A SIMPLE ESTIMATOR OF COINTEGRATING VECTORS IN HIGHER ORDER INTEGRATED SYSTEMS

James H. Stock and Mark W. Watson

Working Paper Series
Macro Economic Issues
Research Department
Federal Reserve Bank of Chicago
February, 1991 (WP-91-3)

A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems

James H. Stock
Department of Economics
University of California, Berkeley
Berkeley, CA 94720

and

Mark W. Watson
Department of Economics
Northwestern University
Evanston, IL 60208
and the Federal Reserve Bank of Chicago

First Draft: September 1989

This revision: January 1991

Abstract

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 (.90, 1.03) and (-.124, -.088), respectively. Estimates based on the postwar data alone, however, are unstable, with variances that suggest substantial sampling uncertainty.

Keywords: Error correction models, unit roots, money demand

JEL classification number: 210

*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 versions of this paper. Helpful suggestions by Lars Hansen and two anonymous referees are gratefully acknowledged. The authors also thank Robert Lucas for kindly providing the annual data analyzed in Section 7. This research was supported in part by the Sloan Foundation and National Science Foundation grants SES-86-18984 and SES-89-10601.

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 that each series is individually integrated of order 1 (is I(1)), typically with no drift term. Johansen (1988a) and Ahn and Reinsel (1990) independently derived the asymptotic distribution of the Gaussian MLE when the cointegrated system is parameterized as a vector error correction model (VECM), and Johansen (1989) extended this result to the case of nonzero drifts. In a series of papers, Phillips and coauthors have considered efficient estimation based on a different model for cointegrated systems, the triangular representation. Phillips (1988a) studied estimation in a cointegrated model with general I(0) errors; Phillips and Hansen (1989) considered a two-step zero frequency seemingly unrelated regression estimator; and Phillips (1988b) used spectral methods to compute efficient estimators in the frequency domain.

This paper proposes two alternative, computationally simple estimators of cointegrating vectors, which readily extend to systems with arbitrary deterministic components and with higher orders of integration and cointegration. The estimators are motivated as Gaussian MLE's for a particular parameterization of the triangular representation. However, under more general conditions they are shown to be asymptotically efficient in Phillips' (1988a) 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. It is argued that the resulting "dynamic OLS" (respectively GLS) estimators are asymptotically equivalent to the Johansen/Ahn-Reinsel MLE. The proposed estimators treat the parameters describing the short-run dynamics of the process as nuisance parameters; the object is to obtain efficient estimates of the cointegrating vector, which are typically of independent interest. If desired, the estimates subsequently can be used to study the short-run dynamics, say by estimating a triangular system (Campbell (1987), Campbell and Shiller (1987, 1989)) or a constrained VECM (King, Plosser, Stock and Watson (1987)).

These estimators are used here to investigate the long-run demand for money (M1) in the U.S. from 1900 to 1989. Other researchers (recently including Hoffman and Rasche [1989] and Baba, Hendry and Starr [1990]) 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 the 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's (1988) suggestion that there is a stable long-run M1 money demand relation spanning the twentieth century.

The paper is organized as follows. The model and estimators are introduced in Section 2 for I(1) variables, and they 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 proposed estimators are related to Johansen's (1988a) MLE, and 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 3-6 with little loss of continuity.[1]

2. Representation and Estimation in I(1) Systems

Suppose that each element of the n-dimensional time series y_t is I(1), that $E\Delta y_t = 0$, and that the n×r matrix of r cointegrating vectors α is $\alpha = (-\theta \;\; I_r)'$, where θ is the r×(n-r) submatrix of unknown parameters to be estimated and I_r is the r×r identity matrix. The triangular representation for y_t is,

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

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

where y_t is partitioned as (y_t^1, y_t^2), where y_t^1 is (n-r)×1 and y_t^2 is r×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 (1988a,1988b), 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 $\{u_t^1\}$. Because u_t is Gaussian and stationary, $E[u_t^2|\{\Delta y_t^1\}]=E[u_t^2|\{u_t^1\}]=d(L)\Delta y_t^1$, where d(L) is in general two-sided and $\{\Delta y_t^1\}$ denotes $\{\Delta y_t^1, t=1,\ldots,T\}$. Thus (2.1b) can be written,

(2.2) $$y_t^2 = \theta y_t^1 + d(L)\Delta y_t^1 + \tilde{u}_t^2$$

where $\tilde{u}_t^2 = u_t^2 - E[u_t^2 | \{u_t^1\}]$. By construction, $\{\Delta y_t^1\}$ and $\{\tilde{u}_t^2\}$ are independent. In addition, u_t^1 and $\tilde{u}_t^2$ have Wold representations $u_t^1 = c_{11}(L)\epsilon_t^1$ and $\tilde{u}_t^2 = c_{22}(L)\epsilon_t^2$, where $\{\epsilon_t^1\}$ and $\{\epsilon_t^2\}$ are independent. Thus (2.1a) and (2.2) can be written,

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

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

where ϵ_t is NIID$(0,\Sigma_\epsilon)$, $\Sigma_\epsilon = \mathrm{diag}(\Sigma_{11}, \Sigma_{22})$, so $\{\epsilon_t^2\}$ is independent of $\{y_t^1\}$.

The two-sided triangular representation (2.3) provides a nonstandard asymptotic factorization of the Gaussian likelihood. Let λ_1 denote the parameters of $c_{11}(L)$ and Σ_{11}, let λ_2 denote the parameters of $d(L)$, $c_{22}(L)$, and Σ_{22}, and let Y^i denote $(y_1^i,\ldots,y_T^i)$, $i=1,2$. Then (2.3) implies that the likelihood can be factored as

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

This differs from the usual prediction-error factorization because the conditional mean of y_t^2 involves future as well as past value of y_t^1.

The representations (2.3) and (2.4) provide concrete guides to 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 can be obtained by maximizing $f(Y^2|Y^1,\theta,\lambda_2)$. This reduces to estimating the parameters of the regression equation (2.3b) 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 (2.3b) 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 against only a constant and y_t^1. Similarly, the feasible GLS estimator of θ in (2.3b) will be referred to as the dynamic GLS (DGLS) estimator.

The representation (2.3) 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 $c_{22}(L)^{-1}$ has finite order p, then (2.3b) can be rewritten as an r-dimensional error correction model, i.e. as a regression of Δy_t^2 onto (Δy_t^1, Δy_{t-1}, $\Delta y_{t-2}, \ldots, \Delta y_{t-p}$, $y_{t-1}-\theta' y_{t-1}$). In this case, the nonlinear least squares estimator (with Δy_t^1 included as a regressor; 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 (2.3). Because ϵ_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 $\theta=\theta_0$ (where rank(θ)=r is maintained) 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 (1988a) provides an insightful discussion of the asymptotic mixed normal property of the MLE of θ, the local asymptotic mixed normal (LAMN) behavior of test statistics, and the efficiency of the Gaussian maximum likelihood estimator of θ.[2]

Third, although the interpretation of (2.3) as a factorization of the likelihood (2.4) assumes Gaussianity, the two-sided triangular representation with $E\epsilon_t^2 \Delta y_{t-j}^{1\prime}=0$ for all j can be constructed under the weaker condition that u_t is linearly regular and covariance stationary. This result, summarized in the following lemma (proven in the Appendix), provides an alternative to the Wold representation theorem.[3] Let Ψ_t^j and Ψ_∞^j denote the linear manifolds spanned by $\{u_s^j\}$, $-\infty<s\le t$ and $\{u_s^j\}$, $-\infty<s<\infty$ (respectively), closed with respect to convergence in mean square, and let $P(u_t^i|\Psi_s^j)$ denote the projection of u_t^i onto Ψ_s^j.

Lemma 2.1. Let $u_t = (u_t^{1\prime}\ u_t^{2\prime})'$ be a mean zero linearly regular weakly stationary stochastic process with $Eu_tu_t' < \infty$, where u_t^1 is $(n-r)\times 1$ and u_t^2 is $r\times 1$. Then u_t has the representation,

$$\begin{bmatrix} u_t^1 \\ u_t^2 \end{bmatrix} = \begin{bmatrix} c_{11}(L) & 0 \\ c_{21}(L) & c_{22}(L) \end{bmatrix} \begin{bmatrix} \epsilon_t^1 \\ \epsilon_t^2 \end{bmatrix} \tag{2.5}$$

or $u_t=c(L)\epsilon_t$ (where $c(L)$ and ϵ_t are partitioned conformably with u_t), where $E\epsilon_t=0$, $E\epsilon_t\epsilon_t'=\text{diag}(\Sigma_{11}, \Sigma_{22})$ and $E\epsilon_t\epsilon_s'=0$ for $t\neq s$, and where $\{\epsilon_t^1\}$ are innovations of $\{u_t^1\}$ and $\{\epsilon_t^2\}$ are innovations of $\{\tilde{u}_t^2=u_t^2-P(u_t^2|\Psi_\infty^1)\}$. Also, $c_{11}(L)$ and $c_{22}(L)$ are one-sided, in general $c_{21}(L)$ is two-sided, and $c(L)$ is square summable.

In the representation (2.5), $c_{21}(z)$ is in general two-sided because $\{\epsilon_t^1\}$ are constructed to be the innovations of $\{u_t^1\}$. Note however that $\{\epsilon_t\}$ are not innovations for $\{u_t\}$.[4]

3. Representation in I(d) Systems

This section extends the triangular representation (2.3) to systems in which variables may be integrated and cointegrated of different orders and in which there are polynomial time trends. The I(d) generalization of (2.1), derived in the Appendix from the Wold representation of an I(d) system with drifts and multiple cointegrating vectors, is

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

for $t=1,\ldots,T$, where the y_t^j are $k_j\times 1$ vectors which form a partition of y_t, i.e. $y_t=(y_t^{1\prime}\ y_t^{2\prime}\ \ldots\ y_t^{d+1\prime})'$, and $\Delta^2=(1-L)(1-L)$, etc. By assumption, $u_t=(u_t^{1\prime}\ u_t^{2\prime}\ \ldots\ u_t^{d+1\prime})'$ is weakly stationary with mean zero. It is assumed that the highest order of integration of the elements of y_t is I(d). However, not all elements of y_t need to be I(d) for (3.1) to apply (see the examples in Section 5(B) and the empirical application to long-run money demand in Section 7). Moreover, some blocks of (3.1) might not appear. For example, with d=2 and n=2, y_t could be CI(2,1) in Engle and Granger's (1987) terminology, in which case $k_1=1$, $k_2=1$, and $k_3=0$; if y_t is CI(2,2), then $k_1=1$, $k_2=0$, and $k_3=1$.

This representation 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 distinct 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 distinct I(d-1) elements in the system; for rows of $\theta_{2,1}^{d-1}$ that 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 distinct I(d-2) components, and so forth. It is straightforward to generalize the representation (3.1) to include higher order polynomials in t, or to specialize it to the common case in which higher-order polynomials are suppressed.

By assumption, (3.1) requires contemporaneous linear combinations of levels of y_t, or levels of y_t plus differences, to be integrated of at least order zero so that u_t has a finite nonsingular spectral density matrix. This assumption is made without loss of generality, and serves to fix the maximum order of integration d. In practice, this is achieved by replacing y_t by $\Delta^{-1}y_t$ or Δy_t as needed for u_t to be I(0) and not cointegrated.

As in the I(1) case, the errors are orthogonalized by projecting onto leads and lags of the errors in the preceding equations. By repeated application of Lemma 2.1, $u_t=(u_t^{1\prime}\ u_t^{2\prime}\ \ldots\ u_t^{d+1\prime})'$ has the representation

$$u_t = C(L)\epsilon_t \tag{3.2}$$

where $\epsilon_t=(\epsilon_t^{1\prime}\ \epsilon_t^{2\prime}\ \ldots\ \epsilon_t^{d+1\prime})'$ and $E\epsilon_t\epsilon_t'=\Sigma_\epsilon=\mathrm{diag}(\Sigma_{11},\Sigma_{22},\ldots,\Sigma_{d+1,d+1})$, and where C(L) is a block lower triangular matrix partitioned conformably with u_t, with diagonal blocks $c_{ii}(L)$ that are one-sided lag polynomials and with lower off-diagonal blocks $c_{ij}(L)$ that in general are two-sided.

