Ann. Inst. Statist. Math.
Vol. 57, No. 1, 83–103 (2005)

JOINT MODELING OF COINTEGRATION AND CONDITIONAL HETEROSCEDASTICITY WITH APPLICATIONS

HEUNG WONG[1], W. K. LI[2] AND SHIQING LING[3]

[1]*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China*
[2]*Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China*
[3]*Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China*

(Received August 15, 2003; revised March 30, 2004)

Abstract. A cointegrated vector AR-GARCH time series model is introduced. Least squares estimator, full rank maximum likelihood estimator (MLE), and reduced rank MLE of the model are presented. Monte Carlo experiments are conducted to illustrate the finite sample properties of the estimators. Its applicability is then demonstrated with the modeling of international stock indices and exchange rates. The model leads to reasonable financial interpretations.

Key words and phrases: Cointegration, full rank maximum likelihood estimator, least squares estimator, partially nonstationary, reduced rank MLE, vector AR-GARCH model.

1. Introduction

Granger (1981) introduced the notion of co-integration and its use in time series analysis. It has then attracted the attention of many researchers, and in particular, analysts of economic time series. It is because the idea can be used to justify that many non-stationary economic variables will reach a long run equilibrium. There is a huge literature on cointegration. A good introduction to the field is Engle and Granger (1991). Banerjee *et al.* (1993) gave a guide to the literature on cointegration and modeling of integrated processes. On the other hand, Engle (1982) introduced the autoregressive conditional heteroscedasticity (ARCH) model. It relaxes the usual constant conditional variance assumption of linear time series models. Very naturally the model is enticing to researchers in finance and economics. It is common sense that many financial prices and economic indicators are volatile and the constant conditional variance assumption is impractical. The ARCH model has many extensions, including generalized autoregressive conditional heteroscedasticity, abbreviated by GARCH (Bollerslev (1986)), and integrated GARCH (Engle and Bollerslev (1986)), etc.

For a long time, however, there is not much research work on the joint modeling of cointegration and conditional heteroscedasticity. Lee and Tse (1996) discussed the issue of testing for cointegration in the presence of GARCH. Their work was based on simulation and no formal model was proposed. Li *et al.* (2001) was possibly the first paper that investigated a model which encompasses these two issues together. In their formulation, the conditional means of the relevant time series follow a partially nonstationary multivariate autoregressive model, which allows its innovations to be generated by a multivariate ARCH-type process, which is an extension of Tsay's (1987) CHARMA model. They also discussed the estimation of the model by least squares, maximum likelihood and reduced rank maximum likelihood approaches. The multivariate ARCH-type process in Li *et al.* (2001) cannot be extended to capture the persistence of the conditional heteroscedasticity. Ling and McAleer (2003) proposed a new vector ARMA-GARCH model which can overcome the drawback in Li *et al.* (2001).

In this paper, we will develop a partially nonstationary vector ARMA-GARCH model. Its applicability will be demonstrated with the modeling of international stock indices and exchange rates. Clearly they are highly important in international finance and macroeconomics. It goes without saying that numerous economists and financial experts publish on these data. We, however, reiterate again that most researchers investigate them from the standpoints of cointegration and conditional heteroscedasticity separately. There is a lot of empirical evidence that the presence of GARCH is ubiquitous in these data, see for example, the survey article by Bollerslev *et al.* (1992), and Bera and Higgins (1993). On the other hand, the empirical evidence on cointegration in these data is rather mixed.

Extensive study of international stocks dated back to early seventies, in papers by Granger and Morgenstern (1970), Grubel and Fadner (1971), and Levy and Samat (1970). From that time to the mid eighties, there were a number of empirical studies: e.g., Agmon (1972), and Jorion and Schwartz (1986), showing that there were no interdependence in international markets. In other words, there was no cointegration. The financial implication is that the national stock markets are segmented and hence international portfolio diversification is worthwhile. Note that these early studies relied largely on simple correlation and regression methodologies. In the nineties, however, with the financial market deregulation and improvements in information technology, there seemed to be more people believing in the interdependence of international equity markets. Another way of saying this is the existence of cointegration is highly probable. Some recent papers are Wu and Su (1998), and Gerrits and Yüce (1999). In these papers, more contemporary techniques like vector autoregressive (VAR) models, vector error correction (VEC) model (Engle and Granger (1987)), and Johansen's cointegration tests (Johansen (1988, 1992)) were used. For exchange rates data, even greater controversy has been reported in the literature. Borthe and Glassman (1987) reported no cointegration for several major currencies using the Engle and Granger (1987) test. Baillie and Pecchenino (1991) also concluded with no cointegration in a dollar-pound model using the Johansen technique (Johansen (1988)). However, MacDonald and Taylor (1992, 1993) reported the presence of the cointegrating vectors using the Johansen technique, while Cushman *et al.* (1996) also concluded the existence of co-integrating vectors in the exchange rates of several OECD countries.

The review above shows that the modeling of financial data has gotten more sophisticated with the development and application of modern techniques. The purpose of this article is to introduce a cointegrated conditional heteroscedastic time series model, which is a combination of two important existing models. Its application will be illustrated with stock prices data.

The rest of the paper is organized as follows: Section 2 describes a partially nonstationary vector AR-GARCH model and their properties. Section 3 considers the least squares estimator. Section 4 considers the full rank estimator. Section 5 considers the reduced rank estimator. A small simulation study is reported in Section 6. The international stock data and its modeling are discussed in Section 7. Section 8 considers the modeling of three exchange rates and Section 9 is the conclusion.

2. A cointegrated AR model with conditional heteroscedasticity

Although many econometricians and financial researchers believe that conditional heteroscedasticity exists in multivariate time series, how to formulate this feature is a difficult problem. Ling and Deng (1993), Wong and Li (1997), and Li *et al.* (2001) extended Tsay's (1987) CHARMA model to the multivariate cases. These models usually have too many parameters so that only some special cases are useful. Another multivariate extensions of Engle (1982) and Bollerslev (1986) was proposed by Engle and Kroner (1995). The general form of Engle and Kroner's (1995) model is quite complicated. Except for some special cases, many basic properties, such as strict stationarity and the positive definiteness of the conditional covariance, are not clear. Recently, Ling and McAleer (2003) proposed a new vector ARMA-GARCH models. As they argued, the vector ARMA-GARCH model seems to be simpler and more reasonable.

Suppose that the m-dimensional data $\{Y_t\}$ is generated by

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_p Y_{t-p} + \varepsilon_t, \tag{2.1}$$

$$\varepsilon_t = D_t \eta_t \quad \text{with} \quad H_t = W + A\tilde{\varepsilon}_{t-1} + BH_{t-1} \tag{2.2}$$

where Φ_i's are constant matrices; $\det\{\Phi(z)\} = |I - \Phi_1 z - \cdots - \Phi_p z^p| = 0$ has $d \le m$ unit roots and the remaining roots are outside the unit circle; rank $\{\Phi(z)\} = r$ with $r = m - d$; $\eta_t = (\eta_{1t}, \ldots, \eta_{mt})'$ is a sequence of independent and identically distributed (i.i.d.) standard normal vector, $D_t = \operatorname{diag}(h_{1t}^{1/2}, \ldots, h_{mt}^{1/2})$, $\tilde{\varepsilon}_t = (\varepsilon_{1t}^2, \ldots, \varepsilon_{mt}^2)'$, $W = (w_1, \ldots, w_m)'$, $H_t = (h_{1t}, \ldots, h_{mt})'$, $A = (\alpha_{ij})_{m\times m}$ and $B = \operatorname{diag}(\beta_1, \ldots, \beta_m)$. We call models (2.1)–(2.2) the partially nonstationary AR-GARCH model. We also assume that each component of the first difference $Y_t - Y_{t-1}$ is stationary. The restriction on the parameters for this can be found in Johansen (1996). The extension to higher lags in (2.2) is straightforward. The results in Ling and McAleer (2003) indicate that ε_t is strictly stationary and $E\varepsilon_{it}^2 < \infty$ if all the roots of $|I - AL - BL| = 0$ are outside the unit circle. The sufficient condition for the existence of the higher-order moment of ε_{it} can be found in the same paper.

