Efficient detrending in cointegrating regression

Authors: Zhijie Xiao, Peter C.B. Phillips

Published in *Econometric Theory*, vol. 15, no. 4, pp. 519-548, August 1999

Econometric Theory, **15**, 1999, 519–548. Printed in the United States of America.

EFFICIENT DETRENDING IN COINTEGRATING REGRESSION

ZHIJIE XIAO
University of Illinois at Urbana-Champaign

PETER C.B. PHILLIPS
Cowles Foundation for Research in Economics, Yale University

This paper studies efficient detrending in cointegrating regression and develops modified tests for cointegration that use efficient detrending procedures. Asymptotics for these tests are derived. Monte Carlo experiments are conducted to evaluate the detrending procedures in finite samples and to compare tests for cointegration based on different detrending procedures. The limit theory allows for increasingly remote initial condition effects as the sample size goes to infinity.

1. INTRODUCTION

There is a large body of work on the theory of unit root tests and their multivariate versions in cointegrated time series. One of the directions in which the subject is presently moving is the development of tests with improved asymptotic properties (Schmidt and Phillips, 1992; Elliot, Rothenberg, and Stock, 1996; Hansen, 1995). Monte Carlo results (see, e.g., Phillips and Perron, 1988; Schwert, 1989; DeJong, Nankervis, Savin, and Whiteman, 1992) indicate that unit root tests often have low power against plausible trend stationary alternatives, and, in consequence, much recent effort has been devoted to the construction of more efficient tests.

One of the mechanisms for increasing the efficiency of these tests is related to point optimal test procedures. When the time series model is a Gaussian AR(1) with unit error variance, the Neyman–Pearson lemma can be used to construct the most powerful test of a unit root against a simple point alternative. King (1988) provided a general discussion of such point optimal invariant tests, and Dufour and King (1991) developed the family of exact most powerful invariant tests. Elliot et al. (1996) applied this idea in the context of unit root tests. When the times series involves a deterministic component, Elliot et al. (1996) showed that power gains can be obtained by detrending under the alternative hypothesis be-

Our thanks go to Katsuto Tanaka and a referee for helpful comments on earlier versions of this paper. We also thank Glena Ames for her skill and effort in keyboarding the first version of this paper. Xiao thanks the Cowles Foundation for support under a C. Anderson Fellowship and Phillips thanks the NSF for research support under grant SBR 94-22922. Address correspondence to: Peter C.B. Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, USA.

 0266-4666/99 $9.50

fore constructing the unit root tests. This detrending procedure is sometimes called generalized least squares (GLS) detrending in the literature. It is more accurate to describe it as detrending after quasi-differencing (QD) (see Phillips and Lee, 1996; Canjels and Watson, 1997) because full GLS is not used in the detrending regression. An analysis of the efficiency gains from QD detrending and its effects on the power of unit root tests has been given by Phillips and Lee (1996).

It is of considerable interest to extend the method of QD detrending to tests for cointegration in multivariable systems. The concept of cointegration was introduced by Granger (1981) and Granger and Weiss (1983) and more systematically studied in Engle and Granger (1987). In the last 10 years, a new body of statistical theory has developed for the analysis of cointegration (see, e.g., Stock, 1987; Stock and Watson, 1988; Phillips and Ouliaris, 1990; Phillips and Hansen, 1990; Phillips, 1991; Johansen, 1988, 1991; Johansen and Juselius, 1990; Sims, Stock, and Watson, 1990). For hypothesis testing, Engle and Granger (1987) suggested testing cointegration by examining whether or not the residuals from the cointegrating regression are stationary, and Phillips and Ouliaris (1990) studied asymptotic properties of the residual based tests. Stock and Watson (1988) proposed the "common trends" approach based on the fact that a vector time series cointegrated with order r can be written as the sum of $n - r$ common trends and an $I(0)$ component. Using the reduced rank regression technique, Johansen (1988, 1991, 1996) has studied likelihood inference based on a Gaussian error correction model and shown that the asymptotic distribution of the likelihood ratio tests for cointegration is determined by a generalized eigenvalue problem and has the form of a multivariate unit root distribution.

This paper applies the idea of QD detrending to tests in cointegrated systems. The key feature of efficient tests for a unit root is that the trend parameters be estimated under the alternative or, more specifically, some plausible alternative hypothesis. Applying the same idea in a cointegration test framework, we perform the detrending procedure under the alternative. Consider an n-dimensional VAR model

$$\Delta y_t = \Pi y_{t-1} + \Phi_1 \Delta y_{t-1} + \cdots + \Phi_k \Delta y_{t-k+1} + Bx_t + \varepsilon_t,$$

where x_t is a deterministic trend. The null hypothesis of interest is H_0 : there are $s = n - r$ unit roots in the system, where r is the dimension of the cointegration space. For alternatives that are distant from unit roots (i.e., y_t is trend stationary), ordinary least squares (OLS) detrending is asymptotically efficient by the Grenander and Rosenblatt (1957) theorem. For alternatives closer to unit roots, such as the local alternative H_c : there are $s = n - r$ roots that are local to unity of the form $1 + (c/n)$, QD detrending is more efficient (see Phillips and Lee, 1996; Canjels and Watson, 1997). When QD detrending is performed under H_c, we need to estimate the trend coefficients based on quasi-differencing in the $n - r$ directions where the data have roots that are local to unity and estimate the trend coefficients in other directions (i.e., stationary directions) by OLS regression.

Because the directions with unit roots and roots less than unity are not immediately distinguishable in the n-dimensional coordinate system in which the time series y_t is observed, we have to transform or rotate the coordinate system so that the $I(1)$ and $I(0)$ components are separated. The first step in our procedure therefore involves a preliminary estimate of the directions of stationarity so that the rotation matrix can itself be estimated. We next run the QD detrending regression on the transformed data. The final test statistics are then constructed based on a reduced rank regression with the detrended data.

The outline of this paper is as follows. The next section presents the model structure. Section 3 develops an efficient detrending procedure for cointegrated systems and gives the limit distribution of the efficient detrended time series. A modified test for cointegration against a local alternative hypothesis is proposed in Section 4, and its asymptotic theory is developed. Section 5 discusses some additional issues concerning the effects of alternative initializations. The results of a small Monte Carlo experiment are reported in Section 6. Section 7 concludes. Proofs are given in Section 8.

A word on notation before we begin. We use the symbol $\Rightarrow$ to signify weak convergence, and $\equiv$ to signify equality in distribution. The term $I(k)$ denotes integration of order k. All limits are taken as $T \to \infty$, unless otherwise specified.

2. STRUCTURE OF THE MODEL

Consider the following VAR process y_t:

$$\Pi(L)y_t = \underline{B}x_t + \varepsilon_t, \qquad t = 1,\ldots,T,$$

where x_t is a deterministic trend, say, $x_t = (1, t, \ldots, t^p)'$, and $\underline{B} = (B_0, \ldots, B_p)$, with $B_p = 0$. The error ε_t is an n-dimensional independent and identically distributed (i.i.d.) random vector with zero mean, variance matrix Λ, and finite fourth cumulant. The initial observations of y_t are taken to be $O_p(1)$. The VAR can be rewritten in the following form using a finite order BN (Beveridge and Nelson, 1981) decomposition:

$$-\Pi y_t + \Pi^*(L)(I - L)y_t = \underline{B}x_t + \varepsilon_t. \tag{1}$$

If y_t is a cointegrated system of order r, then $\text{rank}(\Pi) = r$ and the matrix Π can be decomposed as $\Pi = \alpha\beta'$, where α, β are $n \times r$ matrices of rank r.

Generalize the time series y_t by replacing the difference operator $I - L$ in (1) with a quasi-difference (QD) operator $I - AL$, giving

$$-\Pi y_t + \Pi^*(L)(I - AL)y_t = \underline{B}x_t + \varepsilon_t \tag{2}$$

or

$$\underline{\Pi}(L)y_t = \underline{B}x_t + \varepsilon_t,$$

where $A = \beta\beta' + \beta_\perp \exp(T^{-1}C)\beta_\perp'$, C is a constant matrix of dimension $s \times s$, and $\exp(\cdot)$ is the matrix exponential. (For some subsequent results, more specific assumptions will be made about the localizing parameter matrix C.) The matrix A in the QD operator in (2) can be written

$$A = I + \beta_\perp[\exp(T^{-1}C) - I]\beta_\perp' = I + T^{-1}\beta_\perp C\beta_\perp' + O(T^{-2}),$$

and therefore the QD operator $I - AL$ has a form that is locally different from the operator $I - L$ in the nonstationary directions, i.e., directions other than the cointegrating directions β.

Using the BN decomposition again, we obtain the following representation of (2):

$$-\Psi y_t + \Psi^*(L)\Delta y_t = \underline{B}x_t + \varepsilon_t, \tag{3}$$

where $\Psi = \alpha\beta' + \Pi^*(A - I) \approx \alpha\beta' + T^{-1}\Pi^*\beta_\perp C\beta_\perp'$, $\Pi^* = \Pi^*(1)$, $\Psi^*(L) = \Pi^*(L)A + \Pi^{**}(L)(I - A)$, and $\Pi^{**}(L)$ is defined as $\Pi^*(L) = \Pi^*(1) + \Pi^{**}(L)(I - L)$.

To fix ideas for our analysis, we make the following assumptions on y_t.

Assumption RRR.

(A1) $\{\varepsilon_t\}$ is a sequence of i.i.d. Gaussian vectors with variance matrix Λ.
(A2) The determinantal equation $|\underline{\Pi}(L)| = 0$ has roots on or outside the unit circle.
(A3) $A = \beta_\perp \exp(T^{-1}C)\beta_\perp' + \beta\beta'$, where α and β are orthonormalized matrices of full column rank r, $\beta_\perp$ is $n \times s$ matrix of full column rank that is normalized and orthogonal to β, $s = n - r$.
(A4) $\alpha_\perp'\Pi^*\beta_\perp$ has full rank s, where $\alpha_\perp$ is $n \times s$ matrix of full rank that is normalized and orthogonal to α.

In this model, the n-dimensional time series y_t has s large autoregressive roots that are near unity with localizing parameters that arise from the matrix $I - A = -T^{-1}\beta_\perp C\beta_\perp'$, whose rank is at most s. This is the multivariate version (see Phillips, 1987b, 1988, 1998) of the common univariate local-to-unit roots model. It generalizes the concept of cointegration to cases where individual time series in y_t are highly persistent, with roots close (but not necessarily equal) to unity, but certain linear combinations of the series are stationary. In particular, when $C = 0$, it reduces to the conventional cointegration model with the usual error correction representation. For simplicity, we assume in Assumption A3 that α and β are already orthonormalized. Otherwise we have to orthonormalize them and use another notation in constructing the rotation matrix.