We focus our attention on the first ℓ blocks of equations, and make the additional assumption that $\{c_{ii}(L)\}$, $i=1,\ldots,\ell-1$ are invertible, so that the first ℓ blocks of (3.2) can be written,

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ -d_{21}(L) & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ -d_{\ell 1}(L) & -d_{\ell,\ell-1}(L) & \cdots & 1 \end{bmatrix} u_t = \begin{bmatrix} c_{11}(L) & & 0 \\ & \ddots & \\ 0 & & c_{\ell\ell}(L) \end{bmatrix} \epsilon_t \tag{3.3}$$

where in general $d_{ij}(L)$ are two-sided (for example, $d_{21}(L)=c_{21}(L)c_{11}(L)^{-1}$). Substitution of the ℓ-th equation in (3.3) into (3.1) then yields,

$$(3.4)\quad \Delta^{d-\ell+1}y_t^{\ell} = \sum_{j=0}^{\ell-1} \bar{\mu}_{\ell,j}t^j + \sum_{j=1}^{\ell-1}\sum_{i=j}^{\ell-1}\theta_{\ell,j}^{d-i}(\Delta^{d-i}y_t^j)$$
$$+ \sum_{m=1}^{\ell-1}d_{\ell m}(L)[(\Delta^{d-m+1}y_t^m)-\sum_{j=1}^{m-1}\sum_{i=j}^{m-1}\theta_{m,j}^{d-i}(\Delta^{d-i}y_t^j)] + c_{\ell\ell}(L)\epsilon_t^{\ell}$$

where $\{\bar{\mu}_{\ell,j},\ j=1,\ldots,\ell-1\}$ are functions of $\{\mu_{m,j},\ j=1,\ldots,m-1,\ m=1,\ldots,\ell\}$. The lag polynomials $\{d_{\ell m}(L)\}$, which generalize $d(L)$ in (2.3), arise from the projection of u_t^{ℓ} onto $\{u_t^m\}$ for $m=1,\ldots,\ell-1$. When $\{y_t\}$ is Gaussian, ϵ_t is $NIID(0,\Sigma_\epsilon)$.[5]

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

As in the I(1) case, one motivation for considering (3.4) is as the conditional mean of a nonstandard factorization of the Gaussian likelihood. Let λ_ℓ denote the parameters of $c_{\ell\ell}(L)$ and $d_{\ell m}(L)$, $m=1,\ldots\ell-1$, let θ_ℓ denote $\{\theta_{\ell,j}^{d-i},\ j=1,\ldots,\ell-1,\ i=j,\ldots,\ell-1\}$, let μ_ℓ denote $\{\bar{\mu}_{\ell,j}, j=0,\ldots,\ell-1\}$, and let λ, θ, and μ represent the collection of λ_ℓ, θ_ℓ, and μ_ℓ. The Gaussian likelihood can be written as:

$$(3.5)\qquad f(Y,\theta,\mu,\lambda) = f(Y^{d+1}|Y^1{}_T,\ldots,Y^d,\theta_2,\theta_3,\ldots,\theta_{d+1},\mu_{d+1},\lambda_{d+1})$$
$$\times f(Y^d|Y^1,\ldots,Y^{d-1},\theta_2,\theta_3,\ldots,\theta_d,\mu_d,\lambda_d)\cdots$$
$$\times f(Y^2|Y^1,\theta_2,\mu_2,\lambda_2)f(Y^1,\mu_1,\lambda_1)$$

where $Y=(y_1', y_2', \ldots, y_T')'$ and $Y^i=(y_1^{i\prime}, y_2^{i\prime}, \ldots, y_T^{i\prime})'$ for

$i=1,\ldots,d+1$. If the parameters $(\theta_2,\ldots,\theta_{\ell-1})$ of the higher-order cointegrating vectors in (3.1) are known and if there are no restrictions between $(\lambda_\ell,\theta_\ell,\mu_\ell)$ and $\{(\lambda_j,\theta_j,\mu_j),j<\ell\}$, then $(Y^1,\ldots,Y^{\ell-1})$ are ancillary for θ_ℓ. In this case the MLE of θ_ℓ is obtained by estimating the system (3.4) by GLS. If some of $(\theta_2,\ldots,\theta_{\ell-1})$ are unknown then $(Y^1,\ldots,Y^{\ell-1})$ are no longer ancillary (weakly exogenous) for θ_ℓ, and the GLS estimator of $\{\theta^{d-i}_{\ell,j}\}$ in (3.4) is not the exact Gaussian MLE. However, as is made precise in Section 5 for the I(2) case, the DGLS and DOLS estimators of $\{\theta^{d-i}_{\ell,j}\}$ in (3.4) still have desirable properties.

We therefore consider regression estimators of (3.4). Because the regressors in (3.4) can have stochastic or deterministic trends in common, it is convenient to transform the regressors to isolate these different trends. Let X_t denote the regressors in each equation in (3.4). It is assumed that X_t is a known nonrandom (not data-dependent) linear combination of $Y^1,\ldots,Y^{\ell-1}$. Define $z_t = DX_t$, where D is an invertible matrix of constants (possibly unknown), chosen so that z_t are the canonical regressors in the sense of Sims, Stock and Watson (1990).

The choice of transformation matrix D depends on the specific application (see Section 5(B) for examples). In general, partition z_t as $(z_t^{1\prime}\ z_t^{2\prime}\ \ldots\ z_t^{2\ell\prime})'$, where by construction z_t^1 is an I(0) vector with mean zero (z_t^1 contains the required leads and lags of $\{u_t^m,\ m<\ell\}$, dictated by the polynomials $\{d_{\ell m}(L)\}$), $z_t^2=1$, z_t^3 is dominated by a martingale, $z_t^4=t$, z_t^5 is dominated by integrated martingales, $z_t^6=t^2$, and so forth. In general $\sum_{t=1}^T z_t^i z_t^{i\prime}$ is $O_p(T^{i-1})$ for $i\geq2$. From Sims, Stock, and Watson (1990, Section 2), z_t can be written as $z_t = G(L)\nu_t$, where G(L) is a block lower triangular matrix and $\nu_t=(\xi_t^{0\prime}\ 1\ \xi_t^{1\prime}\ t\ \xi_t^{2\prime}\cdots\xi_t^{\ell-1\prime}\ t^{\ell-1})'$, where $\xi_t^0=(\epsilon_t^{1\prime}\ \epsilon_t^{2\prime}\cdots\epsilon_t^{\ell-1\prime})'$ and where ξ_t^j is defined recursively by $\xi_t^j=\sum_{s=1}^t\xi_s^{j-1}$ for $j\geq1$. Also, let g_i denote the dimension of z_t^i, and let $g=\sum_{i=1}^{2\ell}g_i$ be the dimension of z_t.

With these definitions, the system (3.4) equivalently can be written as, (3.6) $$\Delta^{d-\ell+1}y_t^\ell = (X_t' \otimes I_{k_\ell})\beta + e_t$$

or

(3.7) $$\Delta^{d-\ell+1}y_t^\ell = (z_t' \otimes I_{k_\ell})\delta + e_t$$

where $e_t = c_{\ell\ell}(L)\epsilon_t^\ell$, $E\epsilon_t^\ell\epsilon_t^{\ell\prime} = \Sigma_{\ell\ell}$, and $E[\epsilon_t^\ell|\{z_t\}]=0$. The coefficients β are the coefficients on the regressors appearing in (3.4); these are related to the coefficients on the canonical regressors in (3.7) by $\beta=(D'\otimes I_{k_\ell})\delta$. 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 $\delta=(\delta_1' \ \delta_2' \ \ldots \ \delta_{2\ell}')'$, where δ_i is the $g_i k_\ell$-vector of coefficients on z_t^i.

4. Estimation and Testing

This section examines the least squares estimation of the parameters δ in (3.7). It is assumed that y_t has the triangular representation (3.1) with u_t given by (3.3). We consider the case that z_t and δ are finite dimensional, i.e. in which $\{d_{\ell m}(L), m<\ell\}$ have fixed finite orders. It is assumed that $\{\epsilon_t\}$ in (3.2) is a martingale difference sequence with $E[\epsilon_t\epsilon_t'|\epsilon_{t-1},\epsilon_{t-2},\ldots]=\Sigma_\epsilon=\mathrm{diag}(\Sigma_{11}, \Sigma_{22},\ldots, \Sigma_{d+1,d+1})$ nonsingular and $\max_i \sup_t E[(\epsilon_{it})^4|\epsilon_{t-1},\epsilon_{t-2},\ldots]<\infty$.

There are two natural estimators of the parameters in (3.6): the feasible GLS estimator based on an estimator of $c_{\ell\ell}(L)$, suitably parameterized, and the dynamic OLS estimator (respectively, the DGLS and DOLS estimators). These are

(4.1) $$\hat\delta_{GLS} = [\sum\tilde z_t\tilde z_t']^{-1}[\sum\tilde z_t(\Delta^{d-\ell+1}\tilde y_t^\ell)]$$

(4.2) $$\hat\delta_{OLS} = [(\sum z_t z_t') \otimes I_{k_\ell}]^{-1}[\sum (z_t \otimes I_{k_\ell})(\Delta^{d-\ell+1}y_t^\ell)]$$

where $\tilde{z}_t=[z_t \otimes \hat{\Phi}(L)']$ and $\tilde{y}_t^{\ell} = \hat{\Phi}(L)y_t^{\ell}$, where $\hat{\Phi}(L)$ is an estimator of $\Phi(L)=\Sigma_{\ell\ell}^{-1/2}c_{\ell\ell}(L)^{-1}$.

Associated with the DGLS estimator is the Wald statistic testing the h restrictions $R\delta=r$ (where R and r have dimensions $h\times gk_{\ell}$ and $h\times1$, respectively),

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

Because the disturbance in (3.6) is serially correlated, the Wald statistic for $\hat{\delta}_{OLS}$ must be constructed using a modified covariance matrix. When the hypotheses of interest do not involve the coefficients on the mean-zero stationary regressors in (3.6), this is the spectral density matrix of e_t at frequency zero, $\Omega_{\ell\ell}=c_{\ell\ell}(1)\Sigma_{\ell\ell}c_{\ell\ell}(1)'$, estimated by $\hat{\Omega}_{\ell\ell}$. That is,

$$(4.4) \qquad W_{OLS} = [R\hat{\delta}_{OLS}-r]'\{R[(\sum z_t z_t') \otimes \hat{\Omega}_{\ell\ell}]^{-1}R'\}^{-1}[R\hat{\delta}_{OLS}-r] \ .$$

Define the $g\times g$ scaling matrix T_T to be a block diagonal matrix partitioned conformably with z_t, with diagonal blocks $T_{1T}=T^{1/2}I_{g_1}$ and $T_{iT} = T^{(i-1)/2}I_{g_i}$ for $i>2$. Also, for w_t weakly stationary, define $\Gamma_w(j) = E[w_t-E(w_t)][w_{t-j}-E(w_t)]'$.

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

Theorem 4.1 Suppose that y_t satisfies (3.4) and (3.6) where $c_{jj}(L)$ is d+1-j summable, j=1,...,d+1, that $c_{\ell\ell}(L)^{-1}$ has known order $q<\infty$, and $d_{\ell m}(L)$ has a known finite order. Then $(T_T \otimes I_{k_{\ell}})(\hat{\delta}_{GLS}-\delta) \Rightarrow Q^{-1}\phi$, where after partitioning Q and ϕ conformably with δ:

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

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

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

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

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

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

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

$\phi_1 \sim N(0,E[(z_t^1 \otimes \Phi(L)')\Sigma_{\ell\ell}(z_t^{1\prime} \otimes \Phi(L))])$,

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

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

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

Theorem 4.2. Under the assumptions of Theorem 4.1,

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

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

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

$\omega_m = \int_0^1(G_{mm}(1)W_1^{(m-1)/2}(s) \otimes \Omega_{\ell\ell}^{1/2})dW_2(s)$, $m=3,5,7,\ldots,2\ell-1$,

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

(b) Partition $\delta = (\delta_1{}'\ \delta_*{}')'$ so that δ_1 denotes the g_1k_ℓ elements of δ corresponding to z_t^1 and δ_* corresponds to the remaining $(g-g_1)k_\ell$ elements of δ. Similarly partition $\hat{\delta}_{OLS}$, $\hat{\delta}_{GLS}$, $z_t=(z_t^{1\prime}\ z_t^{*\prime})'$, and $T_T=\text{diag}(T_{1T},T_{*T})$. Then $(T_{*T} \otimes I_{k_\ell})(\hat{\delta}_{*OLS}-\hat{\delta}_{*GLS}) \xrightarrow{p} 0$.

