Econometric Theory, **11**, 1995, 984–1014. Printed in the United States of America.

TESTING FOR COINTEGRATION WHEN SOME OF THE COINTEGRATING VECTORS ARE PRESPECIFIED

MICHAEL T.K. HORVATH
Stanford University
MARK W. WATSON
Princeton University

Many economic models imply that ratios, simple differences, or "spreads" of variables are I(0). In these models, cointegrating vectors are composed of 1's, 0's, and −1's and contain no unknown parameters. In this paper, we develop tests for cointegration that can be applied when some of the cointegrating vectors are prespecified under the null or under the alternative hypotheses. These tests are constructed in a vector error correction model and are motivated as Wald tests from a Gaussian version of the model. When all of the cointegrating vectors are prespecified under the alternative, the tests correspond to the standard Wald tests for the inclusion of error correction terms in the VAR. Modifications of this basic test are developed when a subset of the cointegrating vectors contain unknown parameters. The asymptotic null distributions of the statistics are derived, critical values are determined, and the local power properties of the test are studied. Finally, the test is applied to data on foreign exchange future and spot prices to test the stability of the forward–spot premium.

1. INTRODUCTION

Economic models often imply that variables are cointegrated with simple and known cointegrating vectors. Examples include the neoclassical growth model, which implies that income, consumption, investment, and the capital stock will grow in a balanced way, so that any stochastic growth in one of the series must be matched by corresponding growth in the others. Asset

This paper has benefited from helpful comments by Neil Ericsson, Gordon Kemp, Andrew Levin, Soren Johansen, John McDermott, Pierre Perron, Peter Phillips, James Stock, and three referees. This research was supported by grants from the National Science Foundation (SES-91-22463 and SBR-94-09629) and the Sloan Foundation. Address correspondence to: Professor Mark Watson, Woodrow Wilson School, Princeton University, Princeton, NJ 08540, USA.

  0266-4666/95 $9.00 + .10

pricing models with stable risk premia imply corresponding stable differences in spot and forward prices, long- and short-term interest rates, and the logarithms of stock prices and dividends. Most theories of international trade imply long-run purchasing power parity, so that long-run movements in nominal exchange rates are matched by countries' relative price levels. Certain monetarist propositions are centered around the stability of velocity, implying cointegration among the logarithms of money, prices, and income. Each of these theories has two distinct implications for the properties of economic time series under study: first, the series are cointegrated, and second, the cointegrating vector takes on a specific value. For example, balanced growth implies that the logarithms of income and consumption are cointegrated and that the cointegrating vector takes on the value of $(1 \ -1)$.

The most widely used approach to testing these cointegration propositions is articulated and implemented in Johansen and Juselius (1992), who investigate the empirical support for long-run purchasing power parity. They implement a two-stage testing procedure. In the first stage, the null hypothesis of no cointegration is tested against the alternative that the data are cointegrated with an unknown cointegrating vector using Johansen's (1988) test for cointegration. If the null hypothesis is rejected, a second stage test is implemented with cointegration maintained under both the null and alternative. The null hypothesis is that the data are cointegrated with the specific cointegrating vector implied by the relevant economic theory ($[1 \ -1]$ in the consumption–income example), and the alternative is that data are cointegrated with another unspecified cointegrating vector. Because a consistent test for cointegration is used in the first stage, potential cointegration in the data is found with probability approaching 1 in large samples. Thus, the probability of rejecting the cointegration constraints on the data imposed by the economic model are given by the size of the test in the second step, at least in large samples. An important strength of this procedure is that it can uncover cointegration in the data with a cointegrating vector different from the cointegrating vector imposed by the theory. The disadvantage is that the sample sizes used in economics are often relatively small, so that the first-stage tests may have low power.

This paper discusses an alternative procedure in which the null of no cointegration is tested against the composite alternative of cointegration using a prespecified cointegrating vector. This approach has two advantages. First, and most important, the resulting test for cointegration is significantly more powerful than the test that does not impose the cointegrating vector. For example, in the bivariate example analyzed in Section 3, these power gains correspond to sample size increases ranging from 40 to 70% for a test with power equal to 50%. The second advantage is that the test statistic is very easy to calculate: it is the standard Wald test for the presence of the candidate error correction terms in the first difference vector autoregression. The countervailing disadvantage of the testing approach is that it does not separate the two components of the alternative hypothesis and, thus, may fail to reject the null of no cointegration when the data are cointegrated with a cointegrating vector different from that used to construct the test. We investigate this in Section 3, where it is shown that in situations with weak cointegration (represented by a local-to-unity error correction term) even inexact information on the value of the cointegrating vector often leads to power improvements over the test that uses no information. If the null hypothesis of noncointegration is rejected, one can then determine whether the prespecified cointegrating vector differs significantly from the true cointegrating vector.

The plan of this paper is as follows. In Section 2, we consider the general problem of testing for cointegration in a model in which some of the potential cointegrating vectors are known, and some are unknown, under both the null and the alternative. In particular, we present Wald and likelihood ratio tests for the hypothesis that the data are cointegrated with r_{o_k} known and r_{o_u} unknown cointegrating vectors under the null. Under the alternative, there are r_{a_k} and r_{a_u} additional known and unknown cointegrating vectors, respectively. The tests are constructed in the context of a finite-order Gaussian vector error correction model (VECM) and generalize the procedures of Johansen (1988), who considered the hypothesis testing problem with $r_{o_k} = r_{a_k} = 0$. In Section 2, we also derive the asymptotic null distributions of the test statistics and tabulate critical values. Section 3 focuses on the power properties of the test. First, we present comparisons of the power of likelihood-based tests that do and do not use information about the value of the cointegrating vector. Next, because information about the potential cointegrating vector might be inexact, we investigate the power loss associated with using an incorrect value of the cointegrating vector. Finally, when there are no cointegrating vectors under the null and only one cointegrating vector under the alternative, simple univariate unit root tests provide an alternative to the multivariate VECM-based tests. Section 3 compares the power of these univariate unit root tests to the multivariate VECM-based tests. Section 4 contains an empirical application that investigates the forward premia in foreign exchange markets by examining the cointegration properties of forward and spot prices. Section 5 contains some concluding remarks.

2. TESTING FOR COINTEGRATION IN THE GAUSSIAN VAR MODEL

As in Johansen (1988), we derive tests for cointegration in the context of the reduced rank Gaussian VAR:

$$Y_t = d_t + X_t, \tag{2.1a}$$

$$X_t = \sum_{i=1}^{p} \Pi_i X_{t-i} + \epsilon_t, \tag{2.1b}$$

where Y_t is an $n \times 1$ data vector from a sample of size T, d_t represents deterministic drift in Y_t, X_t is an $n \times 1$ random vector generated by (2.1b), ϵ_t is NIID$(0,\Sigma_\epsilon)$, and, for convenience, the initial conditions X_{-i}, $i = 0, \ldots, p$, are assumed to equal 0. To focus attention on the long-run behavior of the process, it is useful to rewrite (2.1b) as

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta X_{t-i} + \epsilon_t, \tag{2.1c}$$

where $\Pi = -I_n + \sum_{i=1}^{p} \Pi_i$.

Our interest is focused on $r = \text{rank}(\Pi)$, and we consider tests of the hypothesis

$H_o: \text{rank}(\Pi) = r = r_o,$

$H_a: \text{rank}(\Pi) = r = r_o + r_a, \quad \text{with } r_a > 0.$

The alternative is written so that r_a represents the number of additional cointegrating vectors that are present under the alternative. We assume that $r_o = r_{o_k} + r_{o_u}$, where r_{o_k} is the number of cointegrating vectors that are known under the null and r_{o_u} represents the number of cointegrating vectors that are unknown (or, alternatively, unrestricted) under the null. Similarly, $r_a = r_{a_k} + r_{a_u}$, where the subscripts k and u denote known and unknown, respectively. The r_{a_k} prespecified vectors are thought to be cointegrating vectors under the alternative; under the null, they do not cointegrate the series. In spite of this, for expositional ease, they will be referred to as cointegrating vectors.

As in Engle and Granger (1987), Johansen (1988), and Ahn and Reinsel (1990), it is convenient to write the model in vector error correction form by factoring the matrix Π as $\Pi = \delta\alpha'$, where δ and α are $n \times r$ matrices of full column rank and the columns of α denote the cointegrating vectors. The columns of α are partitioned as $\alpha = (\alpha_o \, \alpha_a)$, where α_o is an $n \times r_o$ matrix whose columns are the cointegrating vectors present under the null, α_a is an $n \times r_a$ matrix whose columns are the additional cointegrating vectors present under the alternative. The matrix δ is partitioned conformably as $\delta = (\delta_o \, \delta_a)$, where δ_o is $n \times r_o$ and δ_a is $n \times r_a$. It is also useful to partition α_a to isolate the known and unknown cointegrating vectors. Thus, $\alpha_a = (\alpha_{a_k} \, \alpha_{a_u})$, where the r_{a_k} columns of α_{a_k} are the additional cointegrating vectors known under the alternative, and the r_{a_u} columns of α_{a_u} are the additional cointegrating vectors that are present but unrestricted under the alternative. The matrix δ_a is partitioned conformably as $\delta_a = (\delta_{a_k} \, \delta_{a_u})$. Using this notation, $\Pi X_{t-1} = \delta_o(\alpha_o' X_{t-1}) + \delta_a(\alpha_a' X_{t-1})$, and the competing hypotheses are $H_o: \delta_a = 0$ vs. $H_a: \delta_a \neq 0$, with $\text{rank}(\delta_a \alpha_a') = r_a$.

We develop tests for H_o vs. H_a in three steps. First, we abstract from deterministic components and derive the likelihood ratio statistic and some useful asymptotically equivalent statistics under the maintained assumption that $d_t = 0$. Second, we discuss how these statistics can be modified for nonzero values of d_t. Finally, the asymptotic null distributions of the resulting statistics are derived and critical values based on these asymptotic distributions are tabulated.

2.1. Calculating the LR and Wald Test Statistics When $d_t = 0$

The likelihood ratio statistic for testing $H_o: r = r_{o_k} + r_{o_u}$ vs. $H_a: r = r_{o_k} + r_{a_k} + r_{o_u} + r_{a_u}$ will depend on $r_{o_k}, r_{a_k}, r_{o_u}, r_{a_u}$, and the values of α_{o_k} and α_{a_k}. We write the statistic as $LR_{r_o,r_a}(\alpha_{o_k},\alpha_{a_k})$. The values of r_{o_k} and r_{a_k} appear implicitly as the ranks of α_{o_k} and α_{a_k}, respectively. When $r_{o_k} = 0$, the statistic is written as $LR_{r_o,r_a}(0,\alpha_{a_k})$ and as $LR_{r_o,r_a}(\alpha_{o_k},0)$ when $r_{a_k} = 0$.

To derive the LR statistic, we limit attention to the problem with $r_o = r_{o_k} = r_{o_u} = 0$. For the purposes of deriving the computational formula for the LR statistic, this is without loss of generality because, in the general case, the LR statistic is identically

$$LR_{r_o,r_a}(\alpha_{o_k},\alpha_{a_k}) \equiv LR_{0,r_o+r_a}(0,[\alpha_{o_k}\alpha_{a_k}]) - LR_{0,r_o}(0,\alpha_{o_k}). \tag{2.2}$$

With $r_o = 0$, and ignoring the deterministic components, d_t, the model can be written as

$$\Delta Y_t = \delta_{a_k}(\alpha'_{a_k}Y_{t-1}) + \delta_{a_u}(\alpha'_{a_u}Y_{t-1}) + \beta Z_t + \epsilon_t, \tag{2.3}$$

where $\beta = (\Phi_1\Phi_2 \cdots \Phi_{p-1})$ and $Z_t = (\Delta Y'_{t-1}\Delta Y'_{t-2} \cdots \Delta Y'_{t-p+1})'$. In the context of (2.3), the null hypothesis $H_o: r = 0$ can be written as the composite null $H_o: \delta_{a_k} = 0$, $\delta_{a_u} = 0$.[1] It is convenient to discuss each part of this null separately: we first consider testing $\delta_{a_k} = 0$ maintaining $\delta_{a_u} = 0$, then the converse, and finally the joint hypothesis.

