Estimation and Inference in the Cointegrated System with Stationary Covariates

Byeongseon Seo *

Texas A&M University †

October 2004

*I would like to thank Bruce Hansen, Joon Park, and Pentti Saikkonen for helpful comments and suggestions. The author gratefully acknowledges financial support from the Korea Research Foundation.

†Department of Economics, College Station, TX 77843-4228, USA. E-mail: seo@econmail.tamu.edu. Phone: 979-845-7307. Fax: 979-847-8757.

Abstract

This paper explores the asymptotic distribution of the cointegrating vector estimator and statistical inference on cointegration in the vector error correction model (ECM) with stationary covariates, which are generated from the stationary VAR process. First, the asymptotic distribution of the cointegrating vector estimator is locally asymptotically mixed normal, and asymptotic efficiency improves as the magnitude of the covariate effect increases. Second, the asymptotic distribution of the Wald and the LR statistics for cointegration is a mixed combination of the chi-squared and the nonstandard distributions. The null distribution approaches the chi-squared distribution and the power of the cointegration tests improves significantly as the covariate effect increases. Monte Carlo simulations show that the bootstrap inference, as well as the asymptotic inference, generate moderate size and power performances. Third, the limiting distributions of the cointegrating vector estimator and the cointegration test depend on the stationary covariates, and thus the omission of the covariates leads to the departure from the existing distribution theory, which does not allow for stationary covariates.

Key words: Cointegration Test; Efficient Estimation; Stationary Covariates

JEL classification: C12; C32

1 Introduction

This paper aims to explore the asymptotic theory for estimation and inference in the vector error correction model (ECM) with stationary covariates. The stationary covariates have been used in many economic studies to consider the effect of stationary policy variables or to improve the model fitness on empirical grounds. However, the stationary covariates have been disregarded or treated less importantly in the cointegrated system, and statistical inference on the cointegrating vector and cointegration has been based on the distribution theory, which does not allow for stationary covariates. In this paper, we allow stationary covariates in the cointegrated system and develop the distribution theory for the cointegrating vector estimator and inference on cointegration.

The distribution theory of the cointegrating vector estimator has been developed by Johansen (1988, 1991), Phillips and Hansen (1990), Phillips (1991), Saikkonen (1991), and Stock and Watson (1993). Although extensive literature exists on the cointegrating vector, the literature on theoretical aspects of stationary covariates is still sparse. Saikkonen (1991) considered the stationary covariates in the cointegrated regression model and has shown that the asymptotic distribution is different from the former results. This study explores the distribution of the cointegrating vector estimator in the ECM with stationary covariates.

Statistical inference on the existence of the cointegrating relationship must be one of the crucial issues for the models with nonstationary variables. Johansen (1988, 1991) has developed the likelihood ratio (LR) test statistic for cointegration in the ECM. Because the economic models often use stationary variables in the ECM, it is necessary to develop the appropriate distribution theory. Seo (1998) has shown that the asymptotic distribution of the cointegration rank test depends on the covariate effect. However, the cointegration tests have been based on the nonstandard distribution, which does not allow for the covariate effect. In this study, we propose the Wald test statistic for cointegration and develop the methods to implement proper inference on cointegration based on the bootstrapping and the asymptotic theory.

We also consider the cointegrating vector estimator and inference on cointegration in the misspecified model, which neglects the stationary covariates. The stationary covariates have been overlooked in the model with nonstationary variables, and the effect of the omitted stationary covariates on the distribution of the cointegrating vector has not been explored. Here, we develop the associated distribution theory for estimation and inference in the misspecified model, and thereby contribute to the literature.

The stationary covariates provide information which affects the asymptotic theories of the cointegrating vector estimator and the tests for cointegration. First, the asymptotic distribution of the cointegrating vector estimator depends on the effect of stationary covariates, and asymptotic efficiency improves as the magnitude of the covariate effect increases. Second, the asymptotic distribution of the Wald and the LR statistics for cointegration is a mixed combination of the chi-squared and the nonstandard distributions. The null distribution approaches the chi-squared distribution, and the power of the cointegration tests improves significantly as the covariate effect increases. Monte Carlo simulations show that the bootstrap inference, as well as the asymptotic inference, generate moderate size and power performances. Third, the asymptotic distributions of the cointegrating vector estimator and the cointegration test depend on the stationary covariates, and thus the omission of the covariates leads to the departure from the existing distribution theory, which does not allow for stationary covariates.

Our model is related to the partial system or the structural error correction model because the partial system is specified by conditioning some endogenous variables of interest on the other remaining variables. In this respect, the literature on the structural error correction model such as Boswijk (1995), Ericsson (1995), Johansen (1992), and Harbo et al. (1998) is worth mentioning. However, the cointegrating relationship of our model does not involve the stationary covariates, and thus our model is different from the partial system.

Hansen (1995) and Elliot and Jansson (2003) considered the unit root test with the stationary covariates. The stationary covariates help to increase the power of the unit root tests. Here, we extend the univariate analysis to the cointegration tests. The power of the cointegration tests has been evaluated by Kremers et al. (1992), where the power of the cointegration tests depends on the error correction process. In this paper, we show that the null distribution of the cointegration tests depends on the stationary covariates, and the covariate effect improves the power of the cointegration tests.

We use the following notation: The operation $\otimes$ signifies the Kronecker product, and $\det(A)$ and $\mathrm{tr}(A)$ denote the determinant and the trace of matrix A, respectively. We denote $\mathrm{vec}(A)$ as the column-stacking operator and $|A|$ as the Euclidean norm of the matrix A. Also, we denote $\rightarrow^p$ as convergence in probability and $\Rightarrow$ as weak convergence of probability measures. $W(r) = BM(\Omega)$ represents a multivariate Brownian motion with long-run variance Ω. $\int W$ is an abbreviated form of $\int_0^1 W(r)dr$, W is $W(r)$, $[\cdot]$ is the integer part operator, and L is the lag operator.

The next section deals with the model and assumptions. Section 3 provides the asymptotic theory for the cointegrating vector estimator. Statistical inference on cointegration with stationary covariates is analyzed in Section 4. The omission of the stationary covariates and its effect on the cointegrating vector estimator and cointegration tests are investigated in Section 5. The models with deterministic trends are discussed in Section 6. Section 7 provides simulation evidence on the cointegrating vector estimator and the cointegration tests. An economic application is provided in Section 8.

2 The Model

Consider the p-dimensional nonstationary time series x_t, $t = 1, \ldots, n$, which is generated by an error correction model (ECM). Suppose we also have the k-dimensional stationary variable z_t generated by a stationary vector autoregressive (VAR) model as follows:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{l} \Phi_{11i} \Delta x_{t-i} + \nu_t \tag{1}$$

$$z_t = \sum_{i=1}^{l} \Phi_{21i} \Delta x_{t-i} + \sum_{i=1}^{m} \Phi_{22i} z_{t-i} + e_t \tag{2}$$

where

$$\begin{pmatrix} \nu_t \\ e_t \end{pmatrix} \sim i.i.d.(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_{\nu\nu} & \Sigma_{\nu e} \\ \Sigma_{e\nu} & \Sigma_{ee} \end{pmatrix}).$$

The cointegrating vector β and the adjustment vector α can be estimated without loss of efficiency, when there are no restrictions on the parameters across equations, from the following error correction model:

$$\Delta x_t = \Pi x_{t-1} + \sum_{i=1}^{l} \Gamma_i \Delta x_{t-i} + \sum_{i=0}^{m} \Psi_i z_{t-i} + u_t, \tag{3}$$

where $\Pi = \alpha\beta'$, $u_t = \nu_t - \Psi_0 e_t$, $\Psi_0 = \Sigma_{\nu e}\Omega_{ee}^{-1}$, $\Gamma_i = \Phi_{11i} - \Psi_0\Phi_{21i}$, and $\Psi_j = -\Psi_0\Phi_{22j}$ for $i = 1, 2, \ldots, l$ and $j = 1, 2, \ldots, m$.

We assume that the error u_t is a vector Martingale difference sequence (MDS) with $\Sigma = E(u_t u_t') < \infty$. Note that the ECM error u_t is uncorrelated with the VAR error of the stationary covariates.

Our model is the standard ECM with the stationary covariates. In empirical studies, the estimated errors often do not satisfy the regularity conditions, and they are correlated with stationary economic variables. Also, the stationary policy variables may be considered in macroeconomic models. Thus, the stationary covariates have been used in many studies such as Johansen and Juselius (1992), Baba et al. (1992), and Juselius (1995).

Our model is related to the partial system or the conditional error correction model because the partial system can be specified by conditioning some endogenous variables of interest on the other remaining variables. However, the cointegrating relationship of our model does not involve the stationary covariates, and thus our model is different from the partial system. Our model assumes that the covariates are weakly exogenous to the cointegrating relationship, as in the partial systems. The efficiency of the estimators in the partial system depends on weak exogeneity of the conditioning variables, and this condition can be verified by using the standard testing methods, as discussed by Johansen (1992).

The ECM (3) and the VAR model (2) can be written as follows:

$$\begin{aligned} \Delta x_t &= \Pi x_{t-1} + \Gamma(L)\Delta x_t + \Psi(L) z_t + u_t \\ z_t &= \Phi_1(L)\Delta x_t + \Phi_2(L) z_t + e_t, \end{aligned}$$

where $\Gamma(L) = \sum_{i=1}^{l} \Gamma_i L^i$, $\Psi(L) = \sum_{i=0}^{m} \Psi_i L^i$, $\Phi_1(L) = \sum_{i=1}^{l} \Phi_{1i} L^i$, and $\Phi_2(L) = \sum_{i=1}^{m} \Phi_{2i} L^i$.

If we define $\Pi(L) = (1-L)I - \Pi L - \Gamma(L)(1-L) - \Psi(L)(I - \Phi_2(L))^{-1}\Phi_1(L)(1-L)$, the ECM (3) can be written as

$$\Pi(L)x_t = v_t, \tag{4}$$

where $v_t = u_t + \Psi(L)(I - \Phi_2(L))^{-1}e_t$.

Assumption 1 *(a) All roots of* $\det(\Pi(L)) = 0$ *lie outside or on the unit circle.*

(b) All roots of $\det(I - \Phi_2(L)) = 0$ *lie outside the unit circle.*

To show the main results, we use the representation theorem by Engle and Granger (1987).

Theorem 1 (The Granger Representation Theorem)
Suppose Assumption 1 holds and $\Pi = \alpha\beta'$, where α and β are $p \times r$ full column rank matrices. If $\alpha_\perp$ and $\beta_\perp$ are $p \times (p-r)$ full column rank matrices such that $\alpha_\perp'\alpha = 0$ and $\beta_\perp'\beta = 0$, then the error correction model (3) can be represented by
(a)

$$\Delta x_t = C(L)v_t,$$

with $C(1) = \beta_\perp(\alpha_\perp'\Pi^*(1)\beta_\perp)^{-1}\alpha_\perp'$, where $\Pi^*(L) = \frac{\Pi(L)-\Pi(1)}{1-L}$,
(b)

$$x_t = C(1)\sum_{i=1}^{t} v_i + C^*(L)v_t,$$

where $C^*(L) = \frac{C(L)-C(1)}{1-L}$, and
(c)

$$w_t = \beta' x_t = \beta' C^*(L)v_t.$$

The proof comes directly from Engle and Granger (1987) and Johansen (1991), although our model allows for stationary covariates. The data generating process x_t has stochastic trends and a stationary component. If $\Pi = \alpha\beta'$, the null space of $C(1)$ is spanned by the cointegration space. Hence, $\beta' C(1) = 0$ and $C(1)\alpha = 0$. The stochastic trends in x_t are eliminated if we multiply the cointegrating vector. Thus, the cointegrating relationship $\beta' x_t$ is stationary.

Also, from the representation theorem, the stationary covariate z_t has the following representation:

$$z_t = D_1(L)u_t + D_2(L)e_t,$$

where $D_1(L) = (I - \Phi_2(L))^{-1}\Phi_1(L)C(L)$ and $D_2(L) = (I - \Phi_2(L))^{-1}[I + \Phi_1(L)C(L)\Psi(L)(I - \Phi_2(L))^{-1}]$.

We define $\mathcal{F}_{t-1}$ as the σ-field generated by $\{x_{t-i}, z_t, z_{t-i}, i = 1, 2, \cdots\}$. We denote $\mathcal{G}_{t-1}$ as the sub-σ-field generated by $\{x_{t-i}, z_{t-i}, i = 1, 2, \cdots\}$, which satisfies $\mathcal{G}_{t-1} \subset \mathcal{F}_{t-1}$. We assume the following conditions:

Assumption 2 *(a)* $E(u_t|\mathcal{F}_{t-1}) = 0$ *and* $E(e_t|\mathcal{G}_{t-1}) = 0$.

(b) $\sup_t E|\eta_t|^q < \infty$ *for some* $q > 2$, *where* $\eta_t = (u_t', e_t')'$.