Let $H = (\beta, \beta_\perp)$ and $G = (\alpha, \alpha_\perp)$. Transforming y_t by H', we get

$$\underline{y}_t = H'y_t = \begin{bmatrix} \beta' y_t \\ \beta_\perp' y_t \end{bmatrix} = \begin{bmatrix} \underline{y}_{1t} \\ \underline{y}_{2t} \end{bmatrix}.$$

Notice that $A \approx I + T^{-1}\beta_\perp C\beta_\perp'$ and rewrite (2) as

$$-\alpha\underline{y}_{1t} + \Pi^*(L)\beta(I-L)\underline{y}_{1t} + \Pi^*(L)\beta_\perp\Delta_c\underline{y}_{2t} = \underline{B}x_t + \varepsilon_t,$$

where $\Delta_c\underline{y}_{2t} = \underline{y}_{2t} - (I + C/T)\underline{y}_{2,t-1}$. Multiplying the preceding system by G', we then have

$$\begin{bmatrix} -I_r + \alpha'\Pi^*(L)\beta(I-L) & \alpha'\Pi^*(L)\beta_\perp \\ \alpha_\perp'\Pi^*(L)\beta(I-L) & \alpha_\perp'\Pi^*(L)\beta_\perp \end{bmatrix}\begin{bmatrix} \underline{y}_{1t} \\ \Delta_c\underline{y}_{2t} \end{bmatrix} = G'(\underline{B}x_t + \varepsilon_t).$$

Let

$$\Phi(L) = \begin{bmatrix} -I_r + \alpha'\Pi^*(L)\beta(I-L) & \alpha'\Pi^*(L)\beta_\perp \\ \alpha_\perp'\Pi^*(L)\beta(I-L) & \alpha_\perp'\Pi^*(L)\beta_\perp \end{bmatrix}.$$

Under Assumption RRR, $\Phi(L)$ is invertible. Let $K(L) = \Phi(L)^{-1}G'$ and partition this matrix into blocks as

$$\begin{array}{cl} \begin{pmatrix} K_1(L) \\ K_2(L) \end{pmatrix} & \begin{array}{l} r \\ n-r. \end{array} \end{array}$$

Then

$$\underline{y}_{1t} = K_1(L)(\underline{B}x_t + \varepsilon_t), \tag{4}$$

$$\Delta_c\underline{y}_{2t} = K_2(L)(\underline{B}x_t + \varepsilon_t). \tag{5}$$

Thus, $\underline{y}_t$ can be decomposed into the sum of a deterministic component, $\underline{\Gamma}x_t$, and a stochastic component $\underline{y}_t^s$ as follows:

$$\underline{y}_t = \underline{y}_t^s + \underline{\Gamma}x_t,$$

or, in its partitioned representation, as

$$\underline{y}_{1t} = \underline{y}_{1t}^s + \Gamma_1 x_t,$$

$$\underline{y}_{2t} = \underline{y}_{2t}^s + \Gamma_2 x_t,$$

where $\underline{y}_{1t}^s = K_1(L)\varepsilon_t$ and $\Delta_c\underline{y}_{2t}^s = K_2(L)\varepsilon_t$. In the original coordinate system,

$$y_t = H\underline{y}_t = y_t^s + \Gamma x_t. \tag{6}$$

Remark 1. In the transformed time series $\underline{y}_t$, the stationary component and nonstationary component are separated so that the first r elements, $\underline{y}_{1t}$, are $I(0)$ and the next s elements, $\underline{y}_{2t}$, are nearly $I(1)$ and not cointegrated.

Remark 2. When $C = 0$, the time series has the following conventional Gaussian error correction representation:

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{K-1}\Phi_i\Delta y_{t-i} + \underline{B}x_t + \varepsilon_t.$$

Remark 3. If the model satisfies the condition that $\beta'\Gamma = 0$, then the same cointegrating vector that eliminates stochastic nonstationarity also eliminates deterministic nonstationarity. The random vector y_t is said to be deterministically cointegrated in this case. Otherwise, y_t is stochastically cointegrated, to use the terminology of Park (1992).

3. EFFICIENT DETRENDING IN A COINTEGRATED SYSTEM

We consider two detrending procedures for time series $y_t = y_t^s + \Gamma x_t$ defined by (2) in Section 2. In this section, we assume that the matrix $C = \text{diag}[c,\dots,c]$, where c is a constant. The first detrending regression is the OLS regression of time series y_t on the deterministic trend x_t:

$$y_t = \hat{\Gamma}x_t + \text{residual}. \tag{7}$$

The estimated trend coefficient matrix is $\hat{\Gamma} = Y'X(X'X)^{-1}$, where $Y' = (y_1,\dots,y_t,\dots,y_T)$, $X' = (x_1,\dots,x_t,\dots,x_T)$, and the OLS detrended time series is $\hat{y}_t^s = y_t - \hat{\Gamma}x_t$.

The second detrending procedure, which we call QD detrending, is based on the following regression of the transformed time series $\underline{y}_t$:

$$\begin{bmatrix} \underline{y}_{1t} \\ \Delta_c \underline{y}_{2t} \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma}_1 & 0 \\ 0 & \tilde{\Gamma}_2 \end{bmatrix} \begin{bmatrix} x_t \\ \Delta_c x_t \end{bmatrix} + \text{residual}. \tag{8}$$

The estimated trend coefficient matrices are

$$\tilde{\Gamma}_1 = \underline{Y}_1' X(X'X)^{-1}, \tag{9}$$

$$\tilde{\Gamma}_2 = \Delta_c \underline{Y}_2' \Delta_c X(\Delta_c X' \Delta_c X)^{-1}, \tag{10}$$

where

$$\underline{Y}_1' = (\underline{y}_{11},\dots,\underline{y}_{1t},\dots,\underline{y}_{1T}) = \beta' Y',$$

$$\Delta_c \underline{Y}_2' = (\Delta_c \underline{y}_{21},\dots,\Delta_c \underline{y}_{2t},\dots,\Delta_c \underline{y}_{2T}) = \beta_\perp' \Delta_c Y',$$

$$\Delta_c X' = (\Delta_c x_1,\dots,\Delta_c x_t,\dots,\Delta_c x_T).$$

The QD detrended time series can then be constructed as follows:

$$\tilde{y}_t^s = y_t - \tilde{\Gamma}x_t, \tag{11}$$

where

$$\underline{\tilde{\Gamma}} = \begin{bmatrix} \tilde{\Gamma}_1 \\ \tilde{\Gamma}_2 \end{bmatrix}, \qquad \tilde{\Gamma} = H\underline{\tilde{\Gamma}}.$$

Remark 4. The second detrending procedure, which uses the quasi-differenced data for the nonstationary component of the time series, is asymptotically more efficient than OLS detrending. An analysis of efficiency gains from QD under nonstationarity is given in Phillips and Lee (1996). To understand the efficiency gain of the detrending regression procedure that leads to the detrended time series, we can make a simple comparison between these two detrending procedures. Suppose $\underline{y}_{1t}^s = \underline{u}_{1t} \equiv I(0)$ and $\Delta_c \underline{y}_{2t}^s = \underline{u}_{2t} \equiv I(0)$. Assume that $T^{-1/2}\sum_{t=1}^{[\mathrm{Tr}]}\underline{u}_t \Rightarrow \underline{B}(r) = (\underline{B}_1(r)', B_2(r)')'$, where $\underline{u}_t = (\underline{u}_{1t}', \underline{u}_{2t}')'$, and suppose that $Dx_{[\mathrm{Tr}]} \Rightarrow X(r)$ and $F\Delta_c x_{[\mathrm{Tr}]} \Rightarrow X_c(r)$, where F and D are the corresponding standardizing matrices with $F = TD$, and formulae for $X_c(r)$ are given in Theorem 1, which follows. Notice that the OLS detrending procedure is equivalent to the detrending regression

$$\begin{bmatrix} \underline{y}_{1t} \\ \underline{y}_{2t} \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}_1 \\ \hat{\Gamma}_2 \end{bmatrix} x_t + \text{residual}. \tag{12}$$

The estimates of Γ_1 are identical for the two procedures. The difference occurs in the estimation of Γ_2. The asymptotic distributions of these two estimators are

$$T^{-1/2}(\hat{\Gamma}_2 - \Gamma_2)D^{-1} \Rightarrow \int_0^1 J_{2c}(r)X(r)'\left[\int_0^1 X(r)X(r)'\right]^{-1},$$

$$T^{-1/2}(\tilde{\Gamma}_2 - \Gamma_2)D^{-1} \Rightarrow \int_0^1 dB_2(r)X_c(r)'\left[\int_0^1 X_c(r)X_c(r)'\right]^{-1},$$

where $J_{2c}(r) = \int_0^r \exp\{(r-s)C\}\,dB_2(s)$.

As shown in Phillips and Lee (1996), an efficiency gain is achieved in the QD estimator because $\tilde{\Gamma}_2$ has smaller variance than $\hat{\Gamma}_2$. For instance, when $s = n - r = 1$ and $x_t = t$, the variance of the limit variates of $\tilde{\Gamma}_2$ is $V_{gls} = 3\sigma^2/(3 - 3c + c^2)$, whereas that of the OLS estimator is $V_{ols} = 3\sigma^2[3e^{2c}(c-1)^2 + 2c^3 + 3c^2 - 3]/2c^5$. The relative efficiency, $R_c = V_{ols}/V_{gls}$, is greater than 1 for all finite c.

Remark 5. The QD estimators we propose here ignore the $I(0)$ serial correlation associated with the stationary operator $K(L)$. This brings no loss of asymptotic efficiency by virtue of the Grenander–Rosenblatt theorem.

The asymptotics for $\hat{y}_t^s$ and $\tilde{y}_t^s$ are as follows.

THEOREM 1.

$$T^{-1/2}\hat{y}_{[\mathrm{Tr}]}^s \Rightarrow \underline{J}_c(r) \equiv J_c(r) - \int_0^1 J_c(s)X(s)'\left[\int_0^1 X(s)X(s)'\right]^{-1} X(r),$$

$$T^{-1/2}\tilde{y}_{[\mathrm{Tr}]}^s \Rightarrow \underset{\sim}{J}_c(r) = J_c(r) - \int_0^1 dB(s)X_c(s)'\left[\int_0^1 X_c(s)X_c(s)'\right]^{-1} X(r),$$

where $X(r) = (1, r, \ldots, r^p)$, $g(r) = (0, 1, 2r, \ldots, pr^{p-1})$, $X_c(r) = g(r) - cX(r)$, $J_c(r) = \int_0^r \exp\{(r-s)C\}\,dB(s)$, $B(r) = \beta_\perp(\alpha_\perp'\Pi^*\beta_\perp)^{-1}\alpha_\perp' B_\varepsilon(r)$, $B_\varepsilon(r)$ *is vector Brownian motion with covariance matrix* Λ, *and* $T^{-1/2}\sum_{t=1}^{[\mathrm{Tr}]}\varepsilon_t \Rightarrow B_\varepsilon(r)$.

4. A MODIFIED TEST OF $n-r$ UNIT ROOTS AGAINST A LOCAL ALTERNATIVE

A key feature of the efficient unit root tests in Elliot et al. (1996) is that the trend parameters are estimated under a plausible local alternative hypothesis. The same idea can be used in tests for cointegrated systems, where the detrending procedure can be performed under a plausible local alternative. However, in this case, the formulation is not as straightforward because of the multivariate nature of the procedure.

For the vector time series y_t defined by (2) (or (3)) in Section 2, we consider the null hypothesis of r cointegration vectors, $H_0 : \operatorname{rank}(\Psi) = r$, or equivalently, H_0 : there are $n-r$ unit roots, i.e., $C = 0$. When the alternative is "Ψ is of full rank," then y_t is (trend) stationary, and OLS detrending is asymptotically efficient under the alternative (Grenander and Rosenblatt, 1957).

Next consider the following local alternative, H_c : there are $n-r$ roots local to unity and $C = \operatorname{diag}[c_{r+1}, \dots, c_n]$. In this case, when we perform efficient detrending under H_c, we estimate the trend coefficients based on quasi-differenced data in the $n-r$ directions that are local to unity and estimate the trend coefficients in other directions by OLS. Under the null hypothesis

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{k-1} \Phi_i \Delta y_{t-i} + B x_t + \varepsilon_t,$$

where α, β are $n \times r$ matrices of rank r. As described in Section 2, we rotate[1] the coordinate system by premultiplying the preceding time series by H' and get the transformed data $\underline{y}_t = (\underline{y}_{1t}', \underline{y}_{2t}')'$. Here $\underline{y}_{1t}^s \equiv I(0)$, $\underline{y}_{2t}^s \equiv I(1)$ under H_0.

