Biometrika (2001), **88,** 4, *pp.* 1135–1152

Printed in Great Britain

Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity

By W. K. LI

Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong

hrntlwk@hku.hk

SHIQING LING

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

maling@ust.hk

AND H. WONG

Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong

mathwong@polyu.edu.hk

Summary

This paper investigates a partially nonstationary multivariate autoregressive model, which allows its innovations to be generated by a multivariate ARCH, autoregressive conditional heteroscedastic, process. Three estimators, including the least squares estimator, a full-rank maximum likelihood estimator and a reduced-rank maximum likelihood estimator, are considered and their asymptotic distributions are derived. When the multivariate ARCH process reduces to the innovation with a constant covariance matrix, these asymptotic distributions are the same as those given by Ahn & Reinsel (1990). However, in the presence of multivariate ARCH innovations, the asymptotic distributions of the full-rank maximum likelihood estimator and the reduced-rank maximum likelihood estimator involve two correlated multivariate Brownian motions, which are different from those given by Ahn & Reinsel (1990). Simulation results show that the full-rank and reduced-rank maximum likelihood estimator are more efficient than the least squares estimator. An empirical example shows that the two features of multivariate conditional heteroscedasticity and partial nonstationarity may be present simultaneously in a multivariate time series.

Some key words: Brownian motion; Cointegration; Full-rank and reduced-rank maximum likelihood estimators; Least squares estimator; Multivariate ARCH process; Partially nonstationary; Unit root.

1. Introduction

We consider an m-dimensional autoregressive, AR, process $\{Y_t\}$ which is generated by

$$Y_t = \sum_{i=1}^{p} \Phi_i Y_{t-i} + \varepsilon_t \tag{1·1a}$$

with

$$\varepsilon_t = \sum_{i=1}^{q} \alpha_{it}\varepsilon_{t-i} + e_t, \tag{1·1b}$$

where Φ_i's are constant matrices; $\det\{\Phi(z)\} = |I - \Phi_1 z - \ldots - \Phi_p z^p| = 0$ has $d < m$ unit roots and the remaining roots are outside the unit circle; $\text{rank}(C) = r$ with $r = m - d > 0$, where $C = -\Phi(1)$; $\delta_t = (\alpha_{1t}, \ldots, \alpha_{qt})$ is a sequence of independent and identically distributed $m \times qm$ matrices with mean zero and nonnegative-definite covariance matrix $E\{\text{vec}(\delta_t)\,\text{vec}'(\delta_t)\} = \Omega$; and the e_t are independent identically distributed random vectors with mean zero and positive-definite covariance matrix $E(e_t e_t') = G$. Under certain regularity conditions, the conditional covariance matrix of ε_t can be shown to be changing over time; see (2·7) and (2·8) in § 2. Thus, model (1·1) is a partially nonstationary multivariate AR model with autoregressive conditional heteroscedastic (ARCH) type errors. In view of mathematical complications, we concentrate on the case of diagonal ARCH apart from the results in § 2. In this case, the conditional variance of each component in ε_t has the same form as Engle's (1982) ARCH model; see (2·7)–(2·8) and (4·1) below.

The assumption on $\Phi(z)$ in model (1·1) implies the important feature that, although some component series of Y_t may exhibit nonstationary behaviour, some linear combinations of these component series would be stationary. This phenomenon is the so-called cointegration property with cointegration rank r. When ε_t has a constant conditional covariance matrix, cointegrated time series were proposed first by Engle & Granger (1987) and have been widely investigated. Some estimation methods and asymptotic theories can be found for example in Johansen (1988, 1992, 1995) and Stock & Watson (1993). Rahbek & Mosconi (1999) considered the inference problem with stationary explanatory variables. There is substantial empirical evidence that cointegrated components can exist in various multivariate economic time series. Fountis & Dickey (1988) investigated the AR model with a unit root, that is $d = 1$, and derived a corresponding unit-root test. Ahn & Reinsel (1990) examined the AR model with d unit roots and derived the asymptotic properties of the full-rank least squares estimator and the reduced-rank maximum likelihood estimator. Yap & Reinsel (1995) studied partially nonstationary multivariate ARMA, autoregressive moving average, models.

The class of ARCH models was proposed first by Engle (1982). Various ARCH-type models have been proposed, including generalised ARCH or GARCH (Bollerslev, 1986), and double-threshold ARCH (Li & Li, 1996) models. Many ARCH-type models have been extended to the multivariate case; see the survey by Bollerslev et al. (1994). Tsay (1987) proposed the conditional heteroscedastic ARMA model. It provides a natural alternative for Engle's ARCH model and was extended to multivariate cases by Ling & Deng (1993) and Wong & Li (1997).

Since the inclusion of time-varying conditional variance matrices can improve statistical inference, such as interval estimation and forecasting, it is important to explore cointegration time series with conditional heteroscedasticity. Although our method is a direct extension of Ahn & Reinsel (1990), the estimation procedure for model (1·1) is much more complicated and the asymptotic distributions for the full-rank and reduced-rank maximum likelihood estimators are new. Unlike as in Ahn & Reinsel (1990), these asymptotic distributions involve two correlated multivariate Brownian motions. This type of asymptotic distribution has appeared in Lucas (1997), Hodgson (1998), Seo (1998) and Rahbek & Mosconi (1999). Our technique heavily depends on Theorem 2.1 in Ling & Li (1998) and could be applied to cointegrating time series with other ARCH-type innovations.

In § 2, we discuss some basic properties of model (1·1). In § 3, we derive the asymptotic distribution of the least squares estimator. In §§ 4 and 5, we investigate the full-rank and reduced-rank maximum likelihood estimators, and derive their asymptotic distributions. The finite sample properties of the estimators are examined through simulations in § 6. Finally, in § 7 an empirical example is presented.

2. Basic properties of the models

First we reparameterise the AR part of the model (1·1) as follows:

$$W_t = CY_{t-1} + \Phi_1^* W_{t-1} \ldots + \Phi_{p-1}^* W_{t-p+1} + \varepsilon_t, \tag{2·1}$$

where $W_t = Y_t - Y_{t-1}$ and $\Phi_i^* = -\sum_{k=i+1}^{p} \Phi_k$. Following Ahn & Reinsel (1990), let $m \times m$ matrices P and $Q = P^{-1}$ be such that $Q(\sum_{i=1}^{p} \Phi_i)P = \mathrm{diag}(I_d, \Gamma_r)$, the Jordan canonical form of $\sum_{i=1}^{p} \Phi_i$. Defining $Z_t = QY_t$, we obtain

$$Z_t = \mathrm{diag}(I_d, \Gamma_r) Z_{t-1} + u_t, \tag{2·2}$$

where $u_t = Q(\Phi_1^* W_{t-1} \ldots + \Phi_{p-1}^* W_{t-p+1} + \varepsilon_t)$. Furthermore, if we let

$$g(z) = (1-z)^{-d} \det\{\Phi(z)\}, \quad H(z) = (1-z)^{-d+1} \mathrm{adj}\{\Phi(z)\},$$

we can rewrite u_t as

$$u_t = \left\{ I_m + Q \sum_{j=1}^{p-1} \Phi_j^* g(B)^{-1} H(B) P B^j \right\} a_t = \Psi(B) a_t, \tag{2·3}$$

where $a_t = Q\varepsilon_t$ and

$$\Psi(B) = I_m + Q \sum_{j=1}^{p-1} \Phi_j^* g(B)^{-1} H(B) P B^j = \sum_{k=0}^{\infty} \Psi_k B^k, \tag{2·4}$$

in which $\Psi_0 = I_m$, $\Psi_k = O(\rho^k)$ and $\rho \in (0, 1)$, as in Ahn & Reinsel (1990).

Partition $Q' = [Q_1, Q_2]$ and $P = [P_1, P_2]$ such that Q_1 and P_1 are $m \times d$ matrices, and Q_2 and P_2 are $m \times r$ matrices. Furthermore, partition $u_t = [u_{1t}', u_{2t}']'$ such that u_{1t} is $d \times 1$ and u_{2t} is $r \times 1$. Define $Z_{1t} = Q_1' Y_t$ and $Z_{2t} = Q_2' Y_t$, so that

$$Z_{1t} = Z_{1t-1} + u_{1t}, \quad Z_{2t} = \Gamma_r Z_{2t-1} + u_{2t}. \tag{2·5}$$

Here $\{Z_{1t}\}$ is a nonstationary $d \times 1$ time series with d unit roots. However, under Assumptions 1 and 2 below, $\{Z_{2t}\}$ is a stationary $r \times 1$ time series. The matrix Q_2' is the so-called cointegrated vector with rank r, as in Engle & Granger (1987). The error-correction form of model (2·1) can be found in Ahn & Reinsel (1990).

We now make the following assumptions about the innovations in model (1·1).

Assumption 1. The $\{e_t\}$ and $(\delta_t\}$ are mutually independent.

Assumption 2. All eigenvalues of $E(B_t \otimes B_t)$ are inside the unit circle, where $\otimes$ denotes the Kronecker product and

$$B_t = \begin{pmatrix} \alpha_{1t} & \cdots & \alpha_{q-1,t} & \alpha_{qt} \\ I_m & \cdots & 0 & 0 \\ \vdots & & & \\ 0 & \cdots & I_m & 0 \end{pmatrix},$$

in which I_m is the $m \times m$ identity matrix.

