Determining the Number of Factors in a Multivariate Error Correction–Volatility Factor Model

QIAOLING LI[†] AND JIAZHU PAN[†‡]

[†]*School of Mathematical Sciences, Peking University, Beijing 100871, China*
Email: lql@math.pku.edu.cn

[‡]*Department of Statistics and Modelling Science, University of Strathclyde, U.K.*
Email: jiazhu@stams.strath.ac.uk

Summary   In order to describe the comovements in both conditional mean and conditional variance of high dimensional nonstationary time series by dimension reduction, we introduce the conditional heteroscedasticity with factor structure to the error correction model. The new model is called the error correction–volatility factor model. Some specification and estimation approaches are developed. In particular, the determination of the number of factors is discussed. Our setting is general in the sense that we impose neither i.i.d assumption on idiosyncratic components in the factor structure nor independence between factors and idiosyncratic errors. We illustrate the proposed approach with a Monte Carlo simulation and a real data example.

Key words: *Dimension reduction, Cointegration, Error correction–volatility factor model, Penalized goodness-of-fit criteria, Model selection.*

1 Introduction

The concept of cointegration (Granger (1981), Granger and Weiss (1983), Engle and Granger (1987)) has been successfully applied to modelling multivariate nonstationary time sereis. The literature on cointegration is extensive. The most frequently used representations for a cointegrated system are the error correction model (ECM) of Engle and Granger (1987), the common trends form of Stock and Watson (1988), and the triangular model of Phillips (1991). The error correction model has been applied in various practical problems, such as determining exchange rates, capturing the relationship between expenditure and income, modelling and forecasting inflation, etc. From the equilibrium point of view, the term "error correction" reflects the correction on the long-run relationship by the short-run dynamics.

However, the error correction model ignores the characteristics of time-varying volatility, which plays an important role in various financial areas such as portfolio selection, option evaluation, and risk management. Kroner and Sultan (1993) argued that neglect of either cointegration or time-varying volatility would affect the hedging performance of existing models in the literature for the futures market. Similar conclusion has been given by Ghost (1993) and Lien (1996) through empirical calculation and theoretical analysis respectively. Therefore the traditional error correction model needs to be generalized to have conditional heteroscedasticity for capturing both cointegration and time-varying volatility.

Univariate volatility models have been extended to multivariate cases. Extensions of the generalized autoregressive heteroscedastic (GARCH) model (Bollerslev (1986)) include, for example, vectorized GARCH (VEC-GARCH) model of Bollerslev et al. (1988), the BEKK model of Engle and Kroner (1995) [1], a dynamic conditional correlation (DCC) model of Engle (2002) and Engle and Sheppard (2001), a generalized orthogonal GARCH model of van der Weide (2002); see a survey of multivariate GARCH models by Bauwens, Laurent and Rombouts (2006). These models assume that a vector transformation of the covariance matrix can be written as a linear combination of its lagged values and the innovations. Andersen et al. (1999) showed that these models perform well relatively to competing alternatives. But the curse of dimensionality becomes a major obstacle in application. A useful approach to simplifying the dynamic structure of a multivariate volatility process is to use factor models. As is well known, factor models have been used for performance evaluation and risk measurement in finance. Moreover, it is now widely

---

[1]The early version of this paper was written by Baba, Engle, Kraft and Kroner, which led to the name BEKK of this model.

accepted that financial volatilities move together over time across assets and markets (Anderson et al. (2006)). These make it reasonable that we impose a factor structure on the residual term of a multivariate error correction model. In this sense, an error correction–volatility factor (EC-VF) model can capture the features of comovements in both conditional mean (cointegration) and conditional variance (volatility factors) of a high dimensional time series.

