.

This paper makes three main contributions. First, it develops two alternative, computationally simple estimators of cointegrating vectors. These estimators, which have been independently proposed and studied elsewhere in the case of

[1] The authors thank Manfred Deistler, Robert Engle, Peter Phillips, Danny Quah, Kenneth 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 earlier drafts of this paper. Helpful suggestions by a co-editor and two anonymous referees are gratefully acknowledged. The authors also thank Robert Lucas for kindly providing the annual data analyzed in Section 7. Gustavo Gonzaga and Graham Elliott provided able research assistance. This research was supported in part by the Sloan Foundation and National Science Foundation Grants SES-86-18984, SES-89-10601, and SES-91-22463.

I(1) variates (Hansen (1988), Phillips (1991a), Phillips and Loretan (1991), and Saikkonen (1991)), are developed here for cointegrating regressions among general I(d) variables with general deterministic components. (For an application of this estimator in the I(1) case, see King, Plosser, Stock, and Watson (1991).) The estimators are motivated as Gaussian MLE's for a particular parameterization of the triangular representation. However, under more general conditions they are asymptotically efficient in Saikkonen's (1991) and Phillips' (1991a) sense, having an asymptotic distribution that is a random mixture of normals and producing Wald test statistics with asymptotic chi-squared null distributions. In the I(1) case with a single cointegrating vector, one simply regresses one of the variables onto contemporaneous levels of the remaining variables, leads and lags of their first differences, and a constant, using either ordinary or generalized least squares. The resulting "dynamic OLS" (respectively GLS) estimators are asymptotically equivalent to the Johansen/Ahn-Reinsel estimator.

The second contribution is an examination of the finite sample performance of these estimators in a variety of Monte Carlo experiments. Although all the estimators perform well when the designs incorporate simple short-run dynamics, for designs that mimic the dynamics in U.S. real money (M1) balances, income, and interest rates, there is considerable variation across the estimators and associated confidence intervals. In these designs, the dynamic OLS estimator performs well relative to the other asymptotically efficient estimators.

The third contribution is the use of these procedures to investigate the long-run demand for money (M1) in the U.S. from 1900 to 1989. Other researchers (recently including Hafer and Jansen (1991), Hoffman and Rasche (1991), Miller (1991), and Baba, Hendry, and Starr (1992)) have argued either explicitly or implicitly that long-run money demand can be thought of as a cointegrating relation among real balances, real income, and an interest rate in postwar data. We find this characterization empirically plausible for the longer annual data as well, and therefore use these estimators of cointegrating vectors to examine Lucas' (1988) suggestion that there is a stable long-run M1 money demand relation spanning the twentieth century. Based on the full sample, 95% confidence intervals for the income elasticity and interest rate semielasticity are, respectively, (0.88, 1.06) and (-0.127, -0.075). The postwar data are dominated by a single long-run trend, reflected in growth in income and interest rates and stable real money balances from 1946 to 1982; this results in imprecise estimation of the long-run money demand parameters when only the postwar data are used.

The paper is organized as follows. The model and estimators are introduced in Section 2 for I(1) variables and are extended to I(d) variables in Section 3. The large-sample properties of the estimators and test statistics are summarized in Section 4. In Section 5, the I(2) case is examined in detail. Monte Carlo results are presented in Section 6. The application to long-run M1 demand is given in Section 7. Section 8 concludes. Readers primarily interested in the empirical results can skip Sections 3–5 with little loss of continuity.

2. REPRESENTATION AND ESTIMATION IN I(1) SYSTEMS

Let y_t denote a n-dimensional time series, whose elements are individually I(1). Suppose that $E(\Delta y_t)=0$, and that the $n\times r$ matrix of r cointegrating vectors is $\alpha=(-\theta,\ I_r)'$, where θ is the $r\times(n-r)$ submatrix of unknown parameters to be estimated and I_r is the $r\times r$ identity matrix. We assume that there are no additional restrictions on θ. The triangular representation for y_t is

(2.1a) $\quad \Delta y_t^1 = u_t^1,$

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

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. This representation has been used extensively in theoretical work by Phillips (1991a, 1991b), typically without parametric structure on the I(0) process u_t, and in applications by Campbell (1987) and Campbell and Shiller (1987, 1989) (also see Bewley (1979)). For the moment, u_t is assumed to be Gaussian to permit the development of the Gaussian MLE for θ.

The parameterization that forms the basis for the proposed estimators is obtained by making the error in (2.1b) independent of $\{u_t^1\}$, where $\{u_t^1\}$ denotes $\{u_t^1, t=0,\pm 1,\pm 2,\dots\}$. When u_t is Gaussian, stationary, and linearly regular, $E[u_t^2|\{\Delta y_t^1\}]=E[u_t^2|\{u_t^1\}]=d(L)\Delta y_t^1$, where $d(L)$ is two-sided in general. Thus (2.1b) can be written

(2.2) $\quad y_t^2 = \mu + \theta y_t^1 + d(L)\Delta y_t^1 + v_t^2$

where $v_t^2 = u_t^2 - E[u_t^2|\{u_t^1\}]$. By construction, $\{\Delta y_t^1\}$ and $\{v_t^2\}$ are independent.

The two-sided triangular representation (2.1a) and (2.2) suggests an unconventional factorization of the conditional Gaussian likelihood for a sample of size T. Assume: (i) the data are generated by (2.1), (ii) u_t in (2.1) is Gaussian and stationary with a bounded full rank spectral density matrix, (iii) $d(L)=\Sigma_{j=-\underline{q}}^{\bar{q}} d_j L^j$. The likelihood is conditioned on the required pre- and post-sample values of Δy_t^1 (i.e., $\Delta y_{-\underline{q}}^1,\dots,\Delta y_0^1$ and $\Delta y_{T+1}^1,\dots,\Delta y_{T+\bar{q}}^1$). Let λ_1 denote the parameters of the marginal distribution of $(u_1^1,\dots,u_T^1)$, let λ_2 denote μ and the parameters of $d(L)$ and of the marginal distribution of $(v_1^2,\dots,v_T^2)$, and let Y^i denote $(y_1^i,\dots,y_T^i)$, $i=1,2$. Then the likelihood can be factored as

(2.3) $\quad f(Y^1,Y^2|\theta,\lambda_1,\lambda_2)=f(Y^2|Y^1,\theta,\lambda_2)f(Y^1|\lambda_1).$

This differs from the usual prediction-error factorization because the conditional mean of y_t^2 involves future as well as past values of y_t^1. Assumptions (i) and (ii) allow the triangular representation (2.1) to be represented by its two-sided analogue, (2.2).

The representations (2.2) and (2.3) provide a framework for estimation and inference in these Gaussian systems. If there are no restrictions between λ_1 and $[\theta,\lambda_2]$, then Y^1 is ancillary (in Engle, Hendry, and Richard's (1983) terminology, weakly exogenous, extended to permit conditioning on both leads and lags of Δy_t^1) for θ, so that inference can be carried out conditional on Y^1. In this case, the MLE of θ (conditional on the initial and terminal values) can be obtained by maximizing $f(Y^2|Y^1, \theta, \lambda_2)$. This reduces to estimating the parameters of the regression equation (2.2) by GLS. Because the regressor y_t^1 is I(1), as is shown in Section 4 an asymptotically equivalent estimator of θ can be obtained by estimating θ in (2.2) by OLS; this will be referred to as the dynamic OLS (DOLS) estimator, to distinguish it from the static OLS (SOLS) estimator obtained by regressing y_t^2 on $(1, y_t^1)$. Similarly, the feasible GLS estimator of θ in (2.2) will be referred to as the dynamic GLS (DGLS) estimator.

The representation (2.2) warrants three remarks. First, Sims' (1972) Theorem 2 implies that the projection $d(L)\Delta y_t^1$ will involve only current and lagged values of Δy_t^1 if (and only if) u_t^2 does not Granger-cause Δy_t^1. If so, and if v_t^2 has a finite order autoregressive representation, then (2.2) can be rewritten as an r-dimensional error correction model, i.e. as a regression of Δy_t^2 onto $(\Delta y_t^1, \Delta y_{t-1}, \Delta y_{t-2}, \dots, \Delta y_{t-p}, y_{t-1}^2 - \theta' y_{t-1}^1)$. In this case, the nonlinear least squares estimator of θ (with Δy_t^1 included as a regressor; see Stock (1987)) is the Gaussian MLE.

Second, the large-sample properties of the OLS and GLS estimators of θ are readily deduced from the representation (2.2). Because v_t^2 is uncorrelated with the regressors at all leads and lags, conditional on Y^1 the GLS estimator has a normal distribution and the Wald statistic testing the hypothesis that $\theta = \theta_0$ has a χ^2 distribution. Because y_t^1 is I(1), the conditional covariance matrix of the GLS estimator differs across realizations of Y^1, even in large samples; thus unconditionally the GLS estimator of θ has a large-sample distribution that is a random mixture of normals and the Wald statistic has a χ^2 distribution. Phillips (1991a) and Saikkonen (1991) provide insightful discussions of the asymptotic mixed normal property of the MLE of θ and the local asymptotic mixed normal (LAMN) behavior of test statistics. Also note that results apply even if some rows of θ are equal to zero, so that the corresponding elements of y_t^2 are I(0).

Third, although the interpretation of (2.2) as a factorization of the likelihood (2.3) assumes Gaussianity, a two-sided triangular representation with $Ev_t^2 \Delta y_s^{1'} = 0$ for all t and d can be constructed under weaker conditions as discussed in the next section.

3. REPRESENTATION IN I(d) SYSTEMS

This section extends the framework of Section 2 to systems with maximum order of integration d and with polynomial time trends. First, a linear triangular representation for the n-dimensional time series y_t is derived for the general I(d) case under general conditions on the Wold representation for y_t and the error distribution. This representation is then used to motivate simple OLS and GLS estimators of cointegrating vectors. Properties of these estimators and test statistics based on the estimators are the subject of Section 4.

The maximum order of integration of any element of y_t is assumed to be d. The process y_t is assumed to have the representation

$$\Delta^d y_t = \mu + F(L)\varepsilon_t, \tag{3.1}$$

where $\Delta^d = (1-L)^d$ is the dth differencing operator. The shocks ε_t and the matrix lag polynomial $F(L)$ are assumed to satisfy the following assumption.

ASSUMPTION A: (i) $\{\varepsilon_t\}$ *is a n-dimensional martingale difference sequence, with* $E(\varepsilon_t\varepsilon_t'|\varepsilon_{t-1},\varepsilon_{t-2},\dots)=I_n$;

(ii) $F(L)=\sum_{j=0}^{\infty}F_jL^j$, *where* $F(L)$ *is k-summable* (*that is,* $\sum_{j=0}^{\infty}j^k|F_j|<\infty$, *where* $|A|=\max_{ij}|A_{ij}|$ *for a matrix* A) *for some* $k\geqslant d$;

(iii) $F(e^{i\omega})$ *is nonsingular for* $\omega\neq 0 \pmod{2\pi}$;

(iv) $\operatorname{rank}(F(1))=k_1$, $0<k_1<n$; *and*

(v) *among all linear combinations of* y_t *and its differences which include at least one element of* y_t *in levels, the lowest order of integration is zero.*

The triangular representation for an I(d) process satisfying (3.1) and Assumption A is derived in Appendix A and is:

$$
\begin{aligned}
\text{(3.2)}\qquad \Delta^d y_t^1 &= \mu_{1,0}+u_t^1,\\
\Delta^{d-1}y_t^2 &= \mu_{2,0}+\mu_{2,1}t+\theta_{2,1}^{d-1}\left(\Delta^{d-1}y_t^1\right)+u_t^2,\\
\Delta^{d-2}y_t^3 &= \mu_{3,0}+\mu_{3,1}t+\mu_{3,2}t^2\\
&\quad +\theta_{3,1}^{d-1}\left(\Delta^{d-1}y_t^1\right)+\theta_{3,1}^{d-2}\left(\Delta^{d-2}y_t^1\right)+\theta_{3,2}^{d-2}\left(\Delta^{d-2}y_t^2\right)+u_t^3,\\
&\;\;\vdots\\
y_t^{d+1} &= \sum_{j=0}^{d}\mu_{d+1,j}t^j+\sum_{j=1}^{d}\sum_{i=j}^{d}\theta_{d+1,j}^{d-i}\left(\Delta^{d-i}y_t^j\right)+u_t^{d+1},
\end{aligned}
$$

where

$$u_t=H(L)\varepsilon_t$$

for $t=1,\dots,T$, where $u_t=(u_t^{1\prime},u_t^{2\prime},\dots,u_t^{d+1\prime})'$ and where y_t^j, $j=1,\dots,d+1$, is a $k_j\times 1$ vector, chosen so that y_t can be partitioned as $y_t=(y_t^{1\prime},y_t^{2\prime},\dots,y_t^{d+1\prime})'$. In addition, $H(L)=\sum_{j=0}^{\infty}H_jL^j$, $H(e^{i\omega})$ is nonsingular for all ω, and $H(L)$ is $k-d$ summable for k defined in Assumption A(ii).

Assumption A(iv) ensures that there are at least $n-k_1$ cointegrating vectors in the system. Assumption A(v) serves to fix d, and is made without loss of generality. In practice, A(v) can be achieved by redefining y_t to be $\Delta^{-1}y_t$ or Δy_t as needed for u_t to be I(0) and not cointegrated. Finally, the normalization $E(\varepsilon_t\varepsilon_t')=I_n$ is made without loss of generality because F_0 is not restricted to be diagonal. The assumption of conditional homoskedasticity is made for convenience and could be weakened to admit conditional heterskedasticity; see, for example, Phillips and Solo (1992).

Note that not all elements of y_t need to be I(d) for (3.2) to apply (see the examples in Section 5 and the empirical application to long-run money demand in Section 7). Moreover, some blocks of (3.2) might not appear. For example, with $d=2$ and $n=2$, if y_t is CI(2,2) in Engle and Granger's (1987) terminology, then $k_1=1$, $k_2=0$, and $k_3=1$.

The triangular representation (3.2) partitions y_t into components corresponding to stochastic trends of different orders. Abstracting from the deterministic components, y_t^1 is a k_1-vector corresponding to the k_1 I(d) stochastic trends in the system. In the second block of k_2 equations, $y_t^2-\theta_{2,1}^{d-1}y_t^1$ corresponds to the

k_2 I$(d-1)$ stochastic trends; for rows of $\theta_{2,1}^{d-1}$ which equal zero, y_t^2 is I$(d-1)$, while for nonzero rows of $\theta_{2,1}^{d-1}$, y_t^2 is I(d) and (y_t^1, y_t^2) are CI$(d,1)$. The k_3 equations in the third block describe the I$(d-2)$ components, and so forth. It is straightforward to generalize the representation (3.2) to include higher order polynomials in t, or to specialize it to the leading case in which higher-order polynomials do not appear.

As in the I(1) case, we orthogonalize the errors in (3.2) by projecting onto leads and lags of the errors in the preceding equations. This amounts to premultiplying u_t by an appropriate lower triangular matrix lag polynomial $D(L)$ which is in general two-sided. Let $D(L)H(L) = C(L)$, partitioned conformably with u_t, and let $v_t = D(L)u_t$. Then, since $u_t = H(L)\varepsilon_t$, $v_t = D(L)u_t = D(L)H(L)\varepsilon_t = C(L)\varepsilon_t$, that is,

$$(3.3) \qquad v_t = \begin{bmatrix} I & 0 & \cdots & 0 \\ -d_{21}(L) & I & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ -d_{d+1,1}(L) & -d_{d+1,2}(L) & \cdots & I \end{bmatrix} \begin{bmatrix} u_t^1 \\ u_t^2 \\ \vdots \\ u_t^{d+1} \end{bmatrix}$$

$$= \begin{bmatrix} c_1(L) \\ c_2(L) \\ \vdots \\ c_{d+1}(L) \end{bmatrix} \varepsilon_t .$$

The matrix $D(L)$ is chosen so that the cross spectrum between v_t^l and v_t^m, $c_l(e^{-i\omega})c_m(e^{i\omega})'$, is zero for all $l \neq m$. (Because $H(e^{i\omega})$ is nonsingular and absolutely summable, Brillinger's (1980) Theorem 8.3.1 guarantees that such a $D(L)$ filter can be constructed.) The matrix polynomials $\{d_{lm}(L)\}$ generalize $d(L)$ in (2.2) and are constructed from the projection of u_t^l onto $\{u_t^m\}$ for $m = 1, \ldots, l-1$. For example, $d_{21}(L) = h_2(L)h_1(L^{-1})'[h_1(L)h_1(L^{-1})']^{-1}$, where the rows of $H(L)$ are partitioned conformably with u_t, so

$$H(L) = [h_1(L)', h_2(L)', \ldots, h_{d+1}(L)']'.$$

More generally, $v_t^l = u_t^l - \text{Proj}(u_t^l | \{u_t^1, \ldots, u_t^{l-1}\})$, where $\text{Proj}(x_t | \{z_t\})$ denotes the linear projection of x_t onto $\{z_t\}$.

Substitution of the lth equation in (3.3) into (3.2) yields

$$(3.4) \qquad \Delta^{d-l+1} y_t^l = \sum_{j=0}^{l-1} \bar{\mu}_{l,j} t^j + \sum_{j=1}^{l-1} \sum_{i=j}^{l-1} \theta_{l,j}^{d-i} (\Delta^{d-i} y_t^j)$$

$$+ \sum_{m=1}^{l-1} d_{lm}(L) \left[(\Delta^{d-m+1} y_t^m) - \sum_{j=1}^{m-1} \sum_{i=j}^{m-1} \theta_{m,j}^{d-i} (\Delta^{d-i} y_t^j) \right] + v_t^l$$

where $\{\bar{\mu}_{l,j},\ j=0,\ldots,l-1\}$ are functions of $\{d_{lm}(L), \mu_{m,j},\ j=0,\ldots,m-1,\ m=1,\ldots,l\}$.[2]

The subspaces that cointegrate y_t^l with $(y_t^1,\ldots,y_t^{l-1})$ and its differences are determined by the matrices $\{\theta_{l,j}^{d-i}\}$ appearing in the second term on the right-hand side of (3.4). In general the lth block of equations contains all of the cointegrating vectors for $m<l$, which appear in the higher order error correction terms making up the third term on the right-hand side of (3.4). For example, in a system with $d=2$ the equations describing cointegration in the levels can contain cointegrating relations between the first differences.