The test statistic for $H_o: r = 0$ vs. $H_a: r = r_{a_k}$. When $r_{a_u} = 0$, equation (2.3) simplifies to

$$\Delta Y_t = \delta_{a_k}(\alpha'_{a_k}Y_{t-1}) + \beta Z_t + \epsilon_t. \tag{2.4}$$

Because $\alpha'_{a_k}Y_{t-1}$ does not depend on unknown parameters, (2.4) is a standard multivariate linear regression, so that the LR, Wald, and LM statistics have their standard regression form. Letting $Y = [Y_1Y_2 \cdots Y_T]'$, $Y_{-1} = [Y_0Y_1 \cdots Y_{T-1}]'$, $\Delta Y = Y - Y_{-1}$, $Z = [Z_1Z_2 \cdots Z_T]'$, $\epsilon = [\epsilon_1\epsilon_2 \cdots \epsilon_T]'$, and $M_Z = [I - Z(Z'Z)^{-1}Z']$, the ordinary least-squares (OLS) estimator of δ_{a_k} is $\hat{\delta}_{a_k} = (\Delta Y'M_ZY_{-1}\alpha_{a_k})(\alpha'_{a_k}Y'_{-1}M_ZY_{-1}\alpha_{a_k})^{-1}$, which corresponds to the Gaussian maximum likelihood estimate (MLE). The corresponding Wald test statistic for H_o vs. H_a is

$$\begin{aligned} W &= [\text{vec}(\hat{\delta}_{a_k})]'[(\alpha'_{a_k}Y'_{-1}M_ZY_{-1}\alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon]^{-1}[\text{vec}(\hat{\delta}_{a_k})] \\ &= [\text{vec}(\Delta Y'M_ZY_{-1}\alpha_{a_k})]'[(\alpha'_{a_k}Y'_{-1}M_ZY_{-1}\alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] \\ &\quad \times [\text{vec}(\Delta Y'M_ZY_{-1}\alpha_{a_k})], \end{aligned} \tag{2.5}$$

where $\hat{\Sigma}_\epsilon$ is the usual estimator value of Σ_ϵ, that is, $\hat{\Sigma}_\epsilon = T^{-1}\hat{\epsilon}'\hat{\epsilon}$, and where $\hat{\epsilon}$ is the matrix of OLS residuals from (2.4). For values of δ_{a_k} that are T^{-1} local to $\delta_{a_k} = 0$, the LR and LM statistics are asymptotically equal to W.

The test statistic for $H_o: r = 0$ vs. $H_a: r = r_{a_u}$. The model simplifies to (2.4) with δ_{a_u} and α_{a_u} replacing δ_{a_k} and α_{a_k}. However, the analog of the Wald statistic in (2.5) cannot be calculated because the regressor $\alpha'_{a_u}Y_{t-1}$ depends on unknown parameters. However, the LR statistic can be calculated, and useful formulae for the LR statistic are developed in Anderson (1951) and Johansen (1988). Because $\delta_{a_u} = 0$ under the null hypothesis, the cointegrating vectors α_{a_u} are unidentified, and this complicates the testing problem in ways familiar from the work of Davies (1977, 1987). The problem can be avoided when $r_a = n$, because in this case Π is unrestricted under the alternative and the null and alternative become $H_o: \Pi = 0$ vs. $H_a: \Pi \neq 0$. The problem cannot be avoided when the rank$(\Pi) < n$ under the alternative. Indeed, in the standard classical reduced rank regression, the general form of the asymptotic distribution of the LR statistic has only been derived for the case in which the matrix of regression coefficients has full rank under the alternative. In this case, Anderson (1951) shows that the LR statistic has an asymptotic χ^2 null distribution. When the matrix of regression coefficients has reduced rank under the alternative, the asymptotic distribution of the LR statistic depends on the distribution of the regressors. Still, the special structure of the regressors in the cointegrated VAR allows Johansen (1988) to circumvent this problem and derive the asymptotic distribution of the LR test even when Π has reduced rank under the alternative.

As pointed out by Hansen (1990), when some parameters are unidentified under the null, the LR statistic can be interpreted as a maximized version of the Wald statistic. This interpretation is useful here because it suggests a simple way to compute the statistic. Because this form of the statistic appears as one component in the test statistic for the general $r_a = r_{a_k} + r_{a_u}$ alternative, we derive it here.

Let LR denote the likelihood ratio statistic for testing H_o versus H_a, and let $LR^*(\Sigma_\epsilon)$ denote the (infeasible) LR statistic that would be calculated if Σ_ϵ were known. As usual, $LR = LR^*(\hat{\Sigma}_\epsilon) + o_p(1)$ under H_o and local alternatives (here, T^{-1}). Let $L(\delta_{a_u}, \alpha_{a_u}, \Sigma_\epsilon)$ denote the log likelihood written as a function of δ_{a_u}, α_{a_u}, and Σ_ϵ, with β concentrated out, and let $\hat{\delta}_{a_u}(\alpha_{a_u})$ denote the MLE of δ_{a_u} for fixed α_{a_u}. Then, the well-known relation between the Wald and LR statistic in the linear model implies that

$$\begin{aligned} W(\alpha_{a_u}) &= 2[L(\hat{\delta}_{a_u}(\alpha_{a_u}), \alpha_{a_u}, \hat{\Sigma}_\epsilon) - L(0, \alpha_{a_u}, \hat{\Sigma}_\epsilon)] \\ &= 2[L(\hat{\delta}_{a_u}(\alpha_{a_u}), \alpha_{a_u}, \hat{\Sigma}_\epsilon) - L(0, 0, \hat{\Sigma}_\epsilon)], \end{aligned} \qquad \textbf{(2.6)}$$

where $W(\alpha_{a_u})$ is the Wald statistic in (2.5) written as a function of α_{a_u}; the first equality follows because each log-likelihood function is evaluated using

$\hat{\Sigma}_\epsilon$; and the second equality follows because α_{a_u} does not enter the likelihood when $\delta_{a_u} = 0$. Thus,

$$\operatorname*{Sup}_{\alpha_{a_u}} W(\alpha_{a_u}) = \operatorname*{Sup}_{\alpha_{a_u}} 2[L(\hat{\delta}_{a_u}(\alpha_{a_u}),\alpha_{a_u},\hat{\Sigma}_\epsilon) - L(0,0,\hat{\Sigma}_\epsilon)] = LR^*(\hat{\Sigma}_\epsilon),$$

where the Sup is taken over all $n \times r_{a_u}$ matrices α_{a_u}.

To calculate $\operatorname{Sup}_{\alpha_{a_u}} W(\alpha_{a_u})$, rewrite (2.5) as

$$\begin{aligned}
W(\alpha_{a_u}) &= [\operatorname{vec}(\Delta Y' M_Z Y_{-1}\alpha_{a_u})]'[(\alpha'_{a_u} Y'_{-1} M_Z Y_{-1}\alpha_{a_u})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] \\
&\qquad \times [\operatorname{vec}(\Delta Y' M_Z Y_{-1}\alpha_{a_u})] \\
&= \operatorname{TR}[\hat{\Sigma}_\epsilon^{-1/2}(\Delta Y' M_Z Y_{-1}\alpha_{a_u})(\alpha_{a_u} Y'_{-1} M_Z Y_{-1}\alpha_{a_u})^{-1} \\
&\qquad \times (\alpha'_{a_u} M_Z Y'_{-1}\Delta Y)\hat{\Sigma}_\epsilon^{-1/2\prime}] \\
&= \operatorname{TR}[\hat{\Sigma}_\epsilon^{-1/2}(\Delta Y' M_Z Y_{-1}) D D'(M_Z Y'_{-1}\Delta Y)\hat{\Sigma}_\epsilon^{-1/2\prime}], \\
&\qquad\qquad \text{where } D = \alpha_{a_u}(\alpha'_{a_u} Y'_{-1} M_Z Y_{-1}\alpha_{a_u})^{-1/2} \\
&= \operatorname{TR}[D'(Y'_{-1} M_Z \Delta Y)\hat{\Sigma}_\epsilon^{-1}(\Delta Y' M_Z Y_{-1}) D] \\
&= \operatorname{TR}[F' C C' F], \qquad\qquad \textbf{(2.7)}
\end{aligned}$$

where $F = (Y'_{-1} M_Z Y_{-1})^{1/2}\alpha_{a_u}(\alpha'_{a_u} Y'_{-1} M_Z Y_{-1}\alpha_{a_u})^{-1/2}$ and $C = (Y'_{-1} M_Z Y_{-1})^{-1/2}(Y'_{-1} M_Z \Delta Y)\hat{\Sigma}_\epsilon^{-1/2\prime}$. Notice that $F'F = I_{r_{a_u}}$ and $\operatorname{Sup}_{\alpha_{a_u}} W(\alpha_{a_u}) = \operatorname{Sup}_{F'F=I} TR[F'(CC')F]$. Let $\lambda_i(CC')$ denote the eigenvalues of (CC') ordered so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then,

$$\begin{aligned}
\operatorname*{Sup}_{\alpha_{a_u}} W(\alpha_{a_u}) &= \operatorname*{Sup}_{F'F=1} TR[F'(CC')F] = \sum_{i=1}^{r_{a_u}} \lambda_i(CC') \\
&= LR^*(\hat{\Sigma}_\epsilon) = LR + o_p(1), \qquad\qquad \textbf{(2.8)}
\end{aligned}$$

where the final equality holds under the null and local (T^{-1}) alternatives. Because $\lambda_i(CC') = \lambda_i(C'C)$, the likelihood ratio statistic can then be calculated (up to a term that vanishes in probability) as the largest r_{a_u} eigenvalues of $C'C = [\hat{\Sigma}_\epsilon^{-1/2}(\Delta Y' M_Z Y_{-1})(Y'_{-1} M_Z Y_{-1})^{-1}(Y'_{-1} M_Z \Delta Y)\hat{\Sigma}_\epsilon^{-1/2\prime}]$.

To see the relationship between the expression in (2.8) and the well-known formula for the LR statistic developed in Anderson (1951) and Johansen (1988), note that their formula can be written as $LR = -T\sum_{i=1}^{r_{a_u}} \ln[1 - \gamma_i]$, where γ_i are the ordered squared canonical correlations between ΔY_t and Y_{t-1}, after controlling for $\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}$. Because $\gamma_i = \lambda_i(S'S)$, where $S'S = (\Delta Y' M_Z \Delta Y)^{-1/2}(\Delta Y' M_Z Y_{-1})(Y'_{-1} M_Z Y_{-1})^{-1}(Y'_{-1} M_Z \Delta Y)(\Delta Y' M_Z \Delta Y)^{-1/2\prime}$ (Brillinger, 1980, Ch. 10), $LR = -T\sum_{i=1}^{r_{au}} \ln[1 - \lambda_i(S'S)] = T\sum_{i=1}^{r_{au}} \lambda_i(S'S) + o_p(1) = \sum_{i=1}^{r_{au}} \lambda_i(TS'S) + o_p(1)$. Finally, because $T(S'S) = \tilde{\Sigma}_\epsilon^{-1/2}(\Delta Y' M_Z Y_{-1}) \times (Y'_{-1} M_Z Y_{-1})^{-1}(Y'_{-1} M_Z \Delta Y)\tilde{\Sigma}_\epsilon^{-1/2\prime}$, where $\tilde{\Sigma}_\epsilon = T^{-1}(\Delta Y' M_Z \Delta Y)$, this expression is identical to (2.7), except that Σ_ϵ is estimated under the null.

The test statistic for $H_o: r = 0$ vs. $H_a: r = r_{a_k} + r_{a_u}$. The model now has the general form of (2.3). As earlier, the LR statistic can be approximated up to an $o_p(1)$ term by maximizing the Wald statistic over the unknown parameters in α_{a_u}. Let $M_{zk} = [I - (M_z Y_{-1}\alpha_{a_k})(\alpha'_{a_k} Y'_{-1} M_z Y_{-1}\alpha_{a_k})^{-1}(\alpha'_{a_k} M_z Y'_{-1})] M_z$ denote the matrix that partials both Z and $Y_{-1}\alpha_{a_k}$ out of regression (2.3). The Wald statistic (as a function of α_{a_k} and α_{a_u}) can be written as

$$\begin{aligned} W(\alpha_{a_k}, \alpha_{a_u}) = {} & [\mathrm{vec}(\Delta Y' M_z Y_{-1}\alpha_{a_k})]' [(\alpha'_{a_k} Y'_{-1} M_z Y_{-1}\alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] \\ & \times [\mathrm{vec}(\Delta Y' M_z Y_{-1}\alpha_{a_k})] \\ & + [\mathrm{vec}(\Delta Y' M_{zk} Y_{-1}\alpha_{a_u})]' [(\alpha'_{a_u} Y'_{-1} M_{zk} Y_{-1}\alpha_{a_u})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] \\ & \times [\mathrm{vec}(\Delta Y' M_{zk} Y_{-1}\alpha_{a_u})]. \end{aligned} \qquad \textbf{(2.9)}$$

The first term is identical to equation (2.5), and the second term is the same as (2.7), except that $M_z \Delta Y$ and $M_z Y_{-1}$ are replaced with $M_{zk}\Delta Y$ and $M_{zk} Y_{-1}$.

When maximizing $W(\alpha_{a_k}, \alpha_{a_u})$ over the unknown cointegrating vectors in α_{a_u}, we can restrict attention to cointegrating vectors that are linearly independent of α_{a_k}, so that the LR statistic is obtained by maximizing (2.9) over all $n \times r_{a_u}$ matrices α_{a_u} satisfying $\alpha'_{a_u}\alpha_{a_k} = 0$. Let G denote an (arbitrary) $n \times (n - r_{a_k})$ matrix whose columns span the null space of the columns of α_{a_k}. Then, α_{a_u} can be written as a linear combination of the columns of G, so that $\alpha_{a_u} = G\tilde{\alpha}_{a_u}$, where α_{a_u} is an $(n - r_{a_k}) \times r_{a_u}$ matrix, so that $\alpha'_{a_u}\alpha_{a_k} = \tilde{\alpha}'_{a_u} G'\alpha_{a_k} = 0$ for all $\tilde{\alpha}_{a_u}$. Substituting $G\tilde{\alpha}_{a_u}$ into (2.9) and carrying out the maximization yields

$$\begin{aligned} \operatorname*{Sup}_{\alpha_{au}} W(\alpha_{a_k}, \alpha_{a_u}) &= [\mathrm{vec}(\Delta Y' M_z Y_{-1}\alpha_{a_k})]' [(\alpha'_{a_k} Y'_{-1} M_z Y_{-1}\alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] \\ & \qquad \times [\mathrm{vec}(\Delta Y' M_z Y_{-1}\alpha_{a_k})] + \sum_{i=1}^{r_{au}} \lambda_i (H'H) \\ &= LR + o_p(1), \end{aligned} \qquad \textbf{(2.10)}$$

where $H'H = \hat{\Sigma}_\epsilon^{-1/2}(\Delta Y' M_{zk} Y_{-1} G)(G' Y'_{-1} M_{zk} Y_{-1} G)^{-1}(G' Y'_{-1} M_{zk}\Delta Y)\hat{\Sigma}_\epsilon^{-1/2\prime}$.

