*Econometrica*, Vol. 66, No. 1 (January, 1998), 149–158

# NOTES AND COMMENTS

## ON THE ROBUSTNESS OF COINTEGRATION METHODS WHEN REGRESSORS ALMOST HAVE UNIT ROOTS

BY GRAHAM ELLIOTT[1]

### 1. INTRODUCTION

THIS PAPER EXAMINES the robustness of $\chi^2$ inference on long run relationships between economic variables when we are not sure the variables have exact unit roots. Economic theories often derive relationships between economic quantities that hold over long periods of time. When individual variables have a unit root,[2] Engle and Granger (1987) showed that cointegration can be an empirically useful method to model such relationships. A great number of techniques were devised to estimate and undertake hypothesis testing on cointegrating vectors (cf. survey in Watson (1994)).

The 'knowledge' that there is a unit root usually comes in practice through failure to reject this proposition in a unit root or rank test. It is rare that a unit root is to be expected on (economic) theoretical grounds (see Christiano and Eichenbaum (1990) for example); indeed often variables such as interest rates are modeled in cointegrating relationships even though it is highly unlikely that the interest rate (or its log) could theoretically have a unit root. Tests for unit roots have low power over a range of close alternatives (Elliott et al. (1996) derive theoretical upper bounds on power for univariate tests). Thus correct inference on the cointegrating vector relies critically on the robustness of the methods in this direction.

It is shown analytically, using local to unity asymptotic approximations (Bobkoski (1983), Cavanagh (1985), Phillips (1987)), that whilst point estimates of cointegrating vectors remain consistent, commonly applied hypothesis tests no longer have the usual distribution when roots are near but not one. The size of the effects can be extremely large for even very small deviations from a unit root; indeed it will be shown that rejection rates can be close to one. Hypothesis tests on general restrictions on the cointegrating vector are only affected if the restriction includes coefficients on variables which do not have an exact unit root, and are unaffected asymptotically by the presence of near unit root variables not included in the restriction.

### 2. MODEL AND RESULTS

The model can be written

$$(1)\qquad \begin{aligned} y_{1t} &= d_{1t}^* + Ay_{1t-1} + u_{1t}, \\ y_{2t} &= d_{2t} + \Gamma y_{1t} + u_{2t}, \end{aligned}$$

[1] I thank R. F. Engle, A. Fatas, G. Chamberlain, C. W. J. Granger, J. Hamilton, G. Imbens, S. Johansen, J. H. Stock, M. W. Watson, H. White, and three referees for comments. The paper has benefited from presentations at Harvard University, LSE, ECARE, Toulouse, Tel Aviv, Northwestern, UCSD, Michigan, and Ohio State. All errors are my own.

[2] The inability of tests to reject the presence of a unit root in a number of common macroeconomic time series has previously been shown by Nelson and Plosser (1982).

where $t = 1 \dots T$, $d_{1t}^* = (I - AL)d_{1t}$, $d_{1t} = G_1 z_{1t}$ and $d_{2t} = G_2 z_{2t}$ are deterministic terms, $y_{1t}$ is an $n_1 \times 1$ vector, $y_{2t}$ is an $n_2 \times 1$ vector, $\Phi(L)u_t = \varepsilon_t$, $\varepsilon_t = [\varepsilon_{1t}', \varepsilon_{2t}']'$ is an $n \times 1$ vector of martingale difference sequence errors ($n = n_1 + n_2$) with $E(\varepsilon_t \varepsilon_t') = \Sigma$ and four finite moments, and $\Phi(L)$ is a lag polynomial of known order with all roots outside the unit circle. The scaled (by $2\pi$) spectral density of $u_t$ at frequency zero is $\Omega = \Phi(1)^{-1}\Sigma\Phi(1)^{-1\prime}$ where $\Phi(1) = \Sigma_i \Phi_i$. We consider the cases where (i) $z_{1t} = 0$ and $z_{2t} = 1$, and a constant is included in each equation in the regression run or (ii) $z_{1t} = 1$ or 0 and $z_{2t} = [1, t]'$, and a constant and time trend are included in the regression. Of interest is estimation and hypothesis testing on the cointegrating vector $\Gamma$ when we approximate the process for $y_{1t}$ with unit roots.