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

Theorem 4.4. Suppose that the first g_1k_ℓ columns of R equal zero and that $\hat{\Omega}_{\ell\ell} \xrightarrow{p} \Omega_{\ell\ell}$. Then under the assumptions of Theorem 4.1, $W_{OLS} - W_{GLS} \xrightarrow{p} 0$ and $W_{OLS} \Rightarrow \chi^2_h$.

Note that $c_{\ell\ell}(L)^{-1}$ must be be finitely parameterized to implement the DGLS estimator. Although this is not strictly needed for the DOLS estimator, $\Omega_{\ell\ell}$ and therefore $c_{\ell\ell}(1)$ must be consistently estimated to construct W_{OLS}, which in practice entails estimating a parametric approximation.

The asymptotic equivalence of the DOLS and DGLS estimators of δ_* (Theorem 4.2(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]) and the results of Phillips and Park (1986) for I(1) regressors to the general integrated regression model with regressors of various orders of integration.

Although the theorems are stated in terms of δ, typically the results are translated into results on the coefficients of interest, β. The distribution of $\hat{\beta}_{GLS}$ is obtained from $\hat{\beta}_{GLS}=(D'\otimes I_{k_\ell})\hat{\delta}_{GLS}$ and $\beta=(D'\otimes I_{k_\ell})\delta$, and similarly for $\hat{\beta}_{OLS}$. Moreover, the Wald statistic testing $R\delta=r$ equivalently tests $P\beta=r$, where $P=R(D'^{-1}\otimes I_{k_\ell})$. Theorem 4.3 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 regresssors, Theorem 4.4 implies that the Wald test of $P\beta=r$ based on $\hat{\beta}_{OLS}$ (with a serial correlation-robust covariance matrix) is asymptotically χ^2. Importantly, the result concerning the asymptotic χ^2 distribution of the Wald statistic testing restrictions on cointegrating vectors 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 the case that there are a fixed number of regressors. Conceptually, one could view this estimator as semiparametric by embedding this parametric regression in a sequence of regressions where the number of regressors increase as a function of the sample size. A formal treatment of this extension would entail generalizing the univariate I(0) results of Berk (1974) and the univariate I(1) results of Said and Dickey (1984) to the I(d), vector-valued case, an extension not undertaken here.[6]

5. Examples

This section examines two examples in detail. The first compares the model of Section 2, and the associated dynamic OLS and GLS estimators, with the I(1) VECM formulation studied in Engle and Granger (1987) and Johansen (1988a, 1988b). The second example examines various cases of the I(2) specialization of (3.1).

A. Comparison with the I(1) vector error correction model. One representation of a purely stochastic I(1) cointegrated system is the VECM,

$$(5.1)\quad \Delta y_t = \gamma\alpha' y_{t-1} + A(L)\Delta y_{t-1} + \zeta_t,\ \zeta_t\ \ \text{NIID}(0,\Sigma_\zeta),\ t=1,\ldots,T$$

where y_t is n×1, A(L) has finite order and is unrestricted, α is a n×r matrix of cointegrating vectors and Σ_ζ is unrestricted. Johansen (1988a) and Ahn and Reinsel (1990) derived the limiting distribution of the Gaussian MLE $\hat{\alpha}_{MLE}$ for the unknown parameters of α in (5.1). Here, the MLE for (5.1) is related to the DOLS estimator.

Because the asymptotic information matrix for $(\hat{\alpha}_{MLE}, \hat{A}_{MLE}(z))$ in (5.1) is block diagonal, let A(L)=0. Partition $y_t=(y_t^{1\prime}\ y_t^{2})'$ into a (n-r)- and r-vector, partition $\gamma=(\gamma_1'\ \gamma_2)'$ conformably, where γ_1 is (n-r)×r and γ_2 is r×r and normalize α as $\alpha=(-\theta\ I_r)'$, where θ is r×(n-r). Without loss of generality assume that γ_2 is nonsingular. The block triangular form of (5.1) is,

(5.2a) $$\Delta y_t^1 = u_t^1,\quad u_t^1=\gamma_1 u_{t-1}^2 + \zeta_t^1$$

(5.2b) $$y_t^2 = \theta y_t^1 + u_t^2,\quad u_t^2=\textstyle\sum_{j=0}^{\infty}\rho^j\alpha'\zeta_{t-j},\quad \rho=I_r+\alpha'\gamma\ .$$

Clearly u_t^2 depends on θ and, if $\gamma_1\neq0$, so does u_t^1, so the factorization of the likelihood (2.4) results in restrictions between θ and λ_2 and between λ_1 and (θ,λ_2). In this model, the exact MLE $(\hat{\theta}_{MLE})$ is the system estimator studied by Johansen (1988a) and Ahn and Reinsel (1990), not the single equation estimator examined in Section 4.

To study the behavior of $\hat{\theta}_{MLE}$, it is convenient to reparameterize the VECM (5.1) (with A(L)=0) as:

(5.3a) $$\Delta y_t^1 = \Pi\Delta y_t^2 + \eta_t^1$$

(5.3b) $$\Delta y_t^2 = \beta_1 y_{t-1}^1 + \beta_2 y_{t-1}^2 + \eta_t^2$$

where $\Pi = \gamma_1\gamma_2^{-1}$, $\beta_1=-\gamma_2^{-1}\theta$, $\beta_2=\gamma_2$, $\eta_t^1=\zeta_t^1-\Pi\zeta_t^2$ and $\eta_t^2=\zeta_t^2$. This parameterization is convenient because (5.3) is an unconstrained linear triangular simultaneous equation model so that MLE's correspond to iterated SUR estimators (see Lahiri and Schmidt [1978]). The MLE's of the cointegrating vectors in (5.1) can be recovered from the MLE's of (5.3) as $\hat{\theta}_{MLE}=-\hat{\beta}_{2,MLE}^{-1}\hat{\beta}_{1,MLE}$. Isolating the regressors of different orders of integration, equation (5.3b) can be written as:

$$(5.3c) \qquad \Delta y_t^2 = \delta_1 z_t^1 + \delta_3 z_t^3 + \eta_t^2$$

where $z_t^1=y_{t-1}^2-\theta y_{t-1}^1=\alpha' y_{t-1}$, $z_t^3=y_{t-1}^1$, $\delta_1=\beta_2=\gamma_2$, and $\delta_3=\beta_1+\beta_2\theta$. In this parameterization z_t^3 are canonical I(1) regressors, z_t^1 are mean zero I(0) regressors, the true value of δ_3 is zero, and $\hat{\theta}_{MLE}-\theta=-\hat{\delta}_{1,MLE}^{-1}\hat{\delta}_{3,MLE}$. From (5.a) and (5.c), $\hat{\delta}_{1,MLE}$ and $\hat{\delta}_{3,MLE}$ can be written as OLS estimators from the regression of Δy_t^2 onto z_t^1, z_t^3 and $\hat{\eta}_t^1=\Delta y_t^1-\hat{\Pi}_{MLE}\Delta y_t^2$. Because $\hat{\Pi}_{MLE}$ and $\hat{\delta}_{1,MLE}$ are consistent, $\Sigma y_t^1 y_t^{1\prime}$ is $O_p(T^2)$, and $\Sigma y_{t-1}^1 \Delta y_t'$ is $O_p(T)$, by direct calculation,

$$(5.4) \quad T(\hat{\theta}_{MLE}-\theta) = -\hat{\delta}_{1,MLE}^{-1} T\hat{\delta}_{3,MLE} = -\gamma_2^{-1}(T^{-1}\textstyle\sum_{t=2}^{T} a_t y_{t-1}^{1\prime})(T^{-2}\textstyle\sum_{t=2}^{T} y_{t-1}^1 y_{t-1}^{1\prime})^{-1} + o_p(1)$$

where $a_t=\eta_t^2-E(\eta_t^2|\eta_t^1)$.[7]

The single equation estimators are obtained from the regression,

$$(5.5) \qquad y_t^2 = \theta y_t^1 + d(L)\Delta y_t^1 + e_t \ .$$

For finite order VECM's, d(L) will typically be infinite order, so the finite approximations used in Sections 2-4 will result in misspecified regression equations. If, however, the order of d(L) (say q) is such that $q\to\infty$ as $T\to\infty$ and $q^3/T\to 0$, then the results of Berk (1974) and Said and Dickey (1984) suggest that this misspecification will vanish asymptotically. With this interpretation, the single-equation dynamic OLS estimator can be written,

$$(5.6) \quad T(\hat{\theta}_{OLS}-\theta) = (T^{-1}\sum_{t=1}^{T}\bar{e}_t y_t^{1\prime})(T^{-2}\sum_{t=1}^{T}y_t^1 y_t^{1\prime})^{-1} + o_p(1)$$

where $\bar{e}_t=[-\Omega_{11}^{-1}\Omega_{12} \quad I_2]'\zeta_t$ is the long run component of e_t, where $\Omega=c(1)\Sigma_\zeta c(1)'$ is (2π times) the spectral density matrix of u_t at frequency zero.

A straightforward calculation demonstrates that $\bar{e}_t=-\gamma_2^{-1}a_{t+1}$, so that $T(\hat{\theta}_{MLE}-\hat{\theta}_{OLS}) \xrightarrow{p} 0$. Thus, even though the VECM likelihood cannot be factored as in (2.4), the single equation estimator is asymptotically equivalent to the MLE for this model. An interpretation is that, even though there are constraints across equations and across parameters that appear when the triangular system (5.2) is derived from a VECM, these constraints only involve coefficients that obey conventional $\sqrt{T}$ asymptotics. Thus, asymptotically they convey no information about θ, beyond that contained in the unrestricted second equation. Indeed this is a general property of regressions involving integrated regressors: the asymptotic distribution of estimators of coefficients on canonical integrated regressors remains unaltered when efficient estimators of coefficients on zero mean I(0) regressors are replaced by consistent esimators. One example of this is given by the MLE of θ for (5.1); asymptotically equivalent estimators can be constructed as $\tilde{\theta}=-\tilde{\beta}_2^{-1}\tilde{\beta}_1$, where $\tilde{\beta}_1$ and $\tilde{\beta}_2$ are the OLS estimators from the regression of Δy_t^2 onto y_{t-1}^1, y_{t-1}^2 and $\tilde{\eta}_t^1=\Delta y_t^1-\tilde{\Pi}\Delta y_t^2$, where $\tilde{\Pi}$ is any consistent estimator of Π in equation (5.3a).

Their asymptotic equivalence notwithstanding, it is useful to think of DOLS and $\hat{\theta}_{MLE}$ as applying to two different models. For finite order VECM's, Johansen's (1988a) estimator is the MLE, while for models that support the factorization (2.4) with d(L) finite order, the single equation estimators are the MLE.

B. Examples of I(2) Systems. The following examples concern specification and inference in general I(2) systems. To simplify exposition, all deterministic terms are omitted and their coefficients are taken to be zero. From (3.1), the general I(2) model is,

$$\Delta^2 y_t^1 = u_t^1 \tag{5.7a}$$

$$\Delta y_t^2 = \theta_{2,1}^1 \Delta y_t^1 + u_t^2 \tag{5.7b}$$

$$y_t^3 = \theta_{3,1}^1 \Delta y_t^1 + \theta_{3,1}^0 y_t^1 + \theta_{3,2}^0 y_t^2 + u_t^3 \ . \tag{5.7c}$$

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 (i.e. $k_2=0$). 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. It is assumed that $\{d_{\ell,m}(L)\}$ in (3.3) have known finite order.