Under the local alternative H_c, the components of $\underline{y}_{2t}^s$ are nearly integrated, $\underline{y}_{1t}^s \equiv I(0)$, and we can write

$$\underline{y}_{2t}^s - (I + C/T)\underline{y}_{2,t-1}^s \equiv I(0).$$

Consider the following detrending regression:

$$\begin{bmatrix} \underline{y}_{1t} \\ \Delta_{\bar{c}} \underline{y}_{2t} \end{bmatrix} = \begin{bmatrix} \breve{\Gamma}_1 & 0 \\ 0 & \breve{\Gamma}_2 \end{bmatrix} \begin{bmatrix} x_t \\ \Delta_{\bar{c}} x_t \end{bmatrix} + \text{residual}, \tag{8'}$$

where $\Delta_{\bar{c}}$ denotes the QD operator $I - (I + \bar{C}/T)L$, with $\bar{C} = \operatorname{diag}[\bar{c}, \dots\dots, \bar{c}]$ for some $\bar{c}$. Then, $\breve{\Gamma}_1 = \underline{Y}_1' X (X'X)^{-1}$, $\breve{\Gamma}_2 = \Delta_{\bar{c}} \underline{Y}_2' \Delta_{\bar{c}} X (\Delta_{\bar{c}} X' \Delta_{\bar{c}} X)^{-1}$. Notice that if $c_{r+1} = c_{r+2} = \cdots = c_n = c$, and we set $\bar{c} = c$, then, under the alternative hypothesis, $\Delta_{\bar{c}} \underline{y}_{2t}^s \equiv I(0)$, and the detrending procedure is asymptotically efficient because we are now estimating the trend coefficient Γ_2 with a stationary error.

The detrended time series is $\breve{\underline{y}}_t = \underline{y}_t - \underline{\breve{\Gamma}} x_t$, and, in original coordinates, $\breve{y}_t = H\breve{\underline{y}}_t = y_t - \breve{\Gamma} x_t$, with

$$\breve{\Gamma} = H\breve{\underline{\Gamma}} = H\begin{bmatrix} \breve{\Gamma}_1 \\ \breve{\Gamma}_2 \end{bmatrix}.$$

THEOREM 2. *Under the null hypothesis,*

$$T^{-1/2}\breve{y}_{[\mathrm{Tr}]} \Rightarrow \underline{B}_{\bar{c}}(r) = B(r) - \int_0^1 dB_{\bar{c}}(s)X_{\bar{c}}(s)'\left[\int_0^1 X_{\bar{c}}(s)X_{\bar{c}}(s)'\right]^{-1} X(r),$$

where $X_{\bar{c}}(r) = g(r) - \bar{c}X(r)$, $B_{\bar{c}}(r) = B(r) - \bar{c}\int_0^r B(s)$.

Consider the following reduced rank regression based on the detrended data $\breve{y}_t$:

$$\Delta\breve{y}_t = \alpha\beta'\breve{y}_{t-1} + \sum_{i=1}^{k-1}\Phi_i\Delta\breve{y}_{t-i} + \breve{\varepsilon}_t.$$

To remove the nuisance parameters Φ_i, $i = 1,\dots,k-1$, let R_{0t} = residual from the regression of $\Delta\breve{y}_t$ on $\Delta\breve{y}_{t-1},\dots,\Delta\breve{y}_{t-k+1}$ and R_{1t} = residual from regression of $\breve{y}_{t-1}$ on $\Delta\breve{y}_{t-1},\dots,\Delta\breve{y}_{t-k+1}$. Then we can estimate α, β from a reduced rank regression of R_{0t} on R_{1t},

$$R_{0t} = \alpha\beta' R_{1t} + \text{error}. \tag{13}$$

The likelihood ratio type test for the null

H_0: there are $n - r$ unit roots

against

H_c: there are $n - r$ roots local to unity

is given by $LR_{\bar{c}} = -T\sum_{i=r+1}^{n}\ln(1-\breve{\lambda}_i)$, where $\breve{\lambda}_1,\dots,\breve{\lambda}_n$ are the ordered squared canonical correlations in the regression (13).

Following Johansen (1988, 1991), define $S_{00} = T^{-1}\sum_{t=1}^{T}R_{0t}R_{0t}'$, $S_{01} = T^{-1}\sum_{t=1}^{T}R_{0t}R_{1t}'$, $S_{10} = T^{-1}\sum_{t=1}^{T}R_{1t}R_{0t}'$, $S_{11} = T^{-1}\sum_{t=1}^{T}R_{1t}R_{1t}'$. Then, $\breve{\lambda}_{r+1}, \dots, \breve{\lambda}_n$ can be found by solving the determinantal equation

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0 \tag{14}$$

leading to the ordered eigenvalues $\breve{\lambda}_1 > \cdots > \breve{\lambda}_n$.

The asymptotic properties of this test statistic are given in the following theorem.

THEOREM 3. *Under the null hypothesis that there are* $n - r$ *unit roots, the test statistic* $LR_{\bar{c}} = -T\sum_{i=r+1}^{n}\ln(1-\breve{\lambda}_i)$ *is asymptotically distributed as the sum of the roots of the determinantal equation*

$$\left|\lambda\int_0^1 \underline{W}_{\bar{c}}(r)\underline{W}_{\bar{c}}(r)' - \int_0^1 \underline{W}_{\bar{c}}(r)\,d\underline{W}_{\bar{c}}(r)'\int_0^1 d\underline{W}_{\bar{c}}(r)\,\underline{W}_{\bar{c}}(r)'\right| = 0,$$

i.e., $LR_{\bar{c}} \Rightarrow \operatorname{tr}\{\int_0^1 d\underset{\sim}{W}_{\bar{c}}(r)\underset{\sim}{W}_{\bar{c}}(r)'[\int_0^1 \underset{\sim}{W}_{\bar{c}}(r)\underset{\sim}{W}_{\bar{c}}(r)']^{-1}\int_0^1 \underset{\sim}{W}_{\bar{c}}(r)\, d\underset{\sim}{W}_{\bar{c}}(r)'\}$, *where*

$$\underset{\sim}{W}_{\bar{c}}(r) = W(r) - \int_0^1 dW_{\bar{c}}(s)X_{\bar{c}}(s)'\left[\int_0^1 X_{\bar{c}}(s)X_{\bar{c}}(s)'\right]^{-1} X(r),$$

$$W_{\bar{c}}(r) = W(r) - \bar{c}\int_0^1 W(s),$$

and $W(r)$ *is a standard Brownian motion.*

Remark 6. The derivation of the limit theory depends on the asymptotic behavior of the product moment matrices of the detrended data. These asymptotic properties are verified and the proof of Theorem 3 is given in Section 8.

Remark 7. In the detrending regression, we used OLS regressions based on quasi-differenced data. Joint estimation of the trend coefficients from a seemingly unrelated (SUR) regression on (8′) could also be used, and whether or not there is a further efficiency gain from the use of SUR regression depends on the form of the deterministic trend. For the polynomial trend given in Section 2, no gain is obtained from the SUR regression because the space spanned by x_t is the same as the space spanned by $\Delta_{\bar{c}} x_t$.

Remark 8. If we consider another alternative H_c': there are $n - r - 1$ unit roots and one root local to unity, then an efficient detrending procedure under this alternative would estimate the trend coefficients based on fully differencing in the $n - r - 1$ directions that have unit roots and QD in the direction that is local to unity. In this case, the detrending procedure treats different directions in different ways, which is reflected in the limit distribution of the detrended data. More generally, we can consider local alternative H_c'': there are $n - r - s$ $(s > 0)$ unit roots and s roots local to unity. This is the localized version of the model considered in Johansen (1996, chap. 14) in which there are r cointegrating vectors under the null and $r + s$ cointegrating vectors under the alternative. Again, to apply the idea of efficient detrending under the alternative, we should fully difference the $n - r - s$ directions that have unit roots and quasi-difference the s directions that are local to unity.

The same idea can be employed for the alternative H_c in the way that we perform QD by using different local parameters $c_j (j = r + 1, \dots, n)$ in the different directions. This is efficient under H_c. Let $H = [\beta, \beta_\perp] = [h_1, \dots, h_r, h_{r+1}, \dots, h_n]$; then, under the null, the detrended time series y_t^* has the following asymptotic behavior:

$$T^{-1/2} y^*_{[\mathrm{Tr}]} \Rightarrow B(r) - \sum_{j=r+1}^{n} h_j h_j' \int_0^1 dB_{c_j}(s) X_{c_j}(s)\left[\int_0^1 X_{c_j}(s)X_{c_j}(s)'\right]^{-1} X(r).$$

Note that this limit not only depends on the c_j values used in QD but also depends on the directions h_j.

The preceding procedure requires knowledge of the cointegrating vectors β. When β is unknown we can use a preliminary estimate of β, obtained from a reduced rank regression in the usual way, to construct a rotation of the coordinate system. Then, we reestimate β and calculate the modified test from a new reduced rank regression with the efficiently detrended data. Hence, we suggest the following three-step algorithm for testing in a cointegrated system.

Step 1. Obtain consistent estimates of α and β by running a reduced rank regression in the usual way on the system

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{j=1}^{k-1} A_j^* \Delta y_{t-j} + Bx_t + \varepsilon_t.$$

Step 2. Using the estimated β from step 1, transform the system by premultiplying by $H' = (\hat{\beta}, \hat{\beta}_\perp)'$ and calculate

$$\underline{y}_t = H' y_t = [(\hat{\beta}' y_t)', (\hat{\beta}_\perp' y_t)']' = [\underline{y}_{1t}', \underline{y}_{2t}']'.$$

Run detrending regressions based on the quasi-differences of $\underline{y}_{2t}$ (for some appropriate $\bar{c}$) as in

$$\begin{pmatrix} \underline{y}_{1t} \\ \Delta_{\bar{c}} \underline{y}_{2t} \end{pmatrix} = \begin{pmatrix} \breve{\Gamma}_1 & 0 \\ 0 & \breve{\Gamma}_2 \end{pmatrix} \begin{pmatrix} x_t \\ \Delta_{\bar{c}} x_t \end{pmatrix} + \text{residual}.$$

Construct the detrended time series and transform both back to original coordinates as follows:

$$\breve{\underline{y}}_t = \underline{y}_t - \breve{\underline{\Gamma}} x_t,$$

$$\breve{y}_t = H \breve{\underline{y}}_t = y_t - \breve{\Gamma} x_t,$$

$$\breve{\Gamma} = H \breve{\underline{\Gamma}} = H \begin{bmatrix} \breve{\Gamma}_1 \\ \breve{\Gamma}_2 \end{bmatrix}.$$

Step 3. Reestimate α, β by running a new reduced rank regression on the detrended data $\breve{y}_t$, namely,

$$\Delta \breve{y}_t = \alpha\beta' \breve{y}_{t-1} + \sum_{i=1}^{k-1} \Phi_i \Delta \breve{y}_{t-i} + \breve{\varepsilon}_t$$

and calculate the likelihood ratio (LR) test statistic

$$LR_{\bar{c}} = -T \sum_{i=r+1}^{n} \ln(1 - \breve{\lambda}_i). \tag{15}$$

Remark 9. When a preliminary estimate of the cointegrating vector is used, the estimation error from $\hat{\beta}$ figures in the transformed data as we rotate the coordinate system. As a result, the asymptotic validity of the LR test statistic in the preceding algorithm depends on the magnitude of the error term and the form of the deterministic trend. Because the reduced rank regression estimate of the cointegrating vector converges at the rate T, for the leading cases of a constant term ($x_t = \gamma$) and a linear trend ($x_t = \gamma_0 + \gamma_1 t$), it can be verified that the detrended data $\breve{y}_t$ in the preceding three-step procedure have the same limiting behavior given by Theorem 2, and the likelihood ratio test statistic $LR_{\bar{c}}$ (15) has the same asymptotic distribution as that given in Theorem 3. A proof of this result is given in Section 8.

Remark 10. Obviously, this procedure could be iterated further in the hope of some finite sample advantage.

