

# Working Paper Series

## Testing for Cointegration When Some of the Cointegrating Vectors are Known

Michael T. K. Horvath and Mark W. Watson

Working Papers Series  
Macroeconomic Issues  
Research Department  
Federal Reserve Bank of Chicago  
December 1993 (WP-93-15)

FEDERAL RESERVE BANK  
OF CHICAGO

Testing for Cointegration When Some of the Cointegrating Vectors are Known

by

Michael T.K. Horvath  
Department of Economics, Northwestern University

and

Mark W. Watson  
Department of Economics, Northwestern University  
and Federal Reserve Bank of Chicago

First Draft: October 5, 1993  
This Draft: December 6, 1993

This paper has benefited from comments by Neil Ericsson, Gordon Kemp, Andrew Levin, Soren Johansen, John McDermott, Pierre Perron and James Stock, and was supported by grants from the National Science Foundation (SES-91-22463) and the Sloan Foundation.

## 1. Introduction

Economic models often imply that variables are cointegrated with simple 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 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 countrys' 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 have 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. Since 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 have low power.

This paper discusses an alternative procedure in which the null of no cointegration is tested against the composite alternative of cointegration with a prespecified cointegrating vector. As we will demonstrate, 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%-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 so 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.

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  $r_{0_k}$  known and  $r_{a_u}$  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  $r_{0_k} = r_{a_k} = 0$ . 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, since 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 which 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:

$$(2.1a) \quad Y_t = d_t + X_t$$

$$(2.1b) \quad X_t = \sum_{i=1}^p \Pi_i X_{t-i} + \epsilon_t$$

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),  $\epsilon_t$  is  $\text{NIID}(0, \Sigma_\epsilon)$ , and for convenience, the initial conditions  $X_{-i}$ ,  $i=0, \dots, p$  are assumed to equal zero. To focus attention on the long-run behavior of the process, it is useful to rewrite (2.1b) as:

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

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 hypotheses:

$$H_0: \text{rank}(\Pi) = r = r_0$$

$$H_a: \text{rank}(\Pi) = r = r_0 + r_a, \text{ with } r_a > 0.$$

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  $\Pi$  as  $\Pi = \delta\alpha'$ , where  $\delta$  and  $\alpha$  are  $n \times r$  matrices of full column rank, and the columns of  $\alpha$  denote the cointegrating vectors. The columns of  $\alpha$  are partitioned as  $\alpha = (\alpha_o \alpha_a)$ , where  $\alpha_o$  is an  $n \times r_o$  matrix whose columns are the cointegrating vectors present under the null,  $\alpha_a$  is an  $n \times r_a$  matrix whose columns are the additional cointegrating vectors present under the alternative. The matrix  $\delta$  is partitioned conformably as  $\delta = (\delta_o \delta_a)$ , where  $\delta_o$  is  $n \times r_o$  and  $\delta_a$  is  $n \times r_a$ . It is also useful to partition  $\alpha_a$  to isolate the known and unknown cointegrating vectors. Thus,  $\alpha_a = (\alpha_{a_k} \alpha_{a_u})$ , where the  $r_{a_k}$  columns of  $\alpha_{a_k}$  are the additional cointegrating vectors known under the alternative, and the  $r_{a_u}$  columns of  $\alpha_{a_u}$  are the additional cointegrating vectors that are present but unrestricted under the alternative. The matrix  $\delta_a$  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.

Calculating the LR and Wald Test Statistics when  $d_t=0$ :

The likelihood ratio statistic for testing  $H_0: r=r_{O_k}+r_{O_u}$  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  $\alpha_{O_k}$  and  $\alpha_{a_k}$ . We write the statistic as  $LR_{r_{O_k}, r_{a_k}}(\alpha_{O_k}, \alpha_{a_k})$ . The values of  $r_{O_k}$  and  $r_{a_k}$  appear implicitly as the ranks of  $\alpha_{O_k}$  and  $\alpha_{a_k}$  respectively. When  $r_{O_k}=0$ , the statistic is written as  $LR_{r_{O_k}, r_{a_k}}(0, \alpha_{a_k})$ , and as  $LR_{r_{O_k}, r_{a_k}}(\alpha_{O_k}, 0)$  when  $r_{a_k}=0$ .

To derive the LR statistic, we limit attention to the problem with  $r_0=r_{O_k}=r_{O_u}=0$ . For the purposes of deriving the computational formula for the LR statistic, this is without loss of generality since, in the general case, the LR statistic is identically

$$LR_{r_{O_k}, r_{a_k}}(\alpha_{O_k}, \alpha_{a_k}) \equiv LR_{0, r_{O_k}+r_{a_k}}(0, [\alpha_{O_k} \alpha_{a_k}]) - LR_{0, r_{O_k}}(0, \alpha_{O_k}).$$

With  $r_0=0$ , and ignoring the deterministic components,  $d_t$ , the model can be written as:

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

where  $\beta=(\Phi_1 \Phi_2 \dots \Phi_{p-1})$ , and  $Z_t=(\Delta Y'_{t-1} \Delta Y'_{t-2} \dots \Delta Y'_{t-p+1})'$ . In the context of (2.3) the null hypothesis  $H_0: r=0$  can be written as the composite null  $H_0: \delta_{a_k}=0, \delta_{a_u}=0$ .<sup>1</sup> It is convenient to discuss each part of this null separately, and thus 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  $H_0: r=0$  vs.  $H_a: r=r_{a_k}$ : When  $r_{a_u}=0$ , equation (2.3) simplifies to:

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

Since  $\alpha'_{a_k} Y_{t-1}$  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  $Y=[Y_1 Y_2 \dots Y_T]'$ ,  $Y_{-1}=[Y_0 Y_1 \dots Y_{T-1}]'$ ,  $\Delta Y=Y-Y_{-1}$ ,  $Z=[Z_1 Z_2 \dots Z_T]'$ ,  $\epsilon=[\epsilon_1 \epsilon_2 \dots \epsilon_T]'$ , and  $M_Z=[I-Z(Z'Z)^{-1}Z]$ , the OLS estimator of  $\delta_{a_k}$  is  $\hat{\delta}_{a_k}=(\Delta Y' M_Z Y_{-1} \alpha_{a_k})(\alpha'_{a_k} Y_{-1}' M_Z Y_{-1} \alpha_{a_k})^{-1}$  which corresponds to the Gaussian MLE. The corresponding Wald test statistic for  $H_0$  vs.  $H_a$  is:

$$(2.5) \quad W = [\text{vec}(\hat{\delta}_{a_k})]' [(\alpha_{a_k}' Y_{-1}' M_Z Y_{-1} \alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon]^{-1} [\text{vec}(\hat{\delta}_{a_k})] \\ = [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_k})]' [(\alpha_{a_k}' Y_{-1}' M_Z Y_{-1} \alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_k})].$$

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

The test statistic for  $H_0: r=0$  vs.  $H_a: r=r_{a_u}$ : The model simplifies to (2.4) with  $\delta_{a_u}$  and  $\alpha_{a_u}$  replacing  $\delta_{a_k}$  and  $\alpha_{a_k}$ . However, the analogue of the Wald statistic in (2.5) cannot be calculated since 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). Unknown cointegrating vectors complicate the testing problem because under the null hypothesis  $\delta_{a_u} = 0$ , so that the cointegrating vectors are unidentified. The problem can be avoided when  $r_a = n$ , since in this case  $\Pi$  is unrestricted under the alternative and the null and alternative become  $H_0: \Pi = 0$  vs.  $H_a: \Pi \neq 0$ . The problem cannot be avoided when the  $\text{rank}(\Pi) < n$  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  $\chi^2$  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  $\Pi$  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 the Wald statistic. This interpretation is useful here, because it suggests a simple way to compute the statistic. Since this form of the statistic appears as one component in the test statistic for the general  $r=r_{a_k} + r_{a_u}$



alternative, we derive it here.