The contribution of this paper is to estimate the EC-VF model. The set of parameters is divided into three subsets: structural parameter set including lag order and all autoregressive coefficient vector and matrices, cointegration parameter set including the cointegration vectors and the rank, and factor parameter set including the factor loading matrix and the number of factors. We conduct a two-step procedure to estimate relevant parameters. First, assuming that the structural and cointegration parameters are known, we give the estimation of factor loading matrix in the volatility factor model, and then give a method to determine the number of factors consistently. Our model specification and estimation approaches are general, because we impose neither i.i.d assumption on the idiosyncratic components in the factor structure nor independence between factors and idiosyncratic errors. In contrast to the innovation expansion method in Pan and Yao (2008) and Pan et al. (2007), where they can not prove that their algorithm for the number of factors is consistent, our method in this paper is based on a penalized goodness-of-fit criterion. We prove our estimator of the number of factors is consistent. Secondly, the structural and cointegration parameters will be consistently estimated without knowing the true factor structure. The main distinction between Bai and Ng (2002) and this paper is that their factor model concerned the unconditional mean of economic variables while our factor structure is imposed on the conditional variance to reduce the dimension of volatilities.

The rest of the paper is organized as follows. Section 2 defines the EC-VF model and mentions some practical backgrounds of the model. Section 3 presents an information criterion for determining the number of factors and the consistency of our estimator. In section 4, a simple Monte Carlo simulation is conducted to check the accuracy of the proposed estimation for the factor loading matrix and the number of factors. In section 5, an application to financial risk management is discussed to show the advantages of the EC-VF model to other traditional alternatives. All theoretical proofs are given in the Appendix.

2 Model

2.1 Definition

Suppose that $\{Y_t\}$ is a $d \times 1$ time series. The error correction–volatility factor (EC-VF) model is of the form

$$\begin{cases} \Delta Y_t = \mu + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \cdots + \Gamma_{k-1} \Delta Y_{t-k+1} + \Gamma_0 Y_{t-1} + Z_t \\ Z_t = AF_t + e_t \end{cases} \qquad (2.1)$$

where $\Delta Y_t = Y_t - Y_{t-1}$, μ is a $d \times 1$ vector, Γ_i, $i = 1, ..., k$, are $d \times d$ matrices. The rank of Γ_0, denoted by m, is called the cointegration rank. $\{Z_t\}$ is strictly stationary with $E(Z_t|\mathcal{F}_{t-1}) = 0$ and $Var(Z_t|\mathcal{F}_{t-1}) = \Sigma_z(t)$, where $\mathcal{F}_t = \sigma(Z_t, Z_{t-1}, \cdots)$. F_t is a $r \times 1$ time series, $r < d$ is unknown, A is a $d \times r$ unknown constant matrix. F_t and e_t are assumed to satisfy

$$\begin{cases} E(F_t|\mathcal{F}_{t-1}) = 0, \quad E(e_t|\mathcal{F}_{t-1}) = 0, \\ E(F_t e_t'|\mathcal{F}_{t-1}) = 0, \quad E(e_t e_t'|\mathcal{F}_{t-1}) = \Sigma_e, \end{cases} \qquad (2.2)$$

where Σ_e is a positive definite matrix and independent of t. The components of F_t are called 'factors', and r is the number of factors. Note that F_t and e_t are conditionally uncorrelated. There is no loss of generality in assuming that $E(F_t F_t')$ is a $r \times r$ positive definite matrix (otherwise, the above model may be expressed equivalently in terms of a smaller number of factors).

Remark 1. The error term $\{Z_t\}$ in an EC-VF model is conditionally heteroscedastic and follows a factor structure, while the error term in the traditional error correction model developed by Engle and Granger (1987) is covariance stationary with mean 0. Here the factor structure is not the classical one because we assume neither that the idiosyncratic components e_t are i.i.d with a diagonal covariance matrix nor that the factor components F_t is independent of e_t.

Model (2.1) assumes that the volatility dynamics of ΔY_t is determined by a lower dimensional volatility dynamics of F_t and the static variation of e_t, as

$$\Sigma_y(t) = \Sigma_z(t) = A\Sigma_f(t)A' + \Sigma_e, \qquad (2.3)$$

where $\Sigma_y(t) = Var(\Delta Y_t|\mathcal{F}_{t-1})$ and $\Sigma_f(t) = Var(F_t|\mathcal{F}_{t-1})$. Without loss of generality, we assume $\text{rank}(A) = r$. The lower dimensional volatility dynamics $\Sigma_f(t)$ can be fitted by, for example, the dynamic conditional correlation model of Engle (2002) or the conditionally uncorrelated components model of Fan, Wang and Yao (2005).