We consider estimators of the cointegrating coefficients appearing in the lth block of (3.2). Because the errors $\{v_t^l\}$ in (3.4) are uncorrelated with the variables on the right-hand side of (3.4), (3.4) constitutes a correctly specified regression equation. We therefore consider estimators based on this regression equation. To construct a feasible estimator, we will assume that $d_{lm}(L)$, $m=1,\ldots,l-1$ are finitely parameterized. Specifically, we make the following assumption.

Assumption B: *$d_{lm}(L)=\Sigma_{j=-\underline{q}_{lm}}^{\bar{q}_{lm}} d_{lm,j}L^j$, where $\underline{q}_{lm}$ and $\bar{q}_{lm}$ are finite and known, for $m=1,\ldots,l-1$.*

Under Assumption B, the block of k_l equations in (3.4) can be rewritten as

$$\Delta^{d-l+1}y_t^l=\left(x_t'\otimes I_{k_l}\right)\beta+v_t^l \tag{3.5}$$

where x_t is the vector of regressors in (3.4) and β is the vector of the stacked unknown regression coefficients in (3.4). Specifically, β consists of the elements of $\bar{\mu}_{l,j}$, $j=0,\ldots,l-1$; $\theta_{l,j}^{d-i}$, $j=1,\ldots,l-1$; and $d_{lm,j}$, $m=1,\ldots,l-1$, $j=-\underline{q}_{lm},\ldots,\bar{q}_{lm}$. It is assumed that x_t is a linear combination of $Y^1,Y^2,\ldots,Y^{l-1}$ with known nonrandom weights, where $Y^i=(y_1^i,y_2^i,\ldots,y_T^i)$, $i=1,\ldots,d+1$. In this notation, $Ev_t^lx_\tau'=0$ for all t and τ.

Because the regressors in (3.5) in general have stochastic or deterministic trends in common, they are asymptotically multicollinear. To obtain nondegenerate asymptotic results, the regressors are transformed to isolate these different trends. This is accomplished by defining $z_t=Bx_t$, where B is an invertible matrix of constants (possibly unknown), chosen so that z_t are the canonical regressors of Sims, Stock, and Watson (1990). (The choice of transformation matrix B depends on the specific application.) Partition z_t as $(z_t^1, z_t^{2\prime},\ldots,z_t^{2l\prime})'$, where by construction z_t^1 is an I(0) vector with mean zero (z_t^1 contains the required leads and lags of $\{u_t^m,\ m<l\}$, dictated by the polynomials $\{d_{lm}(L)\}$), $z_t^2=1$, z_t^3 is I(1), $z_t^4=t$, z_t^5 is I(2), $z_t^6=t^2$, and so forth. In general $\Sigma_{t=1}^T z_t^i z_t^{i\prime}$ is $O_p(T^{i-1})$ for $i\geqslant 2$. Using the approach in Sims, Stock, and Watson (1990,

[2] Johansen (1988b, 1992) studied the restrictions on the coefficients of vector autoregressions implied by the existence of cointegration in higher order systems. Johansen (1988b) examined systems with, in Engle and Granger's (1987) terminology, cointegration of the form CI(d,b), where $d\geqslant b$. As Johansen (1992) points out, this excludes cointegration of the general form (3.4), which generalizes what Granger and Lee (1990) term "multicointegration." Johansen's (1992) results complement ours, since both explicitly handle multicointegration: Johansen (1992) relates multicointegration to restrictions on the parameters of the levels VAR, whereas we consider the moving average representation of the dth difference.

Section 2), write z_t as $z_t = G(L)\nu_t$, where $G(L)$ is a block lower triangular matrix and $\nu_t = (\xi_t^{0\prime}, 1, \xi_t^{1\prime}, t, \xi_t^{2\prime}, \dots, \xi_t^{l-1\prime}, t^{l-1})'$, where $\xi_t^0 = \varepsilon_t$ and where ξ_t^j is defined recursively by $\xi_t^j = \Sigma_{s=1}^{t} \xi_s^{j-1}$ for $j \geqslant 1$. Also, let g_i denote the dimension of z_t^i, and let $g = \Sigma_{i=1}^{2l} g_i$ be the dimension of z_t.

With these definitions, the system (3.5) can be rewritten,

$$(3.6) \qquad \Delta^{d-l+1} y_t^l = (z_t' \otimes I_{k_l})\delta + v_t^l$$

where $E(v_t^l z_\tau') = 0$ for all t and τ. The regression coefficients β in (3.5) are related to the coefficients δ in the transformed regression (3.6) by $\beta = (B' \otimes I_{k_l})\delta$. Because the parameters of interest (the cointegrating parameters) are the coefficients on the integrated elements of z_t, it is convenient to partition the gk_l-vector δ as $\delta = (\delta_1', \delta_2', \dots, \delta_{2l}')'$, where δ_i is the $g_i k_l$-vector of coefficients on z_t^i.

4. ESTIMATION AND TESTING

This section examines the least squares estimation of the parameters δ in (3.6). There are two natural estimators of δ. Because the regressors z_t are uncorrelated with the errors v_t^l, the first estimator is the OLS estimator in the dynamic regression (3.6) (the DOLS estimator), which is

$$(4.1) \qquad \hat{\delta}_{\text{OLS}} = \left[\left(\sum_t z_t z_t'\right) \otimes I_{k_l}\right]^{-1} \left[\sum_t (z_t \otimes I_{k_l})(\Delta^{d-l+1} y_t^l)\right]$$

where these and subsequent summations run over the sample used for the regression, leaving sufficient observations for initial and terminal conditions for the leads and lags of the data in z_t.

Because the error term in (3.6), $v_t^l = c_l(L)\varepsilon_t$, is serially correlated and uncorrelated with the regressors at all leads and lags, a second natural estimator is the feasible GLS estimator of the dynamic regression (3.6) (the DGLS estimator). Let $\Phi(L)$ be a $k_l \times k_l$ lag polynomial such that

$$\Phi(L) c_l(L) c_l(L^{-1})' \Phi(L^{-1})' = I_{k_l},$$

and let $\hat{\Phi}(L)$ be an estimator of $\Phi(L)$. Then the DGLS estimator is

$$(4.2) \qquad \hat{\delta}_{\text{GLS}} = \left[\sum_t \tilde{z}_t \tilde{z}_t'\right]^{-1} \left[\sum_t \tilde{z}_t (\Delta^{d-l+1} \tilde{y}_t^l)\right]$$

where $\tilde{z}_t = [\hat{\Phi}(L)(z_t' \otimes I_{k_l})]'$ and $\tilde{y}_t^l = \hat{\Phi}(L) y_t^l$. In principle $\Phi(L)$ can always be constructed as the inverse Fourier transform of the inverse of the Cholesky factor of $c_l(e^{-i\omega}) c_l(e^{i\omega})'$. In general this will yield a two-sided polynomial $\Phi(L)$. In practice, a simple strategy is to model $\Phi(L)$ as being one-sided and finite order and this is the case studied in the formal analysis of the DGLS estimator in this section.

Associated with the DGLS estimator is the Wald statistic testing the h restrictions $R\delta = r$ (where R and r have dimensions $h \times gk_l$ and $h \times 1$, respectively),

$$(4.3) \qquad W_{\text{GLS}} = (R\hat{\delta}_{\text{GLS}} - r)' \left[R\left(\sum_t \tilde{z}_t \tilde{z}_t'\right)^{-1} R'\right]^{-1} (R\hat{\delta}_{\text{GLS}} - r).$$

Because the disturbance in (3.6) is serially correlated, the Wald statistic for $\hat{\delta}_{\text{OLS}}$

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 (3.6), this is the spectral density matrix of v_t^l at frequency zero, $\Omega_{ll} = c_l(1)c_l(1)'$, estimated by (say) $\hat{\Omega}_{ll}$. That is,

$$W_{\text{OLS}} = \left[R\hat{\delta}_{\text{OLS}} - r\right]' \left\{ R\left[\left(\sum_t z_t z_t'\right)^{-1} \otimes \hat{\Omega}_{ll}\right] R' \right\}^{-1} \left[R\hat{\delta}_{\text{OLS}} - r\right]. \tag{4.4}$$

The next four theorems, proven in Appendix A, summarize the asymptotic distributions of these statistics. To prove these theorems, we strengthen somewhat Assumption A, which was used to derive the triangular representation.

Assumption C: (i) $\max_i \sup_t E(\varepsilon_{it}^4 | \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots) < \infty$. (ii) $F(L)$ *in* (3.1) *is* $d+2$ *summable.*

Assumption C(i) strengthens A(i) to include finite fourth moments, and Assumption C(ii) fixes k in A(ii) to be $k = d+2$. Let the matrix $G_{mm}(L)$ denote the mth diagonal block of $G(L)$ (where $z_t = G(L)\nu_t$) and let $\Gamma_m = G_{mm}(1)M'$, for $m = 3, 5, \ldots, 2l-1$, where M is a $(n-k_l) \times n$ matrix with rows that span the null space of the rows of $\Phi(1)c_l(1)$ and $MM' = I_{n-k_l}$. Also let $[\cdot]$ denote the greatest lesser integer function and "$\Rightarrow$" denote weak convergence on $D[0,1]$. Finally, define the $g \times g$ scaling matrix Υ_T to be a block diagonal matrix partitioned conformably with z_t, with diagonal blocks $\Upsilon_{1T} = T^{1/2}I_{g_1}$ and $\Upsilon_{iT} = T^{(i-1)/2}I_{g_i}$ for $i \geqslant 2$.

Theorem 1: *Suppose that y_t has the representation* (3.1), *that Assumptions A, B, and C hold, that $\Phi(L)$ is one-sided with known finite order q, and that $\hat{\Phi}(L)$ is a consistent estimator of $\Phi(L)$ (that is, $\hat{\Phi}_i \xrightarrow{p} \Phi_i$, $i = 0, \ldots, q$). Then $(\Upsilon_T \otimes I_{k_l})(\hat{\delta}_{GLS} - \delta) \Rightarrow Q^{-1}\phi$, where after partitioning Q and ϕ conformably with δ:*

$$Q_{11} = E\bar{z}_t^1 \bar{z}_t^{1\prime}, \quad \textit{where} \quad \bar{z}_t^1 = \left[\Phi(L)\left(z_t^{1\prime} \otimes I_{k_l}\right)\right]', \quad Q_{1j} = Q_{j1}' = 0, \quad j \geqslant 2,$$

and

$$Q_{ij} = V_{ij} \otimes \Omega_{ll}^{-1} \quad \textit{for} \quad i, j \geqslant 2, \quad \textit{where} \quad V_{22} = 1,$$

$$V_{mp} = \Gamma_m \left[\int_0^1 W_1^{(m-1)/2}(s) W_1^{(p-1)/2}(s)' \, ds\right] \Gamma_p',$$
$$m = 3, 5, 7, \ldots, 2l-1; \; p = 3, 5, 7, \ldots, 2l-1,$$

$$V_{mp} = G_{mm}(1)\left[\int_0^1 s^{(m-2)/2} W_1^{(p-1)/2}(s)' \, ds\right] \Gamma_p' = V_{pm}',$$
$$m = 2, 4, 6, \ldots, 2l; \; p = 3, 5, 6, \ldots, 2l-1,$$

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

$$\phi_1 \sim N\left(0, E\left[\bar{z}_t^1 \bar{z}_t^{1\prime}\right]\right),$$

$$\phi_m = \int_0^1 \left(G_{mm}(1) s^{(m-2)/2} \otimes \Phi(1)'\right) dW_2(s), \quad m = 2, 4, 6, \ldots, 2l,$$

$$\phi_m = \int_0^1 \left(\Gamma_m W_1^{(m-1)/2}(s) \otimes \Phi(1)'\right) dW_2(s), \quad m = 3, 5, 7, \ldots, 2l-1,$$

where W_1 *and* W_2 *are independent standard Wiener processes of dimension* $\Sigma_{m=1}^{l-1} k_m$ *and* k_l, *respectively, where* $W_i^{(m)}(\tau) = \int_0^\tau W_i^{(m-1)}(s)\,ds$, $m = 2, 3, \ldots, g$ *initialized with* $W_i^1 = W_i$ *for* $i = 1, 2$, *and where* ϕ_1 *is independent of* W_1 *and* ϕ_m, $m > 1$.

THEOREM 2: *Suppose that* y_t *has the representation* (3.1) *and that Assumptions* A, B, *and* C *hold. Then*:

(a) $(\Upsilon_T \otimes I_{k_l})(\hat{\delta}_{\mathrm{OLS}} - \delta) \Rightarrow [V^{-1} \otimes I_{k_l}]\omega$, *where after partitioning* V *and* ω *conformably with* δ:

$$\omega_1 \sim N(0, \Sigma_{\omega_1}), \quad \textit{where} \quad \Sigma_{\omega_1} = \sum_{j=-\infty}^{\infty} \left[E\left(z_t^1 z_{t-j}^{1\prime}\right) \otimes E\left(v_t^l v_{t-j}^{l\prime}\right)\right],$$

$$\omega_m = \int_0^1 \left(G_{mm}(1) s^{(m-2)/2} \otimes \Omega_{ll}^{1/2}\right) dW_2(s), \quad m = 2, 4, 6, \ldots, 2l,$$

$$\omega_m = \int_0^1 \left(\Gamma_m W_1^{(m-1)/2}(s) \otimes \Omega_{ll}^{1/2}\right) dW_2(s), \quad m = 3, 5, 7, \ldots, 2l-1,$$

where ω_1 *is independent of* W_1 *and* ω_m, $m > 1$, *and where* $V = [V_{ij}]$, $i, j = 1, 2, \ldots, 2l$, *where* $V_{11} = E(z_t^1 z_t^{1\prime})$, $V_{1j} = 0$, $j \geqslant 2$, *and* V_{ij}, $i, j > 2$ *are given in Theorem* 1.

(b) *Partition* $\delta = (\delta_1', \delta_*')$ *so that* δ_1 *denotes the* $g_1 k_l$ *elements of* δ *corresponding to* z_t^1 *and* δ_* *corresponds to the remaining* $(g - g_1)k_l$ *elements of* δ. *Similarly partition* $\hat{\delta}_{OLS}, \hat{\delta}_{GLS}, z_t = (z_t^{1\prime}, z_t^{*\prime})'$, *and* $\Upsilon_T = \mathrm{diag}(\Upsilon_{1T}, \Upsilon_{*T})$. *If in addition* $\Phi(L)$ *and* $\hat{\Phi}(L)$ *satisfy the assumptions of Theorem* 1, *then* $(\Upsilon_{*T} \otimes I_{k_l})(\hat{\delta}_{*OLS} - \hat{\delta}_{*\mathrm{GLS}}) \xrightarrow{p} 0$.

THEOREM 3: *Under the conditions of Theorem* 1, $W_{GLS} \Rightarrow \chi_h^2$.

THEOREM 4: *Suppose that the conditions of Theorem* 2 *hold, that the first* $g_1 k_l$ *columns of* R *equal zero, and that* $\hat{\Omega}_{ll} \xrightarrow{p} \Omega_{ll}$. *Then* $W_{\mathrm{OLS}} \Rightarrow \chi_h^2$. *If in addition the conditions of Theorem* 1 *are satisfied, then* $W_{\mathrm{OLS}} - W_{\mathrm{GLS}} \xrightarrow{p} 0$.

Note that $\Phi(L)$ must be finitely parameterized to implement the DGLS estimator. Although this is not strictly needed to compute the DOLS estimator, $\Omega_{ll} = c_l(1)c_l(1)'$ must be consistently estimated to construct W_{OLS}, which in practice entails estimating a parametric or truncated approximation to Ω_{ll}.

The asymptotic equivalence of the DOLS and DGLS estimators of δ_* (Theorem 2(b)) is a consequence of the trends in z_t: for $m > 2$ the GLS-transformed regressors are asymptotically collinear with their untransformed counterparts. This result extends the familiar result for the case of a constant and polynomial time trend (Grenander and Rosenblatt (1957)). Results similar to Theorems 1–4 are obtained by Phillips and Park (1988) for single equation static OLS regressions with strictly exogenous I(1) regressors.

In practice, the coefficients of interest usually are the original coefficients β in (3.5) rather than δ. The distribution of $\hat{\beta}_{\mathrm{GLS}}$ is obtained using $(\hat{\beta}_{\mathrm{GLS}} - \beta) = (B' \otimes I_{k_l})(\hat{\delta}_{\mathrm{GLS}} - \delta)$, and similarly for $\hat{\beta}_{\mathrm{OLS}}$ (recall that $\beta = (B' \otimes I_{k_l})\delta$). Moreover, the Wald statistic testing $R\delta = r$ equivalently tests $P\beta = r$, where $P = R(B'^{-1} \otimes I_{k_l})$. Theorem 3 implies that W_{GLS} is asymptotically χ^2 for all R, so the Wald statistic testing $P\beta = r$ is asymptotically χ^2 for all P. When $P\beta = r$ places no restrictions on coefficients that can be written as coefficients on mean-zero stationary regressors, Theorem 4 implies that the Wald test of $P\beta = r$ based on $\hat{\beta}_{\text{OLS}}$ (with an autocorrelation-robust covariance matrix) is asymptotically χ^2. Importantly, this result, that the Wald statistic testing restrictions on cointegrating vectors is asymptotically χ^2, 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 specific examples in the I(1) case, see West (1988) and Hansen (1989).

These theorems apply to models with a fixed number of regressors (Assumption B). Conceptually, one could view these estimators as semiparametric by embedding the 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 not undertaken here. In the $d = 1$ case, this result was obtained by Saikkonen (1991) who showed that, if the number of included leads and lags grows at rate T^{λ}, where $0 < \lambda < 1/3$, then the model misspecification induced by the truncation of the $d_{lm}(L)$ polynomials vanishes asymptotically so these theorems continue to hold. He also demonstrated that in the $d = 1$ model, the DOLS and DGLS estimators are asymptotically efficient and asymptotically equivalent to the Johansen (1988a)/Ahn and Reinsel (1990) Gaussian MLE's constructed from a vector error correction model. For additional discussion of efficiency in the $d = 1$ case, see Phillips (1991a).

5. EXAMPLES

The following examples explore specification and inference with DOLS and DGLS in I(2) systems. To simplify exposition, all deterministic terms are omitted and their coefficients are taken to be zero. From (3.2), the general I(2) model is

$$\Delta^2 y_t^1 = u_t^1, \tag{5.1a}$$

$$\Delta y_t^2 = \theta_{2,1}^1 \Delta y_t^1 + u_t^2, \tag{5.1b}$$

$$y_t^3 = \theta_{3,1}^1 \Delta y_t^1 + \theta_{3,1}^0 y_t^1 + \theta_{3,2}^0 y_t^2 + u_t^3. \tag{5.1c}$$

Some of the θ's can have rows of zero, or be zero, and the second block of equations might not be present at all (if so, $k_2 = 0$). These possibilities are examined here by considering a series of special cases with $k_2 = 0$ or $k_2 = 1$; more general cases can be analyzed by combining these special cases.