For the purposes of obtaining asymptotic results, we use the local to unity parameterization $T(A - I) = C$, where $C$ is an $n_1 \times n_1$ matrix with zeros on the offdiagonals[3] and potentially nonzero elements on the diagonal. If $C = 0$, then $y_{1t}$ all have unit roots. If a diagonal element is negative, the corresponding $y_{1t}$ variable is mean reverting. This is a device to obtain asymptotic distributions that provide reasonable approximations when the sample size is small and the variables are (slowly) mean reverting, motivated by the failure of the usual fixed $A$ (pointwise) asymptotic results to provide useful approximations unless the sample size is extremely large. The area of alternatives that unit root tests have difficulty distinguishing from unit roots are precisely those measured by the local alternatives considered here.

Results are derived for the Saikkonen (1992) estimates of the cointegrating vector, which are asymptotically equivalent to full information maximum likelihood for the normalized model. The results also apply to the Johansen (1991) or Ahn and Reinsel (1990) methods. It was shown in a previous version (Elliott (1995)) that the results hold for the Phillips and Hansen (1990) fully modified estimator and the Stock and Watson (1993) DOLS procedure in the bivariate model.

Following Saikkonen (1992) the model in (1) can be rearranged into the VAR:

$$\Delta y_t = \tilde{\beta}_2 z_t + \Psi y_{t-1} + \Pi(L)\Delta y_{t-1} + \varepsilon_t^* \tag{2}$$

where $z_t = 1$ or $z_t = (1, t)'$, $\varepsilon_t^* = P\varepsilon_t$, $\Psi = P\Phi(1)P^{-1}M$,

$$P = \begin{bmatrix} I_{n1} & 0 \\ \Gamma & I_{n2} \end{bmatrix}$$

partitioned after the $n_1$th row and column, and

$$M = \begin{bmatrix} A - I_{n1} & 0 \\ \Gamma A & -I_{n2} \end{bmatrix}.$$

The Saikkonen (1992) method estimates the cointegrating vectors directly from OLS estimates of coefficients in (2), where $\hat{\Gamma} = -\hat{S}_{22}^{-1}\hat{S}_{21}$, $\hat{S} = (\hat{\Psi}'\hat{\Sigma}^{*-1}\hat{\Psi})$, and the partition of $S$ is after the $n_1$th row and column, and $\hat{\Sigma}^* = T^{-1}\Sigma\hat{\varepsilon}_t^*\hat{\varepsilon}_t^{*\prime}$.

[3] Nonzero elements in the offdiagonal positions correspond to $I(2)$ variables if the diagonal is zero. This is not examined.

THEOREM 1: *When the data is generated according to* (1), *then the FIML and Saikkonen* (1992) *estimators of* $\Gamma$ *have the asymptotic normal distribution conditional on* $y_{1t}$ *given by*

$$\hat{V}^{-1/2}\text{vec}\{[T(\hat{\Gamma}-\Gamma)]-(B)\}\sim_a N(0, I_{n1n2})$$

*where*

$$B=-\Omega_{21}\Omega_{11}^{-1}C,$$

$$\hat{V}\Rightarrow V=[\Omega_{11}^{1/2}(\textstyle\int J_c^d J_c^{d\prime})\Omega_{11}^{1/2\prime}]^{-1}\otimes\Omega_{2.1},$$

$$\Omega_{2.1}=\Omega_{22}-\Omega_{21}\Omega_{11}^{-1}\Omega_{12}$$

*and* $J_c^d(\lambda)$ *is a detrended Ornstein Uhlenbeck process where*

$$dJ_c(\lambda)=CJ_c(\lambda)\,d\lambda+dW_1(\lambda),$$

$J_c(0)=0$, $W_1(\lambda)$ *is an* $n_1\times 1$ *standard Brownian Motion defined on* [0 1],

(i) $$J_c^d(\lambda)=J_c(\lambda)-\int_0^1 J_c(s)\,ds \quad \text{if} \quad z_t=1 \quad \text{and}$$

(ii) $$J_c^d(\lambda)=J_c(\lambda)-(4-6\lambda)\int_0^1 J_c(s)\,ds-(12\lambda-6)\int_0^1 sJ_c(s)\,ds \quad \text{if} \quad z_t=(1,t)'.$$

Estimates are consistent but have a bias vec$(B)$ which disappears at rate $T$. The bias is not there if $C=0$, the case where all variables have a unit root. The bias is larger the larger is $\Omega_{12}$, the covariance of $u_{1t}$ and $u_{2t}$ at frequency zero, i.e. for models in which there is a greater degree of simultaneity in the regression. The bias is smaller the larger is $\Omega_{11}$; i.e. if the variance of the $y_{1t}$ variables at frequency zero is large then the bias is small. This has consequences for the choice of the normalization; if it is suspected that $C$ is nonzero, choosing $y_{1t}$ so that this variance is larger makes the bias smaller (see Ng and Perron (1996)). This consistency also means that if a wrong assumption of $C=0$ is made, it is unlikely to show up in point estimates if the sample size is sufficiently large.

