Journal of Econometrics 54 (1992) 139–158. North-Holland

Heteroskedastic cointegration*

Bruce E. Hansen
University of Rochester, Rochester, NY 14627, USA

Received July 1990, final version received October 1991

This paper develops an asymptotic theory of estimation and inference in 'cointegrated' regression models with errors displaying nonstationary variances. Least squares estimates are shown to be consistent at a $T^{1/2}$ rate. Hypothesis testing requires the use of a robust covariance matrix estimate, in contrast to earlier work on cointegrated regressions. The inference theory is not nuisance-free, but preliminary investigations indicate that approximation by the normal distribution may be adequate in practice.

1. Introduction

There has been a recent explosion of theoretical and empirical interest in the model of cointegration proposed by Granger (1981) and developed by Engle and Granger (1987). This model is popular because it allows researchers to take seriously two seemingly contradictory facts: (1) economic data typically appear to possess unit roots (that is, have stochastic trends); yet (2) economic theory often suggests 'equilibrium' or long-run relationships may exist between variables. The model of cointegration reconciles these facts by allowing a linear combination of individually I(1) series to be I(0);[1] equivalently, the residual error in a linear regression is taken to be stationary.

It is not clear, however, that the model of cointegration (which we will refer to as the CI model) as formulated by Engle and Granger is sufficiently general to cover all nonstationary economic models of interest. The CI regression errors differ stochastically from the regressors in that they have a

*This paper is a revised version of chapter four of my Ph.D. dissertation. I would like to thank Don Andrews, Vassilis Hajivassiliou, Peter Phillips, the associate editor, and two referees for helpful advice and comments.

[1]A series X_t is integrated of order d, denoted I(d), if it has no deterministic components and its dth difference, $\Delta^d X_t$, is I(0). A series is I(0) if it has a spectral density which is bounded away from zero and infinite at the zero frequency.

Table 1
Sample split tests on regression error variance.[a]

	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	t-test
Total consumption upon income	688	2322	2.77
NDS consumption upon income	786	1362	1.36
Stock prices upon dividends	0.002	0.012	4.00
Long interest rates upon short rates	0.326	2.642	3.80

[a] $\hat{\sigma}_1^2$ = regression error variance estimate over first half sample; $\hat{\sigma}_2^2$ = regression error variance estimate over second half sample; t-test = t-test of hypothesis that regression error variance is same in two sample halves, computed with Bartlett weights with lag window of five.

fixed mean and a bounded variance. The asymmetry in variance orders is intuitively unsatisfying in some cases. One might expect that as the regressors increase in magnitude, the residual variance would also increase. One might also expect that the variance of the error process might change over time, due to other factors. Essentially, we may wish to allow the variance of the error to be nonstationary.

The empirical relevance of this idea can be illustrated by calculating simple sample split tests on the variance of the regression error in four published cointegrating regressions. The results are given in table 1. The first regression is of aggregate consumption upon disposal income. The second regression is of aggregate nondurables and services consumption upon disposable income. These are taken from Campbell (1987).[2] The third regression is of a stock price index upon dividends, as reported in Campbell and Shiller (1987).[3] The fourth regression is of short-run interest rates (three-month T-bill) on long-run rates (10^+ year).[4] A similar regression appears in Campbell and Shiller (1987). Each regression is computed with only a constant in addition to the stated variable. The regression error variance is calculated for the first and second halves of the sample, and the t-test for the hypothesis that the variance is the same is computed. Since the squared regression errors are serially dependent, the t-statistic is calculated using a robust variance estimate, using a Bartlett window with a lag length of five [see Newey and West (1987)].

[2]The data is originally from Blinder and Deaton (1985). They are seasonally adjusted quarterly series for the period 1953:2–1984:4, measured as per capita real aggregates.

[3]The data, real annual prices and dividends for the period 1871–1986, was courteously provided by Robert Shiller.

[4]The data is quarterly nominal interest rates for the period 1947–1986, extracted from the Citibase files FYGM3 and FYGL.

The t-tests give strong evidence for nonconstancy of the error variance in three of the four cointegrating regressions.[5] The regressions appear to be cointegrating, in the sense that the two variables trend together over the long run, but the regression error appears to violate the asymptotic stationarity assumed in the conventional theory of cointegration. Of course, since the sample split tests have power against a wide variety of alternatives, we cannot conclude from this one piece of evidence anything about the *form* of the nonstationarity in the regression error variance.

If the variance of the regression error is asymptotically nonstationary, then it should not be ignored in our theoretical and empirical investigations. Since the properties of the error process are unlikely to be well known *a priori*, we need a theoretical specification flexible enough to encompass the cases of interest, yet sufficiently tractable to allow the development of an asymptotic theory. Consider the process w_t generated by $w_t = \sigma_t u_t$, where $\sigma_t \equiv \mathrm{I}(1)$ and $u_t \equiv \mathrm{I}(0)$. We call this a bi-integrated (BI) process. We can think of u_t as the 'stationary part' of w_t and of σ_t (or, more precisely, σ_t^2) as the 'variance part'. If y_t is generated by $y_t = \beta' x_t + w_t$, where $x_t \equiv \mathrm{I}(1)$ and $w_t \equiv \mathrm{BI}$, then we say that (y_t, x_t) are heteroskedastically cointegrated. The model, which we call the model of heteroskedastic cointegration (HCI), may be justified on several grounds:

(1) Both x_t and w_t have variances which grow at the same rate. This removes one potential attack on the standard CI model – that the latter derives all its estimation power from the differing stochastic orders of the regressors and residuals. It is well known that regression on trended variables is consistent even under endogeneity. Therefore, the conclusion may have been drawn that the consistency of OLS in the CI model was a consequence of this stochastic order difference. This, in fact, is hinted at in the discussion by Stock (1987). Our results show that the HCI model may be estimated consistently by least squares even under endogeneity. The key requirement is the fixed mean property of the residuals.