When all the roots of $\det\{\Phi(z)\} = |I - \Phi_1 z - \cdots - \Phi_p z^p| = 0$ are outside the unit circle, Y_t is stationary if ε_t is stationary. In this case, the asymptotic theory can be found in Ling and McAleer (2003). For the partially nonstationary AR-GARCH models, we first need to reparameterize model (2.1) as

$$W_t = CY_{t-1} + \Phi_1^* W_{t-1} + \cdots + \Phi_{p-1}^* W_{t-p+1} + \varepsilon_t, \tag{2.3}$$

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

$$Z_t = \operatorname{diag}(I_d, \Gamma_r)Z_{t-1} + u_t \quad \text{and} \quad u_t = Q[\Phi_1^* W_{t-1} + \cdots + \Phi_{p-1}^* W_{t-p-1} + \varepsilon_t].$$

Furthermore, let $g(z) = (1-z)^{-d}det\{\Phi(z)\}$ and $H(z) = (1-z)^{-d+1}adj\{\Phi(z)\}$, we can rewrite u_t as

$$u_t = \left\{ I_m + Q\sum_{j=1}^{p-1}\Phi_j^* g(B)^{-1}H(B)PB^j \right\} a_t = \Psi(B)a_t$$

where $a_t = Q\varepsilon_t$ and $\Psi(B) = I_m + Q\sum_{j=1}^{p-1}\Phi_j^* g(B)^{-1}H(B)PB^j = \sum_{k=0}^{\infty}\Psi_k B^k$, in which $\Psi_0 = I_m$, $\Psi_k = O(\rho^k)$ and $\rho \in (0,1)$, as in Ahn and 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

$$Z_{1t} = Z_{1t-1} + u_{1t} \quad \text{and} \quad Z_{2t} = \Gamma_r Z_{2t-1} + u_{2t}.$$

Thus, $\{Z_{1t}\}$ is a nonstationary $d \times 1$ time series with d unit roots. However, $\{Z_{2t}\}$ is a stationary $r \times 1$ time series if ε_t is stationary. The matrix Q_2' is the so-called cointegrated vector with rank r, as in Engle and Granger (1987). The error correction form of model (2.1) can be found in Ahn and Reinsel (1990).

3. Preliminary estimation

We first consider the least squares estimator (LSE) of the parameters in (2.1). Let $X_{t-1} = [Y_{t-1}', W_{t-1}', \ldots, W_{t-p+1}']'$ and $F = [C, \Phi_1^*, \ldots, \Phi_{p-1}^*]$. Then, the LSE of F is $\bar{F} = (\sum_{t=1}^n W_t X_{t-1}')(\sum_{t=1}^n X_{t-1}X_{t-1}')^{-1}$, and hence $\bar{F} - F = (\sum_{t=1}^n \varepsilon_t X_{t-1}') \cdot (\sum_{t=1}^n X_{t-1}X_{t-1}')^{-1}$. Denote $Q^* = \mathrm{diag}(Q, I_{m(p-1)})$, $P^* = \mathrm{diag}(P, I_{m(p-1)})$, and $X_t^* = Q^*X_t = [Z_{1t}', U_t']'$, with $U_t = [Z_{2t}', W_{t-1}', \ldots, W_{t-p+2}']'$. Then

$$Q(\bar{F} - F)P^* = \left(\sum_{t=1}^n a_t X_{t-1}^{*'}\right)\left(\sum_{t=1}^n X_{t-1}^* X_{t-1}^{*'}\right)^{-1}.$$

Denote $D^* = \mathrm{diag}(D, \sqrt{n}I_{m(p-1)})$, where $D = \mathrm{diag}(nI_d, \sqrt{n}I_r)$. As $n^{-3/2}\sum_{t=1}^n \cdot U_{t-1}Z_{1t-1}' = o_p(1)$ (see Ling and Li (1998)), we can show that

$$\begin{aligned} Q(\bar{F} - F)P^*D^* = \Bigg[& \left(\frac{1}{n}\sum_{t=1}^n a_t Z_{1t-1}\right)\cdot\left(\frac{1}{n^2}\sum_{t=1}^n Z_{1t-1}Z_{1t-1}'\right)^{-1}, \\ & \left(\frac{1}{\sqrt{n}}\sum_{t=1}^n a_t U_{t-1}\right)\cdot\left(\frac{1}{n}\sum_{t=1}^n U_{t-1}U_{t-1}'\right)^{-1}\Bigg] + o_p(1). \end{aligned}$$

Thus, the estimators of the unit roots and the stationary parameters are asymptotically independent. As in Li *et al.* (2001), we can obtain the following result:

THEOREM 3.1. *If ε_t is strictly stationary and $E\varepsilon_{it}^4 < \infty$, then*

$$(\bar{F} - F)P^*D^* \to_{\mathcal{L}} P[M, N],$$

where $M = \Omega_a^{1/2}\{\int_0^1 B_d(u)dB_m(u)'\}'\{B_d(u)B_d(u)'du\}^{-1}\Omega_{a_1}^{-1/2}\Psi_{11}^{-1}$, $B_m(u)$ and $B_d(u) = \Omega_{a_1}^{-1/2}[I_d, 0]\Omega_a^{1/2}B_m(u)$ are standard Brownian motions; $\Omega_a = \mathrm{cov}(a_t) = QV_0Q'$, $V_0 =$

$E(\varepsilon_t\varepsilon_t')$; $\Omega_{a_1} = \mathrm{cov}(a_{1t}) = [I_d, 0]\Omega_a[I_d, 0]'$; $\Psi_{11} = [I_d, 0](\sum_{k=1}^{\infty}\Psi_k)[I_d, 0]'$; *and* $\mathrm{vec}(N)$ *is a normal vector with mean* 0 *and covariance* $E^{-1}(U_{t-1}U_{t-1}' \otimes I_m)E(U_{t-1}U_{t-1}' \otimes QD_t^2Q)E^{-1}(U_{t-1}U_{t-1}' \otimes I_m)$.

Let the residuals $\bar{\varepsilon}_t = Y_t - X_{t-1}\bar{F}$. Using $\bar{\varepsilon}_t$ as the artifical observations of ε_t, we estimate the parameters in (2.2) by the maximum likelihood estimator (MLE). Denote $\delta = (\alpha', \beta')'$ with $\alpha = (W', \mathrm{vec}'(A))'$ and $\beta = (\beta_1, \ldots, \beta_m)'$. The MLE of δ is $\bar{\delta}$, which maximizes the conditional log-likelihood function

$$L(\delta) = \sum_{t=1}^{n} l_t \quad \text{and} \quad l_t = -\frac{1}{2}\sum_{k=1}^{m}\log h_{kt} - \frac{1}{2}\sum_{k=1}^{m}\frac{\varepsilon_{kt}^2}{h_{kt}},$$

where h_{kt} is treated as the function of δ and ε_t. When ε_t in $L(\delta)$ is replaced by the estimated residuals $\bar{\varepsilon}_t$, $L(\delta)$ is denoted by $\bar{L}(\delta)$. Since $\bar{\varepsilon}_t = \varepsilon_t + O_p(D^{*-1})$, it is straightforward to show that the MLE of δ based on $\bar{L}(\delta)$ is asymptotically equivalent to that based on the true likelihood $L(\delta)$, see the proof in Ling *et al.* (2003) for the univariate case. From Ling and McAleer (2003), we know that $\bar{\delta}$ is $\sqrt{n}$–consistent and asymptotically normal if $E\varepsilon_t^6 < \infty$.

4. Full rank ML estimation

From Section 3, we know that $\bar{\delta}$ and $\bar{F}$ are consistent. In fact, $\bar{\delta}$ is also asymptotically efficient, but $\bar{F}$ is not. In order to obtain the efficient estimator of F, we need to use the maximum likelihood method. Rewrite model (2.3) as

$$W_t = CP_1Z_{1t-1} + CP_2Z_{2t-1} + \Phi_1^*W_{t-1} + \cdots + \Phi_{p-1}^*W_{t-p+1} + \varepsilon_t. \tag{4.1}$$