Case 1: $k_2=0$. Then (5.7b) does not appear in the system and y_t^2 is omitted from (5.7c). 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 and nonzero.* Then the estimation equation (3.4) becomes,

$$y_t^3 = \theta_{3,1}^1 \Delta y_t^1 + \theta_{3,1}^0 y_t^1 + \theta_{3,2}^0 y_t^2 + d_{3,1}(L)\Delta^2 y_t^1 + d_{3,2}(L)(\Delta y_t^2 - \theta_{2,1}^1 \Delta y_t^1) + \tilde{u}_t^3 \tag{5.8}$$

where $E\tilde{u}_t^3 u_s^{1\prime}$ and $E\tilde{u}_t^3 u_s^{2\prime}$ 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 $z_t^3=(\Delta y_t^1,\ y_t^2-\theta_{2,1}^1 y_t^1)'$ and $z_t^5=y_t^1$ (other assignments of z_t are possible but yield the same results). The coefficients on z_t^3 and z_t^5 are respectively $\delta^3=(\delta_1^3,\ \delta_2^3)=(\theta_{3,1}^1,\ \theta_{3,2}^0)$ 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 efficient and inference is χ^2.

Case 3: $k_2=1$, $\theta_{2,1}^1$ known to be zero. The estimation equation is (5.8) with $\theta_{2,1}^1=0$. Leads and lags of $\Delta^2 y_t^1$ and Δy_t^2 are I(0) with mean zero, and these comprise z_t^1. Also set $z_t^3=(\Delta y_t^1,\ y_t^2)'$ (y_t^2 is I(1)) and $z_t^5=y_t^1$. Thus $(\hat\theta_{3,1}^0,\hat\theta_{3,2}^0,\hat\theta_{3,1}^1)$ converge at (T^2,T,T) and inference is χ^2.

Case 4: $k_2=1$, $\theta_{2,1}^1$ unknown. Although (Y^1,Y^2) are not weakly exogenous for $(\theta_{3,1}^0,\theta_{3,2}^0,\theta_{3,1}^1)$ in this case, the dynamic OLS and GLS estimators nevertheless have desirable properties. With $\theta_{2,1}^1$ unknown, the estimation equation (5.8) becomes,

$$(5.9)\quad y_t^3 = (\theta_{3,1}^1-d_{3,2}(1)\theta_{2,1}^1)\Delta y_t^1 + d_{3,2}(1)\Delta y_t^2 + \theta_{3,1}^0 y_t^1 + \theta_{3,2}^0 y_t^2 + (d_{3,1}(L)-d_{3,2}^*(L)\theta_{2,1}^1)\Delta^2 y_t^1 + d_{3,2}^*(L)\Delta^2 y_t^2 + \tilde u_t^3$$

where $d_{3,2}^*(L)=(1-L)^{-1}(d_{3,2}(L)-d_{3,2}(1))$. 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 $z_t^1=(\Delta y_t^2-\theta_{2,1}^1\Delta y_t^1)'$, $z_t^3=(\Delta y_t^1,\ y_t^2-\theta_{2,1}^1 y_t^1)'$, and

$z_t^5 = y_t^1$. Evidently $\theta_{3,1}^1$ is not identified from (5.9) alone; using a consistent estimator of $\theta_{2,1}^1$ from (5.7b) would 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 (5.9) 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 $\theta_{3,1}^1$. The exception to this is the special case of $\theta_{3,1}^1$ known to be zero, in which case Δy_t^1 would not enter as a regressor in (5.8) were $\theta_{2,1}^1$ known. Even here, inference on $(\hat{\theta}_{3,1}^0, \hat{\theta}_{3,2}^0)$ 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:

$$\Delta y_t^1 = u_t^1 \tag{6.1a}$$

$$y_t^2 = \theta y_t^1 + u_t^2 \tag{6.1b}$$

with $\Phi(L)u_t = \zeta_t$, $\Phi(L) = I_2 - \Phi L$, ζ_t NIID$(0, \Sigma_\zeta)$, where $u_t = (u_t^1\ u_t^2)'$. 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.

The six estimators considered are the static OLS estimator (SOLS; Engle and Granger [1987], Stock [1987]), the dynamic OLS estimator $\hat{\delta}_{OLS}$ (DOLS), the dynamic

GLS estimator δ_{GLS} (DGLS), the zero frequency band spectrum estimator of Phillips (1988b) (PBSR), the fully modified estimator of Phillips and Hansen (1989) (PHFM), and Johansen's (1988a) VECM maximum likelihood estimator (JOH). t-statistics for JOH were calculated as described in Section 5(A). Two serial correlation-robust estimators of the covariance matrix of the DOLS estimator were considered, one based on a weighted sum of the autocovariances of the errors (DOLS1), the second using an autoregressive spectral estimator (DOLS2). A constant was included in all estimation procedures. The details of the construction of the estimators are given in the notes to Table 1.

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, Δy_t^1 is strictly exogenous in (6.1b) and SOLS is the MLE. 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 and $\Phi_{21}\neq 0$ 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 (the difference again being the unnecessary estimation of some nuisance parameters). In this case, PBSR and PHFM are efficient if interpreted semiparametrically. Third (case C), for general Φ and Σ_{ζ}, JOH with one lag is the exact 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 1 for T=100 and 300. Panel A verifies that the estimation of the nuisance parameters *per se* in the efficient estimators does not substantially reduce 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, .196. The DOLS, DGLS, and JOH estimators eliminate this bias, although when the regressor exhibits strong serial correlation, the DOLS t-statistics tend to have heavier tails than predicted by the asymptotic distribution theory. The PBSR and PHFM estimators tend to have biases comparable to SOLS. When this bias is small (for example when $\phi_{11}=\phi_{21}=0$), their t-statistics have approximately normal distributions.

The final case (Φ, Σ_ζ unrestricted) introduces two additional parameters, and it is beyond the scope of this investigation to explore this case in detail. Rather, case C is examined by generating data from a model relevant to the empirical analysis in Section 7, namely 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.[8] The estimated VAR(1) for the triangular system (v_t, $v_t+.10r_t$) is reported in panel C of Table 1. 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 since it is the exact MLE in this system, JOH is essentially unbiased. The dispersion 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 DOLS and DGLS results, it is useful to write (6.1) in the triangular form given in (2.3). Write the VAR(1) for u_t as $\Psi(L)u_t=a_t$, where $\Psi(L)=\Sigma_\zeta^{-\frac{1}{2}}\Phi(L)$ and $a_t=\Sigma_\zeta^{-\frac{1}{2}}e_t$, so that $E(a_ta_t')=I$. Then Δy_t^1 has a univariate ARMA(2,1) representation and $c_{11}(L)$ in (2.3a) is given by $c_{11}(L)=\kappa(L)|\Psi(L)|^{-1}$, where $\kappa(L)$ is the first degree polynomial with roots outside the unit circle that solves $\kappa(L)\kappa(L^{-1})=\Psi_{22}(L)\Psi_{22}(L^{-1})+\Psi_{12}(L)\Psi_{12}(L^{-1})$. 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, $c_{22}(L)\epsilon_t^2$ in (2.3b), follows an AR(1) with $c_{22}(L)^{-1}=\kappa(L)$. 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, $\kappa(L)=1-.66L$ so the true d(L) is infinite order and the DOLS/DGLS regressions omit 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. The relatively large root of $\kappa(L)$ suggests that the bias is attributable to the truncation of d(L) in the DOLS/DGLS regressions.[9] 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.[10] 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 result in inaccurate spectral density estimates subsquently 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.

This modest Monte Carlo experiment suggests three conclusions. First, all the estimators (except the correctly-specified JOH) exhibit bias in 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 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 two procedures for computing the covariance matrix, the autoregressive estimator produces t-statistics that are more normally-distributed than does the kernel estimator. For this reason, the DOLS standard errors reported in the empirical analysis in Section 7 are based on the autoregressive covariance estimator.

7. Application to the Long-Run Demand for Money in the U.S.

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 these estimated? 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 a paucity of low frequency information in the postwar data. This section examines Lucas's interpretation using the formal 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 for comparison with other studies selected results using postwar monthly data are presented as well.[11]

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).[12] 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 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-characterized as being individually integrated, and formal tests support this view. Specifically, the following characterizations appear consistent with the observed series: m-p is I(1) for the full sample (and both halves of the sample), with drift; r is I(1) with no drift; y is I(1) with drift; and (m-p), y and r are cointegrated. Whether p and m are individually I(1) or I(2) is unclear: the inference depends on the subsample and the test specification. The evidence suggests, but is not conclusive, that $r-\Delta p$ is I(0). Because r_t is nonnegative, characterizing r_t as I(1) raises conceptual difficulty. This decision is driven by the empirical evidence that r_t exhibits considerable persistence, and is consistent with interest rate specifications used by other researchers (e.g., Campbell and Shiller [1987] and Hoffman and Rasche [1989]).[13]

The applicability of the DOLS and DGLS estimators to I(1) and I(2) systems makes it possible to estimate θ_p in the cointegrating relation, $m-\theta_p p-\theta_y y-\theta_r r$, and to test whether $\theta_p=1$. We consider three specifications. 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, modified for nonzero drifts, and inference 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 $y_t^1=p_t$, $y_t^2=(r_t\ y_t)$, and $y_t^3=m_t$, where $\theta_{2,1}^1=0$, $\theta_{3,1}^1=0$, $\theta_{3,1}^0=(\theta_y\ \theta_r)$, and $\theta_{3,1}^0=\theta_p$. This is case 3 in Section 5(B), and inference on $(\theta_y, \theta_r, \theta_p)$ using DOLS or DGLS is χ^2. Third, if p is I(2) and $r-\Delta p$ is I(0), this is a combination of cases 2 and 3 in Section 5(B), with $y_t^1=p_t$, $y_t^2=(\Delta^{-1}r_t,y_t)$, $\theta_{2,1}^1=(1,0)'$, $\theta_{3,1}^1=(\theta_r,0)'$, $\theta_{3,1}^0=\theta_p$, and $\theta_{3,2}^0=(0,\theta_y)'$. Then $\Delta y_t^2-\theta_{2,1}^1\Delta y_t^1=(r_t-\Delta p_t,\Delta y_t)'$. Here, the cointegrating vector (1,-1) is imposed on $(r_t,\Delta p_t)$ as implied by the elementary economic hypothesis that the real interest rate is stationary. Again, inference using DOLS or DGLS is χ^2.

Estimates for the four-variable system are reported in Table 2 for these three specifications.[14] The estimates of θ_p do not differ from one at the 10% (two-sided) level in any of the specifications. In all but two cases, θ_y is statistically indistinguishable from 1 at the 5% level, and in the two exceptions θ_y is estimated imprecisely. To be consistent with 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 M1 demand cointegrating vectors in the system (m-p,y,r) are presented in panel A of Table 3. 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 (1990) 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% 2-sided level. Using only the first half of the sample, the efficient estimators provide smaller income elasticities and larger interest elasticities, but this difference is modest. In sharp contrast to the first-half estimates, the postwar estimates in Table 3 differ greatly across estimators. The SOLS estimate is close to zero, the NLLS elasticities have the "wrong" sign, and the JOH estimator is highly sensitive to the number of lagged first differences used.[15]

The final set of estimates refer to the system $(m-p,y,r^*)$, where r^* is the smoothed commercial paper rate.[16] A smoothed interest rate is used for two reasons. First, the empirical money demand literature is indecisive on whether a long- or short-term interest rate is most appropriate. Because there is no consistent risk-free long-term rate with constant tax treatment over the full sample, r^* can be interpreted as a proxy for a long-term rate which, under the risk-neutral theory of the term structure, is an average of current and expected future short-rates. Second, r^* can be viewed (indeed is constructed as) an estimate of the permanent component in interest rates. The cointegrating vector relates the permanent components of m-p, y, and r; to the extent that r is a particularly noisy measure of its permanent component, the cointegrating regressions will suffer from a small-sample version of errors-in-variables bias, and using r^* could reduce this bias. The results in Table 3 are for a two-sided smoother, but they are typical of results for other smoothed rates. The full-sample estimates change only slightly using r^*. The postwar income elasticities and standard errors are larger with r^* than r. The JOH and NLLS estimates are quite sensitive to using r^*, and the differences across point estimates remain.[17]

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 explore the source of the instability in the postwar 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, using the DOLS estimator. The results, given in panel B of Table 3, are not conclusive. Although two of the four specifications reject constancy at the 10% level, only one rejection is at the 5% level and the shift parameters δ_y and δ_r are imprecisely estimated.