Under Assumptions 1 and 2, Ling & Deng (1993) showed that ε_t is strictly stationary and ergodic, and has the expansion

$$\tilde{\varepsilon}_t = \eta_t + \sum_{k=1}^{\infty} \left(\prod_{i=0}^{k-1} B_{t-i} \right) \eta_{t-k}, \tag{2·6}$$

where $\tilde{\varepsilon}_t = (\varepsilon_t', \ldots, \varepsilon_{t-q+1}')'$ and $\eta_t = (e_t', 0, \ldots, 0)'$. Denote

$$\sigma\{e_s, \alpha_{1s}, \ldots, \alpha_{qs} : s = t, t-1, \ldots\}$$

by $\mathscr{F}_t$. By (2·6), ε_t is $\mathscr{F}_t$-measurable and

$$V_t = E(\varepsilon_t \varepsilon_t' | \mathscr{F}_{t-1}) = (\tilde{\varepsilon}_{t-1}' \otimes I_m) \Omega (\tilde{\varepsilon}_{t-1} \otimes I_m) + G. \tag{2·7}$$

Thus, the conditional covariance matrix $E(\varepsilon_t \varepsilon_t' | \mathscr{F}_{t-1})$ depends on the past information $\mathscr{F}_{t-1}$ and hence $\{\varepsilon_t\}$ generated by (1·2) is a multivariate ARCH-type process. In particular, when the α_{it} are independent for $i = 1, \ldots, q$,

$$V_t = \sum_{i=1}^{q} (\tilde{\varepsilon}_{t-i}' \otimes I_m) E\{\mathrm{vec}(\alpha_{it}) \, \mathrm{vec}'(\alpha_{it})\} (\tilde{\varepsilon}_{t-i} \otimes I_m) + G. \tag{2·8}$$

Furthermore, if $m = 1$ then $V_t = \sum_{i=1}^{q} \sigma_i \varepsilon_{t-i}^2 + G$ with $\sigma_i = E\alpha_{it}^2$, which has the same form as Engle's (1982) ARCH model.

By a direct extension of Theorem 3.1 in Ling (1999), the $2m$-order moment condition of ε_t is $\rho\{E(B_t^{\otimes 2m})\} < 1$, where $B_t^{\otimes 2m}$ denotes $B_t \otimes \ldots \otimes B_t$, with $2m$ factors, and $\rho(A)$ is the largest of the absolute magnitudes of the eigenvalues of A. Under this condition, the $2m$-order moments of Z_{2t} and u_{1t} are finite since u_t has expansion (2·3). These results will be useful for the development of the asymptotic theory in this paper.

3. Preliminary estimation procedure

From this section onward we assume that e_t and $\mathrm{vec}(\delta_t)$ in model (1·1) are normal random vectors and that the conditional covariance matrix V_t of ε_t is diagonal.

We first use least squares estimation for the parameters in the AR part of model (1·1). Let

$$X_{t-1} = [Y_{t-1}', W_{t-1}', \ldots, W_{t-p+1}']', \quad F = [C, \Phi_1^*, \ldots, \Phi_{p-1}^*].$$

From (2·1), the least squares estimator of F is $\hat{F} = (\sum_{t=1}^{n} W_t X_{t-1}')(\sum_{t=1}^{n} X_{t-1} X_{t-1}')^{-1}$.

Write

$$Q^* = \mathrm{diag}(Q, I_{m(p-1)}), \quad P^* = \mathrm{diag}(P, I_{m(p-1)}), \quad X_t^* = Q^* X_t = [Z_{1t}', U_t']',$$

with $U_{t-1} = [Z_{2t-1}', W_{t-1}', \ldots, W_{t-p+1}']'$. Then

$$Q(\hat{F} - F)P^* = \left(\sum_{t=1}^{n} a_t X_{t-1}^{*\prime} \right) \left(\sum_{t=1}^{n} X_{t-1}^* X_{t-1}^{*\prime} \right)^{-1}. \tag{3·1}$$

Furthermore, denote $\mathrm{diag}(D, \sqrt{n} I_{m(p-1)})$ by D^*, where $D = \mathrm{diag}(nI_d, \sqrt{n} I_r)$. Using Lemmas A1 and A2 in the Appendix and a similar argument to that of Ahn & Reinsel (1990), we can obtain the following theorem.

Theorem 1. *Suppose that* (1·1) *has d unit roots and the remaining roots are outside the unit circle, that Assumptions* 1 *and* 2 *hold and that* $\rho\{E(B_t^{\otimes 4})\} < 1$. *Then*

$$(\hat{F} - F)P^* D^* \to P[M, N], \tag{3·2}$$

in distribution, where

$$M=\Omega_a^{1/2}\left\{\int_0^1 B_d(u)\,dB_m(u)'\right\}'\left\{\int_0^1 B_d(u)B_d(u)'\,du\right\}^{-1}\Omega_{a_1}^{-1/2}\Psi_{11}^{-1},$$

$\Omega_a=\text{cov}(a_t)=QV_0Q'$, $V_0=E(\varepsilon_t\varepsilon_t')$ *and* $\Omega_{a_1}=\text{cov}(a_{1t})=[I_d,0]\Omega_a[I_d,0]'$; $B_m(u)$ *denotes an m-dimensional standard Brownian motion,* $B_d(u)=\Omega_{a_1}^{-1/2}[I_d,0]\Omega_a^{1/2}B_m(u)$ *is a d-dimensional standard Brownian motion, and* $\Psi_{11}=[I_d,0](\sum_{k=1}^{\infty}\Psi_k)[I_d,0]'$; *and* $\text{vec}(N)$ *is a normal vector with mean* 0 *and covariance matrix*

$$\{E(U_{t-1}U_{t-1}'\otimes I_m)\}^{-1}E(U_{t-1}U_{t-1}'\otimes V_t)\{E(U_{t-1}U_{t-1}'\otimes I_m)\}^{-1}.$$

The limiting distribution for the nonstationary component corresponding to Z_{1t} is the same as that given by Ahn & Reinsel (1990), but, for the stationary component, the limiting distribution has a different covariance matrix as a result of the conditional heteroscedasticity of the ε_t in (1·1). When $r=0$, all component processes of Y_t in model (1·1) are $I(1)$, that is, each component has a unit root, the limiting distribution M in (3·2) will reduce to that of the multivariate unit-root case.

Using the residual $\hat{\varepsilon}_t=W_t-X_t\hat{F}$ as the artificial observation of ε_t, we can estimate Ω and G in (2·7) by least squares or maximum likelihood. Since the convergent rate of $\hat{F}$ is D^*, it is immediate to show that these estimators are asymptotically equivalent in probability to those based on the true observation of ε_t; see Ling & Deng (1993) and a University of Hong Kong technical report on unit root testing under GARCH errors by S. Ling, W. K. Li and M. McAleer, available on request. From Theorem 4.1 in Nicholls & Quinn (1982), we know that the least squares estimator of (Ω, G) is asymptotically normal if $E\varepsilon_{t,j}^8<\infty$. Ling & Deng (1993) showed that the maximum likelihood estimator of (Ω, G) is asymptotically normal if $E\varepsilon_{t,j}^4<\infty$.

4. Full-rank maximum likelihood estimation

Let

$$V_t=\text{diag}(h_{1t},\ldots,h_{mt}), \tag{4·1}$$

with $h_{kt}=g_k+\sum_{i=1}^{q}\sigma_{ki}\varepsilon_{kt-i}^2$, where $g_k=\text{var}(e_{kt})$, $\sigma_{ki}=\text{var}(\alpha_{kkit})$, e_{kt} is the kth element of e_t and α_{kkit} is the (k,k)th element of α_{it}. Let $\alpha=(\alpha_1',\ldots,\alpha_m')'$, with $\alpha_k=(g_k,\sigma_{k1},\ldots,\sigma_{kq})'$ for $k=1,\ldots,m$.

The maximum likelihood estimators of F and α are $\hat{F}$ and $\hat{\alpha}$, which minimise the conditional loglikelihood function

$$l=\sum_{t=1}^{n} l_t, \tag{4·2a}$$

with

$$l_t=-\frac{1}{2}\sum_{k=1}^{m}\log h_{kt}-\frac{1}{2}\sum_{k=1}^{m}\frac{\varepsilon_{kt}^2}{h_{kt}}. \tag{4·2b}$$

By direct differentiation, we obtain

$$\frac{\partial l_t}{\partial F}=\sum_{i=1}^{q}(X_{t-i-1}\otimes I_m)\eta_{it}+(X_{t-1}\otimes I_m)V_t^{-1}\varepsilon_t, \tag{4·3}$$

where

$$\eta_{it} = \left[\frac{\sigma_{1i}\varepsilon_{1t-i}}{h_{1t}} \left(1 - \frac{\varepsilon_{1t}^2}{h_{1t}} \right), \ldots, \frac{\sigma_{mi}\varepsilon_{mt-i}}{h_{mt}} \left(1 - \frac{\varepsilon_{mt}^2}{h_{mt}} \right) \right]'.$$