(c) $\sum_{k=1}^{\infty} k|B_k| < \infty$, *where* $v_t = u_t + B(L)e_t = u_t + \sum_{k=0}^{\infty} B_k L^k e_t$ *and* $B(L) = \Psi(L)(I - \Phi_2(L))^{-1}$.

(d) $\sum_{k=1}^{\infty} k^2|C_k| < \infty$, *where* $\Delta x_t = C(L)v_t = \sum_{k=0}^{\infty} C_k v_{t-k}$.

(e) $\sum_{k=1}^{\infty} k|D_{1k}| < \infty$ *and* $\sum_{k=1}^{\infty} k|D_{2k}| < \infty$, *where* $D_j(L) = \sum_{k=0}^{\infty} D_{jk}L^k$ *for* $j = 1, 2$.

Assumption 2-(a) implies that the error process $\eta_t = (u_t', e_t')'$ allows for conditional heteroskedasticity. Assumptions 2-(b) and (d) imply that the process $\{\Delta x_t, w_t\}$ is uniformly 2^+ bounded. In the same way, the process $\{z_t\}$ is uniformly 2^+ bounded under Assumptions 2-(b) and (e).

Lemma 1 *Under Assumptions 1-2,*

$$\begin{pmatrix} n^{-1/2}\sum_{t=1}^{[nr]} u_t \\ n^{-1/2}\sum_{t=1}^{[nr]} v_t \end{pmatrix} \Rightarrow \begin{pmatrix} U(r) \\ V(r) \end{pmatrix} = BM\left(\begin{pmatrix} \Sigma & \Sigma \\ \Sigma & \Omega_{vv} \end{pmatrix}\right),$$

where $\Omega_{vv} = \Sigma + \Psi(1)(I - \Phi_2(1))^{-1}\Sigma_{ee}(I - \Phi_2(1))^{-1'}\Psi'(1)$.

3 Cointegrating Vector Estimator

We assume that the matrix Π in (3) is of rank r $(0 < r < p)$. Thus, there exist $p \times r$ full-column rank matrices α and β which satisfy the following:

$$\Pi = -\Pi(1) = \alpha\beta'.$$

The cointegrating relationship $\beta' x_t$ is stationary as defined in Engle and Granger (1987). Our model allows for stationary covariates z_t, and it is assumed that the stationary covariates are weakly exogenous, as defined by Engle et al. (1983), to the cointegrating relationship. As discussed in Johansen (1992) and Boswijk (1995), the cointegrating vector β can be estimated efficiently in the model (3) under weak exogeneity.

We use the following normalization of the cointegrating vector:

$$w_t(\boldsymbol{\beta}) = x_{1t} + \boldsymbol{\beta}' x_{2t}, \tag{5}$$

where x_{1t} is r-dimensional, x_{2t} is $(p-r)$-dimensional, and $\boldsymbol{\beta}$ is a $(p-r) \times r$ matrix.

The cointegrating vector can be identified from the normalization. Our representation of the cointegrating relationship has been used in many studies such as Phillips (1991).

Let $s_t = (s_{1t}', s_{2t}')'$, where $s_{1t} = (\Delta x_{t-1}', \Delta x_{t-2}', \ldots, \Delta x_{t-l}')'$ and $s_{2t} = (z_t', z_{t-1}', \ldots, z_{t-m}')'$. The ECM (3) can be written as follows:

$$\Delta x_t = \alpha w_{t-1}(\boldsymbol{\beta}) + \Gamma s_{1t} + \Psi s_{2t} + u_t,$$

where $\Gamma = (\Gamma_1, \Gamma_2, \ldots, \Gamma_l)$ and $\Psi = (\Psi_0, \Psi_1, \ldots, \Psi_m)$.

We define the parameter vector $\theta = (\text{vec}(\boldsymbol{\beta})', \theta_2')' \in \Theta$, where $\theta_2 = \text{vec}(\alpha, \Gamma, \Psi, \Sigma)$. We denote θ_0 as the true parameter value.

Under the auxiliary condition $u_t \sim N(0, \Sigma)$, the likelihood function can be defined as follows:

$$\mathcal{L}_n(\theta) = -\frac{n}{2} \log |\Sigma| - \frac{1}{2} \sum_{t=1}^{n} u_t'(\theta) \Sigma^{-1} u_t(\theta), \tag{6}$$

where $u_t(\theta)$ is defined as u_t in (3).

We denote $\hat{\theta}$ as the MLE of θ. That is,

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} \mathcal{L}_n(\theta).$$

The maximum likelihood estimator $\hat{\theta}$ maximizes the likelihood function, and the first order condition is given by

$$g_n(\hat{\theta}) = \frac{\partial \mathcal{L}_n(\hat{\theta})}{\partial \theta} = 0,$$

where

$$\begin{aligned}
\frac{\partial \mathcal{L}_n(\theta)}{\partial \boldsymbol{\beta}} &= \sum_{t=1}^{n} x_{2t-1} u_t^{'} \Sigma^{-1} \alpha, \\
\frac{\partial \mathcal{L}_n(\theta)}{\partial \alpha^{'}} &= \sum_{t=1}^{n} w_{t-1} u_t^{'} \Sigma^{-1}, \\
\frac{\partial \mathcal{L}_n(\theta)}{\partial \Gamma^{'}} &= \sum_{t=1}^{n} s_{1t} u_t^{'} \Sigma^{-1}, \\
\frac{\partial \mathcal{L}_n(\theta)}{\partial \Psi^{'}} &= \sum_{t=1}^{n} s_{2t} u_t^{'} \Sigma^{-1}, \text{ and} \\
\frac{\partial \mathcal{L}_n(\theta)}{\partial \Sigma^{-1}} &= \frac{n}{2}\Sigma - \frac{1}{2}\sum_{t=1}^{n} u_t u_t^{'}.
\end{aligned}$$

The MLE of $\boldsymbol{\beta}$ is denoted as $\hat{\boldsymbol{\beta}}$, which can be calculated by reduced rank regression (Ahn and Reinsel, 1988) or canonical analysis (Box and Tiao, 1977). Other slope parameters can be estimated by least squares once the cointegrating vector is estimated.

Lemma 2 *Under Assumptions 1-2 and* $\Pi = \alpha\beta^{'}$,

$$\frac{1}{\sqrt{n}} x_{[nr]} \Rightarrow C(1)V(r), \tag{7}$$

where $C(1) = \beta_{\perp}(\alpha_{\perp}^{'}\Pi^{*}(1)\beta_{\perp})^{-1}\alpha_{\perp}^{'}$.

We define $(p-r)\times(p-r)$ full-rank matrix $\beta_{2\perp}$, which is the partitioned matrix of $\beta_{\perp}$. We denote $C_2 = \beta_{2\perp}(\alpha_{\perp}^{'}\Pi^{*}(1)\beta_{\perp})^{-1}\alpha_{\perp}^{'}$, which corresponds to x_{2t}, so that $\frac{1}{\sqrt{n}} x_{2[nr]} \Rightarrow C_2 V(r)$.

We define the standard Brownian motions $B_1(r)$ and $B_2(r)$ as follows:

$$\begin{pmatrix} (\alpha^{'}\Sigma^{-1}\alpha)^{-1/2}\alpha^{'}\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} u_t \\ (\alpha_{\perp}^{'}\Omega_{vv}\alpha_{\perp})^{-1/2}\alpha_{\perp}^{'}\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} v_t \end{pmatrix} \Rightarrow \begin{pmatrix} B_1(r) \\ B_2(r) \end{pmatrix} = BM(\begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}). \tag{8}$$

We note that $B_1(r)$ and $B_2(r)$ are r and $(p-r)$-dimensional, respectively, and these two Brownian motions are mutually independent.

Theorem 2 *Under Assumptions 1-2 and* $\Pi = \alpha\beta'$*,*

$$n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \;\Rightarrow\; A\,(\int B_2 B_2')^{-1} \int B_2 dB_1'\,(\alpha'\Sigma^{-1}\alpha)^{-1/2'} \tag{9}$$

where $A = [(\alpha_\perp' \Omega_{vv} \alpha_\perp)^{1/2'} (\alpha_\perp' \Pi^*(1) \beta_\perp)^{-1'} \beta_{2\perp}']^{-1}$.

The asymptotic distribution of the cointegrating vector estimator $n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ depends on two independent Brownian motions. Thus, the cointegrating vector estimator follows the locally asymptotically mixed normal (LAMN) distribution:

$$n(\mathrm{vec}(\hat{\boldsymbol{\beta}}) - \mathrm{vec}(\boldsymbol{\beta})) \sim N(0, (\alpha'\Sigma^{-1}\alpha)^{-1} \otimes A(\int B_2 B_2')^{-1} A').$$

It should be noted that the variance of the cointegrating vector estimator depends on the covariate effect. Because $\Omega_{vv} = \Sigma + \Psi(1)(I - \Phi_2(1))^{-1}\Sigma_{ee}(I - \Phi_2(1))^{-1'}\Psi'(1)$, the covariate effect decreases A, which reduces the variance of the cointegrating vector estimator. This result corresponds to the asymptotic efficiency of the cointegrating vector estimator in the nonstationary regression model with stationary covariates, which is shown by Saikkonen (1991).

If there is no covariate effect, the distribution of the cointegrating vector is the same as that found in Johansen (1991). Therefore, the stationary covariates provide information, which causes the asymptotic efficiency of the cointegrating vector estimator. Asymptotic efficiency improves as the magnitude of the covariate effect increases.

In addition, the dependence of the stationary covariates amplifies the covariate effect. Thus, the variance of the cointegrating vector decreases as the matrix $I - \Phi_2(1)$ approaches zero. Also, the variance depends on α and $\alpha_\perp$. As α approaches 0, the cointegrating relationship gets weaker, and the variance of the estimator increases to infinity. On the other hand, the cointegrating relationship becomes evident and its variance decreases as $\alpha_\perp$ approaches 0.

4 Inference on Cointegration

The tests for cointegration can be based on the following null and alternative hypotheses:

$$\mathcal{H}_0 : \Pi = 0 \text{ against } \mathcal{H}_1 : \Pi \neq 0,$$

where Π is in model (3).

The likelihood ratio (LR) statistic for cointegration has been developed by Johansen (1988, 1991) in the error correction model. In this paper, we propose the Wald statistic for cointegration, which is easy to calculate and asymptotically equivalent to the LR statistic. In particular, the Wald statistic for cointegration can be robust to the heteroskedastic errors. The LR statistic for cointegration tends to over-reject the null hypothesis of no cointegration when the errors follow the GARCH process as in the Monte Carlo study by Lee and Tse (1996). Because most economic variables contain heteroskedasticity, the Wald statistic based on the robust covariance estimator can improve the finite sample properties of the cointegration test.

We define the sample moments $S_{00} = n^{-1}\sum_{t=1}^{n} R_{0t}R_{0t}^{'}$, $S_{01} = n^{-1}\sum_{t=1}^{n} R_{0t}R_{1t}^{'}$, and $S_{11} = n^{-1}\sum_{t=1}^{n} R_{1t}R_{1t}^{'}$, where

$$\begin{aligned} R_{0t} &= \Delta x_t - \sum_{t=1}^{n} \Delta x_t s_t^{'} (\sum_{t=1}^{n} s_t s_t^{'})^{-1} s_t, \\ R_{1t} &= x_{t-1} - \sum_{t=1}^{n} x_{t-1} s_t^{'} (\sum_{t=1}^{n} s_t s_t^{'})^{-1} s_t. \end{aligned}$$

The least squares estimator of Π is given by

$$\hat{\Pi}^{'} = S_{11}^{-1} S_{10}.$$

The heteroskedasticity-consistent covariance estimator of $n\text{vec}(\hat{\Pi}^{'})$ is given by

$$H_n = (I \otimes S_{11}^{-1}) \sum_{t=1}^{n} (\hat{u}_t \hat{u}_t^{'} \otimes R_{1t} R_{1t}^{'})(I \otimes S_{11}^{-1}),$$

where $\hat{u}_t$ is the least squares residuals. That is, $\hat{u}_t = \Delta x_t - \sum_{t=1}^{n} \Delta x_t S_t^{'} (\sum_{t=1}^{n} S_t S_t^{'})^{-1} S_t$, where $S_t = (x_{t-1}^{'}, s_t^{'})^{'}$.

The Wald statistic for the null hypothesis of no cointegration is given as follows:

$$\mathrm{W}_n^H = \mathrm{vec}(n\hat{\Pi}')' H_n^{-1} \mathrm{vec}(n\hat{\Pi}'). \tag{10}$$

The standard covariance estimator, which assumes homoskedastic errors, is given by

$$V_n = \hat{\Sigma} \otimes n S_{11}^{-1},$$

where $\hat{\Sigma} = n^{-1} \sum_{t=1}^{n} \hat{u}_t \hat{u}_t'$.