The bias will only appear for those coefficients on variables with near unit roots (this follows from the diagonality of $C$) but in general appears for coefficients on those variables in each equation. Coefficients on variables with an exact unit root are unaffected asymptotically by the inclusion of variables with near unit roots in the cointegrating vector.

Wald tests of the hypothesis $R\,\text{vec}(\Gamma)=r$ where the rank of $R$ is $q$ can be obtained using the usual Wald statistic (Park and Phillips (1988)). For $\hat{V}$ that converges to $V$,

(3) $$W^*=\{R\,\text{vec}[T(\hat{\Gamma}-\Gamma)]\}'\{R\hat{V}R'\}^{-1}\{R\,\text{vec}[T(\hat{\Gamma}-\Gamma)]\}.$$

COROLLARY TO THEOREM 1: *Under the conditions of Theorem* 1, $W^*\Rightarrow\chi_q^2+B^*$, *where*

$$B^*=\{R\,\text{vec}(B)\}'\{RVR'\}^{-1}\{R\,\text{vec}(B)\}+2\{R\,\text{vec}(B)\}'\{RVR'\}^{-1/2\prime}Z$$

*and* $Z$ *has a mixed normal distribution with mean zero and variance* $I_q$ *conditional on* $y_{1t}$.

The limit distribution is a mixture of the usual $\chi_q^2$ and a functional of a vector Brownian motion correlated with this $\chi_q^2$ distribution. The mean shift (and variance) of the bias term $B^*$ depends on $\Omega$ and $C$. When $\Omega_{12}'\Omega_{11}^{-1}$ is nonzero, tests of hypotheses will be affected when the restriction involves a variable that does not have a unit root (cases where $R\,\text{vec}(B) \neq 0$), and will be unaffected for all restrictions only on variables which contain a unit root (cases where $R\,\text{vec}(B) = 0$). Thus even when the model is partially misspecified in the sense that some of the variables have near unit roots, hypothesis tests will be unaffected if they do not include restrictions on coefficients of these near unit root variables. The mean of the bias term is nonnegative; hypothesis tests involving near unit roots will tend to overreject on average when testing $\Gamma$ for its true value, causing researchers to commit larger than expected Type 1 errors. The larger is the bias term $R\,\text{vec}(B)$, the larger is the overrejection on average.

Results are clearest when $n_1 = n_2 = 1$. Testing $\Gamma$ for its true value, $R = 1$. The first term (mean) of $B^*$ is now

$$\frac{C^2\delta^2}{(1-\delta^2)}\left(\int (J_c^d)^2\right), \quad \text{where} \quad \delta = \frac{\Omega_{12}}{\sqrt{\Omega_{11}\Omega_{22}}}.$$

As $|\delta| \to 1$, the size of the bias goes to infinity and so the size of the test goes to one. For $\delta = 0$, there is no reason to use cointegration estimation techniques as they yield the same normally distributed estimates as OLS asymptotically. Cointegrating methods were introduced precisely to deal with large correlations between the residuals, but it is precisely for these large correlations that we do not have robustness.

Figure 1 quantifies this size distortion in the bivariate model for a number of models and a nominal size of 5% for the demeaned and detrended models. For mild endogeneity ($|\delta| < 0.5$), asymptotic size does not rise much above 10% (which is still double nominal size) for $c = -5$. For greater amounts of endogeneity, even for such a slight departure from a unit root asymptotic size is unacceptably large. For $C = -10$ (corresponding to an autoregressive coefficient of 0.9 for a sample size of 100) we see that size is almost 50% when $\delta = 0.7$. As suggested by the analytic results, size goes to one as $\delta$ gets close to one. The effect of the extent of detrending is small.

Table I examines Monte Carlo results in models corresponding to the demeaned case model in Figure 1 where $T = 100$ and $A = 0.9$ ($C = -10$) for a number of popularly employed methods for the estimation and testing of cointegrating vectors.[4] Column 1 gives approximate asymptotic results from Figure 1. The asymptotic results of the corollary give a direct guide as to the expected extent of size distortions in small samples.[5] The results are not dependent on the actual method examined and apply to all methods which optimally use the information in the unit root assumption.