Let LR denote the likelihood ratio statistic for testing  $H_0$  versus  $H_a$ , and let  $LR^*(\Sigma_\epsilon)$  denote the (infeasible) LR statistic that would be calculated if  $\Sigma_\epsilon$  were known. As usual,  $LR = LR^*(\hat{\Sigma}_\epsilon) + o_p(1)$  under  $H_0$  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  $\delta_{a_u}$ ,  $\alpha_{a_u}$ , and  $\Sigma_\epsilon$ , with  $\beta$  concentrated out, and let  $\hat{\delta}_{a_u}(\alpha_{a_u})$  denote the MLE of  $\delta_{a_u}$  for fixed  $\alpha_{a_u}$ . Then the well known relation between the Wald and LR statistic in the linear model implies that

$$(2.6) \quad W(\alpha_{a_u}) = 2[L(\hat{\delta}_{a_u}(\alpha_{a_u}), \alpha_{a_u}, \hat{\Sigma}_\epsilon) - L(0, \alpha_{a_u}, \hat{\Sigma}_\epsilon)] \\ = 2[L(\hat{\delta}_{a_u}(\alpha_{a_u}), \alpha_{a_u}, \hat{\Sigma}_\epsilon) - L(0, 0, \hat{\Sigma}_\epsilon)]$$

where  $W(\alpha_{a_u})$  is the Wald statistic in (2.5) written as a function of  $\alpha_{a_u}$ , and the second equality follows since  $\alpha_{a_u}$  does not enter the likelihood when  $\delta_{a_u} = 0$ . Thus:

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

where the Sup is taken over all  $n \times r_a$  matrices  $\alpha_a$ .

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

$$(2.7) \quad W(\alpha_{a_u}) = [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_u})]' [(\alpha_{a_u}' Y_{-1}' M_Z Y_{-1} \alpha_{a_u})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_u})] \\ = \text{TR}[\hat{\Sigma}_\epsilon^{-1/2} (\Delta Y' M_Z Y_{-1} \alpha_{a_u}) (\alpha_{a_u}' Y_{-1}' M_Z Y_{-1} \alpha_{a_u})^{-1} (\alpha_{a_u}' M_Z Y_{-1}' \Delta Y) \hat{\Sigma}_\epsilon^{-1/2}] \\ = \text{TR}[\hat{\Sigma}_\epsilon^{-1/2} (\Delta Y' M_Z Y_{-1}) D D' (M_Z Y_{-1}' \Delta Y) \hat{\Sigma}_\epsilon^{-1/2}], \text{ where } D = \alpha_{a_u} (\alpha_{a_u}' Y_{-1}' M_Z Y_{-1} \alpha_{a_u})^{-1/2} \\ = \text{TR}[D' (Y_{-1}' M_Z \Delta Y) \hat{\Sigma}_\epsilon^{-1} (\Delta Y' M_Z Y_{-1}) D] \\ = \text{TR}[F' C C' F],$$

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

$$(2.8) \quad \text{Sup}_{\alpha_u} W(\alpha_{a_u}) = \text{Sup}_{F, F=I} \text{TR}[F'(CC')F] = \sum_{i=1}^{r_{au}} \lambda_i(CC') \\ = \text{LR}^*(\hat{\Sigma}_\epsilon) = \text{LR} + o_p(1),$$

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

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  $\text{LR} = -T \sum_{i=1}^{r_{au}} \ln[1 - \gamma_i]$ , where  $\gamma_i$  are the ordered squared canonical correlations between  $\Delta Y_t$  and  $Y_{t-1}$ , after controlling for  $\Delta Y_{t-1}, \dots, \Delta Y_{t-p+1}$ . Since  $\gamma_i = \lambda_i(S'S)$ , where  $S'S = (\Delta Y' M_Z \Delta Y)^{-1/2} (\Delta Y' M_Z Y_{-1})(Y_{-1}' M_Z Y_{-1})^{-1} (Y_{-1}' M_Z \Delta Y) (\Delta Y' M_Z \Delta Y)^{-1/2}$ , (Brillinger (1980), Ch.10),  $\text{LR} = -T \sum_{i=1}^{r_{au}} \ln[1 - \lambda_i(S'S)] = T \sum_{i=1}^{r_{au}} \lambda_i(S'S) + o_p(1) = \sum_{i=1}^{r_{au}} \lambda_i(TS'S) + o_p(1)$ . Finally, since  $T(S'S) = \tilde{\Sigma}_\epsilon^{-1/2} (\Delta Y' M_Z Y_{-1})(Y_{-1}' M_Z Y_{-1})^{-1} (Y_{-1}' M_Z \Delta Y) \tilde{\Sigma}_\epsilon^{-1/2}$ , where  $\tilde{\Sigma}_\epsilon = T^{-1} (\Delta Y' M_Z \Delta Y)$ , this expression is identical to (2.7), except that  $\Sigma_\epsilon$  is estimated under the null.

The test statistic for  $H_0: r=0$  vs.  $H_a: r=r_{a_k} + r_{a_u}$ : The model now has the general form of (2.3).

As above, the LR statistic can be approximated up to an  $o_p(1)$  term by maximizing the Wald statistic over the unknown parameters in  $\alpha_{a_u}$ . 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 the regression (2.8). The Wald statistic (as a function of  $\alpha_{a_k}$  and  $\alpha_{a_u}$ ) can be written as:

$$(2.9) \quad W(\alpha_{a_k}, \alpha_{a_u}) = [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_k})]' [(\alpha_{a_k}' Y_{-1}' M_Z Y_{-1} \alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_k})] \\ + [\text{vec}(\Delta Y' M_{zk} Y_{-1} \alpha_{a_u})]' [(\alpha_{a_u}' Y_{-1}' M_{zk} Y_{-1} \alpha_{a_u})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] [\text{vec}(\Delta Y' M_{zk} Y_{-1} \alpha_{a_u})]$$

The first term is identical to equation (2.5) above, and the second term is the same as (2.6), 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  $\alpha_{a_u}$ , we can restrict

attention to cointegrating vectors that are linearly independent of  $\alpha_{a_k}$ , so that the LR statistic is obtained by maximizing (2.9) over all  $n \times r_{a_u}$  matrices  $\alpha_u$  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  $\alpha_{a_k}$ . Then  $\alpha_{a_u}$  can be written as a linear combination of the columns of  $G$ , so that  $\alpha_{a_u} = G \tilde{\alpha}_{a_u}$ , where  $\tilde{\alpha}_{a_u}$  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:

$$(2.10) \text{Sup}_{\alpha_{a_u}} W(\alpha_{a_k}, \alpha_{a_u}) = [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_k})]' [(\alpha_{a_k}' Y_{-1}' M_Z Y_{-1} \alpha_{a_k})^{-1} \otimes \hat{\Sigma}_\epsilon^{-1}] [\text{vec}(\Delta Y' M_Z Y_{-1} \alpha_{a_k})] \\ + \sum_{i=1}^{r_{a_u}} \lambda_i (H'H) \\ = LR + o_p(1),$$

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

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  $\Sigma_\epsilon$  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.