2.2 Practical background

Factor analysis is an effective way for dimension reduction, and then it is an useful statistical tool for modelling multivariate volatility. Because there might exist cointegration relationship among financial asset prices, the framework given by (2.1) applies to many cases of financial analysis.

1. Value-at-Risk

Value-at-Risk defines the maximum expected loss on an investment over a specified horizon at a given confidence level, and is used by many financial institutions as a key measurement of market risk. The Value-at-Risk of a portfolio of multiple assets can be obtained when the prices are described by an EC-VF model. The EC-VF model can be also used to determine an optimal portfolio based on maximizing expected returns subject to a downside risk constraint measured by Value-at-Risk.

2. Hedge ratio

The importance of incorporating the cointegration relationship into statistical modelling of spot and future prices is well documented in the literature for futures market. It has been shown in Lien and Luo (1994) that although GARCH model may characterize the price behavior, the cointegration relationship is the only indispensable component when comparing ex-post performance of various hedge strategies. A hedger who omits the cointegration relationship will adopt a smaller than optimal futures position, which results in a relatively poor hedge performance; see Lien and Tse (2002) for a survey on hedging and references there.

3. Multi-factor option

A multi-factor option (or multi-asset option) is an option whose payoff depends upon the performance of two or more underlying assets. Basket and rainbow options belong to this category. Duan and Pliska (2004) investigated theoretical and practical aspects of such options when the multiple underlying assets are co-integrated. In particular, they proposed an error correction model with stochastic volatilities that follow a multivariate GARCH process. To avoid introducing too many parameters, they give a parsimonious diagonal model for the volatilities, but it is rather restrictive for the cross-dynamics. In contrast, volatility factor models can be used for reducing dimension as well as for representing the dynamics of both variances and covariances. The EC-VF model, with some modification, is more suitable for valuating the multi-factor options.

3 Estimation of the number of factors

The parameter set of the EC-VF model (2.1) is $\{\Theta; \Gamma_0; A\}$, in which $\Theta = \{\mu, \Gamma_1, \cdots, \Gamma_{k-1}\}$ is called the structural parameter, Γ_0 the cointegration parameter and A the factor parameter. In first two subsections, $\{\Theta, \Gamma_0\}$ is assumed known and its determination will be discussed later in subsection 3.3.

3.1 Determining A

Note that the factor loading matrix A and the vector of factors F_t in (2.1) are not separately identifiable. Our goal is to determine the rank of A and the space spanned by the columns of A. Without loss of generality, we may assume $A'A = I_r$, where I_r denotes the $r \times r$ identity matrix. Let $\mathcal{M}(A)$ be the linear subspace of R^d spanned by the columns of A, which is called the factor loading space. Then we need to estimate $\mathcal{M}(A)$ or its orthogonal complement $\mathcal{M}(B)$, where B is a $d \times (d-r)$ matrix for which (A, B) forms a $d \times d$ orthogonal matrix, i.e. $B'A = 0$ and $B'B = I_{d-r}$. Now it follows from (2.1) that

$$B'Z_t = B'e_t. \tag{3.1}$$

From (3.1) and the assumption that $\{e_t\}$ is a conditional homoscedastic sequence of martingale differences (see (2.2)), we have

$$E(B'Z_t Z_t' B | \mathcal{F}_{t-1}) = B'\Sigma_e B = B'\Sigma_z B,$$

where $\Sigma_z = E(Z_t Z_t')$. This implies that

$$B'E(Z_t Z_t' - \Sigma_e)I(Z_{t-\tau} \in C)B = 0 \quad \text{for any } \tau \geq 1 \text{ and } C \in \mathcal{B}, \tag{3.2}$$

or equivalently

$$\sup_{C \in \mathcal{B}} \|B'E[(Z_t Z_t' - \Sigma_e)I(Z_{t-\tau} \in C)]B\| = 0 \quad \text{for any } \tau \geq 1 \text{ and } C \in \mathcal{B}, \tag{3.3}$$

where $\mathcal{B}$ consists of some subsets in R^d, and $\|M\| = [\text{tr}(M'M)]^{1/2}$ denotes the norm of matrix M. Hence we may estimate B by minimizing