Denote $\beta_0 = \mathrm{vec}(CP_1)$, $\beta_1 = \mathrm{vec}(CP_2, \Phi_1^*, \ldots, \Phi_{p-1}^*)$, $\hat{\beta}_0 = \mathrm{vec}(\hat{C}P_1)$, $\hat{\beta}_1 = \mathrm{vec}(\hat{C}P_2, \hat{\Phi}_1^*, \ldots, \hat{\Phi}_{p-1}^*)$, and $\bar{Q}^* = \mathrm{diag}(Q \otimes I_m, I_{(p-1)m^2})$. Then $\bar{Q}^{*'-1}\mathrm{vec}(\hat{F} - F) = [(\hat{\beta}_0 - \beta_0)', (\hat{\beta}_1 - \beta_1)']'$. Let $\hat{F}$ be the MLE of F, which maximizes the conditional log-likelihood function

$$L(F) = \sum_{t=1}^{n} l_t \quad \text{and} \quad l_t = -\frac{1}{2}\sum_{k=1}^{m}\log h_{kt} - \frac{1}{2}\sum_{k=1}^{m}\frac{\varepsilon_{kt}^2}{h_{kt}}, \tag{4.2}$$

where ε_{kt} and h_{kt} are treated as the function of F and Y_t, and δ is replaced by $\bar{\delta}$. In the following, we use the notation: $\partial F = \partial\,\mathrm{vec}(F)$, where the *vec* operator transforms a matrix into a column vector by stacking the columns of the matrix below each other. By direct differentiation,

$$\frac{\partial l_t}{\partial F} = \frac{1}{2}\frac{\partial H_t'}{\partial F}D_t^{-2}\left[\left(1 - \frac{\varepsilon_{1t}^2}{h_{1t}}\right), \ldots, \left(1 - \frac{\varepsilon_{mt}^2}{h_{mt}}\right)\right]' + (X_t \otimes I_m)D_t^{-2}\varepsilon_t,$$

$$\frac{\partial H_t'}{\partial F} = 2(X_{t-1} \otimes I_m)\tilde{\varepsilon}_{t-1}^*A + \frac{\partial H_{t-1}'}{\partial F}B = 2\sum_{k=1}^{t-1}(X_{t-k} \otimes I_m)\tilde{\varepsilon}_{t-k}^*AB^{k-1},$$

where $\tilde{\varepsilon}_t^* = \mathrm{diag}(\varepsilon_{1t}, \ldots, \varepsilon_{mt})$. Denote $\bar{D}^* = D \otimes I_m = \mathrm{diag}(nI_{dm}, \sqrt{n}I_{rm+(p-1)m^2})$. As in Ling *et al.* (2003), we can show that

$$\bar{D}^{*-1}\bar{Q}^*\frac{\partial^2 L(F)}{\partial F \partial F'}\bar{Q}^{*'}\bar{D}^{*-1}$$

$$= -\sum_{t=1}^{n} \bar{D}^{*-1}\bar{Q}^{*}\left[\frac{1}{2}\frac{\partial H'_{t-1}}{\partial F}D_t^{-4}\frac{\partial H_{t-1}}{\partial F'}\right.$$
$$\left. + (X_{t-1}\otimes I_m)D_t^{-2}(X'_{t-1}\otimes I_m)\right]\bar{Q}^{*'}\bar{D}^{*-1} + o_p(1)$$
$$= \sum_{t=1}^{n} \bar{D}_t^{*-1}\bar{Q}^{*}M_t\bar{Q}^{*'}\bar{D}_t^{*-1} + o_p(1),$$

where $M_t = (\partial H'_{t-1}/\partial F)D_t^{-4}(\partial H_{t-1}/\partial F')/2 + (X_{t-1}\otimes I_m)D_t^{-2}(X'_{t-1}\otimes I_m)$.

As in Ling and Li (1998), we can show that $n^{-1/2}\bar{D}^{*-1}\bar{Q}^{*}\partial L^2(F)/\partial F\partial\delta' = o_p(1)$. Thus, F and δ can be estimated separately without loss of efficiency. The MLE of F can be obtained by the iterative approximate Newton-Raphson relation:

$$\hat{F}^{(i+1)} = \hat{F}^{(i)} + \left[\sum_{t=1}^{n} M_t^{(i)}\right]^{-1}\left[\sum_{t=1}^{n}\frac{\partial l_t}{\partial F}\right]\Bigg|_{F=F^{(i)}}, \tag{4.3}$$

where $F^{(i)}$ is the estimator at the i-th iteration and $\bar{F}$ (the LSE of F) is used as the initial estimator. Since $\bar{D}^{*}\bar{Q}^{*'-1}\text{vec}(\bar{F}-F) = O_p(1)$, using a similar argument as in Li *et al.* (2001), we can obtain the asymptotic representation:

$$\bar{D}^{*}\bar{Q}^{*'-1}\text{vec}(\hat{F}-F) = \left[\sum_{t=1}^{n}\bar{D}^{*-1}\bar{Q}^{*}M_t\bar{Q}^{*'}\bar{D}^{*-1}\right]^{-1}\left[\sum_{t=1}^{n}\bar{D}^{*-1}\bar{Q}^{*}\frac{\partial l_t}{\partial F}\right] + o_p(1).$$

Now partition $\bar{Q}^{*}(X_{t-i}\otimes I_m)$ into two parts corresponding to β_0 and β_1,

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

Furthermore, we can show that,

$$\bar{D}^{*-1}\bar{Q}^{*}\frac{\partial L(F)}{\partial F} = \sum_{t=1}^{n}\begin{pmatrix} N_{1t} \\ N_{2t}\end{pmatrix},$$
$$N_{1t} = -\frac{1}{2}\left[\sum_{i=1}^{t-1}(Z_{1t-i-1}\otimes I_m)\tilde{\varepsilon}_t AB^i\right]D_t^{-2}\xi_t + (Z_{1t-1}\otimes I_m)D_t^{-2}\varepsilon_t,$$
$$N_{2t} = -\frac{1}{2}\left[\sum_{i=1}^{\infty}(U_{t-i-1}\otimes I_m)\tilde{\varepsilon}_t AB^i\right]D_t^{-2}\xi_t + (U_{t-1}\otimes I_m)D_t^{-2}\varepsilon_t,$$

where $\xi_t = (1-\eta_{1t}^2,\ldots,1-\eta_{mt}^2)'$. As $n^{-3/2}\sum_{t=1}^{n}U_{t-i}Z_{1t-j-1} = o_p(1)$, $i,j = 1,\ldots,q$ (see Ling and 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 when multiplied by $n^{-3/2}$. Thus, we have

$$\sum_{t=1}^{n}\bar{D}^{*-1}\bar{Q}^{*}M_t\bar{Q}^{*'}\bar{D}^{*-1}$$
$$= \text{diag}\left\{n^{-2}\sum_{t=1}^{n}\left[\sum_{i=1}^{t-1}(Z_{1t-i-1}Z'_{1t-i-1}\otimes V_{it}) + Z_{1t-1}Z'_{1t-1}\otimes D_t^{-1}\right],\right.$$
$$\left. n^{-1}\sum_{t=1}^{n}\left[\sum_{i=1}^{\infty}(U_{t-i-1}U'_{t-i-1}\otimes V_{it}) + U_{t-1}U'_{t-1}\otimes D_t^{-2}\right]\right\} + o_p(1),$$

where $V_{it} = \tilde{\varepsilon}_t A B^i D_t^{-4} B^i A \tilde{\varepsilon}_t$.

Using the causal expansion in Ling and McAleer (2003) and following the method in Ling and Li (1998), we can obtain the following theorem:

THEOREM 4.1. *Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the full rank MLE obtained from (4.3). If ε_t is strictly stationary and $E\varepsilon_t^6 < \infty$, then*

$$\text{(a)}\ \ n(\hat{C} - C)P_1 \to_{\mathcal{L}} \Omega_1^{-1} \left\{ \int_0^1 B_d(u) d\tilde{W}_m(u)' \right\}' \times \left\{ \int_0^1 B_d(u) B_d(u)' \right\}^{-1} \Omega_{a_1}^{-1/2} \Psi_{11}^{-1};$$

$$\text{(b)}\ \ \sqrt{n}(\hat{\beta}_1 - \beta_1) \to_{\mathcal{L}} N(0, \Omega_u^{-1}),$$

where $B_d = \Omega_{a_1}^{-1/2}[I_d, 0]\Omega_a^{1/2} B_m(u)$ and $B_m(u) = V_0^{-1/2} W_m(u)$ are standard Brownian motions, $(W_m'(u), \tilde{W}_m'(u))'$ is a $2m$-dimensional Brownian motion with covariance given by $u\Omega_b = u \begin{pmatrix} V_0 & I_m \\ I_m & \Omega_1 \end{pmatrix}$, $\Omega_u = \sum_{i=1}^{\infty} E[U_{t-i-1} U_{t-i-1}' \otimes V_{it}) + E[U_{t-1} U_{t-1}' \otimes D_t^{-2}]$ and $\Omega_1 = ED_t^{-2} + \sum_{i=1}^{\infty} EV_{it}$.

When V_t is a constant matrix, $\Omega_1 = V_0$ and hence the limiting distribution of $\hat{C}$ reduces to that given in Theorem 3.1. In the univariate case, Ling and Li (1998) have shown that the MLE of the unit root is more efficient than the LSE of the unit root when the innovations have a time-varying conditional variance.