Second, 95% confidence regions for (θ_y, θ_r) implied by the point estimates in panel A of Table 3 generally overlap, or nearly overlap, near θ_y=1.00 and θ_r=-.11. These regions are plotted in Figure 2a-2d for, respectively, the DOLS, DGLS, PBSR, and PHAN estimators.[18] For each estimator, the only nonoverlapping region is for the postwar estimator based on r; 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.[19]

The third piece of evidence concerns the properties of the cointegrating residuals, $\hat{z}_t = m_t - \hat{\theta}_y y_t - \hat{\theta}_r r_t$. These residuals exhibit quite different properties for the different 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. As shown in Table 4, based on the full-sample point estimates, $\hat{z}_t$ appears stationary. When the cointegrating vector is estimated over the first half and $\hat{\tau}_\mu$ is computed over the second half, asymptotically $\hat{\tau}_\mu$ has the standard univariate Dickey-Fuller (1979) distribution: using $\hat{\tau}_\mu$, non-cointegration is rejected in the postwar data at the 5% one-sided level for the DOLS prewar cointegrating vector. In contrast, the residuals from the postwar money demand equations exhibit more serial correlation but lower variance than the residuals constructed using the prewar or full-sample estimates.

Fourth, the postwar VECM likelihood, concentrated to be a function of (θ_y, θ_r), is bimodal for both the r and r^* data sets. Moreover the JOH point estimates are quite sensitive to the number of nuisance parameters estimated (number of lagged first differences included). Inspection of the concentrated likelihood, plotted in Figure 3 for 3 lags (JOH(3) in Table 3A), indicates 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. This explains, in a mechanical sense, the instability of the JOH estimates with respect to the lag length, and suggests that the JOH estimator might be poorly approximated as normally distributed.

These four pieces of evidence lead us to conclude that, despite the apparently large differences in the prewar and postwar point estimates, the evidence against Lucas's (1988) interpretation of a stable long-run money demand relation is weak, and indeed that the best summary of the evidence is that long-run M1 demand has been stable over 1900-1989. Using the postwar data alone, the elasticities are imprecisely estimated. The postwar data is 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 $\Delta\overline{(m-p)} - \theta_y\Delta\bar{y} - \theta_r\Delta\bar{r}$ (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. Several such trend lines (or low frequency movements) can be drawn from the prewar sample, resulting in tighter confidence regions when only the 1900-1945 sample is used. When the 1900-1945 and 1946-1989 subsamples are combined, the 1900-1945 and 1946-1989 trend lines solve for point estimates θ_y=1.000 and θ_r=-.145. 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 much smaller than that based on the prewar and especially the postwar data.

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 5, panel A. Compared to the postwar annual results, the income elasticities estimated over 1949:1-1988:6 are higher and there is somewhat less disagreement across the efficient estimators, with income elasticities ranging from .30 to .89 based on the commercial paper rate. The estimates are stable across the choice 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 (1990) NLLS estimate of .5 obtained over 1960-1988, strikingly so since they 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.[20]

Although the point estimates are not sensitive to the start date of the regression, they 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 5, whether estimated in logarithms or in levels.

Because we do not provide uniform critical values for tests based on these "recursive" estimates, this observed instability does not provide formal evidence on the stability of the cointegrating vector estimated with the postwar data. This sensitivity to terminal regression dates is, however, consistent with our interpretation of the annual data. Specifically, the data from 1950 to 1982 are dominated by the single upward trend in real balances, income and interest rates, which results in estimated income and interest elasticities that are strongly negatively correlated and are imprecisely estimated, except that they must fall on the trend line. Only with the most recent data, which reflect the 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. This analysis is restricted in several regards. Only one monetary aggregate has been considered, M1. Much of the money demand literature has focused on the search for a stable short-run demand function, an issue avoided here altogether. The analysis has relied heavily on asymptotic distribution theory to construct formal confidence intervals and tests, and the estimation procedures typically entail the estimation of many nuisance parameters relative to the sample size. Although we have only limited evidence, this leads us to suspect that the precision of the foregoing results is overstated.

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, (.90, 1.03), and for the interest semielasticity, (-.124, -.088). Similar intervals are obtained using the other efficient estimators over the full sample.

Second, our results are consistent with Lucas's (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 the first-half estimates of the cointegrating vector.

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 sufficiently to estimate the money demand relation. We suspect that the postwar standard errors understate the sampling variability, particularly for the monthly results, because of this sensitivity to terminal dates, some evidence that the large-sample mixed normal distribution provides a poor approximation to the postwar sampling distributions, and the presence of this problem in the Monte Carlo analysis in Section 6.

8. Conclusions

The empirical investigation suggests some observations that might apply more generally beyond this particular application. First, the precise estimation of long-run money demand appears to require a long span of data: estimates over the full span are more precise than over the first half of the century alone, and the data since 1946 contain quite limited information about long-run money demand when viewed in isolation. Second, the use of several efficient estimators is a valuable check of the sensitivity of the estimates to changes that should be asymptotically negligible. In the case of postwar money demand, the sensitivity of the postwar estimates to the choice of estimator and to the estimation period drew attention to the low-frequency multicollinearity between postwar income, velocity and interest rates in the postwar data.

Appendix

Proof of Lemma 2.1.

The proof is a modification of Anderson's (1971, Theorem 7.6.7) and Rozanov's (1967, Chapter 2.3) proofs of the Wold decomposition. Note that Ψ_t^i and Ψ_∞^i, i=1,2, form Hilbert spaces. Then $u_t^1=c_{11}(L)\epsilon_t^1$ is the Wold representation of u_t^2 and by construction ϵ_t^1 is the innovation process for u_t^1. Let $D_t^1=\Psi_t^1 \ominus \Psi_{t-1}^1$, so that by construction ϵ_t^1 forms a basis for D_t^1. Then $P(u_t^2|\Psi_\infty^1)=P(u_t^2|\bigcup_{s=-\infty}^{\infty}D_s^1)=\sum_{j=-\infty}^{\infty}c_{21j}\epsilon_{t-j}^1=c_{21}(L)\epsilon_t^1$, where $c_{21j}=(Eu_{t+j}^2\epsilon_t^{1\prime})(E\epsilon_t^1\epsilon_t^{1\prime})^{-1}$. Now $\tilde{u}_t^2=u_t^2-P(u_t^2|\Psi_\infty^1)=u_t^2-c_{21}(L)\epsilon_t^1$ is stationary, has $E(\tilde{u}_t^2)^2<\infty$, and is linearly regular, and so possesses the Wold decomposition $\tilde{u}_t^2=c_{22}(L)\epsilon_t^2$, where $\epsilon_t^2=u_t^2-P(u_t^2|\Psi_\infty^1 \oplus \Psi_{t-1}^2)$. By construction, ϵ_t^2 is the innovation process for $\tilde{u}_t^2$, $E\epsilon_t=0$, $E\epsilon_t^1\epsilon_s^{2\prime}=0$ for all t,s, and $E\epsilon_t^i\epsilon_s^{i\prime}=\Sigma_{ii}$ for t=s and equals 0 otherwise. Finally, c(L) is square summable because $Eu_tu_t'<\infty$ by assumption. □

Derivation of (3.1).

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

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

can be represented as:

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

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

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

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

Accumulating (A.1) yields $x_t^2 - \theta'x_t^1 = \sum_{i=0}^{m+1}\bar{\mu}_{2,i}t^i + D_2(L)a_t$, where $D_2(L) = F_2^*(L) - \theta'F_1^*(L)$, where $F_i^*(L) = (1-L)^{-1}(F_i(L)-F_i(1))$, i=1,2. Because $F_i(L)$ is ℓ summable, $F_i^*(L)$ is $(\ell-1)$ 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,...,m) and $D_1(L)=F_1(L)$. □

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

$$\Delta^{d-1}\tilde{y}_t^1 = \tilde{\mu}_{1,0} + F_1^{d-1}(L)a_t$$
$$\Delta^{d-1}\tilde{y}_t^2 = \tilde{\mu}_{2,0} + \tilde{\mu}_{2,1}t + \theta_{2,1}^{d-1}(\Delta^{d-1}\tilde{y}_t^1) + F_2^{d-1}(L)a_t$$

where y_t has been partitioned into k_1×1 and $(n-k_1)$×1 components $\tilde{y}_t^1$ and $\tilde{y}_t^2$. Now assume that $F^{d-1}(1)=[F_1^{d-1}(1)'\ F_2^{d-1}(1)']'$ has rank $k_1+k_2\le n$, and apply the lemma to $x_t = [\Delta^{d-1}\tilde{y}^1_t, (\Delta^{d-2}\tilde{y}^2_t - \theta^{d-1}_{2,1}\Delta^{d-2}\tilde{y}^1_t)]$. Continuing this process yields the triangular representation (3.1), with $u^j_t = D_j(L)a_t$, j=1, ...,d+1, where rank $[D_j(1)] = k_j$ for j=1,...,d. While $\tilde{D} = [D_1(1)'\ D_2(1)'\ \ldots\ D_d(1)']'$ has full row rank by construction, nothing so far ensures that $D_{d+1}(1)$ is linearly independent of the rows of $\tilde{D}$. If it is, then the construction is completed. If it is not, then redefine the variables y_t to be $\Delta^{-1}y_t$ and d to be d+1, and repeat the construction until $[\tilde{D}'\ D_{d+1}(1)]'$ has full rank, so that u_t is I(0) with a full rank spectral density matrix. This yields (3.1) for variables of arbitrary finite orders of integration and cointegration.

Proof of Theorem 4.1.

First consider the infeasible GLS estimator $\bar{\delta}_{GLS}$ constructed using $\Phi(L)$ rather than $\hat{\Phi}(L)$. Note that $(T_T \otimes I)(\bar{\delta}_{GLS} - \delta) = Q_T^{-1}\phi_T$, where $Q_T = (T_T^{-1} \otimes I)\sum_t \bar{z}_t \bar{z}_t' (T_T^{-1} \otimes I)$ and $\phi_T = (T_T^{-1} \otimes I)\sum_t \bar{z}_t \bar{e}_t$, with $\bar{z}_t = [z_t \otimes \Phi(L)']$ and $\bar{e}_t = \Phi(L)e_t$. (Unless otherwise stated, the remaining identity matrices have dimension k_ℓ so this subscript is suppressed.) 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 ≥ 2:

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

where the last two lines follow from Lemma 1 of Sims, Stock and Watson (1990) (SSW) and $\Phi(1)'\Phi(1) = \Omega_{\ell\ell}^{-1}$. For ϕ_{iT}, i ≥ 2:

$$\begin{aligned} \phi_{iT} &= (T_{iT}^{-1} \otimes I)\textstyle\sum_t[\sum_{m=0}^q (z^i_{t-m} \otimes \Phi_m')]\Sigma_{\ell\ell}^{-\frac{1}{2}}\epsilon^\ell_t \\ &= (T_{iT}^{-1} \otimes I)\textstyle\sum_t(z^i_t \otimes \Phi(1)')]\Sigma_{\ell\ell}^{-\frac{1}{2}}\epsilon^\ell_t + o_p(1) \end{aligned}$$

$$\Rightarrow \begin{cases} \int_o^1 (G_{mm}(1) s^{(m-2)/2} \otimes \Phi(1)')dW_2(s), & m=2,4,6,\ldots,2\ell \\ \int_o^1 (G_{mm}(1) W_1^{(m-1)/2}(s) \otimes \Phi(1)')dW_2(s), & m=3,5,7,\ldots,2\ell-1 \end{cases}$$

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

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

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

so $(T_T^{-1}\otimes I)(\hat{\delta}_{GLS}-\bar{\delta}_{GLS}) = \hat{Q}_T(\hat{\phi}_T-\phi_T) + (\hat{Q}_T-Q_T)\phi_T$. Assume that $\hat{\Phi}_j \xrightarrow{p} \Phi_j$ for $j=1,\ldots,q$. Because $\hat{Q}_T - Q \xrightarrow{p} 0$, $\bar{Q}_T - Q \xrightarrow{p} 0$, Q is a.s. invertible, and $\hat{\phi}_T - \phi_T \xrightarrow{p} 0$, GLS and feasible GLS are asymptotically equivalent. □

<u>Proof of Theorem 4.2</u>.