In larger systems, rejection rates can also be arbitrarily large for any nominal size test. For $C$ close to zero many reasonable parameterizations of $\Omega$ result in $R\,\text{vec}(B)$ far from zero. Monte Carlo results are available on request.

## 3. CONCLUSION

This paper shows that hypothesis tests on vectors of trending variables undertaken using the cointegration methodology are in most cases extremely reliant on the assump-

[4] Included are DOLS (Phillips and Loretan (1991), Saikkonen (1991), Stock and Watson (1993)), the Phillips and Hansen (1990) procedure, the Johansen (1988) procedure, and the Saikkonen (1992) method.

[5] Hence the use of local to unity asymptotics; they provide useful approximations to small sample results.

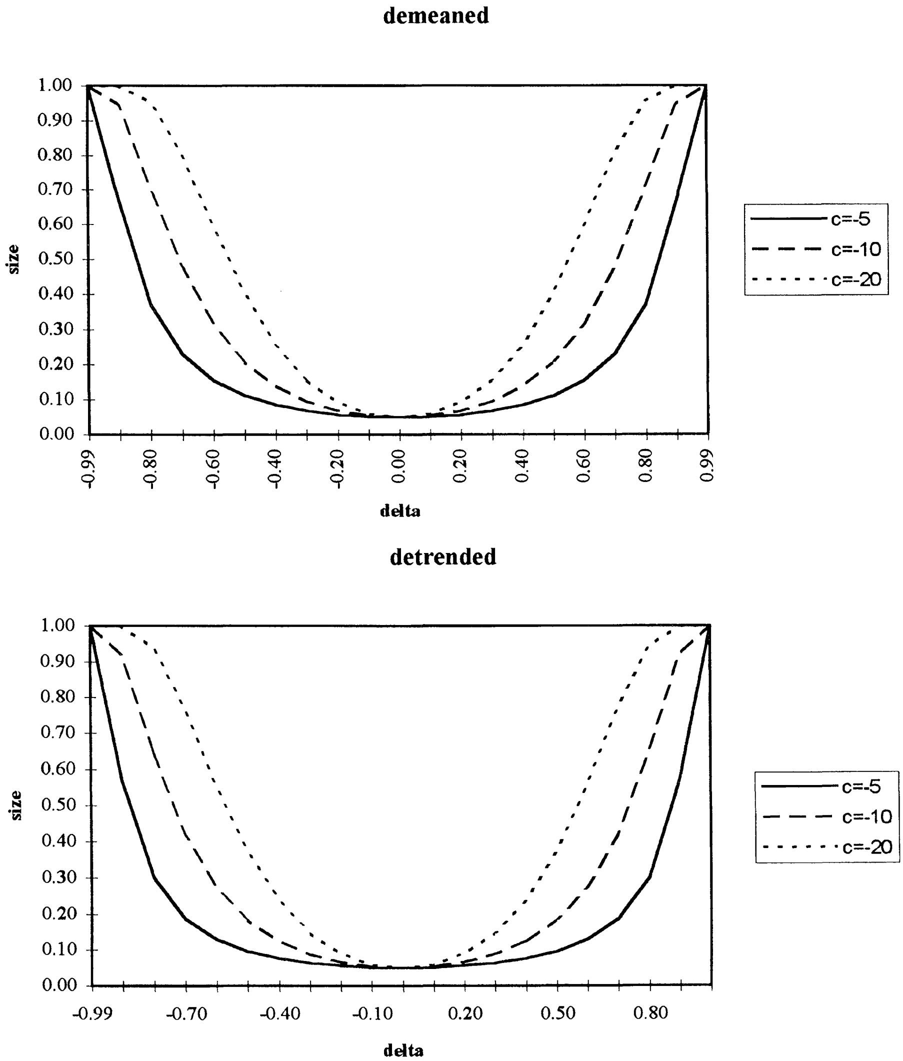


FIGURE 1.—Rejection rates when $A \neq 1$.

tion of exact unit roots in the model. Size distortions can be very large. In many cases the researcher is not interested in cointegration or integration per se, but is interested only in conducting hypothesis tests on the relationship between the variables. It is shown in the paper that for near unit roots, those for which typical unit root or rank tests have low power to reject, extremely large size distortions may occur from approximating slowly mean reverting processes by ones with unit roots.