5. Reduced rank estimation

Using the notation in Section 2, we decompose $C = KB$ with $K = -P_2(I_r - \Gamma_r)Q_{21}'$, $B = [I_r, B_0]$ and $B_0 = Q_{21}'^{-1} Q_{22}'$, where $Q_2' = (Q_{21}', Q_{22}')$ and Q_{21}' is $r \times r$, see Reinsel and 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 and Reinsel (1990) and Yap and Reinsel (1995). Based on this decomposition, model (2.1) can be rewritten further as

$$W_t = KBY_{t-1} + \Phi_1^* W_{t-1} + \cdots + \Phi_{p-1}^* W_{t-p+1} + \varepsilon_t. \tag{5.1}$$

Denote $\gamma_0 = \text{vec}(B_0)$ and $\gamma_1 = \text{vec}(K, \Phi_1^*, \ldots, \Phi_{p-1}^*)$. Then, $\gamma = (\gamma_0', \gamma_1')'$ is the vector of unknown parameters with dimension $b = rd + mr + (p-1)m^2$. Define

$$\tilde{U}_{t-1}^* = [(J'Y_{t-1} \otimes K')', \tilde{U}_{t-1}' \otimes I_m]', \tag{5.2}$$

where $\tilde{U}_{t-1} = [(BY_{t-1})', W_{t-1}', \cdots, W_{t-p+1}']'$. The likelihood function is defined as in (4.2), with parameter F replaced by γ. By directly differentiating (4.2),

$$\frac{\partial l_t}{\partial \gamma} = \frac{1}{2} \frac{\partial H_t'}{\partial \gamma} D_t^{-2} \left[\left(1 - \frac{\varepsilon_{1t}^2}{h_{1t}}\right), \ldots, \left(1 - \frac{\varepsilon_{mt}^2}{h_{mt}}\right) \right]' + \tilde{U}_{t-1}^{*'} D_t^{-2} \varepsilon_t,$$

$$\frac{\partial H_t'}{\partial \gamma} = 2\tilde{U}_{t-1}^{*'} \tilde{\varepsilon}_{t-1}^* A + \frac{\partial H_{t-1}'}{\partial \gamma} B = 2 \sum_{k=1}^{t-1} \tilde{U}_{t-1}^{*'} \tilde{\varepsilon}_{t-k}^* A B^{k-1}.$$

Denote $\bar{D}^{**} = \operatorname{diag}(nI_{rd}, \sqrt{n}I_{b-rd})$. As in Ling and Li (1998), we can show that $n^{-1/2}\bar{D}^{**-1}(\partial^2 l/\partial\gamma\partial\delta') = o_p(1)$. Thus, γ and δ can be estimated separately without loss of efficiency. As in Section 4, we discuss only the estimator of γ. Again, as in Ling and 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 $\tilde{M}_t = (\partial H_t'/\partial\gamma)D_t^{-4}(\partial H_t/\partial\gamma')\tilde{U}_{t-i-1}^{*'} + \tilde{U}_{t-1}^{*}D_t^{-2}\tilde{U}_{t-1}^{*'}$.

Let $\hat{C} = [\hat{C}_1, \hat{C}_2]$ be the full-rank MLE of C, where $\hat{C}_1$ is $m\times r$. Then, using a similar method as in Reinsel and Ahn (1992), we can show that $\hat{K} = \hat{C}_1$ is a consistent estimator of K of order $O_p(n^{-1/2})$ and $\hat{B}_0 = (\hat{K}'\hat{\Omega}_n^{-1}\hat{K})^{-1}\hat{K}'\hat{\Omega}_n^{-1}\tilde{C}_2$ is a consistent estimator of B_0 of order $O_p(n^{-1})$, where $\hat{\Omega}_n = n^{-1}\sum_{t=1}^{n}\varepsilon_t\varepsilon_t'$. With this initial estimator, the MLE of γ can be obtained by iterating the relation:

$$\tilde{\gamma}^{(i+1)} = \tilde{\gamma}^{(i)} + \left[\sum_{t=1}^{n}\tilde{M}_t\right]^{-1}_{\gamma=\gamma^{(i)}}\left[\sum_{t=1}^{n}\frac{\partial l_t}{\partial\gamma}\right]_{\gamma=\gamma^{(i)}}, \tag{5.3}$$

where $\gamma^{(i)}$ is the estimator at the i-th iteration. As in Li *et al.* (2001), we can obtain the asymptotic representation:

$$\bar{D}^{**}(\tilde{\gamma}-\gamma) = \left[\sum_{t=1}^{n}\bar{D}^{**-1}\tilde{M}_t\bar{D}^{**-1}\right]^{-1}\left[\sum_{t=1}^{n}\bar{D}^{**-1}\frac{\partial l_t}{\partial\gamma}\right] + o_p(1). \tag{5.4}$$

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 $\bar{D}^{**-1}(\partial l_t/\partial\gamma)$ will converge to zero, and hence

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

$$\tilde{N}_{1t} = \sum_{k=1}^{t-1}(P_{21}Z_{1t-k-1}\otimes A)\tilde{\varepsilon}_{t-k}^{*}AB^{k-1}D_t^{-2}\xi_t + (P_{21}Z_{1t-1}\otimes A)D_t^{-2}\varepsilon_t,$$

$$\tilde{N}_{2t} = \sum_{k=1}^{\infty}(\tilde{U}_{t-k-1}\otimes I_m)\tilde{\varepsilon}_{t-k}^{*}AB^{k-1}D_t^{-2}\xi_t + (\tilde{U}_{t-1}\otimes I_m)D_t^{-2}\varepsilon_t.$$

As $n^{-3/2}\sum_{t=1}^{n}\tilde{U}_{t-i}Z_{1t-j-1} = o_p(1)$, $i, j = 1, \ldots, q$ (see Ling and 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. Similar to Li *et al.* (2001), we can show the following result:

THEOREM 5.1. *Let $\tilde{B}_0$ and $\tilde{\gamma}_1$ be the reduced-ranked estimators obtained from (5.4). Then, under the same assumptions as in Theorem 4.1,*

$$\text{(a)}\quad n(\hat{B}_0 - B_0) \to_{\mathcal{L}} (A'\Omega_1 A)^{-1}A'\left\{\int_0^1 B_d(u)d\tilde{W}_m(u)'\right\}'$$

$$\times \left\{ \int_0^1 B_d(u) B_d(u)' du \right\}^{-1} \Omega_{a_1}^{-1/2} \Psi_{11}^{-1} P_{21}^{-1};$$

(b) $\sqrt{n}(\hat{\gamma}_1 - \gamma_1) \to_{\mathcal{L}} N(0, \tilde{\Omega}_u^{-1})$,

where $\tilde{\Omega}_u = \sum_{i=1}^{\infty} E(\tilde{U}_{t-i-1}\tilde{U}'_{t-i-1} \otimes V_{it}) + E(\tilde{U}_{t-1}\tilde{U}'_{t-1} \otimes D_t^{-2})$, *and the other notation is defined as in Theorem* 4.1.

As in the full rank MLE 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 and Reinsel (1990). Generalizations of our results to the case with a constant non-zero drift parameter μ and $Q_1'\mu = 0$ in (2.1) is direct.