#### *Modifications Required For Nonzero Drift Component:*

When  $d_t \neq 0$  in (2.1a),  $Y_t$  is not directly observed, and the procedures outlined above 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.<sup>2</sup> In this case the VECM for  $y_t$  becomes:

$$(2.11) \quad \Delta y_t = \theta + \gamma t + \delta(\alpha' y_{t-1}) + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + \epsilon_t$$

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

There are three complications that arise when  $\mu_0$  or  $\mu_1$  are nonzero. First, as discussed in Johansen (1991), (1992a), (1992b) and Johansen and Juselius (1990), relationships between  $\mu_0$ ,  $\mu_1$  and the cointegrating vectors can lead to different interpretations of the drift parameters. For example, some linear combinations of  $\mu_0$  are related to initial conditions in the  $Y_t$  process, while other are related to means of the "error-correction" terms  $\alpha' Y_t$ . Second, these different interpretations can imply different trend properties of 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 (1992) show that different methods for detrending  $Y_t$  (associated with different estimators of  $\mu_0$  and  $\mu_1$ ) 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 the means of the error correction terms,  $\alpha' Y_t$ . Hence  $\gamma = 0$  in (2.11), but  $\theta \neq 0$ , but is constrained because it captures only the non-zero mean of the error correction terms  $\alpha' y_t$ . Imposing the constraint, leads to:

$$(2.12) \quad \Delta y_t = \delta(\alpha' y_{t-1} - \beta) + \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \epsilon_t$$

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  $\theta$  is unrestricted.

#### *Asymptotic Distribution of the Statistics:*

Above, the Gaussian likelihood ratio statistic for testing  $H_0: r = r_{O_k} + r_{O_u}$  vs.  $H_a: r = r_{O_k} + r_{a_k} + r_{O_u} + r_{a_u}$  was defined as  $LR_{r_{O_k}, r_{a_k}}(\alpha_{O_k}, \alpha_{a_k})$ . Let  $W_{r_{O_k}, r_{a_k}}(\alpha_{O_k}, \alpha_{a_k})$  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  $W_{r_o, r_a}(\alpha_{o_k}, \alpha_{a_k}) \equiv W_{0, r_o + r_a}(0, [\alpha_{o_k} \ \alpha_{a_k}]) - W_{0, r_o}(0, \alpha_{o_k})$ .

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

A. The data are generated by (2.1a)-(2.1c) with:

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

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

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

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

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

(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  $\alpha_{a_k}$  annihilates the deterministic drift in the series under both the null and the alternative.

The test statistic will be formed as described above, when  $d_t = 0$ . When  $d_t \neq 0$ , the VECM is augmented with a constant, and the statistic is calculated as above with  $Z_t$  in (2.3) augmented by a constant. Since, under assumption (A.4.iii), the constant term in the 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 the 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  $\alpha$  are known, the unconstrained estimator and test statistics are much easier to calculate than the constrained estimator; the required calculations when  $\alpha$  is known are discussed in Johansen and Juselius (1990) and Johansen (1991). Second, we show that when  $\alpha$  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:  $B(s) = (B_1(s) B_2(s) \dots B_n(s))'$  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;  $s^\mu = s - \int s = s - 1/2$  denote the demeaned time trend; and  $B_{i,j}(s) = (B_i(s), \dots B_j(s))'$  denote a  $(j-i+1) \times 1$  subvector of  $B(s)$ , and  $B_{i,j}^\mu$  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:

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

where  $k = n - r_{o_u}$ ,  $F_2(s) = F_3(s) - \gamma_{3,1} F_1(s)$  with  $\gamma_{3,1} = [\int F_1 F_1']^{-1} \int F_3 F_1'$ ,  $\lambda_i[\cdot]$  is the  $i$ 'th 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:

Case (1), Suppose that (A.1)-(A.3) and (A.4.i) hold, and the statistic is calculated with  $Z_t = (\Delta Y'_{t-1} \Delta Y'_{t-2} \dots \Delta Y'_{t-p+1})'$ , 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 (A.1)-(A.2) and (A.4.ii) hold, and the statistic is calculated with  $Z_t = (1 \Delta Y'_{t-1} \Delta Y'_{t-2} \dots \Delta Y'_{t-p+1})'$ , 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  $j = n - r_{o_k} - r_{o_u}$ .

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

**Proof:** See 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 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  $r_{o_u} = r_{a_u} = 0$ , 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 the VECM (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  $\alpha'_{o_1} \mu_1 = 0$  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. Since 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  $H\alpha_{a_u} = h$ , where  $H$  and  $h$  are known  $\ell \times n$  and  $\ell \times r_{a_u}$  matrices respectively.<sup>3</sup>

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  $T=1000$ . Extended critical values of  $n-r_{o_u} \leq 9$  are tabulated in Horvath and Watson (1993). 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).<sup>4</sup>

### 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, since *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 terms.

For tractability, most of the discussion will focus on a bivariate version of (2.11), with

$$\Phi_1 = \Phi_2 = \dots = \Phi_{p-1} = 0:$$

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

Since the likelihood based procedures are invariant to nonsingular transformations of  $Y_t$ , we set  $\alpha = (0 \ 1)'$  and  $\delta_1 = 0$ , so that the model becomes:

$$(3.2a) \quad \Delta y_{1,t} = \theta_1 + \epsilon_{1,t}$$

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

To investigate the local power of the tests, we suppose that  $\delta_2$  is local to zero; 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 (1987b, 1988) and Stock (1991). To rule out drift in the error correction term we set  $\theta_2 = 0$ , and for simplicity we set  $\Sigma_\epsilon = I_2$ .

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  $dB_{2,c}(s) = -cB_{2,c}(s) + dB_2(s)$ . 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  $D_c(s) = (s^\mu(s) \ B_{2,c}^\mu(s))$  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  $2 \times 1$  vector, and let  $\alpha = (0 \ 1)$  denote the true value of the cointegrating vector. Using the notation introduced above  $W_{0,1}(0, \gamma)$  denotes the test statistic for  $H_0: r = 0$  vs.  $H_a: r = r_{a_k} = 1$  constructed using  $\gamma$  as the cointegrating vector under the alternative; similarly  $W_{0,1}(0, 0)$  denotes the test statistic for  $H_0: r = 0$  vs.  $H_a: r = r_{a_u} = 1$ . The limiting distribution of these statistic is given by:



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

$$W_{0,1}(0,\gamma) \Rightarrow (\gamma' \int B_c dB')' (\gamma' \int B_c B_c' \gamma)^{-1} (\gamma' \int B_c dB');$$

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

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

$$W_{0,1}(0,\gamma) \Rightarrow (\gamma' \int B_c^\mu dB')' (\gamma' \int B_c^\mu B_c^{\mu'} \gamma)^{-1} (\gamma' \int B_c^\mu dB');$$

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

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

$$W_{0,1}(0,\gamma) \Rightarrow (\gamma' \int D_c dB')' (\gamma \int D_c D_c' \gamma')^{-1} (\gamma' \int D_c dB'), \text{ for } \gamma_1 = 0;$$

$$W_{0,1}(0,\gamma) \Rightarrow (\int s^\mu dB')' (\int (s^\mu)^2)^{-1} (\int s^\mu dB'), \text{ for } \gamma_1 \neq 0;$$

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

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 linear function of diffusion process in Cases (1) and (2) for all values of  $\gamma_1$ , 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).