$$\Phi_n(B) = \sup_{1 \leq \tau \leq \tau_0, C \in \mathcal{B}} \|B' \frac{1}{n - \tau_0} \sum_{t=\tau_0+1}^{n} (Z_t Z_t' - \hat{\Sigma}_z)I(Z_{t-\tau} \in C)B\| \tag{3.4}$$

subject to the condition $B'B = I_{d-r}$, where τ_0 is a prescribed positive integer and $\hat{\Sigma}_z = \frac{1}{n-\tau_0}\sum_{t=\tau_0+1}^{n} Z_t Z_t'$. This is a high-dimensional optimization problem, but it does not explicitly address the issue how to determine the number of factors r consistently. We first assume r is

6

known and introduce some properties of the estimator of B derived by Pan et al. (2007) before we present a consistent estimator of r.

Let $\mathcal{H}^r$ be the set of all $d \times (d-r)$ $(d \geq r)$ matrix B satisfying $B'B = I_{d-r}$. We partition $\mathcal{H}^r$ into equivalent classes such that $B_1, B_2 \in \mathcal{H}^r$ belong to the same class if and only if $\mathcal{M}(B_1) = \mathcal{M}(B_2)$, which is equivalent to

$$(I_d - B_1 B_1')B_2 = 0 \text{ and } (I_d - B_2 B_2')B_1 = 0. \tag{3.5}$$

Define

$$D(B_1, B_2) = \|(I_d - B_1 B_1')B_2\|.$$

The equivalent classes can be regarded as the elements of the quotient space $\mathcal{H}_D^r = \mathcal{H}^r/D$ defined by D-distance. It can be shown that D is a well-defined metric distance on the space $\mathcal{H}_D^r$, and thus $(\mathcal{H}_D^r, D)$, which is our parametric space, is a metric space; see Pan and Yao (2008).

Our estimator of B is the minimizer of $\Phi_n(\cdot)$ in $\mathcal{H}_D^r$, i.e.

$$\hat{B} = \arg \min_{B \in \mathcal{H}_D^r} \Phi_n(B).$$

Under the assumptions listed below, the estimator $\hat{B}$ is consistent with a convergence rate $\sqrt{n}$.

Assumption A. $\{Z_t\}$ is a strictly stationary d-dimensional time series with $E\|Z_t\|^{2p} < \infty$ for some $p > 2$. The β-mixing coefficients

$$\beta(n) = E\{ \sup_{B \in \mathcal{F}_n^\infty} |P(B) - P(B|\mathcal{F}_{-\infty}^0)| \}$$

satisfy $\beta_n = O(n^{-b})$ for some $b > \frac{p}{p-2}$, where $\mathcal{F}_i^j$ is the σ-algebra generated by $\{Z_t, i \leq t \leq j\}$.

Assumption B. Denote $\Phi(B) = \sup_{1 \leq \tau \leq \tau_0, C \in \mathcal{B}} \|BE[(Z_t Z_t' - \Sigma_e)I(Z_{t-\tau} \in C)]B\|$. There exists a matrix $B_0 \in \mathcal{H}^r$ which minimizes $\Phi(B)$, and $\Phi(B)$ reaches its minimum value at a matrix $B \in \mathcal{H}^r$ if and only if $D(B, B_0) = 0$.

Assumption C. There exists a positive constant a such that $\Phi(B) - \Phi(B_0) \geq aD(B, B_0)$ for any matrix $B \in \mathcal{H}^r$.

By the similar way to that in proof of Theorem 2 in Pan et al. (2007), we can prove the following result, which is useful in deriving a consistent estimator for the number of factors in next subsection.

Theorem 3.1. *If the collection $\mathcal{B}$ of subsets in R^d is a VC-class[2], and Assumptions A and B hold, then*

$$\sup_{B \in \mathcal{H}_D} \sqrt{n}|\Phi_n(B) - \Phi(B)| = O_p(1) \tag{3.6}$$

If, in addition, Assumption C also holds,

$$\sqrt{n}D(\hat{B}, B_0) = O_p(1). \tag{3.7}$$

3.2   Determining r

Let r_0 be the true number of factors and A_0 the true factor loading matrix with rank r_0. We discuss how to estimate r_0 based on the estimated factor loading matrix $\hat{A}$ (or its counterpart $\hat{B}$) derived in the previous subsection. The basic idea is to treat the number of factors as the "order" of model (2.1) and to determine the order in terms of an appropriate information criterion.