In terms of correcting for these size distortions, the problem disappears when there is no simultaneity in the long run. Phillips and Hansen (1990) and Hsiao (1995) examine instrumenting for the endogenous regressors in the univariate and multivariate models

TABLE I

SIZE DISTORTIONS—BIVARIATE MODEL

| $\delta$ | Asy | DOLS | Fully Modified | Johansen | Saikonnen |
|---|---|---|---|---|---|
| 0.0 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 |
| 0.1 | 0.06 | 0.05 | 0.05 | 0.06 | 0.06 |
| 0.2 | 0.07 | 0.07 | 0.07 | 0.07 | 0.08 |
| 0.3 | 0.10 | 0.10 | 0.11 | 0.10 | 0.11 |
| 0.4 | 0.14 | 0.15 | 0.16 | 0.13 | 0.16 |
| 0.5 | 0.21 | 0.22 | 0.25 | 0.20 | 0.23 |
| 0.6 | 0.32 | 0.34 | 0.37 | 0.29 | 0.33 |
| 0.7 | 0.48 | 0.49 | 0.54 | 0.43 | 0.48 |
| 0.8 | 0.71 | 0.72 | 0.77 | 0.64 | 0.71 |
| 0.9 | 0.95 | 0.95 | 0.97 | 0.91 | 0.94 |

*Notes*: Column 1 gives different values for $\delta = \Omega_{12}/(\Omega_{11}\Omega_{22})^{1/2}$. Entries in column 2 evaluate Pr ($\chi_1^2 + B^* > 3.84$) using 20000 replications and $T = 1000$ for $c = -10$ and each $\delta$. Entries in the final four columns are size adjusted rejection rates for each of the estimators testing the true cointegrating vector over 5000 replications. The data were generated according to equation (1) where $n_1 = n_2 = 1$, there are no deterministic terms in the generated data; however a constant term is included, $A = 0.90$ and $T = 100$ (results are invariant to $\Gamma$). An initial condition of $y_0 = 0$ was imposed and the first five observations dropped. The nominal size is 5%. Similar results were shown for $T = 500$ and $A = 0.98$ in Elliott (1995).

respectively (the latter in a 2SLS framework). If the instruments are exogenous at the zero frequency, then this will eliminate the problems discussed here. It may be difficult to find such instruments; typically in dynamic models lagged regressors are employed but the shocks to these are necessarily correlated at the zero frequency with $u_{2t}$ if they are when they are not lagged. The common approach of using lags of the data as instruments would be ruled out. Cavanagh et al. (1995) show for a simple stylized bivariate model that the uncertainty over the unit root in $y_{1t}$ can be formally incorporated into the hypothesis tests through a Bonferroni procedure. Pesaran and Shin (1995) note that one can still conduct asymptotic normal inference on the long run effect of $y_{1t}$. Stock and Watson (1996) consider a bootstrap method.

*Dept. of Economics, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0508; gelliott@weber.ucsd.edu*

*Manuscript received May, 1995; final revision received April, 1997.*

## APPENDIX

The proofs make use of the convergence results collected in Lemma 1.

The VAR in (2) can be written

$$\text{(A.1)} \qquad \Delta y_t = \tilde{\beta}_2 z_t + \underline{\Psi}_1 y_{1t-1} + \Psi_2 \bar{u}_{2t-1} + \Pi(L)\Delta y_{t-1} + \varepsilon_t^*$$

where $\Psi P = [\underline{\Psi}_1, \Psi_2]$ and the partition is after the $n_1$th column, and $\bar{u}_{2t-1} = y_{2t-1} - \Gamma y_{1t-1}$. Here, we have that $\Psi_2$ is equal to the last $n_2$ columns of $\Psi$ and

$$\underline{\Psi}_1 = P\Phi(1)P^{-1}\begin{bmatrix} A - I \\ \Gamma A - \Gamma \end{bmatrix},$$

which is zero under the maintained hypothesis of $A = I$.

Whilst this is infeasible it can be used as a device (as in Saikkonen (1992) and more generally in Sims et al. (1990) when $C = 0$) to obtain the required limit distributions. We can write the equation as

$$\begin{aligned}\Delta y_t &= \underline{\Psi}_1(y_{1t-1} - d_{1t-1}) + \beta_2 z_t + \Psi_2(\bar{u}_{2t-1} - d_{2t}) + \Pi(L)(\Delta y_{t-1} - K_{t-1}) + \varepsilon_t^* \\ &= \beta x_t + \varepsilon_t^*\end{aligned}$$