If the standard variance estimator is used, the Wald statistic reduces to the following:

$$\begin{aligned} \mathrm{W}_n &= \mathrm{vec}(n\hat{\Pi}')'(\hat{\Sigma} \otimes n S_{11}^{-1})^{-1} \mathrm{vec}(n\hat{\Pi}') \\ &= \mathrm{tr}\ n\hat{\Sigma}^{-1} S_{01} S_{11}^{-1} S_{10}. \end{aligned}$$

The likelihood ratio (LR) statistic for the null hypothesis of no cointegration can be defined as

$$\mathrm{LR}_n = \sum_{i=1}^{p} n\hat{\lambda}_i,$$

where $\hat{\lambda}_i$ is the eigenvalue satisfying $\hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_p$ and for $i = 1, 2, \ldots, p$

$$\mid \hat{\lambda}_i S_{11} - S_{10} S_{00}^{-1} S_{01} \mid = 0. \tag{11}$$

Thus, the LR test statistic for cointegration is given by

$$\mathrm{LR}_n = \mathrm{tr}\ n S_{00}^{-1} S_{01} S_{11}^{-1} S_{10}. \tag{12}$$

Note that S_{00} converges to Σ in probability under the null of no cointegration, and so $\hat{\Sigma}$. Thus, the Wald statistic, which uses the standard covariance estimator, is asymptotically equivalent to the LR statistic.

Under the null hypothesis of no cointegration, the null space of Π amounts to the size of x_t. Thus, $C(1)(= \Pi^*(1)^{-1})$ is a $p \times p$ matrix. As in Lemma 2, we have the following result under the null hypothesis of no cointegration:

$$\frac{1}{\sqrt{n}} R_{1[nr]} \Rightarrow CV(r),$$

where $C = \Pi^*(1)^{-1}$.

Lemma 3 *Under Assumptions 1-2 and the null hypothesis* $\mathcal{H}_0 : \Pi = 0$*, the sample moments have the following asymptotic properties:*

$$\begin{aligned} \hat{\Sigma} &\to^p \Sigma, \\ S_{00} &\to^p \Sigma, \\ S_{10} &\Rightarrow C \int V dU', \quad \textit{and} \\ n^{-1} S_{11} &\Rightarrow C \int V V' C'. \end{aligned}$$

We define the canonical correlation ρ_i as follows:

$$\rho_i = P_i' \Sigma^{1/2'} \Omega_{vv}^{-1/2'} Q_i$$

where $P_i' P_i = 1$, $P_i' P_j = 0$, $Q_i' Q_i = 1$, and $Q_i' Q_j = 0$ for $i \neq j$ and $i, j = 1, 2, \cdots, p$.

The canonical correlation matrix R can be defined as follows:

$$R = \mathrm{diag}(\rho_1, \rho_2, \cdots, \rho_p). \tag{13}$$

The orthonormal matrices P, Q and the canonical correlation matrix R satisfy

$$\begin{aligned} \Sigma^{1/2'} \Omega_{vv}^{-1} \Sigma^{1/2} P &= P R^2, \text{ and} \\ \Omega_{vv}^{-1/2} \Sigma \Omega_{vv}^{-1/2'} Q &= Q R^2. \end{aligned}$$

If we define $W_1(r) = P' \Sigma^{-1/2} U(r)$ and $W_2(s) = Q' \Omega_{vv}^{-1/2} V(r)$, then

$$\begin{pmatrix} W_1(r) \\ W_2(r) \end{pmatrix} = BM \begin{pmatrix} I & R \\ R & I \end{pmatrix}.$$

The Brownian motion W_1 can be decomposed as $W_1 = R W_2 + (I - R^2)^{1/2} W_{1.2}$, where $W_{1.2}$ and W_2 are uncorrelated and hence independent Brownian motions.

Theorem 3 *Under the null hypothesis* $\mathcal{H}_0 : \Pi = 0$ *and Assumptions 1-2 with* $q > 6$*,*

$$\mathrm{W}_n^H \;\Rightarrow\; tr \int dW_1 W_2' (\int W_2 W_2')^{-1} \int W_2 dW_1' \equiv \mathrm{W}(R) \tag{14}$$

where R is defined in equation (13). Also, we have

$$\mathrm{W}(R) = tr\ f(R)'\ (\int W_2 W_2')^{-1}\ f(R), \tag{15}$$

where $f(R) = \int W_2 dW_2' R + \int W_2 dW_{1.2}' (I - R^2)^{1/2}$*, and* $W_{1.2}$ *and* W_2 *are independent standard Brownian motions.*

Theorem 3 uses the condition up to the sixth finite moment to show the asymptotic theory for the Wald statistic for cointegration with the heteroskedasticity-robust covariance estimator. The condition with 2^+ finite moment is sufficient if the cointegration test is based on the standard covariance estimator, which assumes time-invariant variance.

The asymptotic distribution of the Wald statistic is equivalent to that of the LR statistic, which has been found by Seo (1998). The Wald statistic, which uses the heteroskedasticity-robust covariance estimator, has the same asymptotic distribution as the Wald statistic that is based on the standard covariance estimator. Although the asymptotic distribution of the Wald statistic for cointegration is invariant to the heteroskedastic errors, the small sample properties of the cointegration tests depend on heteroskedastic errors. In particular, the standard residual bootstrap inference on cointegration depends heavily on the error condition. Because most economic variables contain time-varying conditional variance, the finite sample performance of the cointegration tests can be improved by using the heteroskedasticity-robust covariance estimator.

We note that the asymptotic distribution depends on two correlated Brownian motions. If there is no covariate effect, then Brownian motions are collinear ($R = I$) and the Wald statistic has the following asymptotic distribution:

$$\mathrm{W}(I) = \mathrm{tr} \int dW_2 W_2' (\int W_2 W_2')^{-1} \int W_2 dW_2', \tag{16}$$

which is the same nonstandard distribution found by Johansen (1988, 1991).

If $W_1(r)$ and $W_2(r)$ are independent ($R = 0$), the stochastic functional $\int dW_{1.2} \otimes W_2$ follows a locally asymptotically mixed normal distribution with covariance matrix $I \otimes \int W_2 W_2'$. In this case, the Wald statistic is asymptotically chi-squared with p^2 degrees of freedom. Therefore, the distribution of the Wald statistic is a mixed combination of the nonstandard and the chi-squared distributions. The limiting distribution approaches the chi-squared distribution as the magnitude of the covariate effect increases.

Hansen (1995) has shown that the asymptotic distribution of the ADF unit root test with stationary covariates is a combination of the standard normal and the Dickey-Fuller nonstandard distributions. This paper extends the univariate analysis to the cointegration tests in the vector error correction model.

Because the stationary covariates affect the asymptotic distribution of the cointegration test, inference on cointegration depends on the canonical correlation. The tabulated critical values, which assume the nonstandard distribution found by Johansen (1988, 1991), do not allow for the covariate effect. This paper proposes two methods to implement inference on cointegration.

The first method simulates the null distribution by using the estimated canonical correlation. Because the matrix R is the canonical correlation of two innovations $U(r)$ and $V(r)$, it can be estimated from the long-run variance of $\{\hat{u}_t, \hat{v}_t\}$, where $\{\hat{u}_t, \hat{v}_t\}$ is the estimated error of $\{u_t, v_t\}$.

The long-run variance of $\hat{v}_t$ can be estimated as follows:

$$\hat{\Omega}_{vv} = \hat{\Sigma} + \hat{\Psi}(1)(I - \hat{\Phi}_2(1))^{-1}\hat{\Sigma}_{ee}(I - \hat{\Phi}_2(1))^{-1'}\hat{\Psi}'(1),$$

where $\hat{\Sigma} = n^{-1}\sum_{t=1}^{n} \hat{u}_t\hat{u}_t'$.

The canonical correlation estimates $\hat{\rho}_i$ for $i = 1, 2, \cdots, p$ can be calculated from the following:

$$\mid \hat{\rho}_i^2 I - \hat{\Sigma}^{1/2'}\hat{\Omega}_{vv}^{-1}\hat{\Sigma}^{1/2} \mid = 0.$$

The null distribution can be simulated by using the estimated canonical correlations, and the p-values can be calculated by simulation.[1]

The second method is based on the bootstrap inference, which does not require estimation of the canonical correlation. In this paper, we consider the standard residual bootstrap algorithm. The residual bootstrap approximates the sampling distribution of the test statistic using the null model and the parameter estimates obtained under the null hypothesis.

[1] A Gauss program which performs the necessary simulation can be provided upon request.

To implement the bootstrap inference, we estimate the error correction model (3) and the VAR model (2). The VAR lag-lengths can be selected by the AIC or the BIC.

The resampled residuals u_t^b and e_t^b are randomly drawn from the sample residuals, and then x_t^b and z_t^b can be constructed using the parameter estimates and the resampled residuals. The W^b statistic can be calculated for each resampled data set, and then we obtain the bootstrap p-value, which is the probability that the simulated statistic exceeds the sample Wald statistic. If the p-value is less than the size chosen, then we reject the null hypothesis in favor of the alternative of the presence of cointegration. The Monte Carlo simulation reveals that the standard residual bootstrap inference generates moderate performance on the size and power of the cointegration test.

5 Omitted Stationary Covariates

We consider a misspecified model where the stationary covariates are excluded.

$$\Delta x_t = \Pi x_{t-1} + \Gamma(L)\Delta x_t + \nu_t. \tag{17}$$

From the representation theorem, the error ν_t in the misspecified model can be written as follows:

$$\nu_t = u_t + \Psi(L)z_t = H(L)v_t,$$

where $H(L) = I + \Psi(L)(I - \Phi_2(L))^{-1}\Phi_1(L)C(L)$.

The error ν_t contains the current and lagged values of stationary covariates. Thus, the error in the misspecified model is serially correlated and correlated with the regressor. Here, we consider the pitfalls in the estimation and inference on cointegration when the stationary covariates are excluded.

5.1 Cointegrating Vector Estimator

Suppose the cointegrating vector is estimated in the following model:

$$\Delta x_t = \alpha w_{t-1}(\boldsymbol{\beta}) + \Gamma(L)\Delta x_t + \nu_t. \tag{18}$$

If there is no covariate effect, the error ν_t is equivalent to u_t. However, the error ν_t is serially correlated in general when there is a covariate effect.

We denote $\tilde{\boldsymbol{\beta}}$ as the cointegrating vector estimator of the misspecified model, which is defined as the solution to the objective function as follows:

$$\tilde{\mathcal{L}}_n(\boldsymbol{\beta}, \alpha, \Gamma, \Sigma_{\nu\nu}) = -\frac{n}{2}\log|\Sigma_{\nu\nu}| - \frac{1}{2}\sum_{t=1}^{n}\nu_t^{'}\Sigma_{\nu\nu}^{-1}\nu_t, \tag{19}$$

where $\Sigma_{\nu\nu} = E(\nu_t\nu_t^{'})$.

Assumption 3 $\sum_{k=1}^{\infty} k|H_k| < \infty$, *where* $\nu_t = H(L)v_t = \sum_{k=0}^{\infty} H_k v_{t-k}$.

In this section, we use the following asymptotic results:

Lemma 4 *Under Assumptions 1-3 and* $\Pi = \alpha\beta^{'}$,

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} x_{t-1}v_t^{'} &\Rightarrow C(1)\int_0^1 VdV^{'} + M \\ n^{-1}\sum_{t=1}^{n} x_{t-1}\nu_t^{'} &\Rightarrow C(1)\int_0^1 VdV^{'}H^{'}(1) + \Upsilon, \end{aligned}$$

where $\Upsilon = MH^{'}(1) + \sum_{k=0}^{\infty} J_k \sum_{i=k}^{\infty} H_i^{'}$, $J_k = E(\Delta x_t v_{t+k}^{'})$, $M = C(1)\Lambda + E((C^*(L)v_{t-1})v_t^{'})$, *and* $\Lambda = \sum_{k=1}^{\infty} E(v_t v_{t+k}^{'})$.

We denote $\Omega_{\nu\nu}$ as the long-run variance of ν_t. To derive the asymptotic distribution of the cointegrating vector estimator of the misspecified model, we define the Brownian motions $\tilde{B}_2(r)$ and $B_2(r)$ as follows:

$$\begin{pmatrix} M_1^{-1/2}\alpha^{'}\Sigma_{\nu\nu}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\nu_t \\ (\alpha_\perp^{'}\Omega_{vv}\alpha_\perp)^{-1/2}\alpha_\perp^{'}\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}v_t \end{pmatrix} \Rightarrow \begin{pmatrix} \tilde{B}_2(r) \\ B_2(r) \end{pmatrix} = BM(\begin{pmatrix} I & G \\ G^{'} & I \end{pmatrix}) \tag{20}$$

where $G = M_1^{-1/2}\alpha^{'}\Sigma_{\nu\nu}^{-1}H(1)\Omega_{vv}\alpha_\perp(\alpha_\perp^{'}\Omega_{vv}\alpha_\perp)^{-1/2'}$ and $M_1 = \alpha^{'}\Sigma_{\nu\nu}^{-1}\Omega_{\nu\nu}\Sigma_{\nu\nu}^{-1}\alpha$.