(2) The use of time-varying parameter (TVP) models in economics has been fairly popular. For instance, in the linear regression $y_t = \beta_t x_t + u_t$, if $\{x_t, \beta_t, u_t\} \equiv \mathrm{I}(0)$, then estimation of $\beta = \mathrm{E}(\beta_t)$ can be handled using standard methods, if the heteroskedasticity of the residuals is taken into account. If, however, $x_t \equiv \mathrm{I}(1)$, the model becomes an HCI model, with $\sigma_t = x_t$. (An extra noise term, u_t, is present, but this is irrelevant asymptotically.) This model is intuitively plausible, for it suggests that the residual will be proportional to the regressor; i.e., big innovations occur more frequently when the regressor process reaches large values.

[5] In the first three regressions (although not the fourth) the apparent heteroskedasticity disappears if the regressions are estimated in logarithms. The cited papers, however, estimated the models in levels.

Andrews (1987) was the first to study an amended CI model with mildly explosive residuals. Specifically, he showed that if $\text{var}(w_t) = O_p(t^{\alpha})$, $0 \leq \alpha < 1$, then least squares is consistent and converges at least as fast as $T^{(1-\alpha)/2}$. The BI errors of our paper, however, are outside the scope of Andrews' analysis as we have $\alpha = 1$. This work is also related to an earlier literature on dynamic regression with nonstandard normalizations. A partial list includes Robinson (1978), Anderson and Taylor (1979), and Wooldridge and White (1988).

This paper covers the following ground. Section 2 outlines the model and assumptions. Section 3 derives an asymptotic theory of least squares estimation and inference for this model. We find that least squares estimation is consistent under general assumptions, yet the asymptotic distribution is neither median unbiased nor mixture-normal under endogeneity. Under an exogeneity assumption, the limiting distribution is mixture-normal, permitting valid chi-square inference if a robust covariance matrix estimate is used. Section 4 examines simulated plots of the distributions under more general assumptions. The conclusion contains some suggestions for future research.

A word on the notation before we begin. The symbol $\Rightarrow$ denotes weak convergence, $\equiv$ signifies equality in distribution, and $\|A\|_p = (\sum_{jk} \mathrm{E}|A_{jk}|^p)^{1/p}$ denotes the L^p-norm for random matrices. Stochastic processes such as the Brownian motion $B(r)$ on $[0,1]$ are frequently written as B to achieve notational economy. Similarly, integrals such as $\int_0^1 B(r)$ are written more simply as $\int_0^1 B$.

2. The model and assumptions

Consider the linear regression model

$$y_t = \beta_0 + \beta_1' x_t + w_t, \tag{1}$$

where the $n \times 1$ regressor vector is I(1):

$$x_t = x_{t-1} + u_{3t}, \tag{2}$$

and the error w_t is a bi-integrated process, as defined in the introduction. That is,

$$w_t = \sigma_t u_{1t}, \tag{3}$$

where the scale process σ_t is I(1):

$$\sigma_t = \sigma_{t-1} + u_{2t}. \tag{4}$$

The initializations x_0 and σ_0 may be any random variables which have finite absolute expectations.

We call model (1)–(4) a model of *heteroskedastic cointegration*, abbreviated by HCI, and is motivated by the issues raised in the introduction.

Note that

$$\operatorname{var}(x_t) \approx C_1 t, \qquad 0 < C_1 < \infty,$$

$$\operatorname{var}(w_t) \approx C_2 t, \qquad 0 < C_2 < \infty,$$

and are thus of the same stochastic order. This is a substantial difference from the standard Engel–Granger cointegration model, where the variance of the regression errors is constant. Although they have variances of the same stochastic order, the variables behave quite differently from one another. The I(1) process x_t wanders around with no tendency to return to any particular value. The BI process w_t, on the other hand, will tend to cross its mean value, Λ_{21}, often.

Define the vector $u_t = (u_{1t}, u_{2t}, u'_{3t})'$ and the partial sum $S_t = \sum_{j=1}^{t} u_j$.

Assumption 1. The process $\{u_t\}$ is mean zero and strong mixing (α-mixing) with mixing coefficients $\{\alpha_m\}$ of size $-3p/(p-3)$, and $\sup_t \mathrm{E}|u_t|^p < \infty$ for some $p > 3$.

Define the long-run covariance matrix

$$\Omega = \lim_{T \uparrow \infty} T^{-1} \mathrm{E}(S_T S'_T)$$

and the matrix

$$\Lambda = \lim_{T \uparrow \infty} T^{-1} \sum_{1}^{T} \mathrm{E}(S_t u'_t).$$

Assumption 1 is sufficient for the invariance principle,

$$T^{-1/2} S_{[Tr]} \Rightarrow B(r) \equiv BM(\Omega), \tag{5}$$

convergence to the matrix stochastic integral,

$$T^{-1} \sum_{1}^{T} S_t u'_t \Rightarrow \int_0^1 B \, \mathrm{d}B' + \Lambda, \tag{6}$$

and convergence to a product stochastic integral

$$T^{-3/2}\sum_{1}^{T}(S_t\otimes S_t)u_t'\Rightarrow\int_0^1(B\otimes B)\,\mathrm{d}B'+\Lambda\otimes\int_0^1 B+\int_0^1 B\otimes\Lambda. \quad (7)$$

Results (5), (6), and (7) have been shown by Herrndorf (1984, corollary 1) and Hansen (1991c, theorems 4.1, 4.2), respectively.

We partition B, Ω, and Λ conformably with u_t. Thus in the case of Ω we write

$$\Omega=\begin{bmatrix}\Omega_{11} & \Omega_{12} & \Omega_{13}\\ \Omega_{21} & \Omega_{22} & \Omega_{23}\\ \Omega_{31} & \Omega_{32} & \Omega_{33}\end{bmatrix}.$$

We will require that $\Omega_{33}>0$, $\Omega_{22}>0$, and $\Omega_{11}>0$, and normalize $\mathrm{E}u_{1t}^2=1$.