Let $\bar{D}^* = D \otimes I_m = \mathrm{diag}(nI_{dm}, \sqrt{n}I_{rm+(p-1)m^2})$, $\bar{Q}^* = \mathrm{diag}(Q \otimes I_m, I_{(p-1)m^2})$ and

$$V_{it} = 2\ \mathrm{diag}\left(\frac{\sigma_{1i}^2\varepsilon_{1t-i}^2\varepsilon_{1t}^2}{h_{1t}^3}, \ldots, \frac{\sigma_{mi}^2\varepsilon_{mt-i}^2\varepsilon_{mt}^2}{h_{mt}^3} \right).$$

As in Ling & Li (1998), we can show that

$$\bar{D}^{*-1}\bar{Q}^* \frac{\partial^2 l}{\partial F\, \partial F'} \bar{Q}^{*\prime}\bar{D}^{*-1} = -\sum_{t=1}^{n} \bar{D}^{*-1}\bar{Q}^* M_t \bar{Q}'^{*}\bar{D}^{*-1} + o_p(1),$$

where

$$M_t = \sum_{i=1}^{q} (X_{t-i-1}X'_{t-i-1} \otimes V_{it}) + X_{t-1}X'_{t-1} \otimes V_t^{-1}.$$

Since $n^{-1/2}\bar{D}^{*-1}\bar{Q}^*(\partial^2 l/\partial F\, \partial\alpha_k') = o_p(1)$ (Ling & Li, 1998), both F and α_k can be estimated separately without loss of efficiency. The maximum likelihood estimator of α through (4·2) is asymptotically equivalent to the maximum likelihood estimator mentioned in § 3. Thus, in the following α is assumed to be known or estimated while we only discuss the estimator of F.

The maximum likelihood estimator of F is obtained by the iterative approximate Newton–Raphson relation

$$\hat{F}^{(i+1)} = \hat{F}^{(i)} + \left(\sum_{t=1}^{n} M_t \right)^{-1}_{F=\hat{F}^{(i)}} \left(\sum_{t=1}^{n} \frac{\partial l_t}{\partial F} \right)_{F=\hat{F}^{(i)}}, \qquad (4{\cdot}4)$$

where $\hat{F}^{(i)}$ is the estimator at the ith iteration. It is straightforward to show that

$$\sum_{t=1}^{n} \bar{D}^{*-1}\bar{Q}^*(M_t|_{F=\tilde{F}} - M_t)\bar{Q}'^{*}\bar{D}^{*-1} = o_p(1), \qquad (4{\cdot}5)$$

$$\sum_{t=1}^{n} \bar{D}^{*-1}\bar{Q}^* \left(\left.\frac{\partial l_t}{\partial F}\right|_{F=\tilde{F}} - \frac{\partial l_t}{\partial F} \right) = -\left(\sum_{t=1}^{n} \bar{D}^{*-1}\bar{Q}^* M_t \right) \mathrm{vec}(\tilde{F} - F) + o_p(1), \qquad (4{\cdot}6)$$

uniformly in the ball $\Theta_n = \{\tilde{F} : \|\bar{D}^*\bar{Q}^{*\prime -1}\, \mathrm{vec}(\tilde{F} - F)\| \leqslant M\}$ for any fixed positive constant M, where $\|.\|$ denotes the Euclidean norm. Similar details and the proof of consistency can be found in the aforementioned University of Hong Kong technical report by Ling et al. Thus, the estimator of F obtained by (4·4) satisfies $\bar{D}^*\bar{Q}^{*\prime -1}\, \mathrm{vec}(\hat{F} - F) = O_p(1)$ if the initial estimator also satisfies this condition. The least squares estimator, denoted by $\bar{F}$, of F in § 3 can be used as such an initial estimator. With this initial estimator and (4·4)–(4·6), we can obtain the asymptotic representation

$$\bar{D}^*\bar{Q}^{*\prime -1}\, \mathrm{vec}(\hat{F} - F) = \left(\sum_{t=1}^{n} \bar{D}^{*-1}\bar{Q}^* M_t \bar{Q}^{*\prime}\bar{D}^{*-1} \right)^{-1} \left(\sum_{t=1}^{n} \bar{D}^{*-1}\bar{Q}^* \frac{\partial l_t}{\partial F} \right) + o_p(1). \qquad (4{\cdot}7)$$

Note that we can partition $\bar{Q}^*(X_{t-i} \otimes I_m)$ into two parts:

$$\bar{Q}^*(X_{t-i} \otimes I_m) = \begin{pmatrix} Z_{1t-i} \otimes I_m \\ U_{t-i} \otimes I_m \end{pmatrix}.$$

We have

$$\bar{D}^{*-1}\bar{Q}^*\frac{\partial l}{\partial F}=\sum_{t=1}^{n}\begin{pmatrix}N_{1t}\\ N_{2t}\end{pmatrix},$$

$$N_{1t}=\sum_{i=1}^{q}(Z_{1t-i-1}\otimes I_m)\eta_{it}+(Z_{1t-1}\otimes I_m)V_t^{-1}\varepsilon_t,$$

$$N_{2t}=\sum_{i=1}^{q}(U_{t-i-1}\otimes I_m)\eta_{it}+(U_{t-1}\otimes I_m)V_t^{-1}\varepsilon_t.$$

As $n^{-3/2}\sum_{t=1}^{n}U_{t-i}Z_{1t-j-1}=o_p(1)$, for $i, j=1,\ldots,q$ (Ling & Li, 1998), the cross-product terms in $\sum_{t=1}^{n}\bar{D}^{*-1}\bar{Q}^*M_t\bar{Q}^*\bar{D}^{*-1}$ involving U_{t-i} and Z_{1t-j} converge to zero in probability. Thus, we have

$$\begin{aligned}\sum_{t=1}^{n}&\bar{D}^{*-1}\bar{Q}^*M_t\bar{Q}^{*\prime}\bar{D}^{*-1}\\ &=\operatorname{diag}\left[n^{-2}\sum_{t=1}^{n}\left\{\sum_{i=1}^{q}(Z_{1t-i-1}Z'_{1t-i-1}\otimes V_{it})+Z_{1t-1}Z'_{1t-1}\otimes V_t^{-1}\right\},\right.\\ &\qquad\left. n^{-1}\sum_{t=1}^{n}\left\{\sum_{i=1}^{q}(U_{t-i-1}U'_{t-i-1}\otimes V_{it})+U_{t-1}U'_{t-1}\otimes V_t^{-1}\right\}\right]+o_p(1).\end{aligned}$$

The following is a basic and important lemma for Theorems 2 and 3. Its proof can be found in the Appendix.

LEMMA 1. *Assume the same conditions as in Theorem* 1. *Then*

(a) $$n^{-2}\sum_{t=1}^{n}\left\{\sum_{i=1}^{q}(Z_{1t-i-1}Z'_{1t-i-1}\otimes V_{it})+Z_{1t-1}Z'_{1t-1}\otimes V_t^{-1}\right\}$$
$$\to\Psi_{11}\Omega_{a_1}^{1/2}\left\{\int_0^1 B_d(u)B_d(u)'\,du\right\}\Omega_{a_1}^{1/2}\Psi'_{11}\otimes\Omega_1,$$

in distribution;

(b) $$n^{-1}\sum_{t=1}^{n}N_{1t}\to\operatorname{vec}\left[\left\{\int_0^1 B_d(u)\,d\tilde{W}_m(u)'\right\}'\Omega_{a_1}^{1/2}\Psi'_{11}\right],$$

in distribution;

(c) $$n^{-1}\sum_{t=1}^{n}\left\{\sum_{i=1}^{q}(U_{t-i-1}U'_{t-i-1}\otimes V_{it})+U_{t-1}U'_{t-1}\otimes V_t^{-1}\right\}\to\Omega_u,$$

in probability;

(d) $$n^{-\frac{1}{2}}\sum_{t=1}^{n}N_{2t}\to N(0,\Omega_u),$$

in distribution.

Here $B_d=\Omega_{a_1}^{-1/2}[I_d, 0]\Omega_a^{1/2}B_m(u)$ *and* $B_m(u)=V_0^{-\frac{1}{2}}W_m(u)$ *are standard Brownian motions and*

$(W_m'(u), \tilde{W}_m'(u))'$ *is a* $2m$*-dimensional Brownian motion with covariance matrix given by*

$$u\Omega_b = u\begin{pmatrix} V_0 & I_m \\ I_m & \Omega_1 \end{pmatrix}, \quad V_0 = EV_t,$$

$$\Omega_1 = \mathrm{diag}\left(E\frac{1}{h_{1t}} + 2\sum_{i=1}^{q} \sigma_{1i}^2 E\frac{\varepsilon_{1t-i}^2}{h_{1t}^2}, \ldots, E\frac{1}{h_{mt}} + 2\sum_{i=1}^{q} \sigma_{mi}^2 E\frac{\varepsilon_{mt-i}^2}{h_{mt}^2}\right),$$

$$\Omega_u = 2\sum_{i=1}^{q} E\left\{U_{t-i-1}U_{t-i-1}' \otimes \mathrm{diag}\left(\frac{\sigma_{1i}^2\varepsilon_{1t-i}^2}{h_{1t}^2}, \ldots, \frac{\sigma_{mi}^2\varepsilon_{mt-i}^2}{h_{mt}^2}\right)\right\} + E(U_{t-1}U_{t-1}' \otimes V_t^{-1}).$$

To obtain the asymptotic distribution of F, we rewrite model (2·1) as

$$W_t = CP_1 Z_{1t-1} + CP_2 Z_{2t-1} + \Phi_1^* W_{t-1} \ldots + \Phi_{p-1}^* W_{t-p+1} + \varepsilon_t. \qquad (4\cdot 8)$$

Let $\beta_0 = \mathrm{vec}(CP_1)$ and $\beta_1 = \mathrm{vec}(CP_2, \Phi_1^*, \ldots, \Phi_{p-1}^*)$. Furthermore, let

$$\hat{\beta}_0 = \mathrm{vec}(\hat{C}P_1), \quad \hat{\beta}_1 = \mathrm{vec}(\hat{C}P_2, \hat{\Phi}_1^*, \ldots, \hat{\Phi}_{p-1}^*).$$

Then $\bar{Q}^{*\prime -1}\,\mathrm{vec}(\hat{F} - F) = \{(\hat{\beta}_0 - \beta_0)', (\hat{\beta}_1 - \beta_1)'\}'$. The following theorem comes directly from Lemma 1.