Case 1: $k_2 = 0$. Then (5.1b) does not appear in the system and y_t^2 does not enter (5.1c). Elements of y_t^3 corresponding to rows of zeros in $\theta_{3,1}^0$ and $\theta_{3,1}^1$ are I(0) (these variables do not enter any cointegrating relations), those corresponding to rows of zeros in $\theta_{3,1}^0$ but not $\theta_{3,1}^1$ are I(1), and the remaining elements are I(2). The dynamic OLS and GLS estimators of $(\theta_{3,1}^1, \theta_{3,1}^0)$ are asymptotically efficient and inference is χ^2.

Case 2: $k_2 = 1$, $\theta_{2,1}^1$ *known*. Then the estimation equation (3.4) becomes

$$y_t^3 = \theta_{3,1}^1 \Delta y_t^1 + \theta_{3,1}^0 y_t^1 + \theta_{3,2}^0 y_t^2 + d_{31}(L)\Delta^2 y_t^1 + d_{32}(L)\left(\Delta y_t^2 - \theta_{2,1}^1 \Delta y_t^1\right) + v_t^3 \tag{5.2}$$

where $Ev_t^3 u_s^{1\prime}$ and $Ev_t^3 u_s^{2\prime}$ are zero for all t, s. Because $\theta_{2,1}^1$ is known, the regressors $\Delta^2 y_t^1, \Delta y_t^2 - \theta_{2,1}^1 \Delta y_t^1$, and their leads and lags are I(0) with mean zero, so these comprise z_t^1. Because y_t^1 and y_t^2 are CI(2,1), we can set $z_t^3 = (\Delta y_t^{1\prime}, y_t^2 - \theta_{2,1}^1 y_t^1)'$ and $z_t^5 = y_t^1$ (other assignments of z_t are possible but they produce the same distributional results). The coefficients on z_t^3 and z_t^5 are respectively $\delta^3 = (\delta_1^{3\prime}, \delta_2^3)' = (\theta_{3,1}^{1\prime}, \theta_{3,2}^0)'$ and $\delta^5 = \theta_{3,1}^0 + \theta_{2,1}^1 \theta_{3,2}^0$, so $\hat{\theta}_{3,1}^1 = \hat{\delta}_1^3$, $\hat{\theta}_{3,2}^0 = \hat{\delta}_2^3$, and $\hat{\theta}_{3,1}^0 = \hat{\delta}^5 - \theta_{2,1}^1 \hat{\delta}_2^3$. Because $(\hat{\delta}^3, \hat{\delta}^5)$ converge at rates (T, T^2), $\hat{\theta}_{3,1}^1$, $\hat{\theta}_{3,2}^0$, and $\hat{\theta}_{3,1}^0$ individually converge at the rate T. Jointly, $(\hat{\theta}_{3,1}^0 + \theta_{2,1}^1 \hat{\theta}_{3,2}^0, \hat{\theta}_{3,2}^0, \hat{\theta}_{3,1}^1)$ converge at rates (T^2, T, T). The estimators are asymptotically efficient and inference is χ^2. These results hold for $\theta_{2,1}^1$ known, whether or not $\theta_{2,1}^1 = 0$.

Case 3: $k_2 = 1$, $\theta_{2,1}^1$ *unknown*. In this case there are cross-equation restrictions between (5.1b) and (5.1c) so that in general the DOLS and DGLS estimators are not efficient. Nonetheless, the dynamic OLS and GLS estimators have desirable properties. With $\theta_{2,1}^1$ unknown, $\Delta y_t^2 - \theta_{21}' \Delta y_t'$ cannot be used as a regressor and the estimation equation (5.2) becomes

$$y_t^3 = \left(\theta_{3,1}^1 - d_{32}(1)\theta_{2,1}^1\right)\Delta y_t^1 + d_{32}(1)\Delta y_t^2 + \theta_{3,1}^0 y_t^1 + \theta_{3,2}^0 y_t^2 + \left(d_{31}(L) - d_{32}^*(L)\theta_{2,1}^1\right)\Delta^2 y_t^1 + d_{32}^*(L)\Delta^2 y_t^2 + v_t^3 \tag{5.3}$$

where $d_{32}^*(L) = (1-L)^{-1}(d_{32}(L) - d_{32}(1))$. Because $\Delta^2 y_t^1$ is I(0) and $\Delta^2 y_t^2$ is either I(0) (if $\theta_{2,1}^1 \neq 0$) or I(-1) (if $\theta_{2,1}^1 = 0$), and because both have mean zero, their presence does not affect the asymptotic distribution of the other estimators and they will be ignored in this discussion. Whether or not $\theta_{2,1}^1 = 0$, a valid assignment of z_t is $z_t^1 = \Delta y_t^2 - \theta_{2,1}^1 \Delta y_t^1$, $z_t^3 = (\Delta y_t^{1\prime}, y_t^2 - \theta_{2,1}^1 y_t^1)'$, and $z_t^5 = y_t^1$. Evidently $\theta_{3,1}^1$ is not identified from (5.3) alone. Using a consistent estimator of $\theta_{2,1}^1$ from (5.1b) to estimate $\theta_{3,1}^1$ would in general result in loss of χ^2 inference (although the resulting estimator would be consistent). However, $\theta_{3,1}^0$ and $\theta_{3,2}^0$ are separately identified in (5.3) and individually converge at rate T. Together, the coefficients on (z_t^3, z_t^5) have an asymptotic mixed normal distribution. Moreover, the distribution of $(\hat{\theta}_{3,1}^0, \hat{\theta}_{3,2}^0)$ is the same as in Case 2, when the true value of $\theta_{2,1}^1$ is known. Thus $(\hat{\theta}_{3,1}^0, \hat{\theta}_{3,2}^0)$ are asymptotically efficient even if $\theta_{2,1}^1$ is unknown, for general $\theta^1_{3,1}$. The exception to this is the special case of $\theta^1_{3,1}$ known to be zero, in which case Δy^1_t would not enter as a regressor in (5.2) were $\theta^1_{2,1}$ known. Even in this case, however, inference on $(\hat{\theta}^0_{3,1}, \hat{\theta}^0_{3,2})$ is χ^2.

6. MONTE CARLO RESULTS

This section summarizes a study of the sampling properties of seven estimators of cointegrating vectors in three bivariate probability models. The data were generated by the model:

(6.1a) $\Delta y^1_t = u^1_t,$

(6.1b) $y^2_t = \theta y^1_t + u^2_t,$

with $\Phi(L)u_t = \zeta_t$, $\Phi(L) = I_2 - \Phi L, \zeta_t$ NIID$(0, \Sigma_\zeta)$, where $u_t = (u^1_t, u^2_t)'$. The true drift in the series is zero. Because u_t follows a VAR(1), y_t follows a VAR(2). Under (6.1), $T(\hat{\theta} - \theta)$ is invariant to θ for all the estimators considered, so without loss of generality θ is set to zero. This design, or variants with moving average rather than autoregressive errors, forms the basis of several previous Monte Carlo studies of estimators of cointegrating vectors (Banerjee et al. (1986), Phillips and Loretan (1991), and Hansen and Phillips (1990)).

The six estimators considered are the static OLS estimator (SOLS; Engle and Granger (1987), Stock (1987)), the dynamic OLS estimator $\hat{\delta}_{\mathrm{OLS}}$ (DOLS), the dynamic GLS estimator $\hat{\delta}_{\mathrm{GLS}}$ (DGLS), the zero frequency band spectrum estimator of Phillips (1991b) (PBSR), the fully modified estimator of Phillips and Hansen (1990) (PHFM), and Johansen's (1988a) VECM maximum likelihood estimator (JOH). Two serial correlation-robust estimators of the covariance matrix of the DOLS estimator were considered, one using a weighted sum of the autocovariances of the errors (DOLS1), the second using an autoregressive spectral estimator (DOLS2). To make results invariant to initial conditions for the level of y_t, a constant was included in all estimation procedures. All of the estimators relied on lead and lag lengths that depended only on sample size. This makes it possible to examine the consequences of overparameterization and truncation bias without the complications which would arise with data dependent lag lengths. The details of the construction of the estimators are given in the notes to Table I.

The design (6.1) parsimoniously nests several important special cases. First (Case A), when all elements of Φ except Φ_{11} equal zero and Σ_ζ is diagonal, y^1_t is strictly exogenous in (6.1b) and SOLS is the MLE (except that the zero intercept is not imposed). In this case, all the efficient estimators are asymptotically equivalent to SOLS, although they estimate nuisance parameters that in fact are zero. Second (Case B), if the second column of Φ is zero, but $\Phi_{21} \neq 0$ or Σ_ζ is not diagonal or both, then SOLS is no longer the MLE and does not have an asymptotic mixed normal distribution, but the DOLS, DGLS, and JOH estimators are correctly specified and are asymptotically MLE's (again except for the estimation of some parameters which have true values zero). In this case, PBSR and PHFM are efficient if interpreted semiparametrically. Third (Case

C), for general Φ and Σ_ζ, JOH with one lag of Δy_t is the MLE and DOLS, DGLS, PBSR, and PHFM are asymptotically efficient when interpreted semiparametrically.

Results for Cases A, B, and C are reported in the respective panels of Table I for $T = 100$ and 300. Panel A verifies that the estimation of the nuisance parameters in the asymptotically efficient estimators does not substantially impair performance in the special case that OLS is the MLE. Panel B explores the performance of the estimators in 22 models in which DOLS, DGLS, and JOH are correctly specified. Even when $\phi_{21} = 0$, SOLS can have substantial bias; for example, for $T = 100$ and $\phi_{11} = -.90$, the 5%, 50%, and 95% points of the SOLS distribution are $-.001$, .076, and .196. The DOLS, DGLS, and JOH estimators eliminate this bias. The DOLS t statistics tend to have heavier tails than predicted by the asymptotic distribution theory, particularly when the regressor is positively autocorrelated. The PBSR and PHFM estimators tend to have biases comparable to SOLS, evident in Table I from the shift in the distribution of their t statistics. When this bias is small (for example when $\phi_{11} = \phi_{21} = 0$), their t statistics have approximately normal distributions.

Case C (Φ, Σ_ζ unrestricted) introduces two additional parameters, and it is beyond the scope of this investigation to explore all aspects of this case. Rather, it is examined by generating data from a model relevant to the empirical analysis in Section 7, specifically a bivariate model of log M1 velocity (v) and the commercial paper rate (r), estimated using annual data from 1904–1989 (earlier observations were used for initial lags) imposing a long-run interest semielasticity of .10. The data are discussed in Section 7. While this simple model does not provide a full characterization of these data—that is the subject of Section 7—it is a useful way to calibrate the Monte Carlo design so that it informs our subsequent empirical analysis. The estimated VAR(1) for the triangular system $(\Delta v_t, v_t - .10 r_t)$ is reported in panel C of Table I. The results for this system indicate large bias in SOLS and, to a lesser extent, in DGLS, PBSR, and PHFM. DOLS exhibits less bias and, not surprisingly because it is the (overparameterized) MLE in this system, JOH is essentially unbiased. The dispersion

TABLE I

MONTE CARLO RESULTS

A: $\Phi = \begin{bmatrix} .4 & 0 \\ .0 & 0 \end{bmatrix}, \quad \Sigma_\zeta = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

	$T = 100$					$T = 300$				
Estimator	Bias($\hat\theta$)	$\sigma(\hat\theta)$	$t_{.05}$	$t_{.95}$	$P(W > 3.84)$	Bias(θ)	$\sigma(\hat\theta)$	$t_{.05}$	$t_{.95}$	$P(W > 3.84)$
SOLS	.000	.021	-1.67	1.68	.054	.000	.007	-1.63	1.70	.052
DOLS1	.000	.023	-1.86	1.87	.083	.000	.007	-1.69	1.79	.062
DOLS2	.000	.023	-1.87	1.86	.087	.000	.007	-1.66	1.78	.061
DGLS	.000	.024	-1.80	1.76	.073	.000	.007	-1.64	1.75	.056
PBSR	.000	.021	-1.78	1.81	.073	.000	.007	-1.67	1.76	.060
PHFM	.000	.022	-1.88	1.88	.086	.000	.007	-1.71	1.81	.065
JOH	.000	.025	-1.98	1.96	.077	.000	.007	-1.84	1.67	.057

TABLE 1. (*Continued*)

B. $T = 100$, $\Phi = \begin{bmatrix} \Phi_{11} & .0 \\ \Phi_{21} & .0 \end{bmatrix}$, $\Sigma_\zeta = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix}$

		SOLS	DOLS1		DOLS2		DGLS		PBSR		PHFM		JOH	
Φ_{21}	Φ_{11}	bias($\hat{\theta}$)	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$
0.0	−.90	.084	−1.80	1.84	−1.84	1.84	−1.77	1.77	−1.46	1.69	−1.06	2.98	−1.95	1.83
	−.80	.092	−1.81	1.84	−1.85	1.86	−1.77	1.76	−1.49	1.74	−1.17	2.74	−1.95	1.83
	−.70	.089	−1.82	1.84	−1.84	1.85	−1.77	1.76	−1.52	1.77	−1.25	2.55	−1.95	1.83
	−.60	.081	−1.83	1.84	−1.85	1.84	−1.77	1.76	−1.56	1.78	−1.31	2.40	−1.94	1.84
	−.50	.071	−1.83	1.84	−1.84	1.84	−1.77	1.76	−1.58	1.79	−1.38	2.32	−1.95	1.86
	.00	.026	−1.85	1.83	−1.86	1.83	−1.77	1.77	−1.76	1.72	−1.67	2.05	−1.97	1.90
	.50	.000	−1.87	1.89	−1.90	1.87	−1.80	1.77	−2.05	1.46	−2.01	1.65	−1.99	2.01
	.60	−.002	−1.88	1.89	−1.89	1.89	−1.81	1.80	−2.15	1.34	−2.11	1.55	−1.97	2.01
	.70	−.003	−1.88	1.91	−1.91	1.91	−1.82	1.82	−2.25	1.25	−2.21	1.42	−1.96	2.05
	.80	−.003	−1.92	1.94	−1.91	1.94	−1.81	1.82	−2.36	1.13	−2.33	1.28	−1.97	2.04
	.90	−.002	−1.90	1.93	−1.93	1.94	−1.85	1.83	−2.45	1.04	−2.42	1.15	−1.99	2.02
0.8	−.90	−2.83	−1.80	1.84	−1.84	1.84	−1.77	1.77	−0.58	1.23	−3.80	0.69	−1.90	1.79
	−.80	−.078	−1.81	1.84	−1.85	1.85	−1.77	1.76	−0.77	1.46	−1.80	1.34	−1.90	1.79
	−.70	.007	−1.82	1.84	−1.84	1.85	−1.77	1.75	−0.86	1.62	−1.31	1.75	−1.91	1.79
	−.60	.048	−1.83	1.84	−1.85	1.84	−1.77	1.76	−0.94	1.73	−1.15	2.00	−1.91	1.79
	−.50	.068	−1.83	1.84	−1.84	1.84	−1.77	1.76	−0.98	1.82	−1.09	2.13	−1.92	1.79
	.00	.065	−1.85	1.83	−1.86	1.83	−1.77	1.77	−1.08	2.08	−1.09	2.28	−1.97	1.80
	.50	.028	−1.87	1.89	−1.89	1.87	−1.80	1.77	−1.14	2.18	−1.16	2.30	−2.00	1.84
	.60	.021	−1.88	1.89	−1.89	1.89	−1.81	1.80	−1.15	2.20	−1.18	2.30	−2.01	1.85
	.70	.015	−1.88	1.91	−1.91	1.91	−1.82	1.82	−1.17	2.19	−1.19	2.31	−2.00	1.87
	.80	.010	−1.92	1.94	−1.91	1.94	−1.81	1.82	−1.24	2.16	−1.23	2.30	−2.03	1.84
	.90	.005	−1.90	1.93	−1.93	1.94	−1.85	1.83	−1.45	2.18	−1.38	2.36	−1.98	1.87

C. $\Phi = \begin{bmatrix} -.103 & -.039 \\ -.062 & .643 \end{bmatrix}$, $\Sigma_\zeta = \begin{bmatrix} .951 & .499 \\ .499 & 1.374 \end{bmatrix} \times 10^{-2}$

	$T = 100$					$T = 300$				
Estimator	Bias($\hat{\theta}$)	$\sigma(\hat{\theta})$	$t_{.05}$	$t_{.95}$	$P(W > 3.84)$	Bias($\hat{\theta}$)	$\sigma(\hat{\theta})$	$t_{.05}$	$t_{.95}$	$P(W > 3.84)$
SOLS	.085	.120	−1.95	5.16	.466	.033	.045	−1.90	5.29	.483
DOLS1	.026	.125	−2.10	2.71	.188	.007	.041	−1.79	2.32	.118
DOLS2	.026	.125	−1.72	2.25	.111	.007	.040	−1.55	1.97	.071
DGLS	.045	.131	−1.52	2.35	.111	.012	.042	−1.43	2.08	.076
PBSR	.039	.123	−1.83	2.79	.180	.012	.041	−1.64	2.41	.122
PHFM	.041	.122	−1.91	3.01	.206	.011	.041	−1.69	2.46	.131
JOH	.003	.330	−2.40	2.07	.095	−.001	.044	−1.97	1.75	.064

Notes: Bias ($\hat{\theta}$) and $\sigma(\hat{\theta})$ are the Monte Carlo bias and standard deviation of $\hat{\theta}$, respectively. $t_{.05}$ and $t_{.95}$ are the empirical 5% and 95% critical values of the t ratios, and $P(W > 3.84)$ is the percent rejections at the asymptotic 5% level of the test statistic testing $\theta = \theta_0$ which, for all but JOH, is the square of the t statistic, and for JOH is the likelihood ratio statistic. 5000 Monte Carlo replications were used. The number of observations (100 and 300) refer to the span of the regressions; 100 startup observations, plus terminal conditions as needed, were also generated. All regressions included a constant in addition to the terms listed below. The estimators are:

SOLS—Static OLS regression of y_t^2 on y_t^1.

DOLS1—Dynamic OLS regression of y_t^2 on $(y_t^1, \Delta y_t^1, \Delta y_{t\pm 1}^1, \ldots, \Delta y_{t\pm k}^1)$, where $k = 2$ for $T = 100$, $k = 3$ for $T = 300$. The covariance matrix is estimated by averaging the first r error autocovariances using the Bartlett kernel, where $r = 5$ for $T = 100$, $r = 8$ for $T = 300$.