3. Estimation and inference

In this section we examine the asymptotic distributions of the least squares estimate of β in the HCI model. Denote this estimator by $\hat\beta=(\hat\beta_0\ \hat\beta_1')'$. Define the matrix $\delta_T=\mathrm{diag}(1,\sqrt{T})$. The main finding is reported in Theorem 1.

Theorem 1. Under Assumption 1,

$$\delta_T(\hat\beta-\beta^*)\Rightarrow\left(\int_0^1 XX'\right)^{-1}\left(\int_0^1 XB_2\,\mathrm{d}B_1+\begin{pmatrix}0\\ \Lambda_{31}\int_0^1 B_2\end{pmatrix}\right),$$

where $\beta^*=(\beta_0^*\ \beta_1')'$, $\beta_0^*=\beta_0+\Lambda_{21}$, *and* $X(r)=(1\ B_3(r'))'$.

Remark 1. The cointegrating slope parameter β_1 is estimated consistently by least squares. The rate of convergence is the square root of sample size, as in standard asymptotic theory. This obtains regardless of the nature of the dependence between the innovations u_{1t}, u_{2t}, and u_{3t}. That is, no assumption of exogeneity is required to obtain consistency, even though the regressors and regression error are of the same stochastic order.

Remark 2. The intercept β_0 is not estimated consistently by least squares. As β_0 is not typically of economic interest, this is potentially only a problem for the estimation of covariance parameters, but we will show later that the inconsistency of $\hat\beta_0$ is irrelevant asymptotically.

Remark 3. The asymptotic distributions in Theorem 1 are nonstandard and are not generally mixtures of normals. The unusual structure of the regressors and regression errors implies that we cannot apply existing techniques which obtain mixture normality in cointegrated models, such as MLE [see Johansen (1988)] or semiparametric estimation [see Phillips and Hansen (1990)]. In order to progress further, we need to impose a stronger set of covariance assumptions.

Assumption 2. $\mathrm{E}(x_t u_{1t}) = 0$.

This states that the stationary part of the regression error is orthogonal to the regressor. Assumption 2 directly implies that $\Lambda_{31} = 0$, so Theorem 1 reduces to

$$\delta_T(\hat{\beta} - \beta^*) \Rightarrow \left(\int_0^1 XX'\right)^{-1} \int_0^1 XB_2\,\mathrm{d}B_1. \tag{8}$$

An improved understanding of the distribution in (8) is obtained by defining the stochastic process

$$J(r) = \int_0^r B_2\,\mathrm{d}B_1. \tag{9}$$

$J(r)$, like the Brownian motion $B_1(r) = \int_0^r \mathrm{d}B_1$, is a continuous time martingale. We can contrast their behavior via their quadratic variation processes. To review, for any semimartingale X the quadratic variation process is defined as

$$[X, X]_r = X(r)^2 - 2\int_0^r X\,\mathrm{d}X.$$

For the Brownian motion $B_1(r)$, Ito's lemma gives

$$[B_1, B_1]_r = B_1(r)^2 - 2\int_0^r B_1\,\mathrm{d}B_1 = r\Omega_{11}.$$

In contrast, for the process $J(r)$ we find

$$[J, J]_r = \int_0^r B_2(s)\,\mathrm{d}[B_1, B_1]_s = \int_0^r B_2\Omega_{11},$$

as shown in Protter (1990, theorem 29). This gives rise to the heuristic expressions

$$(\mathrm{d}B_1)^2 = \Omega_{11}\,\mathrm{d}r \quad \text{and} \quad (\mathrm{d}J)^2 = \Omega_1 B_2(r)^2\,\mathrm{d}r.$$

What is especially useful about the process $J(r)$ is that by the associative property of semimartingales [Protter (1990, theorem 19)],

$$\int_0^1 XB_2\,\mathrm{d}B_1 = \int_0^1 X\,\mathrm{d}J.$$

This allows us to write the limit distribution in (8) as

$$\left(\int_0^1 XX'\right)^{-1} \int_0^1 XB_2\,\mathrm{d}B_1 = \left(\int_0^1 XX'\right)^{-1} \int_0^1 X\,\mathrm{d}J, \tag{10}$$

which resembles in form the distributions obtained in standard cointegration theory [see, for example, Park and Phillips (1988)].

The distribution in (10) is not a mixture of normals since B_1 may be correlated with B_2 and/or B_3. We now impose this requirement.

Assumption 3. $\Omega_{21} = 0, \quad \Omega_{31} = 0.$

This assumption states that the stationary part of the regression error is 'long-run orthogonal' (spectral cohesion is zero at the zero frequency) to the innovations driving the regressors and the variance scale of the regression error. This is quite a restrictive assumption and is relaxed in the next section. It is helpful, however, in understanding the form of the distributional theory and the type of covariance matrix estimator needed.

Define $\mathscr{F}_{23} = \sigma(B_2(r), B_3(r)\colon 0 \le r \le 1)$, the smallest σ-field containing the history of the Brownian motions B_2 and B_3. Under Assumption 3, B_1 is independent of $\mathscr{F}_{23}$, so conditional on $\mathscr{F}_{23}$,

$$\int_0^1 X\,\mathrm{d}J = \int_0^1 XB_2\,\mathrm{d}B_1 \equiv N(0, V), \qquad V = \int_0^1 XX'B_2^2\Omega_{11}.$$

Unconditionally, we find

$$\int_0^1 X\,\mathrm{d}J \equiv \int_{V>0} N(0,V)\,\mathrm{d}P(V),$$

where $P(V)$ is the probability measure over the random matrix V defined above. To simplify notation, we will write the unconditional distribution simply as $N(0, V)$.