Before proceeding, we make three computational notes about (2.10). First, when $r_{a_u} = 0$, the statistic is just the standard Wald statistic testing for the presence of the error correction terms $\alpha'_{a_k} Y_{t-1}$ that is calculated by most econometric software packages. Second, any consistent estimator of Σ_ϵ can be used as $\hat{\Sigma}_\epsilon$. A particularly easy estimator, consistent under the most general hypothesis considered here, is the residual covariance matrix from the regression of Y_t onto p lagged levels of Y_t. Third, the columns of the matrix G (appearing in the definition of H) can be formed in a number of ways, for example, using the Gram–Schmidt orthogonalization procedure.

2.2. Modifications Required for the Nonzero Drift Component

When $d_t \neq 0$ in (2.1a), Y_t is not directly observed, and the procedures already outlined require modification. The necessary modification depends on the precise form of drift function. Here we assume that $d_t = \mu_0 + \mu_1 t$ and, thus, allow Y_t to have a nonzero mean and, when $\mu_1 \neq 0$, a nonzero trend. While more general drift functions are certainly possible, this formulation of d_t has proved to be adequate for most applications.[2] In this case, the VECM for y_t becomes

$$\Delta Y_t = \theta + \gamma t + \delta(\alpha' Y_{t-1}) + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \epsilon_t, \tag{2.11}$$

where $\theta = \left(I - \sum_{i=1}^{p-1} \Phi_i\right)\mu_1 - \delta\alpha'\mu_0$ and $\gamma = -\delta\alpha'\mu_1$.

There are three complications that arise when μ_0 or μ_1 are nonzero. First, as discussed in Johansen (1991, 1992a, 1992b) and Johansen and Juselius (1990), relationships between μ_0, μ_1 and the cointegrating vectors can lead to different interpretations of the drift parameters. For example, some linear combinations of μ_0 are related to initial conditions in the Y_t process, whereas others are related to means of the "error-correction" terms $\alpha' Y_t$. The second complication is that these different interpretations can imply different trend properties in the data, and this leads to changes in the asymptotic distribution of test statistics. Third, in the context of the univariate unit root model, Elliott, Rothenberg, and Stock (1995) show that different methods for detrending Y_t (associated with different estimators of μ_0 and μ_1) can lead to large differences in the power of unit root test statistics, and Elliott (1993) shows that the tests' power depends on assumptions concerning initial conditions of the process.

Rather than investigate all of the possible methods here, we present results for what are arguably the three most important cases. The first is simply the baseline case with $\mu_0 = \mu_1 = 0$; in this case, $\theta = \gamma = 0$ in (2.11). In the second case, $\mu_1 = 0$ so that the data are not "trending," but $\mu_0 \neq 0$ and is unrestricted. This is appropriate when there are no restrictions on the initial conditions of the X_t process or on the means of the error correction terms, $\alpha' Y_t$. Because $\mu_1 = 0$ in this case, then $\gamma = 0$ in (2.11); the parameter θ is nonzero but is constrained because it captures only the nonzero mean of the error correction terms $\alpha' Y_t$. Imposing the constraint leads to

$$\Delta Y_t = \delta(\alpha' Y_{t-1} - \beta) + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \epsilon_t, \tag{2.12}$$

where $\beta = \alpha'\mu_0$. In the third case, $\mu_0 \neq 0$ and is unrestricted and $\mu_1 \neq 0$ but is restricted by the requirement that $\alpha'\mu_1 = 0$; in this case, $\gamma = 0$ in (2.11) and θ is unrestricted.

2.3. Asymptotic Distribution of the Statistics

Earlier, the Gaussian likelihood ratio statistic for testing $H_o: r = r_{o_k} + r_{o_u}$ versus $H_a: r = r_{o_k} + r_{a_k} + r_{o_u} + r_{a_u}$ was defined as $LR_{r_o, r_a}(\alpha_{o_k}, \alpha_{a_k})$. Let

$W_{r_o,r_a}(\alpha_{o_k},\alpha_{a_k})$ define the corresponding Wald statistic constructed by maximizing over all values of the unknown cointegrating vectors. In particular, defining $W_{0,r_a}(0,\alpha_{a_k}) \equiv \text{Sup}_{\alpha_{au}} W(\alpha_{a_k},\alpha_{a_u})$ from (2.10), then $W_{r_o,r_a}(\alpha_{o_k},\alpha_{a_k}) \equiv W_{0,r_o+r_a}(0,[\alpha_{o_k}\alpha_{a_k}]) - W_{0,r_o}(0,\alpha_{o_k})$. Writing the statistic as $W_{r_o,r_a}(\alpha_{o_k},\alpha_{a_k})$ completely describes the null and alternative hypotheses: $r_{o_k} = \text{rank}(\alpha_{o_k})$, $r_{o_u} = r_o - \text{rank}(\alpha_{o_k})$ and similarly for r_{a_k} and r_{a_u}. Using this notation, the well-known likelihood ratio tests developed in Johansen (1988) are denoted as $LR_{r_o,r_a}(0,0)$ and the associated Wald statistics are $W_{r_o,r_a}(0,0)$.

To derive the asymptotic distribution of $W_{r_o,r_a}(\alpha_{o_k},\alpha_{a_k})$, we make four sets of assumptions.

Assumption A. The data are generated by (2.1a)–(2.1c) with the following:

(A.1) $E(\epsilon_t | \epsilon_{t-1},\ldots,\epsilon_1) = 0,$

$E(\epsilon_t\epsilon_t' | \epsilon_{t-1},\ldots,\epsilon_1) = \Sigma_\epsilon,$

$E(\epsilon_{i,t}^4) < \kappa < \infty$ for all i and t.

(A.2) Letting $\Phi(z) = I - \Phi_1 z - \cdots - \Phi_{p-1}z^{p-1}$, then the roots of $|\Phi(z)|$ are all outside the unit circle.

(A.3) $X_{-i} = 0,\ i = 0,\ldots,p-1.$

(A.4) Three alternative assumptions are made about d_t:

(A.4.i) $d_t = 0$ for all t;

(A.4.ii) $d_t = \mu_0$ for all t;

(A.4.iii) $d_t = \mu_0 + \mu_1 t$ for all t, with $\alpha_o'\mu_1 = 0$ and $\alpha_{a_k}'\mu_1 = 0$.

Note that under Assumption (A.4.iii) we assume that α_{a_k} annihilates the deterministic drift in the series under both the null and the alternative.

The test statistic will be formed as already described, when $d_t = 0$. When $d_t \neq 0$, the VECM is augmented with a constant, and the statistic is calculated as earlier with Z_t in (2.3) augmented by a constant. Because, under Assumption (A.4.iii), the constant term in VECM (2.11) is unrestricted, augmenting Z_t with a constant and carrying out least squares produces the Gaussian maximum likelihood estimator. However, under Assumption (A.4.ii), the constant term in VECM (2.11) is constrained (see (2.12)), and thus the least-squares estimator does not correspond to the Gaussian MLE. We nevertheless consider test statistics based on this formulation for two reasons. First, when some columns of α are known, the unconstrained estimator and test statistics are much easier to calculate than the constrained estimator; the required calculations when α is known are discussed in Johansen and Juselius (1990) and in Johansen (1991). Second, we show that when α is unknown, the test based on the unconstrained estimator has somewhat better local power than the test based on the constrained estimator.

Convenient representations for the asymptotic null distribution can be derived using the following notation. Let $B(s) = (B_1(s)B_2(s)\cdots B_n(s))'$ denote an $n \times 1$ dimensional standard Wiener process; $\int_0^1 F(s)\,ds = \int F$ and $\int_0^1 F(s)\,dB(s) = \int F\,dB$, for arbitrary function $F(s)$; $B^\mu(s) = B(s) - \int B$ denote the corresponding "demeaned" process; $s^\mu = s - \int s = s - \frac{1}{2}$ denote the demeaned time trend; and let $B_{i,j}(s) = (B_i(s) \dots B_j(s))'$ denote a $(j - i + 1) \times 1$ subvector of $B(s)$, and let $B_{i,j}^\mu$ be defined analogously.

THEOREM 1. *The asymptotic null distribution of $W_{r_o,r_a}(\alpha_{o_k}, \alpha_{a_k})$ can be represented as*

$$W_{r_o,r_a}(\alpha_{o_k}, \alpha_{a_k}) \Rightarrow \text{Trace}\left[\left(\int F_1\, dB_{1,k}\right)'\left(\int F_1 F_1'\right)^{-1}\left(\int F_1\, dB_{1,k}\right)\right] + \sum_{i=1}^{r_{au}} \lambda_i\left[\left(\int F_2\, dB_{1,k}'\right)'\left(\int F_2 F_2'\right)^{-1}\left(\int F_2\, dB_{1,k}'\right)\right],$$

where $k = n - r_{o_u}$, $F_2(s) = F_3(s) - \gamma_{3,1} F_1(s)$ with $\gamma_{3,1} = \int F_3 F_1' [\int F_1 F_1']^{-1}$, $\lambda_i[.]$ is the ith largest eigenvalue of the matrix in brackets, and the definition of $F_1(s)$ and $F_3(s)$ depends on the particular assumptions employed.

In particular, we have the following cases.

Case (1). Suppose that Assumptions (A.1)–(A.3) and (A.4.i) hold, and the statistic is calculated with $Z_t = (\Delta Y_{t-1}' \Delta Y_{t-2}' \cdots \Delta Y_{t-p+1}')'$, then $F_1(s) = B_{1,m}(s)$ with $m = r_{a_k}$ and $F_3(s) = B_{i,j}(s)$ with $i = r_{a_k} + 1$ and $j = n - r_{o_k} - r_{o_u}$.

Case (2). Suppose that Assumptions (A.1)–(A.2) and (A.4.ii) hold, and the statistic is calculated with $Z_t = (1\, \Delta Y_{t-1}' \Delta Y_{t-2}' \cdots \Delta Y_{t-p+1}')'$, then $F_1(s) = B_{1,m}^\mu(s)$ with $m = r_{a_k}$ and $F_3(s) = B_{i,j}^\mu(s)$ with $i = r_{a_k} + 1$ and $j = n - r_{o_u} - r_{o_k}$.

Case (3). Suppose that Assumptions (A.1)–(A.2) and (A.4.iii) hold, and the statistic is calculated with $Z_t = (1\, \Delta Y_{t-1}' \Delta Y_{t-2}' \cdots \Delta Y_{t-p+1}')'$, then $F_1(s) = B_{1,m}^\mu(s)$ with $m = r_{a_k}$, and $F_3(s) = (s^\mu(s)'\, B_{i,j}^{\mu\prime}(s))$ with $i = r_{a_k} + 1$ and $j = n - r_{o_k} - r_{o_u} - 1$.

Proof. See the Appendix.

We make six remarks about these results. First, Theorem 1 is a generalization of the results in Johansen (1988, 1991), who considered the problem with $r_{o_k} = r_{a_k} = 0$. Second, when a constant is included in Z_t, the test statistic is invariant to the initial conditions for X_t, $t = 0, \dots, -p + 1$ under the null hypothesis. Thus, Assumption (A.3) is not necessary under Cases (2) and (3) in Theorem 1. Third, when $r_{a_u} = 0$, the limiting distributions in Cases (2) and (3) are the same. Fourth, under Cases (1) and (3), the $W_{r_o,r_a}(\alpha_{o_k}, \alpha_{a_k})$ statistic is asymptotically equivalent to the LR statistics; this equivalence fails to obtain in Case (2) because the constraint on the constant term in VECM's (2.11) and (2.12) is imposed when the LR statistic is calculated, but the W statistic is calculated using an unconstrained estimator. Fifth, while the case with $d_t = \mu_0 + \mu_1 t$ for all t, with $\alpha_o' \mu_1 = 0$, and $\alpha_{a_k}' \mu_1 \neq 0$ is not covered by the theorem, the limiting distribution of the test statistic is readily deduced in this case as well. Because we did not tabulate critical values for this case, we did not include the limiting distribution in the theorem. As a practical matter, our calculations indicated that the critical values for the test statistic under the assumption that $\alpha_{a_k}'\mu_1 = 0$ are larger than those under the assumption $\alpha_{a_k}'\mu_1 \neq 0$, and so using the Case (3) distribution results in conservative inference. Finally, it is also straightforward to generalize the theorem to accommodate linear restrictions on the cointegrating vector of the form $R\alpha_{a_u} = 0$, where R is a known $\ell \times n$ matrix. Specifically, the statistic is formed as in (2.10), where now the matrix G is $n \times (n - r_{a_k} - \ell)$ with columns spanning the null space of the columns of $(\alpha_{a_k} H')$; the asymptotic distribution Theorem 1 continues to hold except that the index j in the definition of $F_3(s)$ becomes $j = n - r_{o_u} - r_{a_k} - \ell$. General linear restrictions of the form $R[\text{vec}(\alpha_{a_u})] = h$ are not covered by the theorem.