In the following, we always assume that Assumptions A-C hold. Let M^l denote a matrix with rank $d - l$. In particular, $B_0^{r_0}$ and $\hat{B}^r$ ($0 \leq r \leq d$) denote the matrices B_0 and $\hat{B}$ with ranks $d - r_0$ and $d - r$ respectively.

Let

$$\Phi_n(r, \hat{B}^r) = \sup_{1 \leq \tau \leq \tau_0, C \in \mathcal{B}} \|\hat{B}^{r'} \hat{D}_{n,\tau}(C)\hat{B}^r\|, \tag{3.8}$$

$$\Phi(r, B_0^r) = \sup_{1 \leq \tau \leq \tau_0, C \in \mathcal{B}} \|B_0^{r'} D_\tau(C)B_0^r\|,$$

where

$$\hat{D}_{n,\tau}(C) = \frac{1}{n - \tau_0} \sum_{t=\tau_0+1}^{n} (Z_t Z_t' - \hat{\Sigma}_z)I(Z_{t-\tau} \in C),$$

$$D_\tau(C) = E[(Z_t Z_t' - \Sigma_e)I(Z_{t-\tau} \in C)],$$

$$\hat{B}^r = \arg\min_{B \in \mathcal{H}_D^r} \Phi_n(r, B), \qquad B_0^r = \arg\min_{B \in \mathcal{H}_D^r} \Phi(r, B).$$

Our penalized goodness-of-fit criterion is defined as

$$PC(r) = \Phi_n(r, \hat{B}^r) + rg(n), \tag{3.9}$$

where $g(n)$ is a penalty for "overfitting". We may estimate r_0 by minimizing $PC(r)$, i.e.

$$\hat{r} = \arg\min_{0 \leq r \leq d} PC(r).$$

---

[2]The definition of Vapnik-Červonenkis (VC) class can be found in van der Vaart and Wellner (1996).

We call (3.9) a penalized goodness-of-fit criterion because of Lemma A.1.

Remark 2. $\Phi_n(\cdot)$ can be regarded as fitting error, because a model with $r + 1$ factors can fit no worse than a model with r factors, while Lemma A.1 shows that $\Phi_n(\cdot)$ is a non-increasing function of r. But the efficiency is lost as more factors are estimated. For example, there is neither error nor efficiency in the extreme case when $r = d$, $\Phi_n(d, \hat{B}^d) = 0$ with $\hat{B}^d = 0$.

The following theorem shows that $\hat{r}$ is a consistent estimator of r_0 provided that the penalty function $g(n)$ satisfies some mild conditions. Then, the problem left in Pan and Yao (2007) is solved.

Theorem 3.2. *Under assumptions A-C, as* $n \to \infty$, $\hat{r} \overset{P}{\to} r_0$ *provided that* $g(n) \to 0$ *and* $\sqrt{n}g(n) \to \infty$.

3.3 Determining $\{\Theta, \Gamma_0\}$

In this subsection, we give an estimation of the structural and cointegration parameter sets without knowledge of the true factor structure for Z_t. By the Grange representation theorem, if there are exactly m cointegration relations among the components of Y_t, and Γ_0 admits the decomposition $\Gamma_0 = \gamma\alpha'$, then α is a $d \times m$ matrix with linearly independent columns and $\alpha'Y_t$ is stationary. In this sense, α consists of m cointegration vectors. Since α and γ are not separately identifiable, our goal is to determine the rank of α, i.e. the dimension of the space spanned by the columns of α. Besides Assumptions A-C on $\{Z_t\}$, we need an additional assumption on $\{Y_t\}$ as follows.

Assumption D. The process Y_t satisfies the basic assumptions of the Granger representation theorem given by Engle and Granger (1987), and $E\|\alpha'Y_{t-1}\|^4 < \infty$.