This discussion can be summarized in the following result:

Theorem 2. Under Assumptions 1, 2, and 3,

$$\delta_T(\hat{\beta} - \beta^*) \Rightarrow N(0, M^{-1}VM^{-1}), \tag{11}$$

where

$$M \equiv \int_0^1 XX' \quad \textit{and} \quad V \equiv \int_0^1 XX'B_2^2\,\mathrm{d}\Omega_{11} = \int_0^1 XX'(\mathrm{d}J)^2.$$

Remark 4. The distribution given in Theorem 2 is a variance mixture of normals. The random covariance matrix $M^{-1}VM^{-1}$ is of the form often found in estimators with heteroskedasticity and/or serial correlation. Interestingly, the covariance matrix for the least squares estimator in the standard cointegration model can also be put in this form, for in this case the quadratic variation is $(\mathrm{d}B_1)^2 = \Omega_{11}\,\mathrm{d}r$, and the matrix V is a scale multiple of M.

Remark 5. Consider the problem of testing linear hypotheses upon the slope coefficients of the form

$$\mathrm{H}_0: \quad R'\beta_1 = r, \quad \mathrm{rank}(R) = q.$$

First, observe that by partitioned matrix inversion we can obtain from (11)

$$\sqrt{T}\left(\hat{\beta}_1 - \beta_1\right) \Rightarrow N\left(0, M_1^{-1}V_1M_1^{-1}\right),$$

where $M_1 = \int_0^1 B_3^* B_3^{*\prime}$, $B_3^*(r) = B_3(r) - \int_0^1 B_3$, and $V_1 = \int_0^1 B_3^* B_3^{*\prime}(\mathrm{d}J)^2$. ($B_3^*$ is demeaned B_3.) Therefore, under H_0,

$$\sqrt{T}\left(R'\hat{\beta}_1 - r\right) \Rightarrow N\left(0, R'M_1^{-1}V_1M_1^{-1}R\right).$$

Inference upon the slope parameters β_1 can proceed conventionally if we can find a consistent estimate $\hat{V}_1$ of V_1, in the sense that $\hat{V}_1 \Rightarrow V_1$ (jointly with the slope parameter estimates). Since V_1 is not a scale multiple of M_1, the standard 'OLS' covariance matrix estimator will not achieve this goal. We therefore need to consider a covariance matrix estimator which is robust to heteroskedasticity and autocorrelation, as in White and Domowitz (1984). Define the Wald statistic

$$W = T\left(R'\hat{\beta}_1 - r\right)'\left(R'\hat{M}_1^{-1}\hat{V}_1\hat{M}_1^{-1}R\right)^{-1}\left(R'\hat{\beta}_1 - r\right),$$

where

$$\hat{M}_1 = T^{-1} \sum_1^T (x_1 - \bar{x})(x_t - \bar{x})',$$

$$\hat{w}_t = y_t - \hat{\beta}_0 - x_t' \hat{\beta}_1,$$

$$\hat{V}_1 = \frac{1}{T} \sum_{m=-B}^{B} k_m \sum_t (x_{t+m} - \bar{x})(x_t - \bar{x})' \hat{w}_{t+m} \hat{w}_t,$$

and the second summation is over t such that $1 \leq t + m \leq T$. The weights $\{k_m\}$ are selected so that, for each m, $\lim_{B \uparrow \infty} k_m = 1$. The bandwidth, or lag truncation number, B is selected to grow to infinity such that $B = \mathrm{o}(T^{1/4})$. We need the following strengthening of Assumption 1:

Assumption 4. The process $\{u_t\}$ *is mean zero and strong mixing* (α*-mixing*) *with mixing coefficients* $\{\alpha_m\}$ *of size* $-6p/(p-6)$, *and* $\sup_t \mathrm{E}|u_t|^p < \infty$ *for some* $p > 6$.

Theorem 3. Under Assumptions 2, 3, and 4, and $B = \mathrm{o}(T^{1/4})$,

$$\text{(a)} \quad T^{-2}\hat{V}_1 \Rightarrow V_1, \qquad \text{(b)} \quad \hat{W} \underset{\mathrm{d}}{\rightarrow} \chi_q^2$$

Remark 6. In the HCI model, the OLS covariance matrix estimator does not generate an appropriate metric, but an heteroskedasticity and autocorrelation-consistent covariance matrix estimator allows inference to proceed conventionally. In the literature on cointegrated models, little attention has been paid to robust covariance matrix estimation, since it is unnecessary in the standard model. Theorem 3 points out that there exist situations where the robust estimators are necessary.

4. Estimates of the nonstandard distributions

If Assumption 3 is violated, the inference procedures outlined in the previous section may be biased, as the test statistics will not have chi-square asymptotic distributions. We can get a feel for the magnitude of this bias by displaying the graphs of t-statistics for the simple case of one regressor. To review, in the CI model, the divergence from the $N(0,1)$ can be quite substantial. As shown by Park and Phillips (1988), the t-statistic in this case has an asymptotic distribution which is a mixture of an independent normal random variable and a Dickey–Fuller t-variate:

$$\hat{t} = N \cdot (1 - \underline{\Omega}_{13}^2)^{1/2} + DF \cdot \underline{\Omega}_{13}, \tag{12}$$

where $N \equiv N(0,1)$ and

$$DF \equiv \int_0^1 \overline{W}\, \mathrm{d}\overline{W} \Big/ \left(\int_0^1 \overline{W}^2 \right)^{1/2}, \qquad \overline{W}_a \equiv W_a - \int_0^1 W_a.$$

Here and elsewhere in this section we will use the compact notation

$$\underline{\Omega}_{ab} = \Omega_{aa}^{-1/2} \Omega_{ab} \Omega_{bb}^{-1/2}, \qquad a, b = 1, 2, 3.$$

The DF distribution is well known to have negative mean and skewness, and is not well approximated by the standard normal. For moderate values of $\underline{\Omega}_{31}$, therefore, the distribution in (12) will diverge considerably from the standard normal.