DOLS2—Same as DOLS1 except that the covariance matrix is estimated by an autoregressive spectral estimator with 2 lags for $T = 100$, 3 lags for $T = 300$.

DGLS—Dynamic GLS regression of y_t^2 on $(y_t^1, \Delta y_t^1, \Delta y_{t\pm 1}^1, \ldots, \Delta y_{t\pm k}^1)$, where $k = 2$ for $T = 100$, $k = 3$ for $T = 300$. The errors were modeled as an AR(2) for $T = 100$ and AR(3) for $T = 300$.

PBSR—Phillips (1991b) band spectral regression, where the spectral density at frequency zero was estimated using the Bartlett kernel with 5 lead and lags for $T = 100$ and 8 lead and lags for $T = 300$.

PHFM—Phillips-Hansen (1990) fully modified estimator using the Bartlett kernel with 5 lead and lags for $T = 100$ and 8 lead and lags for $T = 300$.

JOH—Johansen (1991) VECM MLE based on the estimated model $y_t = \gamma\alpha' y_{t-1} + \Sigma_{i=1}^{k} A_i \Delta y_{t-i} + a_t$, where $k = 4$ for $T = 100$ and $k = 6$ for $T = 300$.

The DOLS and DGLS standard errors were computed using a degrees-of-freedom adjustment, specifically df = number of periods in the regression minus number of regressors in the DGLS or DOLS regression minus number of autoregressive lags in the GLS transform (DGLS) or AR spectral estimator (DOLS). The JOH standard errors were computed as described in Watson (1992) with a degrees-of-freedom adjustment (df = number of periods in the regression minus number of regressors in a single equation of the VECM). The degrees-of-freedom corrections are motivated by analogy to the classical linear regression model. No adjustments were made for PBSR or PHFM.

of the distributions are comparable, except for the JOH estimator which has some large outliers for $T=100$. The χ^2 approximation to the Wald statistic (testing $\theta=.10$) works best for JOH, next best for DOLS2 and DGLS, less well for the remaining efficient estimators.

To interpret the DOLS and DGLS results, it is useful to write (6.1) in the triangular form (2.1a) and (2.2). Write the VAR(1) for u_t as $\Psi(L)u_t=\varepsilon_t$, where $\Psi(L)=\Sigma_\zeta^{-1/2}\Phi(L)$ and $\varepsilon_t=\Sigma_\zeta^{-1/2}\zeta_t$, so that $E(\varepsilon_t\varepsilon_t')=I$. Then Δy_t^1 has a univariate ARMA(2,1) representation, say $|\Psi(L)|\Delta y_t^1=\kappa(L)a_t^1$, where $\kappa(L)$ is the first degree polynomial with its root outside the unit circle that solves $\kappa(L)\Sigma_{11}\kappa(L^{-1})=\Psi_{22}(L)\Psi_{22}(L^{-1})+\Psi_{12}(L)\Psi_{12}(L^{-1})$. The projection of $y_t^2-\theta y_t^1$ onto $\{\Delta y_t^1\}$ is $d(L)\Delta y_t^1$; for this design $d(L)=-[\Psi_{21}(L)\Psi_{22}(L^{-1})+\Psi_{11}(L)\Psi_{12}(L^{-1})][\kappa(L)\kappa(L^{-1})]^{-1}$. Finally, the residual from this regression, v_t^2, follows the AR(1) model $\kappa(L)v_t^2=a_t^2$. Thus $\kappa(L)$ dictates both how quickly the coefficients on leads and lags of Δy_t^1 in the DOLS/DGLS regressions die out and the degree of serial correlation in the regression error. In Cases A and B, $\kappa(L)=1$, and the DOLS/DGLS regressions have no omitted variables. In Case C, $\kappa(L)\cong 1-.66L$ so the true $d(L)$ is infinite order but the DOLS/DGLS regressions include on $2(T=100)$ or $3(T=300)$ leads and lags of Δy_t^1.

The results from the experiments can be summarized as follows. First, SOLS is biased in almost all trials, with nonstandard distributions for the estimator and test statistics. Second, DOLS and DGLS are unbiased for Cases A and B, but exhibit bias in Case C, although this diminishes when the sample size increases. The relatively large root of $\kappa(L)$ suggests that the bias is attributable to the truncation of $d(L)$ in the DOLS/DGLS regressions.[3] Third, in results not shown in the table, doubling the number of leads and lags for DOLS and DGLS and the order of the AR correction for DGLS has little effect in Cases A and B and reduces the bias in Case C. However, doubling the number of lags and the AR order increased the dispersion of the DOLS and DGLS statistics. Fourth, the PBSR and PHFM bias has the same sign as, but is somewhat less than, the SOLS bias. A possible explanation is that both PBSR and PHFM rely on initial biased SOLS estimates of θ, which results in inaccurate spectral density estimates subsequently used to compute PBSR and PHFM. Fifth, for Case C (where the error is highly serially correlated) the autoregressive spectral estimator used in DOLS2 produces a more normally-distributed t statistic than does the kernel estimator used in DOLS1. Sixth, tripling the sample size noticeably improves the quality of the asymptotic approximations.

These results suggest four conclusions. First, each estimator (except the correctly-specified JOH) has substantial bias in at least some of the simulations, although the bias is in each case less than for SOLS: no single estimator is a panacea. Second, the distributions of the t ratios tend to be spread out relative to the normal distribution, suggesting that the usual confidence intervals will

[3] This interpretation is supported by an additional Monte Carlo experiment in which Δy_t^1 was replaced by $[\kappa(L)\kappa(L^{-1})]\Delta y_t^1$. (Of course in an empirical application $\kappa(L)$ would be unknown.) This eliminates nearly all of the bias: for $T=100$, the bias falls from .026 to $-.006$ for DOLS and from .045 to $-.008$ for DGLS.

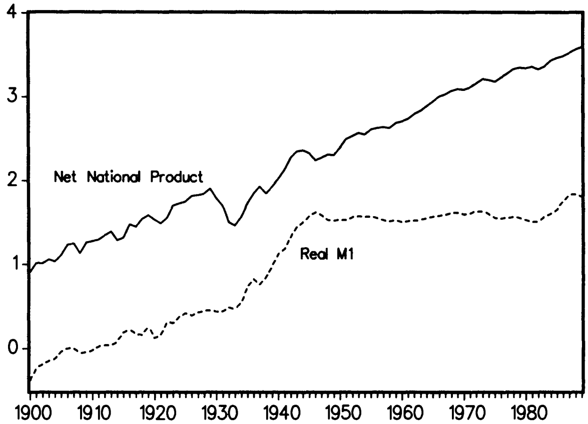
overstate precision. Third, in Case C each estimator has shortcomings: the DGLS, PBSR, and PHFM estimators are substantially biased, and the JOH estimator, while unbiased, has an empirical distribution with a much greater dispersion than the other efficient estimators; DOLS has the lowest RMSE. Fourth, of the autoregressive and kernel estimators used to compute the DOLS covariance matrix, t statistics based on the former have distributions closer to their asymptotic N(0, 1) approximation. The DOLS standard errors reported in the empirical analysis in Section 7 therefore are based on the autoregressive covariance estimator.

7. APPLICATION TO THE LONG-RUN DEMAND FOR MONEY IN THE U.S.

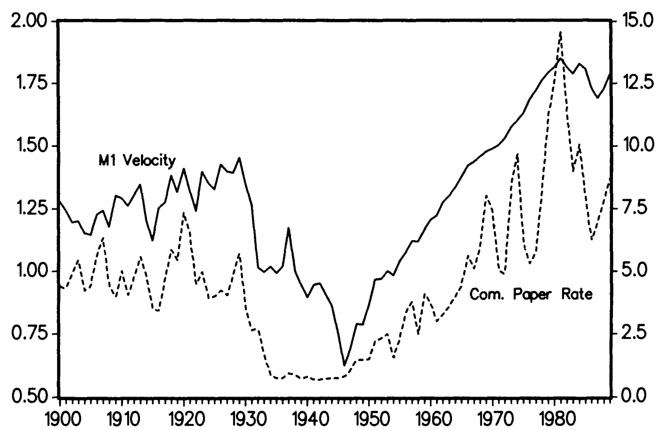
The long-run demand for money plays an important role in the quantitative analysis of the effects of monetary policy. Unfortunately, estimates of long-run income and interest elasticities obtained using postwar data have been sensitive to the sample period and specification (see the reviews by Laidler (1977), Judd and Scadding (1982), and Goldfeld and Sichel (1990)). In his review of this research and of early work by Meltzer (1963), Lucas (1988) presented informal but highly suggestive evidence that this apparent sensitivity resulted not from a breakdown of the prewar long-run M1 demand relation, but from the lack of low frequency variation in the postwar data. This section examines this interpretation using the econometric techniques for the analysis of cointegrating relations developed in this paper and elsewhere. Our analysis focuses on the annual data studied by Lucas (1988), extended to cover 1900–1989, although results for postwar monthly data are also presented to permit a comparison with other studies. This section addresses two questions. First, is there a stable long-run M1 demand equation spanning 1900–1989 in the United States? Second, what are the income elasticity and interest semielasticity, and how precisely are they estimated?[4]

A. *Results for annual data*. The annual time series are M1 (in logarithms, m), real net national product (in logarithms, y), the net national product price deflator (in logarithms, p), and the commercial paper rate (in percent at an annual rate, r). Data sources are given in Appendix B. Real M1 balances ($m-p$, plotted with y in Figure 1a) grew strongly over the first half of the century, but experienced almost no net growth over most of the postwar period. Over the entire period, velocity ($y-m+p$) and r (plotted in Figure 1b) exhibit

[4] Most empirical analyses of money demand predate the literature on cointegration. Exceptions are Hoffman and Rasche (1991), who apply Johansen's (1991) estimator to monthly U.S. M1 data from 1953 to 1987, and Johansen and Jeuslius (1990), who apply Johansen's procedure to the long-run demand for money in Denmark and Finland. Baba, Hendry, and Starr (1992) focus on short-run U.S. M1 demand (1960–1988, quarterly), but a preliminary step is their estimation of long-run M1 demand using a single equation error correction model (the "NLLS" estimator). With the same purpose and methodology, Hendry and Ericsson (1991) present results for the U.K. as well as the U.S. Related results for the U.S. are reported in Miller (1991). Hafer and Jansen (1991) investigate long-run money demand (M1 and M2) for the U.S. using Balke and Gordon's (1986) quarterly data and Johansen's estimator.



A. U.S. real net national product and real M1, in logarithms, 1900–1989.



B. Logarithm of U.S. M1 velocity (left scale), and short-term commercial paper rate (right scale), 1900–1989.

FIGURE 1

strikingly similar long-run trends, dropping from the 1920's to the 1930's, growing from 1950 to 1980, then declining after 1981.

Inspection of these figures suggest that real balances, output, velocity, and interest rates might be well modeled as being individually integrated, and formal tests summarized in Appendix B support this view. Specifically, Dickey-Fuller (1979) tests of one or two unit roots, augmented Dickey-Fuller tests for cointegration, and Stock-Watson (1988) tests of the number of unit roots in multivariate systems are consistent with the following specifications: $m-p$ is I(1) with drift; r is I(1) with no drift; y is I(1) with drift; and $(m-p)$, y, and r are cointegrated. The tests also suggest that $r-\Delta p$ is I(0). Whether p and m are individually I(1) or I(2) is unclear: the inference depends on the subsample and the test specification. Because r_t is nonnegative, characterizing r_t as I(1) raises conceptual difficulty. Our decision to do so is driven by the empirical evidence that r_t exhibits considerable persistence; in any event, this I(1) specification is consistent with interest rate specifications used by other researchers (e.g., Campbell and Shiller (1987) and Hoffman and Rasche (1991)).

The applicability of the DOLS and DGLS estimators to I(1) and I(2) systems makes it possible to estimate θ_p in the nominal M1 cointegrating relation, $m = \theta_p p - \theta_y y - \theta_r r$, and to test whether $\theta_p = 1$. Based on the foregoing characterizations of the integration and cointegration properties of these series, we consider three specifications. In each r_t is modeled as I(1) and $m_t - \theta_p p_t - \theta_y y_t - \theta_r r_t$ is modeled as I(0). First, if m and p are I(1), then (m, p, y, r) constitute the I(1) system analyzed in Section 2 with one cointegrating vector, extended to nonzero drifts, and inference on $(\theta_y, \theta_r, \theta_p)$ using DOLS or DGLS is χ^2. Second, if m and p are I(2) and $(r, \Delta p)$ are not cointegrated, then this is an I(2) system with $y_t^1 = p_t$, $y_t^2 = (y_t, r_t)'$, and $y_t^3 = m_t$, where $\theta_{2,1}^1 = 0$, $\theta_{3,1}^1 = 0$, $\theta_{3,2}^0 = (\theta_y \theta_r)$, and $\theta_{3,1}^0 = \theta_p$. This is Case 2 in Section 5 (with $\theta_{2,1}^1 = 0$), and inference is χ^2. Third, if p is I(2) and if, as the evidence suggests, the real rate $r - \Delta p$ is I(0), then this is an extension of Case 2 in Section 5, with $y_t^1 = p_t$, $y_t^2 = (y_t, \Delta^{-1} r_t)'$, $\theta_{2,1}^1 = (0,1)'$, $\theta_{3,1}^1 = \theta_r$, $\theta_{3,1}^0 = \theta_p$, and $\theta_{3,2}^0 = (\theta_y, 0)'$. Then $\Delta y_t^2 - \theta_{2,1}^1 \Delta y_t^1 = (\Delta y_t, r_t - \Delta p_t)'$. Following the discussion in Section 5, inference based on DOLS or DGLS is χ^2.

Estimates for the four-variable system are reported in Table II for these three specifications. The sample periods in Table II and subsequent tables refer to the dates over which the regression are run, with earlier and later observations used for initial and terminal leads and lags as necessary. The estimates of θ_p do not differ from 1 at the 10% (two-sided) level in any of the specifications. In all cases, θ_y is statistically indistinguishable from 1 at the 10% level. In most cases θ_y is imprecisely estimated, with standard errors in the range 0.12–0.27. To be consistent with economic theory and with the rest of the money demand literature, we henceforth impose $\theta_p = 1$ and study in more detail the estimation of θ_y and θ_r.

Estimates of cointegrating vectors in the system $(m - p, y, r)$ are presented in panel A of Table III. The estimators are those studied in the Monte Carlo experiment, plus the single-equation nonlinear least squares estimator (NLLS), which is used by Baba, Hendry, and Starr (1992) to estimate their long-run M1 demand equation. The full-sample estimates are similar across estimators and none of the efficient estimators reject the hypothesis that $\theta_y = 1$ at the 10% level.[5] The remaining columns examine the stability of the estimates over two subsamples, 1900–1945 and 1946–1989. The subsamples were chosen both because of the natural break at the end of World War II and because they divide the full sample nearly in two. Using only the first half of the sample, with the exception of DGLS and JOH the efficient estimators provide smaller

[5] Our static OLS estimates differ slightly from those presented in Lucas (1988) because of transcription errors in his original data set, now corrected. We thank Lucas for bringing these errors to our attention.

TABLE II

ESTIMATED COINTEGRATING RELATIONS: $m_t = \mu + \theta_p p_t + \theta_y y_t + \theta_r r_t$

Specifications:

I. p_t, r_t, y_t, I(1) and not cointegrated:

$$m_t = \mu + \theta_p p_t + \theta_y y_t + \theta_r r_t + d_p(L)\Delta p_t + d_y(L)\Delta y_t + d_r(L)\Delta r_t + e_t.$$

II. p_t I(2), r_t and y_t I(1), and $(r_t, \Delta p_t)$ not cointegrated:

$$m_t = \mu + \theta_p p_t + \theta_y y_t + \theta_r r_t + d_p(L)\Delta^2 p_t + d_y(L)\Delta y_t + d_r(L)\Delta r_t + e_t.$$

III. For p_t I(2), r_t and y_t I(1), and $r_t - \Delta p_t$ I(0):

$$m_t = \mu + \theta_p p_t + \theta_y y_t + \theta_r r_t + d_p(L)\Delta^2 p_t + d_y(L)\Delta y_t + d_r(L)(r_t - \Delta p_t) + e_t.$$

Specification	Estimator	Sample Period	No. Leads and Lags	Estimates (Standard Errors) θ_p	θ_y	θ_r
I	DOLS	1903–1987	2	1.119	0.858	−0.114
				(0.202)	(0.168)	(0.017)
	DOLS	1904–1986	3	1.159	0.831	−0.122
				(0.234)	(0.191)	(0.018)
	DGLS	1903–1987	2	0.997	0.685	−0.034
				(0.194)	(0.237)	(0.015)
	DGLS	1904–1986	3	1.105	0.890	−0.115
				(0.159)	(0.133)	(0.015)
II	DOLS	1904–1987	2	1.163	0.841	−0.114
				(0.249)	(0.208)	(0.021)
	DOLS	1905–1986	3	1.277	0.754	−0.125
				(0.290)	(0.238)	(0.023)
	DGLS	1904–1987	2	1.022	0.725	−0.032
				(0.205)	(0.241)	(0.016)
	DGLS	1905–1986	3	1.140	0.723	−0.062
				(0.228)	(0.265)	(0.023)
III	DOLS	1904–1987	2	0.981	0.972	−0.086
				(0.190)	(0.158)	(0.017)
	DOLS	1905–1986	3	1.051	0.922	−0.095
				(0.185)	(0.151)	(0.016)
	DGLS	1904–1987	2	0.854	0.671	−0.002
				(0.217)	(0.263)	(0.014)
	DGLS	1905–1986	3	1.087	0.917	−0.098
				(0.141)	(0.115)	(0.013)

Notes: $d_i(L) = \sum_{j=-k}^{k} d_{ij} L^j$, where k is the number of leads and lags listed in the fourth column. Standard errors are in parentheses. An AR(2) error process was used to implement the GLS transformation for the DGLS estimator and to estimate the DOLS covariance matrix when $k = 2$, and an AR(3) was used for $k = 3$. The shorter regression periods for $k = 3$ than for $k = 2$, and for specifications II and III than for specification I, allow for necessary initial and terminal conditions (leads and lags).

income elasticities and comparable interest elasticities than over the full sample, but the differences are slight. In sharp contrast to the first-half estimates, the postwar estimates in Table III differ greatly across estimators. The SOLS, DOLS, PBSR, and PHFM estimates are close to zero, and the NLLS and JOH(3) elasticities have the "wrong" sign. The JOH estimator is highly sensitive to the number of lagged first differences used. (Likelihood ratio statistics testing 2 vs. 3 lags in the VECM reject for the full and postwar samples, and thus suggest relying on the JOH(3) estimates.)