In Figure 3.1, we plot the local power curves associated with these limiting random variables for  $\alpha = \gamma$ .<sup>5</sup> Thus, the  $W_{0,1}(0,\alpha_{2k})$  plot shows the power of the test that imposes the true value of the cointegrating vector, while the  $W_{0,1}(0,0)$  plots 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. Panel B of the figure also shows the local power of the LR analogue 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 analogue. The reduction in power for the case 2 LR statistic relative to case 1 arises because, under the null that  $\delta=0$ , the constant term  $\beta$  in (2.12) is unidentified. The LR statistic maximizes over this parameter, leading to an increase in critical value relative to case 1. The reduction in power for the  $W_{0,1}(0,0)$  statistic in case 2 relative to case 1 arises because the data are demeaned in case 2, leading to a reduction in the variance of the regressor. Apparently, the reduction in power from demeaning the data is less than the reduction from maximizing over the parameter  $\beta$ .

Since 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 3.1. For fixed alternatives, the  $W_{0,1}(0,0)$  and corresponding LR statistics are consistent. On the other hand, since  $\gamma'y_t$  is  $I(1)$  when  $\gamma$  is not proportional to  $\alpha$ , 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, since it applies in a situation when the power of the  $W_{0,1}(0,0)$  test is unity. An arguably more meaningful comparison obtains from the local-to-unity results where cointegration is weak. Figure 3.2 shows the power results for  $W_{0,1}(0,\gamma)$  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 non-trending data Cases 1 and 2; results for Case 3 will be discussed shortly. It is apparent from Figure 3.2 that for values of  $\gamma_1$  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  $\gamma_1 < .1$ . On the other hand, when  $\gamma_1 = 0.5$ , the  $W_{0,1}(0,0)$  test dominates for most values of  $c$ .

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  $\gamma_1$  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 results for Case 3 given above. When  $\gamma_1 \neq 0$  and  $c=0$  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  $\chi^2_2$  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  $\gamma$  and assuming  $c=0$  is  $P(\chi^2_2 > 10.18) = 0.6\%$ . The rejection probabilities for other values of  $c$  are also close to this value.

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  $\theta_1$  or  $\gamma_1$  local to zero, say  $\theta_1 = c_{\theta_1}/T^{1/2}$  or  $\gamma_1 = c_{\gamma_1}/T^{1/2}$ . Since these power functions depend critically on the assumed values of the constant  $c_{\theta_1}$  and  $c_{\gamma_1}$ , and since 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)-(3.2b) and values of  $\gamma_1$  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  $\theta_1 = .004$ ,  $\sigma_1 = .006$ ,  $\sigma_2 = .011$ ,  $\text{cor}(\epsilon_{1,t}, \epsilon_{2,t}) = 0.21$  and  $T = 174$ .<sup>6</sup> Results are shown for values of  $\gamma_1$  ranging from 0 to .10. For comparison with previous graphs  $\delta_2$  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. To highlight these differences, consider the VECM

$$(3.3) \quad \Delta Y_t = \delta \alpha' Y_{t-1} + \sum_{i=1}^{p-1} \Phi_i \Delta Y_{t-i} + \epsilon_t,$$

and let  $w_t = \alpha' Y_t$  and  $v_t = v' Y_t$  where  $v$  is chosen so that  $P = [\alpha \ v]$  has full rank. The elements of  $w_t$  are the I(0) error correction terms and the elements of  $v_t$  are the remaining I(1) non-cointegrated linear combinations of the data. After multiplying (3.3) by  $P'$ , the system can be partitioned as:

$$(3.4a) \quad \Delta w_t = \rho w_{t-1} + \sum_{i=1}^{p-1} \phi_{ww,i} \Delta w_{t-i} + \sum_{i=1}^{p-1} \phi_{wv,i} \Delta v_{t-i} + e_{w,t}$$

$$(3.4b) \quad \Delta v_t = \xi w_{t-1} + \sum_{i=1}^{p-1} \phi_{vw,i} \Delta w_{t-i} + \sum_{i=1}^{p-1} \phi_{vv,i} \Delta v_{t-i} + e_{v,t}$$

where  $\rho = \alpha' \delta$ ,  $\xi = v' \delta$ ,  $e_{w,t} = \alpha' \epsilon_t$ ,  $e_{v,t} = v' \epsilon_t$  and  $\phi_{ww,i}$ ,  $\phi_{wv,i}$ , etc., are the appropriate elements of  $P' \Phi_i$ . The multivariate tests considered in this paper test  $H_0: \delta = 0$  in (3.3), or equivalently  $H_0: \rho = 0$  and  $\xi = 0$  in (3.4a) and (3.4b). In contrast, univariate tests concentrate their attention on the equation for  $\Delta w_t$ , written as:

$$(3.5) \quad \Delta w_t = \rho w_{t-1} + u_t,$$

and test the null hypothesis  $H_0: \rho = 0$ .

Thus, the first major difference in the univariate and multivariate tests is the null hypothesis: univariate tests consider the simple null,  $\rho = 0$ , while multivariate tests consider the composite null,  $\rho = 0$  and  $\xi = 0$ . Since  $w_t$  is I(1) under the null, the lagged level of  $w_t$  will not help predict any I(0) variable after controlling for lagged first differences of  $w_t$ . The univariate tests look for predictive power for future values of  $\Delta w_t$ , while the multivariate tests look for predictive power for future values of all of the I(0) variables,  $\Delta w_t$  and  $\Delta v_t$ . The second difference between the univariate and multivariate tests is that the univariate tests typically use a one-sided alternative ( $\rho < 0$ ), while the multivariate tests consider two-sided alternatives. The third major difference is

the conditioning set used to estimate  $\rho$  in (3.5) or (3.4a). The multivariate tests include lagged values of  $\Delta w_t$  and  $\Delta v_t$  in the regression; univariate procedures, such as augmented Dickey-Fuller regression, include only lags of  $\Delta w_t$ . Thus, when lags of  $\Delta v_t$  help predict  $\Delta w_t$ , the error term in the multivariate regression will have smaller error than the error term in the univariate regression. This leads to a more efficient estimator of  $\rho$  and a more powerful test.

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 the potential power gains and losses associated with the first two differences in the univariate and multivariate procedures. We do this in the context of the bivariate example introduced in (3.1) and (3.2) above, except now  $\delta_1 \neq 0$ , so that (3.2a) becomes:

$$(3.2a') \quad \Delta y_{1,t} = \delta_1 y_{2,t-1} + \epsilon_{1,t}.$$

As before,  $y_{2,t}$  is generated by (3.2b) with  $\delta_1 = -c/T$ , and the errors are mutually independent iid(0,1). Since lagged first differences do not appear in the model and since the errors are mutually uncorrelated,  $E(\epsilon_{2,t} | \{\Delta y_{1,t}\}) = 0$ , so that the standard univariate Dickey-Fuller test for (3.2b) and the generalization of this procedure suggested by Hansen (1993) coincide.