6. Simulation results

Our simulation experiment considers data generated from the following two equations.

$$Y_t = \Phi_1 Y_{t-1} + \varepsilon_t \tag{6.1}$$

$$\varepsilon_t = D_t \eta_t \quad \text{with} \quad H_t = W + A\tilde{\varepsilon}_{t-1} + BH_{t-1}. \tag{6.2}$$

These are special cases of equations (2.1) and (2.2) and the definitions of the relevant symbols were given in Section 2. We consider two models:

Model 1.

$$\Phi_1 = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

$$W = \begin{pmatrix} 0.15 \\ 0.1 \end{pmatrix}, \quad A = \begin{pmatrix} 0.2 & 0.1 \\ 0.15 & 0.1 \end{pmatrix}, \quad B = \begin{pmatrix} 0.6 & 0 \\ 0 & 0.7 \end{pmatrix}$$

Model 2.

$$\Phi_1 = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$$

$$W = \begin{pmatrix} 0.04 \\ 0.01 \end{pmatrix}, \quad A = \begin{pmatrix} 0.6 & 0.1 \\ 0.2 & 0.5 \end{pmatrix}, \quad B = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.1 \end{pmatrix}.$$

Remark. It is not difficult to see that the reduced rank parameters for the matrix $\begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$ and $\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$ are $(-0.25, 0.25, -1)$ and $(-0.1, 0.1, -1)$ respectively. Moreover, in both models, we have

$$\tilde{\varepsilon}_{t-1} = \begin{pmatrix} \varepsilon_{1t-1}^2 \\ \varepsilon_{2t-1}^2 \end{pmatrix} \quad \text{and} \quad H_{t-1} = \begin{pmatrix} h_{1t-1} \\ h_{2t-1} \end{pmatrix}.$$

For each model, three sample sizes, $n = 300$, 500, and 800 are considered. These sample sizes can be regarded as small to moderate in financial applications. For Model 1,

Table 1. Empirical means and sample standard errors for Model 1.

n	Model Parameters	LS Mean	LS SSE	FR Mean	FR SSE	RR Mean	RR SSE
300	0.75	0.7421	0.0485	0.7445	0.0373		
	0.25 (−0.25)	0.2542	0.0482	0.2531	0.0392	−0.2528	0.0370
	0.25 (0.25)	0.2519	0.0507	0.2511	0.0374	0.2558	0.0387
	0.75 (−1)	0.7385	0.0504	0.7419	0.0388	−1.0018	0.0301
	0.15			0.2066	0.1025	0.2064	0.1025
	0.2			0.2149	0.0718	0.2163	0.0703
	0.1			0.1264	0.0569	0.1249	0.0554
	0.6			0.5279	0.1196	0.5296	0.1181
	0.1			0.1505	0.0856	0.1504	0.0847
	0.15			0.1811	0.0621	0.1798	0.0617
	0.1			0.1088	0.0568	0.1086	0.0544
	0.7			0.6357	0.1019	0.6366	0.1009
500	0.75	0.7440	0.0378	0.7459	0.0285		
	0.25 (−0.25)	0.2528	0.0382	0.2524	0.0303	−0.2521	0.0287
	0.25 (0.25)	0.2525	0.0387	0.2517	0.0286	0.2532	0.0297
	0.75 (−1)	0.7432	0.0388	0.7447	0.0303	−1.0002	0.0162
	0.15			0.1813	0.0704	0.1857	0.0740
	0.2			0.2088	0.0549	0.2098	0.0535
	0.1			0.1141	0.0419	0.1125	0.0415
	0.6			0.5594	0.0895	0.5561	0.0906
	0.1			0.1322	0.0617	0.1301	0.0623
	0.15			0.1673	0.0454	0.1661	0.0456
	0.1			0.1029	0.0420	0.1017	0.0411
	0.7			0.6636	0.0724	0.6668	0.0765
800	0.75	0.7470	0.0301	0.7477	0.0217		
	0.25 (−0.25)	0.2520	0.0301	0.2508	0.0241	−0.2516	0.0221
	0.25 (0.25)	0.2508	0.0321	0.2508	0.0219	0.2518	0.0241
	0.75 (−1)	0.7456	0.0322	0.7476	0.0241	−1.0003	0.0100
	0.15			0.1713	0.0584	0.1699	0.0574
	0.2			0.2054	0.0430	0.2051	0.0428
	0.1			0.1065	0.0320	0.1068	0.0328
	0.6			0.5744	0.0730	0.5756	0.0730
	0.1			0.1024	0.0474	0.1191	0.0470
	0.15			0.1597	0.0373	0.1605	0.0374
	0.1			0.1000	0.0328	0.1002	0.0324
	0.7			0.6785	0.0608	0.6786	0.0617

the eigenvalues for Φ_1 are 0.5 and 1, whereas the corresponding eigenvalues for Model 2 are 0.8 and 1 , so that both models represent systems of bivariate time series with partial nonstationarity. We calculate the least-squares (LS), full-rank (FR) and reduced-rank (RR) estimates for each possible combination of model and sample size. The number

Table 2. Empirical means and sample standard errors for Model 2.

n	Model Parameters	LS Mean	LS SSE	FR Mean	FR SSE	RR Mean	RR SSE
300	0.9	0.8823	0.0552	0.8964	0.0233		
	0.1 (−0.1)	0.1059	0.0437	0.1028	0.0176	−0.1028	0.0228
	0.1 (0.1)	0.1091	0.0562	0.1008	0.0237	0.1030	0.0177
	0.9 (−1)	0.8889	0.0447	0.8958	0.0181	−0.9994	0.0463
	0.04			0.0419	0.0110	0.0422	0.0111
	0.6			0.6043	0.1177	0.6004	0.1140
	0.1			0.1375	0.0876	0.1388	0.0905
	0.2			0.1826	0.0864	0.1832	0.0870
	0.01			0.0101	0.0036	0.0103	0.0036
	0.2			0.2006	0.0455	0.1966	0.0442
	0.5			0.4919	0.1073	0.4880	0.1055
	0.1			0.1093	0.0626	0.1071	0.0625
500	0.9	0.8894	0.0457	0.8972	0.0169		
	0.1 (−0.1)	0.1040	0.0379	0.1008	0.0133	−0.1017	0.0173
	0.1 (0.1)	0.1052	0.0455	0.1009	0.0171	0.1015	0.0128
	0.9 (−1)	0.8926	0.0377	0.8982	0.0134	−0.9992	0.0263
	0.04			0.0414	0.0084	0.0413	0.0084
	0.6			0.6028	0.0978	0.6011	0.0941
	0.1			0.1207	0.0639	0.1229	0.0681
	0.2			0.1839	0.0717	0.1863	0.0721
	0.01			0.0100	0.0027	0.0101	0.0027
	0.2			0.1995	0.0364	0.2009	0.0360
	0.5			0.4891	0.0829	0.4880	0.0833
	0.1			0.1052	0.0492	0.1051	0.0498
800	0.9	0.8931	0.0364	0.8979	0.0134		
	0.1 (−0.1)	0.1049	0.0361	0.1005	0.0104	−0.1018	0.0136
	0.1 (0.1)	0.1036	0.0342	0.1012	0.0134	0.1012	0.0102
	0.9 (−1)	0.8932	0.0341	0.8990	0.0104	−1.0001	0.0151
	0.04			0.0413	0.0068	0.0411	0.0069
	0.6			0.6014	0.0805	0.5986	0.0810
	0.1			0.1100	0.0498	0.1109	0.0504
	0.2			0.1884	0.0612	0.1920	0.0619
	0.01			0.0099	0.0022	0.0099	0.0022
	0.2			0.2028	0.0279	0.2018	0.0272
	0.5			0.4949	0.0679	0.4947	0.0680
	0.1			0.1002	0.0391	0.1009	0.0391

of replications for each combination is 1000. The empirical means and sample standard errors of the estimates of Model 1 and Model 2 are computed and listed in Table 1 and Table 2 respectively.

From Tables 1 and 2, the following properties are noted. First, the full-rank and reduced-rank estimators are better than the least-squares in terms of bias and efficiency. The efficiency improvement should probably be emphasized. Note that in Table 2, the sample standard errors of the Φ_1 matrix from least-squares estimates are at least double those of the full-rank and reduced-rank. The reason is that in FR and RR estimates, we have taken care of the conditional heteroscedasticity in the data, which are neglected in the LS estimates. Such a big difference in the standard errors may cause problem in statistical inference about the parameters. Secondly, there is not much difference between the FR and RR estimates, in terms of both bias and efficiency. The same observation was made by Ahn and Reinsel (1990) and Li *et al.* (2001). However, as argued by Ahn and Reinsel (1990), the reduced-rank model may provide better forecasting performance.

7. An example: Standard and Poor's (SP500) and Sydney's All Ordinary (AO) indices

To illustrate the presence of both co-integration and GARCH, we consider the SP500 and the AO indices during the period January 1993 to June 1997. There are 1136 observation's for each series. The two series are of different magnitude, with the mean of AO series about 3.5 times that of SP500. To get a graphical understanding of the data, we multiply the SP500 series by 3.5 and plot it with the AO series in Fig. 1. It is interesting to observe that the two series looked quite different in the first 400 observations or so, but they show rather clear co-movements in roughly the last 200 observations.