where $\beta = [\underline{\Psi}_1, \beta_2, \beta_3]$, $\beta_2$ is $n \times s$ ($s = 1$ if $z_t = 1, s = 2$ if $z_t = (1,t)'$), $\beta_3$ is $n \times l$ ($l = n_2$ plus $n$ times the number of lags in $\Phi(L)$), $x_t = [(y_{1t-1} - d_{1t-1})', x_{2t}', x_{3t}']'$ and $K_t = E[\Delta y_t]$ where now all of the deterministic variables are collected in $x_{2t}$ and the mean zero stationary variables (except $y_{1t}$) are collected into $x_{3t}$. Define $\Lambda$ as a square scaling matrix with upper left-hand block $TI_{n1}$, middle block $s \times s$ with zeros in off diagonals, $(1,1)$ element $\sqrt{T}$ and $(2,2)$ element $T^{3/2}$ (if time is included as a regressor) and lower right-hand block $\sqrt{T}I_l$, with all other elements zero; then we have

$$(\hat{\beta} - \beta)\Lambda = \sum \varepsilon_t^* x_t' \Lambda^{-1}\left(\Lambda^{-1} \sum x_t x_t' \Lambda^{-1}\right)^{-1}.$$

LEMMA 1: *The following convergence results hold*:[6]

(i) $$\frac{1}{\sqrt{T}} \sum_{t=1}^{[T\lambda]} \varepsilon_t \Rightarrow \Sigma^{1/2} W(\lambda),$$

(ii) $$\frac{1}{\sqrt{T}}(y_{1t} - d_{1t}) \Rightarrow \Omega^{1/2} J_c(\lambda),$$

(iii) $$\left(\Lambda^{-1} \sum x_t x_t' \Lambda^{-1}\right) \Rightarrow \begin{bmatrix} \Omega_{11}^{1/2} \int J_c J_c' \Omega_{11}^{1/2\prime} & \Omega_{11}^{1/2} \int J_c D' & 0 \\ \int D J_c' \Omega_{11}^{1/2\prime} & \int DD' & 0 \\ 0 & 0 & M_{33} \end{bmatrix},$$

(iv) $$\sum \varepsilon_t^* x_t' \Lambda^{-1} \Rightarrow \left[ P\Phi(1)\Omega^{1/2} \int dW J_c' \Omega_{11}^{1/2\prime}, N_1, N_2 \right],$$

*where $W(\lambda)$ is an $n \times 1$ standard Brownian motion, $\Rightarrow$ denotes weak convergence, $[.]$ is the greatest lessor integer function, the limit distributions in (iii) are partitioned after the $n_1$th row and column and after the $(n_1 + s)$th row and column, the limit results in (iv) are partitioned by columns conformably with the results in (iii), $D = 1$ if a constant only is included and $D' = (1, \lambda)$ when a time trend is included, and $N_1$ and $N_2$ are (mutually correlated) multivariate normal distributions.*

PROOF: The first result follows from the multivariate FCLT (see Wooldridge (1994) for review). The second follows from Phillips (1987).

The results in (iii) can be taken block by block. The lower right block is $(1/T)\Sigma x_{3t} x_{3t}' \xrightarrow{p} M_{33}$ where $M_{33}$ is the variance covariance matrix of these variables as $x_{3t}$ are mean zero stationary variables. The $(2,3)$ block is 0 as $x_{3t}$ is mean zero. The $(2,2)$ block follows directly from calculation. The result in the $(1,1)$ block follows directly from (ii). The $(1,3)$ block is $(1/T^{3/2})\Sigma(y_{1t-1} - d_{1t-1})x_{3t}'$. As $x_{3t}$ is stationary this term is $o_p(1)$. The $(1,2)$ block in the demeaned case $z_t = 1$ converges to $\Omega_{11}^{1/2} \int J_c \, d\lambda$ by (ii) and the continuous mapping theorem. In the detrended case $z_t' = [1, t]$ and the term converges to $[\Omega_{11}^{1/2} \int J_c \, d\lambda, \Omega_{11}^{1/2} \int \lambda J_c \, d\lambda]$. The results in (iv) also can be taken block by block. The first is

$$T^{-1} \sum \varepsilon_t^* (y_{1t-1} - d_{1t-1})' = PT^{-1} \sum \varepsilon_t (y_{1t-1} - d_{1t-1})' \Rightarrow P\Phi(1)\Omega^{1/2} \int (dW J_c') \Omega_{11}^{1/2\prime}$$

[6] For brevity the $(\lambda)$ will be dropped from Brownian motion terms in longer expressions.

(Phillips (1987)). The second block converges to $P\Phi(1)\Omega^{1/2}\int dWD'$ which has a normal distribution. The final piece follows by the stationarity of $x_{3t}$.