Critical values for $n - r_{o_u} \leq 5$ are provided in Table 1. These critical values were calculated by simulation using 10,000 replications and $T =$ 1,000. Extended critical values of $n - r_{o_u} \leq 9$ are tabulated in Horvath and Watson (1995). When $r_{o_k} = r_{a_k} = 0$, these correspond to the critical values tabulated in Johansen (1988), Johansen and Juselius (1990), and Osterwald-Lenum (1992).[3]

3. COMPARISON OF TESTING PROCEDURES

In this section, we carry out three power comparisons. First, we compare the local power of the W/LR tests that impose the value of the cointegrating vector under the alternative to the corresponding tests that do not use this information. Second, because a priori information about the cointegrating vector may only be approximately correct, we investigate the power implications of imposing an incorrect value of the cointegrating vector. Finally, for the special case with $r_o = r_{a_u} = 0$ and $r_{a_k} = 1$, we compare the power of the VECM-based tests to univariate unit root tests applied to the error correction term.

For tractability, our discussion will focus on a bivariate version of (2.11), with $\Phi_1 = \Phi_2 = \cdots = \Phi_{p-1} = 0$:

$$\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} (\alpha' y_{t-1}) + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}. \tag{3.1}$$

Because the likelihood-based procedures are invariant to nonsingular transformations of Y_t, we can set $\alpha = (0\ 1)'$ and $\delta_1 = 0$ when studying these tests. This will also prove convenient when studying univariate testing procedures. Thus, the model that we consider is

$$\Delta y_{1,t} = \theta_1 + \epsilon_{1,t} \tag{3.2a}$$

$$\Delta y_{2,t} = \theta_2 + \delta_2 y_{2,t-1} + \epsilon_{2,t}. \tag{3.2b}$$

Table 1. Critical values for tests for cointegration

				Case 1			Case 2			Case 3		
$n - r_{o_u}$	r_{o_k}	r_{a_k}	r_{a_u}	1%	5%	10%	1%	5%	10%	1%	5%	10%
1	0	0	1	7.26	4.12	2.95	12.18	8.47	6.63	6.84	3.98	2.73
1	0	1	0	7.26	4.12	2.95	12.18	8.47	6.63	12.18	8. 47	6.63
2	0	0	1	14.83	11.03	9.35	19.14	14.93	13.01	18.13	14.18	12.36
2	0	0	2	16.10	12.21	10.45	22.43	18.17	15.87	19.66	15.41	13.54
2	0	1	0	9.43	6.28	4.73	13.73	10.18	8.30	13.73	10.18	8.30
2	0	1	1	16.10	12.21	10.45	22.43	18.17	15.87	19.66	15.41	13.54
2	0	2	0	16.10	12.21	10.45	22.43	18.17	15.87	22.43	18.17	15.87
2	1	0	1	9.43	6.28	4.73	13.73	10.18	8.30	8.94	6.02	4.64
2	1	1	0	9.43	6.28	4.73	13.73	10.18	8.30	13.73	10.18	8.30
3	0	0	1	22.25	17.51	15.42	25.93	21.19	19.12	26.17	21.14	18.62
3	0	0	2	28.02	23.28	20.81	35.98	29.46	26.79	34.84	28.75	26.08
3	0	0	3	29.31	23.91	21.52	37.72	31.66	28.82	35.83	29.62	27.05
3	0	1	0	11.44	7.94	6.43	15.41	11.62	9.72	15.41	11.62	9.72
3	0	1	1	24.91	20.30	18.05	31.42	26.08	23.67	30.67	25.70	23.04
3	0	1	2	29.31	23.91	21.52	37.72	31.66	28.82	35.83	29.62	27.05
3	0	2	0	19.75	15.20	13.04	25.35	20.74	18.51	25.35	20.74	18.51
3	0	2	1	29.31	23.91	21.52	37.72	31.66	28.82	35.83	29.62	27.05
3	0	3	0	29.31	23.91	21.52	37.72	31.66	28.82	37.72	31.66	28.82
3	1	0	1	16.84	12.89	11.03	21.62	16.65	14.51	20.36	15.93	13.93
3	1	0	2	19.75	15.20	13.04	25.35	20.74	18.51	22.90	18.18	16.25
3	1	1	0	11.44	7.94	6.43	15.41	11.62	9.72	15.41	11.62	9.72
3	1	1	1	19.75	15.20	13.04	25.35	20.74	18.51	22.90	18.18	16.25
3	1	2	0	19.75	15.20	13.04	25.35	20.74	18.51	25.35	20.74	18.51
3	2	0	1	11.44	7.94	6.43	15.41	11.62	9.72	11.39	7.87	6.36
3	2	1	0	11.44	7.94	6.43	15.41	11.62	9.72	15.41	11.62	9.72
4	0	0	1	28.33	23.82	21.51	32.35	27.40	24.94	32.19	27.07	24.84
4	0	0	2	40.14	34.35	31.63	47.03	40.50	37.78	46.00	40.27	37.17
4	0	0	3	44.62	39.17	35.90	54.25	47.31	44.03	53.14	46.30	43.32
4	0	0	4	45.66	39.91	36.58	56.17	49.16	45.61	54.34	47.33	44.09
4	0	1	0	13.60	9.73	7.93	17.16	13.20	11.16	17.16	13.20	11.16
4	0	1	1	32.75	27.86	25.43	39.55	33.55	30.73	39.47	33.22	30.45
4	0	1	2	42.47	36.93	33.81	51.82	44.98	41.45	50.96	43.78	40.94
4	0	1	3	45.66	39.91	36.58	56.17	49.16	45.61	54.34	47.33	44.09
4	0	2	0	22.85	17.92	15.81	28.62	23.41	21.10	28.62	23.41	21.10
4	0	2	1	38.43	33.36	30.69	47.26	40.98	38.11	46.82	40.76	37.50
4	0	2	2	45.66	39.91	36.58	56.17	49.16	45.61	54.34	47.33	44.09
4	0	3	0	33.53	27.80	25.24	41.08	35.33	32.33	41.08	35.33	32.33
4	0	3	1	45.66	39.91	36.58	56.17	49.16	45.61	54.34	47.33	44.09
4	0	4	0	45.66	39.91	36.58	56.17	49.16	45.61	56.17	49.16	45.61
4	1	0	1	24.15	19.28	17.30	27.09	22.73	20.61	28.06	22.74	20.36

(*continued*)

Table 1 (*continued*)

				Case 1			Case 2			Case 3		
$n - r_{o_u}$	r_{o_k}	r_{a_k}	r_{a_u}	1%	5%	10%	1%	5%	10%	1%	5%	10%
4	1	0	2	31.30	26.19	23.82	37.76	32.45	29.49	38.01	31.74	28.65
4	1	0	3	33.53	27.80	25.24	41.08	35.33	32.33	40.07	33.57	30.41
4	1	1	0	13.60	9.73	7.93	17.16	13.20	11.16	17.16	13.20	11.16
4	1	1	1	28.04	23.19	20.82	33.83	28.87	26.10	33.45	28.25	25.73
4	1	1	2	33.53	27.80	25.24	41.08	35.33	32.33	40.07	33.57	30.41
4	1	2	0	22.85	17.92	15.81	28.62	23.41	21.10	28.62	23.41	21.10
4	1	2	1	33.53	27.80	25.24	41.08	35.33	32.33	40.07	33.57	30.41
4	1	3	0	33.53	27.80	25.24	41.08	35.33	32.33	41.08	35.33	32.33
4	2	0	1	18.59	14.60	12.78	23.09	18.37	16.12	21.92	17.52	15.51
4	2	0	2	22.85	17.92	15.81	28.62	23.41	21.10	25.82	21.00	18.74
4	2	1	0	13.60	9.73	7.93	17.16	13.20	11.16	17.16	13.20	11.16
4	2	1	1	22.85	17.92	15.81	28.62	23.41	21.10	25.82	21.00	18.74
4	2	2	0	22.85	17.92	15.81	28.62	23.41	21.10	28.62	23.41	21.10
4	3	0	1	13.60	9.73	7.93	17.16	13.20	11.16	12.81	9.54	7.85
4	3	1	0	13.60	9.73	7.93	17.16	13.20	11.16	17.16	13.20	11.16
5	0	0	1	35.29	30.51	27.76	39.10	33.87	31.08	38.95	33.51	30.89
5	0	0	2	51.50	45.84	42.75	59.27	52.05	48.77	57.99	51.53	48.24
5	0	0	3	61.05	54.42	51.22	70.75	63.29	59.44	70.30	62.45	58.82
5	0	0	4	65.54	58.65	55.23	77.46	69.37	65.20	75.64	67.89	64.37
5	0	0	5	66.00	59.39	55.80	78.85	70.93	66.58	76.36	68.62	65.15
5	0	1	0	15.32	11.41	9.46	19.00	14.53	12.49	19.00	14.53	12.49
5	0	1	1	41.09	35.77	32.98	47.18	41.36	38.44	46.58	40.78	38.15
5	0	1	2	56.00	49.75	46.41	64.19	57.55	53.98	63.59	56.60	53.43
5	0	1	3	63.52	56.83	53.56	74.61	66.88	63.00	73.49	65.73	62.31
5	0	1	4	66.00	59.39	55.80	78.85	70.93	66.58	76.36	68.62	65.15
5	0	2	0	26.01	20.92	18.55	31.26	26.15	23.51	31.26	26.15	23.51
5	0	2	1	48.36	42.54	39.54	56.90	50.15	46.93	56.23	49.55	46.51
5	0	2	2	60.54	54.27	50.93	71.63	64.20	60.46	70.31	62.86	59.64
5	0	2	3	66.00	59.39	55.80	78.85	70.93	66.58	76.36	68.62	65.15
5	0	3	0	37.35	31.75	28.94	44.87	39.03	36.03	44.87	39.03	36.03
5	0	3	1	57.01	50.44	47.36	67.41	60.14	56.68	66.72	59.62	55.85
5	0	3	2	66.00	59.39	55.80	78.85	70.93	66.58	76.36	68.62	65.15
5	0	4	0	50.02	44.42	41.43	61.04	53.88	50.14	61.04	53.88	50.14
5	0	4	1	66.00	59.39	55.80	78.85	70.93	66.58	76.36	68.62	65.15
5	0	5	0	66.00	59.39	55.80	78.85	70.93	66.58	78.85	70.93	66.58
5	1	0	1	30.10	25.62	23.21	34.36	29.09	26.61	33.87	28.72	26.37
5	1	0	2	42.91	37.30	34.70	50.23	43.52	40.60	49.21	42.92	40.02
5	1	0	3	48.63	42.91	40.13	58.91	51.22	47.80	57.59	50.31	47.15
5	1	0	4	50.02	44.42	41.43	61.04	53.88	50.14	59.39	51.95	48.67
5	1	1	0	15.32	11.41	9.46	19.00	14.53	12.49	19.00	14.53	12.49
5	1	1	1	36.01	30.74	28.25	41.68	36.30	33.62	41.37	35.94	33.11

(*continued*)

TABLE 1 (*continued*)

				Case 1			Case 2			Case 3		
$n - r_{o_u}$	r_{o_k}	r_{a_k}	r_{a_u}	1%	5%	10%	1%	5%	10%	1%	5%	10%
5	1	1	2	46.54	40.78	37.76	55.99	48.54	45.25	54.54	47.42	44.73
5	1	1	3	50.02	44.42	41.43	61.04	53.88	50.14	59.39	51.95	48.67
5	1	2	0	26.01	20.92	18.55	31.26	26.15	23.51	31.26	26.15	23.51
5	1	2	1	42.58	37.40	34.60	50.71	44.76	41.71	50.25	44.34	41.27
5	1	2	2	50.02	44.42	41.43	61.04	53.88	50.14	59.39	51.95	48.67
5	1	3	0	37.35	31.75	28.94	44.87	39.03	36.03	44.87	39.03	36.03
5	1	3	1	50.02	44.42	41.43	61.04	53.88	50.14	59.39	51.95	48.67
5	1	4	0	50.02	44.42	41.43	61.04	53.88	50.14	61.04	53.88	50.14
5	2	0	1	25.44	20.91	18.95	28.77	24.48	22.09	29.62	24.41	21.83
5	2	0	2	34.64	29.41	26.66	40.57	35.03	32.20	40.73	34.50	31.42
5	2	0	3	37.35	31.75	28.94	44.87	39.03	36.03	43.65	37.21	34.13
5	2	1	0	15.32	11.41	9.46	19.00	14.53	12.49	19.00	14.53	12.49
5	2	1	1	31.01	25.99	23.64	36.35	31.39	28.72	36.34	30.99	28.34
5	2	1	2	37.35	31.75	28.94	44.87	39.03	36.03	43.65	37.21	34.13
5	2	2	0	26.01	20.92	18.55	31.26	26.15	23.51	31.26	26.15	23.51
5	2	2	1	37.35	31.75	28.94	44.87	39.03	36.03	43.65	37.21	34.13
5	2	3	0	37.35	31.75	28.94	44.87	39.03	36.03	44.87	39.03	36.03
5	3	0	1	20.52	16.39	14.39	24.46	19.95	17.70	23.82	19.16	16.94
5	3	0	2	26.01	20.92	18.55	31.26	26.15	23.51	28.71	23.83	21.25
5	3	1	0	15.32	11.41	9.46	19.00	14.53	12.49	19.00	14.53	12.49
5	3	1	1	26.01	20.92	18.55	31.26	26.15	23.51	28.71	23.83	21.25
5	3	2	0	26.01	20.92	18.55	31.26	26.15	23.51	31.26	26.15	23.51
5	4	0	1	15.32	11.41	9.46	19.00	14.53	12.49	15.02	11.23	9.31
5	4	1	0	15.32	11.41	9.46	19.00	14.53	12.49	19.00	14.53	12.49

To investigate the local power of the tests, we suppose that δ_2 is local to 0; specifically, we set $\delta_2 = \delta_{2,T} = -c/T$. This allows us to study local power using local-to-unity asymptotics familiar from the work of Bobkowsky (1983), Cavanagh (1985), Chan and Wei (1987), Chan (1988), Phillips (1987, 1988), and Stock (1991). To rule out drift in the error correction term, we set $\theta_2 = 0$. Finally, our initial comparisons are made with $\Sigma_\epsilon = \mathrm{I}$; the case of correlated errors is discussed later.