If there is no covariate effect, $H(L) = I$ and $\Sigma_{\nu\nu} = \Omega_{vv} = \Sigma$. Thus, G becomes zero, and two Brownian motions $\tilde{B}_2(r)$ and $B_2(r)$ are mutually independent. However, these two Brownian motions are correlated if the covariate effect is non-zero. We define the standard Brownian motion $B_{1.2}(r)$, which is independent of $B_2(r)$ and satisfies $\tilde{B}_2(r) = GB_2(r) + (I - GG^{'})^{1/2}B_{1.2}(r)$.

Theorem 4 *Under Assumptions 1-3 and* $\Pi = \alpha\beta'$,

$$n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow A(\int B_2 B_2')^{-1}(\int B_2 d\tilde{B}_2' + \Delta) M_1^{1/2'} M_2^{-1}, \tag{21}$$

where $\Delta = \Omega_{vv}^{-1/2} C_2^{-1} \Upsilon \Sigma_{\nu\nu}^{-1} \alpha M_1^{-1/2'}$ *and* $M_2 = \alpha' \Sigma_{\nu\nu}^{-1} \alpha$.

Also, we have

$$n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow A(\int B_2 B_2')^{-1}(f(G) + \Delta) M_1^{1/2'} M_2^{-1}, \tag{22}$$

where $f(G) = \int B_2 dB_2' G' + \int B_2 d\tilde{B}_{1.2}' (I - GG')^{1/2'}$.

First, the distribution of the cointegrating vector estimator $n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})$ depends on the functional $f(G)$, which is a mixed combination of the mixture normal and the nonstandard distributions. If there is no covariate effect, then the correlation matrix G becomes zero, and the cointegrating vector follows a mixed normal distribution. As the covariate effect increases, the correlation increases, and the limiting distribution tends to depart from normality. Because the nonstandard distribution is skewed and leptokurtic, standard inference generates invalid results.

Second, the error of the misspecified model is serially correlated, and thus the cointegrating vector estimator entails the asymptotic bias Δ. If there is no covariate effect, the asymptotic bias disappears as the error ν_t becomes independent. However, the covariate effect generates the asymptotic bias, which leads to inferential difficulty when the standard distribution theory is applied to the cointegrating vector estimator.

The simulation evidence shows that the distribution of the t-statistic based on $\tilde{\boldsymbol{\beta}}$ is close to the normal distribution if the covariate effect is not strong. However, as the covariate effect increases, the distribution of the t-statistic shows wide variance, asymmetry, and fat-tailed behavior. Therefore, the asymptotic distribution of the cointegrating vector estimator depends on the covariate effect if the covariates are omitted. The limiting distribution is nonstandard, and thus the misspecified model leads to the departure from the standard distribution theory.

5.2 Inference on Cointegration

Next, we consider the cointegration tests in the misspecified model (17) where the stationary covariates are excluded. The error in the misspecified model is serially correlated because the error ν_t contains the covariates. In this section we consider the Wald statistic, which is based on the standard covariance estimator.

$$\tilde{\mathrm{W}}_n = \mathrm{tr}\ n\tilde{\Sigma}_{\nu\nu}^{-1}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10},$$

where

$$\begin{aligned}
\tilde{\Sigma}_{\nu\nu} &= \frac{1}{n}\sum_{t=1}^{n}\Delta x_t\Delta x_t^{'} - \frac{1}{n}\sum_{t=1}^{n}\Delta x_t\tilde{s}_{1t}^{'}(\sum_{t=1}^{n}\tilde{s}_{1t}\tilde{s}_{1t}^{'})^{-1}\sum_{t=1}^{n}\tilde{s}_{1t}\Delta x_t^{'} \\
\tilde{S}_{10} &= \frac{1}{n}\sum_{t=1}^{n}x_{t-1}\Delta x_t^{'} - \frac{1}{n}\sum_{t=1}^{n}x_{t-1}s_{1t}^{'}(\sum_{t=1}^{n}s_{1t}s_{1t}^{'})^{-1}\sum_{t=1}^{n}s_{1t}\Delta x_t^{'}, \\
\tilde{S}_{11} &= \frac{1}{n}\sum_{t=1}^{n}x_{t-1}x_{t-1}^{'} - \frac{1}{n}\sum_{t=1}^{n}x_{t-1}s_{1t}^{'}(\sum_{t=1}^{n}s_{1t}s_{1t}^{'})^{-1}\sum_{t=1}^{n}s_{1t}x_{t-1}^{'}
\end{aligned}$$

and where $\tilde{s}_{1t} = (x_{t-1}^{'}, s_{1t}^{'})^{'}$.

The LR test statistic for cointegration in the misspecified model is given by

$$\tilde{\mathrm{LR}}_n = \mathrm{tr}\ n\tilde{S}_{00}^{-1}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10},$$

where $\tilde{S}_{00} = \frac{1}{n}\sum_{t=1}^{n}\Delta x_t\Delta x_t^{'} - \frac{1}{n}\sum_{t=1}^{n}\Delta x_t s_{1t}^{'}(\sum_{t=1}^{n}s_{1t}s_{1t}^{'})^{-1}\sum_{t=1}^{n}s_{1t}\Delta x_t^{'}$.

If the stationary covariates are omitted, the Wald statistic for cointegration has the limiting distribution as follows:

Theorem 5 *Under the null hypothesis $\mathcal{H}_0 : \Pi = 0$ and Assumptions 1-3,*

$$\tilde{\mathrm{W}}_n \Rightarrow \mathrm{tr}\ (\int dW_2W_2^{'} + J^{'})(\int W_2W_2^{'})^{-1}(\int W_2dW_2^{'} + J)\ K, \tag{23}$$

where $J = \Omega_{vv}^{-1}C^{-1}\Upsilon H(1)^{-1'}\Omega_{vv}^{-1/2'}$ and $K = \Omega_{vv}^{1/2'}H(1)^{'}\Sigma_{\nu\nu}^{-1}H(1)\Omega_{vv}^{1/2}$.

Because $\tilde{S}_{00}$ converges to $\Sigma_{\nu\nu}$ in probability under the null of no cointegration, the LR statistic $\tilde{\mathrm{LR}}_n$ and the Wald statistic $\tilde{\mathrm{W}}_n$ follow the same asymptotic distribution. The

Wald and the LR statistics for cointegration in the misspecified model are asymptotically distributed as Johansen's nonstandard distribution up to the scale effect K and the location shift effect J. If $K = 1$ and $J = 0$, then the asymptotic distribution is exactly the same as Johansen's nonstandard distribution. If the covariate effect is zero, the scale effect and the location shift effect disappear, and so there is no size distortion when the nonstandard distribution is applied. However, the covariate effect is likely to increase the scale effect and the location shift effect, and thus the Wald and the LR statistics tend to over-reject the null hypothesis if the cointegration test is based on Johansen's nonstandard distribution.

6 Models with Deterministic Trends

We consider the vector error correction model with deterministic trends as follows:

$$\Delta x_t = \mu k_t + \Pi x_{t-1} + \sum_{i=1}^{l} \Gamma_i \Delta x_{t-i} + \sum_{i=0}^{m} \Psi_i z_{t-i} + u_t.$$

If $k_t = 1$, then the model allows an intercept. If $k_t = (1, t)'$, then the model allows an intercept and a linear trend. For the general $(p \times q)$ coefficient matrix μ, we assume that $\alpha_{\perp}' \mu_q = 0$, where μ_q is the q-th column of μ, which corresponds to the trend of the highest order. Thus, we preclude the possibility that the demeaned or detrended variables contain the deterministic trend.

If the information on the deterministic trend is given, efficiency can be improved by using the model with the restriction on the deterministic trend. However, this restriction may cost a specification error, and thus we do not impose the restriction on the deterministic trend. Because our model uses the detrended data, our results are robust to the deterministic trend.

We define the detrended variable x_t^* as follows.

$$x_t^* = x_t - (\sum_{t=1}^{n} x_t s_t^{*'})(\sum_{t=1}^{n} s_t^* s_t^{*'})^{-1} s_t^*,$$

where $s_t^* = (k_t', s_{1t}', s_{2t}')'$.

Because x_t^* removes deterministic trends, the asymptotic distribution of the cointegrating vector estimator based on the detrended variable is invariant to the deterministic trends. In the model with detrended variables, we can extend the previous analysis without difficulty.

$$n^{-1/2}x^*_{[nr]} \Rightarrow C(1)V(r) - \int_0^1 C(1)V(r)K'(r)dr[\int_0^1 K(r)K'(r)dr]^{-1}K(r) \equiv C(1)V^*(r),$$

where $K(r) = 1$ for the model with an intercept, and $K(r) = (1, r)'$ for the model with an intercept and a linear trend.

Lemma 5 *Suppose $\alpha'_\perp \mu_q = 0$. Under Assumptions 1-2 and $\Pi = \alpha\beta'$,*

$$n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \Rightarrow A(\int B_2^* B_2^{*'})^{-1} \int B_2^* dB_1'(\alpha'\Sigma^{-1}\alpha)^{-1/2'} \tag{24}$$

where $B_2^*(r) = B_2(r) - \int_0^1 B_2(r)K'(r)dr[\int_0^1 K(r)K'(r)dr]^{-1}K(r)$.

In the models with deterministic trends, the asymptotic distribution of the cointegrating vector estimator is a mixed normal with a variance of $(\alpha'\Sigma^{-1}\alpha)^{-1} \otimes A(\int B_2^* B_2^{*'})^{-1}A'$. Thus, the asymptotic distribution of the cointegrating vector estimator is invariant to the deterministic trend.

In the model with the deterministic trend, the Wald statistic for the cointegration test can be defined in the same way by using the sample moments, which are projected on $s_t^* = (k_t', s_t')'$. If we assume that $\alpha'_\perp \mu_q = 0$, then the detrending removes the deterministic trends. Since $n^{-1/2}x^*_{[nr]} \Rightarrow CV^*(r)$, where $V^*(r) = V(r) - \int V(r)dK'(r)(\int K(r)K'(r)dr)^{-1}K(r)$, the proof of the following result is analogous to that of Theorem 3.

Lemma 6 *Suppose $\alpha'_\perp \mu_q = 0$. Under $\mathcal{H}_0 : \Pi = 0$ and Assumptions 1-2 with $q > 6$,*

$$\mathrm{W}_n^H \Rightarrow \mathrm{W}^*(R) = \mathrm{tr} \int dW_1 W_2^{*'} (\int W_2^* W_2^{*'})^{-1} \int W_2^* dW_1', \tag{25}$$

where $W_2^*(r) = W_2(r) - \int W_2(r)dK'(r)(\int K(r)K'(r)dr)^{-1}K(r)$,

$$\begin{pmatrix} W_1(r) \\ W_2(r) \end{pmatrix} = BM \begin{pmatrix} I & R \\ R & I \end{pmatrix},$$

and R is defined in equation (13).

7 Simulation Evidence

In this section, we examine the small sample performances of the cointegrating vector estimator and the cointegration tests by using Monte Carlo simulation. The simulations are based on a bivariate error correction model with the stationary covariates z_t, which follow a bivariate VAR process.

$$\Delta x_t = \mu + \Pi x_{t-1} + \Gamma \Delta x_{t-1} + \begin{pmatrix} \psi & 0 \\ 0 & \psi \end{pmatrix} z_t + \begin{pmatrix} \psi & 0 \\ 0 & \psi \end{pmatrix} z_{t-1} + u_t, \tag{26}$$

$$z_t = \begin{pmatrix} \phi_1 & 0 \\ 0 & \phi_1 \end{pmatrix} \Delta x_{t-1} + \begin{pmatrix} \phi_2 & 0 \\ 0 & \phi_2 \end{pmatrix} z_{t-1} + e_t, \tag{27}$$

where $x_t = (x_{1t}, x_{2t})'$, $z_t = (z_{1t}, z_{2t})'$, $u_t = (u_{1t}, u_{2t})'$, and $e_t = (e_{1t}, e_{2t})'$.

7.1 Cointegrating Vector Estimator

First, we design the experiments on the distribution of the cointegrating vector estimators in the model (26) with

$$\Pi = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \begin{pmatrix} 1 \\ \beta \end{pmatrix}'.$$

The error process (u_t, e_t) is generated by the Gauss random number generator. We assume that the errors are independently $N(0, I)$-distributed. In the simulation, we principally focus on the effect of stationary covariates, and the issue regarding heteroskedasticity will be treated in separate research. The experiments on the distribution of the estimators are based on a sample size of 250 and 10,000 simulation replications. In the simulation, we fix $\mu = 0$, $\beta = -1$, $\alpha_1 = -1$, $\alpha_2 = 0$, and $\Gamma = 0$. We vary ϕ_1 among $(0, 0.2)$, ϕ_2 among $(0, 0.25, 0.5, 0.75, 0.95)$, and ψ among $(0, 0.25, 0.5)$.