The final set of estimates refer to the system $(m - p, y, r^*)$, where r^* is the commercial paper rate passed through a low-pass filter; the results in Table III

TABLE III

MONEY DEMAND COINTEGRATING VECTORS: ESTIMATES AND TESTS, ANNUAL DATA

Dynamic OLS/GLS estimation equation:

$$m_t - p_t = \mu + \theta_y y_t + \theta_r r_t + d_y(L)\Delta y_t + d_r(L)\Delta r_t + e_t.$$

A. Point Estimates (Standard Errors)

Estimator	1903–1987 θ_y	1903–1987 θ_r	1903–1945 θ_y	1903–1945 θ_r	1946–1987 θ_y	1946–1987 θ_r	1946–1987 θ_y	1946–1987 θ_{r*}
SOLS	0.943	−0.083	0.919	−0.085	0.192	−0.016	0.410	−0.046
NLLS	0.856	−0.110	0.897	0.014	−0.435	0.082	0.297	−0.023
DOLS	0.970	−0.101	0.887	−0.104	0.269	−0.027	0.410	−0.047
	(0.046)	(0.013)	(0.197)	(0.038)	(0.213)	(0.025)	(0.319)	(0.042)
DGLS	0.829	−0.051	1.166	−0.084	0.951	−0.020	1.170	−0.091
	(0.135)	(0.015)	(0.199)	(0.031)	(0.307)	(0.009)	(0.131)	(0.013)
PBSR	0.965	−0.097	0.866	−0.098	0.216	−0.020	0.366	−0.042
	(0.035)	(0.010)	(0.099)	(0.018)	(0.091)	(0.011)	(0.147)	(0.019)
PHFM	0.963	−0.097	0.899	−0.094	0.204	−0.018	0.391	−0.045
	(0.034)	(0.009)	(0.083)	(0.015)	(0.054)	(0.006)	(0.106)	(0.014)
JOH(2)	0.975	−0.114	0.886	0.058	163.93	−23.52	−2.174	0.316
	(0.042)	(0.013)	(0.193)	(0.116)	(38167.1)	(5479.1)	(4.024)	(0.568)
JOH(3)	0.994	−0.113	0.839	0.081	−3.140	0.472	−0.393	0.073
	(0.040)	(0.012)	(0.220)	(0.160)	(12.63)	(1.860)	(0.559)	(0.082)

B. Tests for Breaks in the Cointegrating Vector Base on DOLS, Break Date = 1946

Interest Rate	No. Leads and Lags	χ_2^2 Wald statistics (p-value)	Point Estimates (Standard Errors) θ_y	θ_r	δ_y	δ_r
r	2	1.75	1.047	−0.090	−0.500	0.040
		(0.42)	(0.164)	(0.034)	(0.413)	(0.045)
	3	1.38	1.059	−0.090	−0.525	0.040
		(0.50)	(0.177)	(0.037)	(0.481)	(0.047)
r^*	2	0.57	0.945	−0.116	−0.349	0.055
		(0.75)	(0.183)	(0.043)	(0.597)	(0.081)
	3	1.16	0.943	−0.117	−0.551	0.079
		(0.56)	(0.180)	(0.042)	(0.582)	(0.079)

Notes: *Panel A*: NLLS is the nonlinear least squares estimator; the other estimators are defined in the notes to Table I (DOLS here and in subsequent tables is DOLS2 in Table I). JOH(k) is the JOH estimator evaluated using k lagged first differences. JOH(3) was computed over regression dates 1904–1987, 1904–1945, and 1946–1987. For the NLLS estimator, $\Delta(m-p)_t$ was regressed on $(m-p)_{t-1}, y_{t-1}, r_{t-1}$, and 2 lags each of $\Delta(m-p)_{t-1}$, Δy_{t-1}, and Δr_{t-1}; θ_y and θ_r were estimated from the coefficients on the lagged levels. DOLS and DGLS used 2 leads and lags of the first differences in the regressions and an AR(2) process for the error. The frequency zero spectral estimators required for PBSR and PHFM were computed using a Bartlett kernel with 5 lags. All regressions included a constant.

Panel B: The statistics are based on the regression, $(m-p)_t = \mu + \theta_y y_t + \theta_r r_t + \delta_y(y_t - y_\tau)1(t > \tau) + \delta_r(r_t - r_\tau)1(t > \tau) + d_y(L)\Delta y_t + d_r(L)\Delta r_t$, where $1(\cdot)$ is the indicator function, $d_y(L)$ and $d_r(L)$ have the number of leads and lags stated in the second column, and $\tau = 1945$. Regressions with $k = 2$ were run over 1903–1987, with $k = 3$, over 1904–1986. The Wald statistic tests the hypothesis that $\delta_y = \delta_r = 0$ and has a χ_2^2 distribution. The covariance matrix was computed using an AR(2) spectral estimator for $k = 2$ and an AR(3) estimator for $k = 3$.

refer to r^* produced using a two-sided filter based on the Kalman smoothing algorithm and are typical of results based on other low-pass filters.[6] This smoothed interest rate can be interpreted as a proxy for a long-term rate which,

[6] r_t^* was constructed as the two-sided estimate of the permanent component of r_t calculated using the Kalman smoother for the model $r_t = r_t^* + \mu_{1t}$, $\Delta r_t^* = \mu_{2t}$, with (μ_{1t}, μ_{2t}) independent and $\text{var}(\mu_{1t})/\text{var}(\mu_{2t}) = 3$. Other filters that yield similar results are a one-sided exponentially weighted moving average filter with coefficient .95 and the Hodrick-Prescott filter.

under the risk-neutral theory of the term structure, is an average of current and expected future short rates. The empirical money demand literature is inconclusive on whether a long- or short-term interest rate is more appropriate. Because there is no consistent risk-free long-term rate with constant tax treatment over the full sample, using r^* provides a way to compare specifications with long-term and short-term rates. Although the full-sample estimates (not tabulated here) change little using r^*, using r^* rather than r changes the postwar estimates substantially. The postwar OLS, DOLS, DGLS, PBSR, and PHFM elasticities and (except for DGLS) standard errors are all larger using r^*. The JOH and NLLS estimates are quite sensitive to using r^* rather than r, and the differences across estimators remain.[7]

The differences between the prewar and postwar estimates raise the possibility that there has been a shift in the long-run money demand relation. To evaluate this and to ascertain the source of the instability in the postwar point estimates, we examine four related pieces of evidence. The first consists of formal tests of the null hypothesis of a constant cointegrating relation, against the alternative of different cointegrating vectors over 1900–1945 and 1946–1989, under the maintained hypothesis that the parameters describing the short-run relations are constant. These tests, implemented using the DOLS estimator and summarized in panel B of Table III, fail to reject the null of no break at all conventional significance levels. Although the point estimates of the changes, $\hat{\delta}_y$ and $\hat{\delta}_r$, are large—the income elasticity point estimates are 0.94–1.06 prewar and 0.39–0.60 postwar—these shifts are imprecisely estimated and are not statistically significant.[8]

The second piece of evidence concerns the properties of the cointegrating residuals, $\hat{z}_t = (m_t - p_t) - \hat{\theta}_y y_t - \hat{\theta}_r r_t$, which are quite different for the prewar and postwar point estimates. Residuals constructed using either the full-sample or first-half point estimates are consistent with cointegration, while the residuals based on the postwar estimates are not. For example, when $\hat{z}_t$ is constructed using the prewar DOLS elasticities, the sum of the coefficients in a levels AR(3) specification for $\hat{z}_t$ (with a constant) estimated over 1946–1989 is .49. In contrast, the postwar DOLS elasticities yield residuals with greater persistence: the corresponding sum of coefficients is .73. More formal evidence is obtained by computing the detrended Dickey-Fuller $\hat{\tau}_\mu$ statistic using postwar $\hat{z}_t$, where $\hat{z}_t$ is constructed using prewar DOLS elasticities. Because these elasticities are estimated from data prior to those used for the test, under the null of noncointegration this $\hat{\tau}_\mu$ statistic has the asymptotic univariate demeaned Dickey-Fuller distribution. This $\hat{\tau}_\mu$ statistic is -3.29 (2 lagged first differences;

[7] The results in Table III are robust to changes in the details of the computation of each of the estimators, in particular: using a Bartlett kernel with 7 lags for PBSR and PHFM, using 3 rather than 2 leads and lags for DOLS and DGLS, and using 1 rather than 2 or 3 lags for JOH. The only exception is the postwar instability of the JOH estimates, as discussed below.

[8] The statistics in Table IIIB only consider the possibility of a break in 1946. Recently, Gregory and Nason (1991) have computed sequential tests for the stability of the cointegrating coefficients with these data, treating the break date as unknown. Using Monte Carlo critical values, they also fail to reject the null hypothesis of constant coefficients.

with 4 lags, it is -3.31), and so rejects at the 5% one-sided level, supporting the view that the prewar money demand relation cointegrates the postwar series. In short, the postwar cointegrating residuals based on the first-half estimates exhibit less persistence than those computed using the postwar elasticities. However, postwar cointegrating residuals have a smaller standard deviation (0.054 over 1946–1987) when computed with the postwar DOLS elasticities than they do when computed with the prewar DOLS elasticities (the standard deviation is 0.166 over 1946–1987).

Third, 95% confidence regions for (θ_y, θ_r) estimated using DOLS over the full sample, the first half, and the second half using r^* overlap, and the DOLS region computed using r over 1946–1987 nearly overlaps, near the full-sample DOLS estimates of $(.970, -.101)$. Because Wald statistics testing hypotheses about (θ_y, θ_r) using the efficient estimators have large-sample χ^2 distributions, standard formulae can be used to construct confidence ellipses for (θ_y, θ_r). The estimators for the two subsamples are independent asymptotically (but not in finite samples because of short-run dependence in the data and the presence of initial and terminal leads and lags). Confidence sets for the subsamples in Table III are plotted in Figure 2a–2d for, respectively, the DOLS, DGLS, PBSR, and PHFM estimators. In almost all cases, the major axes of the prewar and postwar ellipses are approximately orthogonal and the confidence region for the full sample is much smaller than for either half. Comparing the DOLS with the other confidence regions, however, produces two qualitative differences. First, the postwar DGLS region computed using r has different axes and location than the other postwar regions. This arises because the estimated GLS transformation for DGLS approximately differences the postwar data (the estimated AR(2) filter is $1 - 1.39L + .41L^2$). In effect, this DGLS point estimate is determined by covariances between first differences of the data, not their levels, which leads us to conclude that this DGLS estimator is not estimating a cointegrating relation. Second, although the DOLS confidence sets contain or nearly contain $(.970, -.101)$, the postwar sets constructed using the other estimators do not. This might correctly reflect better finite-sample precision of these estimators relative to DOLS. Alternatively, because the postwar regions are based on only 42 observations, the postwar regions for DGLS, PBSR, PHFM, and perhaps DOLS, might overstate the precision of these estimators. To investigate these possibilities, we performed an additional Monte Carlo experiment to check the empirical confidence coefficient of asymptotic 95% confidence intervals based on these estimators. Specifically, 42 observations of Gaussian pseudo-data (plus initial conditions) were generated from a trivariate VECM(2), estimated using the full sample with $(m-p, y, r)$ (including a constant) imposing the full-sample DOLS elasticities $(.970, -.101)$. Wald statistics testing these values of (θ_y, θ_r) were constructed using the same kernels, number of lags, etc. as in Table III. For each of the estimators, the asymptotic χ_2^2 approximation was found to understate substantially the dispersion under the null hypothesis: the Monte Carlo coverage rates for asymptotic 95% confidence regions for (θ_y, θ_r) for the DOLS, DGLS, PBSR, PHFM, and JOH statistics are,

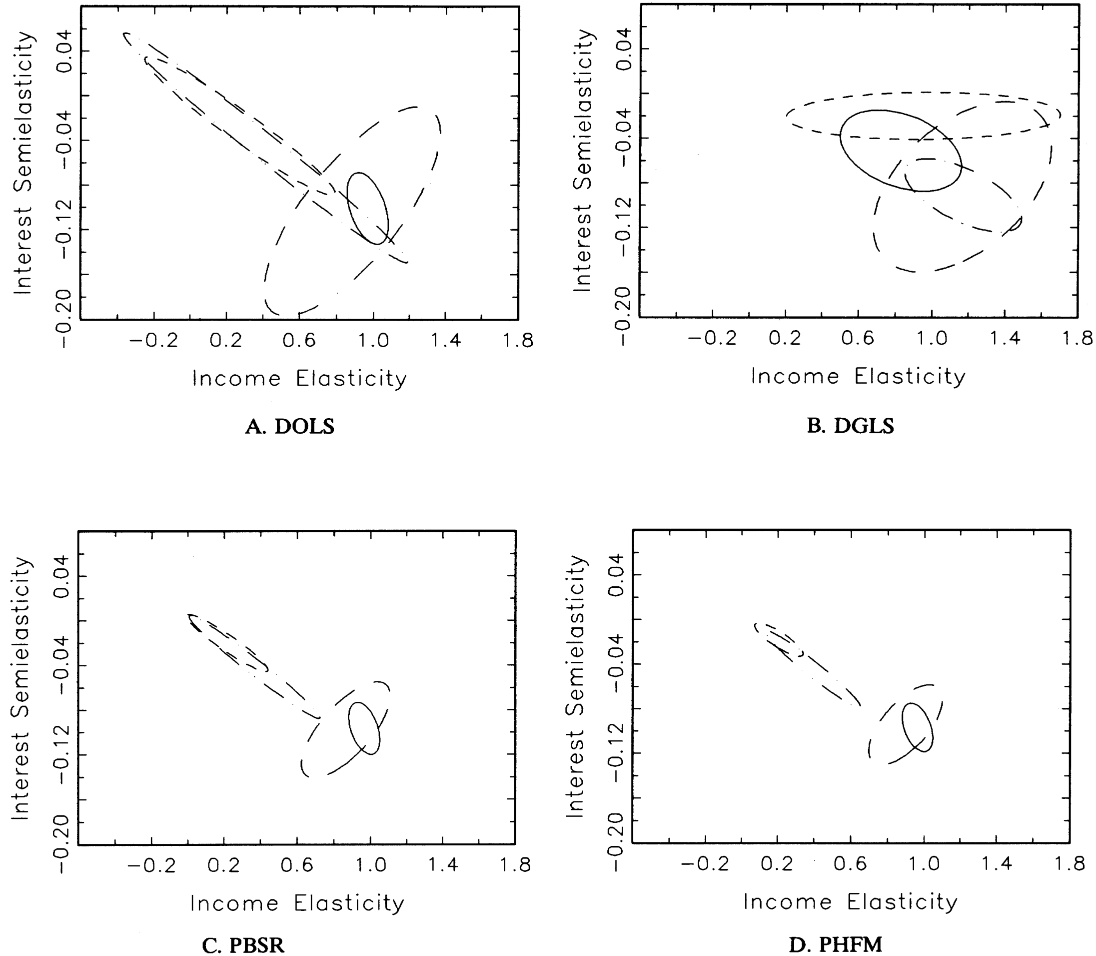


FIGURE 2.—95% confidence regions for the income elasticity θ_y and the interest semielasticity θ_r, estimated over 1903–1987 (solid line), 1903–1945 (dashes), 1946–1987 (short dashes), and, using the smoothed interest rate r^*, 1946–1987 (dash-dots), based on the DOLS, DGLS, PBSR, and PHFM estimators.

respectively, 57%, 32%, 33%, 14%, and 61%. These large coverage distortions diminish as the sample size, and with it the number of lags, increase. For the sample size at hand, however, these results strongly suggest that the postwar confidence regions in Figure 2—particularly the DGLS, PBSR, and PHFM regions—considerably understate the true sampling variability.[9]

The fourth piece of evidence concerns the most extreme of the postwar estimates, the JOH estimates, and the instability of the JOH estimator with respect to the subsample. For example, for samples ending in 1987 and starting in 1940, 1942, 1944, 1948, and 1950, the JOH(2) estimates of $\hat{\theta}_y$ are respectively 0.04, -0.36, 0.89, 0.56, and 1.33. The postwar VECM likelihood for JOH(3) in Table IIIA, concentrated to be a function of (θ_y, θ_r), is bimodal for both the $(m-p, y, r)$ and $(m-p, y, r^*)$ data sets. Inspection of the concentrated likelihood, plotted in Figure 3 for the $(m-p, y, r)$ data set, yields two conclusions: that the JOH MLE's for 2 and 3 lags lie on a ridge that corresponds to the major axis of the postwar confidence ellipses in Figure 2, and that the likelihood is not well approximated as a quadratic. The shape of this surface is typical for JOH estimators for starting dates ranging from 1940 to 1950. This ridge in the likelihood therefore explains, in a mechanical sense, the sensitivity of the JOH estimates in Table III to the lag length and to the precise subsample.

These four pieces of evidence lead us to conclude that, despite the large differences between the prewar and postwar point estimates, the results support Lucas' (1988) conclusion that long-run M1 demand has been stable over 1900–1989. Using the postwar data alone, the elasticities are imprecisely estimated. The postwar data are dominated by the 1950–1980 trends in velocity and interest rates; as Lucas (1988) pointed out, this requires the estimates to lie on the "trend line" given by $\Delta\overline{(m-p)} - \theta_y \Delta\bar{y} - \theta_r \Delta\bar{r} = 0$ (where $\Delta\bar{y}$ is the average annual growth rate of y_t, etc.). This line constitutes the major axis of the postwar confidence ellipses in Figure 2 and the ridge in the postwar VECM likelihood in Figure 3. To make this concrete, consider the estimator obtained by solving the 1900–1945 and 1946–1989 trend lines uniquely for (θ_y, θ_r); the resulting estimators are 1.00 and $-.147$, respectively, strikingly close to the full-sample efficient estimates. This "trend line" analysis emphasizes three conclusions from the more formal results. First, because the efficient estimators of cointegrating vectors exploit this same low-frequency information, albeit in a more sophisticated way, the sampling uncertainty of the full-sample estimates is considerably less than that based on the prewar and especially the postwar data. Second, several such trend lines (or low frequency movements) can be seen in the prewar sample, resulting in tighter prewar than postwar confidence regions.

[9] Increasing the number of lags and kernel length does not appreciably improve the coverage rates, although increasing the number of observations does. For autoregressive lag length 4 and kernel truncation lag 7 (Bartlett kernel) and 42 observations, the DOLS, DGLS, PBSR, PHFM, and JOH coverage rates (for asymptotic 95% regions) are: 52%, 46%, 45%, 14%, and 53%. For 250 observations and the same lag lengths, the respective empirical coverage rates improve to: 83%, 62%, 61%, 44%, and 87%. These results are based on 1000 Monte Carlo replications. The low coverage rates in this trivariate system, particularly for PHFM, are consistent with the size distortions in the $P(W > 3.84)$ columns of Table I, panel C, discussed in Section 6.