5. ADDITIONAL ISSUES ON INITIALIZATION

We assumed in previous sections that the initial values of y_t have finite variance and are $O_p(1)$ as $T \to \infty$. However, in nonstationary models when the initial conditions are allowed to go into the remote past, this no longer holds, and information on the initial condition plays a role in the limit distributions. The following construction for the initial conditions of the nonstationary part allows for this possibility and is used in Canjels and Watson (1997) and Phillips and Lee (1996):

$$\underline{y}_{20} = \sum_{i=0}^{[\theta T]} K_2(L)\varepsilon_{-i}.$$

The variance $\underline{y}_{20}$ is determined by the new "remote past" initialization parameter θ. When $\theta = 0$, $\underline{y}_{20} = O_p(1)$, whereas for $\theta > 0$, $\underline{y}_{20} = O_p(\sqrt{T})$. To make this initial condition consistent with the structure of the model described in Section 2, we transform to original coordinates using

$$y_0 = \beta K_1(L)\varepsilon_0 + \beta_\perp K_2(L)\sum_{i=0}^{[\theta T]} \varepsilon_{-i}. \tag{16}$$

In stationary directions, the initial conditions do not influence the asymptotic theory, so these initializations can be ignored here.

Initial conditions of this type change the limit distributions of our test statistics in two ways. First, they change the limit of $T^{-1/2}y^s_{[\mathrm{Tr}]}$. Second, and more importantly, they can affect the asymptotic properties of the trend coefficient estimator $\tilde{\Gamma}$.

When the initial condition is given by (16), under the null hypothesis H_0, we have

$$\begin{aligned} y_t^s &= H\underline{y}_t^s = \beta \underline{y}_{1t}^s + \beta_\perp \underline{y}_{2t}^s \\ &= \beta K_1(L)\varepsilon_t + \beta_\perp K_2(L)\left[\sum_{i=0}^{[\theta T]} \varepsilon_{-i} + \sum_{j=0}^{t} \varepsilon_j\right]. \end{aligned}$$

Thus

$$\begin{aligned} T^{-1/2}y^s_{[\mathrm{Tr}]} &= \beta_\perp K_2(L)\left[T^{-1/2}\sum_{i=0}^{[\theta T]} \varepsilon_{-i} + T^{-1/2}\sum_{j=0}^{[\mathrm{Tr}]} \varepsilon_j\right] + o_p(1) \\ &\Rightarrow B^*(\theta) + B(r), \end{aligned}$$

where

$$B^*(\theta) = \beta_\perp(\alpha_\perp'\Pi^*\beta_\perp)^{-1}\alpha_\perp' B_\varepsilon^*(\theta),$$

$$B(r) = \beta_\perp(\alpha_\perp'\Pi^*\beta_\perp)^{-1}\alpha_\perp' B_\varepsilon(r),$$

$$B_\varepsilon^*(\theta) = \Lambda^{1/2} W^*(\theta),$$

$$B_\varepsilon(r) = \Lambda^{1/2} W(r).$$

Here, $W^*(\theta)$ and $W(r)$ are independent standard Brownian motions.

The effect of the new initial values on the estimated trend coefficients depends on both the form of deterministic trend and the type of estimation procedure employed. Without loss of generality, we will study the case of a linear trend $x_t = (1,t)$ here to illustrate how initial conditions and different estimators can affect the asymptotic results. In this case, $D = \mathrm{diag}(1, T^{-1})$, $F = TD = \mathrm{diag}(T, 1)$, and then

$$Dx_{[\mathrm{Tr}]} \to (1,r)' = X(r),$$

$$F\Delta_{\bar{c}} x_{[\mathrm{Tr}]} \to (-\bar{c}, 1 - \bar{c}r)' = X_{\bar{c}}(r).$$

Partition the coefficient matrix $\underline{\Gamma}$ as

$$\underline{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix},$$

so that

$$\underline{y}_{1t} = \gamma_{11} + \gamma_{12} t + \underline{y}_{1t}^s = \Gamma_1 x_t + \underline{y}_{1t}^s,$$

$$\underline{y}_{2t} = \gamma_{21} + \gamma_{22} t + \underline{y}_{2t}^s = \Gamma_2 x_t + \underline{y}_{2t}^s.$$

As in Canjels and Watson (1997), we consider the following four estimators for Γ. Because the estimate of Γ_1 in all of these procedures is the same OLS estimate, we focus our discussion on the estimation of Γ_2. The estimate of Γ_2 are as follows.

E1. OLS estimation of Γ_2:

$$\underline{y}_{2t} = \hat{\Gamma}_2^{OLS} x_t + \underline{y}_{2t}^s, \qquad t = 1, 2, \ldots, T.$$

E2. Cochrane–Orcutt (1949) GLS estimation of Γ_2:

$$\Delta_{\bar{c}} \underline{y}_{2t} = \tilde{\Gamma}_2^{CO} \Delta_{\bar{c}} x_t + \Delta_{\bar{c}} \underline{y}_{2t}^s, \qquad t = 2, \ldots, T.$$

E3. QD estimation including the levels information in the first observation:

$$\underline{y}_{21} = \tilde{\Gamma}_2^{GLS} x_1 + \underline{y}_{21}^s,$$

$$\Delta_{\bar{c}} \underline{y}_{2t} = \tilde{\Gamma}_2^{GLS} \Delta_{\bar{c}} x_t + \Delta_{\bar{c}} \underline{y}_{2t}^s, \qquad t = 2, \ldots, T.$$

E4. Prais–Winsten (1954) estimation of Γ_2:

$$(\delta \underline{y}_{21}) = \tilde{\Gamma}_2^{PW}(\delta x_1) + \text{error}, \qquad t = 1,$$

$$\Delta_{\bar{c}} \underline{y}_{2t} = \tilde{\Gamma}_2^{PW} \Delta_{\bar{c}} x_t + \text{error}, \qquad t = 2,\dots,T,$$

where $\delta^2 = 1 - (1 + \bar{c}/T)^2 = 2|\bar{c}|/T + o(T^{-1})$.

The limit distributions of the preceding estimators are given in the following theorem.

THEOREM 4.

$$T^{-1/2}(\hat{\Gamma}_2^{OLS} - \Gamma_2)D^{-1} \Rightarrow \beta_\perp' \int_0^1 (B(\theta)^* + B(r))X(r)' \left[\int_0^1 X(r)X(r)' \right]^{-1},$$

$$T^{-1/2}(\tilde{\Gamma}_2^{CO} - \Gamma_2)D^{-1} \Rightarrow \beta_\perp' \int_0^1 dB_{\bar{c}}(r)X_{\bar{c}}(r)' \left[\int_0^1 X_{\bar{c}}(r)X_{\bar{c}}(r)' \right]^{-1},$$

$$T^{-1/2}(\tilde{\Gamma}_2^{PW} - \Gamma_2)D^{-1} \Rightarrow \left[(2|\bar{c}|K_2(1)B_\varepsilon^*(\theta), \quad 0) + \beta_\perp' \int_0^1 dB_{\bar{c}}(r)X_{\bar{c}}(r)' \right] \times \left[\begin{pmatrix} 2|\bar{c}| & 0 \\ 0 & 0 \end{pmatrix} + \int_0^1 X_{\bar{c}}(r)X_{\bar{c}}(r)' \right]^{-1}.$$

When $\theta > 0$,

$$T^{-1/2}(\tilde{\Gamma}_2^{GLS} - \Gamma_2)D^{-1} \Rightarrow \left(K_2(1)B_\varepsilon^*(\theta), \quad K_2(1)B_\varepsilon^*(\theta) + \beta_\perp' \int_0^1 dB_{\bar{c}}(r)(1 - \bar{c}r) \right) \times \begin{bmatrix} 1 & 1 - \bar{c} + \bar{c}^2/2 \\ 0 & [1 - (1 - \bar{c})^3]/(3\bar{c}) \end{bmatrix}^{-1}.$$

When $\theta = 0$,

$$(\tilde{\Gamma}_2^{GLS} - \Gamma_2)G^{-1} \Rightarrow \left(\underline{y}_{21}^s, \quad \frac{3\bar{c}}{1 - (1 - \bar{c})^3} \beta_\perp' \int_0^1 dB_{\bar{c}}(r)(1 - \bar{c}r) \right),$$

where $G = \operatorname{diag}(1, T^{-1/2})$.

Remark 11. In all of these cases, the estimator for the intercept term γ_{21} is inconsistent. In most of the cases, the estimator of γ_{21} is of order $T^{1/2}$. However, for the estimation procedure (E3), when there is finite variance, the rate of divergence in the estimator of γ_{21} is lower than that of the other estimators and we have to use another scaling matrix G. As a result, the variance of $\tilde{\gamma}_{21}$ in this case is lower than that of the other cases.

Remark 12. The invertibility of $\int X_{\bar{c}} X_{\bar{c}}'$ depends on $\bar{c}$ not equaling 0 because the constant term and the coefficient of the linear trend t are unidentified when $\bar{c} = 0$. For values of $\bar{c}$ close to zero, the trend coefficients will be poorly estimated.

Remark 13. The limit distribution of the Cochrane–Orcutt GLS estimator is invariant to initial conditions because the estimator ignores the levels information of the first observation. Consequently, the asymptotics for $\tilde{\Gamma}^{CO}$ are the same as those obtained in Section 4. The effect of the initial condition on the tests comes from the limit of $T^{-1/2}y^s_{[Tr]}$. Specifically,

$$
\begin{aligned}
T^{-1/2}\tilde{y}_{[Tr]} &= T^{-1/2}y^s_{[Tr]} - T^{1/2}(\tilde{\Gamma}-\Gamma)x_{[Tr]} \\
&\Rightarrow B^*(\theta) + B(r) - \int_0^1 dB_{\bar{c}}(s)X_{\bar{c}}(s)'\left[\int_0^1 X_{\bar{c}}(s)X_{\bar{c}}(s)'\right]^{-1}X(r) \\
&\equiv \underset{\sim}{B}{}^*_\theta(r), \text{ say,}
\end{aligned}
$$

and the test statistic for the null hypothesis H_0 against H_c, namely, $-T\sum_{i=r+1}^{n}\ln(1-\tilde{\lambda}_i)$, is asymptotically distributed as the trace statistic from the following determinantal equation:

$$
\left|\lambda\int_0^1 \underset{\sim}{W}{}^*_\theta(r)\underset{\sim}{W}{}^*_\theta(r)' - \int_0^1 \underset{\sim}{W}{}^*_\theta(r)\,d\underset{\sim}{W}{}^*_\theta(r)'\int_0^1 d\underset{\sim}{W}{}^*_\theta(r)\underset{\sim}{W}{}^*_\theta(r)'\right| = 0, \tag{17}
$$

where $\underset{\sim}{W}{}^*_\theta(r) = W^*(\theta) + W(r) - \int_0^1 dW_{\bar{c}}(s)X_{\bar{c}}(s)'[\int_0^1 X_{\bar{c}}(s)X_{\bar{c}}(s)']^{-1}X(r)$.

Remark 14. The limit distribution of the trend coefficient estimator (E3) is dependent on the initial observation. The initial condition affects the distribution of the test statistic through both $y^s_{[Tr]}$ and $\tilde{\Gamma}$. As a result, the test statistic $-T\sum_{i=r+1}^{n}\ln(1-\tilde{\lambda}_i)$ is asymptotically distributed as the trace statistic from the following determinantal equation:

$$
\left|\lambda\int_0^1 \underset{\sim}{W}{}^{**}_\theta(r)\underset{\sim}{W}{}^{**}_\theta(r)' - \int_0^1 \underset{\sim}{W}{}^{**}_\theta(r)\,d\underset{\sim}{W}{}^{**}_\theta(r)'\int_0^1 d\underset{\sim}{W}{}^{**}_\theta(r)\underset{\sim}{W}{}^{**}_\theta(r)'\right| = 0, \tag{18}
$$

where $\underset{\sim}{W}{}^{**}_\theta(r) = W^*(\theta) + W(r) - [W^*(\theta)h' + \int_0^1 dW_{\bar{c}}(s)X_{\bar{c}}(s)'] \times [\int_0^1 X_{\bar{c}}(s)X_{\bar{c}}(s)']^{-1}X(r)$ and h is a vector dependent on the first observation.