Table 1 shows the root mean squared error (RMSE) and the mean absolute error (MAE) of the cointegrating vector estimators. When there is no covariate effect, the cointegrating vector with stationary covariates is almost equivalent to the estimator without covariates. As the covariate effect increases, the RMSE and MAE of the MLE $\hat{\beta}$ decrease sharply, while those of the estimator $\tilde{\beta}$ remain stable. For example, at $(\phi_1, \phi_2) = (0.2, 0.25)$ both the RMSE and MAE of the MLE $\hat{\beta}$ decrease about 28% as the covariate effect ψ increases from 0 to 0.5. Thus, the efficiency gain of the MLE of the cointegrating vector improves in the covariate effect.

We examine the size performance of the t-statistics for the null hypothesis $\mathcal{H}_0 : \beta = -1$. Table 1 shows the descriptive statistics and the coverage rates of the t-statistics based on the cointegrating vector estimators $\hat{\beta}$ and $\tilde{\beta}$. The coverage rate is defined as $P(T < \upsilon_{0.05})$ for the lower 5% size and $P(T > \upsilon_{0.95})$ for the upper 5% size, where T is the t-statistic and υ is the critical value.

The descriptive statistics of the t-statistics based on the estimator $\hat{\beta}$ are consistent with the properties of the normal distribution. The coverage rates are very close to the true size for most parameter values, and thus statistical inference on the cointegrating vector can be based on the standard theory. However, the t-statistic based on the estimator $\tilde{\beta}$ reveals a large amount of size distortion, large variance, asymmetry, and leptokurtic behavior as the covariate effect increases. For example, at $(\phi_1, \phi_2, \psi) = (0.2, 0.75, 0.5)$ the t-statistic based on the estimator $\tilde{\beta}$ shows that the standard deviation is larger than unity, and so the lower and upper 5% coverage rates are overly stated. Because the cointegrating vector estimator follows the nonstandard distribution if the stationary covariates are omitted, inference based on the standard distribution may bring invalid results.

7.2 Size of Cointegration Tests

Next, we set up the experiment on the null distribution of the cointegration tests. The null hypothesis of no cointegration is set as $\Pi = 0$ in (26).

The small sample performance of the cointegration tests is analyzed by using the asymptotic distribution theory and the bootstrap inference. The sample canonical correlations are estimated by using the estimated coefficients to simulate the asymptotic null distribution. The simulation on the size of the cointegration tests is based on a sample size of 250 and 1,000 replications, and for each replication 500 bootstrap replications are made to calculate the bootstrap p-values in the bootstrap inference.

We allow for an intercept, one lagged variable ($l = 1$), and the covariates z_t and z_{t-1}. We fix $\mu = 0$ and $\Gamma = 0$. We vary ϕ_1 among $(0, 0.2)$, ϕ_2 among $(0, 0.25, 0.5)$, and ψ among $(0, 0.25, 0.5)$. The error process (u_t, e_t) is randomly drawn from $N(0, I)$.

Table 2 reports the rejection frequencies of the cointegration tests with or without covariates at the nominal sizes 5% and 10%. The rejection frequencies are the percentage of the simulated p-values which are smaller than the nominal size. If we use the correct null distribution $W(R)$, the rejection frequencies of the cointegration tests with stationary covariates are very close to the true size for most parameter values, and they do not appear to be affected by the covariate effect. However, if the null distribution is simulated by the asymptotic distribution $W(I)$, which does not account for the covariate effect, the rejection frequencies are lower than the true size. Thus, the cointegration tests with stationary covariates tend to reveal under-sized performances if we use the tabulated critical values, which do not allow for stationary covariates.

Table 2 also shows the small sample performance on the size of the bootstrap cointegration tests with stationary covariates. The rejection frequencies of the bootstrap inference are very close to the true size for most parameter values, as in the asymptotic inference.

On the other hand, the cointegration tests in the misspecified model, which neglect the stationary covariates, tend to over-reject the null hypothesis as the covariate effect increases. For example, at $(\phi_1, \phi_2, \psi) = (0.2, 0.75, 0.5)$, the test statistics $\tilde{\mathrm{W}}_n$ and $\tilde{\mathrm{LR}}_n$ reject 13.9% and 12.6% of the null hypothesis, respectively, at the 5% size.

7.3 Power of Cointegration Tests

Next, we consider the experiment on the power of the cointegration tests. The null and the local alternative hypotheses are postulated as follows:

$$\mathcal{H}_0 : \alpha_n = 0 \ \text{ against } \ \mathcal{H}_n : \alpha_n = -\frac{\delta}{n},$$

where

$$\Pi = \begin{pmatrix} \alpha_n \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ \beta \end{pmatrix}.$$

In the simulation, we fix $\mu = 0$, $\Gamma = 0$, and $\beta = -1$ in (26). The model allows for an intercept, one lagged variable ($l = 1$), and the covariates z_t. We vary the parameters ϕ_1, ϕ_2, and ψ to take on several values. The experiment on the power is based on a sample size of 250 and 1,000 replications.

We increase the shift parameter δ from 0 to 5, 10, and 15. If $\delta = 0$, then the null hypothesis is maintained and there is no cointegration. However, if $\delta > 0$, then the alternative hypothesis holds and the Wald and the LR statistics tend to reject the null hypothesis of no cointegration.

Table 3 shows the rejection frequency of the cointegration tests at the 5% size. As Table 3 shows, the rejection frequency of the tests increases as the shift parameter δ deviates from the null hypothesis. In particular, the power of the cointegration tests with stationary covariates improves significantly as the covariate effect increases. For example, at $(\phi_1, \phi_2, \psi) = (0.2, 0.75, 0.5)$, the Wald and the LR statistics reject all but about 95 percent of the null hypothesis at the local alternative $\delta = 5$.

Table 3 also reports the small sample performance on the power of the bootstrap inference on cointegration with stationary covariates. In the bootstrap cointegration tests, the calculation of the canonical correlations is not required. The power performance of the bootstrap inference is close to that of the asymptotic inference. Although the bootstrap refinement does not appear to be strong, the bootstrap inference generates moderate performances on the size and power of the cointegration tests.

The cointegration tests in the misspecified model, where the covariates are omitted, reject about 40 percent of the null hypothesis at $(\phi_1, \phi_2, \psi) = (0.2, 0.75, 0.5)$ and $\delta = 5$. However, this outcome depends largely on the scale effect and the location shift effect, which tend to inflate the test statistics. Thus, inference based on the tabulated nonstandard distribution, which ignores the scale effect, is likely to over-reject the null distribution as the magnitude of the covariate effect increases.

8 Empirical Application

This section provides an empirical analysis on the cointegrating relationship of the U.S. money demand equation. The money demand relation has been assessed in many empirical studies such as Baba et al. (1992). Here, we consider the estimation and cointegration inference of the money demand relation with the stationary covariates.

The data set is extracted from the Federal Reserve Economic Database (FRED) for the period 1960Q1-1999Q4: m_t is M2, p_t the price index, y_t real GDP, and r_t the TB 3-month interest rate. All variables are in logarithms except the short-run interest rate. We also consider the stationary covariates: z_{1t} is the oil price change, and z_{2t} is the risk factor, which is defined as the difference between the Moody's Baa and Aaa corporate bond rates. The monthly data are converted to the quarterly data by taking the three-month average.

Because the real money balances and real income contain growth terms, we include a constant and a linear trend in the error correction model. The VAR model of (z_{1t}, z_{2t}) allows a constant. The augmented Dickey-Fuller test rejects the unit root hypothesis of the oil price change and the risk factor. The empirical results show that the covariates are weakly exogenous to the cointegrating relationship. The lag lengths of the ECM and the VAR model picked sufficiently large by Akaike information criterion (AIC) are $l = 5$ and $m = 1$, respectively.

The long-run money demand relationship and adjustment coefficients are estimated in Table 4. When the stationary covariates are used, the standard error of the long-run income elasticity decreases compared to the estimate without covariates. The income elasticity of the M2 demand is estimated as greater than unity. The estimated interest semi-elasticity is statistically significant and corresponds to the economic model. On the other hand, the interest elasticity is positive if the stationary covariates are not used. When we use the covariates, the adjustment coefficient of the money equation becomes larger, which implies that the adjustment process becomes faster. The R-squared coefficients and the log-likelihoods increase in each equation when we use the stationary covariates.

The Ljung-Box Q-statistics show that the estimated errors are not serially correlated at the lag length 1 in each equation. However, at the lag length 12, the serial correlation appears in the interest rate equation when the stationary covariates are not used. The residual of the interest rate equation shows irregular movement around the early 1980s, when the Fed operating system was changed. The ARCH LM test rejects the null hypothesis of no ARCH effect in the interest rate equation. Although the stationary covariates help reduce the ARCH effect, the interest rate equation still contains heteroskedastic errors because of the policy change.

Table 5 shows that the null hypothesis of no cointegration is maintained at the 5% size if we use the LR test statistic without stationary covariates. When the stationary covariates are used, the LR statistic rejects the null at the 5% size based on the asymptotic p-value. The magnitude of the covariate effect can be measured by the canonical correlation, which is estimated at $(0.247, 0.973, 1.000)$. However, the bootstrap p-value does not support the cointegrating relationship. The Wald statistic, which uses the standard covariance estimator, generates the same result. If we use the Wald statistic based on the heteroskedasticity-consistent covariance estimator, the bootstrap p-value, as well as the asymptotic p-value, improve empirical evidence in favor of the presence of the long-run money demand relation.

9 Concluding Remarks

In this study, we find that the stationary covariates affect the asymptotic distribution of the estimation and inference in the cointegrated system. The distribution of the cointegrating vector estimator is a mixed normal, and the efficiency of the estimator improves as the covariate effect increases. The test for cointegration has the asymptotic distribution, which is a combination of the chi-squared and the nonstandard distributions. As the covariate effect increases, the limiting distribution approaches the chi-squared distribution. Therefore, the covariate effect generates a significant power gain. Furthermore, this study shows that the omission of the stationary covariates affects the distribution theory of the cointegrating vector estimator and the cointegration test.

Although stationary covariates have been used in economic models, the estimation and inference in the cointegrated system have been based on the distribution theory, which does not allow stationary covariates. Furthermore, in the econometric model with nonstationary variables, the theoretical aspects of the stationary covariates have been disregarded or treated less importantly. Because it shows the distribution theory and proper inference in the cointegrated system with stationary covariates, this study is useful and required.

References

[1] Ahn, S. K., and G. C. Reinsel, 1988, Nested Reduced-Rank Autoregressive Models for Multiple Time Series, *Journal of the American Statistical Association*, 83, 849-856.

[2] Andrews, D. W. K., 1991, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation, *Econometrica*, 59, 817-858.

[3] Baba, Y., D. Hendry, and R. Starr, 1992, The Demand for M1 in the U. S. A., 1960-1988, *Review of Economic Studies*, 59, 25-61.

[4] Billingsley, P., 1968, *Convergence of Probability Measures*, New York: John Wiley.

[5] Boswijk, H. P., 1995, Efficient Inference on Cointegration Parameters in Structural Error Correction Models, *Journal of Econometrics*, 69, 133-158.

[6] Box, G. E. P., and G. C. Tiao, 1977, A Canonical Analysis of Multiple Time Series, *Biometrika*, 64, 355-365.

[7] Chambers, M., 2003, The Asymptotic Efficiency of Cointegration Estimators under Temporal Aggregation, *Econometric Theory*, 19, 49-77.

[8] Elliot, G., and M. Jansson, 2003, Testing for Unit Root with Stationary Covariates, *Journal of Econometrics*, 115, 75-89.

[9] Engle, R. F., and C. W. J. Granger, 1987, Cointegration and Error Correction Representation, Estimation, and Testing, *Econometrica*, 55, 251-276.

[10] Engle, R. F., and D. F. Hendry, 1993, Testing Superexogeneity and Invariance in Regression Models, *Journal of Econometrics*, 56, 119-139.

[11] Engle, R. F., D. F. Hendry, and J. F. Richard, 1983, Exogeneity, *Econometrica*, 51, 277-304.

[12] Ericsson, N. R., 1995, Conditional and Structural Error Correction Models, *Journal of Econometrics*, 69, 159-171.

[13] Granger, C.J., 1981, Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics*, 16, 121-130.

[14] Hall, P., and C. Heyde, 1980, *Martingale Limit Theory and Its Application*, New York: Academic Press.

[15] Hansen, B. E., 1992, Convergence to Stochastic Integrals for Dependent Heterogeneous Processes, *Econometric Theory*, 8, 489-500.

[16] Hansen, B. E., 1995, Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power, *Econometric Theory*, 11, 1148-1171.