Our estimation of cointegration vectors is the solution to the following optimization problem

$$\max_{\alpha'S_{11}\alpha=I_m} \mathrm{tr}(\alpha'S_{10}S_{01}\alpha), \tag{3.10}$$

where $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{it}R_{jt}'$, $R_{0t} = \Delta Y_t - \Theta_1 X_t$, $R_{1t} = Y_{t-1} - \Theta_2 X_t$, $X_t = (1, \Delta Y_{t-1}', \ldots, \Delta Y_{t-k+1}')'$, $\Theta_1 = \sum_{t=1}^{T} \Delta Y_t X_t'(\sum_{t=1}^{T} X_t X_t')^{-1}, \Theta_2 = \sum_{t=1}^{T} Y_{t-1} X_t'(\sum_{t=1}^{T} X_t X_t')^{-1}$. The solution of (3.10) is $\hat{\alpha} \equiv (\hat{\alpha}_1, \cdots, \hat{\alpha}_m)$, where $\hat{\alpha}_1, \cdots, \hat{\alpha}_m$ are the m generalized eigenvectors of $S_{10}S_{01}$ with respect to S_{11} corresponding to the m largest generalized eigenvalues.

The estimated cointegration vectors are consistent with the standard root-n convergence rate. The corresponding estimator $\hat{\gamma} = S_{01}\hat{\alpha}$ of the cointegration loading matrix and the estimator

$\hat{\Theta} = \Theta_1 - \hat{\gamma}\hat{\alpha}'\Theta_2$ of the structural parameter are also consistent. These conclusions are obtained by Li, Pan and Yao (2006), who also give a joint estimation for the cointegration rank and the lag order of the error correction model by a penalized goodness-of-fit measure

$$\mathbb{M}(m, k) = \mathbb{R}(m, k, \hat{\alpha}) + n_{m,k}g_1(n), \tag{3.11}$$

where

$$\mathbb{R}(m, k, \hat{\alpha}) = \text{tr}(S_{00} - S_{01}\hat{\alpha}(\hat{\alpha}'S_{11}\hat{\alpha})^{-1}\hat{\alpha}'S_{10}), \tag{3.12}$$

$g_1(n)$ is the penalty for "overfitting" and $n_{m,k}$ is the number of free parameters. Note that $n_{m,k} = d + d^2(k-1) + 2dm - m^2$ for model (2.1). We may estimate m_0 by minimizing $\mathbb{M}(m, k)$, i.e.

$$(\hat{m}, \hat{k}) = \arg \min_{0 \leq m \leq d, 1 \leq k \leq K} \mathbb{M}(m, k).$$

where K is a prescribed positive integer. Let k_0 be the true lag order. The theorem below ensures that $(\hat{m}, \hat{k})$ is a consistent estimator for (m_0, k_0).

Theorem 3.3. *Under assumptions A-D, as $n \to \infty$, $(\hat{m}, \hat{k}) \xrightarrow{P} (r_0, k_0)$ provided that $g_1(n) \to 0$ and $ng_1(n) \to \infty$.*

In practice, the choice of penalty function $g(\cdot)$ is flexible, e.g. $\ln(n)/\sqrt{n}$ or $2\ln(\ln(n))/\sqrt{n}$.

4   Monte Carlo simulation

We present a simple Monte Carlo experiment to illustrate the proposed approach in this section. Particularly we check the accuracy of our estimation for the factor loading matrix A and the number of factors r.

Consider a simple EC-VF model with $d = 6$, $m = 1$, $r = 1$,

$$\begin{cases} \Delta Y_t = \mu + \gamma\alpha'Y_{t-1} + Z_t, \\ Z_t = AF_t + e_t, \\ F_t|\mathcal{F}_{t-1} \sim N(0, \sigma_t^2), \quad e_t|\mathcal{F}_{t-1} \sim N(0, I_6), \end{cases} \tag{4.1}$$

where $\sigma_t^2 = \beta_0 + \beta_1 F_{t-1}^2 + \beta_2 \sigma_{t-1}^2$, e_t is independent of F_t, and the values of parameters are given as follows: $\mu = (0.2028, 0.1987, 0.6038, 0.2722, 0.1988, 0.0153)'$, $\gamma = (0.1, 0.2, 0.3, 0.4, 0.5, 0.6)'$, $\alpha = (1, 2, -1, -1, -2, 3)'$, $A = (\frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6})'$ and $\beta = (\beta_0, \beta_1, \beta_2)' = (0.02, 0.10, 0.76)'$.