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

(b) Theorems 4.1 and 4.2 imply that $T_{1T}^{-1}\sum_t z_t^1 z_t^{*\prime} T_{*T}^{-1} \xrightarrow{p} 0$. First consider the infeasible GLS estimator $\bar{\delta}_{GLS}$, defined in the proof of Theorem 4.1. Theorem 4.1 implies that

$$(T_{*T} \otimes I)(\bar{\delta}_{*GLS}-\delta_*) = B_{*T}^{-1}[(T_{*T}^{-1}\otimes I)\sum_t(z_t\otimes\Phi(L)')\Sigma_{\ell\ell}^{-\frac{1}{2}}\epsilon_t^{\ell}] + o_p(1)$$

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

$$B_{*T} = (T_{*T}^{-1}\otimes I)(I_{(g-g_1)k_\ell}\otimes\Phi(1)'\Phi(1))[\sum_t(z_t^* z_t^{*\prime}\otimes I)](T_{*T}^{-1}\otimes I)$$

so

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

Also,

$$(T_{*T}\otimes I)(\hat{\delta}_{*OLS}-\delta_*) = [(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes I)(z_t^{*\prime}\otimes I)(T_{*T}^{-1}\otimes I)]^{-1}$$

$$\times\ (T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes I)c_{\ell\ell}(L)\epsilon_t^\ell + o_p(1)$$
$$= B_{*T}^{-1}(I\otimes\Phi(1)'\Phi(1))(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes I)c_{\ell\ell}(L)\epsilon_t^\ell + o_p(1)$$
$$= B_{*T}^{-1}(T_{*T}^{-1}\otimes I)\sum_t(z_t^*\otimes\Phi(1)'\Phi(1))c_{\ell\ell}(L)\epsilon_t^\ell + o_p(1)\ .$$

Thus

$$(T_{*T}\otimes I)(\hat{\delta}_{*OLS}-\bar{\delta}_{*GLS}) = B_{*T}^{-1}(T_{*T}^{-1}\otimes I)\{\sum_t(z_t^*\otimes\Phi(1)'\Phi(1))c_{\ell\ell}(L)\epsilon_t^\ell - \sum_t(z_t^*\otimes\Phi(L)')\Sigma_{\ell\ell}^{-\frac{1}{2}}\epsilon_t^\ell\} + o_p(1)$$
$$= B_{*T}^{-1}(T_{*T}^{-1}\otimes I)\sum_t\{(z_t^*\otimes\Phi(1)')\Phi(1)[c_{\ell\ell}(1)\epsilon_t^\ell + c_{\ell\ell}^*(L)\Delta\epsilon_t^\ell] - (z_t^*\otimes\Phi(1)')\Sigma_{\ell\ell}^{-\frac{1}{2}}\epsilon_t^\ell + [z_t^*\otimes(\Phi(1)-\Phi(L))']\Sigma_{\ell\ell}^{-\frac{1}{2}}\epsilon_t^\ell\} + o_p(1)$$
$$= B_{*T}^{-1}(A_{1T} + A_{2T}) + o_p(1)$$

where the final equality follows from $\Phi(1)c_{\ell\ell}(1)=\Sigma_{\ell\ell}^{-\frac{1}{2}}$, and where $\Phi^*(L)=(1-L)^{-1}(\Phi(L)-\Phi(1))$, $c_{\ell\ell}^*(L)=(1-L)^{-1}(c_{\ell\ell}(L)-c_{\ell\ell}(1))$, and

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

Because $B_{*T} \Rightarrow Q_*$ (the (*,*) block of Q given in Theorem 4.1), the result follows if $A_{1T} \xrightarrow{p} 0$ and $A_{2T} \xrightarrow{p} 0$. Because $\Phi^*(L)$ has a finite order by assumption and $E(\epsilon_t^\ell|\{z_t^*\})=0$, standard telescoping arguments imply that $A_{1T} \xrightarrow{p} 0$. In addition, $A_{2T} \xrightarrow{p} 0$ as a consequence of the results for ϕ_{it} in Theorem 4.1. □

Proof of Theorem 4.3.

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

Proof of Theorem 4.4.

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

Footnotes

1. Since submitting this paper it has come to our attention that the estimator proposed here was independently developed by Phillips (1988a) (also see Phillips and Loretan [1989]) and Saikkonen (1989). The earliest reference of which we have become aware is Hansen (1988). The current paper extends previous results to higher, differing orders of integration, handles deterministic time trends, and applies the results to the estimation of long-run money demand.

2. Similar results hold in the Gaussian model with explosive roots, see Domowitz and Muus (1988).

3. This lemma has antecedents (but to our knowledge no previous formal statement and proof) in the literature on optimal filtering, for example Whittle (1983) Chapter 5 or Brillinger (1980) Section 8.3, or Sims' (1972) discussion of Granger causality.

4. Although the construction leading to (2.5) makes (2.5) unique, alternative triangular representations exist. For example, it is possible to construct a one-sided triangular representation analogous to (2.5), except that ϵ_t^1 will not be innovations of u_t^1 and in general $c_{11}(z)$ will not be invertible. Such a representation was derived by Hansen, Roberds and Sargent (1990) to study restrictions on the consumption and labor income process implied by the balanced budget constraint.

5. Johansen (1988b, 1990) 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 Granger and Engle's (1987) terminology, cointegration of the form CI(d,b), where $d \geq b$. As Johansen (1990) points out, this excludes cointegration of the general form (3.4), which generalizes what Granger and Lee [1988] term "multicointegration". Johansen (1990) complements our derivation, since it explicitly handles multicointegration; it relates multicointegration to restrictions on the parameters of the levels VAR, whereas the current derivation refers to the moving average representation of the d-th difference.

6. This result has recently been provided by Saikkonen (1989) in the d=1 case.

7. Equations (5.3) and (5.4) provide an easy way to construct standard errors for $\hat{\theta}_{MLE}$, namely from (5.4) or alternatively from the usual NLLS formula from the regression of Δy_t^2 onto $(y_{t-1}^1, y_{t-1}^2, \hat{\eta}_t^1 = \Delta y_t^1 - \hat{\Pi}_{MLE}\Delta y_t^2)$. This extends directly to the case of general finite-order A(L) by including lags of Δy_t in the regression. As discussed below, an asymptotically equivalent estimator of θ and its standard error can be constructed by replacing $\hat{\Pi}_{MLE}$ used in the construction of $\hat{\eta}_t^1$ by any consistent estimator of Π.

8. See Section 7 for a description of the data.

9. 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.

10. For Model 3 and T=100, the DOLS bias was reduced to .007. The main effect of doubling the number of leads and lags and order of the autoregression for Models 1 and 2 was to increase the dispersion of the t-statistic; for example, for T=100 $t_{.95}-t_{.05}$ for DOLS1 increased from 3.86 to 4.13. A similar increase in dispersion of the distribution of JOH t-statistics occurred when the number of lags in the VECM was doubled.

11. Most empirical analyses of money demand predate the literature on cointegration. Exceptions are Hoffman and Rasche (1989), who apply Johansen's (1988a) estimator to monthly U.S. M1 data from 1953 to 1987, and Johansen and Jesulius (1990), who apply Johansen's (1988b) procedure to the long-run demand for money in Denmark and Finland. Baba, Hendry, and Starr (1990) 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 (1990) present results for the U.K. as well as the U.S.

12. Data sources and construction: M1: 1947-1989: The monthly Citibase M1 series (FM1) was used for 1959-1989; the earlier M1 data was 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 the quarterly or annual observations. Data prior to 1947 are those used by Lucas; 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 3. Y: U.S. Net National Product 1947-1989, Citibase GNNP. Prior to 1947, Lucas's (1988) data (Friedman and Schwartz real net national product (1982 dollars), Table 4.8). For the monthly data 1959-1989, we used personal income (GMPY). P: Price deflator for NNP. 1947-1989, Citibase GDNNP. Prior to 1947, Lucas's (1988) data -- same source as NNP. For monthly data we used the price deflator for GMPY. Interest rates: Commercial Paper rate. 1947-1989, 6-month commercial paper, Federal Reserve Board (FYCP), prior to 1947, Lucas's (1988) data (Friedman and Schwartz (1982), Table 4.8, column 6).

13. 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, y, r, m-p, and log 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, Δr, and Δ(m-p) are each rejected at the 1% level. Similar inferences obtain when the sample is split 1900-1945, 1946-1989. Whether m and p have two unit roots is less clear: for m, two unit roots are rejected in favor of one at the 5% level for both subsamples, but not the full sample, while the reverse is true for p. For r-Δp (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 ($\hat{\tau}_\mu$ rejects at 10% in both subsamples).

The Stock-Watson (1988) $q_{\tau}^{f}(3,1)$ statistic, applied to the system (m-p,y,r) over the full sample, rejects the hypothesis of three unit roots in favor of one unit root at the 5% level over the full sample with 1-4 lags. The evidence on three vs. two unit roots is less strong: the $q_{\tau}^{f}(3,2)$ statistic (2 lags) has a p-value of .33. However, the Engle-Granger (1987) augmented Dickey Fuller test based on the residual from regressing m-p on y and r (with a constant and time trend in the regression and using the appropriate critical values for a trivariate detrended system, two lags) rejects non-cointegration at the 5% level over the full sample. The details are available from the authors on request.

14. The dates in Table 2 and henceforth refer to the dates over which regressions are run; earlier and later observations are used as initial and terminal conditions as needed.

15. Likelihood ratio tests of 2 vs. 3 lags in the VECM are, respectively, 12.08, 11.12, and 32.88 over the periods 1904-1986, 1904-1945, and 1946-1986. With asymptotic χ_9^2 distributions, these suggest specifying p=3 over 1946-1986.

16. The smoothed interest rate was constructed to be the two-sided estimate of the permanent component 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.

17. The results in Table 3 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/lags for DOLS and DGLS, using 1 or 2 rather than 3 lags for JOH. The only exception is the postwar instability of the JOH estimates, discussed in more detail below.

18. Because Wald statistics testing hypotheses about (θ_y, θ_r) using the efficient estimators have large-sample χ^2 distributions, the usual approach can be used to construct confidence regions for (θ_y, θ_r). Asymptotically the estimators in the two subsamples are independent, but for the small samples considered here, the short-run dependence in the data, the presence of initial and terminal leads and lags, and possible deviations from the large-sample mixed normal distribution will result in a lack of independence.

19. The anomalous region is the postwar unsmoothed interest rate region for DGLS. This is best understood by noting that the estimated GLS transformation for DGLS approximately differenced the data (the estimated AR(2) filter is $1-1.39L+.41L^2$), so that the DGLS point estimates are in effect determined by covariances between first differences of the data, not their levels.

20. Using the JOH estimator with monthly data on log real personal income, log real M1, and the logarithm of the 90-day Treasury bill rate, 1953-1988 (3 lags), Hoffman and Rasche [1989] estimate the income elasticity to be .78; the difference between their estimate and the corresponding value from Table 5 (.462) arises from our use of levels, not logarithms, of interest rates.

References

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

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

Baba, Y., D.F. Hendry, and R.M. Starr (1990), "The Demand for M1 in the USA, 1960-1988," manuscript, Department of Economics, University of California, San Diego (revision of University of California, San Diego Working Paper #88-8, "U.S. Money Demand, 1960-84," December 1987).

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

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

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

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

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

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

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.

Domowitz, I.H. and L. Muus (1988), "Likelihood Inference in the Nonlinear Regression Model with Explosive Linear Dynamics," in W. Barnett, E. Berndt and H. White (eds), *Dynamic Econometric Modelling*, Cambridge University Press, Cambridge.

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

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

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: Unversity 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 B.M. Friedman and F.H. Hahn (eds.), *Handbook of Monetary Economics, vol. 1*, Amsterdam: North-Holland, 299-356.

Granger, C.W.J. and T-H. Lee (1988), "Multicointegration," Discussion Paper #24, Department of Economics, University of California, San Diego.

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

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

Hansen, B.E. (1989), "Efficient Estimation of Cointegrating Vectors in the Presence of Deterministic Trends," manuscript, University of Rochester.

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

Hansen, L.P., W. Roberds, and T.J. Sargent (1987), "Time Series Implications of Present Value Budget Balance and of Martingale Models of Consumption and Taxes," manuscript, Department of Economics, University of Chicago.

Hendry, D.F. and N.R. Ericsson (1990), "Modeling the Demand for Narrow Money in the United Kingdom and the United States," International Finance Discussion Paper no. 383, Board of Governors of the Federal Reserve System, Washington, D.C.

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

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

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

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

Johansen, S. (1990), "A Representation of Vector Autoregressive Processes Integrated of Order 2," Preprint 1990 no. 3, Institute of Mathematical Statistics, University of Copenhagen.