[17] Harbo, I., S. Johansen, B. Nielson, and A. Rahbek, 1998, Asymptotic Inference on Cointegration Rank in Partial Systems, *Journal of Business and Economic Statistics*, 16, 388-399.

[18] Johansen, S., 1988, Statistical Analysis of Cointegrating Vectors, *Journal of Economic Dynamics and Control*, 12, 231-254.

[19] Johansen, S., 1991, Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica*, 59, 1551-1580.

[20] Johansen, S., 1992, Cointegration in Partial Systems and the Efficiency of Single-Equation Analysis, *Journal of Econometrics*, 52, 389-402.

[21] Johansen, S., and K. Juselius, 1992, Testing Structural Hypotheses in a Multivariate Cointgeration Analysis of the PPP and the UIP for UK, *Journal of Econometrics*, 53, 211-244.

[22] Juselius, K., 1995, Do Purchasing Power Parity and Uncovered Interest Parity Hold in the Long Run? An Example of Likelihood Inference in a Multivariate Time Series Model, *Journal of Econometrics*, 69, 211-240.

[23] Kremers, J.M., N.R. Ericsson, and J.J. Dolado, 1992, The Power of Cointegration Tests, *Oxford Bulletin of Economics and Statistics*, 54, 325-348.

[24] Lee, T.H., and Y. Tse, 1996, Cointegration Tests with Conditional Heteroskedastcity, *Journal of Econometrics*, 73, 401-410.

[25] Park, J. Y., and P. C. B. Phillips, 1988, Statistical Inference in Regressions with Integrated Processes: Part 1, *Econometric Theory*, 4, 468-497.

[26] Pham, T.D., and L.T. Tran, 1985, Some Mixing Properties of Time Series Models, *Stochastic Processes and their Applications*, 19, 297-303.

[27] Phillips, P. C. B., 1988, Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations, *Econometric Theory*, 4, 528-533.

[28] Phillips, P.C.B., 1991, Optimal Inference in Cointegrated System, *Econometrica*, 59, 283-306.

[29] Phillips, P.C.B., and S.N. Durlauf, 1986, Multiple Time Series with Integrated Variables, *Review of Economic Studies*, 53, 473-495.

[30] Phillips, P.C.B., and B.E. Hansen, 1990, Statistical Inference in Instrumental Variables Regression with I(1) Process, *Review of Economic Studies*, 57, 99-124.

[31] Rao, C.R., 1973, *Linear Statistical Inference and Its Applications*, New York: John Wiley.

[32] Saikkonen, P., 1991, Asymptotically Efficient Estimation of Cointegration Regressions, *Econometric Theory*, 7, 1-21.

[33] Saikkonen, P., 1992, Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation, *Econometric Theory*, 8, 1-27.

[34] Seo, B., 1998, Statistical Inference on Cointegration Rank in Error Correction Models with Stationary Covariates, *Journal of Econometrics*, 85, 339-386.

[35] Stock, J. and M. Watson, 1993, A Simple Estimator of Cointegrating Vectors in Highly Order Integrated Systems, *Econometrica*, 61, 783-820.

Appendix: Mathematical Proofs

Proof of Lemma 1:

The invariance principle of Phillips and Durlauf (1986) implies

$$\begin{pmatrix} n^{-1/2}\sum_{t=1}^{[nr]} u_t \\ n^{-1/2}\sum_{t=1}^{[nr]} e_t \end{pmatrix} \Rightarrow \begin{pmatrix} U(r) \\ E(r) \end{pmatrix} = BM(\begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma_{ee} \end{pmatrix}).$$

We show $V_{[nr]} \Rightarrow V(r) = BM(\Omega_{vv})$, where $V_{[nr]} = n^{-1/2}\sum_{t=1}^{[nr]} v_t$ and $v_t = u_t + B(L)e_t$. Define $B^*(L) = \frac{B(L)-B(1)}{1-L}$. Since

$$\sup_t \|B^*(L)e_t\|_q \leq \sum_{j=0}^{\infty} |B_j^*| \ \sup_t \|e_t\|_q \leq \sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty} |B_k| \ \sup_t \|e_t\|_q = \sum_{k=1}^{\infty} k|B_k| \ \sup_t \|e_t\|_q < \infty$$

for some $q > 2$,

$$P(sup_{r\in[0,1]} n^{-1/2}|V_{[nr]} - \sum_{t=1}^{[nr]} u_t - B(1)\sum_{t=1}^{[nr]} e_t| > \epsilon) \leq P(sup_{r\in[0,1]} n^{-1/2}|B^*(L)e_{[nr]}| > \epsilon) \to 0.$$

Thus, Assumptions 1-2 imply

$$n^{-1/2}\sum_{t=1}^{[nr]} v_t \Rightarrow V(r) = BM(\Omega_{vv}),$$

where $\Omega_{vv} = \Sigma + B(1)\Sigma_{ee}B(1)^{'}$.

Proof of Lemma 2:

We show $n^{-1/2}x_{[nr]} \Rightarrow C(1)V(r)$. We need to show

$$P(sup_{r\in[0,1]} n^{-1/2}|x_{[nr]} - C(1)\sum_{t=1}^{[nr]} v_t| > \epsilon) \leq P(sup_{r\in[0,1]} n^{-1/2}|C^*(L)v_{[nr]}| > \epsilon) \to 0.$$

We can show that $\{C^*(L)v_t\}$ is uniformly square integrable because Assumptions 1-2 imply

$$\sup_t \|C^*(L)v_t\|_q \leq \sum_{j=0}^{\infty} |C_j^*| \ \sup_t \|v_t\|_q \leq \sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty} |C_k| \ \sup_t \|v_t\|_q = \sum_{k=1}^{\infty} k|C_k| \ \sup_t \|v_t\|_q < \infty$$

for some $q > 2$.

Thus, we get the desired result.

Proof of Theorem 2:

The Hessian matrix $H_n(\theta)$ can be defined as

$$H_n(\theta) = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_n(\theta)}{\partial b \partial b'} & \frac{\partial^2 \mathcal{L}_n(\theta)}{\partial b \partial \theta_2'} \\ \frac{\partial^2 \mathcal{L}_n(\theta)}{\partial \theta_2 \partial b'} & \frac{\partial^2 \mathcal{L}_n(\theta)}{\partial \theta_2 \partial \theta_2'} \end{pmatrix},$$

where $b = \text{vec}(\boldsymbol{\beta})$.

We denote $D_{1n} = \text{diag}(n, \ldots, n)$ and $D_{2n} = \text{diag}(\sqrt{n}, \ldots, \sqrt{n})$, which correspond to the parameter vectors $\text{vec}(\boldsymbol{\beta})$ and θ_2, respectively. Define a diagonal matrix $D_n = \text{diag}(D_{1n}, D_{2n})$.

From the representation theorem, $\Delta x_t = C(L)v_t$, $w_t = \beta' C^*(L)v_t$, and $z_t = D_1(L)u_t + D_2(L)e_t$. Assumptions 1-2 imply that $\sup_t \|\Delta x_t\|_q < \infty$, $\sup_t \|w_t\|_q < \infty$, and $\sup_t \|z_t\|_q < \infty$ for some $q > 2$.

First, we show that the normalized Hessian matrix $D_n^{-1} H_n(\theta) D_n^{-1}$ is asymptotically block-diagonal, which can be implied by $n^{-1}\sum_{t=1}^n x_{2t-1}s_t' = O_p(1)$. This result has been proved by Phillips (1988) for the linear process and by Hansen (1992) for the strong mixing process.

Thus, $n^{-1}\sum_{t=1}^n x_{2t-1}s_t' = O_p(1)$ and $\frac{1}{n^{3/2}}\frac{\partial^2 \mathcal{L}_n(\theta)}{\partial b \partial \theta_2'} \to^p 0$, and therefore the normalized Hessian matrix is block-diagonal.

Next, we get the following asymptotic results:

$$\begin{aligned}
-\frac{1}{n^2}\frac{\partial^2 \mathcal{L}_n(\theta)}{\partial b \partial b'} &= \alpha' \Sigma^{-1}\alpha \otimes \frac{1}{n^2}\sum_{t=1}^n x_{2t-1}x_{2t-1}' \\
&\Rightarrow \alpha' \Sigma^{-1}\alpha \otimes C_2 \int VV' C_2' \\
\frac{1}{n}\sum_{t=1}^n x_{2t-1}u_t' &\Rightarrow \int C_2 V dU'.
\end{aligned}$$

Therefore, we can show that

$$\begin{aligned}
n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (\frac{1}{n^2}\sum_{t=1}^n x_{2t-1}x_{2t-1}')^{-1}\frac{1}{n}\sum_{t=1}^n x_{2t-1}u_t' \Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1} + o_p(1) \\
&\Rightarrow (\int C_2 VV' C_2)^{-1}\int C_2 V dU' \Sigma^{-1}\alpha(\alpha'\Sigma^{-1}\alpha)^{-1} \\
&= A(\int B_2 B_2')^{-1}\int B_2 dB_1'(\alpha'\Sigma^{-1}\alpha)^{-1/2'},
\end{aligned}$$

where $A = [(\alpha_\perp' \Omega_{vv}\alpha_\perp)^{1/2'}(\alpha_\perp' \Pi^*(1)\beta_\perp)^{-1'}\beta_{2\perp}']^{-1}$.

<u>Proof of Lemma 3</u>:

First, we show $n^{-1/2}R_{1t} \Rightarrow C(1)V(r)$, where $R_{1t} = x_{t-1} - \sum_{t=1}^n x_{t-1}s_t'(\sum_{t=1}^n s_t s_t')^{-1}s_t$. We need to show

$$\sup_{r\in[0,1]} |n^{-1/2}\sum_{t=1}^n x_{t-1}s_t'(\sum_{t=1}^n s_t s_t')^{-1}s_{[nr]}| \to^p 0.$$

Since s_t is uniformly square integrable, we can appeal to Hall and Heyde (1980), p. 143, to show that

$$sup_{t\leq n} n^{-1/2}|s_t| \to^p 0.$$

Because $n^{-1}\sum_{t=1}^n x_{2t-1}s_t' = O_p(1)$ and $n^{-1}\sum_{t=1}^n s_t s_t' \to^p E(s_t s_t')$,

$$sup_{r\in[0,1]}|n^{-1}\sum_{t=1}^n x_{2t-1}s_t'(n^{-1}\sum_{t=1}^n s_t s_t')^{-1}n^{-1/2}s_{[nr]}| \to^p 0,$$

which implies $n^{-1/2}R_{1t} \Rightarrow C(1)V(r)$.

Under the null hypothesis $\mathcal{H}_0 : \Pi = 0$, we can show that

$$\begin{aligned} R_{0t} &= \Delta x_t - \sum_{t=1}^{n} \Delta x_t s_t^{'} (\sum_{t=1}^{n} s_t s_t^{'})^{-1} s_t \\ &= u_t + o_p(\sqrt{n}). \end{aligned}$$

Therefore,

$$\begin{aligned} S_{00} &= \frac{1}{n}\sum_{t=1}^{n} R_{0t}R_{0t}^{'} \rightarrow^p \Sigma \\ S_{10} &= \frac{1}{n}\sum_{t=1}^{n} R_{1t}R_{0t}^{'} \Rightarrow C\int VdU^{'} \\ \frac{1}{n}S_{11} &= \frac{1}{n^2}\sum_{t=1}^{n} R_{1t}R_{1t}^{'} \Rightarrow C\int VV^{'}C^{'}. \end{aligned}$$

Also, $\hat{\Sigma} \rightarrow^p \Sigma$ because

$$\hat{\Sigma} = \frac{1}{n}\sum_{t=1}^{n}\Delta x_t \Delta x_t^{'} - \frac{1}{n}\sum_{t=1}^{n}\Delta x_t S_t^{'}(\sum_{t=1}^{n} S_t S_t^{'})^{-1}\sum_{t=1}^{n} S_t \Delta x_t^{'} = \frac{1}{n}\sum_{t=1}^{n} u_t u_t^{'} + o_p(1),$$

where $S_t = (x_{t-1}^{'}, s_t^{'})^{'}$.

Proof of Theorem 3:

First, we show that $Q_n \Rightarrow (\Sigma \otimes C\int VV^{'}C^{'})$, where $Q_n = n^{-2}\sum_{t=1}^{n}(\hat{u}_t\hat{u}_t^{'} \otimes R_{1t}R_{1t}^{'})$.

Define $\epsilon_t = u_t u_t^{'} - \Sigma$. Clearly, ϵ_t is a zero mean and strong mixing process with $\sup_t E|\epsilon_t|^{q/2} < \infty$ for some $q > 6$.