Remark 15. When $\theta = 0$, our results reduce to the case of $O_p(1)$ initial observations, $\underset{\sim}{W}{}^*_\theta(r)$, $\underset{\sim}{W}{}^{**}_\theta(r)$ reduce to $\underset{\sim}{W}(r)$, and the limits are the same as those obtained in Section 4.

6. MONTE CARLO RESULTS

A small Monte Carlo experiment was conducted to evaluate the efficient detrending procedures on tests for cointegration and to provide a comparison of the QD detrended tests with the OLS detrended test. The data generating processes (DGP's) are

$$
\begin{aligned}
y_{1t} &= \rho y_{1,t-1} + \varepsilon_{1t}, \\
y_{2t} &= 0.2y_{1,t-1} + \varepsilon_{2t},
\end{aligned}
$$

where ε_{1t} and ε_{2t} are both i.i.d. $N(0,1)$ variates and are independent of each other. The sample size considered is $T = 100$, and $\rho = 1 + c/T$. The null hypothesis is H_0 : there is one cointegration vector, i.e., there is one unit root in the system. The alternative considered here is H_c : there is one root local to unity.

We consider the following detrending procedures:

1. OLS detrending,
2. QD detrending with a choice of $\bar{c} = -7.5$, and
3. QD detrending with a choice of $\bar{c} = -13.5$.

These values of $\bar{c}$ were chosen because in tests for a unit root, the c values for which local asymptotic power is 50% are approximately -13.5 for the case with a linear trend and -7 for the demeaned case. Readers are referred to Elliot et al. (1996) for further discussion on this matter.

Two kinds of deterministic trends were considered:

Case 1: $x_t = t$, and
Case 2: $x_t = (1, t)'$.

About generating the random variables and calculating the test statistics, pseudorandom normal variates are generated using the GAUSS subroutine RNDN, and trace statistic $LR_{\bar{c}}$ is calculated using COINT 2.0, the software developed for unit root and cointegration testing (Ouliaris and Phillips, 1994). The power of the LR test based on these detrending procedures is examined. The finite sample critical values of the tests are calculated from simulations based on 10,000 replications (Table 1). We also calculated the approximated critical values for the asymptotic distribution based on a Monte Carlo simulation using Gaussian random variates and $T = 400$ (Table 2). Table 3 reports the empirical size of these tests using the critical values given in Table 2. To compare the power of different tests, we calculated the size corrected power, i.e., the Monte Carlo rejection rates when the actual 5% critical value computed for that model is used to calculate the rejections. Table 4 reports the empirical power of the cointegration tests for Case 1, and Table 5 reports the corresponding results for Case 2. Figures 1 and 2 depict the power functions.

For the time series considered here, the efficient detrending procedures perform reasonably well. Although the relative performance of different detrend-

TABLE 1. Finite sample critical values, size = 5%

	Case 1: $x_t = t$	Case 2: $x_t = (1, t)$
OLS detrended test	8.230238	11.580656
QD detrended test, $\bar{c} = -7.5$	8.157201	8.4625046
QD detrended test, $\bar{c} = -13.5$	8.259264	8.9291147

Table 2. (Approximated) asymptotic critical values, size = 5%

	Case 1: $x_t = t$	Case 2: $x_t = (1, t)$
OLS detrended test	7.9738310	11.745856
QD detrended test, $\bar{c} = -7.5$	8.0059878	9.5076413
QD detrended test, $\bar{c} = -13.5$	7.9915987	10.588673

ing procedures depends on the distributional form of the process, more efficiency gain is generally achieved in Case 2, which is consistent with the Monte Carlo results in unit root tests. An explanation of this phenomenon in unit root testing is given by Phillips and Lee (1996). A choice of $\bar{c}$ value around -13.5 has been found in simulations to be a generally good default choice, whereas choices of $\bar{c}$ closer to 0 provide less favorable results. One of the reasons for this phenomenon is that the constant term and the coefficient of the trend t are unidentified when $\bar{c} = 0$, as discussed earlier. As a result, for $\bar{c}$ close to 0, the inverse matrix $[\int X_{\bar{c}} X_{\bar{c}}']^{-1}$ is unstable and the trend coefficients are poorly estimated.

7. CONCLUSION

This paper analyzes efficient detrending procedures in cointegrated time series regression. Tests for cointegration based on these detrending procedures are developed, and the limit distributions of these new tests are derived. Some limited Monte Carlo evidence indicates that the efficient detrending procedures and the efficiently detrended tests for cointegration perform reasonably well in finite samples.

Because the directions of nonstationarity and stationarity are usually not known a priori in multivariate time series, some preliminary estimation and transformation of the system need to be performed before detrending. The LR test statistics, which are constructed from a reduced rank regression with the detrended data, are functions of certain eigenvalues of the product moment matrices corresponding to the smallest squared canonical correlations. The asymptotics of these tests are

Table 3. Empirical size of the tests

	Case 1: $x_t = t$	Case 2: $x_t = (1, t)$
OLS detrended test	0.0562	0.0422
QD detrended test, $\bar{c} = -7.5$	0.0540	0.0276
QD detrended test, $\bar{c} = -13.5$	0.0568	0.0213

TABLE 4. Power of tests for cointegration: Case 1, $x_t = t$

True ρ	OLS detrending	QD detrending, $\bar{c} = -7.5$	QD detrending $\bar{c} = -13.5$
1	0.05	0.05	0.05
0.975	0.0647	0.0646	0.0652
0.95	0.108	0.1081	0.1089
0.925	0.18	0.181	0.1819
0.9	0.2867	0.2884	0.2884
0.875	0.4148	0.4161	0.4156
0.85	0.5467	0.5475	0.5485
0.825	0.6733	0.6714	0.6743
0.8	0.7785	0.7796	0.7782
0.775	0.8606	0.8586	0.8614
0.75	0.9122	0.9104	0.9138
0.725	0.9491	0.9467	0.9493
0.7	0.9709	0.9683	0.9712

generally dependent on the specific directions in which the QD is performed, and the cancellation of nuisance parameters in the limit distributions depends crucially on the asymptotic behavior of these product moment matrices of the detrended data. Certain relationships among the limiting product moment matrices have been verified in this paper to validate these tests for cointegration.

TABLE 5. Power of tests for cointegration: Case 2, $x_t = (1, t)'$

True ρ	OLS detrending	QD detrending, $\bar{c} = -7.5$	QD detrending $\bar{c} = -13.5$
1	0.05	0.05	0.05
0.975	0.0594	0.0624	0.0647
0.95	0.0859	0.1036	0.1073
0.925	0.1244	0.1679	0.1732
0.9	0.1834	0.2573	0.2653
0.875	0.2596	0.3589	0.3719
0.85	0.3523	0.4641	0.4797
0.825	0.4493	0.5674	0.5863
0.8	0.5545	0.6598	0.6868
0.775	0.6533	0.7359	0.7697
0.75	0.7434	0.7965	0.8324
0.725	0.8159	0.8456	0.8854
0.7	0.8697	0.8796	0.9204

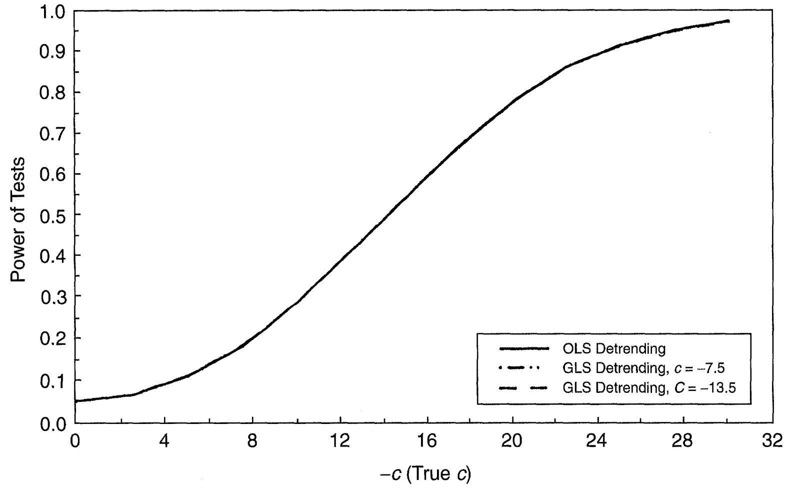

FIGURE 1. Empirical power of tests for cointegration, trend $= t$.

When the initial observation has variance that grows with the sample size, the initial value plays a role in the asymptotic theory of the estimate of the trend coefficients and can influence the limit distribution of the test statistics, as in unit root tests. Some differences between models with and without intercepts are also found in the analysis, and, again, this extends earlier findings on unit root tests.

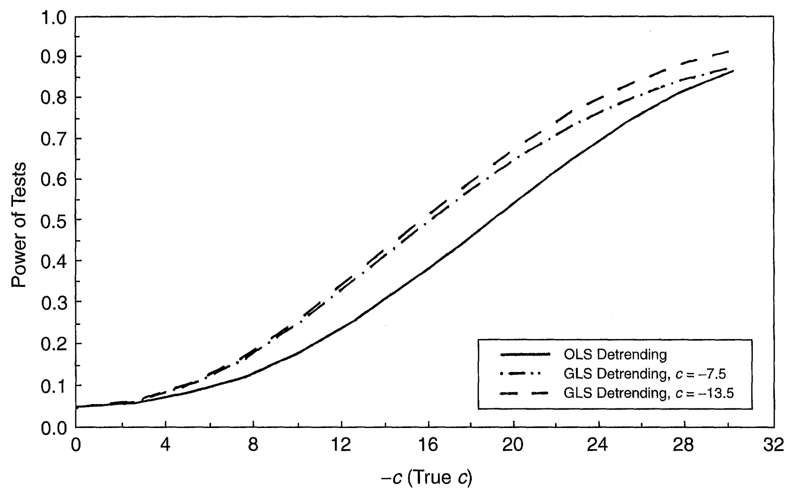

FIGURE 2. Empirical power of tests for cointegration, trend $= (1, t)$.