Theorem 2. *Let* $\hat{\beta}_0$ *and* $\hat{\beta}_1$ *denote the full-rank maximum likelihood estimator obtained from* (4·4) *using an initial estimator* $\bar{F}$ *such that* $\bar{D}^*\bar{Q}^{*-1}\,\mathrm{vec}(\bar{F} - F) = O_p(1)$. *Then, under the same assumptions as in Theorem* 1,

(a) $n(\hat{C} - C)P_1 \to \Omega_1^{-1}\left\{\int_0^1 B_d(u)\, d\tilde{W}_m(u)'\right\}' \times \left\{\int_0^1 B_d(u)B_d(u)'\, du\right\}^{-1} \Omega_{a_1}^{-\frac{1}{2}}\Psi_{11}^{-1}$,

in distribution;

(b) $\sqrt{n}(\hat{\beta}_1 - \beta_1) \to N(0, \Omega_u^{-1})$, *in distribution.*

The notation is defined as in Lemma 1 *and* $\hat{C}$ *is the estimator of* C *corresponding to* $\hat{\beta}_0$.

Note that the limiting distribution of $\hat{C}$ is different from that given by the least squares estimator in Theorem 1. Here it involves two correlated m-dimensional Brownian motions. Let

$$\tilde{B}_m(\tau) = -(\Omega_1 - V_0^{-1})^{\frac{1}{2}} V_0^{-1} W_m(\tau) + (\Omega_1 - V_0^{-1})^{\frac{1}{2}} \tilde{W}_m(\tau).$$

Then $B_m(\tau)$ and $\tilde{B}_m(\tau)$ are two independent m-dimensional standard Brownian motions and $\tilde{W}_m(\tau) = V_0^{-\frac{1}{2}} B_m(\tau) + (\Omega_1 - V_0^{-1})^{-\frac{1}{2}} \tilde{B}_m(\tau)$. Furthermore, we obtain

$$n(\hat{C} - C)P_1 \to \Omega_1^{-1}\left[V_0^{-\frac{1}{2}}\left\{\int_0^1 B_d(u)\, dB_m(u)'\right\}' + (\Omega_1 - V_0^{-1})^{\frac{1}{2}}\left\{\int_0^1 B_d(u)\, d\tilde{B}_m(\tau)'\right\}'\right]$$
$$\times \left\{\int_0^1 B_d(u)B_d(u)'\, du\right\}^{-1} \Omega_{a_1}^{-\frac{1}{2}}\Psi_{11}^{-1},$$

in distribution. When V_t is a constant matrix, $\Omega_1 = V_0^{-1}$ and hence the limiting distribution of $\hat{C}$ reduces to that given in Theorem 1 since $V_0^{1/2}B_m(u)$ has the same distribution as $P\Omega_a^{1/2}B_m(u)$ in (3·2). As in the least squares case, when $r = 0$, $P_1 = I_m$ and furthermore, when $m = 1$, the asymptotic distribution is the same as that given by Ling & Li (1998) and Seo (1999). Ling & Li (1998) have shown that the maximum likelihood estimator of the unit root is more efficient than the least squares estimator when the innovations have a time-varying conditional covariance matrix, in the sense defined in Ling & McAleer (2002). Similarly, when V_t is a time-varying conditional covariance matrix, it can be shown that the maximum likelihood estimator of $\hat{C}$ is more efficient than its least squares estimator counterpart in the same sense. The maximum likelihood estimator of β_1 is also more efficient than its least squares estimator. Some comparisons can be found in Engle (1982) and Wong & Li (1997).

5. Reduced-rank estimation

The reduced-rank structure matrix C can be decomposed as $C = AB$, where A and B are full-rank matrices of dimensions $m \times r$ and $r \times m$, respectively. Such a decomposition is not unique. In particular, Johansen (1995) in his Theorem 4.2 showed that there exist A and B such that BY_t is stationary and $A'_\perp Y_t$ is a unit root process, where $A_\perp$ is an $m \times d$ matrix such that $A'_\perp A = 0$ and $\operatorname{span}(A, A_\perp) = I_m$. This decomposition is quite natural in explaining the cointegration relations, and the parameters in A and B can be estimated by using canonical correlation analysis. For model (1·1), this technique seems not to be very convenient since the conditional covariance matrix is changing from time to time.

Using the notation in § 2, we decompose $C = AB$ with $A = -P_2(I_r - \Gamma_r)Q'_{21}$, $B = [I_r, B_0]$ and $B_0 = Q'^{-1}_{21} Q'_{22}$, where $Q'_2 = (Q'_{21}, Q'_{22})$ and Q'_{21} is $r \times r$; see Reinsel & Ahn (1992). Such a decomposition is unique and B_0 is an $r \times d$ matrix of unknown parameters. For this decomposition, it is assumed that the components of series Y_t are arranged so that $J'Y_t$ is purely nonstationary, where $J' = [0, I_d]$. This assumption was used in Ahn & Reinsel (1990) and Yap & Reinsel (1995). As a referee has pointed out, it may not be easy to detect the order of the variables with data; see Maddala & Kim (1998, pp. 156–9). Based on this decomposition, model (2·1) can be rewritten further as

$$W_t = ABY_{t-1} + \Phi_1^* W_{t-1} \ldots + \Phi_{p-1}^* W_{t-p+1} + \varepsilon_t. \tag{5·1}$$

Let $\gamma_0 = \operatorname{vec}(B_0)$ and $\gamma_1 = \operatorname{vec}(A, \Phi_1^*, \ldots, \Phi_{p-1}^*)$. Then $\gamma = (\gamma'_0, \gamma'_1)'$ is the vector of unknown parameters with dimension $b = rd + mr + (p-1)m^2$. Furthermore, let $\hat{\gamma}_0 = \operatorname{vec}(\hat{B}_0)$, $\hat{\gamma}_1 = \operatorname{vec}(\hat{A}, \hat{\Phi}_1^*, \ldots, \hat{\Phi}_{p-1}^*)$ and $\hat{\gamma} = (\hat{\gamma}'_0, \hat{\gamma}'_1)'$. Define

$$\tilde{U}_{t-1}^* = [(J'Y_{t-1} \otimes A'), \tilde{U}'_{t-1} \otimes I_m]', \tag{5·2}$$

where $\tilde{U}_{t-1} = [(BY_{t-1})', W'_{t-1}, \ldots, W'_{t-p+1}]'$. The loglikelihood function is defined as in (4·3), with the parameter F replaced by γ. By directly differentiating (4·3), we obtain

$$\frac{\partial l_t}{\partial \gamma} = \sum_{i=1}^{q} \tilde{U}_{t-i-1}^* \eta_{it} + \tilde{U}_{t-1}^* V_t^{-1} \varepsilon_t. \tag{5·3}$$

Let $\bar{D}^{**} = \operatorname{diag}(nI_{rd}, \sqrt{n} I_{b-rd})$. As in Ling & Li (1998), we can show that $n^{-\frac{1}{2}} \bar{D}^{**-1} (\partial^2 l / \partial\gamma\, \partial\alpha') = o_p(1)$. Thus, γ and α can be estimated separately without loss of efficiency. As in § 4, we discuss only the estimator of γ. Again, as in Ling & Li (1998), we can show that

$$\bar{D}^{**-1} \frac{\partial^2 l}{\partial\gamma\, \partial\gamma'} \bar{D}^{**-1} = -\sum_{t=1}^{n} \bar{D}^{**-1} \tilde{M}_t \bar{D}^{**-1} + o_p(1),$$

where $\tilde{M}_t = \sum_{i=1}^{q} \tilde{U}_{t-i-1}^* V_{it} \tilde{U}_{t-i-1}^{*\prime} + \tilde{U}_{t-1}^* V_t^{-1} \tilde{U}_{t-1}^{*\prime}$. The maximum likelihood estimator of γ can be obtained by iterating the Newton–Raphson relation

$$\hat{\gamma}^{(i+1)} = \hat{\gamma}^{(i)} + \left(\sum_{t=1}^{n} \tilde{M}_t \right)^{-1}_{\gamma = \hat{\gamma}^{(i)}} \left(\sum_{t=1}^{n} \frac{\partial l_t}{\partial \gamma} \right)_{\gamma = \hat{\gamma}^{(i)}}, \tag{5·4}$$

where $\hat{\gamma}^{(i)}$ is the estimate at the ith iteration. As argued for (4·7), the estimator of γ obtained by (5·4) satisfies $\bar{D}^{**}(\hat{\gamma}-\gamma)=O_p(1)$ if the initial estimator also satisfies this condition.