We consider the local power of the one-sided Dickey-Fuller univariate t-test, the corresponding two-sided test based on the squared Dickey-Fuller t-test statistic, and the  $W(0,\alpha)$  statistics. The power functions for  $W(0,\alpha)$  are calculated for various values of  $\delta_1$  local to zero, parameterized as  $\delta_1 = (\kappa c)/T$ . (When  $\delta_1$  is not  $T^{-1}$  local to zero, e.g.,  $\delta_1 = \kappa c$  for fixed non-zero values of  $\kappa$  and  $c$ , then the asymptotic power of the  $W(0,\alpha)$  test is unity.) Figure 3.4 plots the power functions. We make two remarks. 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  $\kappa$ . When  $\kappa = 0$ , 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. On the other hand, when  $\kappa$  is large, the  $W(0,\alpha)$  statistic dominates the Dickey-Fuller test. For example, when  $\kappa=30$ , the power gain in the  $W(0,\alpha)$  statistic relative to the univariate test corresponds to a sample size increase of approximately 35% at 50% power.

#### 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, 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 two-day/non-holiday delivery lag for spot market exchange orders as described in Bekaert and Hodrick (1993).<sup>7</sup> 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 4.1.

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. (See Ng and Perron (1993) for an analysis of step-down testing in the context of testing for unit roots.) Results for testing for cointegration between forward and spot rates are presented in Table 4.1. 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  $\alpha$  is unspecified (denoted by  $W_{0,1}(0,0)$ ), the cointegrating vector estimated in this case (denoted by  $\hat{\alpha}_{a_k}$ ), and the ADF statistic calculated from the forward premium. All statistics are reported for the optimal lag length chosen via the step-down procedure. Constant 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)). Since nominal exchange rates exhibit some trending behavior over the sample period, the p-values for the  $W_{0,1}(0,0)$  statistic are reported for 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  $\alpha$ , differ little from the  $W_{0,1}(0,\alpha_{a_k})$  statistics. Their p-values are much greater however, since 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), at least to two decimal places.<sup>8</sup> 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.

### 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 one 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.

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 in Elliott, Rothenberg and Stock (1992) 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.

## 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 (1988); the second is Phillips' (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  $(n - r_{O_u} - r_{O_k} - r_{a_k}) \times 1$ . Since the cointegration test statistic is invariant to nonsingular transformations on  $Y_t$ , we set  $\alpha_{O_k} = [0 \ I_{r_{O_k}} \ 0 \ 0]'$  and  $\alpha_{a_k} = [0 \ 0 \ I_{r_{a_k}} \ 0]'$ , 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  $\alpha_{O_u}' = [I_{r_{O_u}} \ \omega_2 \ \omega_3 \ \omega_4]$  and  $\alpha_{a_u}' = [0 \ 0 \ 0 \ \tilde{\alpha}_{a_u}']$ , which insures 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  $\omega_2$ ,  $\omega_3$  and  $\omega_4$  are zero. 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  $I(1)$ , non-cointegrated components.

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

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

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

where  $Q_t = (Y_{2,t}' \ Y_{3,t}' \ Y_{4,t}')'$ ,  $S_t = (\Delta Y_{1,t}' \ Y_{2,t-1}' \ Z_t)'$ , and

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

These equations follow from writing the first  $r_{O_u}$  equations in (2.3) as:

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



and the last  $(n-r_{\alpha_u})$  equations as:

$$(A.4) \Delta Q_t = \delta_{Q,\alpha_u} \alpha_u' Y_{t-1} + \delta_{Q,\alpha_k} Y_{2,t-1} + \delta_{Q,\alpha_k} Y_{3,t-1} + \delta_{Q,\alpha_u} (\tilde{\alpha}_{\alpha_u}' Y_{4,t-1}) + \beta_Q Z_t + \epsilon_{Q,t}$$

In equation (A.1), the term  $\theta' Y_{t-1}$  captures the effect of all of the error correction terms on  $\Delta Y_{1,t}$ . Since  $\omega_2$ ,  $\omega_3$  and  $\omega_4$  are unknown,  $\theta$  is unrestricted. To obtain (A.2), equation (A.3) is solved for  $\alpha_u' Y_{t-1}$  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,  $e_t = \epsilon_{Q,t} - \delta_{Q,\alpha_u} \delta_{1,\alpha_u}^{-1} \epsilon_{1,t}$  in (A.2). In terms of the reparameterized model (A.1)-(A.2), the only constraints on the parameters are those imposed by the null hypothesis:  $H_0: \tilde{\delta}_a = 0$ .

Equations (A.1) and (A.2) are useful because, for given  $\tilde{\alpha}_{\alpha_u}$ , 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  $Q = [Q_1 \ Q_2 \ \dots \ Q_T]'$ ,  $V_{-1} = [v_0 \ v_1 \ \dots \ v_{T-1}]'$ ,  $S = [S_1 \ S_2 \ \dots \ S_T]'$ ,  $C = [C_1 \ C_2 \ \dots \ C_T]'$ ,  $e = [e_1 \ e_2 \ \dots \ e_T]'$ ,  $\hat{S} = C(C'C)^{-1}C'S$ , and  $M_S = I - \hat{S}(\hat{S}'\hat{S})^{-1}\hat{S}'$ , the Wald statistic for testing  $H_0: \tilde{\delta}_a = 0$  using a fixed  $\tilde{\alpha}_{\alpha_u}$ , is:

$$(A.5) \quad W(\tilde{\alpha}_{\alpha_u}) = [\text{vec}(\Delta Q' M_S V_{-1} H')] [(H V_{-1}' M_S V_{-1} H')^{-1} \otimes \hat{\Sigma}_e^{-1}] [\text{vec}(\Delta Q' M_S V_{-1} H')] \\ = [\text{vec}(e' M_S V_{-1} H')] [(H V_{-1}' M_S V_{-1} H')^{-1} \otimes \hat{\Sigma}_e^{-1}] [\text{vec}(e' M_S V_{-1} H')],$$

where the second equality holds under  $H_0$ .

To asymptotic distribution of  $\sup_{\tilde{\alpha}_{\alpha_u}} W(\tilde{\alpha}_{\alpha_u})$  depends on the behavior of the regressors and instruments, which is readily deduced from the triangular moving average representation of the model:

$$(A.6) \quad u_t = D_u(L)a_t + \mu_u$$

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

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. Since 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. Since  $v_t$  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 since 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  $W(\tilde{\alpha}_{a_u})$ . (For example, see 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

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

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

$$(A.8.iii) \quad \text{plim}(\hat{\Sigma}_e) = \Sigma_e = FF'$$

so that

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

From the partitioned inverse formula:

$$(A.9) \quad [\text{vec}(T^{-1}e'V_{-1}H')] [(T^{-2}HV_{-1}'V_{-1}H')^{-1} \otimes (D'D)^{-1}] [\text{vec}(T^{-1}e'V_{-1}H')] = \\ [\text{vec}(T^{-1}e'V_{1,-1})] [(T^{-2}V'_{1,-1}V_{1,-1})^{-1} \otimes (D'D)^{-1}] [\text{vec}(T^{-1}e'V_{1,-1})] + \\ [\text{vec}(T^{-1}e'M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})] [(T^{-2}\tilde{\alpha}'_{a_u}V'_{2,-1}M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})^{-1} \otimes (D'D)^{-1}] [\text{vec}(T^{-1}e'M_{V_1}V_{2,-1}\tilde{\alpha}_{a_u})]$$