Thus, Theorem 4.2 of Hansen (1992) can be used to show $n^{-3/2}\sum_{t=1}^{n}(\epsilon_t \otimes R_{1t}R_{1t}^{'}) = O_p(1)$.

$$\begin{aligned} n^{-2}\sum_{t=1}^{n}(u_t u_t^{'} \otimes R_{1t}R_{1t}^{'}) &= n^{-2}\sum_{t=1}^{n}(\Sigma \otimes R_{1t}R_{1t}^{'}) + n^{-2}\sum_{t=1}^{n}(\epsilon_t \otimes R_{1t}R_{1t}^{'}) \\ &= n^{-2}\sum_{t=1}^{n}(\Sigma \otimes R_{1t}R_{1t}^{'}) + o_p(1) \\ &\Rightarrow (\Sigma \otimes C\int VV^{'}C^{'}). \end{aligned}$$

Because $Q_n = n^{-2}\sum_{t=1}^{n}(u_t u_t^{'} \otimes R_{1t}R_{1t}^{'}) + o_p(1)$, $Q_n \Rightarrow (\Sigma \otimes C\int VV^{'}C^{'})$.

Using Lemmas 1-2, we can show that

$$\begin{aligned} \mathrm{W}_n^H &= \mathrm{vec}(S_{10})^{'} Q_n^{-1}\mathrm{vec}(S_{10}) \\ &\Rightarrow \mathrm{vec}(C\int VdU^{'})^{'}(\Sigma \otimes C\int VV^{'}C^{'})^{-1}\mathrm{vec}(C\int VdU^{'}) \\ &= \mathrm{tr}\ \Sigma^{-1/2}\int dUV^{'}C^{'}(C\int VV^{'}C^{'})^{-1}\int CVdU^{'}\Sigma^{-1/2'} \\ &= \mathrm{tr}\int dW_1W_2^{'}(\int W_2W_2^{'})^{-1}\int W_2dW_1^{'}. \end{aligned}$$

The Wald statistic, which is based on the standard covariance estimator, has the asymptotic distribution as follows:

$$\begin{aligned} \mathrm{W}_n &= \mathrm{tr}\ n\hat{\Sigma}^{-1}S_{01}S_{11}^{-1}S_{10} \\ &\Rightarrow \mathrm{tr}\ \Sigma^{-1/2}\int dUV'C'(C\int VV'C')^{-1}\int CVdU'\Sigma^{-1/2'} \\ &= \mathrm{tr}\int dW_1W_2'(\int W_2W_2')^{-1}\int W_2dW_1'. \end{aligned}$$

The LR statistic also has the same distribution. First, define $\hat{\mu}_i = n\hat{\lambda}_i$ for $i = 1, 2, \ldots, p$, so that $\hat{\mu}_i$ satisfies the eigenvalue equation

$$\left|\hat{\mu}_i n^{-1}S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0.$$

Thus, $\hat{\mu}_i$ converges weakly to the solution to the equation

$$\mid \mu_i C\int VV'C' - C\int VdU'\Sigma^{-1}\int dUV'C' \mid = 0.$$

Since $C = \Pi^*(1)^{-1}$ is nonsingular, this can be written as

$$\mid \mu_i \int W_2W_2' - \int W_2dW_1'\int dW_1W_2' \mid = 0.$$

Hence, $\mathrm{LR}_n = \sum_{i=1}^{p} n\hat{\lambda}_i = \sum_{i=1}^{p}\hat{\mu}_i \Rightarrow \sum_{i=1}^{p}\mu_i = \mathrm{LR}$.

Proof of Lemma 4:

Show that $n^{-1}\sum_{t=1}^{n} x_{t-1}v_t' \Rightarrow C(1)\int_0^1 VdV' + M$, where $M = C(1)\Lambda + E((C^*(L)v_{t-1})v_t')$ and $\Lambda = \sum_{k=1}^{\infty} E(v_tv_{t+k}')$.

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} x_{t-1}v_t' &= n^{-1}\sum_{t=1}^{n} C(1)V_{t-1}v_t' + n^{-1}\sum_{t=1}^{n} C^*(L)v_{t-1}v_t' \\ &\Rightarrow C(1)(\int_0^1 VdV' + \Lambda) + E((C^*(L)v_{t-1})v_t'), \end{aligned}$$

where $V_t = \sum_{i=1}^{t} v_i$.

Set $\xi_{t,k} = \sum_{j=1}^{k}\Delta x_{t-j}$ and $J_k = E(\Delta x_t v_{t+k}')$.

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} x_{t-1}\nu_t' &= n^{-1}\sum_{t=1}^{n} x_{t-1}(H(L)v_t)' \\ &= \sum_{k=0}^{\infty} n^{-1}\sum_{t=1}^{n} V_{t-k-1}v_{t-k}'H_k' + \sum_{k=1}^{\infty} n^{-1}\sum_{t=1}^{n}(\xi_{t,k}v_{t-k}'H_k') \\ &\Rightarrow C(1)\int_0^1 VdV'H'(1) + \Upsilon, \end{aligned}$$

where $\Upsilon = MH^{'}(1) + \sum_{k=0}^{\infty} J_k \sum_{i=k}^{\infty} H_i^{'}$.

Proof of Theorem 4:

The estimator $\tilde{\theta} = (\text{vec}(\tilde{\boldsymbol{\beta}})^{'}, \text{vec}(\tilde{\alpha})^{'}, \text{vec}(\tilde{\Gamma})^{'}, \text{vec}(\tilde{\Sigma}_{\nu\nu})^{'})^{'}$ satisfies the first order condition $\frac{\partial \tilde{\mathcal{L}}_n(\tilde{\theta})}{\partial \theta} = 0$, where

$$\begin{aligned}
\frac{\partial \tilde{\mathcal{L}}_n(\theta)}{\partial \boldsymbol{\beta}} &= \sum_{t=1}^{n} x_{2t-1}\nu_t^{'}\Sigma_{\nu\nu}^{-1}\alpha, \\
\frac{\partial \tilde{\mathcal{L}}_n(\theta)}{\partial \alpha^{'}} &= \sum_{t=1}^{n} w_{t-1}\nu_t^{'}\Sigma_{\nu\nu}^{-1}, \\
\frac{\partial \tilde{\mathcal{L}}_n(\theta)}{\partial \Gamma^{'}} &= \sum_{t=1}^{n} s_{1t}\nu_t^{'}\Sigma_{\nu\nu}^{-1}, \text{ and} \\
\frac{\partial \tilde{\mathcal{L}}_n(\theta)}{\partial \Sigma_{\nu\nu}^{-1}} &= \frac{n}{2}\Sigma_{\nu\nu} - \frac{1}{2}\sum_{t=1}^{n} \nu_t\nu_t^{'}.
\end{aligned}$$

Because $n^{-1}\sum_{t=1}^{n} x_{2t-1}w_{t-1}^{'} = O_p(1)$ and $n^{-1}\sum_{t=1}^{n} x_{2t-1}s_{1t}^{'} = O_p(1)$, the normalized Hessian matrix of (19) is asymptotically block-diagonal, and thus under Assumptions 1-3 and $\mathcal{H}_0 : \Pi = \alpha\beta^{'}$, we get the following result.

$$\begin{aligned}
n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= (n^{-2}\sum_{t=1}^{n} x_{2t-1}x_{2t-1}^{'})^{-1} n^{-1}\sum_{t=1}^{n} x_{2t-1}\nu_t^{'}\Sigma_{\nu\nu}^{-1}\alpha(\alpha^{'}\Sigma_{\nu\nu}^{-1}\alpha)^{-1} + o_p(1) \\
&\Rightarrow (C_2\int VV^{'}C_2^{'})^{-1}(\int C_2 V dV^{'}H(1)^{'} + \Upsilon)\Sigma_{\nu\nu}^{-1}\alpha(\alpha^{'}\Sigma_{\nu\nu}^{-1}\alpha)^{-1} \\
&= A(\int B_2 B_2^{'})^{-1}(\int B_2 d\tilde{B}_2^{'} + \Delta)M_1^{1/2^{'}}M_2^{-1}
\end{aligned}$$

where $\Delta = \Omega_{vv}^{-1/2}C_2^{-1}\Upsilon\Sigma_{\nu\nu}^{-1}\alpha M_1^{-1/2^{'}}$, $M_1 = \alpha^{'}\Sigma_{\nu\nu}^{-1}\Omega_{\nu\nu}\Sigma_{\nu\nu}^{-1}\alpha$, and $M_2 = \alpha^{'}\Sigma_{\nu\nu}^{-1}\alpha$.

Proof of Theorem 5:

We define $\tilde{R}_{1t} = x_{t-1} - \sum_{t=1}^{n} x_{t-1}s_{1t}^{'}(\sum_{t=1}^{n} s_{1t}s_{1t}^{'})^{-1}s_{1t}$. As in Lemma 3, we can show that $n^{-1/2}\tilde{R}_{1t} \Rightarrow CV(r)$.

Under the null hypothesis $\mathcal{H}_0 : \Pi = 0$,

$$\begin{aligned}
\tilde{R}_{0t} &= \Delta x_t - \sum_{t=1}^{n} \Delta x_t s_{1t}^{'}(\sum_{t=1}^{n} s_{1t}s_{1t}^{'})^{-1}s_{1t} \\
&= \nu_t + o_p(\sqrt{n}).
\end{aligned}$$

Thus, we get the following asymptotic results:

$$\begin{aligned}
\tilde{S}_{00} &= \frac{1}{n}\sum_{t=1}^{n} \tilde{R}_{0t}\tilde{R}_{0t}^{'} \to^p \Sigma_{\nu\nu} \\
\tilde{S}_{10} &= \frac{1}{n}\sum_{t=1}^{n} \tilde{R}_{1t}\tilde{R}_{0t}^{'} \Rightarrow C\int V dV^{'}H(1)^{'} + \Upsilon \\
\frac{1}{n}\tilde{S}_{11} &= \frac{1}{n^2}\sum_{t=1}^{n} \tilde{R}_{1t}\tilde{R}_{1t}^{'} \Rightarrow C\int VV^{'}C^{'}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\tilde{\mathrm{W}}_n &= \mathrm{tr}\ n\tilde{\Sigma}_{\nu\nu}^{-1}\tilde{S}_{01}\tilde{S}_{11}^{-1}\tilde{S}_{10} \\
&\Rightarrow \mathrm{tr}\ \Sigma_{\nu\nu}^{-1}(H(1)\int dVV^{'}C^{'}+\Upsilon^{'})(C\int VV^{'}C^{'})^{-1}(\int CVdV^{'}H^{'}(1)+\Upsilon) \\
&= \mathrm{tr}\ (\int dW_2W_2^{'}+J^{'})(\int W_2W_2^{'})^{-1}(\int W_2dW_2^{'}+J)\ K,
\end{aligned}$$

where $J=\Omega_{vv}^{-1}C^{-1}\Upsilon H(1)^{-1^{'}}\Omega_{vv}^{-1/2^{'}}$ and $K=\Omega_{vv}^{1/2^{'}}H(1)^{'}\Sigma_{\nu\nu}^{-1}H(1)\Omega_{vv}^{1/2}$.

Because $\tilde{S}_{00}\rightarrow^{p}\Sigma_{\nu\nu}$, the LR statistic has the same asymptotic distribution.