The local power results are conveniently stated in terms of a two-dimensional Wiener/diffusion process, $B_c(s) = (B_{1,c}(s)\ B_{2,c}(s))'$. Let $B(s) = (B_1(s)B_2(s))'$ denote a two-dimensional standardized Wiener process, let $B_{1,c}(s) = B_1(s)$, and let $B_{2,c}(s)$ evolve as $dB_{2,c}(s) = -cB_{2,c}(s)\,ds + dB_2(s)$. Thus, the first element of $B_c(s)$ is a random walk, and the second element is generated by a diffusion process with parameter c. Let $B_c^\mu(s) = B_c(s) - \int B_c$ denote the demeaned version of this bivariate process, and let $D_c(s) = (s^\mu(s)\ B^\mu_{2,c}(s))'$ denote the bivariate process composed of the demeaned values of the time trend and $B_{2,c}$. Corresponding to the three cases in Theorem 1, it is straightforward to derive limiting representations for the cointegration test statistics under local departures from the null. Let $\gamma = (\gamma_1\ \gamma_2)'$ denote an arbitrary 2×1 vector, and let $\alpha = (0\ 1)$ denote the true value of the cointegrating vector. Using the notation already introduced, $W_{0,1}(0,\gamma)$ (with $\gamma \neq 0$) denotes the test statistic for $H_o: r = 0$ vs. $H_a: r = r_{a_k} = 1$ constructed using γ as the cointegrating vector under the alternative; similarly, $W_{0,1}(0,0)$ denotes the test statistic for $H_o: r = 0$ vs. $H_a: r = r_{a_u} = 1$. The limiting distribution of this statistic is given by the following.

Case 1. Suppose that the data are generated by (3.2a) and (3.2b) with $\theta_1 = \theta_2 = 0$, $\delta_2 = -c/T$, and ϵ_t satisfies Assumption (A.1) with $\Sigma_\epsilon = \mathrm{I}$. If the test statistic is calculated without including a constant in Z_t, then

$$W_{0,1}(0,\gamma) \Rightarrow \mathrm{Trace}\left[\left(\gamma'\int B_c\,dB'\right)'\left(\gamma'\int B_c B_c'\gamma\right)^{-1}\left(\gamma'\int B_c\,dB'\right)\right];$$

$$W_{0,1}(0,0) \Rightarrow \lambda_1\left[\left(\int B_c\,dB'\right)'\left(\int B_c B_c'\right)^{-1}\left(\int B_c\,dB'\right)\right].$$

Case 2. Suppose that the data are generated by (3.2a) and (3.2b) with $\theta_1 = \theta_2 = 0$, $\delta_2 = -c/T$, and ϵ_t satisfies Assumption (A.1) with $\Sigma_\epsilon = \mathrm{I}$. If the test statistic is calculated including a constant in Z_t, then

$$W_{0,1}(0,\gamma) \Rightarrow \mathrm{Trace}\left[\left(\gamma'\int B_c^\mu\,dB'\right)'\left(\gamma'\int B_c^\mu B_c^{\mu\prime}\gamma\right)^{-1}\left(\gamma'\int B_c^\mu\,dB'\right)\right];$$

$$W_{0,1}(0,0) \Rightarrow \lambda_1\left[\left(\int B_c^\mu\,dB'\right)'\left(\int B_c^\mu B_c^{\mu\prime}\right)^{-1}\left(\int B_c^\mu\,dB'\right)\right].$$

Case 3. Suppose that the data are generated by (3.2a) and (3.2b) with $\theta_1 \neq 0$, $\theta_2 = 0$, $\delta_2 = -c/T$, and ϵ_t satisfies Assumption (A.1) with $\Sigma_\epsilon = \mathrm{I}$. If the test statistic is calculated including a constant in Z_t, then

$$W_{0,1}(0,\gamma) \Rightarrow \mathrm{Trace}\left[\left(\gamma'\int D_c\,dB'\right)'\left(\gamma'\int D_c D_c'\gamma\right)^{-1}\left(\gamma'\int D_c\,dB'\right)\right], \qquad \text{for } \gamma_1 = 0;$$

$$W_{0,1}(0,\gamma) \Rightarrow \mathrm{Trace}\left[\left(\int s^\mu\,dB'\right)'\left(\int (s^\mu)^2\right)^{-1}\left(\int s^\mu\,dB'\right)\right], \qquad \text{for } \gamma_1 \neq 0;$$

$$W_{0,1}(0,0) \Rightarrow \lambda_1\left[\left(\int D_c\,dB'\right)'\left(\int D_c D_c'\right)^{-1}\left(\int D_c\,dB'\right)\right].$$

In Case 3, when $\theta_1 \neq 0$ and $\gamma_1 \neq 0$, the regressor $\gamma' y_{t-1}$ is dominated by the linear trend $\gamma_1 \theta_1 t$. In contrast, $\gamma' y_{t-1}$ is a linear function of a diffusion process in Cases (1) and (2) for all values of γ_1, and in Case (3) when $\gamma_1 = 0$. This difference leads to the two possible limiting representations for $W_{0,1}(0,\gamma)$ in Case (3). When $\gamma_1 = 0$, the limiting distributions of $W_{0,1}(0,\gamma)$ coincide in Cases 2 and 3, because the second elements of B_c^{μ} and D_c are identical.

In Figure 1, we plot the local power curves associated with these limiting random variables for $\alpha = \gamma$.[4] Thus, the $W_{0,1}(0,\alpha_{a_k})$ plot shows the power of the test that imposes the true value of the cointegrating vector, and the $W_{0,1}(0,0)$ plot shows the power of the test that does not use this information. The power gains from incorporating the true value of the cointegrating vector are substantial: at 50% power they correspond to sample size increases of approximately 70, 50, and 40% for Cases 1-3, respectively. Figure 1B also shows the local power of the LR analog of $W_{0,1}(0,0)$ that imposes the constraint on the constant term shown in (2.12). As discussed in Johansen and Juselius (1990) and Johansen (1991), this statistic is calculated by augmenting the matrix Y_{-1} in (2.10) by a column of 1's and excluding the constant from Z_t. Letting $F_c(s)$ denote $(1\, B_c(s))$, this statistic has a limiting distribution given by $\lambda_1[(\int F_c\, dB')'(\int F_c F_c')^{-1}(\int F_c\, dB')]$. Interestingly, the power curve lies below the corresponding $W_{0,1}(0,0)$ power curve that does not impose this constraint on the constant term, and of course both curves lie below their Case 1 analog. The reduction in power for the LR statistic in Figure 1B relative to Figure 1A arises because, under the null that $\delta = 0$, the constant term β in (2.12) is unidentified. The LR statistic maximizes over this parameter, leading to an increase in the test's critical value. The reduction in power for the $W_{0,1}(0,0)$ statistic in Figure 1B relative to Figure 1A arises because the data are demeaned in Figure 1B, leading to a reduction in the variance of the regressor. Apparently, more powerful tests are obtained from using demeaned data rather than maximizing over the unidentified parameter β.

Because the a priori knowledge of the cointegrating vector may be inexact, it is also of interest to consider the behavior of the statistics constructed from incorrect values of the cointegrating vector. Asymptotic results for fixed values of $\delta_2 < 0$ imply that using the correct value of the cointegrating vector is critical to the power gains apparent in Figure 1. For fixed alternatives, the $W_{0,1}(0,0)$ and corresponding LR tests are consistent. On the other hand, because $\gamma' y_t$ is I(1) when γ is not proportional to α, the test based on $W_{0,1}(0,\gamma)$ for $\gamma \neq \alpha$ will not be consistent. Thus, imposing the incorrect value of the cointegrating vector would seem to have disastrous effects on the power of the test.

However, this drawback is somewhat artificial, because it applies in a situation when the power of the $W_{0,1}(0,0)$ test is unity. An arguably more meaningful comparison is obtained from the local-to-unity results where

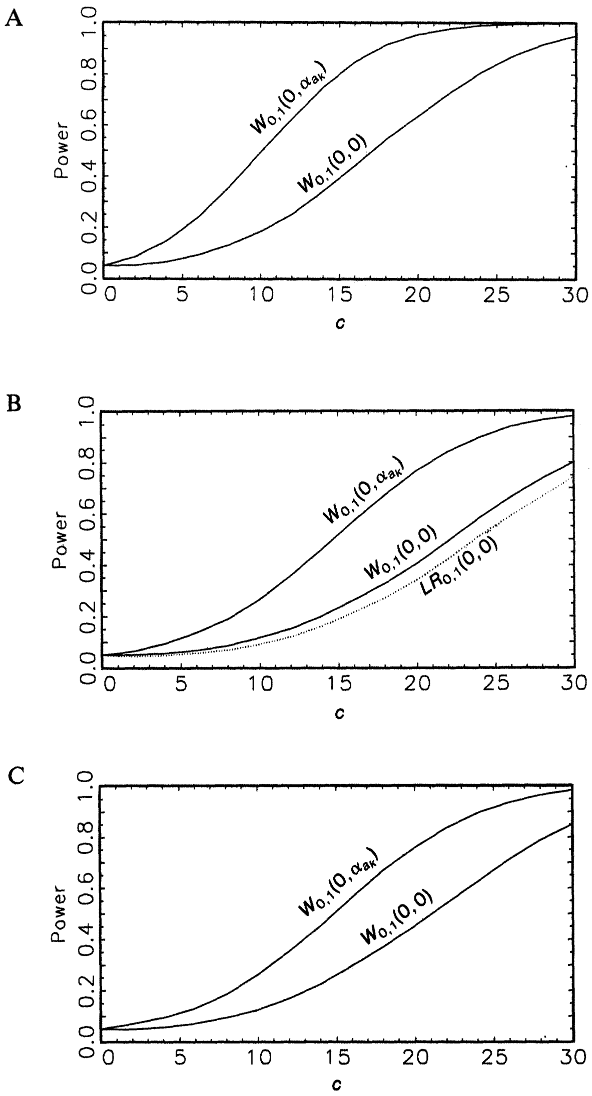

FIGURE 1. Local asymptotic power. Panels plot local power curves for a two-variable system. Curves labeled $W_{0,1}(0, \alpha_{a_k})$ show the power of the test that imposes the true value of the cointegrating vector. Curves labeled $W_{0,1}(0,0)$ show the power of the test that does not use this information. A. Case 1: Data contain zero drift terms, and statistics are calculated without inclusion of explanatory constant terms. B. Case 2: Data contain zero drift terms, but statistics are calculated with explanatory constant terms in regressions. The curve labeled $LR_{0,1}(0,0)$ shows the local power of the LR analog of $W_{0,1}(0,0)$ that imposes the constraint on the constant term in (2.12). C. Case 3: Data contain nonzero drift terms, and statistics are calculated with explanatory constant terms in regressions.

cointegration is weak. Figure 2 shows the power results for the $W_{0,1}(0,\gamma)$ test for a variety of values of $\gamma = (\gamma_1\ 1)$; also plotted are the power results for $W_{0,1}(0,0)$. Results are presented for the nontrending data Cases 1 and 2; results for Case 3 will be discussed shortly. It is apparent from Figure 2 that for values of γ_1 reasonably close to the true value of 0, the $W_{0,1}(0,\gamma)$ test continues to dominate the $W_{0,1}(0,0)$ test. For example, for the entire range of values of c considered, the $W_{0,1}(0,\gamma)$ test dominates the $W_{0,1}$ test for $\gamma_1 < 0.1$. On the other hand, for larger values of γ_1, the $W_{0,1}(0,0)$ test dominates for large values of c, in line with the results for the fixed alternative already described.