Let $\tilde{C}=[\tilde{C}_1, \tilde{C}_2]$ be the full-rank least squares estimator or maximum likelihood estimator of C, where $\tilde{C}_1$ is $m \times r$. Then, using a similar method to that of Ahn & Reinsel (1990), we can show that $\tilde{A}=\tilde{C}_1$ is a consistent estimator of A of order $O_p(n^{-\frac{1}{2}})$ and $\tilde{B}_0=(\tilde{A}'\hat{\Omega}_n^{-1}\tilde{A})^{-1}\tilde{A}'\hat{\Omega}_n^{-1}\tilde{C}_2$ is a consistent estimator of B_0 of order $O_p(n^{-1})$, where $\hat{\Omega}_n=n^{-1}\sum_{t=1}^n \hat{\varepsilon}_t\hat{\varepsilon}_t'$. With these initial estimators and using a similar argument as for (4·7), we can obtain the asymptotic representation

$$\bar{D}^{**}(\hat{\gamma}-\gamma)=\left(\sum_{t=1}^{n} \bar{D}^{**-1}\tilde{M}_t\bar{D}^{**-1}\right)^{-1}\left(\sum_{t=1}^{n} \bar{D}^{**-1}\frac{\partial l_t}{\partial \gamma}\right)+o_p(1). \tag{5·5}$$

Let $J'P=[P_{21}, P_{22}]$, where P_{21} is $d \times d$ and P_{22} is $d \times r$. Then, $J'Y_t=[0, I_d]PZ_t=P_{21}Z_{1t}+P_{22}Z_{2t}$. Here, $J'Y_t$ and Z_{1t} are purely nonstationary, Z_{2t} is stationary, and P_{21} is nonsingular. Thus, terms involving Z_{2t} in the first rd components of $\bar{D}^{**-1}(\partial l/\partial \gamma)$ will converge to zero, and hence

$$\bar{D}^{**-1}\frac{\partial l}{\partial \gamma}=\sum_{t=1}^{n}\begin{pmatrix} n^{-1}\tilde{N}_{1t} \\ n^{-\frac{1}{2}}\tilde{N}_{2t}\end{pmatrix}+o_p(1),$$

$$\tilde{N}_{1t}=\sum_{i=1}^{q}(P_{21}Z_{1t-i-1}\otimes A)\eta_{it}+(P_{21}Z_{1t-1}\otimes A)V_t^{-1}\varepsilon_t,$$

$$\tilde{N}_{2t}=\sum_{i=1}^{q}(\tilde{U}_{t-i-1}\otimes I_m)\eta_{it}+(\tilde{U}_{t-1}\otimes I_m)V_t^{-1}\varepsilon_t.$$

As $n^{-3/2}\sum_{t=1}^n \tilde{U}_{t-i}Z_{1t-j-1}=o_p(1)$, for $i, j=1, \ldots, q$ (Ling & Li, 1998), the cross-product terms in $\sum_{t=1}^n \bar{D}^{**-1}\tilde{M}_t\bar{D}^{**-1}$ involving $\tilde{U}_{t-i}$ and Z_{1t-j} converge to zero in probability. Thus, using Lemma 1, we have the following theorem.

Theorem 3. *Let $\hat{B}_0$ and $\hat{\gamma}_1$ be the reduced-ranked estimators obtained from* (5·4) *using initial values $\tilde{B}_0$ and $\tilde{\gamma}_1$ such that $n^{-1}(\tilde{B}_0-B_0)=O_p(1)$ and $\sqrt{n}(\tilde{\gamma}_1-\gamma_1)=O_p(1)$. Then, under the same assumptions as in Theorem* 1,

(a) $$n(\hat{B}_0-B_0)\rightarrow(A'\Omega_1A)^{-1}A'\left\{\int_0^1 B_d(u)\,d\tilde{W}_m(u)'\right\}'\times\left\{\int_0^1 B_d(u)B_d(u)'\,du\right\}^{-1}\Omega_{a_1}^{-1/2}\Psi_{11}^{-1}P_{21}^{-1},$$

in distribution,

(b) $\sqrt{n}(\hat{\gamma}_1-\gamma_1)\rightarrow N(0, \tilde{\Omega}_u^{-1})$, *in distribution,*

where $\tilde{\Omega}_u=E(\sum_{i=1}^q \tilde{U}_{t-i-1}\tilde{U}_{t-i-1}'\otimes V_{it}+\tilde{U}_{t-i}\tilde{U}_{t-i}'\otimes V_t^{-1})$, and the other notation is as defined in Lemma 1.

As in the full-rank case, we can show that, when V_t is a constant matrix, the limiting distribution of $\hat{B}_0$ is the same as that given in Ahn & Reinsel (1990). Furthermore, generalisation of our results to the case with a constant nonzero drift parameter μ and $Q_1'\mu=0$ in (1·1) is direct.

In § 4 and in this section, we have constrained the conditional covariance matrix to be diagonal. From our derivations, the essential argument is to derive the limiting distributions of the information matrix and the score function. The former involves an

m-dimension Brownian motion and the latter involves the stochastic integral of two m-dimension Brownian motions as shown in Lemma 1(a) and (b). The difficult part is the former, which heavily depends on Theorem 2.1 in Ling & Li (1998) since $\sum_{i=1}^{q} V_{it} + V_t^{-1} - \Omega_1$ in (A·6) is not a martingale difference. To use this theorem, we need Lemma A3, which depends on expansion (2·6) of $\tilde{\varepsilon}_t$. For other types of ARCH model, the problem of under what conditions one can obtain the limiting distributions as in Lemma 1 seems difficult.

6. Simulation results

Data are generated from the following bivariate ARCH models.

Model 1:

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} 0{\cdot}75 & 0{\cdot}25 \\ 0{\cdot}25 & 0{\cdot}75 \end{pmatrix} \begin{pmatrix} Y_{1t-1} \\ Y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \tag{6·1}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} \delta_{11,t} & 0 \\ 0 & \delta_{22,t} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \tag{6·2}$$

In (6·2), we assume that $E(\varepsilon_{1t}\varepsilon_{2t}) = 0$ and $E(e_{1t}e_{2t}) = 0$. Furthermore, $\delta_{11,t}$, $\delta_{22,t}$, e_{1t} and e_{2t} are zero-mean, independent random variables with variances 0·36, 0·49, 0·01 and 0·09 respectively.

Thus, the equivalent representation for the innovation equation is

$$\begin{pmatrix} h_{1t} \\ h_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1}^2 \\ \varepsilon_{2t-1}^2 \end{pmatrix}, \tag{6·3}$$

where $\beta_{01} = 0{\cdot}01$, $\beta_{11} = 0{\cdot}36$, $\beta_{02} = 0{\cdot}09$ and $\beta_{22} = 0{\cdot}49$. It is easy to see that the reduced-rank parameters for the coefficient matrix in (6·1) are $-0{\cdot}25$, 0·25 and -1.

Model 2:

$$\begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix} = \begin{pmatrix} 0{\cdot}9 & 0{\cdot}1 \\ 0{\cdot}1 & 0{\cdot}9 \end{pmatrix} \begin{pmatrix} X_{1t-1} \\ X_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}, \tag{6·4}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} \eta_{11,t} & 0 \\ 0 & \eta_{22,t} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \tag{6·5}$$

The assumptions are the same as for Model 1. The variances of $\eta_{11,t}$, $\eta_{22,t}$, e_{1t} and e_{2t} are 0·49, 0·25, 0·16 and 0·04, respectively. Thus the corresponding coefficients in (6·3) are $\beta_{01} = 0{\cdot}16$, $\beta_{11} = 0{\cdot}49$, $\beta_{02} = 0{\cdot}04$ and $\beta_{22} = 0{\cdot}25$. The reduced-rank parameters for (6·4) are $-0{\cdot}1$, 0·1 and -1.

For each model, three sample sizes, $n = 200$, 400 and 800, are considered although the results for $n = 400$ are omitted, for brevity. These sample sizes can be regarded as small to moderate in financial applications. For Model 1, the eigenvalues of the matrix in (6·1) are 0·5 and 1. Similarly the eigenvalues for (6·4) are 0·8 and 1, so that both models represent systems of bivariate time series with partial nonstationarity. We calculate the least-squares, full-rank and reduced-rank estimates for each possible combination of model and sample size. The number of replications for each combination is 1000. The empirical means, sample standard errors and asymptotic standard errors of the estimates are also computed. The sample standard error of the estimates is just the standard error of the estimates in the 1000 replications. The asymptotic standard error is defined as the average of the theoretical standard errors, which in each simulation are obtained from the second derivatives of the loglikelihood. From maximum likelihood theory, the negative of the inverse of the matrix of second derivatives can be used to estimate the covariance matrix of the model parameters.

Table 1 lists the results for Model 1 and Model 2 respectively. The least-squares estimates provide initial values for the full-rank and reduced-rank estimation.