Table 1. Distribution of Cointegrating Vector Estimators

Parameters		Estimator		T-statistic				Coverage Rates	
(ϕ_1, ϕ_2, ψ)		RMSE	MAE	Mean	S.D.	Skewness	Kurtosis	5%	95%
(0.0, 0.00, 0.00)	$\hat{\beta}$	0.0138	0.0102	-0.0108	1.0620	0.0115	3.1390	0.0605	0.0588
	$\tilde{\beta}$	0.0137	0.0101	-0.0107	1.0556	0.0074	3.1591	0.0592	0.0571
(0.0, 0.25, 0.25)	$\hat{\beta}$	0.0131	0.0096	-0.0121	1.0635	-0.0077	3.1719	0.0618	0.0582
	$\tilde{\beta}$	0.0137	0.0101	-0.0113	1.0620	-0.0025	3.1655	0.0608	0.0584
(0.0, 0.25, 0.50)	$\hat{\beta}$	0.0114	0.0084	-0.0082	1.0566	-0.0161	3.1225	0.0593	0.0586
	$\tilde{\beta}$	0.0137	0.0102	-0.0060	1.0691	0.0047	3.1248	0.0594	0.0602
(0.0, 0.50, 0.25)	$\hat{\beta}$	0.0123	0.0091	-0.0104	1.0639	-0.0160	3.1277	0.0606	0.0584
	$\tilde{\beta}$	0.0137	0.0102	-0.0090	1.0859	0.0001	3.1239	0.0638	0.0619
(0.0, 0.50, 0.50)	$\hat{\beta}$	0.0097	0.0072	-0.0019	1.0538	-0.0105	3.0644	0.0578	0.0588
	$\tilde{\beta}$	0.0138	0.0101	0.0024	1.1210	0.0053	3.1201	0.0678	0.0709
(0.0, 0.75, 0.25)	$\hat{\beta}$	0.0100	0.0074	-0.0012	1.0613	-0.0108	3.0446	0.0606	0.0627
	$\tilde{\beta}$	0.0137	0.0101	0.0026	1.2283	0.0060	3.1177	0.0870	0.0884
(0.0, 0.75, 0.50)	$\hat{\beta}$	0.0063	0.0047	0.0070	1.0478	-0.0127	3.0230	0.0564	0.0580
	$\tilde{\beta}$	0.0137	0.0101	0.0144	1.3497	-0.0113	3.1750	0.1057	0.1112
(0.0, 0.95, 0.25)	$\hat{\beta}$	0.0037	0.0027	0.0083	1.0531	-0.0047	3.0429	0.0593	0.0582
	$\tilde{\beta}$	0.0140	0.0103	0.0458	2.5884	-0.0089	3.3045	0.2485	0.2631
(0.0, 0.95, 0.50)	$\hat{\beta}$	0.0019	0.0014	0.0134	1.0436	-0.0068	3.0386	0.0558	0.0565
	$\tilde{\beta}$	0.0165	0.0108	0.0603	2.6066	-0.0308	3.7403	0.2439	0.2602
(0.2, 0.00, 0.25)	$\hat{\beta}$	0.0127	0.0093	-0.0136	1.0606	-0.0064	3.1927	0.0623	0.0576
	$\tilde{\beta}$	0.0131	0.0096	-0.0129	1.0563	-0.0052	3.1897	0.0607	0.0579
(0.2, 0.00, 0.50)	$\hat{\beta}$	0.0110	0.0081	-0.0131	1.0550	-0.0226	3.1431	0.0599	0.0564
	$\tilde{\beta}$	0.0123	0.0091	-0.0117	1.0551	-0.0057	3.1166	0.0588	0.0565
(0.2, 0.25, 0.25)	$\hat{\beta}$	0.0122	0.0090	-0.0140	1.0622	-0.0131	3.1669	0.0619	0.0577
	$\tilde{\beta}$	0.0128	0.0095	-0.0371	1.0534	-0.0093	3.1613	0.0615	0.0548
(0.2, 0.25, 0.50)	$\hat{\beta}$	0.0099	0.0073	-0.0119	1.0547	-0.0253	3.1139	0.0590	0.0576
	$\tilde{\beta}$	0.0120	0.0088	-0.0513	1.0518	-0.0086	3.1153	0.0628	0.0526
(0.2, 0.50, 0.25)	$\hat{\beta}$	0.0111	0.0082	-0.0135	1.0628	-0.0230	3.1212	0.0611	0.0581
	$\tilde{\beta}$	0.0125	0.0092	-0.0859	1.0702	-0.0114	3.1184	0.0726	0.0522
(0.2, 0.50, 0.50)	$\hat{\beta}$	0.0078	0.0058	-0.0078	1.0518	-0.0217	3.0560	0.0586	0.0586
	$\tilde{\beta}$	0.0113	0.0083	-0.1357	1.0866	-0.0201	3.1162	0.0791	0.0494
(0.2, 0.75, 0.25)	$\hat{\beta}$	0.0081	0.0060	-0.0073	1.0600	-0.0185	3.0280	0.0605	0.0618
	$\tilde{\beta}$	0.0116	0.0085	-0.2315	1.2094	-0.0155	3.1099	0.1166	0.0592
(0.2, 0.75, 0.50)	$\hat{\beta}$	0.0039	0.0029	-0.0055	1.0443	-0.0234	3.0086	0.0573	0.0574
	$\tilde{\beta}$	0.0096	0.0070	-0.4324	1.2860	-0.0462	3.1891	0.1658	0.0516

Table 2. Size of Cointegration Tests

Parameters	Tests	With Covariates						Without Covariates	
		$W(R)$		$W(I)$		Bootstrap		$W(I)$	
(ϕ_1, ϕ_2, ψ)		5%	10%	5%	10%	5%	10%	5%	10%
(0.0, 0.00, 0.00)	Wald	0.059	0.118	0.058	0.115	0.056	0.116	0.059	0.113
	LR	0.050	0.104	0.048	0.105	0.058	0.114	0.053	0.103
(0.0, 0.25, 0.25)	Wald	0.056	0.119	0.048	0.098	0.065	0.110	0.063	0.127
	LR	0.048	0.108	0.039	0.088	0.064	0.110	0.055	0.114
(0.0, 0.25, 0.50)	Wald	0.058	0.108	0.037	0.068	0.058	0.104	0.060	0.114
	LR	0.051	0.102	0.030	0.059	0.054	0.098	0.054	0.102
(0.0, 0.50, 0.25)	Wald	0.061	0.121	0.044	0.085	0.065	0.110	0.058	0.116
	LR	0.054	0.113	0.036	0.076	0.064	0.110	0.053	0.103
(0.0, 0.50, 0.50)	Wald	0.063	0.122	0.027	0.051	0.058	0.104	0.068	0.115
	LR	0.059	0.112	0.025	0.048	0.054	0.098	0.052	0.106
(0.0, 0.75, 0.25)	Wald	0.067	0.123	0.033	0.052	0.065	0.110	0.067	0.107
	LR	0.062	0.112	0.030	0.048	0.064	0.110	0.058	0.103
(0.0, 0.75, 0.50)	Wald	0.064	0.114	0.010	0.028	0.058	0.104	0.071	0.113
	LR	0.061	0.107	0.009	0.023	0.054	0.098	0.066	0.104
(0.0, 0.95, 0.25)	Wald	0.064	0.118	0.005	0.014	0.065	0.110	0.460	0.513
	LR	0.061	0.114	0.004	0.012	0.064	0.110	0.443	0.506
(0.0, 0.95, 0.50)	Wald	0.067	0.114	0.002	0.006	0.058	0.104	0.402	0.467
	LR	0.061	0.112	0.002	0.006	0.054	0.098	0.391	0.589
(0.2, 0.25, 0.25)	Wald	0.059	0.120	0.048	0.097	0.065	0.110	0.061	0.120
	LR	0.046	0.108	0.041	0.089	0.064	0.110	0.053	0.105
(0.2, 0.25, 0.50)	Wald	0.059	0.109	0.034	0.069	0.053	0.107	0.058	0.104
	LR	0.052	0.102	0.029	0.061	0.054	0.107	0.054	0.098
(0.2, 0.50, 0.25)	Wald	0.064	0.123	0.045	0.085	0.060	0.113	0.056	0.108
	LR	0.058	0.111	0.037	0.075	0.059	0.113	0.051	0.094
(0.2, 0.50, 0.50)	Wald	0.064	0.119	0.026	0.053	0.057	0.104	0.065	0.107
	LR	0.059	0.112	0.025	0.048	0.057	0.100	0.059	0.102
(0.2, 0.75, 0.25)	Wald	0.076	0.132	0.037	0.061	0.053	0.101	0.091	0.137
	LR	0.070	0.118	0.032	0.056	0.052	0.100	0.083	0.128
(0.2, 0.75, 0.50)	Wald	0.069	0.119	0.011	0.025	0.051	0.091	0.139	0.196
	LR	0.065	0.111	0.010	0.022	0.050	0.092	0.126	0.185

Table 3. Power of Cointegration Tests

Parameters	Tests	With Covariates						Without Covariates		
		$W(R)$			Bootstrap			$W(I)$		
(ϕ_1, ϕ_2, ψ)	δ	5	10	15	5	10	15	5	10	15
(0.0, 0.00, 0.00)	Wald	0.174	0.445	0.703	0.136	0.348	0.636	0.162	0.438	0.693
	LR	0.154	0.421	0.676	0.136	0.344	0.634	0.144	0.413	0.668
(0.0, 0.25, 0.25)	Wald	0.202	0.543	0.763	0.151	0.415	0.713	0.169	0.473	0.705
	LR	0.185	0.502	0.718	0.154	0.412	0.713	0.147	0.421	0.649
(0.0, 0.25, 0.50)	Wald	0.293	0.712	0.893	0.227	0.631	0.880	0.175	0.452	0.662
	LR	0.274	0.694	0.880	0.225	0.625	0.880	0.160	0.430	0.643
(0.0, 0.50, 0.25)	Wald	0.259	0.615	0.826	0.175	0.489	0.786	0.162	0.449	0.661
	LR	0.231	0.583	0.802	0.175	0.485	0.785	0.151	0.421	0.630
(0.0, 0.50, 0.50)	Wald	0.431	0.864	0.981	0.374	0.814	0.966	0.168	0.421	0.625
	LR	0.413	0.851	0.975	0.369	0.814	0.966	0.151	0.391	0.600
(0.0, 0.75, 0.25)	Wald	0.428	0.848	0.966	0.335	0.750	0.946	0.177	0.423	0.611
	LR	0.408	0.828	0.956	0.331	0.751	0.945	0.169	0.402	0.574
(0.0, 0.75, 0.50)	Wald	0.819	0.991	1.000	0.774	0.987	0.998	0.210	0.451	0.594
	LR	0.811	0.991	1.000	0.775	0.986	0.998	0.200	0.429	0.565
(0.0, 0.95, 0.25)	Wald	0.973	1.000	1.000	0.952	0.998	1.000	0.540	0.717	0.771
	LR	0.970	1.000	1.000	0.953	0.998	1.000	0.524	0.708	0.756
(0.0, 0.95, 0.50)	Wald	1.000	1.000	1.000	1.000	1.000	1.000	0.523	0.677	0.708
	LR	1.000	1.000	1.000	1.000	1.000	1.000	0.511	0.662	0.693
(0.2, 0.25, 0.25)	Wald	0.216	0.571	0.798	0.171	0.467	0.767	0.180	0.492	0.732
	LR	0.194	0.547	0.777	0.174	0.466	0.762	0.160	0.455	0.698
(0.2, 0.25, 0.50)	Wald	0.347	0.790	0.950	0.280	0.726	0.925	0.199	0.528	0.760
	LR	0.321	0.773	0.934	0.277	0.725	0.924	0.182	0.505	0.733
(0.2, 0.50, 0.25)	Wald	0.265	0.650	0.862	0.218	0.572	0.843	0.175	0.489	0.713
	LR	0.246	0.613	0.840	0.221	0.568	0.841	0.166	0.465	0.685
(0.2, 0.50, 0.50)	Wald	0.552	0.933	0.993	0.473	0.907	0.987	0.205	0.536	0.736
	LR	0.528	0.924	0.991	0.468	0.904	0.988	0.193	0.503	0.711
(0.2, 0.75, 0.25)	Wald	0.470	0.892	0.981	0.427	0.856	0.973	0.250	0.550	0.720
	LR	0.443	0.877	0.977	0.424	0.852	0.972	0.241	0.532	0.699
(0.2, 0.75, 0.50)	Wald	0.955	0.999	1.000	0.923	0.998	1.000	0.393	0.666	0.795
	LR	0.949	0.999	1.000	0.922	0.998	1.000	0.382	0.650	0.776

Table 4. Money Demand Equation: 1960Q1-1999Q4

	$m_t - p_t$		y_t		r_t	
	With Stationary Covariates					
$\hat{\beta}$	1		-1.3903	(0.2345)	0.0124	(0.0058)
$\hat{\alpha}$	-0.0593	(0.0124)	0.0115	(0.0166)	-0.0927	(1.5100)
R^2	0.6989		0.4178		0.4645	
Log-likelihood	590.77		553.56		-141.39	
$Q(1)$	0.0563	[0.812]	0.0677	[0.795]	0.1247	[0.724]
$Q(12)$	9.5024	[0.660]	7.3507	[0.834]	15.791	[0.201]
ARCH LM(1)	0.6818	[0.410]	0.0003	[0.986]	0.0320	[0.858]
ARCH LM(12)	0.3911	[0.965]	0.6055	[0.834]	2.1492	[0.018]
	Without Stationary Covariates					
$\tilde{\beta}$	1		-1.1188	(0.3166)	-0.0144	(0.0042)
$\tilde{\alpha}$	-0.0411	(0.0136)	0.0300	(0.0129)	2.7507	(1.4304)
R^2	0.6552		0.3725		0.3625	
Log-likelihood	580.36		547.79		-154.81	
$Q(1)$	0.0087	[0.926]	0.0083	[0.927]	0.0006	[0.981]
$Q(12)$	12.042	[0.442]	7.1259	[0.849]	20.272	[0.062]
ARCH LM(1)	0.7832	[0.378]	0.0177	[0.894]	4.6330	[0.033]
ARCH LM(12)	0.8926	[0.556]	0.7394	[0.711]	10.225	[0.000]

The standard errors are in the parentheses, and the p-values are in the square brackets.

Table 5. Cointegration Tests: 1960Q1-1999Q4

	W_n^H	W_n	LR_n	$\tilde{\mathrm{W}}_n$	$\tilde{\mathrm{LR}}_n$
Statistics	46.795	42.733	39.515	36.376	34.609
Asymptotic p-value	0.002	0.003	0.010	0.108	0.153
Bootstrap p-value	0.072	0.089	0.095	0.295	0.290

The canonical correlation R is estimated at $(0.247, 0.973, 1.000)$.