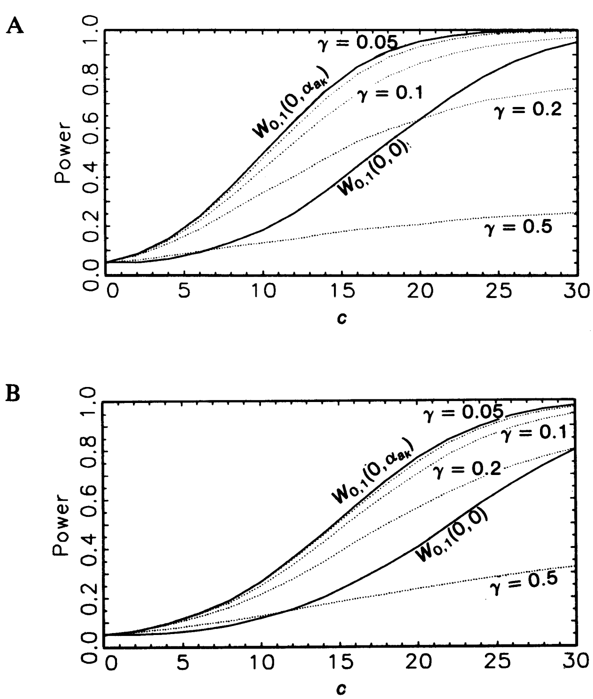

FIGURE 2. Local asymptotic power: Incorrectly specified cointegrating vector $(\gamma, 1)$. Panels plot local power curves for a two-variable system. Curves labeled $W_{0,1}(0,\alpha_{a_k})$ show the power of the test that imposes the true cointegrating vector $(0,1)$. Curves labeled $W_{0,1}(0,0)$ show the power of the test that does not use this information. Dotted curves show the power of the test that imposes an incorrect cointegrating vector $(\gamma, 1)$ for particular values of γ. A. Case 1: Data contain zero drift terms, and statistics are calculated without inclusion of explanatory constant terms. B. Case 2: Data contain zero drift terms, and statistics are calculated with explanatory constant terms in regressions.

The results are quite different in Case 3. These results are not shown because the rejection probability for the test constructed from incorrect values of γ_1 for the $W_{0,1}(0,\gamma)$ test are very small for all values of c. The reason for this can be seen from the limiting representation for $W_{0,1}(0,\gamma)$ in Case 3 that was already given. When $\gamma_1 \neq 0$ the $W_{0,1}(0,\gamma)$ statistic converges to $(\int s^{\mu}\, dB')'(\int (s^{\mu})^2)^{-1}(\int s^{\mu}\, dB')$, which has a χ_2^2 distribution. From Table 1, the 5% critical value for the $W_{0,1}(0,\gamma)$ test is 10.18, so that the corresponding rejection probability for the $W_{0,1}(0,\gamma)$ test using the incorrect value of γ is $P(\chi_2^2 > 10.18) = 0.6\%$.

Arguably, these results for Case 3 have little relevance. After all, when $\theta_1 \neq 0$, $\gamma' y_t$ will be trending when $\gamma_1 \neq 0$. This behavior would be obvious in a large sample, and so the hypothesis that $\gamma' y_t$ is $I(0)$ could easily be dismissed. This suggests that the comparison should be made, for example, with θ_1 or γ_1 local to 0, say $\theta_1 = c_{\theta_1}/T^{1/2}$ or $\gamma_1 = c_{\gamma_1}/T^{1/2}$. Because these power functions depend critically on the assumed values of the constant c_{θ_1} and c_{γ_1}, and because reasonable values of these parameters will differ from application to application, we do not report these functions. Instead, we carry out an experiment for a fixed sample size and Gaussian errors, using values for the parameters in (3.2a) and (3.2b) and values of γ_1 that are relevant for a typical application: the analysis of postwar U.S. quarterly data on income and consumption. Letting $y_{1,t}$ denote the logarithm of per capita consumption and $y_{2,t}$ denote the logarithm of the consumption/income ratio, then $\theta_1 = 0.004$, $\sigma_1 = 0.006$, $\sigma_2 = 0.011$, $\text{cor}(\epsilon_{1,t}\epsilon_{2,t}) = 0.21$, and $T = 175$.[5] In Figure 3, results are shown for values of γ_1 ranging from 0 to 0.10. For com-

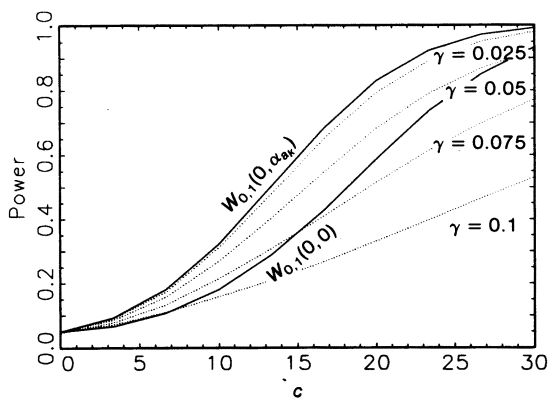

FIGURE 3. Power in the income–consumption system: Incorrectly specified cointegrating vector $(\gamma, 1)$. Panel plots local power curves for a two-variable system with parameters chosen to match the postwar U.S. quarterly data on income and consumption. Notation on curves matches that of Figure 2. See notes for Figure 2 for clarification.

parison with previous graphs, δ_2 is written as $-c/T$, and the power is plotted against c. For this example, the $W_{0,1}(0,\gamma)$ dominates the $W_{0,1}(0,0)$ statistic for all values of c considered when the error in the postulated cointegrating vector is 5% or less.

When there is only one cointegrating vector under the alternative, simple univariate tests provide an alternative to the likelihood-based tests. Thus, if the cointegrating vector is assumed to be known, then the error correction term $\alpha' y_t$ can be formed and cointegration tested by employing a standard unit root test. The final task of this section is to compare the VECM likelihood-based test to standard univariate tests.

There are three distinct differences between the multivariate tests considered in this paper and standard univariate unit root tests. These are easily discussed in terms of the bivariate example summarized in (3.1) and (3.2). First, univariate tests concentrate on equation (3.2b) and test the simple null, $\delta_2 = 0$. Multivariate tests consider the whole system (3.1) and test the composite null, $\delta_1 = \delta_2 = 0$. This has both positive and negative effects: because $\delta_1 = 0$ (from (3.2a)), the multivariate tests lose power through an extra degree of freedom. In this sense, the univariate test is more powerful because it is focused in the right direction. On the other hand, the multivariate tests utilize any covariance between $\epsilon_{1,t}$ and $\epsilon_{2,t}$ to increase test power. This potential covariance is ignored in the univariate tests. The second difference between the univariate and multivariate tests is that the univariate tests typically use a one-sided alternative ($\delta_2 < 0$), whereas the multivariate tests consider two-sided alternatives. The third major difference is the conditioning set used to estimate δ_2 in (3.2b). In general, lagged first differences enter equation (3.1), so that both the univariate and multivariate tests must be constructed from regressions "augmented" with lags of the variables. The multivariate tests include lagged values of $\Delta y_{1,t}$ and $\Delta y_{2,t}$ in the regression; univariate procedures, such as augmented Dickey–Fuller regression, include only lags of $\Delta y_{2,t}$. Thus, when lags of $\Delta y_{1,t}$ help predict $\Delta y_{2,t}$, the error term in the multivariate regression will have a smaller variance than the error term in the univariate regression. When $\Delta y_{1,t}$ and $\Delta y_{2,t}$ are I(0), as assumed here, this leads to a more efficient estimator of δ_2 and a more powerful test. (Of course, this final point has force only when it is known that $\Delta y_{1,t}$ and $\Delta y_{2,t}$ are I(0).)

This last point is the subject of recent papers by Kremers, Ericsson, and Dolado (1992) and Hansen (1993). These papers carefully document the power gains associated with augmenting standard Dickey–Fuller regressions with additional I(0) regressors and allow us to focus instead on the potential power gains and losses associated with the first two differences in the univariate and multivariate procedures. Specifically, Figure 4 compares the power of the univariate and multivariate tests using the same design discussed earlier, but now for various values of $\rho = \text{cor}(\epsilon_{1,t}\epsilon_{2,t})$. All statistics are computed using demeaned values of the data. Two results stand out from the figure. First, the power functions of the one-sided Dickey–Fuller t-test and the two-sided test based on the squared t-statistic are nearly identical. This is a reflection of the skewed distribution of the Dickey–Fuller t-statistic. Thus, the two-sided nature of the W statistics has little impact on the power relative to the one-sided univariate test. Second, the relative performance of the $W(0, \alpha)$ statistic depends critically on the value of ρ^2, the squared correlation between $\epsilon_{1,t}$ and $\epsilon_{2,t}$. When $\rho^2 = 0$, the power loss in the $W(0, \alpha)$ statistic relative to the univariate test corresponds to a sample size reduction of 10% at 50% power. This is the loss of power associated with the extra degree of freedom in the multivariate test. However, the power gains from exploiting nonzero values for ρ are large. For example, when $\rho^2 = 0.10$, the multivariate and univariate tests have essentially identical power. For larger values of ρ^2, the multivariate dominate the univariate tests. For example, when $\rho^2 = 0.50$, the power gain corresponds to a sample size increase of over 60% at 50% power. The reason for this power gain follows from standard seemingly unrelated regression logic: nonzero values of ρ^2 essentially allow the multivariate procedure to partial out part of the error term in (3.2b) and increase the power of the test.

Of course, the results shown in Figure 4 apply to a design with one cointegrating vector in a bivariate system. In a higher dimensional system with only one cointegrating vector, the power of the multivariate test will fall

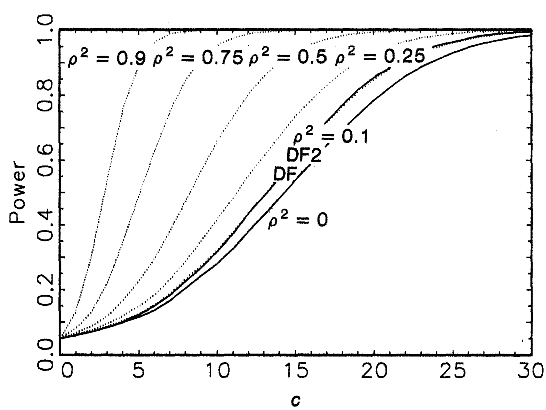

FIGURE 4. Local asymptotic power. Panel plots local power curves for a two-variable system where the covariance between the error terms is allowed to be different from zero. Solid curves labeled DF and DF2 show the power of one- and two-sided Dickey–Fuller univariate tests for a unit root. The solid curve labeled $\rho^2 = 0$ shows the power of the Wald test imposing the correct cointegrating vector when the (squared) correlation between the error terms is zero. Dotted curves show the power of the Wald test for different nonzero levels of the squared correlation in the error terms.

because of the extra degrees of freedom. Univariate tests could still be used in this case, but these tests become difficult to use and interpret when there are multiple cointegrating vectors.

4. STABILITY OF THE FORWARD-SPOT FOREIGN EXCHANGE PREMIUM

In this section, we examine forward and spot exchange rates, focusing on whether the forward-spot premium, defined as the forward exchange rate minus the spot exchange rate (in logarithms), is I(0). The data come from Citicorp Database Services, are sampled weekly for the period January 1975 through December 1989 (for a total of 778 observations), and are adjusted for transactions costs induced by bid–ask spreads and for the 2-day/nonholiday delivery lag for spot market exchange orders, as described in Bekaert and Hodrick (1993).[6] The forward–spot premia for the British pound, Swiss franc, German mark, and Japanese yen, the currencies used in our analysis, are shown in Figure 5.

The tests for cointegration are performed on bivariate systems of forward and spot rates in levels, currency by currency. In each case, the number of lagged first differences in the VECM was determined by step-down testing, beginning with a lag length of 18 and using a 5% test for each lag length (for an analysis of step-down testing in the context of testing for unit roots, see Ng and Perron, 1993). Results for testing for cointegration between forward and spot rates are presented in Table 2. For each currency, we report the test statistic for the case where we impose $\alpha = (1\ -1)'$ (denoted by $W_{0,1}(0,\alpha_{a_k})$), the test statistic for the case where α is unspecified (denoted by $W_{0,1}(0,0)$), the cointegrating vector estimated in this case (denoted by $\hat{\alpha}_{a_u}$), and the ADF statistic calculated from the forward premium. All statistics are reported for the optimal lag length chosen via the step-down procedure. Con-

TABLE 2. Tests for cointegration between spot and forward exchange rates (weekly data, January 1975 to December 1989)

Currency	$W_{0,1}(0,\alpha_{a_k})$		$W_{0,1}(0,0)$		$\hat{\alpha}_{a_u}$	ADF	
British pound	10.95	(0.04)	10.97	(0.21)	[1 − 1.001 (0.004)]	−3.12	(0.03)
Swiss franc	12.73	(0.02)	13.67	(0.08)	[1 − 0.998 (0.003)]	−3.33	(0.02)
German mark	23.38	(<0.01)	25.00	(<0.01)	[1 − 0.999 (0.002)]	−3.58	(<0.01)
Japanese yen	15.00	(<0.01)	15.02	(0.05)	[1 − 1.001 (0.003)]	−2.99	(0.04)

Note: The statistics $W_{0,1}(0,\alpha_{a_k})$ were calculated using $\alpha_{a_k} = (1\ -1)'$. The numbers in parentheses next to the test statistics are p-values. The estimated cointegrating vector $\hat{\alpha}_{a_u}$ is normalized as $(1\ \hat{\beta})$, and the numbers in parentheses are the standard errors for $\hat{\beta}$ computed under the maintained hypothesis that the data are cointegrated.