8. PROOF OF THEOREMS

8.1. Proof of Theorem 1

$$\begin{aligned}
T^{-1/2}\hat{y}^s_{[Tr]} &= T^{-1/2}y_{[Tr]} - T^{-1/2}\hat{\Gamma}x_{[Tr]} \\
&= T^{-1/2}y^s_{[Tr]} - T^{-1/2}(\hat{\Gamma}-\Gamma)x_{[Tr]} \\
&= T^{-1/2}y^s_{[Tr]} - \left[\frac{1}{T}\sum \frac{y^s_t}{\sqrt{T}}x_t'D\right]\left[\frac{1}{T}\sum Dx_tx_t'D\right]^{-1} Dx_{[Tr]} \\
&\Rightarrow J_c(r) - \int J_c(s)X(s)'\left[\int X(s)X(s)'\right]^{-1} X(r) \\
&\equiv \underline{J}_c(r), \\
T^{-1/2}\tilde{y}^s_{[Tr]} &= T^{-1/2}y_{[Tr]} - T^{-1/2}\tilde{\Gamma}x_{[Tr]} \\
&= T^{-1/2}y^s_{[Tr]} - T^{-1/2}(\tilde{\Gamma}-\Gamma)x_{[Tr]} \\
&= T^{-1/2}y^s_{[Tr]} - T^{-1/2}\beta(\tilde{\Gamma}_1-\Gamma_1)x_{[Tr]} - T^{-1/2}\beta_\perp(\tilde{\Gamma}_2-\Gamma_2)x_{[Tr]} \\
&= T^{-1/2}y^s_{[Tr]} - T^{-1}\beta\left[\sum \frac{y^s_{1t}}{\sqrt{T}}x_t'D\right]\left[\frac{1}{T}\sum Dx_tx_t'D\right]^{-1} Dx_{[Tr]} \\
&\quad -\beta_\perp\left[\sum \frac{\Delta_c y^s_{2t}}{\sqrt{T}}\Delta_c x_t'F\right]\left[\frac{1}{T}\sum F\Delta_c x_t\Delta_c x_t'F\right]^{-1} Dx_{[Tr]} \\
&\Rightarrow J_c(r) - \int dB(s)X_c(s)'\left[\int X_c(s)X_c(s)'\right]^{-1} X(r) \\
&\equiv \underset{\sim}{J}_c(r). \qquad \blacksquare
\end{aligned}$$

8.2. Proof of Theorem 2

Under the null hypothesis, $C = 0, T^{-1/2}y^s_{[Tr]} \Rightarrow B(r)$. Notice that

$$\begin{aligned}
\breve{y}_t &= y_t - \breve{\Gamma}x_t = y^s_t - (\breve{\Gamma}-\Gamma)x_t = y^s_t - \beta(\breve{\Gamma}_1-\Gamma_1)x_t - \beta_\perp(\breve{\Gamma}_2-\Gamma_2)x_t \\
&= y^s_t - \beta \underline{y}^{s\prime}_1 X(X'X)^{-1}x_t - \beta_\perp \Delta_{\bar{c}}\underline{y}^{s\prime}_2\Delta_{\bar{c}}X(\Delta_{\bar{c}}X'\Delta_{\bar{c}}X)^{-1}x_t
\end{aligned}$$

and

$$T^{-1/2}\sum_{t=1}^{[Tr]}\Delta_{\bar{c}}y^s_t = T^{-1/2}\sum_{t=1}^{[Tr]}\left(\Delta y^s_t - \frac{\bar{c}}{T}y^s_{t-1}\right) \Rightarrow B(r) - \bar{c}\int_0^r B(s) = B_{\bar{c}}(r),$$

$$F\Delta_{\bar{c}}x_{[Tr]} = F\Delta x_{[Tr]} - \bar{c}Dx_{[Tr]} \to g(r) - \bar{c}X(r) = X_{\bar{c}}(r).$$

By a similar argument to that in Theorem 1, we get

$$T^{-1/2}\breve{y}_{[Tr]} \Rightarrow B(r) - \int_0^1 dB_{\bar{c}}(s)X_{\bar{c}}(s)'\left[\int_0^1 X_{\bar{c}}(s)X_{\bar{c}}(s)'\right]^{-1} X(r) = \underset{\sim}{B}_{\bar{c}}(r). \qquad \blacksquare$$

To prove Theorem 3, we need some preliminary lemmas.

8.3. Lemma 1

$$T^{-1}\sum_{t=1}^{T} \breve{\varepsilon}_t \breve{\varepsilon}'_{t+i} \xrightarrow{p} 0.$$

8.4. Proof of Lemma 1

Because $\breve{\varepsilon}_t = \varepsilon_t - (\breve{\Gamma} - \Gamma)\Delta x_t + \sum_{j=1}^{k-1} \Phi_j(\breve{\Gamma} - \Gamma)\Delta x_{t-j} + \alpha\beta'(\breve{\Gamma} - \Gamma)x_{t-1}$, we substitute it into the product and get the following representation:

$$\begin{aligned}
&T^{-1}\sum_{t=1}^{T} \breve{\varepsilon}_t \breve{\varepsilon}'_{t+i} \\
&\quad = T^{-1}\sum_{t=1}^{T} \varepsilon_t \varepsilon'_{t+i} - (\breve{\Gamma} - \Gamma)T^{-1}\sum_{t=1}^{T} \Delta x_t \varepsilon'_{t+i} \\
&\qquad + \alpha\beta'(\breve{\Gamma} - \Gamma)T^{-1}\sum_{t=1}^{T} x_{t-1}\varepsilon'_{t+i} + \sum_{j=1}^{k-1} \Phi_j(\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t=1}^{T} \Delta x_{t-j}\varepsilon'_{t+i}\right) \\
&\qquad - T^{-1}\sum_{t=1}^{T} \varepsilon_t \Delta x'_{t+i}(\breve{\Gamma} - \Gamma)' + (\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t} \Delta x_t \Delta x'_{t+i}\right)(\breve{\Gamma} - \Gamma)' \\
&\qquad - \alpha\beta'(\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t} x_{t-1}\Delta x'_{t+i}\right)(\breve{\Gamma} - \Gamma)' - \sum_{j=1}^{k-1} \Phi_j(\breve{\Gamma} - \Gamma) \\
&\qquad \times \left(T^{-1}\sum_{t} \Delta x_{t-j}\Delta x'_{t+i}\right)(\breve{\Gamma} - \Gamma)' \\
&\qquad + T^{-1}\sum_{t=1}^{T} \varepsilon_t x'_{t+i-1}(\breve{\Gamma} - \Gamma)'\beta\alpha' - (\breve{\Gamma} - \Gamma) \\
&\qquad \times \left(T^{-1}\sum_{t} \Delta x_t x'_{t+i-1}\right)(\breve{\Gamma} - \Gamma)'\beta\alpha' \\
&\qquad + \alpha\beta'(\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t} x_{t-1}x'_{t+i-1}\right)(\breve{\Gamma} - \Gamma)\beta\alpha' \\
&\qquad + \sum_{j=1}^{k-1} \Phi_j(\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t} \Delta x_{t-j}x'_{t+i-1}\right)(\breve{\Gamma} - \Gamma)'\beta\alpha' \\
&\qquad + \sum_{j=1}^{k-1}\left[T^{-1}\sum_{t=1}^{T} \varepsilon_t \Delta x'_{t+i-j}(\breve{\Gamma} - \Gamma)'\Phi'_j\right] \\
&\qquad - \sum_{j=1}^{k-1}\left[(\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t=1}^{T} \Delta x_t \Delta x'_{t+i-j}\right)(\breve{\Gamma} - \Gamma)'\Phi'_j\right] \\
&\qquad + \alpha\beta'\sum_{j=1}^{k-1}\left[(\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t=1}^{T} x_{t-1}\Delta x'_{t+i-j}\right)(\breve{\Gamma} - \Gamma)'\Phi'_j\right] \\
&\qquad + \sum_{j=1}^{k-1}\sum_{\ell=1}^{k-1}\left[\Phi_\ell(\breve{\Gamma} - \Gamma)\left(T^{-1}\sum_{t=1}^{T} \Delta x_{t-\ell}\Delta x'_{t+i-j}\right)(\breve{\Gamma} - \Gamma)'\Phi'_j\right].
\end{aligned}$$

It can be verified that all these terms go to zero in probability as $T \to \infty$. In particular, because $T^{1/2}(\breve{\Gamma}_1 - \Gamma_1)D^{-1} = O_p(1)$ and $T^{1/2}(\breve{\Gamma}_2 - \Gamma_2)F^{-1} = O_p(1)$, we have

$$T^{-1} \sum_{t=1}^{T} \varepsilon_t \varepsilon'_{t+i} \xrightarrow{p} 0,$$

$$(\breve{\Gamma} - \Gamma)\left(T^{-1} \sum_{t=1}^{T} \Delta x_t \varepsilon'_{t+i}\right)$$
$$= [\beta(\breve{\Gamma}_1 - \Gamma_1) + \beta_\perp(\breve{\Gamma}_2 - \Gamma_2)]\left(T^{-1} \sum_{t=1}^{T} \Delta x_t \varepsilon'_{t+i}\right) \xrightarrow{p} 0,$$

$$\alpha\beta'(\breve{\Gamma} - \Gamma)T^{-1} \sum_{t=1}^{T} x_{t-1} \varepsilon'_{t+i}$$
$$= \alpha\beta'\beta(\breve{\Gamma}_1 - \Gamma_1)\left(T^{-1} \sum_{t=1}^{T} x_{t-1} \varepsilon'_{t+i}\right)$$
$$= \alpha\beta'\beta \cdot T^{-1/2}(T^{1/2}(\breve{\Gamma}_1 - \Gamma_1)D^{-1})\left(T^{-1} \sum_{t=1}^{T} D x_{t-1} \varepsilon'_{t+1}\right) \xrightarrow{p} 0,$$

$$\sum_{j=1}^{k-1} \Phi_j(\breve{\Gamma} - \Gamma)\left(T^{-1} \sum_{t=1}^{T} \Delta x_{t-j} \varepsilon'_{t+i}\right) \xrightarrow{p} 0,$$

$$\left(T^{-1} \sum_{t=1}^{T} \varepsilon_t \Delta x'_{t+i}\right)(\breve{\Gamma} - \Gamma)' \xrightarrow{p} 0,$$

$$(\breve{\Gamma} - \Gamma)\left(T^{-1} \sum_{t} \Delta x_t \Delta x'_{t+i}\right)(\breve{\Gamma} - \Gamma)'$$
$$= [\beta(\breve{\Gamma}_1 - \Gamma_1) + \beta_\perp(\breve{\Gamma}_2 - \Gamma_2)]\left(T^{-1} \sum_{t} \Delta x_t \Delta x'_{t+i}\right)$$
$$\times [(\breve{\Gamma}_1 - \Gamma_1)'\beta' + (\breve{\Gamma}_2 - \Gamma_2)'\beta'_\perp] \xrightarrow{p} 0,$$

$$\alpha\beta'(\breve{\Gamma} - \Gamma)\left(T^{-1} \sum_{t} x_{t-1} \Delta x'_{t+i}\right)(\breve{\Gamma} - \Gamma)'$$
$$= \alpha\beta'\beta(\breve{\Gamma}_1 - \Gamma_1)\left(T^{-1} \sum_{t} x_{t-1} \Delta x'_{t+i}\right)$$
$$\times [(\breve{\Gamma}_1 - \Gamma_1)'\beta' + (\breve{\Gamma}_2 - \Gamma_2)'\beta'_\perp] \xrightarrow{p} 0,$$

$$\sum_{j=1}^{k-1} \Phi_j(\breve{\Gamma} - \Gamma)\left(T^{-1} \sum_{t} \Delta x_{t-j} \Delta x'_{t+i}\right)(\breve{\Gamma} - \Gamma)' \xrightarrow{p} 0,$$

$$T^{-1} \sum_{t=1}^{T} \varepsilon_t x'_{t+i-1}(\breve{\Gamma} - \Gamma)'\beta\alpha' \xrightarrow{p} 0,$$

$$(\check{\Gamma} - \Gamma)\left(T^{-1} \sum_t \Delta x_t x_{t+i-1}\right)(\check{\Gamma} - \Gamma)'\beta\alpha' \xrightarrow{p} 0,$$

$$\alpha\beta'(\check{\Gamma} - \Gamma)\left(T^{-1} \sum_t x_{t-1} x'_{t+i-1}\right)(\check{\Gamma} - \Gamma)\beta\alpha'$$

$$= \alpha\beta'\beta(\check{\Gamma}_1 - \Gamma_1)\left(T^{-1} \sum_t x_{t-1} x'_{t+i-1}\right)(\check{\Gamma}_1 - \Gamma_1)'\beta'\beta\alpha' \xrightarrow{p} 0,$$

$$\sum_{j=1}^{k-1} \Phi_j(\check{\Gamma} - \Gamma)\left(T^{-1} \sum_t \Delta x_{t-j} x'_{t+i-j}\right)(\check{\Gamma} - \Gamma)'\beta\alpha' \xrightarrow{p} 0,$$

$$\sum_{j=1}^{k-1}\left[T^{-1} \sum_{t=1}^{T} \varepsilon_t \Delta x'_{t+i-1}(\check{\Gamma} - \Gamma)'\Phi'_j\right] \xrightarrow{p} 0,$$

$$\sum_{j=1}^{k-1}\left[(\check{\Gamma} - \Gamma)\left(T^{-1} \sum_{t=1}^{T} \Delta x_t \Delta x'_{t+i-j}\right)(\check{\Gamma} - \Gamma)'\Phi'_j\right] \xrightarrow{p} 0,$$

$$\alpha\beta' \sum_{j=1}^{k-1}\left((\check{\Gamma} - \Gamma)\left(T^{-1} \sum_{t=1}^{T} x_{t-1} \Delta x'_{t+i-j}\right)(\check{\Gamma} - \Gamma)'\Phi'_j\right] \xrightarrow{p} 0,$$

and

$$\sum_{j=1}^{k-1}\sum_{\ell=1}^{k-1}\left[\Phi_\ell(\check{\Gamma} - \Gamma)\left(T^{-1} \sum_{t=1}^{T} \Delta x_{t-\ell} \Delta x'_{t+i-j}\right)(\check{\Gamma} - \Gamma)'\Phi'_j\right] \xrightarrow{p} 0. \qquad \blacksquare$$