In both the full-rank and reduced-rank estimation, the parameters are estimated separ-

Table 1. *Simulation results. Empirical means, sample standard errors and asymptotic standard errors for Models* 1 *and* 2

	Model	LS		FR			RR		
n	parameters	Mean	SSE	Mean	SSE	ASE	Mean	SSE	ASE
				Model 1					
200	0·75	0·7480	0·0184	0·7491	0·0163	0·0153			
	0·25 (−0·25)	0·2493	0·0973	0·2453	0·0595	0·0576	−0·2509	0·0150	0·0153
	0·25 (0·25)	0·2503	0·0167	0·2494	0·0149	0·0144	0·2560	0·0538	0·0556
	0·75 (−1)	0·7337	0·0943	0·7449	0·0558	0·0551	−1·0009	0·0231	0·0188
	0·01	0·0108	0·0018	0·0099	0·0014	0·0014	0·0099	0·0015	0·0015
	0·36	0·2826	0·1270	0·3489	0·1238	0·1262	0·3500	0·1362	0·1268
	0·09	0·1236	0·0260	0·0924	0·0151	0·0152	0·0934	0·0151	0·0154
	0·49	0·3373	0·1357	0·4633	0·1349	0·1349	0·4558	0·1347	0·1392
800	0·75	0·7496	0·0081	0·7498	0·0069	0·0069			
	0·25 (−0·25)	0·2516	0·0591	0·2503	0·0279	0·0271	−0·2505	0·0069	0·0069
	0·25 (0·25)	0·2500	0·0080	0·2500	0·0068	0·0068	0·2507	0·0263	0·0267
	0·75 (−1)	0·7441	0·0589	0·7475	0·0278	0·0268	−0·9997	0·0052	0·0044
	0·01	0·0105	0·0012	0·0100	0·0007	0·0007	0·0100	0·0007	0·0007
	0·36	0·3181	0·0849	0·3546	0·0620	0·0630	0·3576	0·0654	0·0631
	0·09	0·1199	0·0191	0·0904	0·0074	0·0074	0·0905	0·0072	0·0074
	0·49	0·3999	0·0981	0·4860	0·0692	0·0696	0·4875	0·0687	0·0698
				Model 2					
200	0·9	0·8822	0·0549	0·8893	0·0374	0·0358			
	0·1 (−0·1)	0·1019	0·0180	0·1013	0·0167	0·0161	−0·1084	0·0381	0·0362
	0·1 (0·1)	0·1009	0·0589	0·0994	0·0428	0·0396	0·1031	0·0166	0·0162
	0·9 (−1)	0·8944	0·0210	0·8955	0·0194	0·0181	−1·0049	0·0874	0·0715
	0·16	0·1931	0·0397	0·1597	0·0265	0·0253	0·1625	0·0243	0·0255
	0·49	0·3451	0·1355	0·4747	0·1346	0·1311	0·4597	0·1328	0·1389
	0·04	0·0415	0·0063	0·0401	0·0059	0·0057	0·0400	0·0059	0·0058
	0·25	0·2051	0·1173	0·2365	0·1147	0·1132	0·2409	0·1154	0·1149
800	0·9	0·8955	0·0272	0·8981	0·0172	0·0173			
	0·1 (−0·1)	0·1006	0·0087	0·1005	0·0081	0·0077	−0·1008	0·0175	0·0172
	0·1 (0·1)	0·1004	0·0277	0·0994	0·0177	0·0177	0·1003	0·0077	0·0076
	0·9 (−1)	0·8987	0·0090	0·8989	0·0083	0·0079	−1·0000	0·0200	0·0169
	0·16	0·1834	0·0256	0·1596	0·0123	0·0124	0·1607	0·0123	0·0125
	0·49	0·4023	0·0980	0·4864	0·0691	0·0697	0·4873	0·0672	0·0698
	0·04	0·0407	0·0038	0·0400	0·0029	0·0029	0·0400	0·0029	0·0028
	0·25	0·2346	0·0734	0·2484	0·0552	0·0568	0·2487	0·0591	0·0569

Values in parentheses are parameters for reduced rank estimation.
LS, least-squares; FR, full-rank maximum likelihood; RR, reduced-rank maximum likelihood; SSE, sample standard error; ASE, asymptotic standard error.

ately, as stated in previous sections. This greatly reduces the computation as the dimensions of the matrices computed are highly reduced. In our case, we work with 4 by 4 and 2 by 2 matrices, instead of 8 by 8 matrices, and this advantage will be even greater if we are to work with higher-dimensional multivariate time series.

From Table 1, the following properties are noted. First, the full-rank and reduced-rank estimators are clearly better than the least-squares estimator in terms of both bias and efficiency. For instance, take Model 2 and $n=200$. Simple calculation gives the norms of the bias vectors of the least-squares and full-rank estimates to be 0·156 and 0·0235 respectively. The least-squares bias vector has a norm that is 6·6 times that of the full-rank case. The result is similar in other cases. Moreover, the sample standard error of the full-rank estimator is uniformly smaller than that of the least-squares estimator. In some cases the difference can exceed 30%. Note also that the asymptotic standard error for least-squares estimation is not listed. The method for calculating the theoretical standard error for the least-squares estimators is standard; see for example Lutkepohl (1993, Ch. 3). In our computation, however, it is found that the covariance matrices of the estimators of the parameters in the innovation equation are often negative definite, thus giving rise to complex values for the asymptotic standard error. This phenomenon can naturally be attributed to the imprecision of the least-squares estimates. Secondly, note that there is not much difference between the full-rank and reduced-rank estimates, in terms of both bias and efficiency. The same observation was made by Ahn & Reinsel (1990). However, as argued by Ahn & Reinsel (1990) the reduced rank formulation may provide better forecasting performance because of the imposition of the unit root on the series.

7. An example

To illustrate the presence of both cointegration and conditional heteroscedasticity, we consider the daily closing price of the Malaysian and Thai stock indices during the period 1989–91. This is a three-year span with a total of 695 obervations for each series. We try to model the centred data of the log prices. The data were also multiplied by 100 in Fig. 1; the two series are more separated in the first half of the plot, with the lower graph being the Thai series. They are much closer to each other and show many crossings in the second half. This agrees well with the usual error-correcting interpretation of cointegration; the two series can appear to be independent of each other for some time but tend to reach an equilibrium state in the long run. Using the Johansen Test from the package CATS in RATS (Hansen & Juselius, 1995), we find that there is one cointegrating vector; see Table 2.

To understand the conditional heteroscedasticity dynamics of the two series, we perform some preliminary analysis for the individual series. This time their first differences of logs are considered to ensure stationarity; the first differences are also scaled by a factor of 100. They can then be fitted by the univariate ARCH (1) model. In other words, if W_{1t} and W_{2t} are the scaled first differences of logs of the Malaysian and Thai stock indices respectively, then we have

$$W_{1t} = \underset{(0\cdot0370)}{0\cdot2794}\, W_{1t-1} + \varepsilon_{1t}, \quad h_{1t} = E(\varepsilon_{1t}^2 \mid \mathscr{F}_{t-1}) = \underset{(0\cdot3266)}{1\cdot1362} + \underset{(0\cdot1177)}{0\cdot3178}\, \varepsilon_{1t-1}^2,$$

$$W_{2t} = \underset{(0\cdot0378)}{0\cdot1521}\, W_{2t-1} + \varepsilon_{2t}, \quad h_{2t} = E(\varepsilon_{2t}^2 \mid \mathscr{F}_{t-1}) = \underset{(0\cdot3306)}{2\cdot2044} + \underset{(0\cdot1259)}{0\cdot5097}\, \varepsilon_{2t-1}^2;$$

the values in parentheses are the estimated standard errors. It is quite well known that the linear part, the observation equation, of stock prices usually contains a lag-1 term at most. We thus try a bivariate first-order ARCH model for the centred and scaled log data.

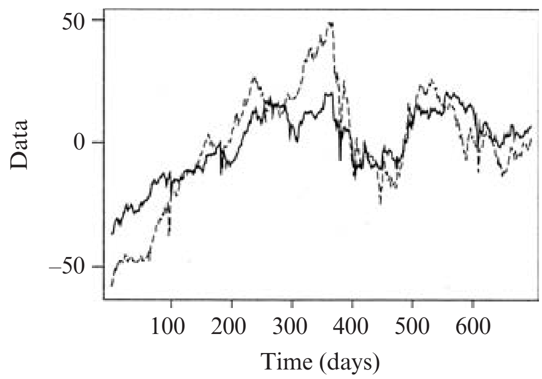

Fig. 1. Time series plot of transformed stock data, Malaysian stock (solid line) and Thai stock (broken line).

Table 2. *Malaysian and Thai stock indices: $I(1)$ analysis using the Johansen Test*

					Upper 90% critical value	
Eigenv.	L-max	Trace	$H_0:r$	$m-r$	L-max	Trace
0·0163	11·41	16·78	0	2	10·60	13·31
0·0077	5·37	5·37	1	1	2·71	2·71

Eigenv., eigenvalues corresponding to the maximised likelihood function; H_0, hypothesis about the cointegrating rank r; L-max, the likelihood ratio test statistic for testing r cointegrating vectors versus the alternative of $r+1$ cointegrating vector; Trace, the likelihood ratio test statistic for testing the hypothesis of at most r cointegrating vectors.