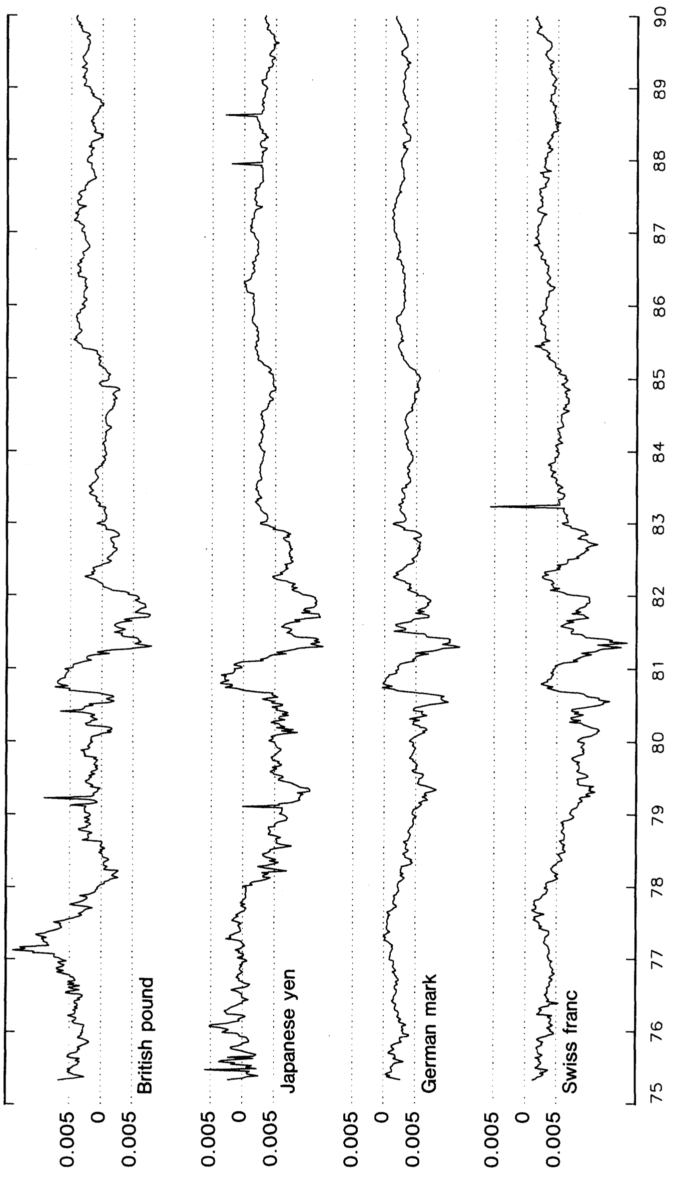

FIGURE 5. Forward-spot premia.

stant terms were included in all regressions, and so the p-values for the $W_{0,1}(0,\alpha_{a_k})$ statistic are from the Case (3) asymptotic null distribution (equivalently Case (2), because $\alpha_{a_u}=0$). Because nominal exchange rates exhibit some trending behavior over the sample period, the p-values for the $W_{0,1}(0,0)$ statistic are reported from the Case (3) asymptotic null distribution.

Looking first at the $W_{0,1}(0,\alpha_{a_k})$ column, the null of no cointegration is rejected for all currencies at the 5% level. The $W_{0,1}(0,0)$ statistics, which can be interpreted as $W_{0,1}(0,\alpha)$ maximized over all values of α, differ little from the $W_{0,1}(0,\alpha_{a_k})$ statistics. Their p-values are much greater, however, because their null distribution must account for the fact that they are maximized versions of $W_{0,1}(0,\alpha_{a_k})$. The next column shows why the two statistics are so similar: the estimated values of the cointegrating vector are equal to $(1\ -1)$, out to two decimal places.[7] The final column shows the ADF test statistic applied directly to the forward-spot premium. Like the $W_{0,1}(0,\alpha_{a_k})$ statistic, the ADF tests reject the null at the 5% level for all of the currencies. This application clearly shows the power advantage of testing for cointegration using a prespecified value of the cointegrating vector. Using the $W_{0,1}(0,0)$ statistic, the null of no cointegration is rejected at the 5% level for only two of the four currencies.

5. CONCLUDING REMARKS

In this paper, we have generalized VECM-based tests for cointegration to allow for known cointegrating vectors under both the null and alternative hypotheses. The results presented in Section 3 suggest that the power gains associated with these new methods can be substantial. These power gains were evident in the tests for cointegration involving forward and spot exchange rates. Cointegration was found in all currencies using tests that imposed a cointegrating vector of $(1\ -1)$, but the null of cointegration was rejected in only half of the cases when this information was not used. Yet, in these bivariate exchange rate models, the univariate ADF test applied to the forward premium $(F_t - S_t)$ yielded roughly the same inference as the multivariate VECM-based tests that imposed the cointegrating vector. Arguably, a more interesting application of the new procedures will be in larger systems with some known and some unknown cointegrating vectors. As argued in Section 3, the power trade-offs in the multivariate and univariate tests for cointegration are more interesting in higher dimensional systems.

The tests developed here rely on simple methods for eliminating trends in the data—incorporating unrestricted constants in the VECM. In the unit root context, the work by Elliott et al. (1995) suggests that large power gains can be achieved using alternative detrending methods. Hence, one extension of the current research will be a thorough investigation of alternative methods of detrending and their effects on tests for cointegration.

NOTES

1. Formally, the restriction $\text{rank}(\delta_a'\alpha_a) = r_a$ should be added to the alternative. Because this constraint is satisfied almost surely by the estimators under the alternative, it can be ignored when constructing the likelihood ratio test statistics.

2. The formulation used here is not as general as that used in Johansen (1992a), who considers a model of the form $\Delta Y_t = \beta_0 + \beta_1 t + \Pi Y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \epsilon_t$. Johansen's formulation allows for the possibility of quadratic trends in Y_t, which are ruled out in our formulation of d_t. For more discussion, see Johansen (1992a).

3. There are many repeated entries in Table 1. For example, as already noted, when $r_{a_u} = 0$, the Case (2) and Case (3) critical values are identical. Furthermore, within each case, the critical values are the same for all combinations of r_{a_k} and r_{a_u} with $r_{a_k} + r_{a_u} = n - r_{o_u}$. In this situation when $r_{o_u} = 0$, these hypotheses all correspond to $H_o: \Pi = 0$ in equation (2.2). There are a number of other examples of identical critical values that are not listed here.

4. These power curves were computed using 10,000 replications and $T = 1{,}000$.

5. These parameter values were calculated using consumption and output from the Citibase Database Services, spanning the quarters 1947:1 through 1990:4, and are in constant (1987) dollar, per capita terms. The consumption series is the sum of consumption expenditures on nondurables and services. The output series corresponds to gross, private sector, nonresidential, and domestic product and is constructed as gross domestic product minus farm, nonfarm housing, and government production.

6. We thank Robert Hodrick for making the data available to us.

7. Evans and Lewis (1992) using monthly data over the 1975–1989 period also found estimates of cointegrating vectors very close to $(1\ -1)$. While their estimated standard errors suggest that the cointegrating vectors may be different from $(1\ -1)$, Evans and Lewis argued that this arises from large outliers or "regime shifts" that are evident in the data (see Figure 5). Recent work on robust estimation of cointegrating vectors reported in Phillips (1993) suggests potential efficiency gains for data sets such as the one examined here. Further work is required to determine how the presence of outliers affects the cointegration tests discussed here.

REFERENCES

Ahn, S.K. & G.C. Reinsel (1990) Estimation for partially nonstationary autoregressive models. *Journal of the American Statistical Association* 85, 813–823.

Anderson, T.W. (1951) Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327–351.

Bekaert, G. & R.J. Hodrick (1993) On biases in the measurement of foreign exchange risk premiums. *Journal of International Money and Finance* 12(2), 115–138.

Bobkowsky, M.J. (1983) *Hypothesis Testing in Nonstationary Time Series.* Ph.D. Thesis, Department of Statistics, University of Wisconsin.

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

Cavanagh, C.L. (1985) Roots Local to Unity. Manuscript, Department of Economics, Harvard University.

Chan, N.H. (1988) On parameter inference for nearly nonstationary time series. *Journal of the American Statistical Association* 83, 857–862.

Chan, N.H. & C.Z. Wei (1987) Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics* 15, 1050–1063.

Chan, N.H. & C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16(1), 367–401.

Davies, R.B. (1977) Hypothesis testing when a parameter is present only under the alternative. *Biometrika* 64, 247–254.

Davies, R.B. (1987) Hypothesis testing when a parameter is present only under the alternative. *Biometrika* 74, 33–43.

Elliott, G. (1993) Efficient Tests for a Unit Root When the Initial Observation Is Drawn from Its Unconditional Distribution. Manuscript, Harvard University.

Elliott, G., T.J. Rothenberg, & J.H. Stock (1995) Efficient tests of an autoregressive unit root. *Econometrica*, forthcoming.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276. Reprinted in R.F. Engle & C.W.J. Granger (eds.), *Long-Run Economic Relations: Readings in Cointegration*. New York: Oxford University Press, 1991.

Evans, M.D.D. & K.K. Lewis (1992) Are Foreign Exchange Rates Subject to Permanent Shocks. Manuscript, University of Pennsylvania.

Hansen, B.E. (1990) Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis. Manuscript, University of Rochester.

Hansen, B.E. (1993) Testing for Unit Roots Using Covariates. Manuscript, University of Rochester.

Horvath, M. & M. Watson (1995) Critical Values for Likelihood Based Tests for Cointegration When Some of the Cointegrating Vectors Are Prespecified. Manuscript, Stanford University.

Johansen, S. (1988) Statistical analysis of cointegrating vectors. *Journal of Economic Dynamics and Control* 12, 231–254. Reprinted in R.F. Engle & C.W.J. Granger (eds.), *Long-Run Economic Relations: Readings in Cointegration*. New York: Oxford University Press, 1991.

Johansen, S. (1991) Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregression models. *Econometrica* 59, 1551–1580.

Johansen, S. (1992a) Determination of cointegration rank in the presence of a linear trend. *Oxford Bulletin of Economics and Statistics* 54, 383–397.

Johansen, S. (1992b) The Role of the Constant Term in Cointegration Analysis of Nonstationary Variables. Preprint 1, Institute of Mathematical Statistics, University of Copenhagen.

Johansen, S. & K. Juselius (1990) Maximum likelihood estimation and inference on cointegration—With applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52(2), 169–210.

Johansen, S. & K. Juselius (1992) Testing structural hypotheses in a multivariate cointegration analysis of the PPP and UIP of UK. *Journal of Econometrics* 53, 211–244.

Kremers, J.J.M., N.R. Ericsson, & J.J. Dolado (1992) The power of cointegration tests. *Oxford Bulletin of Economics and Statistics* 54(3), 325–348.

Ng, S. & P. Perron (1993) Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag. Manuscript, University of Montreal.

Osterwald-Lenum, M. (1992) A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics. *Oxford Bulletin of Economics and Statistics* 54, 461–471.

Park, J.Y. (1990) Maximum Likelihood Estimation of Simultaneous Cointegrating Models. Manuscript, Institute of Economics, Aarhus University.

Park, J.Y. & P.C.B. Phillips (1988) Statistical inference in regressions with integrated regressors I. *Econometric Theory* 4, 468–497.

Phillips, P.C.B. (1987) Toward a unified asymptotic theory for autoregression. *Biometrika* 74, 535–547.

Phillips, P.C.B. (1988) Multiple regression with integrated regressors. *Contemporary Mathematics* 80, 79–105.

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59(2), 283–306.

Phillips, P.C.B. (1993) Robust Nonstationary Regression. Manuscript, Cowles Foundation, Yale University.

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.
Sims, C.A., J.H. Stock, & M.W. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58(1), 113–144.
Stock, J.H. (1991) Confidence intervals of the largest autoregressive root in U.S. macroeconomic time series. *Journal of Monetary Economics* 28, 435–460.
Tsay, R.S. & G.C. Tiao (1990) Asymptotic properties of multivariate nonstationary processes with applications to autoregressions. *Annals of Statistics* 18, 220–250.

APPENDIX

Proof of Theorem 1. To prove the theorem, it is useful to introduce two alternative representations for the model. The first is a triangular simultaneous equations model used by Park (1990); the second is Phillips's (1991) triangular moving average representation. The first representation is useful because it allows the test statistic to be written in a particularly simple form; the second representation is useful because it neatly separates the regressors into I(0) and I(1) components.