Returning to the HCI model, the distribution of the t-statistic in the simple one-regressor model can be written as

$$\frac{\int_0^1 \overline{B}_3 B_2\, \mathrm{d}B_1}{\left(\int_0^1 \overline{B}_3^2 B_2^2 \Omega_{11} \right)^{1/2}} \equiv \frac{\int_0^1 \underline{\overline{B}}_3 \underline{B}_2\, \mathrm{d}\underline{B}_1}{\left(\int_0^1 \underline{\overline{B}}_3^2 \underline{B}_2^2 \right)^{1/2}}, \tag{13}$$

where

$$\underline{B}_1 \equiv \left(1 - (1 - \underline{\Omega}_{23}^2)^{-1} (\underline{\Omega}_{21} - \underline{\Omega}_{23}\underline{\Omega}_{31})^2 - \underline{\Omega}_{13}^2\right)^{1/2} W_1$$

$$+ (1 - \underline{\Omega}_{23}^2)^{-1/2} (\underline{\Omega}_{21} - \underline{\Omega}_{23}\underline{\Omega}_{31}) W_2 + \underline{\Omega}_{13} W_3,$$

$$\underline{B}_2 \equiv (1 - \underline{\Omega}_{23}^2)^{1/2} W_2 + \underline{\Omega}_{23} W_3,$$

$$\underline{B}_3 \equiv W_3,$$

and

$$(W_1, W_2, W_3)' \equiv BM(I_3).$$

In general, this distribution seems rather complicated. The information we need, however, can be extracted by examining the leading cases: $B_2 = B_3$, and B_2 and B_3 independent. The first occurs when $\underline{\Omega}_{23} = 1$ (and therefore $\underline{\Omega}_{21} = \underline{\Omega}_{31}$). We can expect B_2 and B_3 to be related in this way, for example,

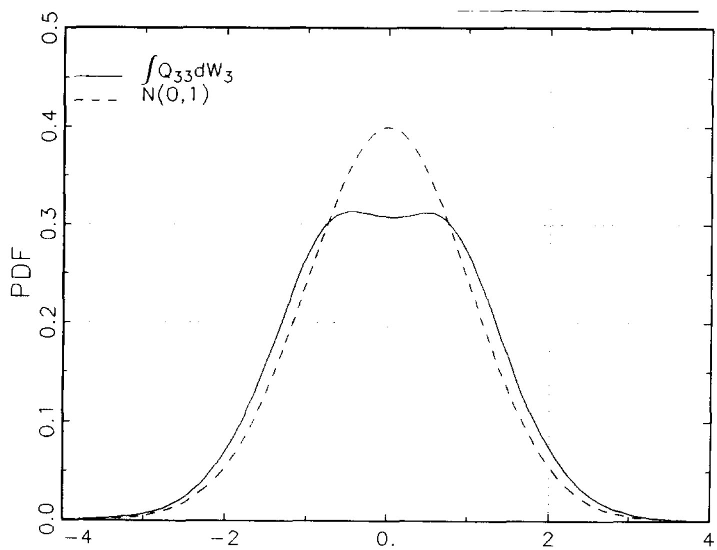

Fig. 1

when the BI error emerges due to a TVP process (see the remarks in the Introduction). The distribution in (13) then equals

$$N(0,1)\left(1-\underline{\Omega}_{13}^2\right)^{1/2}+\int_0^1 Q_{33}\,\mathrm{d}W_3\,\underline{\Omega}_{13},$$

where

$$Q_{33}(r)=\overline{W}_3(r)W_3(r)\Big/\left(\int_0^1\overline{W}_3^2W_3^2\right)^{1/2}.$$

Fig. 1 displays[6] the distributions $N(0,1)$ and $\int_0^1 Q_{33}\,\mathrm{d}W_3$, which are the two independent components of the distribution. The distribution of $\int_0^1 Q_{33}\,\mathrm{d}W_3$ has an symmetric, slightly bi-modal shape. The relevant fact for inference, however, is its behavior in the tails. The distribution is slightly more dispersed than the $N(0,1)$, and thus some size distortion will occur, if critical values are set using the $N(0,1)$. The degree of distortion, however, seems negligible.

The picture changes somewhat for the second leading case: B_2 and B_3 independent, which occurs when $\underline{\Omega}_{23}=0$. In this case the distribution in (13)

[6]The distributions were approximated by Monte Carlo simulation. 64,000 samples of size 1000 were drawn using pseudo-normal increments. A normal kernel estimated the density function.

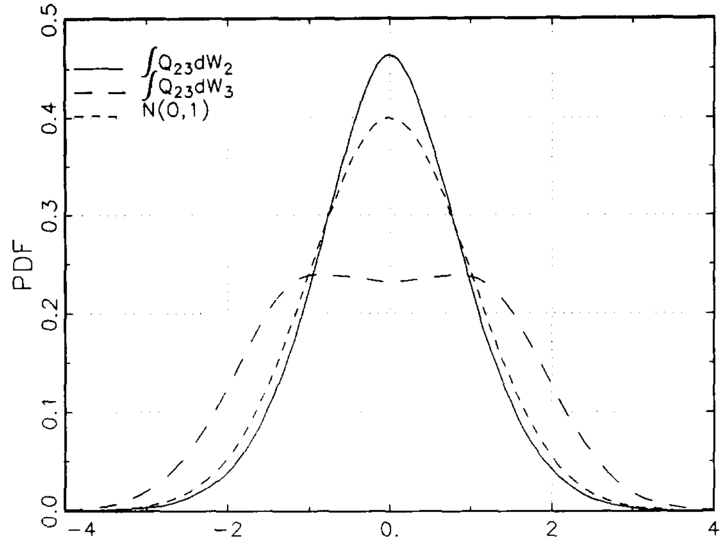

Fig. 2

equals

$$N(0,1)\left(1-\underline{\Omega}_{21}^2-\underline{\Omega}_{13}^2\right)^{1/2}+\int_0^1 Q_{23}(\underline{\Omega}_{21}\,\mathrm{d}W_2+\Omega_{31}\,\mathrm{d}W_3), \tag{14}$$

where

$$Q_{23}(r)=\overline{W}_3(r)W_2(r)\Big/\left(\int_0^1\overline{W}_3^2W_2^2\right)^{1/2}.$$