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, no. 2, 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*, Vol. XX, pp. 993-1023.

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

Lahiri, K. and P. Schmidt (1978), "On the Estimation of Triangular Structural Systems," *Econometrica 46*, 1217-1222.

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., "The Demand for Money: The Evidence from the Time Series," *Journal of Political Economy*, 71, 219-246.

Phillips, P.C.B. (1988a), "Optimal Inference in Cointegrated Systems," Cowles Foundation Discussion Paper No. 866 (revised August 1989).

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

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

Phillips, P.C.B. and M. Loretan (1989), "Estimating Long Run Economic Equilibria," Cowles Foundation Discussion Paper no. 928, Yale University.

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

Phillips, P.C.B and P. Perron (1988) " Testing for a Unit Roots in a Time Series Regression," *Biometrika*, 75, 335-346.

Rozanov, Y. (1967), *Stationary Random Processes*. San Francisco: Holden Day.

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. (1989), "Asymptotically Efficient Estimation of Cointegrating

Regressions," manuscript, Department of Statistics, University of Helsinki.

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*, Vol. 58, No. 1..

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.

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

Whittle, P. (1983), "Prediction and Regulation by Least Squares," 2nd Edition, revised, University of Minnesota Press, Minneapolis.

Table 1. Monte Carlo Results

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

	T=100					T=300				
Estimator	Bias(θ)	$\sigma(\theta)$	$t_{.05}$	$t_{.95}$	P(W>3.84)	Bias(θ)	$\sigma(\theta)$	$t_{.05}$	$t_{.95}$	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	2.81	.065
JOH	.000	.025	-1.98	1.96	.077	.000	.007	-1.84	1.67	.057

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

		SOLS	DOLS1		DOLS2		DGLS		PBSR		PHFM		JOH	
Φ_{21}	Φ_{11}	bias(θ)	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$	$t_{.05}$	$t_{.95}$
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	-.283	-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. $\Phi = \begin{bmatrix} -.103 & -.039 \\ -.062 & .643 \end{bmatrix}, \quad \Sigma_\epsilon = \begin{bmatrix} .951 & .499 \\ .499 & 1.374 \end{bmatrix} \times 10^{-2}$

	T=100					T=300				
Estimator	Bias(θ)	$\sigma(\theta)$	$t_{.05}$	$t_{.95}$	P(W>3.84)	Bias(θ)	$\sigma(\theta)$	$t_{.05}$	$t_{.95}$	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 to Table 1:

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 $P(\hat{W}>3.84)$ 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; additional observations were used for initial and terminal conditions. All regression include a constant term.

The estimators are:

SOLS -- Static OLS regression of y^1_t on y^2_t.

DOLS1 -- Dynamic OLS regression of y^1_t on $(y^2_t, \Delta y^2_t, \Delta y^2_{t\pm1}, \ldots, \Delta y^2_{t\pm k})$, where k=2 for T=100, k=3 for T=300. The covariance matrix is estimated by averaging the first k error autocovariances using the Bartlett kernel, where k=5 for T=100, k=8 for T=300.

DOLS2 -- Dynamic OLS regression of y^1_t on $(y^2_t, \Delta y^2_t, \Delta y^2_{t\pm1}, \ldots, \Delta y^2_{t\pm k})$, where k=2 for T=100, k=3 for T=300. 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^1_t on $(y^2_t, \Delta y^2_t, \Delta y^2_{t\pm1}, \ldots, \Delta y^2_{t\pm k})$, 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 (1988b) band spectral regression, where the spectral density at frequency zero was estimated using the Bartlett kernel with 5 lead/lags for T=100 and 8 lead/lags for T=300.

PHFM -- Phillips-Hansen (1989) fully modified estimator using the Bartlett kernel with 5 lead/lags for T=100 and 8 lead/lags for T=300.

JOH -- Johansen (1988a) VECM MLE based on the estimated model $y_t = \gamma\alpha' y_{t-1} + \sum_{i=1}^{k} A_i \Delta y_{t-i} + a_t$, where k=4 for T=100 and k=6 for T=300. Standard errors were computed using the formulas given in Section 5(A).

The DOLS and DGLS standard errors were computed using a degrees-of-freedom adjustment, specifically df=number of periods in the regression - number of regressors in the DGLS or DOLS regression - number of autoregressive lags in the GLS transform (DGLS) or AR spectral estimator (DOLS). The JOH standard errors were computed as described in Section 5(A) with a degrees-of-freedom adjustment (df=number of periods in the regression - 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 such adjustments were made for PBSR or PHFM.

Table 2

Estimated Cointegrating Relations: $m_t = \alpha + \theta_p p_t + \theta_y y_t + \theta_r r_t$

Specifications:

I. For p_t I(1): $m_t = \mu+\theta_p p_t+\theta_y y_t+\theta_r r_t+d_p(L)\Delta p_t+d_y(L)\Delta y_t+d_r(L)\Delta r_t+e_t$

II. For p_t I(2) and r, Δp_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) and $r-\Delta p_t$ 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	Period	no. leads/lags	Estimates (Standard Errors) θ_p	θ_y	θ_r
I	DOLS	1903-87	2	1.143	.838	-.119
				(.185)	(.154)	(.016)
	DOLS	1904-86	3	1.205	.794	-.128
				(.177)	(.145)	(.014)
	DGLS	1903-87	2	1.000	.322	-.042
				(.213)	(.289)	(.019)
	DGLS	1904-86	3	1.219	.798	-.133
				(.152)	(.125)	(.013)
II	DOLS	1904-87	2	1.183	.820	-.119
				(.190)	(.159)	(.016)
	DOLS	1905-86	3	1.304	.732	-.132
				(.200)	(.165)	(.016)
	DGLS	1904-87	2	1.041	.932	-.100
				(.166)	(.143)	(.016)
	DGLS	1905-86	3	1.292	.763	-.134
				(.180)	(.147)	(.016)
III	DOLS	1904-87	2	1.011	.949	-.096
				(.165)	(.138)	(.015)
	DOLS	1905-86	3	1.100	.887	-.106
				(.145)	(.118)	(.013)
	DGLS	1904-87	2	.982	.355	-.024
				(.209)	(.289)	(.017)
	DGLS	1905-86	3	1.180	.842	-.115
				(.128)	(.103)	(.011)

Notes: $d_i(L)=\sum_{j=-k}^{k} d_{ij} L^j$, where k is the number of leads/lags listed in the third 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 in panel B relative to k=2 in panel B, and for k=2 in panel A relative to k=2 in Panel B, allow for necessary initial and terminal conditions (leads and lags).

Table 3
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	.929	-.083	.916	-.089	.193	-.016	.412	-.046
NLLS	.898	-.093	1.104	-.093	-.445	.084	.298	-.023
DOLS	.965	-.106	.859	-.117	.270	-.027	.413	-.047
	(.031)	(.009)	(.142)	(.028)	(.213)	(.025)	(.320)	(.042)
DGLS	.960	-.100	.972	-.100	.945	-.020	1.171	-.091
	(.037)	(.010)	(.170)	(.030)	(.308)	(.009)	(.132)	(.013)
PBSR	.960	-.101	.903	-.103	.216	-.020	.367	-.042
	(.033)	(.009)	(.103)	(.018)	(.091)	(.011)	(.147)	(.019)
PHFM	.956	-.100	.911	-.102	.205	-.018	.393	-.045
	(.032)	(.008)	(.082)	(.015)	(.054)	(.006)	(.106)	(.014)
JOH(2)	.971	-.109	.878	-.111	35.588	-5.088	-2.344	.340
	(.031)	(.009)	(.094)	(.018)	(1787.6)	(266.0)	(4.581)	(.645)
JOH(3)	.976	-.115	.940	-.121	-.473	.075	-.131	.033
	(.030)	(.009)	(.119)	(.022)	(.390)	(.055)	(.274)	(.039)

B. Tests for Breaks in the Cointegrating Vector Based on DOLS, break date = 1946

Interest rate	no. leads/ lags	Wald statistic (p-value)	Point estimates (standard errors) θ_y	θ_r	δ_y	δ_r
r	2	6.05	.969	-.111	-.452	.061
		(.05)	(.107)	(.022)	(.269)	(.029)
	3	5.03	.983	-.111	-.446	.056
		(.08)	(.103)	(.022)	(.280)	(.027)
r*	2	2.90	.862	-.141	-.197	.068
		(.24)	(.108)	(.025)	(.352)	(.048)
	3	3.82	.862	-.144	-.285	.076
		(.15)	(.099)	(.023)	(.320)	(.043)

Notes to Table 3:

Panel A: NLLS is the nonlinear least squares estimator; the other estimators are defined in the notes to Table 1 (DOLS here and in subsequent tables is DOLS2 in Table 1). JOH(k) is the JOH estimator evaluated using k lagged first differences. JOH(3) was computed over regression dates 1904-1986, 1904-1945, and 1946-1986. For the NLLS estimator, $\Delta(m-p)_t$ is 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 are 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, $(m-p)_t=\mu+\theta_y y_t+\theta_r r_t+\delta_y(y_t-y_\tau)1(t\geq\tau)+\delta_r(r_t-r_\tau)1(t\geq\tau)+d_y(L)\Delta y_t+d_r(L)\Delta r_t$, where $1(\cdot)$ is the indicator function and $d_y(L)$ and $d_r(L)$ have the number of leads and lags stated in the second column. 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 and AR(2) spectral estimator.

Table 4

Properties of Error Corrections Terms $\hat{z}_t = m_t - \theta_y y_t - \theta_r r_t$

Estimator	Estimation Period	θ_y	θ_r	1904-86 $\hat{\tau}_\mu$	1904-86 $\hat{\rho}$	1904-86 $\hat{\sigma}$	1904-45 $\hat{\tau}_\mu$	1904-45 $\hat{\rho}$	1904-45 $\hat{\sigma}$	1946-86 $\hat{\tau}_\mu$	1946-86 $\hat{\rho}$	1946-86 $\hat{\sigma}$
DOLS	1903-87	0.965	-0.106	-4.646	0.397	0.156	-3.618	0.334	0.136	-3.496	0.360	0.172
DGLS	1903-87	0.960	-0.100	-4.542	0.411	0.151	-3.685	0.301	0.133	-3.314	0.407	0.166
JOH	1903-87	0.971	-0.109	-4.673	0.396	0.159	-3.549	0.361	0.139	-3.553	0.347	0.176
DOLS	1903-45	0.859	-0.116	-3.289	0.674	0.195	-3.654	0.282	0.137	-3.211	0.456	0.192
DGLS	1903-45	0.972	-0.100	-4.531	0.409	0.151	-3.667	0.312	0.134	-3.215	0.432	0.167
JOH	1903-45	0.878	-0.111	-3.592	0.617	0.179	-3.688	0.271	0.134	-3.438	0.389	0.180
DOLS	1946-87	0.270	-0.027	-1.512	0.972	0.468	1.136	1.065	0.347	-3.134	0.463	0.050
DGLS	1946-87	0.945	-0.020	-1.047	0.951	0.240	-1.114	0.848	0.187	-1.526	0.970	0.279
JOH(2)	1946-87	35.588	-5.088	-1.541	0.958	23.401	0.222	1.009	19.997	-3.377	0.444	9.014
JOH(3)	1946-86	-0.473	0.075	-1.327	0.978	1.017	-1.453	0.964	1.017	-0.078	0.998	1.017
DOLS*	1946-87	0.413	-0.047	-1.331	0.968	0.379	0.656	1.052	0.279	-2.969	0.664	0.042
DGLS*	1946-87	1.170	-0.091	-1.740	0.897	0.207	-2.898	0.520	0.157	-1.641	0.938	0.173
JOH(2)*	1946-87	-2.344	0.340	-1.583	0.986	2.152	1.607	1.038	1.761	-2.542	0.811	0.257
JOH(3)*	1946-86	-0.131	0.033	-1.447	0.980	0.773	-1.446	0.964	0.773	0.201	1.001	0.773

Notes: The point estimates for the indicated estimator and sample period are taken from Table 3. DOLS*, DGLS*, and JOH* refer to these estimators evaluated using the smoothed interest rate r_t^*. The summary statistics $\hat{\tau}_\mu$, $\hat{\rho}$, and $\hat{\sigma}$ are respectively the Dickey-Fuller t-statistic testing $\rho=1$ with a constant and 3 lags in the autoregression, the sum of the autoregressive coefficients in the regression of $\hat{z}_t$ on a constant and 3 lags, and the standard devation of $\hat{z}_t$. The reported entries are these statistics, computed for $\hat{z}_t$ constructed using the the point estimates in the first columns for each row, with regressions run (and statistics computed) over the subsample given in the column heading.