Note that $A'A = 1$. We conduct 2000 replications, and for each replication, the sample sizes are $n = 500$ and 1000 respectively. We estimate the transformation matrix B by minimizing $\Phi_n(B)$ defined by (3.4), and measure the estimation error of the factor loading space $\mathcal{M}(A)$ by

$$D_1(A, \hat{A}) = ([\text{tr}\{\hat{A}'(I_d - AA')\hat{A}\} + \text{tr}(\hat{B}'AA'\hat{B})]/d)^{1/2}.$$

The coefficients $\beta_i, i = 0, 1, 2$, are estimated by quasi-MLE based on a Gaussian likelihood. The resulting estimates are summarized in Table 1.

The mean of estimation errors $D_1(A, \hat{A})$ is less than 0.06 while it decreases over 15% as the sample size increases from 500 to 1000. The negative biases indicate a slight underestimation for the heteroscedastic coefficients. The relative frequencies for $\hat{r}$ taking different values are listed in Table 2. It shows that when the sample size n increases, the estimation of r becomes more accurate.

5   An application

The Value-at-Risk (VaR) is widely adopted by banks and other financial institutions to measure and manage market risk, as it reflects downside risk of a given portfolio or investment. Specifically, at a given confidence level $1 - a$, the VaR of a portfolio with weight ω_t is defined as the solution to

$$P(\omega_t \Delta Y_t < VaR_a | \mathcal{F}_{t-1}) = a, \tag{5.1}$$

where ΔY_t is a vector of log returns of assets in the portfolio. In the case when the conditional density $f(\Delta Y_t | \mathcal{F}_{t-1})$ is normal, (5.1) reduces to the well known formula

$$VaR_a = \omega_t \mu_y(t) + (\omega_t' \Sigma_y(t)\omega_t)^{1/2} z_a, \tag{5.2}$$

where z_a is the a-th quantile of the univariate standard normal distribution.

In this section, we attempt to compare the VaR forecasting results by assuming three different models: AR-DCC, EC-DCC, EC-VF-DCC for the asset price series $\{Y_t\}$. The DCC refers to dynamic conditional correlation, a volatility model proposed by Engle (2002). Focusing on the methodology, we only consider the case when the conditional multivariate density $f(\Delta Y_t | \mathcal{F}_{t-1})$ is normal, while the impact of other distributions (like Student-t and some nonparametric densities) on VaR computation is beyond our scope here.

5.1 Data set and estimation of the EC-VF-DCC model

Our data set consists of 2263 daily log prices of CSCO, DELL, INTC, MSFT and ORCL, the five most active stocks in US market, from June 19, 1997 to June 16, 2006. The plots of log returns (in percentage) are presented in Figure 1 which shows significant time-varying volatilities. Descriptive statistics are listed in Table 3. All unconditional distributions of these series exhibit excessive kurtosis and nonzero skewness, indicating significant departure from the normal distribution.

The estimation procedure for the EC-VF-DCC model is given step by step as follows.

Step 1. Fit an error correction model for Y_t to determine the structural and cointegration parameters. Compute the estimate of conditional mean vector $\hat{\mu}_y(t) = \hat{\Theta}X_t + \hat{\gamma}\hat{\alpha}'Y_{t-1}$.

Step 2. Conduct a multivariate portmanteau test for the squared residuals obtained from the previous step to detect conditional heteroscedasticity. If there exists serial dependence, fit a volatility factor model for the residual series $\{Z_t\}$ to determine the factor loading matrix $\hat{A}$, otherwise switch to Step 3 with $\hat{A} = I_r$ and $r = d$.