The $N(0,1)$ term in (14) is independent of the second term. In fig. 2 we display $N(0,1)$, $\int_0^1 Q_{23}\,\mathrm{d}W_2$, and $\int_0^1 Q_{23}\,\mathrm{d}W_3$. Interestingly, the second and third are quite different. $\int_0^1 Q_{23}\,\mathrm{d}W_2$ is bell-shaped and less dispersed than $N(0,1)$. $\int_0^1 Q_{23}\,\mathrm{d}W_3$, on the other hand, is significantly different from the standard normal. It is symmetric about zero, yet bi-modal and more dispersed than the $N(0,1)$. If the distribution in (14) is close to this shape, then inferences based upon the standard normal will be misleading. This will occur when both $\underline{\Omega}_{31}$ is large (the regressors and the stationary part of the error are driven by the same process) and the persistent movement in the variance of the error is nearly independent of both x_t and u_{1t} ($\underline{\Omega}_{23}$ and $\underline{\Omega}_{21}$ are small). It is hard to think of a standard economic model which would produce this configuration. We can conclude that heteroskedasticity of the BI form plus endogeneity can distort the size of hypothesis tests constructed with conventional critical values, yet this distortion will not be large in most cases.

5. Conclusion

Modeling dynamic economic variables using the framework of cointegration is attractive because it allows researchers to specify meaningful regression relationships between nonstationary series, permitting the broad range of analysis and hypothesis testing which permeates applied economics. This paper has shown that much of the statistical theory developed for the standard model of cointegration carries over to a broader class of models which can be characterized by heteroskedastic regression errors whose variances are potentially unbounded.

In the standard model of cointegration, the regressors differ stochastically from the regression errors in two respects: the regressors' variance grows linearly over time, and the regressors possess a stochastic trend. The bi-integrated errors of this paper are similar to conventional errors in that they do not have a stochastic trend, yet their unconditional second moments grow linearly as do the regressors. The fact that coefficient estimates are consistently estimated even under arbitrary endogeneity assumptions suggests that the asymptotic results of the standard cointegration model are driven by the differing trend properties of the regressors and the regression errors, not the differing variance properties.

There are several questions left unanswered by this paper. First, although it may seem evident that empirical second moments exhibit nonstationary characteristics, it is not at all clear whether the bi-integrated processes introduced here are the most useful approximation to the data. Second, in the presence of heteroskedasticity, least squares estimation, although consistent, is inefficient. Some alternative estimator may be more efficient. Harvey and Robinson (1988) suggest in a different context an adaptive estimation procedure. Essentially, the authors suggest estimating the underlying variance process by smoothing (via a kernel) the squared residuals from the OLS regression. In the context of heteroskedastic cointegration, what would be necessary is to show that the variance estimate obtained in this manner converges weakly to the variance process $B_2(r)^2$. Proving such a theorem appears to be quite challenging and will be left to future research.

Appendix

Proof of Theorem 1

Define $x_{Tt} = \delta_T^{-1}(1\ x_t')' = (1\ T^{-1/2}x_t')'$. Eq. (1) may be rewritten as

$$y_t = \beta^{*\prime}\begin{pmatrix} 1 \\ x_t \end{pmatrix} + (w_t - \Lambda_{21}).$$

Thus,

$$\delta_T(\hat{\beta}-\beta^*) = \left(T^{-1}\sum_{t=1}^{T} x_{Tt}x_{Tt}\right)^{-1}\left(T^{-1}\sum_{t=1}^{T} x_{Tt}(w_t-\Lambda_{21})\right)$$

$$= \left(T^{-1}\sum_{t=1}^{T} x_{Tt}x_{Tt}\right)^{-1}\begin{pmatrix} T^{-1}\sum_1^T (w_t-\Lambda_{21}) \\ T^{-3/2}\sum_1^T x_t(w_t-\Lambda_{21}) \end{pmatrix}. \quad \text{(A.1)}$$

From (5) and the assumption that $\mathrm{E}|x_0|<\infty$,

$$x_{T[Tr]} = \begin{pmatrix} 1 \\ T^{-1/2}x_0 + T^{-1/2}S_{3[Tr]} \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ B_3(r) \end{pmatrix} = X(r). \quad \text{(A.2)}$$

By the continuous mapping theorem and (A.2),

$$T^{-1}\sum_{t=1}^{T} x_{Tt}x_{Tt} \Rightarrow \int_0^1 XX'. \quad \text{(A.3)}$$

From (6) we can deduce

$$T^{-1}\sum_1^T (w_t-\Lambda_{21}) = \sigma_0 T^{-1}\sum_1^T u_{1t} + T^{-1}\sum_1^T S_{2t}u_{1t} - \Lambda_{21} \Rightarrow \int_0^1 B_2\,\mathrm{d}B_1. \quad \text{(A.4)}$$