Table 5

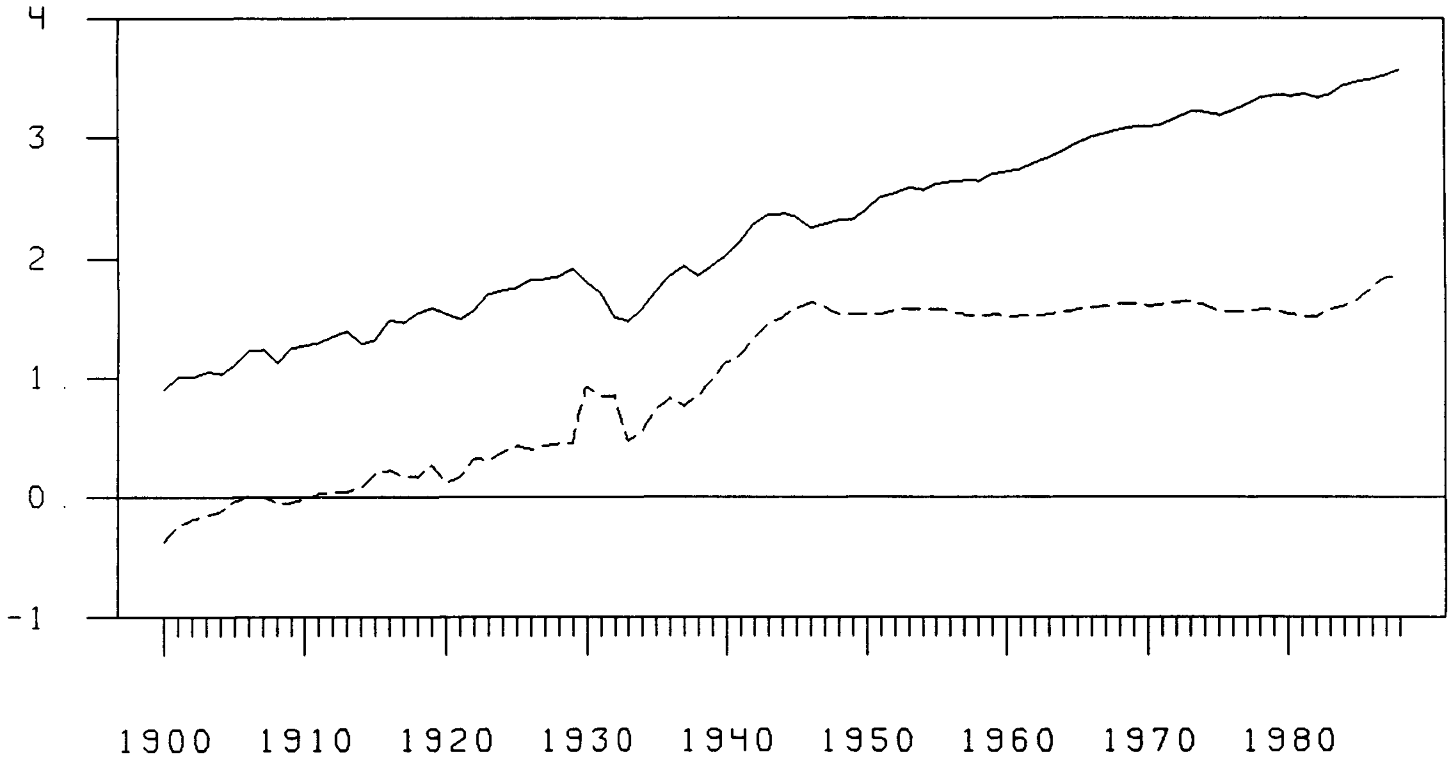
Money Demand Cointegrating Vectors: Estimates, Monthly Data

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	.889	-.008	.525	-.026	1.139	-.009	1.195	-.011	1.046	-.019
	(.203)	(.003)	(.109)	(.007)	(.289)	(.003)	(.268)	(.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 assumed 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.

Figure 1

A. U.S. real net national product (solid line) and real M1, 1900-1989



B. U.S. short-term commercial paper rate (solid line; left scale) and the logarithm of M1 velocity, 1900-1989

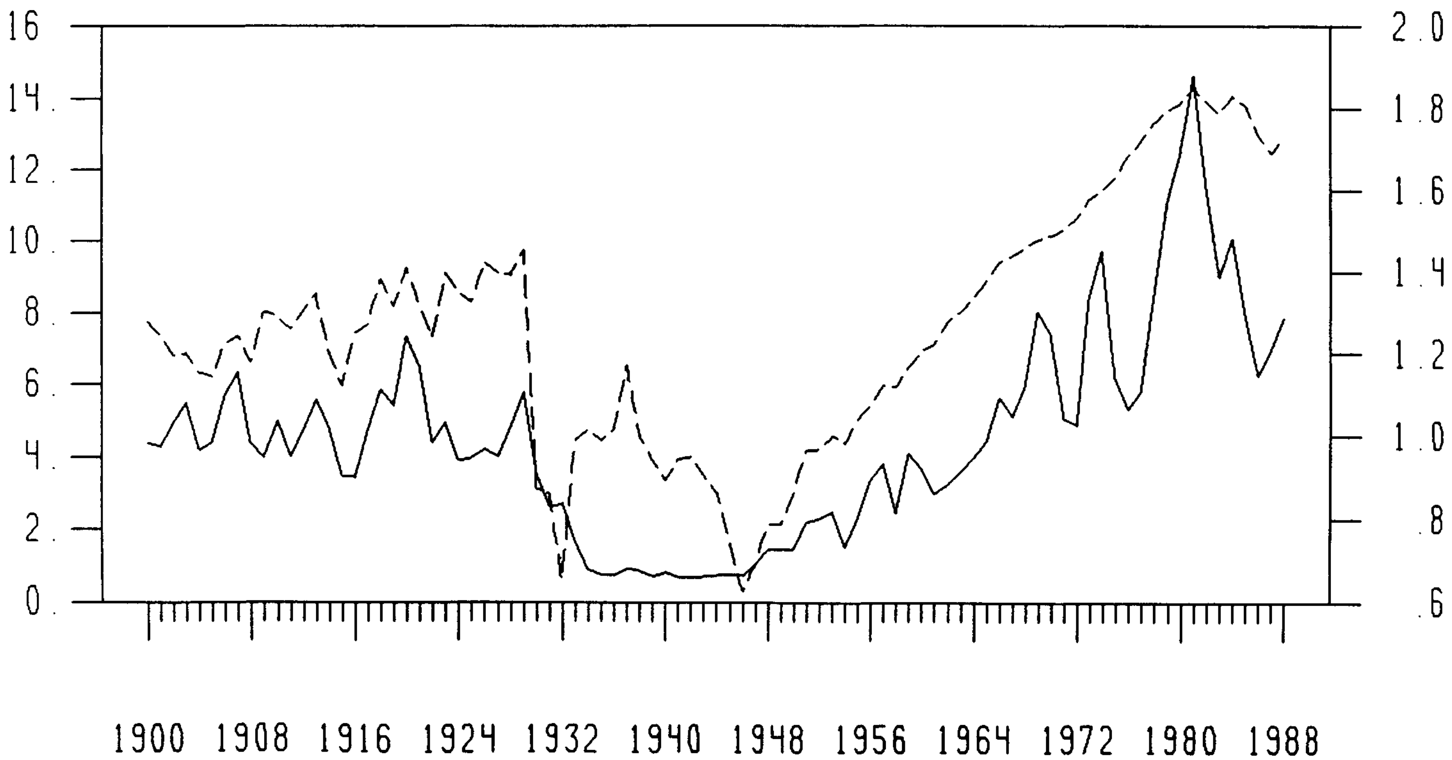


Figure 2. 95% confidence regions for the income elasticity θ_y and the interest semielasticity θ_r, estimated over 1903-1987 (solid line), 1903-1945 (dashes), 1946-1987 (short dashes), and, using the smoothed interest rate r^*, 1946-1987 (dash-dots), based on the DOLS, DGLS, PBSR, and PHFM estimators.

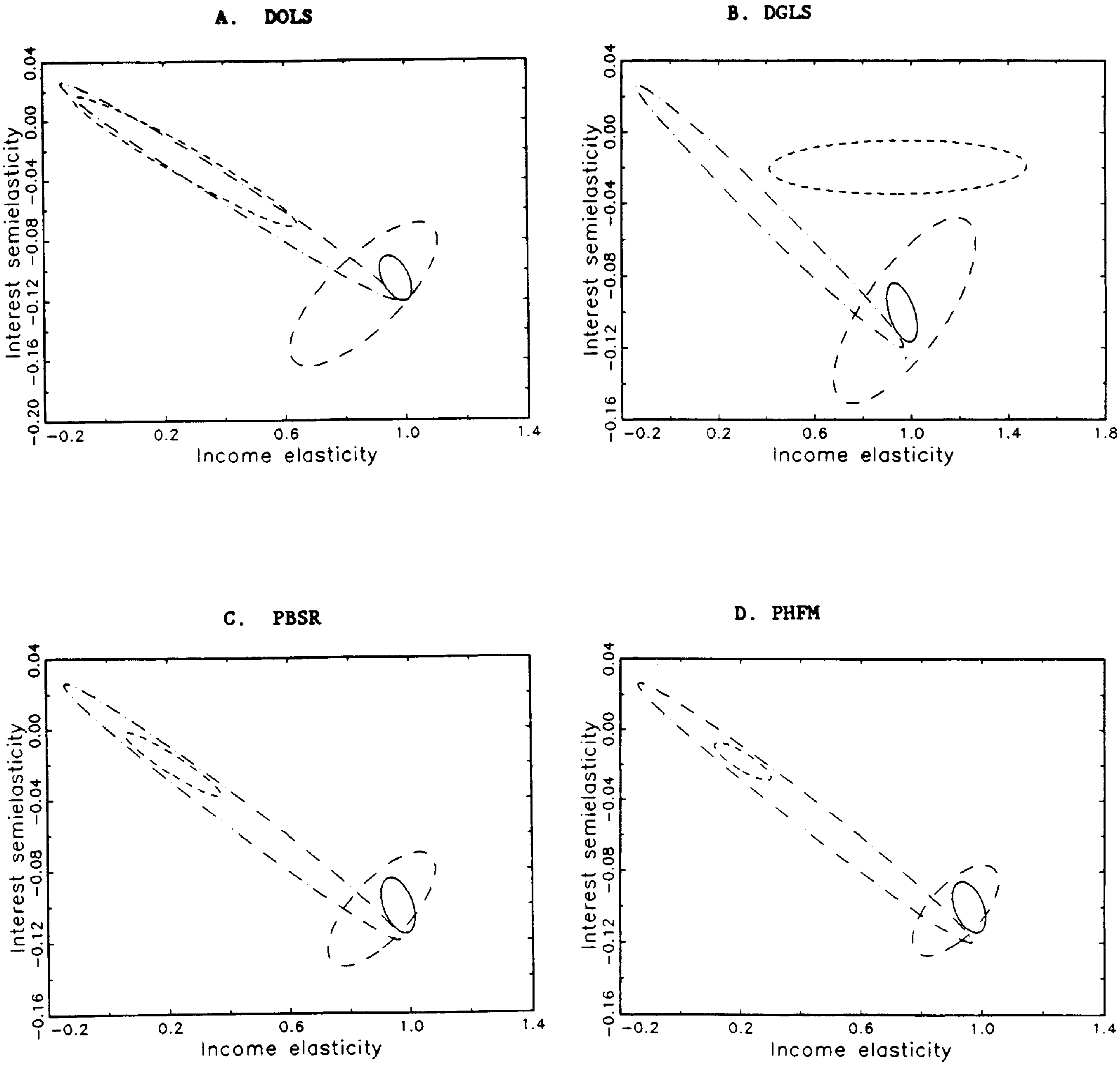


Figure 3.
Concentrated vector error correction model (VECM(3)) likelihood surface in (θ_y, θ_r) space, 1946-1986