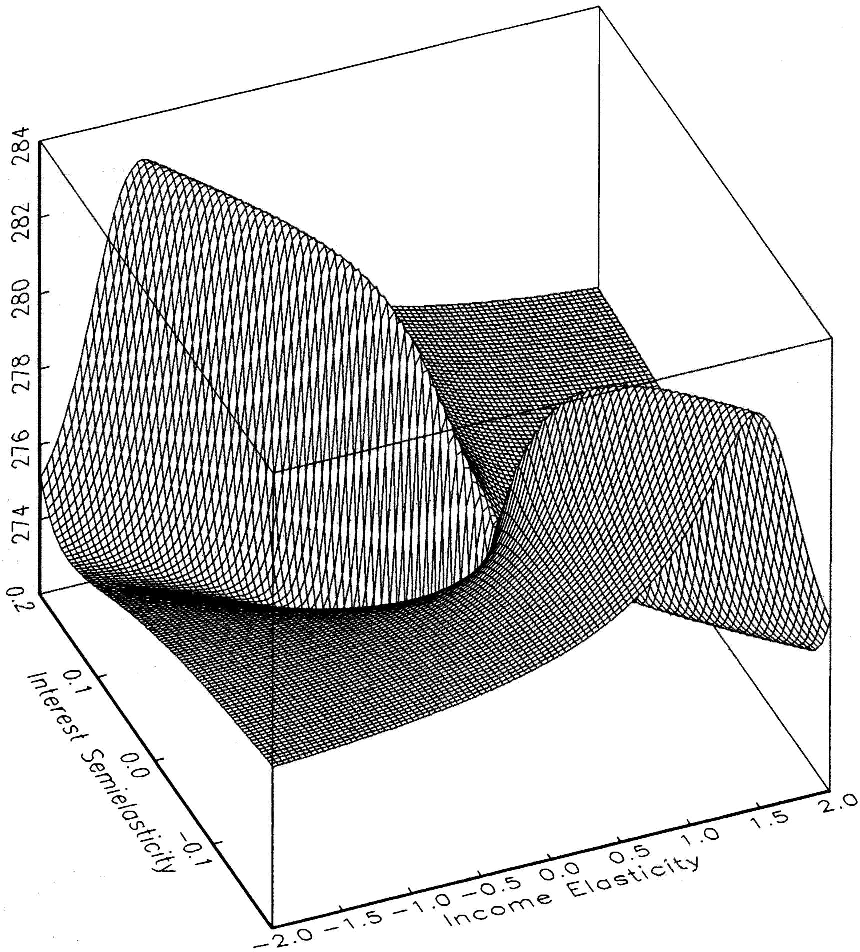


FIGURE 3.—Concentrated vector error correction model likelihood surface in (θ_y, θ_r) space, 1946–1987 (JOH(3) estimator).

Third, because the postwar data are dominated by a single trend and because the growth rate of real M1 balances is nearly zero postwar, a linear combination involving y and r alone can eliminate much of their long-term trend; the JOH estimator selects such a linear combination, placing small weight on $m-p$. When normalized so that $m-p$ has a unit coefficient, the resulting postwar money demand relations are unstable, with the ratio θ_y/θ_r, but not the individual elasticities, well determined.

B. *Results for postwar monthly data.* Cointegrating vectors estimated using postwar monthly data on M1, real personal income, the personal income price deflator, and a variety of interest rates are reported in Table IV. Compared to the postwar annual results, the income elasticities estimated over 1949:1–1988:6 are large and there is somewhat less disagreement across the efficient estimators, with income elasticities ranging from .30 to .89 based on the commercial

TABLE IV

MONEY DEMAND COINTEGRATING VECTORS: ESTIMATES, POSTWAR MONTHLY DATA

Period:	49:1–88:6		49:1–88:6		60:1–88:6		60:1–88:6		60:1–88:6	
Interest rate:	Comm. Paper		Comm. Paper*		Comm. Paper		90-day T-bill		10-yr T-bond	
Estimator	θ_y	θ_r	θ_y	θ_r	θ_y	θ_r	θ_y	θ_r	θ_y	θ_r
SOLS	.272	−.016	.398	−.035	.339	−.017	.362	−.021	.480	−.031
NLLS	.539	−.044	.259	.034	.570	−.030	−.483	−.026	.353	.012
DOLS	.326	−.025	.457	−.044	.398	−.027	.415	−.030	.529	−.037
	(.187)	(.026)	(.136)	(.019)	(.208)	(.025)	(.165)	(.020)	(.210)	(.023)
DGLS	.890	−.008	.525	−.026	1.139	−.009	1.196	−.011	1.046	−.019
	(.203)	(.003)	(.109)	(.007)	(.289)	(.003)	(.269)	(.003)	(.173)	(.003)
PBSR	.302	−.021	.404	−.036	.367	−.022	.389	−.025	.500	−.034
	(.037)	(.005)	(.045)	(.006)	(.053)	(.005)	(.049)	(.005)	(.052)	(.005)
PHFM	.302	−.021	.412	−.037	.370	−.022	.393	−.025	.511	−.035
	(.033)	(.004)	(.042)	(.006)	(.048)	(.004)	(.045)	(.004)	(.047)	(.005)
JOH	.561	−.068	.629	−.076	.520	−.075	.462	−.060	.631	−.067
	(.199)	(.032)	(.129)	(.020)	(.202)	(.039)	(.137)	(.024)	(.144)	(.021)

Notes: NLLS and JOH used 8 lagged differences of the variables; DOLS and DGLS used 8 leads and lags of the first differences in the regressions. An AR(6) error was used for DGLS and for the calculation of the standard errors for DOLS. The frequency zero spectral estimators required for PBSR and PHFM were computed using a Bartlett kernel with 18 (monthly) lags. All regressions included a constant. The "Commercial Paper*" rate was constructed using the Kalman filter as described in footnote 6 in the text.

paper rate. The estimates are stable across the choice of interest rate (the exception is the DGLS estimates, for which GLS effectively first-differences the data, as in the postwar annual estimates). The point estimates agree closely with Baba, Hendry, and Starr's (1992) NLLS estimate of .5 obtained over 1960–1988, strikingly so since Baba, Hendry, and Starr (1992) used GNP rather than personal income, quarterly rather than monthly data, and several additional regressors designed to account for shifts in short-run money demand relation. They are also comparable to Hoffman and Rasche's (1991) JOH results based on M1, personal income, and 90-day T-Bill data for 1953–1988; the difference between their income elasticity of .78 and the JOH estimate in Table IV for 60:1–88:6 (.46) mainly arises from our use of levels and their use of logarithms of the interest rate.

The relative stability of these estimates across estimators and initial dates contrasts with the results based on postwar annual data. Further examination, however, reveals that the monthly results are quite sensitive to the final regression date. For example, JOH estimates of the income elasticity, estimated over 60:1 to the last month in each quarter from March 1984 through June 1988 using the commercial paper rate (8 lags), range from -3.00 to 3.54; for the NLLS estimator, this range is .29 to 1.08. When computed over 60:1–78:12, the JOH, NLLS, and DOLS income elasticities are $-.27$, $-.13$, and .11. Comparable instability is present for each of the interest rates studied in Table IV, whether in logarithms or in levels. Because we do not provide uniform critical values for tests based on these "recursive" estimates, they do not provide formal evidence on the stability of the cointegrating vector estimated with the postwar data. Still, this sensitivity to terminal regression dates is consistent with our conclusions from the annual data. That is, the data from 1950 to 1982 are dominated by the single upward trend in real balances, income, and interest rates, which results in income and interest elasticity estimates which are highly negatively correlated and are imprecisely estimated, except that they must lie on the trend line which determines their ratio. Only when the most recent data are used—the data since 1982 follow a second trend (increasing income, declining velocity, and interest rates)—are the estimates more precise with values that are comparable across estimators.

C. *Discussion and summary*. Our analysis has relied heavily on asymptotic distribution theory to construct formal confidence regions and tests, and the estimation procedures typically entail the estimation of many nuisance parameters relative to the sample size. This and the Monte Carlo evidence leads us to suspect that the asymptotic standard errors reported here overstate the precision of the estimated elasticities, at least for the postwar data. Also, of course, this work has not examined the long-run demand for other monetary quantity aggregates.

Even with these caveats, these results suggest three conclusions. First, when viewed over 1900–1989, there appears to be a stable long-run M1 demand function. Estimated over the entire sample, 95% confidence intervals based on the DOLS estimator are, for the income elasticity, (.88, 1.06), and for the interest semielasticity, (−.127, −.075). Qualitatively similar intervals are obtained using the other efficient estimators over the full sample.

Second, these results are consistent with Lucas' (1988) suggestion that there is a stable long-run money demand relation over the pre- and postwar periods. A key piece of evidence for this is the apparent stationarity of the postwar residuals computed using long-run elasticities estimated from the prewar data.

Third, in isolation the postwar evidence says little about the parameters of the cointegrating vector: the estimates have large standard errors and moreover are sensitive to the subsample and estimator used. The main reason for this is that the postwar data are dominated by steadily rising income and interest rates and effectively no growth in real balances. Only after 1982 is there a decline in interest rates that reduces multicollinearity between interest rates and income sufficiently to estimate the money demand relation, rather than simply the ratio θ_y/θ_r. We suspect that the postwar standard errors understate the sampling variability, particularly for the monthly results, both because of the evident sensitivity to terminal dates and because of Monte Carlo evidence in Sections 6 and 7A that the asymptotic distributions provide poor approximations to the postwar sampling distributions in designs that approximate the empirical multivariate models.

8. CONCLUSIONS

As do other asymptotically efficient estimators of cointegrating vectors, the procedures proposed here require at least partial knowledge of which variables cointegrate and of the orders of integration of individual series. With currently available techniques, this entails pretesting for unit roots or, when possible, relying on economic theory for guidance. A plausible suspicion is that this pretesting introduces additional sampling uncertainty, at least in finite samples, beyond that which is formally studied here. It is worth emphasizing, however, that one advantage of our estimators is that Wald statistics testing restrictions on the cointegrating vectors will have asymptotic χ^2 distributions under a variety of assumptions about the orders of integration of the various estimators. This feature made it possible to perform inference on the price and interest elasticities in long run nominal money demand, relating log nominal M1 to the log price level, log income, and the interest rate, even though the evidence was inconclusive about whether money and prices were individually integrated of orders one or two.

Our emphasis has been on inference about the cointegrating relations, and the proposed estimators treat the parameters describing the short-run dynamics of the process as nuisance parameters. In many applications, however, the short-run dynamics are of independent or even primary interest. For example, much of the empirical money demand literature has focused on the search for a stable short-run money demand function. In such cases, these efficient estimators can be used in subsequent stages of the analysis by imposing the estimated cointegrating vectors, for example in a variant of the triangular form (Campbell (1987), Campbell and Shiller (1987, 1989)), in a VECM (King, Plosser, Stock and Watson (1991)), or in a single-equation error correction framework (Hendry and Ericsson (1991)).

A lesson suggested by the empirical investigation of M1 demand and by the Monte Carlo results is that, when estimating cointegrating vectors, it can be valuable to use more than one of the currently available asymptotically efficient estimators. In the prewar data and in the full data set, the efficient estimators produced qualitatively and statistically close estimates, but when only the postwar annual data were used the estimates differed substantially. This divergence signaled the need for further analysis of the postwar relations. The postwar data contain a single dominant trend, which results in θ_y/θ_r but not the individual elasticities being well-determined. A Monte Carlo investigation suggested that, in this application, the DOLS estimator was preferable to the other estimators.

The empirical analysis suggests that the precise estimation of long-run M1 demand requires a long span of data: estimates over the full 90 years are considerably more precise than over the first half of the century alone, and when used in isolation the data since 1946 contain quite limited information about long-run M1 demand. Overall, the evidence is consistent with there being a single stable long-run demand for money, with an income elasticity near one and an interest semielasticity near -0.10.

Kennedy School of Government, Harvard University, Cambridge, MA 02138, U.S.A.

and

Dept. of Economics, Northwestern University, Evanston, IL 60208, U.S.A.

Manuscript received September, 1989; final revision received December, 1992.

APPENDIX A

Derivation of the $I(d)$ Triangular Representation (3.2)

The triangular representation (3.2) is constructed under Assumption A by repeated application of the following Lemma:

LEMMA A.1: *Assume that the $n \times 1$ vector x_t is generated by $\Delta x_t = \Sigma_{j=0}^{m} \mu_j t^j + F(L)\varepsilon_t$, where $F(L)$ is l-summable and $\operatorname{rank}[F(1)] = k \leqslant n$. Without loss of generality arrange x_t so that the upper $k \times n$ block of $F(1)$ has full row rank. Then x_t can be represented as*:

$$\Delta x_t^1 = \sum_{j=0}^{m} \bar{\mu}_{1,j} t^j + J_1(L)\varepsilon_t,$$

$$x_t^2 = \sum_{j=0}^{m+1} \bar{\mu}_{2,j} t^j + \theta x_t^1 + J_2(L)\varepsilon_t,$$

where $x_t = (x_t^{1\prime}, x_t^{2\prime})'$, where x_t^1 is $k \times 1$, x_t^2 is $(n-k)\times 1$, $J(L) = [J_1(L)', J_2(L)']'$ is $(l-1)$ summable, and $J_1(1)$ has full row rank k. When μ_m lies in the column space of $F(1)$, $\bar{\mu}_{2,m+1} = 0$.

PROOF OF LEMMA A.1: 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 $k \times n$, $f_2(L)$ is $(n-k)\times n$, and $f_1(1)$ has full row rank. The expression for Δx_t^1 follows by setting $J_1(L) = f_1(L)$. Because $\operatorname{rank}[F(1)] = \operatorname{rank}[f_1(1)]$, $f_2(1) = \theta f_1(1)$, for some $(n-k)\times k$ matrix θ. Now partition μ_i as $(\mu'_{1,i}, \mu'_{2,i})'$, so that

$$\Delta x_t^2 - \theta \Delta x_t^1 = \sum_{i=0}^{m} (\mu_{2,i} - \theta\mu_{1,i}) t^i + [f_2(L) - \theta f_1(L)]\varepsilon_t. \tag{A.1}$$

Accumulating (A.1) yields $x_t^2 - \theta x_t^1 = \Sigma_{i=0}^{m+1} \bar{\mu}_{2,i} t^i + J_2(L)\varepsilon_t$, where $\{\bar{\mu}_{2,i}\}$ are linear combinations of $\{\mu_{2,i} - \theta\mu_{1,i}\}$, $J_2(L) = f_2^*(L) - \theta f_1^*(L)$, where $f_i^*(L) = (1-L)^{-1}(f_i(L) - f_i(1))$, $i = 1, 2$, and where $\bar{\mu}_{2,0}$ subsumes the initial conditions of x^1 and x^2. It is readily shown that, because $f_i(L)$ is l-summable, $f_i^*(L)$ is $(l-1)$-summable. If μ_m lies in the column space of $F(1)$, then $\mu_{2,m} - \theta\mu_{1,m} = 0$ so $\bar{\mu}_{2,m+1} = 0$. The Lemma follows by setting $\bar{\mu}_{1,i} = \mu_{1,i}$ $(i = 0, \ldots, m)$. *Q.E.D.*

To construct the triangular representation (3.2) for $\Delta^d y_t = \mu + F(L)\varepsilon_t$, apply Lemma A.1 to $x_t \equiv \Delta^{d-1} y_t$ to yield the decomposition:

$$\Delta^d y_t^1 = \mu_{1,0} + F_1^{d-1}(L)\varepsilon_t,$$

$$\Delta^{d-1}\tilde{y}_t^2 = \tilde{\mu}_{2,0} + \tilde{\mu}_{2,1} t + \tilde{\theta}_{2,1}^{d-1}\Delta^{d-1} y_t^1 + F_2^{d-1}(L)\varepsilon_t,$$

where y_t has been partitioned (and possibly reordered) into $k_1 \times 1$ and $(n-k_1)\times 1$ components y_t^1 and $\tilde{y}_t^2$ such that $F_1^{d-1}(1)$ has full row rank k_1. Some or all rows of $\tilde{\theta}_{2,1}^{d-1}$ can be rows of zeros. Let $H(L) = [h_1(L)', h_2(L)', \ldots, h_{d+1}(L)']'$ where $h_m(L)$ is $k_m \times n$, and set $h_1(L) = F_1^{d-1}(L)$. This determines the first k_1 equations in (3.2). Without loss of generality let $F^{d-1}(1) = [F_1^{d-1}(1)', F_2^{d-1}(1)']'$ have rank $k_1 + k_2 \leqslant n$. Next, apply the lemma to $x_t \equiv [\Delta^{d-1} y_t^{1\prime}, (\Delta^{d-2}\tilde{y}_t^2 - \tilde{\theta}_{2,1}^{d-1}\Delta^{d-2} y_t^1)']'$, reordering as necessary the final $n - k_1$ rows of x_t; this determines the next k_2 equations of (3.2). Continuing this process yields (3.2), with $u_t^j = h_j(L)\varepsilon_t$, $j = 1, \ldots, d+1$, where $\operatorname{rank}[h_j(1)] = k_j$ for $j = 1, \ldots, d$. Note that $\tilde{H} = [h_1(1)', h_2(1)', \ldots, h_d(1)']'$ has full row rank by construction, and assumption $A(v)$ implies that $h_{d+1}(1)$ is linearly independent of the rows of $\tilde{H}$ and has full row rank.

By construction, $u_t = H(L)\varepsilon_t$ where $H(1)$ has full rank, and so the spectrum of u_t is nonsingular at frequency zero. At frequencies other than zero, the spectrum of u_t is nonsingular because u_t is $F(L)\varepsilon_t$ passed through a linear filter which is nonsingular at frequencies $\omega \neq 0 \pmod{2\pi}$, and because $F(e^{i\omega})$ is nonsingular at $\omega \neq 0 \pmod{2\pi}$ by Assumption A(iii). Finally, the assumed k-summability of $F(L)$ implies that $H(L)$ is $k-d$ summable for $k \geqslant d$.

The proofs of the theorems rely on an extension of Sims, Stock, and Watson's (1990) (SSW) Lemma 1, which gives limiting properties of moment matrices involving general I(d) regressors. This extension weakens one of the conditions of SSW Lemma 1. (The statement of this weaker condition and the revised proof of SSW Lemma 1 which follows will use the notation of SSW.) Specifically, replace SSW Condition 1 with the following condition.