From (7) we find

$$T^{-3/2}\sum_1^T x_t(w_t-\Lambda_{21})$$

$$= T^{-3/2}\sum_1^T (x_0+S_{3t})(\sigma_0+S_{2t})u_{1t} - T^{-3/2}\sum_1^T x_{3t}\Lambda_{21}$$

$$= T^{-3/2}\sum_1^T S_{3t}S_{2t}u_{1t} - T^{-3/2}\sum_1^T x_{3t}\Lambda_{21} + \mathrm{o}_p(1)$$

$$\Rightarrow \int_0^1 B_3B_2\,\mathrm{d}B_1 + \int_0^1 B_2\Lambda_{31} + \int_0^1 B_3\Lambda_{21} - \int_0^1 B_3\Lambda_{21}$$

$$= \int_0^1 B_3B_2\,\mathrm{d}B_1 + \int_0^1 B_2\Lambda_{31}, \quad \text{(A.5)}$$

respectively. (A.3), (A.4), and (A.5) combine to yield the result. □

Proof of Theorem 3

The proof will be facilitated by a few inequalities. Define

$$Y_{Ttm} = T^{-2}(x_{t+m} - \bar{x})(x_t - \bar{x})'\sigma_{t+m}\sigma_t \tag{A.6}$$

and

$$\eta_{tm} = u_t u'_{t+m} - \mathrm{E}(u_t u'_{t+m}). \tag{A.7}$$

The inequalities we need are

$$\left\| \sum_{j=i}^{i+t} u_j \right\|_6 \leq K_1 \sqrt{t}, \tag{A.8}$$

$$\| Y_{Tt+k,m} - Y_{Ttm} \|_{3/2} \leq K_2 (k/T)^{1/2}, \tag{A.9}$$

$$\left\| \sum_{j=i}^{i+t} \eta_{jm} \right\|_3 \leq K_3 \sqrt{t}, \tag{A.10}$$

for finite K_1, K_2, and K_3.

Define $\mathscr{F}_t = \sigma(u_i\colon i \leq t)$, the smallest σ-field containing the past history of u_i. (A.8) and (A.10) will follow from Lemma 2 of Hansen (1991a) if $\{u_t, \mathscr{F}_t\}$ and $\{\eta_{tm}, \mathscr{F}_t\}$ are an L_6-mixingale and an L_3-mixingale, respectively, with summable mixingale coefficients (independent of m for η_{tm}). Indeed, by McLeish's strong mixing inequality,

$$\| \mathrm{E}(u_t | \mathscr{F}_{t-k}) \|_6 \leq 6 \alpha_k^{1/6 - 1/p} \| u_t \|_p,$$

and under Assumption 4, $\sum_1^\infty \alpha_k^{1/6-1/p} < \infty$, so (A.8) holds. (A.9) follows from (A.8) and a few tedious algebraic manipulations. To establish (A.10) we note that, by Lemma 1 in Hansen (1991b),

$$\| \mathrm{E}(\eta_{tm} | \mathscr{F}_{t-k}) \|_3 \leq 12 \alpha_k^{1/3 - 2/p} \| u_t \|_p \| u_{t+m} \|_p,$$

and $\sum_1^\infty \alpha_k^{1/3-2/p} < \infty$ under Assumption 4, so $\{\eta_{tm}, \mathscr{F}_t\}$ is an L_3-mixingale with summable mixingale coefficients independent of m.

(a) First, we establish

$$T^{-3} \sum_{m=-B}^{B} k_m \sum_t (x_{t+m} - \bar{x})(x_t - \bar{x})' \sigma_t \sigma_{t+m} u_{1t+m} u_{1t}$$

$$\Rightarrow \int_0^1 B_3^* B_3^{*\prime} B_2^2 \Omega_{11}, \tag{A.11}$$

which follows from

$$T^{-1}\sum_{m=-B}^{B}k_m\sum_t Y_{Ttm}\eta_{tm}\underset{\mathrm{p}}{\to}0. \tag{A.12}$$

Set $N=[\sqrt{T}]$, $t_k=[kT/N]+1$, and $t_k^*=t_{k+1}-1$. Then

$$\begin{aligned}
&\mathrm{E}\left|\sum_{t=1}^{T}Y_{Ttm}\eta_{tm}\right| \\
&=\mathrm{E}\left|\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*}Y_{Ttm}\eta_{tm}\right| \\
&\le \mathrm{E}\left|\sum_{k=0}^{N-1}Y_{Tt_km}\sum_{t=t_k}^{t_k^*}\eta_{tm}\right|+\mathrm{E}\left|\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*}\left(Y_{Ttm}-Y_{Tt_km}\right)\eta_{tm}\right| \\
&\le \sum_{k=0}^{N-1}\mathrm{E}\left|Y_{Tt_km}\sum_{t=t_k}^{t_k^*}\eta_{tm}\right|+\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*}\mathrm{E}\left|\left(Y_{Ttm}-Y_{Tt_km}\right)\eta_{tm}\right| \\
&\le \sum_{k=0}^{N-1}\|Y_{Ttm}\|_{3/2}\left\|\sum_{t=t_k}^{t_k^*}\eta_{tm}\right\|_3+\sum_{k=0}^{N-1}\sum_{t=t_k}^{t_k^*}\|Y_{Ttm}-Y_{Tt_km}\|_{3/2}\|\eta_{tm}\|_3 \\
&\le NK_2K_3(T/N)^{1/2}+N\sum_{t=1}^{T/N}K_2(t/T)^{1/2}K_3 \\
&\le K_2K_3\left((TN)^{1/2}+T/\sqrt{N}\right)=2K_2K_3T^{3/4},
\end{aligned}$$

by repeated use of the triangle inequality, Holder's inequality, (A.9), and (A.10). Therefore,

$$\begin{aligned}
&\mathrm{E}\left|T^{-1}\sum_{m=-B}^{B}k_m\sum_t Y_{Ttm}\eta_{tm}\right| \\
&\le T^{-1}(2B+1)\max_{|m|\le B}\mathrm{E}\left|\sum_t Y_{Ttm}\eta_{tm}\right| \\
&\le 2K_2K_3(2B+1)T^{-1/4}\to 0 \quad \text{as} \quad T\to\infty.
\end{aligned}$$

This establishes (A.12) by Markov's inequality, and hence (A.11).