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  $D_v(1)$ ,

$$(A.10) \quad [\text{vec}(T^{-1}e'V_{1,-1})] [(T^{-2}V'_{1,-1}V_{1,-1})^{-1} \otimes (D'D)^{-1}] [\text{vec}(T^{-1}e'V_{1,-1})] =$$

$$\begin{aligned} & \text{Trace}[(D'D)^{-1/2}(T^{-1}e'V_{1,-1})((T^{-2}V_{1,-1}'V_{1,-1})^{-1}(T^{-1}V_{1,-1}'e)(D'D)^{-1/2})] \\ & \Rightarrow \text{Trace}[(D'D)^{-1/2}(D_1 \int BdB'F')'(D_1 \int BB'D_1)^{-1}(D_1 \int BdB'F')(D'D)^{-1/2}] \\ & = \text{Trace}[(\int F_1dB_{1,n-r_{ou}})'(\int F_1F_1')^{-1}(\int F_1dB_{1,n-r_{ou}})] \end{aligned}$$

where  $B(s)$  denotes an  $n \times 1$  standard Brownian motion process,  $F_1(s) = B_{1,r_{ak}}(s)$  (the first  $r_{ak}$  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  $\alpha_{au}$  yields:

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

where,

$$\text{(A.12)} \quad R = (D'D)^{-1/2}[T^{-1}e'M_{V_1}V_{2,-1}][T^{-2}V_{2,-1}'M_{V_1}V_{2,-1}]^{-1}[T^{-1}e'M_{V_1}V_{2,-1}](D'D)^{-1/2}.$$

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

$$\text{(A.13)} \quad R \Rightarrow (\int F_2dB_{1,n-r_{ou}})'(\int F_2F_2')^{-1}(\int F_2dB_{1,n-r_{ou}})$$

where  $F_2(s) = F_3(s) - \gamma F_1(s)$ , with  $\gamma = [\int F_1F_1']^{-1}[\int F_1F_3']$  where  $F_3 = B_{r_{ak}+1,n-r_o}(s)$ . 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 the limiting representation (A.10) and (A.13).

Case 3: In Case (3), both  $\mu_u$  and  $\mu_v \neq 0$ . However, since  $E(\alpha_{ak}' Y_t) = 0$  is assumed in Case 3, the first  $r_{ak}$  elements of  $\mu_v = 0$ . Thus the first term of the statistic (the analogue 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).

## Footnotes

1. Formally, the restriction  $\text{rank}(\delta'_a \alpha_a) = r_a$  should be added to the alternative. Since 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 (1992b), 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  $d_t$ . See Johansen (1992b) for more discussion.
3. For example, when  $h=0$ , 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  $j = n - r_{O_u} - r_{a_k} - \ell$ . Similar results obtain when  $h \neq 0$ .
4. There are many repeated entries in Table 1. For example, as noted above, 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  $H_0: \Pi = 0$  in equation (2.2). There are a number of other examples of identical critical values that are not listed here.
5. These power curves were computed using 10,000 replications and  $T=1000$ .
6. These parameter values were calculated using consumption and output from the Citibase database, spanning the quarters 1947:1 through 1992:4, 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, domestic product and is constructed as gross domestic product minus farm, nonfarm housing, and government production.
7. We thank Robert Hodrick for making the data available to us.
8. Evans and Lewis (1993) using monthly data over the 1975-1989 period also find 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 argue that this arises from large outliers or "regime shifts" in the data.

## References

- Ahn, S.K. and G.C. Reinsel (1990), "Estimation for Partially Nonstationary Autoregressive Models," *Journal of the American Statistical Association*, 85, pp. 813-823.
- Anderson, T.W. (1951), "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 22, pp. 327-51.
- Bekaert, G. and R.J. Hodrick (1993), "On Biases in the Measurement of Foreign Exchange Risk Premiums", *Journal of International Money and Finance*, Vol. 12, No. 2, pp. 115-38.
- 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 Edition, Holden-Day, San Francisco.
- 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, pp. 857-62.
- Chan, N.H. and Wei, C.Z. (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics* 15, pp. 1050-63.
- Chan, N.H. and Wei, C.Z. (1988), "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *The Annals of Statistics* 16, no. 1, pp. 367-401.
- Davies, R. B. (1977), "Hypothesis Testing When a Parameter is Present Only Under the Alternative," *Biometrika*, Vol. 64, pp. 247-54.
- Davies, R. B. (1987), "Hypothesis Testing When a Parameter is Present Only Under the Alternative," *Biometrika*, Vol. 74, pp. 33-43.
- Elliott, Graham, Thomas J. Rothenberg and James H. Stock (1992), "Efficient Tests of an Autoregressive Unit Root," NBER Technical Working Paper 130.
- Elliott, Graham (1993), "Efficient Tests for a Unit Root When the Initial Observation is Drawn from Its Unconditional Distribution," manuscript, Harvard University.
- Engle, R.F. and C.W.J. Granger (1987), "Cointegration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, pp. 251-276. Reprinted in *Long-Run Economic Relations: Readings in Cointegration*, edited by R.F. Engle and C.W.J. Granger, Oxford University Press, New York, 1991.
- Engle, R.F. and B.S. Yoo (1987), "Forecasting and Testing in Cointegrated Systems," *Journal of Econometrics*, 35, pp. 143-59. Reprinted in *Long-Run Economic Relations: Readings in Cointegration*, edited by R.F. Engle and C.W.J. Granger, Oxford University Press, New York, 1991.
- Hansen, B.E. (1990b), "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, Michael and Mark Watson (1993), "Critical Values for Likelihood Based Tests for Cointegration When Some Cointegrating Vectors May be Known," manuscript, Northwestern University.
- Johansen, S., (1988), "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12, pp. 231-54. Reprinted in *Long-Run Economic Relations: Readings in Cointegration*, edited by R.F. Engle and C.W.J. Granger, Oxford University Press, New York, 1991.
- Johansen, S. (1991), "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregression Models," *Econometrica*, 59, 1551-1580.
- Johansen, S. (1992a), "The Role of the Constant Term in Cointegration Analysis of Nonstationary Variables," Preprint No. 1, Institute of Mathematical Statistics, University of Copenhagen
- Johansen, S. (1992b), "Determination of Cointegration Rank in the Presence of a Linear Trend," *Oxford Bulletin of Economics and Statistics*, 54, pp. 383-397.
- Johansen, S. and K. Juselius (1990), "Maximum Likelihood Estimation and Inference on Cointegration -- with Applications to the Demand for Money," *Oxford Bulletin of Economics and Statistics*, 52, no. 2, pp. 169-210.
- Johansen, S. and K. Juselius (1992), "Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and UIP of UK," *Journal of Econometrics*, 53, pp. 211-44.
- Kremers, J.J.M., N.R. Ericsson and J.J. Dolado (1992), "The Power of Cointegration Tests," *Oxford Bulletin of Economics and Statistics*, 54, no. 3, pp. 325-48.
- MacKinnon, James G. (1991), "Critical Values for Cointegration Tests," in R.F. Engle and C.W.J. Granger (eds) *Long-Run Economic Relations, Readings in Cointegration*, Oxford University Press: New York.
- Osterwald-Lenum, Michael (1992), "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics," *Oxford Bulletin of Economics and Statistics*, 54, pp. 461-71.
- Ng, Serena and P. Perron (1993). "Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag," manuscript, University of Montreal.
- Park, J.Y. (1990), "Maximum Likelihood Estimation of Simultaneous Cointegrating Models," manuscript, Institute of Economics, Aarhus University.
- Park, J.Y. and P.C.B. Phillips (1988), "Statistical Inference in Regressions with Integrated Regressors I," *Econometric Theory* 4, pp. 468-97.
- Phillips, P.C.B. (1987b), "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, pp. 535-47.