CONDITION 1′: (i) *Let* η_t *be a* $n \times 1$ *martingale difference sequence with* $E(\eta_t|\eta_{t-1},\eta_{t-2},\dots)=0$, $E(\eta_t\eta_t'|\eta_{t-1},\eta_{t-2},\dots)=I_n$, *and* $\max_t \sup_t E(\eta_{it}^4|\eta_{t-1},\eta_{t-2},\dots)<\infty$.
(ii) $\sum_{j=0}^{\infty} j|F_{m1j}|<\infty$, $m=1,\dots$.

Condition 1′(i) is the same as SSW Condition 1(i). Condition 1′(ii) weakens the g summability in SSW Condition 1(ii) to 1-summability. For completeness, we state the revised lemma:

LEMMA A.2: *Lemma* 1 *of SSW holds under Condition* (1′).

PROOF OF LEMMA A.2: SSW Lemma 1 contains parts (a)–(h). The proofs of parts (a)–(f) require only 1-summability of $F(L)$ and thus they hold under Condition 1′ automatically. To prove part (g) of the lemma, notice that the result follows if the last term in SSW equation (A.2) converges to 0 in probability. In turn, this follows if $T^{-1/2-p}\sum_{t=1}^{T-1} t^k F_{m1}^*(L)\eta_t \xrightarrow{p} 0$ for $k \leqslant p-1$, where $F_{m1}^*(L)=(1-L)^{-1}(F_{m1}(L)-F_{m1}(1))$. To show this, write:

$$\begin{aligned}
T^{-1/2}E\left|T^{-p}\sum_{t=1}^{T-1} t^k F_{m1}^*(L)\eta_t\right| &\leqslant T^{-1/2}T^{-p}\sum_{t=1}^{T-1} t^k E\left|F_{m1}^*(L)\eta_t\right| \\
&\leqslant T^{-1/2}T^{-(p-k-1)}E\left|F_{m1}^*(L)\eta_t\right| \\
&\leqslant T^{-1/2}T^{-(p-k-1)}\left(\sum_{j=0}^{\infty}|F_{m1j}^*|\right)E|\eta_t| \to 0,
\end{aligned}$$

where the last line follows $p-k-1 \geqslant 0$ and Condition 1′, noting that the 1-summability of $F_{m1}(L)$ implies the absolute summability of $F_{m1}^*(L)$. Thus $T^{-1/2-p}\sum_{t=1}^{T-1} t^k F_{m1}^*(L)\eta_t \xrightarrow{p} 0$ by Markov's inequality.

To prove part (h), notice that the result follows if the last term in SSW equation (A.6) converges to 0 in probability. In turn, this follows if $T^{-p}\sum_{t=1}^{T}\xi_{t-1}^k(F_{m1}^*(L)\eta_{t-1})' \xrightarrow{p} 0$ for $k \leqslant p-1$. To show this, again use Markov's inequality and assume for notational convenience that ξ_t and $F_{m1}^*(L)\eta_t$ are scalars. Then,

$$\begin{aligned}
E\left|T^{-p}\sum_{t=1}^{T}\xi_{t-1}^k\left(F_{m1}^*(L)\eta_{t-1}\right)\right| &= T^{-(p-k+1/2)}E\left|\sum_{t=1}^{T}\left(T^{-k+1/2}\xi_{t-1}^k\right)\left(F_{m1}^*(L)\eta_{t-1}\right)\right| \\
&\leqslant T^{-(p-k+1/2)}\sum_{t=1}^{T}E\left(\left|T^{-k+1/2}\xi_{t-1}^k\right|\left|F_{m1}^*(L)\eta_{t-1}\right|\right) \\
&\leqslant T^{-(p-k+1/2)}\sum_{t=1}^{T}\left[E\left(T^{-k+1/2}\xi_{t-1}^k\right)^2\right]^{1/2} \\
&\quad\times\left[E\left(F_{m1}^*(L)\eta_{t-1}\right)^2\right]^{1/2} \\
&\leqslant T^{-(p-k-1)}T^{-1}\sum_{t=1}^{T}\left[E\left(T^{-k+1/2}\xi_{t-1}^k\right)^2\right]^{1/2} \\
&\quad\times T^{-1/2}\left(\sum_{j=0}^{\infty}|F_{m1j}^*|\sigma_\eta\right) \\
&\to 0
\end{aligned}$$

where the last line follows from $T^{-1}\sum_{t=1}^{T}[E(T^{-k+1/2}\xi_{t-1}^{k})^2]^{1/2} \Rightarrow \int_0^1\{E(W^k(s)^2)\}^{1/2}ds < \infty$; from $F_{m1}^*(L)$ being absolutely summable (because $F_{m1}(L)$ is 1-summable by Condition 1′); and from $k \leqslant p-1$. *Q.E.D.*

To apply Lemma A.2 in the proofs below, it must be shown that the canonical regressors have the representation $z_t = G(L)\nu_t$, where $G(L)$ and ν_t satisfy Condition 1′. ($G(L)$ here corresponds to $F(L)$ in Lemma A.2, and so it must be shown that $G_{mj}(L)$ is 1-summable for all m, j.) Condition 1′(i) follows from Assumptions A(i) and C(i). To show that Condition 1′(ii) holds, it suffices to show that $G(L)$ can always be chosen so that $G_{m1}(L)$ is 1-summable and $G_{mj}(L) = G_{mj}$ for $j \geqslant 2$ and all m. This result is straightforward for the canonical regressors dominated by deterministic terms. For the canonical regressors dominated by stochastic components (z_t^m, $m = 1, 3, \ldots, 2l-1$), it suffices to show this for the canonical regressor with the highest possible order of integration. That regressor has the representation $z_t^{2l-1} = M^+ \Delta^{d-l}y_t$ for some matrix M^+. To simplify the expressions we omit initial conditions and deterministic terms. Then, from (3.1), z_t^{2l-1} can be written

$$(\mathrm{A.2}) \qquad \begin{aligned} z_t^{2l-1} &= \sum_{s_1=1}^{t} \cdots \sum_{s_d=1}^{s_{d-1}} M^+F(L)\varepsilon_{s_d} \\ &= M^+F(1)\xi_t^{l-1} + M^+F^{*1}(1)\xi_t^{l-2} + \cdots + M^+F^{*l}(L)\varepsilon_t \\ &= G_{2l-1,2l-1}\xi_t^{l-1} + G_{2l-1,2l-3}\xi_t^{l-2} + \cdots + G_{2l-1,1}(L)\varepsilon_t \end{aligned}$$

where $F^{*l}(L)$ has coefficients $F_i^{*l} = -\sum_{j=i+1}^{\infty}F_j^{*l-1}$, defined recursively with $F^{*0}(L) = F(L)$, and where $G_{2l-1,2l-1} = M^+F(1)$, $G_{2l-1,2l-3} = M^+F^{*1}(1), \ldots, G_{2l-1,1}(L) = M^+F^{*l}(L)$. It is readily shown that the k-summability of $F(L)$ (Assumption A(ii)) implies that $F^{*l}(L)$ is $(k-l)$-summable. By assumption C(ii), $F(L)$ is $(d+2)$ summable. Because $l \leqslant d+1$, $F^{*l}(L)$ and $G_{2l-1,1}(L)$ are (at least) 1-summable for all l. Thus Assumptions A and C imply that $z_t = G(L)\nu_t$ can always be chosen so that Condition 1′ applies. In this construction all elements of $G(L)$ except $G_{11}(L)$ are one-sided. $G_{11}(L)$ will be two-sided because z_t^1 contains finitely many leads of u_t^j, $j < l$.

The following two lemmas are used in the proofs of the theorems.

LEMMA A.3:

$$\left[T^{-1/2}z_{[T\cdot]}^3, T^{-3/2}z_{[T\cdot]}^5, \ldots, T^{-l+3/2}z_{[T\cdot]}^{2l-1}, T^{-1/2}\sum_{t=1}^{[T\cdot]}\Phi(1)c_l(1)\varepsilon_t\right]$$
$$\Rightarrow \left[\Gamma_3W_1(\cdot), \Gamma_5W_1^2(\cdot), \ldots, \Gamma_{2l-1}W_1^{(l-1)/2}(\cdot), W_2(\cdot)\right],$$

where $W(s) = (W_1(s)'W_2(s)')'$ is standard n-dimensional Brownian motion and $\Gamma_m = G_{mm}(1)M'$, where M is a $(n-k_l)\times n$ matrix with rows that span the null space of the rows of $\Phi(1)c_l(1)$ and $MM' = I_{n-k_l}$.

PROOF OF LEMMA A.3: Let $B(s)$ be defined as the limiting process $T^{-1/2}\sum_{t=1}^{[T\cdot]}\varepsilon_t \Rightarrow B(\cdot)$; because ε_t is a martingale difference sequence and $E(\varepsilon_t\varepsilon_t') = I_n$, $B(s)$ is a n-dimensional standard Brownian motion. Define M to be a fixed $(n-k_l)\times n$ matrix with rows that span the null space of the rows of $\Phi(1)c_l(1)$ and $MM' = I_{n-k_l}$. Also define $W_1(s) = MB(s)$, $W_2(s) = \Phi(1)c_l(1)B(s)$, and $W(s) = (W_1(s)', W_2(s)')'$. Because $MM' = I_{n-k_l}$, $Mc_l(1)'\Phi(1)' = 0$, and $\Phi(1)c_l(1)c_l(1)'\Phi(1)' = I_{k_l}$, $W(s)$ is a standard Brownian motion.

From the discussion following the proof of Lemma A.2, the mth canonical regressor can be represented as

$$z_t^m = G_{m1}(L)\xi_t^0 + G_{m2} + G_{m3}\xi_t^1 + \cdots + G_{mm}\xi_t^{1/2(m-1)} \qquad m = 3, 5, \ldots, 2l-1,$$

so that $G_{mj}(1) = G_{mj}$ for $j \geqslant 2$. Note that $T^{-(m-2)/2}\xi_{[T\cdot]}^{(m-1)/2} \Rightarrow B^{(m-1)/2}(\cdot)$, where $B^1(s) = B(s)$ and $B^j(s) = \int_{r=0}^{s}B^{j-1}(r)\,dr$, for $m = 3, 5, \ldots, 2l-1$. Because $G_{m1}(L)$ is at least 1-summable (this follows from Assumption A as discussed following the proof of Lemma A.2), $T^{-(m-2)/2}z_{[T\cdot]}^m \Rightarrow G_{mm}(1)B^{(m-1)/2}(\cdot)$. Because $Ez_{t+k}v_t^{l\prime} = 0$ for all t and k, the cross spectrum of v_t^l and z_t is 0 at frequency 0 so that $G_{mm}(1)c_l(1)'\Phi(1)' = 0$, $m = 3, 5, \ldots, 2l-1$. Thus $G_{mm}(1)$ can be written $G_{mm}(1) = \Gamma_mM$, so $T^{-(m-2)/2}z_{[T\cdot]}^m \Rightarrow G_{mm}(1)B^{(m-1)/2}(\cdot) = \Gamma_mMB^{(m-1)/2}(\cdot) = \Gamma_mW_1^{(m-1)/2}(\cdot)$. The $g_m \times (n-k_l)$ matrix Γ_m can be obtained (given M) as $\Gamma_m = G_{mm}(1)M'$, $m = 3, 5, \ldots, 2l-1$. *Q.E.D.*

LEMMA A.4 (Beveridge-Nelson (1981) decomposition for two-sided filters): *Suppose* $\zeta(L)=\sum_{j=-\infty}^{\infty}\zeta_j L^j$ *is m-summable. Then* $\zeta(L)=\zeta(1)+\zeta^*(L)(1-L)$, *where* $\zeta^*(L)$ *is* $(m-1)$*-summable and* $\zeta_j^*=-\sum_{i=j+1}^{\infty}\zeta_i$, $j\geqslant 0$ *and* $\zeta_j^*=\sum_{i=-\infty}^{j}\zeta_i$, $j<0$.

PROOF OF LEMMA A.4: Write

$$\zeta(L)=\sum_{j=-\infty}^{-1}\zeta_j L^j+\sum_{j=0}^{\infty}\zeta_j L^j=\zeta(1)+\sum_{j=-\infty}^{-1}\zeta_j(L^j-1)+\sum_{j=0}^{\infty}\zeta_j(L^j-1).$$

The results obtain by using $L^j-1=(L-1)\sum_{i=0}^{j-1}L^i$, $j\geqslant 1$, and collecting terms. *Q.E.D.*

PROOF OF THEOREM 1: First consider the infeasible GLS estimator $\tilde{\delta}_{\text{GLS}}$ constructed using $\Phi(L)$ rather than $\hat{\Phi}(L)$. Note that $(\Upsilon_T\otimes I)(\tilde{\delta}_{\text{GLS}}-\delta)=Q_T^{-1}\phi_T$, where $Q_T=(\Upsilon_T^{-1}\otimes I)\sum_t\bar{z}_t\bar{z}_t'(\Upsilon_T^{-1}\otimes I)$ and $\phi_T=(\Upsilon_T^{-1}\otimes I)\sum_t\bar{z}_t\Phi(L)v_t^l$, with $\bar{z}_t=[\Phi(L)(z_t'\otimes I)]'$. (Identity matrices have dimension $k_l\times k_l$ unless otherwise stated.) 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 $j\geqslant 2$, we use Lemma A.2. Note that $\bar{z}_t=\sum_{m=0}^{q}(z_{t-m}\otimes\Phi_m')$, so

$$\begin{aligned}
Q_{ijT}&=\left(\Upsilon_{iT}^{-1}\otimes I\right)\sum_t\left[\sum_{m=0}^{q}\left(z_{t-m}^i\otimes\Phi_m'\right)\right]\left[\sum_{h=0}^{q}\left(z_{t-h}^j\otimes\Phi_h'\right)\right]'\left(\Upsilon_{jT}^{-1}\otimes I\right)\\
&=\left(\Upsilon_{iT}^{-1}\otimes I\right)\sum_t\left[\sum_{m=0}^{q}\sum_{h=0}^{q}\left(z_{t-m}^i z_{t-h}^{j\prime}\otimes\Phi_m'\Phi_h\right)\right]\left(\Upsilon_{jT}^{-1}\otimes I\right)\\
&=\left(\Upsilon_{iT}^{-1}\otimes I\right)\sum_t\left[\sum_{m=0}^{q}\sum_{h=0}^{q}\left(z_t^i z_t^{j\prime}\otimes\Phi_m'\Phi_h\right)\right]\left(\Upsilon_{jT}^{-1}\otimes I\right)+o_p(1)\\
&\Rightarrow V_{ij}\otimes(\Phi(1)'\Phi(1))=V_{ij}\otimes\Omega_{ll}^{-1}
\end{aligned}$$

where the last two lines follow from Lemma A.2 and $\Phi(1)'\Phi(1)=\Omega_{ll}^{-1}$.

For ϕ_{iT}, $i\geqslant 2$, let $\Phi(L)c_l(L)=\zeta(L)$ and use the two-sided Beveridge-Nelson decomposition (Lemma A.4) to write $\Phi(L)v_t^l=\zeta(L)\varepsilon_t=\zeta(1)\varepsilon_t+\zeta^*(L)\Delta\varepsilon_t$, where $\zeta^*(L)$ is defined in Lemma A.4. Because $c_l(L)=d_l(L)H(L)$, where $d_l(L)$ is the lth block of rows of $D(L)$, because $H(L)$ is 2-summable (by Assumption C(ii)), and because $d_l(L)$ has finite order (Assumption B), $c_l(L)$ is 2-summable. Because $\Phi(L)$ has finite order by assumption, $\zeta(L)$ is 2-summable, so by Lemma A.4, $\zeta^*(L)$ is 1-summable. Thus write $\phi_{iT}=\phi_{iT}^1+\phi_{iT}^2$, where

$$\phi_{iT}^1=\left(\Upsilon_{iT}^{-1}\otimes I\right)\sum_t\left[\sum_{m=0}^{q}\left(z_{t-m}^i\otimes\Phi_m'\right)\right]\Phi(1)c_l(1)\varepsilon_t,$$

$$\phi_{iT}^2=\left(\Upsilon_{iT}^{-1}\otimes I\right)\sum_t\left[\sum_{m=0}^{q}\left(z_{t-m}^i\otimes\Phi_m'\right)\right]\zeta^*(L)\,\Delta\varepsilon_t.$$

First consider ϕ_{iT}^1. The limit of ϕ_{iT}^1 follows from Lemmas A.2 and A.3:

$$\phi_{iT}^1\Rightarrow\begin{cases}\displaystyle\int_0^1\left(G_{ii}(1)s^{(i-2)/2}\otimes\Phi(1)'\right)dW_2(s) & i=2,4,6,\ldots,2l,\\[2ex] \displaystyle\int_0^1\left(F_iW_1^{(i-1)/2}(s)\otimes\Phi(1)'\right)dW_2(s) & i=3,5,7,\ldots,2l-1.\end{cases}$$

Next consider ϕ_{iT}^2. By telescoping arguments,

$$\begin{aligned}
\text{(A.3)}\qquad \phi_{iT}^2&=\left(\Upsilon_{iT}^{-1}\otimes I\right)\sum_{t=t_0}^{t_1-1}\left[\sum_{m=0}^{q}\left(\Delta z_{t-m}^i\otimes\Phi_m'\right)\right]\zeta^*(L)\varepsilon_t\\
&\quad+\left(\Upsilon_{iT}^{-1}\otimes I\right)\sum_{m=0}^{q}\left\{\left(z_{t-m}^i\otimes\Phi_m'\right)\zeta^*(L)\varepsilon_{t_1}-\left(z_{t-m}^i\otimes\Phi_m'\right)\zeta^*(L)\varepsilon_{t_0-1}\right\},
\end{aligned}$$

where the limits of the summation over t reflect a fixed number of initial and terminal observations used as initial and terminal conditions. The second term in this expression $\xrightarrow{p} 0$ by the Cauchy-Schwartz inequality. To show that the first term $\xrightarrow{p} 0$, first consider the case with $i=3$. Then $\Delta z^3_{t-m+1} = \gamma(L)\varepsilon_t$, where $\gamma(L) = (G_{33} + \Delta G_{31}(L))L^m$ from (A.2). Because $G_{31}(L)$ is 1-summable, $\gamma(L)$ is 1-summable; also $\Upsilon_{3T}^{-1} = T^{-1}$ and $\zeta^*(L)$ is 1-summable. Thus, the first term in (A.3) is a linear combination of finitely many terms of the form $T^{-1}\Sigma(\gamma(L)\varepsilon_t)(\zeta^*(L)\varepsilon_t)$. The absolute summability of $\gamma(L)$ and $\zeta^*(L)$, together with the assumed fourth moments of ε_t, imply that $T^{-1}\Sigma(\gamma(L)\varepsilon_t)(\zeta^*(L)\varepsilon_t) \xrightarrow{p} \kappa$, where κ is a constant. But $E\phi_{3T} = E\phi^1_{3T} + E\phi^2_{3T}$; because $E\phi_{3T} = 0$ and $E\phi^1_{3T} \to 0$, $\kappa = 0$, so $\phi^2_{3T} \xrightarrow{p} 0$.