We next show that

$$T^{-3}\sum_{m=-B}^{B}k_m\sum_t(x_{t+m}-\bar{x})(x_t-\bar{x})'\sigma_t\sigma_{t+m}u_{1t+m}u_{1t}$$

$$=T^{-3}\sum_{m=-B}^{B}k_m\sum_t(x_{t+m}-\bar{x})(x_t-\bar{x})'\hat{w}_{t+m}w_t+\mathrm{o}_p(1). \tag{A.13}$$

To simplify the presentation, assume x_t is scalar. Now,

$$\left|T^{-3}\sum_{m=-B}^{B}k_m\sum_t(x_{t+m}-\bar{x})(x_t-\bar{x})(w_{t+m}-\hat{w}_{t+m})w_t\right|$$

$$\leq T^{-3}\sum_{m=-B}^{B}\left|\sum_t(x_{t+m}-\bar{x})(x_t-\bar{x})x'_{Tt+m}\delta_T(\hat{\beta}-\beta)\sigma_t u_{1t}\right|$$

$$\leq\frac{2B+1}{\sqrt{T}}\sup_{|m|\leq B}\left(T^{-4}\sum_t(x_{t+m}-\bar{x})^2(x_t-\bar{x})^2\right.$$

$$\left.\times\left(x'_{Tt+m}\delta_T(\hat{\beta}-\beta)\right)^2\sigma_t^2\right)^{1/2}\left(T^{-1}\sum_t u_{1t}^2\right)^{1/2}$$

$$=\frac{2B+1}{\sqrt{T}}\mathrm{O}_p(1)\underset{\mathrm{p}}{\rightarrow}0.$$

This establishes (A.13). Similarly, we can replace w_t by $\hat{w}_t$ to find that

$$T^{-3}\sum_{m=-B}^{B}k_m\sum_t(x_{t+m}-\bar{x})(x_t-\bar{x})'\sigma_t\sigma_{t+m}u_{1t+m}u_{1t}$$

$$=T^{-2}\hat{V}_1+\mathrm{o}_p(1). \tag{A.14}$$

(A.14) and (A.11) complete the proof of part (a).

(b) Note that $T^{-1}M_1 \Rightarrow \int_0^1 B_3^* B_3^{*\prime}$. Thus

$$\hat{W} = \sqrt{T}\left(R'\hat{\beta}_1 - r\right)'\left(R'\left(\frac{1}{T}\hat{M}_1\right)^{-1}\left(T^{-2}\hat{V}_1\right)\left(\frac{1}{T}\hat{M}_1\right)^{-1}R\right)^{-1}$$

$$\times\sqrt{T}\left(R'\hat{\beta}_1 - r\right)$$

$$\Rightarrow N\left(0, R'M_1^{-1}VM_1^{-1}R\right)'\left(R'M_1^{-1}VM_1^{-1}R\right)^{-1}N\left(0, R'M_1^{-1}VM_1^{-1}R\right).$$

Conditional on $M_1^{-1}V_1^{-1}$, this distribution is χ_q^2. Since this does not depend upon $M_1^{-1}V_1^{-1}$, it is the unconditional distribution as well. This completes the proof. □

References

Anderson, T.W. and J.B. Taylor, 1979, Strong consistency of least squares estimates in dynamic models, Annals of Statistics 7, 484–489.

Andrews, D.W.K., 1987, Least squares regression with integrated or dynamic regressors under weak error assumptions, Econometric Theory 3, 98–116.

Blinder, A.S. and A.S. Deaton, 1985, The time series consumption function revisited, Brookings Papers on Economic Activity, 465–511.

Campbell, J.Y., 1987, Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis, Econometrica 55, 1249–1273.

Campbell, J.Y. and R.J. Shiller, 1987, Cointegration and tests of present value models, Journal of Political Economy 95, 1062–1088.

Engle, R.F. and C.W.J. Granger, 1987, Co-integration 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.

Hansen, B.E., 1991a, Strong laws for dependent heterogeneous processes, Econometric Theory 7, 213–221; and Errata, Econometric Theory 8, forthcoming.

Hansen, B.E., 1991b, Consistent covariance matrix estimation for dependent heterogeneous processes, Econometrica, forthcoming.

Hansen, B.E., 1991c, Convergence to stochastic integrals for dependent heterogeneous processes, Econometric Theory, forthcoming.

Harvey, A.C. and P.M. Robinson, 1988, Efficient estimation of nonstationary time series regression, Journal of Time Series Analysis, 201–214.

Herrndorf, N., 1984, A functional central limit theorem for weakly dependent sequences of random variables, Annals of Probability 12, 141–153.

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

McLeish, D.L., 1975, A maximal inequality and dependent strong laws, Annals of Probability 5, 829–839.

Newey, W. and K.D. West, 1987, A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703–708.

Park, J.Y. and P.C.B. Phillips, 1988, Statistical inference in regressions with integrated processes: Part I, Econometric Theory 4, 468–497.

Phillips, P.C.B. and S.N. Durlauf, 1986, Multiple time series with integrated variables, Review of Economic Studies 3, 473–496.

Protter, P., 1990, Stochastic integration and differential equations: A new approach (Springer-Verlag, New York, NY).

Robinson, P., 1978, On consistency in time series analysis, Annals of Statistics 6, 214–223.
Stock, J.H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, Econometrica 55, 1035–1056.
White, H. and I. Domowitz, 1984, Nonlinear regression with dependent observations, Econometrica 52, 143–162.
Wooldridge, J.M. and H. White, 1988, Some invariance principles and central limit theorems for dependent heterogeneous processes, Econometric Theory 4, 210–230.