If Y_{1t} and Y_{2t} are the transformed data for the Malaysian and Thai stock indices, then our model is

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} Y_{1t-1} \\ Y_{2t-1} \end{pmatrix} + \begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix}, \tag{7·1}$$

$$\begin{pmatrix} \eta_{1t} \\ \eta_{2t} \end{pmatrix} = \begin{pmatrix} \delta_{11t} & 0 \\ 0 & \delta_{22t} \end{pmatrix} \begin{pmatrix} \eta_{1t-1} \\ \eta_{2t-1} \end{pmatrix} + \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \tag{7·2}$$

Based on our early discussion, (7·2) is equivalent to

$$\begin{pmatrix} h_{1t} \\ h_{2t} \end{pmatrix} = \begin{pmatrix} E(\eta_{1t}^2 | \mathscr{F}_{t-1}) \\ E(\eta_{2t}^2 | \mathscr{F}_{t-1}) \end{pmatrix} = \begin{pmatrix} \beta_{01} \\ \beta_{02} \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 \\ 0 & \beta_{22} \end{pmatrix} \begin{pmatrix} \eta_{1t-1}^2 \\ \eta_{2t-1}^2 \end{pmatrix}. \tag{7·3}$$

Let

$$\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}.$$

The estimation results are summarised in Table 3. Note that, for the full-rank estimation assuming no conditional heteroscedasticity the loglikelihood value is $-1291{\cdot}57$. It is interesting to make the following observations.

Table 3. *Estimates of parameters, with estimated standard errors in parentheses, for the Malaysian and Thai stock indices*

Parameter	Least-squares	Full-rank	Reduced-rank
ϕ_{11}	0·9780 (0·0071)	0·9771 (0·0062)	—
ϕ_{21}	−0·0050 (0·0041)	−0·0219 (0·0106)	—
ϕ_{12}	0·0085 (0·0123)	0·0095 (0·0036)	—
ϕ_{22}	0·9940 (0·0072)	1·0067 (0·0062)	—
β_{01}	0·9379 (0·2383)	0·8788 (0·0668)	0·8793 (0·0667)
β_{11}	0·3112 (0·0652)	0·4030 (0·0846)	0·4022 (0·0860)
β_{02}	2·5459 (0·4733)	2·1361 (0·1637)	2·1354 (0·1643)
β_{22}	0·3841 (0·0410)	0·5455 (0·0900)	0·5477 (0·0894)
a_1	—	—	−0·0220 (0·0062)
a_2	—	—	−0·0252 (0·0094)
b	—	—	−0·3866 (0·0605)
Loglikelihood	—	−1167·146	−1167·502

Remark 1. The full-rank and reduced-rank results are very similar. Note also the reasonable agreement between the univariate and bivariate results in the ARCH parameters. For the Malaysian series, the ARCH parameter estimates from univariate and bivariate estimation, in the reduced-rank case, are 0·3178 and 0·4022. Similarly, for the Thai series, the estimates for ARCH are 0·5097 and 0·5477. These results can serve as a crossvalidation for both methods.

Remark 2. The likelihood ratio test for ARCH versus no ARCH is highly significant. Standard errors for $\hat{\Phi}$ in the least-squares case are also higher, suggesting wider confidence intervals. In particular, both $\hat{\phi}_{21}$ and $\hat{\phi}_{12}$ are more than twice the standard error in magnitude under full-rank estimation while under least-squares estimation these are just about equal to their respective standard errors. Note that $\hat{\Phi}$ is asymptotically the maximum likelihood estimate of Φ with no ARCH.

Acknowledgement

The authors would like to thank the referees, the associate editor, the editor and Professor M. McAleer for their careful comments which led to improvements of this paper. W. K. Li's research is partially supported by the Hong Kong Research Grants Council. H. Wong's research is supported by the Central Research Grant of the Hong Kong Polytechnic University.

Appendix

Proof of Lemma 1

As mentioned in §§ 3 and 4, we only prove the result with V_t defined by (4·1); thus, V_t is diagonal in Lemmas 1, A1, A2 and A3.

LEMMA A1. *Suppose that the process ε_t is defined as in model* (1·1) *and that Assumptions* 1 *and* 2 *are satisfied. Then*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]}\begin{pmatrix}\varepsilon_t\\ \varepsilon_t^*\end{pmatrix}\to\begin{pmatrix}W_m(\tau)\\ \tilde{W}_m(\tau)\end{pmatrix}$$

in distribution in D^{2m}, *where* $\varepsilon_t^* = \sum_{i=1}^q \eta_{it} + V_t^{-1}\varepsilon_t$, $(W_m'(\tau), \tilde{W}_m'(\tau))'$ *are defined as in Lemma* 1, *and* $D^n = D \times D \ldots \times D$, *n factors, with D denoting the space of functions on* $[0, 1]$ *defined and equipped with the Skorokhod topology.*

Proof. Let λ_1 and λ_2 be constant $m \times 1$ vectors and $\lambda = (\lambda_1', \lambda_2')'$, with $\lambda\lambda' \neq 0$. Let $\xi_t = \lambda_1'\varepsilon_t + \lambda_2'\varepsilon_t^*$ and $S_n = \sum_{t=1}^n \xi_t$. It is obvious that ξ_t is a martingale difference sequence with respect to $\mathscr{F}_t$. By Assumptions 1 and 2, $\sigma^{*2} = n^{-1}ES_n^2 = \lambda_1' E(V_t)\lambda_1 + 2\lambda_1'\lambda_2 + \lambda_2'\Omega_1\lambda_2 < \infty$. When V_t is not a constant matrix, it is easy to show that Ω_b is positive definite, and hence $\sigma^{*2} > 0$. Furthermore,

$$E(\xi_t^2 | \mathscr{F}_{t-1}) = \lambda_1' V_t \lambda_1 + 2\lambda_1'\lambda_2 + \lambda_2' \sum_{i=1}^q V_{1t-i}^* \lambda_2,$$

where $V_{1t-i}^* = \text{diag}(2\sigma_{1i}^2\varepsilon_{1t-i}^2/h_{1t}^2 + 1/h_{1t}, \ldots, 2\sigma_{mi}^2\varepsilon_{mt-i}^2/h_{mt}^2 + 1/h_{mt})$. Note that $\{E(\xi_t^2|\mathscr{F}_{t-1})\}$ is a strictly stationary and ergodic time series. By the ergodic theorem and Assumption 2, we have that $(ES_n^2)^{-1}\sum_{t=1}^n (\xi_t^2|\mathscr{F}_{t-1}) = 1 + o_p(1)$. Since $\{\xi_t\}$ is also a strictly stationary and ergodic time series with finite variance for any small $\varepsilon > 0$,

$$\frac{1}{n}\sum_{t=1}^n E\{\xi_t^2 I(\xi_t \geqslant \sqrt{n}\sigma^*\varepsilon)\} = E\{\xi_t^2 I(\xi_t \geqslant \sqrt{n}\sigma^*\varepsilon)\} = \int_{x > \sqrt{n}\sigma^*\varepsilon} x^2\, dP(x) = o(1), \qquad (\text{A}\cdot 1)$$

where $P(x)$ is the distribution function of ξ_t. By the usual invariance principle for martingales, $n^{-1/2}\sum_{t=1}^{[n\tau]} \xi_t \to \sigma^* B(\tau)$ in distribution in D, where $B(\tau)$ is a univariate standard Brownian motion.

When V_t is a constant matrix, $\sigma^* = 0$. In a similar manner to the proof of Theorem 3.2 in Ling & Li (1998), we can still show that $n^{-1/2}\sum_{t=1}^{[n\tau]} \xi_t \to \sigma^* B(\tau)$ in distribution in D. By Proposition 4.1 of Wooldridge & White (1988), we complete the proof. □

Note that, for Theorem 1, we need the limiting distribution of $n^{-1/2}\sum_{t=1}^{[n\tau]} \varepsilon_t$ in Lemma A1 which holds without the diagonal assumption on V_t.

LEMMA A2. *Under the same assumptions as in Theorem* 1,

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{[n\tau]} u_{1k} \to \Psi_{11}\Omega_{a_1}^{1/2} B_d(\tau)$$

in distribution in D, where u_{1t} *is defined as in* (2·5), $B_d = ([I_d, 0]\Omega_a[I_d, 0]')^{-1/2}[I_d, 0]\Omega_a^{1/2} B_m$, $B_m = \Omega_a^{-1/2} Q W_m$ *and* $\Psi_{11} = [I_d, 0]\ (\sum_{k=1}^\infty \Psi_k)[I_d, 0]'$.

The proof of Lemma A2 can be found in a University of Hong Kong technical report by the authors. In order to prove Lemma 1, the following lemma plays an important role.