This can be interpreted by the idea of cointegration, i.e., the two series are reaching an equilibrium near the end of the period. To explore their variance structures, we consider their first differences. The first differences of the two raw data series are shown in Figs. 2 and 3 respectively. Now it is quite clear that SP shows more conditional heteroscedasticity than AO. The SP graph shows small fluctuations at the beginning and larger fluctuations near the end. The AO graph does not show such behavior. We then fit a GARCH $(1, 1)$ model to the individual return series, i.e. first difference of logs times by 100. Let Z_t be the return series. We find $Z_t = a_t$ and a_t is first order white noise. Let $E(a_t^2 \mid \Psi_{t-1}) = \delta_0 + \delta_1 a_{1t-1}^2 + \beta h_{t-1}$, and the results for the two series are summarized Table 3, with values in brackets being standard errors. For convenience, δ_1

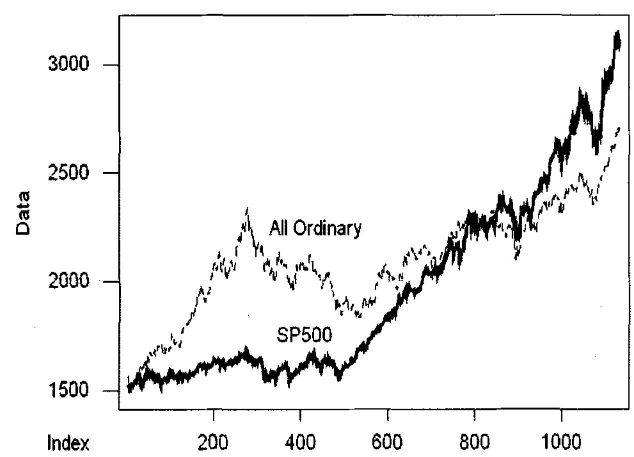

Fig. 1. Time plot of SP500 (transformed) and All Ordinary series.

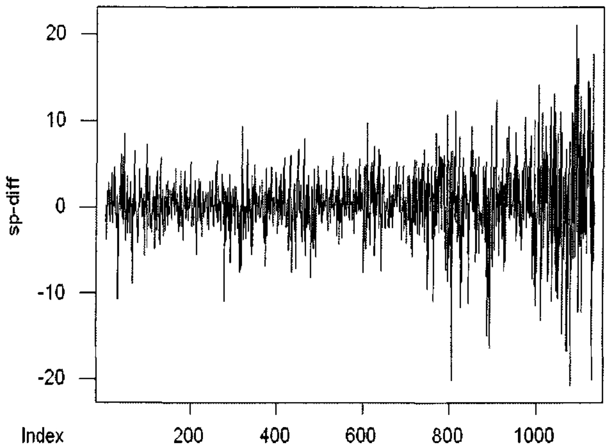

Fig. 2. Time plot of first differences of SP500 series.

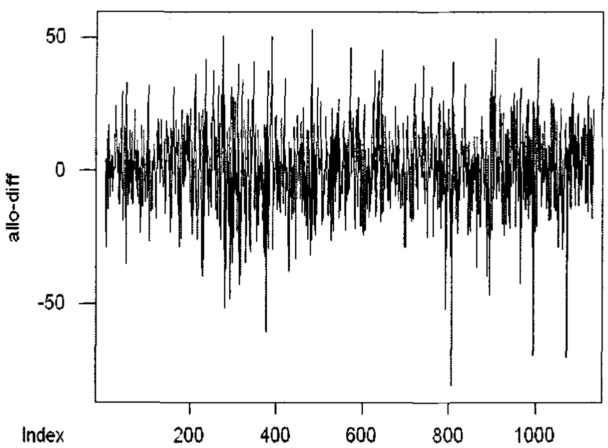

Fig. 3. Time plot of first differences of All Ordinary series.

Table 3.

	δ_0	δ_1	β
SP500	0.009 (0.0046)	0.0483 (0.012)	0.9329 (0.0177)
All Ordinary	0.1875 (0.167)	0.0716 (0.0299)	0.5991 (0.2875)

and β will be called the ARCH and GARCH coefficients.

Note that using t-ratio of 2 as yardstick, the parameters in the GARCH models are significant.

To check the presence of co-integration in the two series, we use Johansen's test from the package CATS in RATS (Hansen and Juselius (1995)). Note that like the Dickey-Fuller test, Johansen's test could be inefficient in the presence of GARCH (see

Table 4. SP500 and All Ordinary indices: I(1) analysis using the Johansen test.

					Upper 90% critical value	
Eigenv.	L-max	Trace	$H_0: r$	$m-r$	L-max	Trace
0.0113	12.90	15.70	0	2	10.60	13.31
0.0025	2.80	2.80	1	1	2.71	2.71

Eigenv., eigenvalues corresponding to the maximized 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

for example Sin and Ling (2004), Ling and Li (1998), and Ling *et al.* (2003)). There is one cointegrating vector found. See Table 4.

Thus the preliminary analysis shows that it is worthwhile to consider a conditional heteroscedastic model for the data. We try to model the centred data of the log prices. The centred data are also multiplied by 100. It is quite well-known that the observation equation of stock prices contains a lag-1 term at most. Thus a first order model for (2.1) is tried. If Y_{1t} and Y_{2t} are the transformed data for the SP500 and All Ordinary indices, then our model is

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

$$\begin{pmatrix} h_{1t} \\ h_{2t} \end{pmatrix} = \begin{pmatrix} E(\varepsilon_{1t}^2 \mid \mathcal{F}_{t-1}) \\ E(\varepsilon_{2t}^2 \mid \mathcal{F}_{t-1}) \end{pmatrix} = \begin{pmatrix} \alpha_{01} \\ \alpha_{02} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1}^2 \\ \varepsilon_{2t-1}^2 \end{pmatrix}$$
$$+ \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} h_{1t-1} \\ h_{2t-1} \end{pmatrix}.$$

Let $\Phi = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix}$. Recall $C = \Phi - I$ and $C = \binom{a_1}{a_2}(1 \ \ b)$. The results of the full rank and reduced rank estimation are summarised in Table 5.

Using a t-ratio of 2 as a yardstick again, it is interesting to observe that the volatility of the SP index is driven by itself but not the AO index. Note that α_{12} is not significant.

On the other hand, the volatility of the AO index is driven by the SP index but not by itself. This is because α_{21} is significant but α_{22} is not. Li *et al.* (2001) considered a cointegrated model with diagonal ARCH. The present example shows the need of a non-diagonal GARCH model. It should also be emphasized that both α_{22} and β_2 are not significant. This is an intriguing observation. In the univariate GARCH model, the AO index's volatility is driven by itself, since the ARCH and GARCH coefficients are significant. This is not true now. The change can be from the effect of cointegration, the ARCH effect from SP500 index, or from both. We shall come back to this point in the next example.

Finally, the likelihood ratio test for no GARCH versus GARCH is highly significant. The log-likelihood for the constant variance model is -949.962, which is much smaller

Table 5. Full rank and reduced rank estimates of parameters SP500 and All Ordinary indices.

Parameter	Full rank	Reduced rank
ϕ_{11}	1.0017 (0.0023)	
ϕ_{21}	0.0054 (0.0037)	
ϕ_{12}	−0.0019 (0.0025)	
ϕ_{21}	0.9902 (0.0044)	
α_{01}	0.0665 (0.0307)	0.0669 (0.0308)
α_{11}	0.1173 (0.0404)	0.1189 (0.0407)
α_{12}	0.0046 (0.0182)	0.0048 (0.0183)
β_1	0.8693 (0.0426)	0.8680 (0.0429)
α_{02}	0.7154 (0.1648)	0.7168 (0.1639)
α_{21}	0.4026 (0.1075)	0.4021 (0.1073)
α_{22}	0.0708 (0.0588)	0.0715 (0.0589)
β_2	0.0870 (0.1652)	0.0854 (0.1643)
a_1		0.0013 (0.0022)
a_2		0.0055 (0.0025)
b		−1.7752 (0.4929)
Loglikelihood	−916.828	−916.948

Note: the values in brackets are the standard errors.

than the two conditional heteroscedastic model.

8. Three exchange rates

8.1 *Preliminary analysis*

The data in this example are the daily closing rates of the Japanese Yen (Yen), German Mark (Mark) and the UK Sterling (Ster) against U.S. dollars during the period May 5, 1986 to June 5, 1995. For each series, there are 2354 data. Descriptive statistics of the three return series, i.e. first difference of logs, are shown in Table 6.

From Table 6, we observe that the kurtosis of Yen is larger than the other two exchange rates.

Next, we fit GARCH $(1,1)$ models to the individual return series. To avoid unnecessary numerical problems due to small variance, they are again multiplied by 100. All three individual series can be described by the same model, i.e., $Z_t = a_t$, a_t is the first order white noise such that

$$E(a_t^2 \mid \Psi_{t-1}) = h_t = \delta_0 + \delta_1 a_{t-1}^2 + \beta h_{t-1}.$$

Estimation results are summarized in Table 7.