PROOF OF THEOREM 1: From the results of Lemma 1 and rearranging,

$$T(\underline{\hat{\Psi}}_1 - \underline{\Psi}_1) \Rightarrow P\Phi(1)\Omega^{1/2}\int dWJ_c^{d\prime}\left(\int J_c^d J_c^{d\prime}\right)^{-1}\Omega_{11}^{-1/2}. \tag{A.2}$$

The Saikkonen (1992) estimator of the cointegrating vector is given by

$$\hat{\Gamma} = -(\hat{\Psi}_2'\hat{\Sigma}^{*-1}\hat{\Psi}_2)^{-1}(\hat{\Psi}_2'\hat{\Sigma}^{*-1}\hat{\Psi}_1).$$

The assumption $A = I$ yields $\underline{\Psi}_1 = \Psi_1 + \Psi_2\Gamma = 0$ so with scaling we have

$$\begin{aligned} T(\hat{\Gamma} - \Gamma) &= -(\hat{\Psi}_2'\hat{\Sigma}^{*-1}\hat{\Psi}_2)^{-1}(\hat{\Psi}_2'\hat{\Sigma}^{*-1}T\underline{\hat{\Psi}}_1) \\ &= -(\hat{\Psi}_2'\hat{\Sigma}^{*-1}\hat{\Psi}_2)^{-1}(\hat{\Psi}_2'\hat{\Sigma}^{*-1}T(\underline{\hat{\Psi}}_1 - \underline{\Psi}_1)) \\ &\quad -(\hat{\Psi}_2'\hat{\Sigma}^{*-1}\hat{\Psi}_2)^{-1}(\hat{\Psi}_2'\hat{\Sigma}^{*-1}T\underline{\Psi}_1) \\ &= -(\Psi_2'\Sigma^{*-1}\Psi_2)^{-1}(\Psi_2'\Sigma^{*-1}T(\underline{\hat{\Psi}}_1 - \underline{\Psi}_1)) \\ &\quad -(\Psi_2'\Sigma^{*-1}\Psi_2)^{-1}(\Psi_2'\Sigma^{*-1}T\underline{\Psi}_1) + o_p(1) \end{aligned} \tag{A.3}$$

where the last line follows due to the consistency of $\hat{\Psi}_2$ and $\hat{\Sigma}^*$ (which follows from the consistency of $\hat{\beta}$). Let the $n_1 \times n$ selector matrix partitioned after the $n_1$th column $e_2' = [0, I_{n2}]$; then

$$\Psi_2 = P\Phi(1)P^{-1}Me_2,$$

$$(\Psi_2'\Sigma^{*-1}\Psi_2) = \Omega_{2.1}^{-1},$$

$$(\Psi_2'\Sigma^{*-1}) = -e_2'\Phi(1)'\Sigma^{-1}P^{-1}, \quad \text{and}$$

$$(\Psi_2'\Sigma^{*-1}T\underline{\Psi}_1) = \Omega_{2.1}^{-1}\Omega_{21}\Omega_{11}^{-1}T(A - I).$$

Substituting these results and (A.2) into (A.3) gives the result

$$T(\hat{\Gamma} - \Gamma) \Rightarrow \Omega_{2.1}^{1/2}\int dW_{2.1}J_c^{d\prime}\left(\int J_c^d J_c^{d\prime}\right)^{-1}\Omega_{11}^{-1/2} + B$$

where $W_{2.1} = W_2 - \Omega_{21}\Omega_{11}^{-1}W_1$, $W$ is a standard Brownian motion on [0 1], and $W = [W_1', W_2']'$ and the partition is after the $n_1$th row. We can rearrange this for the result

$$\hat{V}^{-1/2}\text{vec}\{T(\hat{\Gamma} - \Gamma) - B\} \sim_a \text{vec}\left[dW_{2.1}J_c^{d\prime}\left(\int J_c^d J_c^{d\prime}\right)^{-1/2}\right]$$

where $\sim_a$ is "approximately distributed." Conditional on $y_{1t}$, the (vectorization) of the approximate distribution is a standard multivariate normal distribution (Park and Phillips (1988)). Rearranging yields the result in the statement of the theorem.

Saikkonen (1992) shows that the likelihood for the model assuming $A = I$ (with no deterministics) yields the first order conditions

$$\Psi_2'\Sigma^{*-1}\sum(\Delta y_t - \Psi_2\theta' y_{t-1} - \Pi x_{3t})'y_{1t-1}$$

where the summation is over available datapoints and $\theta = [\Gamma, -I]$. The above estimate solves this first order equation and is hence the MLE; thus the MLE has the same asymptotic distribution.