Following Johansen (1988), we define the matrices

$$\psi(j) = \operatorname{Var}(\Delta\check{y}_t \Delta\check{y}'_{t+j}), \qquad \mu_{00} = E[\Delta\check{y}_t \Delta\check{y}'_t],$$

$$\mu_{0z} = E[\Delta\check{y}_t Z'_t], \qquad \mu_{zz} = E[Z_t Z'_t], \qquad \mu_{j1} = \sum_{i=1-j}^{\infty} \psi(i),$$

$$\mu_{z1} = \begin{bmatrix} \mu_{11} \\ \cdots \\ \mu_{k-1,1} \end{bmatrix},$$

and use the following limits and notation:

$$\Psi = I + \alpha\beta' - \sum \Phi_i,$$

$$Z'_t = [\Delta\check{y}'_{t-1}, \ldots, \Delta\check{y}'_{t-k+1}],$$

$$M_{00} = T^{-1} \sum_{t=1}^{T} \Delta \breve{y}_t \cdot \Delta \breve{y}_t' \xrightarrow{p} \mu_{00},$$

$$M_{0z} = T^{-1} \sum_{t=1}^{T} \Delta \breve{y}_t \cdot Z_t' \xrightarrow{p} \mu_{0z},$$

$$M_{zz} = T^{-1} \sum_{t=1}^{T} Z_t \cdot Z_t' \xrightarrow{p} \mu_{zz},$$

$$M_{01} = T^{-1} \sum_{t=1}^{T} \Delta \breve{y}_t \cdot \breve{y}_{t-1}' \Rightarrow \int_0^1 d\underset{\sim}{B}_{\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)' + \mu_{01},$$

$$M_{z1} = T^{-1} \sum_{t=1}^{T} Z_t \, \breve{y}_{t-1}' \Rightarrow \begin{bmatrix} \vdots \\ \int_0^1 d\underset{\sim}{B}_{\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)' \\ \vdots \end{bmatrix} + \mu_{z1},$$

$$M_{11} = T^{-1} \sum_{t=1}^{T} \breve{y}_{t-1} \breve{y}_{t-1}' = O_p(T), \qquad \beta' M_{11} \beta \xrightarrow{p} \beta' \mu_{11} \beta,$$

$$M_{z0} = M_{0z}' \xrightarrow{p} \mu_{z0} = \mu_{0z}',$$

$$M_{10} = M_{01}' \xrightarrow{p} \mu_{10} + \int_0^1 \underset{\sim}{B}_{\bar{c}}(r) \, d\underset{\sim}{B}_{\bar{c}}(r)', \qquad \mu_{10} = \mu_{01}',$$

$$M_{1z} = M_{z1}' \xrightarrow{p} \left[\ldots, \int_0^1 \underset{\sim}{B}_{\bar{c}}(r) \, d\underset{\sim}{B}_{\bar{c}}(r)', \ldots \right] + \mu_{1z}, \qquad \mu_{1z} = \mu_{z1}'.$$

Also

$$S_{00} = M_{00} - M_{0z} M_{zz}^{-1} M_{z0} \xrightarrow{p} \mu_{00} - \mu_{0z} \mu_{zz}^{-1} \mu_{z0} = \Sigma_{00},$$

$$S_{01} = M_{01} - M_{0z} M_{zz}^{-1} M_{z1}$$

$$\Rightarrow \int_0^1 d\underset{\sim}{B}_{\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)' + \mu_{01} - \mu_{0z} \mu_{zz}^{-1} \left(\mu_{z1} + \begin{bmatrix} \vdots \\ \int_0^1 d\underset{\sim}{B}_{\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)' \\ \vdots \end{bmatrix} \right),$$

$$T^{-1} S_{11} = T^{-1} (M_{11} - M_{1z} M_{zz}^{-1} M_{z1}) \Rightarrow \int_0^1 \underset{\sim}{B}_{\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)',$$

$$\beta' S_{11} \beta \xrightarrow{p} \beta' \Sigma_{11} \beta,$$

$$S_{01} \beta \xrightarrow{p} (\mu_{01} - \mu_{0z} \mu_{zz}^{-1} \mu_{z1}) \beta = \Sigma_{01} \beta.$$

We have the following results for the product moment matrices.

8.5. Lemma 2

$$\Sigma_{00} = \alpha\beta' \Sigma_{10} + \Lambda,$$

$$\Sigma_{01}\beta = \alpha\beta' \Sigma_{11}\beta.$$

8.6. Proof of Lemma 2

The proof of these two relationships uses the following limits:

$$T^{-1}\varepsilon' XD = T^{-1}\sum_{t=1}^{T} \varepsilon_t x_t' D \xrightarrow{p} 0,$$

$$T^{-1}\underline{y}_1^{s\prime} XD = T^{-1}\sum_{t=1}^{T} \underline{y}_{1t}^{s\prime} x_t' D \xrightarrow{p} 0,$$

$$T^{-1}\Delta \underline{y}_2^{s\prime} \Delta XF = T^{-1}\sum_{t=1}^{T} \Delta \underline{y}_{2t}^{s\prime} \Delta x_t' F \xrightarrow{p} 0,$$

which hold because $\underline{y}_{1t}^{s} = K_1(L)\varepsilon_t$ and $\Delta \underline{y}_{2t}^{s} = K_2(L)\varepsilon_t$. ■

8.7. Proof of Theorem 3

Under the null hypothesis,

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{k-1} \Phi_i \Delta y_{t-i} + Bx_t + \varepsilon_t.$$

For the QD detrended time series, the following properties hold:

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \breve{\varepsilon}_t \Delta \breve{y}_{t-i}' = 0, \qquad i = 1,2,\dots,k-1,$$

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \breve{\varepsilon}_t \Delta \breve{y}_t' = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \breve{\varepsilon}_t \breve{\varepsilon}_t' = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \varepsilon_t \varepsilon_t' = \Lambda,$$

$$\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} \breve{\varepsilon}_t \breve{y}_{t-1}' \beta = 0.$$

By an argument similar to that in Johansen (1988), it can be shown that the test statistic $-T\sum_{i=r+1}^{n} \ln(1 - \breve{\lambda}_i)$ depends on the limit in the eigenvalues of deterministic equation

$$\left| \beta_\perp' \int_0^1 \underset{\sim}{B}_{\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)'\beta_\perp - \mu\beta_\perp' S_{10}\alpha_\perp(\alpha_\perp'\Lambda\alpha_\perp)^{-1}\alpha_\perp' S_{01}\beta_\perp \right| = 0, \tag{19}$$

and we can show that the $\breve{\varepsilon}_t$ are asymptotically uncorrelated (Lemma 1). Then

$$w \lim_{T\to\infty} \alpha_\perp' S_{01} = w \lim_{T\to\infty} \alpha_\perp' T^{-1} \sum_{t=1}^{T} \breve{\varepsilon}_t \breve{y}_{t-1}'$$

$$= \alpha_\perp' \int_0^1 d\underset{\sim}{B}_{\varepsilon\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)',$$

where "w lim" signifies the limit in weak convergence as $T \to \infty$. Thus (19) simplifies as

$$\left| \beta_\perp' \int_0^1 \underset{\sim}{B}_{\bar{c}}\underset{\sim}{B}_{\bar{c}}'\beta_\perp - \mu\beta_\perp' \int_0^1 \underset{\sim}{B}_{\bar{c}}(r)\, d\underset{\sim}{B}_{\varepsilon\bar{c}}(r)'\alpha_\perp(\alpha_\perp'\Lambda\alpha_\perp')^{-1}\alpha_\perp' \int_0^1 d\underset{\sim}{B}_{\varepsilon\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)'\beta_\perp \right|$$
$$= 0.$$

Because

$$(\alpha_\perp'\Lambda\alpha_\perp)^{-1/2}\alpha_\perp'\underset{\sim}{B}_{\varepsilon\bar{c}}(r) = \underset{\sim}{W}_{\bar{c}}(r),$$

deterministic equation (19) can be further simplified to

$$\left| \beta_\perp' \int_0^1 \underset{\sim}{B}_{\bar{c}}\underset{\sim}{B}_{\bar{c}}'\beta_\perp - \mu\beta_\perp' \int_0^1 \underset{\sim}{B}_{\bar{c}}(r) d\underset{\sim}{W}_{\bar{c}}(r)' \int_0^1 d\underset{\sim}{W}_{\bar{c}}(r)\underset{\sim}{B}_{\bar{c}}(r)'\beta_\perp \right| = 0,$$

where

$$\underset{\sim}{B}_{\bar{c}}(r) = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}(\alpha_\perp'\Lambda\alpha_\perp)^{1/2}$$
$$\times \left[W(r) - \int_0^1 dW_{\bar{c}}(s)X_{\bar{c}}(s)' \left[\int_0^1 X_{\bar{c}}X_{\bar{c}}' \right]^{-1} X(r) \right]$$
$$= \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}(\alpha_\perp'\Lambda\alpha_\perp)^{1/2}\underset{\sim}{W}_{\bar{c}}(r).$$

Thus, we have

$$\Bigg| \beta_\perp'\beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}(\alpha_\perp'\Lambda\alpha_\perp)^{1/2}$$
$$\times \left[\int_0^1 \underset{\sim}{W}_{\bar{c}}(r)\underset{\sim}{W}_{\bar{c}}(r)' - \mu \int_0^1 \underset{\sim}{W}_{\bar{c}}(r)d\underset{\sim}{W}_{\bar{c}}(r)' \int_0^1 d\underset{\sim}{W}_{\bar{c}}(r)\underset{\sim}{W}_{\bar{c}}(r)' \right]$$
$$\times (\alpha_\perp'\Lambda\alpha_\perp)^{1/2}(\beta_\perp'\Psi\alpha_\perp)^{-1}\beta_\perp'\beta_\perp) \Bigg| = 0.$$

The test statistic $-T\sum_{r+1}^{n}\ln(1-\check{\lambda}_i) = T\sum_{i=r+1}^{n}\check{\lambda}_i + o_p(1)$ is then asymptotically distributed as the sum of the eigenvalues of the equation

$$\left|\lambda\int_0^1 \underset{\sim}{W}_{\bar{c}}(r)\underset{\sim}{W}_{\bar{c}}(r)' - \int_0^1 \underset{\sim}{W}_{\bar{c}}(r)\,d\underset{\sim}{W}_{\bar{c}}(r)'\int_0^1 d\underset{\sim}{W}_{\bar{c}}(r)\underset{\sim}{W}_{\bar{c}}(r)'\right| = 0. \qquad \blacksquare$$