Similar results have been widely reported in the literature. Both δ_1 and β are significant in all cases and $\delta_1 + \beta$ is close to 1. Test for cointegration is the final step in the preliminary analysis. The Johansen test as implemented in Hansen and Juselius (1995) package is used again. In parallel with our earlier transformation, the raw data are logged, centered and then multiplied by 100. There is no cointegration detected and the results are shown in Table 8.

Table 6. Descriptive statistics of returns.

Yen	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.00028128	0.0069645	−0.045473	0.039550	−0.66185	0.000048504	−0.26207	6.39885
Mark	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.00018689	0.0073068	−0.032339	0.036088	−0.43974	0.000053389	0.0048518	4.75041
Ster	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.000013718	0.0071065	−0.033163	0.047943	−0.032277	0.000050503	0.24489	5.78971

Table 7.

	δ_0	δ_1	β
Yen	0.047 (0.0086)	0.103 (0.0158)	0.806 (0.024)
Mark	0.038 (0.0079)	0.069 (0.0124)	0.860 (0.0221)
Ster	0.023 (0.0046)	0.065 (0.0100)	0.890 (0.0155)

N.B. Values in small brackets are standard errors.

Table 8.

Eigenv	L-max	Trace	$H_0 : r$	$p - r$	L-max90	Trace 90
0.0044	10.27	15.72	0	3	13.39	26.70
0.0023	5.41	5.46	1	2	10.60	13.31
0.0000	0.04	0.04	2	1	2.71	2.71

8.2 *Further analysis*

Up to now, there seems to be nothing very unusual. It confirms some earlier reports of no cointegration in exchange rates. We know that there were several important economic and political events in the early nineteen nineties. The economic bubble in Japan started to blow up and the communist countries in Europe toppled one after the other. With this in mind, we divide our data into roughly two equal parts. The first part is from May 5, 1986 to December 31, 1990, and the 2nd part is from January 2, 1991 to June 5, 1995. The two parts consist of 1204 and 1150 data, respectively. Descriptive statistics for the two parts are shown in Table 9.

It is clear that the kurtosis of the series in the 2nd part is in general larger than that in the first part, meaning that the conditional heteroscedasticity is larger. As before, we now test for cointegration. Results are shown in Tables 10(a) and 10(b).

From Table 10(a), by Johansen's test using a 5% significance level, there is clear cointegration. Note that both the L-max test and the trace test reject the H_0 of no cointegration. From Table 10(b), we see that the L-max test and the trace test are insignificant at 5% level. We thus conclude that there is no cointegration. There is a plausible explanation emerging from Tables 9 and 10. Conditional heteroscedasticity and cointegration in exchange rates may be striving for a delicate balance. When conditional heteroscedasticity is strong, cointegration will be weak, and if conditional heteroscedasticity is weak, then cointegration may prevail. On the other hand, as the two are related, it is reasonable to fit the data with a partially stationary vector AR-GARCH model. Models for the full data and the 2nd part of the data are tried. The model is defined as in (2.1)–(2.2).

Table 9(a). Descriptive statistics of returns, 1986–1990.

Yen1	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.00016396	0.0070326	−0.030104	0.039550	−0.19724	0.000049458	−0.027654	5.49386
Mark1	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.00032070	0.0068636	−0.028593	0.032256	−0.38580	0.000047109	−0.095485	4.51995
Ster1	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.000188802	0.0067203	−0.027287	0.027795	−0.22619	0.000045162	0.13162	4.88755

Table 9(b). Descriptive statistics of returns, 1991–1995.

Yen2	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.00039855	0.0068939	−0.045473	0.037274	−0.45793	0.000047526	−0.52725	7.41425
Mark2	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	−0.000046660	0.0077472	−0.032339	0.036088	−0.053612	0.000060019	0.066136	4.78591
Ster2	Mean	Std Dev	Minimum	Maximum	Sum	Variance	Skewness	Kurtosis
	0.00017280	0.0074898	−0.033163	0.047943	0.19855	0.000056097	0.31372	6.28681

Table 10(a). Co-integration test for the first part data, 1986–1990.

Eigenv	L-max	Trace	$H_0 : r$	$p - r$	L-max95	Trace 95
0.0171	20.76	37.64	0	3	17.89	24.31
0.0079	9.59	16.89	1	2	11.44	12.53
0.0060	7.29	7.29	2	1	3.84	3.84

Table 10(b). Co-integration test for the second part data, 1991–1995.

Eigenv	L-max	Trace	$H_0 : r$	$p - r$	L-max95	Trace 95
0.0124	14.35	22.52	0	3	17.89	24.31
0.0070	8.08	8.17	1	2	11.44	12.53
0.0001	0.09	0.09	2	1	3.84	3.84

Results for the full data are:
Loglikelihood = −938.22

$$\hat{\Phi} = \begin{bmatrix} 0.9975(0.0014) & 0.0013(0.0024) & -0.0027(0.0020) \\ -0.0018(0.0015) & 0.9988(0.0026) & -0.0038(0.0022) \\ -0.0032(0.0014) & 0.0039(0.0025) & 0.9947(0.0021) \end{bmatrix}$$

$$\hat{G} = [0.0525(0.014), 0.417(0.0088), 0.0228(0.0052)]'$$

$$\hat{A} = \begin{bmatrix} 0.1048(0.0171) & 0 & 0.0053(0.0059) \\ 0 & 0.0663(0.0122) & 0.0054(0.0092) \\ 0 & 0.0001(0.0073) & 0.0644(0.0112) \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 0.7869(0.0338) & 0 & 0 \\ 0 & 0.8523(0.0209) & 0 \\ 0 & 0 & 0.891(0.016) \end{bmatrix}.$$

N.B.1: Values in small brackets are standard errors.
N.B.2: The BHHH algorithm is used in the maximum likelihood estimation.
N.B.3: Small negative values in the A matrix are set to zero.
Results for the 2nd part of the data are:
Loglikelihood $= -555.29$

$$\hat{\Phi} = \begin{bmatrix} 0.9961(0.0029) & 0.004(0.0054) & -0.0066(0.0043) \\ 0.0066(0.0035) & 0.9849(0.0065) & 0.0063(0.0054) \\ 0.0002(0.0036) & -0.0078(0.0067) & 0.9988(0.0058) \end{bmatrix}$$

$$\hat{G} = [0.0809(0.0264), 0.081(0.0257), 0.0346(0.0091)]'$$

$$\hat{A} = \begin{bmatrix} 0.1361(0.0312) & 0 & 0.0083(0.0094) \\ 0 & 0.0671(0.0243) & 0.0208(0.0208) \\ 0 & 0 & 0.1021(0.0186) \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 0.6947(0.0736) & 0 & 0 \\ 0 & 0.7816(0.0504) & 0 \\ 0 & 0 & 0.8399(0.0273) \end{bmatrix}.$$

Following a similar argument as in Section 7, we fit the ordinary AR1 model to the full data and the data in the second part.

Results for the full data are:
The loglikelihood value is -1126.68.

$$\hat{\Phi} = \begin{pmatrix} 0.9979(0.0015) & 0.0038(0.0026) & -0.0042(0.0023) \\ -0.0013(0.0016) & 0.9992(0.0027) & -0.0038(0.0024) \\ -0.0023(0.0016) & 0.0015(0.0027) & 0.9951(0.0023) \end{pmatrix}.$$

Results for the data of the second part are:
The loglikelihood value is -657.88.

$$\hat{\Phi} = \begin{pmatrix} 0.9962(0.0031) & 0.0056(0.0058) & -0.008(0.0047) \\ 0.0065(0.0035) & 0.9848(0.0064) & 0.007(0.0053) \\ 0.0052(0.0034) & -0.0161(0.0062) & 1.0032(0.0051) \end{pmatrix}.$$

Therefore, for the full data and the second part data, the likelihood ratio statistic for the presence of GARCH in the AR1 model are 376.92 and 102.59 respectively. Obviously both statistics are highly significant. Also note that for the full data, the standard errors of the parameters in the observation equation are uniformly smaller than their counterparts in the linear AR1 model. This suggests asymptotically we have smaller confidence intervals and hence better power in statistical inference.

The same model is tried with the data in first part. There are no convergent estimates. A possible explanation is that there is not much GARCH effects in the data with the presence of cointegration, as in the stocks example. The model is misspecified, so there is no convergence when we try to fit a vector GARCH model. To verify the conjecture, we fit the three parts by least squares and then use the estimated squared residuals to test for ARCH disturbances, using the Lagrange multiplier test suggested by Engle (1982). The procedure is as follows.