PROOF OF COROLLARY TO THEOREM 1: From Theorem 1 and $\hat{V} \Rightarrow V$,

$$\hat{V}^{-1/2}(\text{vec}[T(\hat{\Gamma} - \Gamma)] - \text{vec}(B)) \sim_a N(0, I_{n1n2}).$$

When $R\,\text{vec}(\Gamma) = r$ is true, then we have

$$(R\hat{V}R')^{-1/2}\{R\,\text{vec}[T(\hat{\Gamma} - \Gamma)] - R\,\text{vec}(B)\} \sim_a N(0, I_q). \tag{A.4}$$

It follows directly that

$$\{R\,\text{vec}[T(\hat{\Gamma} - \Gamma)] - R\,\text{vec}(B)\}'\{R\hat{V}R'\}^{-1}\{R\,\text{vec}[T(\hat{\Gamma} - \Gamma)] - R\,\text{vec}(B)\} \sim_a \chi_q^2. \tag{A.5}$$

This is equivalent to the usual Wald statistic only if $R\,\text{vec}(B) = 0$.

Rearranging (A.5), and using the result (A.4), we obtain the result stated in the theorem.

## REFERENCES

AHN, S. K., AND G. C. REINSEL (1990): "Estimation for Partially Nonstationary Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

BOBKOSKI, M. J. (1983): "Hypothesis Testing in Nonstationary Time Series," Unpublished Ph.D. Thesis, Department of Statistics, University of Wisconsin.

CAVANAGH, C. (1985): "Roots Local to Unity," Manuscript, Department of Economics, Harvard University.

CAVANAGH, C. L., G. ELLIOTT, AND J. H. STOCK (1995): "Inference in Models with Nearly Nonstationary Regressors," *Econometric Theory*, 11, 1131–1147.

CHRISTIANO, L., AND M. E. EICHENBAUM (1990): "Unit Roots in GNP: Do We Know and Do We Care?" Carnegie Rochester Conference Series on Public Policy, 32, 7–62.

ELLIOTT, G. (1995): "On the Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," UCSD Discussion Paper 95-18.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

ENGLE, R. F., AND C. W. GRANGER (1987): "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica*, 55, 251–276.

HSIAO, C. (1995): "Statistical Properties of the Two Stage Least Squares Estimator Under Cointegration," Manuscript, University of Southern California.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregression Models," *Econometrica*, 59, 1551–1580.

NELSON, C. R., AND C. I. PLOSSER (1982): "Trends and Random Walks in Macroeconomic Time Series; Some Evidence and Implications," *Journal of Monetary Economics*, 10, 139–162.

NG, S., AND P. PERRON (1996): "Estimation and Inference in Nearly Unbalanced, Nearly Cointegrated Systems," Working Paper 3295 CRDE, Université de Montréal.

PARK, J. Y., AND P. C. B. PHILLIPS (1988): "Statistical Inference in Regressions with Integrated Processes: I," *Econometric Theory*, 4, 467–497.

PESARAN, M. H., AND Y. SHIN (1995): "An Autoregrressive Distributed Lag Modeling Approach to Cointegration Analysis," D.A.E. Working Paper No. 9514, University of Cambridge.

PHILLIPS, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrica*, 74, 535–547.

PHILLIPS, P. C. B., AND B. E. HANSEN (1990): "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

PHILLIPS, P. C. B., AND M. LORETAN (1991): "Estimating Long-Run Economic Equilibria," *Review of Economic Studies*, 59, 407–436.

Saikkonen, P. (1991): "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1–21.

——— (1992): "Estimation and Testing of Cointegrated Systems by an Autoregressive Approximation," *Econometric Theory*, 8, 1–27.

Sims, C. A., J. H. Stock, and M. W. Watson (1990): "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

Stock, J. H., and M. W. Watson (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

——— (1996): "Confidence Sets in Regressions with Highly Serially Correlated Regressors," Manuscript, Harvard University.

Watson, M. W. (1994): "Vector Autoregressions and Cointegration," in *Handbook of Econometrics*, Vol. 4, ed. by D. McFadden and R. F. Engle. Amsterdam: North Holland, 2843–2915.

Wooldridge, J. (1994): "Estimation and Inference for Dependent Processes," in *Handbook of Econometrics*, Vol. 4, ed. by D. McFadden and R. F. Engle. Amsterdam: North Holland, 2639–2738.