- Phillips, P.C.B. (1988), "Multiple Regression with Integrated Regressors," *Contemporary Mathematics*, 80, pp. 79-105.
- Phillips, P.C.B. (1991), "Optimal Inference in Cointegrated Systems," *Econometrica*, Vol. 59, No. 2, pp. 283-306.
- Phillips, P.C.B. and B.E. Hansen (1990), "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, pp. 99-125.
- Phillips, P.C.B. and S. Ouliaris, (1990), "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, pp. 165-94.
- Phillips P.C.B. and V. Solo (1992), "Asymptotics for Linear Processes," *Annals of Statistics*, 20, pp. 971-1001.
- Sims, C.A., J.H. Stock, and M.W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, Vol. 58, No. 1, pp. 113-44.
- Stock, James H. (1991), "Confidence Intervals of the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics* 28, pp. 435-60.
- Stock, J.H. and M.W. Watson (1988), "Testing for Common Trends," *Journal of the American Statistical Association*, 83, pp. 1097-1107. Reprinted in *Long-Run Economic Relations: Readings in Cointegration*, edited by R.F. Engle and C.W.J. Granger, Oxford University Press, New York, 1991.
- Tsay, R.S and G.C Tiao (1990), "Asymptotic Properties of Multivariate Nonstationary Processes with Applications to Autoregressions," *Annals of Statistics*, 18. pp. 220-50.

Table 2.1  
Critical Values for Tests for Cointegration

n-r	r <sub>ou</sub>	r <sub>ok</sub>	r <sub>ak</sub>	r <sub>au</sub>	----- Case 1 -----			----- Case 2 -----			----- Case 3 -----		
					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
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

Table 2.1 (Continued)  
Critical Values for Tests for Cointegration

n-r	r <sub>ou</sub>	r <sub>ok</sub>	r <sub>ak</sub>	r <sub>au</sub>	----- Case 1 -----			----- Case 2 -----			----- Case 3 -----		
					1%	5%	10%	1%	5%	10%	1%	5%	10%
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
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

Table 2.1 (Continued)  
 Critical Values for Tests for Cointegration

n-r	r <sub>ou</sub>	r <sub>ok</sub>	r <sub>ak</sub>	r <sub>au</sub>	----- Case 1 -----			----- Case 2 -----			----- Case 3 -----		
					1%	5%	10%	1%	5%	10%	1%	5%	10%
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

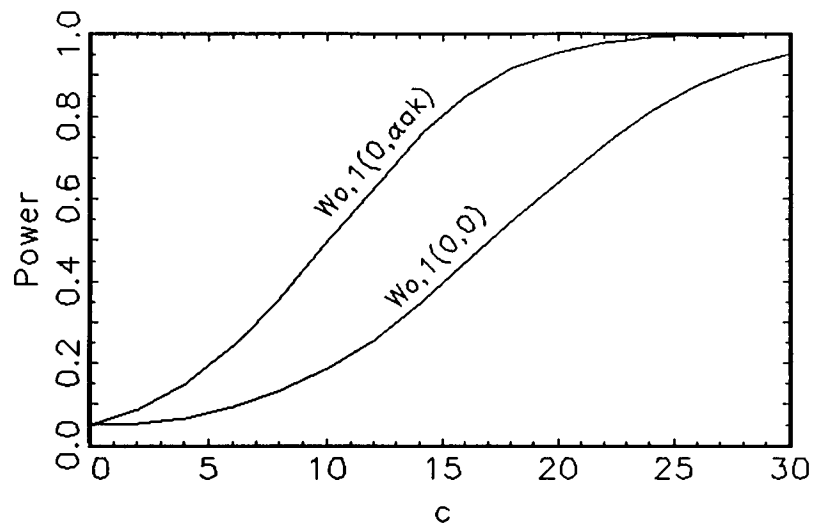
Table 4.1  
 Tests for Cointegration  
 Between Spot and Forward Exchange Rates  
 (Weekly Data, January 1975 - December 1989)

<u>Currency</u>	$\frac{W_{0,1}(0, \alpha_{a_k})}{}$	$\frac{W_{0,1}(0,0)}{}$	$\frac{\hat{\alpha}_{a_u}}{}$	$\frac{ADF}{}$
British Pound	10.95 (0.04)	10.97 (0.21)	[1 -1.001 (.004)]	-3.12 (0.03)
Swiss Franc	12.73 (0.02)	13.67 (0.08)	[1 -0.998 (.003)]	-3.33 (0.02)
German Mark	23.38 (<.01)	25.00 (<.01)	[1 -0.999 (.002)]	-3.58 (<.01)
Japanese Yen	15.00 (<.01)	15.02 (0.05)	[1 -1.001 (.003)]	-2.99 (0.04)

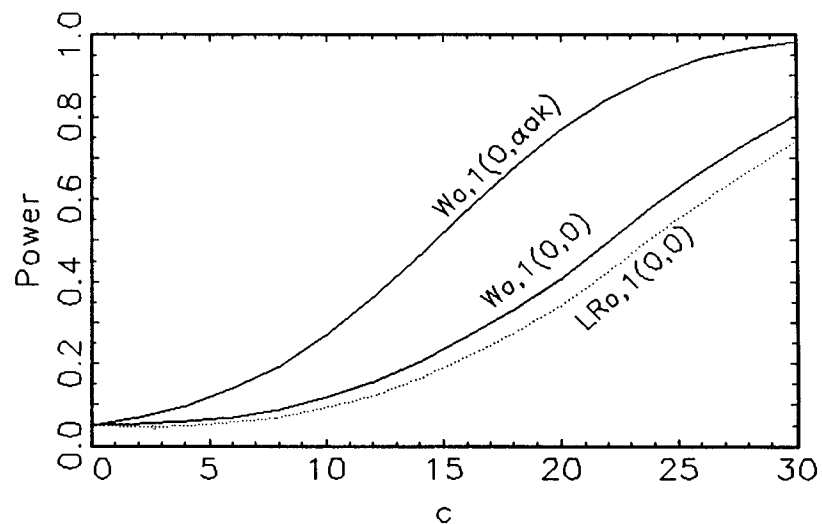
Notes: The statistics  $W_{0,1}(0, \alpha_{a_k})$  were calculated using  $\alpha_{a_k} = (1 \ -1)'$ . The numbers in parentheses next to the test statistics are p-values. The estimated cointegrating vector  $\hat{\alpha}_{a_u}$  is normalized as  $(1 \ \hat{\beta})$ , and the numbers in parentheses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