Table 11.

			p		
			1	2	3
		First	18.82*	32.91*	37.60*
Yen	Data	Second	28.22*	28.32*	30.55*
	Portion	Whole	23.52*	25.86*	28.20*
		First	0.85	4.70	5.07
Mark	Data	Second	21.17*	28.21*	28.22*
	Portion	Whole	11.76*	18.81*	19.45*
		First	0.28	0.76	2.35
Ster	Data	Second	14.11*	32.09*	35.44*
	Portion	Whole	9.41	39.97*	46.06*

N.B. * denotes a value significant at 5% level.

1. Fit the model

$$Y_t = \Phi Y_{t-1} + \varepsilon_t.$$

Estimate Φ by least squares. Then $\hat{\varepsilon}_t = Y_t - \hat{\Phi} Y_{t-1}$ as the estimated residuals, and $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t}, \hat{\varepsilon}_{3t})'$.

2. As indicated in Engle ((1982), p. 100), we regress $\hat{\varepsilon}_{1t}^2$ on a constant and p lags and test nR^2 as a χ_p^2. Here n is the sample size and R^2 is the coefficient of determination. Similar tests are constructed for $\hat{\varepsilon}_{2t}^2$ and $\hat{\varepsilon}_{3t}^2$. In the study, we try with the values $p = 1$, 2 and 3 and the results are summarized in Table 11. Entries in the table are nR^2 values from the regressions.

The values are consistent with our belief. In the first half of the data, only Yen has conditional heteroscedasticity; whereas for the second part of the whole data, all the exchange rates have conditional heteroscedasticity. It is thus worthwhile to redo the estimation for the first half of the data, letting Yen to be conditionally heteroscedastic and the other two to be conditionally homoscedastic. It is interesting that the iterations now converge and the results are reported below.

Loglikelihood $= -398.79$

$$\hat{\Phi} = \begin{bmatrix} 0.9833(0.0039) & -0.0091(0.0042) & 0.0207(0.0061) \\ -0.0061(0.0040) & 1.0014(0.0042) & -0.0021(0.0063) \\ -0.0035(0.0039) & 0.0097(0.0041) & 0.9915(0.0062) \end{bmatrix}$$

$$\hat{G} = [0.0881(0.0216), 0.4685(0.0191), 0.4481(0.0183)]'$$

$$\hat{A} = \begin{bmatrix} 0.1182(0.0253) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 0.7069(0.0548) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

From Johansen's test in Table 10(a), there should be two unit roots and one cointegrating vector for this data set. Using the reduced rank estimation in Section 5, the cointegrating vector is found to be $[1.37, 0.27, -1.03]'$.

9. Conclusion

A cointegrated vector AR-GARCH model is proposed to investigate cointegration and conditional heteroscedasticity simultaneously. Least squares and maximum likelihood estimation of the model parameters are presented. In a set of stocks data, we observe that when cointegration exists, conditional heteroscedasticity seems to be weakened. In a set of exchange rates data, we observe the phenomenon that when conditional heteroscedasticity is strong, cointegration seems to be weakened, and the converse is also true. This relationship cannot be revealed if the series are being investigated for the two properties separately. The phenomenon is interesting and it deserves further work to study the joint modeling of cointegration and conditional heteroscedasticity.

Acknowledgements

We thank the referees for their constructive comments that lead to improvement of the paper. The research of W. K. Li is supported by Competitive Earmarked Research Grant HKU7149/98H. Shiqing Ling's research is supported partially by Competitive Earmarked Research Grants HKUST6113/02P and HKUST6273/03H. The research of Heung Wong is supported by a grant of the Hong Kong Polytechnic University, under RGC direct allocation.

References

Agmon, T. (1972). The relations among equity markets: A study of share price comovements in the United States, United Kingdom, Germany and Japan, *Journal of Finance*, **27**, 193–208.

Ahn, S. K. and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate models, *Journal of the American Statistical Association*, **85**, 813–823.

Baillie, R. and Pecchenino, R. A. (1991). The search for equilibrium relationships in international finance: The case of the monetary model, *Journal of International Money and Finance*, **10**, 582–593.

Banerjee, A., Dolado, J., Galbraith, J. and Hendry, D. F. (1993). *Co-integration, Error-correction, and the Econometric Analysis of Non-stationary Data*, Oxford University Press, New York.

Bera, A. K. and Higgins, M. L. (1993). ARCH models: Properties, estimating and testing, *Journal of Economic Surveys*, **7**, 305–362.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T., Chou, R. Y. and Kroner, K. F. (1992). ARCH modeling in finance, *Journal of Econometrics*, **52**, 5–59.

Borthe, P. and Glassman, D. (1987). Off the Mark: Lessons for exchange rate modeling, *Oxford Economic Papers*, **39**, 443–457.

Cushman, D., Lee, S. and Thorgeirsson, T. (1996). Maximum likelihood estimation of cointegration in exchange rate models for seven inflationary OECD countries, *Journal of International Money and Finance*, **15**, 337–369.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of variance of U. K. inflation, *Econometrica*, **50**, 987–1008.

Engle, R. F. and Bollerslev, T. (1986). Modeling the persistence of conditional variances, *Econometric Reviews*, **5**, 1–50.

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

Engle, R. F. and Granger, C. W. J. (1991). *Long-run Economic Relationship, Reading in Cointegration*, Oxford University Press, New York.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH, *Econometric Theory*, **11**, 122–150.

Gerrits, R. and Yüce, A. (1999). Short- and long-term links among European and US stock markets, *Applied Financial Economics*, **9**, 1–9.

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

Granger, C. W. J. and Morgenstern, O. (1970). *Predictability of Stock Market Price*, Lexington Books, Lexington, Massachusetts.

Grubel, H. G. and Fadner, K. (1971). The interdependence of international equity markets, *Journal of Finance*, **26**, 89–94.

Hansen, H. and Juselius, K. (1995). CATS in RATS, Estima, Evanston, Illinois.

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

Johansen, S. (1992). Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551–1580.

Johansen, S. (1996). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, New York.

Jorion, P. and Schwartz, E. (1986). Integration versus segmentation in the Canadian stock market, *Journal of Finance*, **41**, 603–616.

Lee, T. H. and Tse, Y. (1996). Cointegration tests with conditional heteroskedasticity, *Journal of Econometrics*, **73**, 401–410.

Levy, H. and Samat, M. (1970). International diversification of investment portfolios, *American Economic Review*, **60**, 668–675.

Li, W. K., Ling, S. and Wong, H. (2001). Estimation for partially nonstationary multivariate autoregressive models with conditional heteroscedasticity, *Biometrika*, **88**, 1135–1152.

Ling, S. and Deng, W. C. (1993). Parametric estimate of multivariate autoregressive models with conditional heteroscovariance matrix errors, *Acta Mathematicae Applicitae Sinica*, **16**, 517–533.

Ling, S. and Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors, *Annals of Statistics*, **26**, 84–125.

Ling, S. and McAleer, M. (2003). Asymptotic theory for a new vector ARMA-GARCH model, *Econometric Theory*, **19**, 280–310.

Ling, S., Li, W. K. and McAleer, M. (2003). Estimation and testing for a unit root process with GARCH (1, 1) errors, *Econometric Reviews*, **22**, 179–202.

MacDonald, R. and Taylor, M. (1992). Exchange rate economics: A survey, *IMF Staff Papers*, **39**, 1–57.

MacDonald, R. and Taylor, M. (1993). The monetary approach to the exchange rate: Rational expectation, long-run equilibrium, and forecasting, *IMF Staff Papers*, **40**, 89–107.

Reinsel, G. C. and Ahn, S. K. (1992). Vector AR models with unit root and reduced rank structure: Estimation, likelihood ratio test, and forecasting, *Journal of Time Series Analysis*, **13**, 133–145.

Sin, C. and Ling, S. (2004). Estimation and testing for partially nonstationary vector autoregressive models with GARCH (submitted).

Tsay, R. S. (1987). Conditional heteroscedastic time series models, *Journal of the American Statistical Association*, **7**, 590–604.

Wong, H. and Li, W. K. (1997). On a multivariate conditional heteroscedasticity model, *Biometrika*, **84**, 111–123.

Wu, C. and Su, Y. C. (1998). Dynamic relations among international stock markets, *International Review of Economics and Finance*, **7**, 63–84.

Yap, S. F. and Reinsel, G. C. (1995). Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model, *Journal of the American Statistical Association*, **90**, 253–267.