LEMMA A3. *Let* $s_{kt} = 2\sum_{i=1}^q \sigma_{ki}^2\varepsilon_{kt-i}^2\varepsilon_{kt}^2/h_{kt}^3 + h_{kt}^{-1}$. *Then, under the assumptions as in Theorem* 1, $n^{-1/2}\sum_{t=1}^{[n\tau]}\{s_{kt} - E(s_{kt})\} \to \sigma_k w_k(\tau)$ *in distribution, where* $k = 1, \ldots, m$, σ_k *is a nonnegative constant, and* w_k *is a standard Brownian motion.*

Proof. Let

$$G_{t-m} = \sigma\{\alpha_{1s}, \ldots, \alpha_{qs}, e_s | s = t-m, \ldots, t\}, \quad \varepsilon_{t-m}^* = \eta_{t-m} + \sum_{j=1}^m \left(\prod_{i=0}^{j-1} B_{t-m-i}\right)\eta_{t-m-j}$$

and $R_{t,m} = \varepsilon_t - \varepsilon_{t-m}^*$. We first show that

$$E(\| R_{t,m} \|^4) = O(\rho^m). \qquad (\text{A}\cdot 2)$$

Let $r_{t,j} = \eta_{t-j}'(\prod_{i=0}^j B_{t-i})'\prod_{i=0}^j B_{t-i}\eta_{t-j}$. Then, by (2·6) and a direct calculation,

$$E(r_{t,j}^2) = E\left[\operatorname{vec}'(\eta_{t-j}\eta'_{t-j})\operatorname{vec}\left\{\left(\prod_{i=0}^{j} B_{t-i}\right)'\prod_{i=0}^{j} B_{t-i}\right\}\right]^2$$
$$= c'_1 E\left(\prod_{i=0}^{j} B_{t-i}\otimes\prod_{i=0}^{j} B_{t-i}\otimes\prod_{i=0}^{j} B_{t-i}\otimes\prod_{i=0}^{j} B_{t-i}\right)'\operatorname{vec}\{\operatorname{vec}(I_m)\operatorname{vec}'(I_m)\}$$
$$= c'_1\{E(B_t^{\otimes 4})\}^{j+1}c_2 = O(\rho^j), \tag{A·3}$$

where c_1 and c_2 are constant vectors and $O(.)$ holds uniformly in t. Furthermore, we can show that

$$E(\|R_{t,m}\|^4) = E(\|R'_{t,m}R_{t,m}\|^2)$$
$$= \sum_{j=m+1}^{\infty} E(r_{t,j}^2) + \sum_{r=1}^{\infty}\sum_{j=m+1}^{\infty} E(r_{t,j-1}r_{t,j-1+r} + r_{t,j-1+r}r_{t,j-1}) = O(\rho^m), \tag{A·4}$$

so that (A·2) holds.

Since s_{kt} is a function of ε_{t-i} and $\varepsilon'_{t-i}\varepsilon_{t-j}$ $(i, j = 1, \ldots, q)$ by a direct calculation, we can show that $E\{s_{kt} - E(s_{kt}|G_{t-m})\}^2 = O(\rho^m)$. It is not difficult to verify that conditions C1–C4 of Theorem 3.2 in Ling & Li (1998) hold. Thus, $\sum_{t=1}^{[n\tau]}[s_{kt} - E(s_{kt})] \to \sigma_k w_k(\tau)$, where $w_k(\tau)$ is a standard Brownian motion. This completes the proof. □

Proof of Lemma 1. Note that

$$Z_{1t-i-1} = \sum_{k=1}^{t-i-1} u_{1k} = Z_{1t} - \sum_{k=t-i}^{t} u_{1k} = Z_{1t} + r_{it},$$

where $r_{it} = -\sum_{k=t-i}^{t} u_{1k}$. It is not difficult to show that

$$n^{-2}\sum_{t=1}^{n}\sum_{i=1}^{q}(Z_{1t-1}r'_{it} + r_{it}Z_{1t-1} + r_{it}r'_{it})\otimes\frac{\varepsilon_{kt-i}^2\varepsilon_{kt}^2}{h_{kt}^3} = o_p(1).$$

Furthermore, since $\sum_{i=1}^{q} V_{it} + V_t^{-1} = \operatorname{diag}(s_{1t}, \ldots, s_{mt})$, we have that

$$n^{-2}\sum_{t=1}^{n}\left(\sum_{i=1}^{q} Z_{1t-i-1}Z'_{1t-i-1}\otimes V_{it} + Z_{1t-1}Z'_{1t-1}\otimes V_t^{-1}\right)$$
$$= n^{-2}\sum_{t=1}^{n}\left\{Z_{1t-1}Z'_{1t-1}\otimes\left(\sum_{i=1}^{q} V_{it} + V_t^{-1}\right)\right\} + o_p(1). \tag{A·5}$$

By Theorem 3.1 in Ling & Li (1998), (2.5), Lemma A2 and Lemma A3, (A·5) is given by

$$n^{-2}\sum_{t=1}^{n}(Z_{1t-1}Z'_{1t-1}\otimes\Omega_1) + n^{-2}\sum_{t=1}^{n}\left\{Z_{1t-1}Z'_{1t-1}\otimes\left(\sum_{i=1}^{q} V_{it} + V_t^{-1} - \Omega_1\right)\right\} + o_p(1)$$
$$\to \Psi_{11}\Omega_{a_1}^{1/2}\left\{\int_0^1 B_d(u)B_d(u)'\,du\right\}\Omega_{a_1}^{1/2}\Psi'_{11}\otimes\Omega_1, \tag{A·6}$$

in distribution. By (A·5)–(A·6) part (a) holds. Similarly, we can show that

$$\frac{1}{n}\sum_{t=1}^{n} N_{1t} = \frac{1}{n}\sum_{t=1}^{n}(Z_{1t-1}\otimes I_m)\varepsilon_t^* + o_p(1). \tag{A·7}$$

By Theorem 2.2 in Kurtz & Protter (1991), Lemma A1 and (A·7), part (b) holds. Part (c) holds by the ergodic theorem and part (d) can be proved by the standard martingale central limit theorem. This completes the proof. □

References

Ahn, S. K. & Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate models. *J. Am. Statist. Assoc.* **85**, 813–23.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Economet.* **31**, 307–27.
BOLLERSLEV, T., ENGLE, R. F. & NELSON, D. B. (1994). ARCH models. In *Handbook of Econometrics*, **4**, Ed. R. Engle and D. McFadden, pp. 2959–3038. Amsterdam: North-Holland.
ENGLE, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of U.K. inflation. *Econometrica* **50**, 987–1008.
ENGLE, R. F. & GRANGER, C. W. J. (1987). Cointegrated and error correction: Representation, estimation and testing. *Econometrica* **55**, 251–76.
ENGLE, R. F. & KRONER, K. F. (1995). Multivariate simultaneous generalized ARCH. *Economet. Theory* **11**, 122–50.
FOUNTIS, N. G. & DICKEY, D. A. (1988). Testing for a unit root nonstationarity in multivariate time series. *Ann. Statist.* **17**, 419–28.
HANSEN, H. & JUSELIUS, K. (1995). *Cats in RATS*. Evanston, IL: Estima.
HODGSON, D. J. (1998). Adaptive estimation of error correction models. *Economet. Theory* **14**, 44–69.
JOHANSEN, S. (1988). Statistical analysis of cointegration vectors. *J. Econ. Dynam. Contr.* **12**, 231–54.
JOHANSEN, S. (1992). Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models. *Econometrica* **59**, 1551–80.
JOHANSEN, S. (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. New York: Oxford University Press.
KURTZ, T. G. & PROTTER, P. (1991). Weak limit theorems to stochastic integrals and stochastic differential equations. *Ann. Prob.* **19**, 1035–70.
LI, C. W. & LI, W. K. (1996). On a double threshold autoregressive conditional heteroskedastic time series model. *J. Appl. Economet.* **11**, 253–74.
LING, S. (1999). On the stationarity and the existence of moments of conditional heteroskedastic ARMA models. *Statist. Sinica* **9**, 1119–30.
LING, S. & DENG, W. C. (1993). Parametric estimate of multivariate autoregressive models with conditional heterocovariance matrix errors. *Acta Math. Appl. Sinica* **16**, 517–33.
LING, S. & LI, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Ann. Statist.* **26**, 84–125.
LING, S. & MCALEER, M. (2002). On adaptive estimation in nonstationary ARMA models with GARCH errors. *Ann. Statist.* To appear.
LUCAS, A. (1997). Cointegration testing using pseudolikelihood ratio tests. *Economet. Theory* **13**, 149–69.
LUTKEPOHL, H. (1993). *Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag.
MADDALA, G. S. & KIM, I-M. (1998). *Unit Roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press.
NICHOLLS, D. F. & QUINN, B. G. (1982). *Random Coefficient Autoregressive Model: An Introduction*, Lecture Notes in Statistics, **11**. New York: Springer.
RAHBEK, A. & MOSCONI, R. (1999). Cointegration rank inference with stationary regressors in VAR models. *Economet. J.* **2**, 76–91.
REINSEL, G. C. & AHN, S. K. (1992). Vector AR models with unit root and reduced rank structure: estimation, likelihood ratio test, and forecasting. *J. Time Ser. Anal.* **13**, 353–75.
SEO, B. (1998). Statistical inference on cointegration rank in error correction models with stationary covariates. *J. Economet.* **85**, 339–85.
SEO, B. (1999). Distribution theory for unit root tests with conditional heteroskedasticity. *J. Economet.* **91**, 113–44.
STOCK, J. H. & WATSON, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* **61**, 783–820.
TSAY, R. S. (1987). Conditional heteroscedastic time series models. *J. Am. Statist. Assoc.* **82**, 590–604.
WOOLDRIDGE, J. M. & WHITE, H. (1988). Some invariance principles and central limit theorems for dependent heterogeneous processes. *Economet. Theory* **4**, 210–30.
WONG, H. & LI, W. K. (1997). On a multivariate conditional heteroscedasticity model. *Biometrika* **84**, 111–23.
YAP, S. F. & REINSEL, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *J. Am. Statist. Assoc.* **90**, 253–67.

[*Received October* 1999. *Revised May* 2001]