For $i = 5, 7, \ldots, 2l-1$, the first term in (A.3) is a linear combination of finitely many terms of the form $T^{-1/2(i-1)}\Sigma_{t=t_0}^{t_1-1} \Delta z^i_{r,t-k}(\zeta^*_h(L)\varepsilon_t)'$, for fixed k, where $z^i_{r,t-k}$ is the rth element of z^i_{t-k} and $\zeta^*_h(L)$ is the hth row of $\zeta^*(L)$. These terms are readily shown to converge to zero in probability using the Cauchy-Schwartz inequality and SSW Lemma 1(h) (modified in Lemma A.2) because ε_t has finite fourth moments and $\zeta^*(L)$ is 1-summable.

For $i = 2, 4, \ldots, 2l$, the result follows from SSW Lemma (g) (modified above). Thus $\phi^2_{it} \xrightarrow{p} 0$ for $i = 1, 2, \ldots, 2l$.

The joint convergence $(Q_T, \phi_T) \Rightarrow (Q, \phi)$ follows from SSW Lemmas 1 and 2.

To prove that the feasible GLS estimator has the same limit, let

$$\hat{Q}_T = (\Upsilon_T^{-1} \otimes I) \sum_t \left[\sum_{m=0}^{q} \left(z_{t-m} \otimes \hat{\Phi}'_m \right) \right] \left[\sum_{h=0}^{q} \left(z_{t-h} \otimes \hat{\Phi}'_h \right) \right]' (\Upsilon_T^{-1} \otimes I),$$

$$\hat{\phi}_T = (\Upsilon_T^{-1} \otimes I) \sum_t \left[\sum_{m=0}^{q} \left(z_{t-m} \otimes \hat{\Phi}'_m \right) \right] \hat{\Phi}(L) v^l_t,$$

so $(\Upsilon_T \otimes I)(\hat{\delta}_{\mathrm{GLS}} - \bar{\delta}_{\mathrm{GLS}}) = \hat{Q}_T^{-1}(\hat{\phi}_T - \phi_T) + (\hat{Q}_T^{-1} - Q_T^{-1})\phi_T$. Because $\Phi(L)$ has known finite order q by assumption, $\hat{\Phi}_j \xrightarrow{p} \Phi_j$ for $j = 1, \ldots, q$, $\hat{Q}_T - Q \xrightarrow{p} 0$ and $\hat{\phi}_T - \phi_T \xrightarrow{p} 0$. Finally, since Q is a.s. invertible, $(\Upsilon_T^{-1} \otimes I)(\hat{\delta}_{\mathrm{GLS}} - \bar{\delta}_{\mathrm{GLS}}) \xrightarrow{p} 0$. *Q.E.D.*

Proof of Theorem 2: (a) Write $(\Upsilon_T \otimes I)(\hat{\delta}_{\mathrm{OLS}} - \delta) = V_T^{-1}\omega_T$, where $V_T = [\Upsilon_T^{-1} \otimes I][(\Sigma_t z_t z'_t) \otimes I][\Upsilon_T^{-1} \otimes I]$ and $\omega_T = [\Upsilon_T^{-1} \otimes I][\Sigma_t(z_t \otimes I)v^l_t]$. The proof that $V_T \Rightarrow V$ follows from SSW Lemma 1. The proof that $\omega_{iT} \Rightarrow \omega_i$, $i \geq 2$ parallels the proof in Theorem 1: apply the 2-sided Beveridge-Nelson decomposition to $v^l_t = c_l(L)\varepsilon_t$, apply SSW Lemma 1 to the term involving $c_l(1)\varepsilon_t$, and use the absolute summability of $c^*_l(L)$ to show that the term involving $\Delta c^*_l(L)\varepsilon_t$ converges to 0 in probability. For $i = 1$, the result is a consequence of the central limit theorem for summable processes.

(b) Theorems 1 and 2 imply that $\Upsilon_{1T}^{-1}\Sigma_t z^1_t z^{*\prime}_t \Upsilon_{*T}^{-1} \xrightarrow{p} 0$. First consider the infeasible GLS estimator $\bar{\delta}_{\mathrm{GLS}}$, defined in the proof of Theorem 1. It is a consequence of the proof of Theorem 1 that

$$(\Upsilon_{*T} \otimes I)\left(\bar{\delta}_{*\mathrm{GLS}} - \delta_*\right) = B_{*T}^{-1}\left(\Upsilon_{*T}^{-1} \otimes I\right) \sum_t \left(z^*_t \otimes \Phi(1)'\right)\Phi(1)c_l(1)\varepsilon_t + o_p(1)$$

where $B_{*T} = (\Upsilon_{*T}^{-1} \otimes I)[\Sigma_t(z^*_t \otimes \Phi(1)')(z^{*\prime}_t \otimes \Phi(1))](\Upsilon_{*T}^{-1} \otimes I)$. Now

$$B_{*T} = \left(\Upsilon_{*T}^{-1} \otimes I\right)\left(I_{(g-g_1)k_l} \otimes \Phi(1)'\Phi(1)\right)\left[\sum_t \left(z^*_t z^{*\prime}_t \otimes I\right)\right]\left(\Upsilon_{*T}^{-1} \otimes I\right)$$

so

$$B_{*T}^{-1} = \left[\left(\Upsilon_{*T}^{-1} \otimes I\right)\sum_t \left(z^*_t \otimes I\right)\left(z^{*\prime}_t \otimes I\right)\left(\Upsilon_{*T}^{-1} \otimes I\right)\right]^{-1} \left[I \otimes \Phi(1)'\Phi(1)\right]^{-1}.$$

Also,

$$
\begin{aligned}
(\Upsilon_{*T}\otimes I)\left(\hat{\delta}_{*\mathrm{OLS}}-\bar{\delta}_{*\mathrm{GLS}}\right) &= B_{*T}^{-1}\left(\Upsilon_{*T}^{-1}\otimes I\right) \\
&\quad\times\left\{\sum_t \left(z_t^{*}\otimes \Phi(1)'\Phi(1)\right)c_l(L)\varepsilon_{tz}\right. \\
&\quad\left.-\sum_t \left(z_t^{*}\otimes \Phi(1)'\Phi(1)\right)c_l(1)\varepsilon_t\right\}+o_p(1) \\
&= B_{*T}^{-1}K_{*T}+o_p(1)
\end{aligned}
$$

where $K_{*T}=(\Upsilon_{*T}^{-1}\otimes I)\{\Sigma_t(z_t^{*}\otimes\Phi(1)'\Phi(1))\,\Delta c_l^{*}(L)\varepsilon_t\}$ and where the final line uses the two-sided Beveridge-Nelson decomposition, $c_l(L)\varepsilon_t=[c_l(1)+\Delta c_l^{*}(L)]\varepsilon_t$. The lag polynomial $c_l^{*}(L)$ is 1-summable (since $c_l(L)$ is 2-summable). Telescoping arguments were used in the proof of Theorem 1 to show that $(\Upsilon_{*T}^{-1}\otimes I)\{\Sigma_t(z_t^{*}\otimes\Phi(1)'\Phi(1))\,\Delta\zeta^{*}(L)\varepsilon_t\}\xrightarrow{p}0$, and those arguments applied here show that $K_{*T}\xrightarrow{p}0$. The result $(\Upsilon_{*T}\otimes I)(\hat{\delta}_{*\mathrm{OLS}}-\bar{\delta}_{*\mathrm{GLS}})\xrightarrow{p}0$ follows from $Q_{ijT}\Rightarrow Q_{ij}$ (from Theorem 1). *Q.E.D.*

PROOF OF THEOREM 3: The result follows from Theorem 1, which provides a limiting representation for the estimated coefficients; Theorem 2 of SSW, which provides a limiting representation for the Wald test statistic when the restrictions are on coefficients whose estimators converge at different rates; and Johansen (1988a) or alternatively Phillips (1991a), who show that the resulting limiting distributions are χ^2. *Q.E.D.*

PROOF OF THEOREM 4: This follows from Theorem 3 and the proof of Theorem 2. *Q.E.D.*

APPENDIX B

Data Sources

Money Supply (M1): The monthly Citibase M1 series (FM1) was used for 1959–1989; monthly data for 1947–1958 were formed by splicing the M1 series reported in *Banking and Monetary Statistics, 1941–1970* (Board of Governors of the Federal Reserve System), to the Citibase data in January 1959. The monthly data were averaged to obtain annual observations. Data prior to 1947 are those used by Lucas (1988); from 1900–1914 the data are from *Historical Statistics*, series X267 and from 1915–1946 they are from Friedman and Schwartz (1970, pp. 704–718, column 7).

Real Output: The annual data are U.S. Net National Product. For 1947–1989, they are the Citibase series GNNP82. Prior to 1947, we used Lucas' (1988) data (Friedman and Schwartz real net national product (1982, Table 4.8)). The Friedman and Schwartz series was linked to the annual Citibase series in 1947. For the monthly data analyzed in Section 7B, we used real personal income (Citibase Series GMPY82).

Prices: The annual data are the price deflator for U.S. Net National Product. For 1947–1989, they are the Citibase Series GDNNP. Prior to 1947, we used Lucas' (1988) data (Friedman and Schwartz (1982, Table 4.8)). The Friedman and Schwartz series was linked to the annual Citibase Series in 1947. For the monthly data analyzed in Section 7B, we used the price deflator for real personal income, formed as the ratio of the Citibase series GMPY to GMPY82.

Interest rates: The annual data are the rate on commercial paper. For 1947–1989, we used the 6-month commercial paper rate (Citibase Series FYCP). The monthly data were averaged to obtain annual observations. Prior to 1947, we used Lucas' (1988) data (Friedman and Schwartz (1982, Table 4.8, column 6)). For our analysis of the postwar monthly data, we also used the 90-day U.S. Treasury bill rate (Citibase Series FYGM3) and the 10-year U.S. Treasury bond rate (Citibase series FYGT10).

Summary of Results from Unit Root and Cointegration Tests

Univariate Dickey-Fuller (1979) $\hat{\tau}_\mu$ and $\hat{\tau}_\tau$ statistics, computed with 2 and 4 lags on the full data set, fail to reject a single unit root in each of $m, p, r, m-p$, and the logarithm of velocity at the 10% level; the unit root hypothesis is not rejected for y with 4 lags, but is rejected at the 10% (but not 5%) level with 2 lags. A unit root in Δy and Δr are each rejected at the 1% level, and a unit root in $\Delta m - \Delta p$ is rejected at the 10% level. Similar inferences obtain for the 1900–1945 and 1946–1989 subsamples. Whether m and p have two unit roots is less clear: for m, two unit roots are rejected in favor of one at the 10% level for the second but not the first subsample; for p, two unit roots are rejected for neither subsample, but this null is rejected at the 10% level for the full sample. For $r - \Delta p$ (Δp in percentages), one unit root is rejected (vs. zero) for the full sample at the 10% level, but not in either subsample using the $\hat{\tau}_\tau$ statistic; however, $\hat{\tau}_\mu$ rejects at 10% in both subsamples and the full sample.

Turning to the evidence on cointegration, Johansen's (1991) $J_T(0)$ test of the null of at most zero cointegrating vectors, against one or more cointegrating vector in the $(m-p, y, r)$ system, rejects at the 10% level using 1 or 2 lags over the full sample. Using the Stock-Watson (1988) $q_T^f(3,2)$ test for 3 vs. 2 unit roots (with 2 lags), the evidence is less strong: the p-value is .43. However, the Engle-Granger (1987) augmented detrended Dickey Fuller test (one lagged first difference) based on the residual from regressing $m-p$ on y and r rejects noncointegration at the 5% level over the full sample. Finally, demeaned ADF tests of the residual from the regression of the logarithm of velocity on r reject noncointegration at the 10% level (with one or two lagged first differences) over the full sample.

Details of these results are available from the authors on request.

REFERENCES

Ahn, S. K., and G. C. Reinsel (1990): "Estimation for Partially Nonstationary Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

Baba, Y., D. F. Hendry, and R. M. Starr (1992): "The Demand for M1 in the USA: 1960–1988," *The Review of Economic Studies*, 59, 25–60.

Balke, N. S., and R. J. Gordon (1986): "Historical Data," in *The American Business Cycle: Continuity and Change*, edited by R. J. Gordon. Chicago: University of Chicago Press, for the NBER.

Banerjee, A., J. J. Dolado, D. F. Hendry, and G. W. Smith (1986): "Exploring Equilibrium Relations 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.

Beveridge, Stephen, and Charles R. Nelson (1981): "A New Approach to Decomposition of Time Series in Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle'," *Journal of Monetary Economics*, 7, 151–174.

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

Brillinger, D. R. (1980): *Time Series, Data Analysis and Theory*, Expanded Edition. San Francisco: Holden-Day.

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.

——— (1989): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies*, 1, 195–228.

Dickey, D. A., and W. A. Fuller (1979): "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Engle, 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, 277–304.

Friedman, M., and A. J. Schwartz (1970): *Monetary Statistics of the United States*. New York: Columbia University Press for the National Bureau of Economic Research.

FRIEDMAN, M., AND A. J. SCHWARTZ (1982): *Monetary Trends in the United States and the United Kingdom*. Chicago: University of Chicago Press for the National Bureau of Economic Research.

GOLDFELD, S. M., AND D. E. SICHEL (1990): "The Demand for Money," Chapter 8 in *Handbook of Monetary Economics*, Vol. 1, ed. by B. M. Friedman and F. H. Hahn. Amsterdam: North-Holland, 299–356.

GRANGER, C. W. J., AND T-H. LEE (1990): "Multicointegration," *Advances in Econometrics*, 8, 71–84.

GREGORY, A. W., AND J. M. NASON (1991): "Testing for Structural Breaks in Cointegrated Relationships," Discussion Paper #827, Institute for Economic Research, Queen's University.

GRENANDER, U., AND M. ROSENBLATT (1957): *Statistical Analysis of Stationary Time Series*. New York: John Wiley and Sons.

HAFER, R. W., AND D. W. JANSEN (1991): "The Demand for Money in the United States: Evidence from Cointegration Tests," *Journal of Money, Credit and Banking*, 23, 155–168.

HANSEN, B. E. (1988): "Robust Inference in General Models of Cointegration," manuscript, Yale University.

——— (1989): "Efficient Estimation of Cointegrating Vectors in the Presence of Deterministic Trends," manuscript, University of Rochester.

HANSEN, B. E., AND P. C. B. PHILLIPS (1990): "Estimation and Inference in Models of Cointegration: A Simulation Study," *Advances in Econometrics*, 8, 225–248.

HENDRY, D. F., AND N. R. ERICSSON (1991): "Modeling the Demand for Narrow Money in the United Kingdom and the United States," *European Economic Review*, 35, 833–848.

HOFFMAN, D., AND R. H. RASCHE (1991): "Long-Run Income and Interest Elasticities of Money Demand in the United States," *Review of Economics and Statistics*, 73, 665–674.

JOHANSEN, S. (1988a): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–255.

——— (1988b): "The Mathematical Structure of Error Correction Models," *Contemporary Mathematics, Vol. 80: Structural Inference from Stochastic Processes*, ed. by N. U. Prabhu. Providence, RI: American Mathematical Society.

——— (1991): "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregression Models," *Econometrica*, 59, 1551–1580.

——— (1992): "A Representation of Vector Autoregressive Processes Integrated of Order 2," *Econometric Theory*, 8, 188–202.

JOHANSEN, S., AND K. JUSELIUS (1990): "Maximum Likelihood Estimation and Inference on Cointegration—with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, 169–210.

JUDD, JOHN P., AND J. L. SCADDING (1982): "The Search for a Stable Demand Function: A Survey of the Post-1973 Literature," *Journal of Economic Literature*, 20, 993–1023.

KING, R., C. PLOSSER, J. H. STOCK, AND M. W. WATSON (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–840.

LAIDLER, D. E. W. (1977): *The Demand for Money: Theories and Empirical Evidence*. New York: Dun-Donnelly.

LUCAS, R. E. (1988): "Money Demand in the United States: A Quantitative Review," *Carnegie-Rochester Conference Series on Public Policy*, 29, 137–168.

MELTZER, A. H. (1963): "The Demand for Money: The Evidence from the Time Series," *Journal of Political Economy*, 71, 219–246.

MILLER, STEPHEN (1991): "Monetary Dynamics: An Application of Cointegration and Error Correction," *Journal of Money, Credit and Banking*, 23, 139–154.

PARK, J. Y. (1992): "Canonical Cointegrating Regressions," *Econometrica*, 60, 119–144.

PHILLIPS, P. C. B. (1991a): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

——— (1991b): "Spectral Regression for Cointegrated Time Series," in *Nonparametric and Semiparametric Methods in Economics and Statistics*, ed. by W. Barnett. Cambridge: Cambridge University Press, 413–436.

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

PHILLIPS, P. C. B., AND M. LORETAN (1991): "Estimating Long Run Economic Equilibria," *The Review of Economic Studies*, 58, 407–436.

PHILLIPS, P. C. B., AND J. Y. PARK (1988): "Asymptotic Equivalence of OLS and GLS in Regression with Integrated Regressors," *Journal of the American Statistical Association*, 83, 111–115.

PHILLIPS, P. C. B., AND V. SOLO (1992): "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971–1001.

SAID, S. E., AND D. A. DICKEY (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–608.

SAIKKONEN, P. (1991): "Asymptotically Efficient Estimation of Cointegrating Regressions," *Econometric Theory*, 7, 1–21.

SIMS, C. A. (1972): "Money, Income and Causality," *American Economic Review*, 62, 540–552.

SIMS, C. A., J. H. STOCK, AND M. W. WATSON (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

STOCK, J. H. (1987): "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1035–1056.

STOCK, J. H., AND M. W. WATSON (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

WATSON, M. W. (1992): "Vector Autoregressions and Cointegration," forthcoming, *Handbook of Econometrics*, Volume IV.

WEST, K. D. (1988): "Asymptotic Normality when Regressors Have a Unit Root," *Econometrica*, 56, 1397–1418.