We begin by defining some additional notation. First, partition Y_t as $Y_t = (Y'_{1,t}\ Y'_{2,t}\ Y'_{3,t}\ Y'_{4,t})'$, where $Y_{1,t}$ is $r_{o_u} \times 1$, $Y_{2,t}$ is $r_{o_k} \times 1$, $Y_{3,t}$ is $r_{a_k} \times 1$, and $Y_{4,t}$ is $(n - r_{o_u} - r_{o_k} - r_{a_k}) \times 1$. Because the cointegration test statistic is invariant to nonsingular transformations on Y_t, we set $\alpha_{o_k} = [0\ I_{r_{ok}}\ 0\ 0]'$ and $\alpha_{a_k} = [0\ 0\ I_{r_{ak}}\ 0]'$, where these matrices are partitioned conformably with Y_t. Thus, $\alpha'_{o_k} Y_t = Y_{2,t}$ and $\alpha'_{a_k} Y_t = Y_{3,t}$. Without loss of generality, we write $\alpha'_{o_u} = [I_{r_{ou}}\ \omega_2\ \omega_3\ \omega_4]$ and $\alpha'_{a_u} = [0\ 0\ 0\ \tilde{\alpha}'_{a_u}]$, which ensures that the columns of $\alpha = [\alpha_{o_u}\ \alpha_{o_k}\ \alpha_{a_k}\ \alpha_{a_u}]$ are linearly independent. Finally, we assume that the true (but unknown) values of ω_2, ω_3, and ω_4 are 0. These normalizations imply that $u_t = (Y'_{1,t}\ Y'_{2,t})'$ denotes the I(0) components of Y_t and $v_t = (Y'_{3,t}\ Y'_{4,t})'$ denotes the I(1), noncointegrated components.

Using this notation, the VECM in equation (2.3) can be reparameterized as the simultaneous equations models

$$\Delta Y_{1,t} = \theta' Y_{t-1} + \beta_1 Z_t + \epsilon_{1,t}, \tag{A.1}$$

$$\Delta Q_t = \tilde{\delta}_a (H v_{t-1}) + \gamma' S_t + e_t, \tag{A.2}$$

where $Q_t = (Y'_{2,t}\ Y'_{3,t}\ Y'_{4,t})'$, $S_t = (\Delta Y'_{1t}\ Y'_{2,t-1}\ Z'_t)'$, and

$$H = \begin{bmatrix} I_{r_{ak}} & 0 \\ 0 & \tilde{\alpha}'_{a_u} \end{bmatrix}.$$

These equations follow from writing the first r_{o_u} equations in (2.3) as

$$\Delta Y_{1,t} = \delta_{1,o_u} \alpha'_{o_u} Y_{t-1} + \delta_{1,o_k} Y_{2,t-1} + \delta_{1,a_k} Y_{3,t-1} + \delta_{1,a_u} (\tilde{\alpha}'_{a_u} Y_{4,t-1}) + \beta_1 Z_t + \epsilon_{1,t} \tag{A.3}$$

and the last $(n - r_{o_u})$ equations as

$$\Delta Q_t = \delta_{Q,o_u}\alpha'_{o_u}Y_{t-1} + \delta_{Q,o_k}Y_{2,t-1} + \delta_{Q,a_k}Y_{3,t-1} + \delta_{Q,a_u}(\tilde{\alpha}'_{a_u}Y_{4,t-1}) + \beta_Q Z_t + \epsilon_{Q,t}. \quad \textbf{(A.4)}$$

In equation (A.1), the term $\theta' Y_{t-1}$ captures the effect of all of the error correction terms on $\Delta Y_{1,t}$. Because ω_2, ω_3, and ω_4 are unknown, θ is unrestricted. To obtain (A.2), equation (A.3) is solved for $\alpha'_{o_u}Y_{t-1}$ as a function of $\Delta Y_{1,t}$, the other error correction terms, Z_t, and $\epsilon_{1,t}$; this expression is then substituted into (A.4). Thus, for example, $e_t = \epsilon_{Q,t} - \delta_{Q,o_u}\delta_{1,o_u}^{-1}\epsilon_{1,t}$ in (A.2). In terms of reparameterized models (A.1) and (A.2), the only constraints on the parameters are those imposed by the null hypothesis: $H_o: \tilde{\delta}_a = 0$.

Equations (A.1) and (A.2) are useful because, for given $\tilde{\alpha}_{a_u}$, the parameters in (A.2) can be efficiently estimated by 2SLS using $C_t = (u'_{t-1}, v'_{t-1}, Z'_t)'$ as instruments. Thus, letting $Q = [Q_1\ Q_2 \cdots Q_T]'$, $V_{-1} = [v_0\ v_1 \cdots v_{T-1}]'$, $S = [S_1\ S_2 \cdots S_T]'$, $C = [C_1\ C_2 \cdots C_T]'$, $e = [e_1\ e_2 \cdots e_T]'$, $\hat{S} = C(C'C)^{-1}C'S$, and $M_{\hat{S}} = I - \hat{S}(\hat{S}'\hat{S})^{-1}\hat{S}'$, the Wald statistic for testing $H_o: \delta_a = 0$ using a fixed $\tilde{\alpha}_{a_u}$ *is*

$$\begin{aligned} W(\tilde{\alpha}_{a_u}) &= [\mathrm{vec}(\Delta Q' M_{\hat{S}} V_{-1} H')]' [(H V'_{-1} M_{\hat{S}} V_{-1} H')^{-1} \otimes \hat{\Sigma}_e^{-1}] [\mathrm{vec}(\Delta Q' M_{\hat{S}} V_{-1} H')] \\ &= [\mathrm{vec}(e' M_{\hat{S}} V_{-1} H')]' [(H V'_{-1} M_{\hat{S}} V_{-1} H')^{-1} \otimes \hat{\Sigma}_e^{-1}] [\mathrm{vec}(e' M_{\hat{S}} V_{-1} H')], \end{aligned} \quad \textbf{(A.5)}$$

where the second equality holds under H_o.

The asymptotic distribution of $\mathrm{Sup}_{\tilde{\alpha}_{au}} W(\tilde{\alpha}_{a_u})$ depends on the behavior of the regressors and instruments, which is readily deduced from the triangular moving average representation of the model

$$u_t = D_u(L)a_t + \mu_u, \quad \textbf{(A.6)}$$

$$\Delta v_t = D_v(L)a_t + \mu_v, \quad \textbf{(A.7)}$$

where $a_t = \Sigma_\epsilon^{-1/2}\epsilon_t$, where $\mu_u = 0$ in Case 1 and $\mu_v = 0$ in Case 1 and Case 2. Because the variables are generated by a finite order VAR, the matrix coefficients in the lag polynomials $D_u(L)$ and $D_v(L)$ eventually decay at an exponential rate. Because v_t is I(1) and not cointegrated, $D_v(1)$ has full row rank. Furthermore, the error term e_t in (A.2) can be written as $e_t = Da_t$, and $D_v(1)D'$ has full row rank because only the first differences of $Y_{1,t}$ enter (A.2).

The theorem now follows from applying standard results from the analysis of integrated regressors to the components $W(\tilde{\alpha}_{a_u})$ (see, e.g., Chan and Wei, 1988; Park and Phillips, 1988; Phillips, 1988; Sims, Stock, and Watson, 1990; Tsay and Tiao, 1990; or the comprehensive summary in Phillips and Solo, 1992). We now consider the theorem's three cases in turn.

Case 1. In this case, $\mu_u = 0$ and $\mu_v = 0$ in (A.6) and (A.7), and it is readily verified that

$$T^{-2}V'_{-1}M_{\hat{S}}V_{-1} = T^{-2}V'_{-1}V_{-1} + o_p(1) \quad \textbf{(A.8.i)}$$

$$T^{-1}V'_{-1}M_{\hat{S}}e = T^{-1}V'_{-1}e + o_p(1), \quad \textbf{(A.8.ii)}$$

$$\mathrm{plim}(\hat{\Sigma}_e) = \Sigma_e = DD' \quad \textbf{(A.8.iii)}$$

so that

$$W(\tilde{\alpha}_{a_u}) = [\text{vec}(T^{-1}e'V_{-1}H')]'[(T^{-2}HV'_{-1}V_{-1}H')^{-1} \otimes (DD')^{-1}] \times [\text{vec}(T^{-1}e'V_{-1}H')] + o_p(1).$$

From the partitioned inverse formula,

$$\begin{aligned}
&[\text{vec}(T^{-1}e'V_{-1}H')]'[(T^{-2}HV'_{-1}V_{-1}H')^{-1} \otimes (DD')^{-1}][\text{vec}(T^{-1}e'V_{-1}H')] \\
&\quad = [\text{vec}(T^{-1}e'V_{1,-1})]'[(T^{-2}V'_{1,-1}V_{1,-1})^{-1} \otimes (DD')^{-1}][\text{vec}(T^{-1}e'V_{1,-1})] \\
&\qquad + [\text{vec}(T^{-1}e'M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})]'[(T^{-2}\tilde{\alpha}'_{a_u}V'_{2,-1}M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})^{-1} \otimes (DD')^{-1}] \\
&\qquad \times [\text{vec}(T^{-1}e'M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})],
\end{aligned} \tag{A.9}$$

where $V_{1,-1}$ denotes the first r_{a_k} columns of V_{-1}, and $V_{2,-1}$ denotes the remaining $n - r_{o_u} - r_{o_k} - r_{a_k}$ columns. Letting D_1 denote the first r_{a_k} rows of $D_v(1)$,

$$\begin{aligned}
&[\text{vec}(T^{-1}e'V_{1,-1})]'[(T^{-2}V'_{1,-1}V_{1,-1})^{-1} \otimes (DD')^{-1}][\text{vec}(T^{-1}e'V_{1,-1})] \\
&\quad = \text{Trace}[(DD')^{-1/2}(T^{-1}e'V_{1,-1})(T^{-2}V'_{1,-1}V_{1,-1})^{-1}(T^{-1}V'_{1,-1}e)(DD')^{-1/2\prime}] \\
&\quad \Rightarrow \text{Trace}\Big[(DD')^{-1/2}\Big(D_1\int B\,dB'\,D'\Big)'\Big(D_1\int BB'D_1'\Big)^{-1}\Big(D_1\int B\,dB'\,F'\Big) \\
&\qquad \times (DD')^{-1/2\prime}\Big] \\
&\quad = \text{Trace}\Big[\Big(\int F_1\,dB_{1,n-r_{ou}}\Big)'\Big(\int F_1F_1'\Big)^{-1}\Big(\int F_1\,dB_{1,n-r_{ou}}\Big)\Big],
\end{aligned} \tag{A.10}$$

where $B(s)$ denotes an $n \times 1$ standard Brownian motion process, $F_1(s) = B_{1,r_{a_k}}(s)$ (the first r_{a_k} elements of $B(s)$), and the last equality denotes equality in distribution.

As shown in equation (2.7), maximizing the second terms in (A.9) over all values of α_{a_u} yields

$$\begin{aligned}
&\sup_{\tilde{\alpha}_{au}}[\text{vec}(T^{-1}e'M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})]'[(T^{-2}\tilde{\alpha}'_{a_u}V'_{2,-1}M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})^{-1} \otimes (DD')^{-1}] \\
&\qquad \times [\text{vec}(T^{-1}e'M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})] \\
&\quad = \sum_{i=1}^{r_{au}} \lambda_i(R)
\end{aligned} \tag{A.11}$$

where

$$R = (DD')^{-1/2}[T^{-1}e'M_{V_1}V_{2,-1}][T^{-2}V'_{2,-1}M_{V_1}V_{2,-1}]^{-1}[T^{-1}e'M_{V_1}V_{2,-1}]'(DD')^{-1/2}. \tag{A.12}$$

Using notation borrowed from Phillips and Hansen (1990), R is readily seen to converge to

$$R \Rightarrow \Big(\int F_2\,dB'_{1,n-r_{ou}}\Big)'\Big(\int F_2F_2'\Big)^{-1}\Big(\int F_2\,dB'_{1,n-r_{ou}}\Big), \tag{A.13}$$

where $F_2(s) = F_3(s) - \gamma F_1(s)$, with $\gamma = [\int F_3F_1'][\int F_1F_1']^{-1}$ where $F_3(s) = B_{r_{a_k}+1,n-r_o}(s)$. Case (1) of the theorem follows from (A.10) and (A.13).

Case 2. In Case (2), $\mu_u \neq 0$ but $\mu_v = 0$. Letting $\bar{V}_{-1} = T^{-1} \sum v_{t-1}$, the proof follows as in Case (1) with $(V_{-1} - \bar{V}_{-1})$ replacing V_{-1} in (A.8)-(A.12) and $\beta^{\mu}(s)$ replacing $B(s)$ in limiting representations (A.10) and (A.13).

Case 3. In Case (3), both μ_u and $\mu_v \neq 0$. However, because $E(\alpha'_{a_k} Y_t) = 0$ is assumed in Case 3, the first r_{a_k} elements of $\mu_v = 0$. Thus, the first term of the statistic (the analog of (A.10)) is identical to the corresponding term in Case 2. The last $n - r_{o_u} - r_{o_k} - r_{a_k}$ elements of v_t contain a linear trend, and so, appropriately transformed, this set of regressors behaves like a single time trend and $n - r_{o_u} - r_{o_k} - r_{a_k} - 1$ martingale components. With this modification, the result for Case (3) follows as in Case (2).