8.8. Proof of Remark 9

In the three-step algorithm suggested in Section 4, we may use a preliminary estimate of the cointegrating vector in constructing the rotation. To obtain a uniquely determined representative of the cointegrating vector, we may use the normalization $b' = [I_r, -\Upsilon]$ of β. As discussed in Phillips (1994, p. 76), this normalization corresponds to the priori requirement that there be r structural relations of a form that explicitly recognizes a subvector of full rank integrated regressors and is sufficient for the unique determination of β. Once the matrix of cointegrating vectors is estimated from a first stage reduced rank regression, ex post estimation of Υ can be accomplished, and it is shown (Phillips, 1991) that $\hat{\Upsilon}$ converges to Υ at rate T. We may then reparameterize this estimator as

$$\hat{\beta} = \begin{bmatrix} I_r \\ -\hat{\Upsilon}' \end{bmatrix}(I_r + \hat{\Upsilon}\hat{\Upsilon}')^{-1/2}$$

so that it is orthonormalized, and the corresponding estimator of $\beta_\perp$ is

$$\hat{\beta}_\perp = \begin{bmatrix} \hat{\Upsilon} \\ I_{n-r} \end{bmatrix}(I_{n-r} + \hat{\Upsilon}'\hat{\Upsilon})^{-1/2}.$$

We then have $\hat{\beta} - \beta = O_p(T^{-1})$ and $\hat{\beta}_\perp - \beta_\perp = O_p(T^{-1})$. The transformed time series are

$$\underset{\sim}{\hat{y}}_{1t} = \hat{\beta}'y_t = \beta'y_t + (\hat{\beta}-\beta)'y_t = \underset{\sim}{y}_{1t} + (\hat{\beta}-\beta)'y_t,$$

$$\underset{\sim}{\hat{y}}'_{2t} = \hat{\beta}'_\perp y_t = \beta'_\perp y_t + (\hat{\beta}_\perp-\beta_\perp)'y_t = \underset{\sim}{y}_{2t} + (\hat{\beta}_\perp-\beta_\perp)'y_t.$$

Next, we run detrending regressions based on the quasi-differences of $\underset{\sim}{y}_{2t}$ as in

$$\begin{pmatrix} \underset{\sim}{\hat{y}}_{1t} \\ \Delta_{\bar{c}}\underset{\sim}{\hat{y}}_{2t} \end{pmatrix} = \begin{pmatrix} \Gamma_1^* & 0 \\ 0 & \Gamma_2^* \end{pmatrix}\begin{pmatrix} x_t \\ \Delta_{\bar{c}}x_t \end{pmatrix} + \text{residual}.$$

In the following analysis, we use the affix * to signify the feasible versions of the estimators. For example, Γ_i^* represents the estimate of Γ_i based on the transformed data $\underset{\sim}{\hat{y}}_t$ using $\hat{\beta}$, and $\check{\Gamma}_1$ and $\check{\Gamma}_2$ correspond to the infeasible version of them using the true cointegrating vector β. Thus, we have

$$
\begin{aligned}
\Gamma_1^* &= \left[\sum_t \hat{\underline{y}}_{1t} x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} \\
&= \left[\sum_t \{\underline{y}_{1t} + (\hat{\beta} - \beta)' y_t\} x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} \\
&= \left[\sum_t \underline{y}_{1t} x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} + \left[(\hat{\beta} - \beta)' \sum_t y_t x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} \\
&= \breve{\Gamma}_1 + (\hat{\beta} - \beta)'\left[\sum_t y_t x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} \\
&= \breve{\Gamma}_1 + d_1, \\
\Gamma_2^* &= \left[\sum_t \Delta_c \hat{\underline{y}}_{2t} \Delta_c x_t'\right]\left[\sum_t \Delta_c x_t \Delta_c x_t'\right]^{-1} \\
&= \breve{\Gamma}_2 + (\hat{\beta}_\perp - \beta_\perp)'\left[\sum_t \Delta_c \underline{y}_{2t} \Delta_c x_t'\right]\left[\sum_t \Delta_c x_t \Delta_c x_t'\right]^{-1} \\
&= \breve{\Gamma}_2 + d_2,
\end{aligned}
$$

where $\breve{\Gamma}_1$ and $\breve{\Gamma}_2$ are the infeasible version of Γ_i^*.

The QD detrended data then have the following decomposition:

$$
\begin{aligned}
y_t^* &= y_t - \Gamma^* x_t \\
&= y_t - (\hat{\beta}\Gamma_1^* + \hat{\beta}_\perp \Gamma_2^*) x_t \\
&= y_t - (\beta\breve{\Gamma}_1 + \beta_\perp \breve{\Gamma}_2) x_t - [(\hat{\beta} - \beta)\breve{\Gamma}_1 + (\hat{\beta}_\perp - \beta_\perp)\breve{\Gamma}_2] x_t - (\hat{\beta} d_1 + \hat{\beta}_\perp d_2) x_t \\
&= \breve{y}_t - [(\hat{\beta} - \beta)\breve{\Gamma}_1 + (\hat{\beta}_\perp - \beta_\perp)\breve{\Gamma}_2] x_t - (\hat{\beta} d_1 + \hat{\beta}_\perp d_2) x_t,
\end{aligned}
$$

where $\breve{y}_t$ is the infeasible QD detrended data.

We need to show that $T^{-1/2} y_{[\mathrm{Tr}]}^*$ and $T^{-1/2} \breve{y}_{[\mathrm{Tr}]}$ have the same limit distribution. This will be so if

$$
T^{-1/2}[(\hat{\beta} - \beta)\breve{\Gamma}_1 + (\hat{\beta}_\perp - \beta_\perp)\breve{\Gamma}_2] x_{[\mathrm{Tr}]} = o_p(1), \tag{20}
$$

$$
T^{-1/2}(\hat{\beta} d_1 + \hat{\beta}_\perp d_2) x_{[\mathrm{Tr}]} = o_p(1). \tag{21}
$$

Notice that

$$
\begin{aligned}
T^{-1/2}\hat{\beta} d_1 x_{[\mathrm{Tr}]} &= T^{-1/2}\hat{\beta}\left[(\hat{\beta} - \beta)' \sum_t y_t x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} x_{[\mathrm{Tr}]} \\
&= T^{-1/2}\hat{\beta}(\hat{\beta} - \beta)'\left[\sum_t (y_t^s + \Gamma x_t) x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} x_{[\mathrm{Tr}]} \\
&= T^{-1/2}\hat{\beta}(\hat{\beta} - \beta)'\left[\sum_t y_t^s x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} x_{[\mathrm{Tr}]} \\
&\quad + T^{-1/2}\hat{\beta}(\hat{\beta} - \beta)'\Gamma\left[\sum_t x_t x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} x_{[\mathrm{Tr}]} \\
&= T^{-1/2}\hat{\beta}(\hat{\beta} - \beta)'\left[\sum_t y_t^s x_t'\right]\left[\sum_t x_t x_t'\right]^{-1} x_{[\mathrm{Tr}]} \\
&\quad + T^{-1/2}\hat{\beta}(\hat{\beta} - \beta)'\Gamma x_{[\mathrm{Tr}]}.
\end{aligned}
$$

The first term, $T^{-1/2}\hat{\beta}(\hat{\beta}-\beta)'[\sum_t y_t^s x_t'][\sum_t x_t x_t']^{-1}x_{[\mathrm{Tr}]}$, goes to zero as T goes to infinity. The order of magnitude of the second term, $T^{-1/2}\hat{\beta}(\hat{\beta}-\beta)'\Gamma x_{[\mathrm{Tr}]}$, is determined by the convergence rate of $\hat{\beta}$ and the form of x_t. Because $\hat{\beta}-\beta = O_p(T^{-1})$, when $x_t = (1,t)'$ is a linear trend, $T^{-1/2}\hat{\beta}(\hat{\beta}-\beta)'\Gamma x_{[\mathrm{Tr}]} = O_p(T^{-1/2})$. (When there is deterministic cointegration, the order of magnitude for $T^{-1/2}\hat{\beta}(\hat{\beta}-\beta)'\Gamma x_{[\mathrm{Tr}]}$ is further reduced.) Thus

$$T^{-1/2}\hat{\beta}d_1 x_{[\mathrm{Tr}]} = o_p(1).$$

Similarly,

$$T^{-1/2}\hat{\beta}_\perp d_2 x_{[\mathrm{Tr}]} = o_p(1),$$

$$T^{-1/2}(\hat{\beta}-\beta)\check{\Gamma}_1 x_{[\mathrm{Tr}]} = o_p(1),$$

$$T^{-1/2}(\hat{\beta}_\perp-\beta_\perp)\check{\Gamma}_2 x_{[\mathrm{Tr}]} = o_p(1),$$

and (20) and (21) follow. ■

NOTE

1. To simplify the presentation of our approach, we proceed here as if H (and hence β) were known. Later on in this section, we provide a feasible stepwise RRR procedure that utilizes preliminary estimates of β and H at this stage.

REFERENCES

Beveridge, S. & C.R. Nelson (1981) A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics* 7, 151–174.

Canjels, E. & M.W. Watson (1997) Estimating deterministic trend in the presence of serially correlated errors. *Review of Economics and Statistics* LXXIX(2), 184–200.

Cochrane, D. & G.H. Orcutt (1949) Applications of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* 44, 32–61.

DeJong, D.N., J.C. Nankervis, N.E. Savin, & C.H. Whiteman (1992) The power problems of unit root tests in time series with autoregressive errors. *Journal of Econometrics* 53, 323–43.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–31.

Dufour, J.M. & M.L. King (1991) Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *Journal of Econometrics* 47, 115–143.

Elliot, G., T.J. Rothenberg, & J.H. Stock (1996) Efficient tests for an autoregressive unit root. *Econometrica* 64, 813–836.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121–130.

Granger, C.W.J. & A.A. Weiss (1983) Time series analysis and error correcting models. In S. Karlin, T. Amemiya, & L.A. Goodman (eds.), *Studies in Econometrics, Time Series and Multivariate Statistics*, 255–278. New York: Academic Press.

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

Hansen, B.E. (1995) Rethinking the univariate approach to unit root testing: Using covariates to increase power. *Econometric Theory* 11, 1148–1173.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

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

Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

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

King, M.L. (1988) Towards a theory of point optimal testing. *Econometric Reviews* 6, 169–218.

Ouliaris, S. & P.C.B. Phillips (1994) COINT 2.0. Maple Valley, Washington: Aptech Systems.

Park, J.Y. (1992) Canonical cointegrating regressions. *Econometrica* 60 119–143.

Phillips, P.C.B. (1987a) Time series regression with a unit root. *Econometrica* 55, 277–301.

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

Phillips, P.C.B. (1988) Regression theory for near-integrated time series. *Econometrica* 56 (5), 1021–1044.

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

Phillips, P.C.B. (1994) Some exact distribution theory for maximum likelihood estimators for cointegrating coefficients in error correction models. *Econometrica* 62, 73–93.

Phillips, P.C.B. (1998) Impulse response and forecast error variance asymptotics in nonstationary VAR's. *Journal of Econometrics* 83, 21–56.

Phillips, P.C.B. & B.E. Hansen (1990). Statistical inference with $I(1)$ process. *Review of Economic Studies* 57, 99–124.

Phillips, P.C.B. & C.C. Lee (1996) Efficiency gains from quasi-differencing under nonstationarity. In P.M. Robinson & M. Rosenblatt (eds.), *Essays in Memory of E.J. Hannan*, pp. 300–314. New York: Springer-Verlag.

Phillips, P.C.B. & S. Ouliaris (1990) Asymptotic properties of residual based tests for cointegration. *Econometrica* 58, 165–193.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrica* 74, 535–547.

Prais, S.J. & C.B. Winsten (1954) Trend Estimators and Serial Correlation. Cowles Foundation Discussion paper 383.

Sargan, P. & A. Bhargava (1983) Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* 51, 153–174.

Schmidt, P. & P.C.B. Phillips (1992) LM tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics* 54, 257–287.

Schwert, G.W. (1989) Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7, 147–160.

Sims, C.A., J.H. Stock, & M.W. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.

Stock, J.H. (1987) Asymptotic properties of least squares estimators of cointegration vectors. *Econometrica* 55, 1035–1056.

Stock, J.H. & M.W. Watson (1988) Testing for common trends. *Journal of the American Statistical Association* 83, 1097–1107.