Figure 3.1  
Local Asymptotic Power

A. Case 1



B. Case 2



C. Case 3

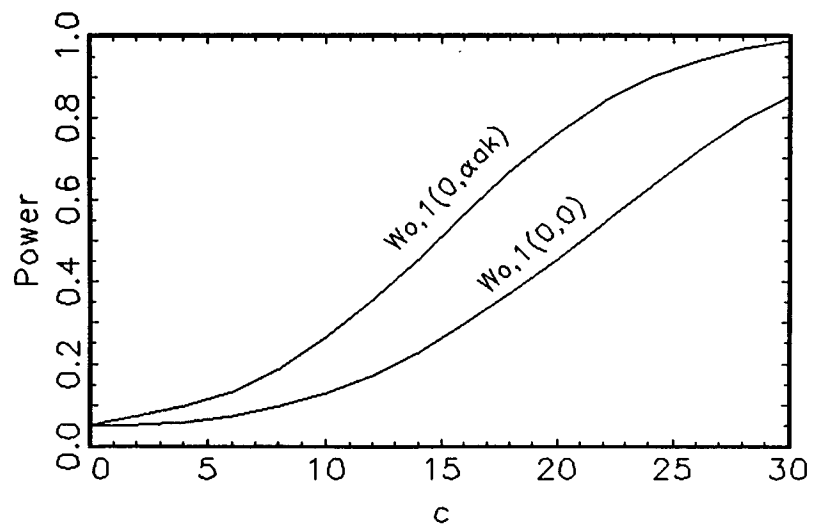


Figure 3.2  
 Local Asymptotic Power  
 Incorrectly Specified Cointegrating Vector

A. Case 1

B. Case 2

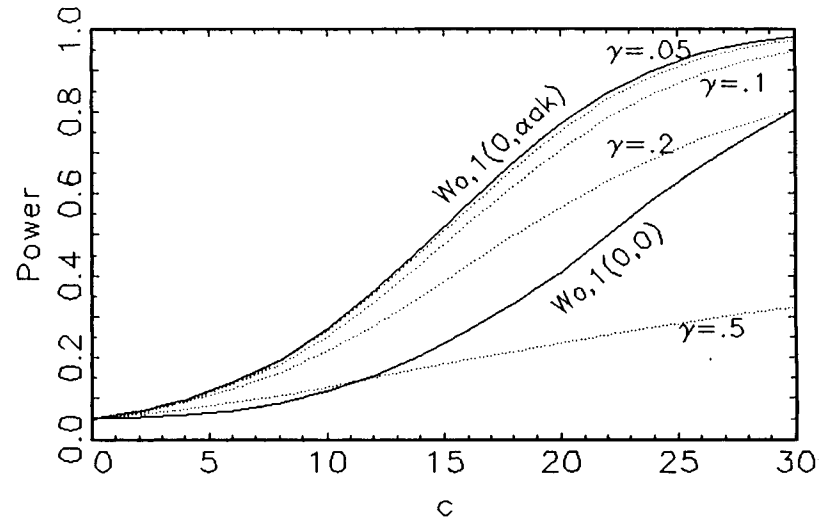
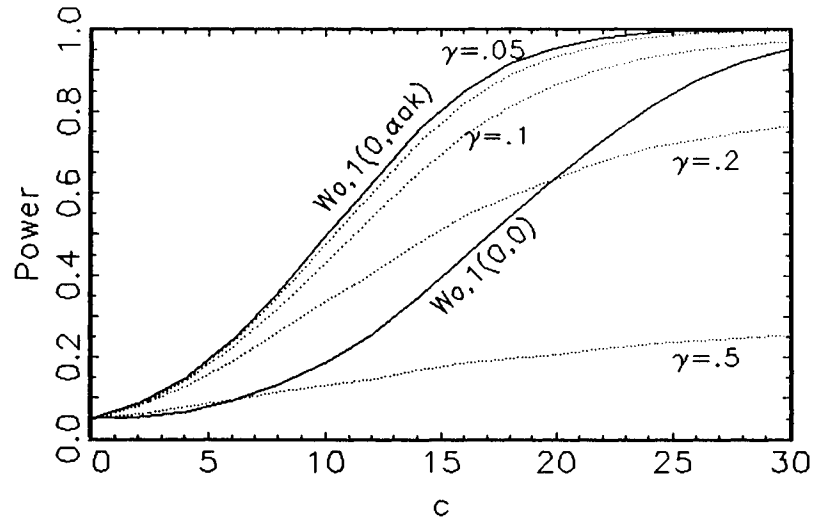


Figure 3.3  
 Power in the Income/Consumption System  
 Incorrectly Specified Cointegrating Vector

Consumption/Income System

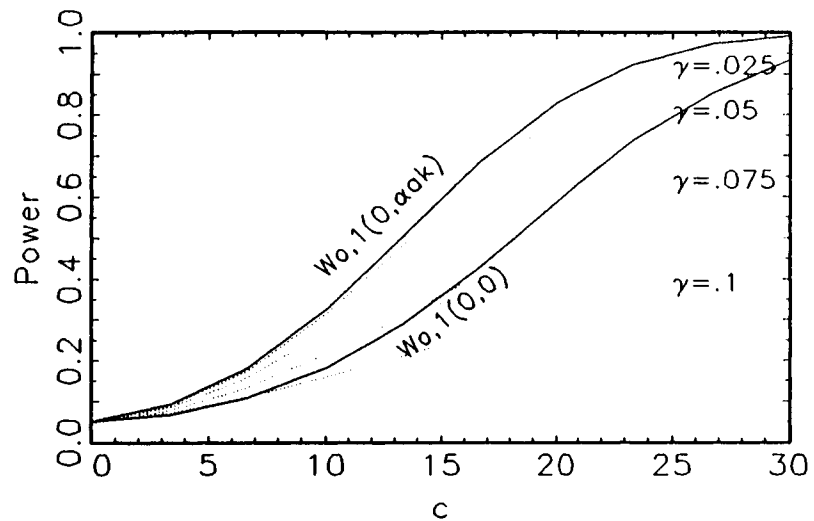


Figure 3.4  
 Local Asymptotic Power

Univariate and Multivariate Tests

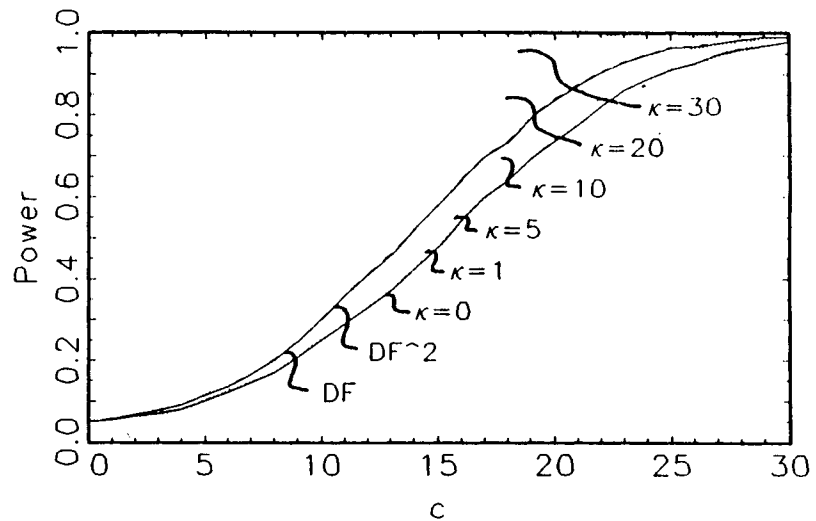




Figure 4.1  
Forward-Spot Premia