Denote $B = (b_1, b_2, \cdots, b_{d-r})$, the objective function (3.4) can be modified to

$$\Psi_n(B) = \sum_{\tau=1}^{\tau_0} \sum_{C \in \mathcal{B}} w(C)\|B'\frac{1}{n-\tau_0}\sum_{t=\tau_0+1}^{n}(Z_t Z_t' - \hat{\Sigma}_z)I(Z_{t-\tau} \in C)B\|^2$$

where $w(C) \geq 0$ are weights which ensure that the sum over $C \in \mathcal{B}$ converges. In numerical implementation, we simply take $\mathcal{B}$ as the collection of all the balls centered at the origin in R^d and $w(C) = \{\#(\mathcal{B})\}^{-1}$.

An algorithm for estimating B and r is given as follows. Put

$$\Psi(b) = \sum_{\tau=1}^{\tau_0} \tilde{\Phi}_\tau(b), \quad \tilde{\Phi}_\tau(b) = \sum_{C \in \mathcal{B}} w(C)[b'\frac{1}{n-\tau_0}\sum_{t=\tau_0+1}^{n}(Z_t Z_t' - \hat{\Sigma}_z)I(Z_{t-\tau} \in C)b]^2,$$

$$\Psi_l(b) = \sum_{\tau=1}^{\tau_0}\left\{\sum_{i=1}^{l-1}\sum_{C \in \mathcal{B}} w(C)[\hat{b}_i'\frac{1}{n-\tau_0}\sum_{t=\tau_0+1}^{n}(Z_t Z_t' - \hat{\Sigma}_z)I(Z_{t-\tau} \in C)b]^2 + \tilde{\Phi}_\tau(b)\right\}.$$

Compute $\hat{b}_1$ by minimizing $\Psi(b)$ subject to the constraint $b'b = 1$. For $l = 2, \cdots, d$, compute $\hat{b}_l$ which minimizes $\Psi_l(b)$ subject to the constraint $b'b = 1$, $b'\hat{b}_i = 0$ for $i = 1, 2, \cdots, l-1$. Let $\hat{r} = \arg\min_{0 \leq r \leq d} PC(r)$ with $\hat{B}^r = (\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_r)$, where $PC(r)$ is defined by (3.9). Note that $\hat{B}^{r'}\hat{B}^r = I_{d-\hat{r}}$. Let $\hat{A}$ consist of the $\hat{r}$ (orthogonal) unit eigenvectors, corresponding to the common eigenvalue 1, of matrix $I_d - \hat{B}^r\hat{B}^{r'}$.

Step 3. Fit a dynamic conditional correlation (DCC) volatility model (Engle (2002)) for $\{\hat{A}'Z_t\}$ and compute its conditional covariance $\tilde{\Sigma}_z(t) = D_t^{1/2} R_t D_t^{1/2}$.

To this end, we first fit each element of D_t with a univariate GARCH(1,1) model using the i-th component of $\hat{A}'Z_t$ only, and then model the conditional correlation matrix R_t by

$$R_t = S(1 - \theta_1 - \theta_2) + \theta_1(\varepsilon_{t-1}\varepsilon'_{t-1}) + \theta_2 R_{t-1},$$

where ε_t is a $\hat{r} \times 1$ vector of the standardized residuals obtained from the separate GARCH(1,1) fittings for the $\hat{r}$ components of $\hat{A}'Z_t$, and S is the sample correlation matrix of $\hat{A}'Z_t$.

If $\hat{A} = I_d$, the estimate of conditional covariance matrix $\hat{\Sigma}_y(t)$ of ΔY_t is equal to $\tilde{\Sigma}_z(t)$ and terminate the algorithm. Otherwise proceed to Step 4.

Step 4. The factor structure in (2.1) and the facts $B'A = 0$, $B'e_t = B'Z_t$, $AA' + BB' = I_d$ lead to a dynamics for $\Sigma_y(t) \equiv \Sigma_z(t)$ as follows

$$\hat{\Sigma}_y(t) = \hat{A}\tilde{\Sigma}_z(t)\hat{A}' + \hat{A}\hat{A}'\hat{\Sigma}_z\hat{B}\hat{B}' + \hat{B}\hat{B}'\hat{\Sigma}_z, \tag{5.3}$$

where $\hat{\Sigma}_z = \frac{1}{n-\tau_0}\sum_{t=\tau_0+1}^{n} Z_t Z_t'$.

We determine the cointegration rank by minimizing $\mathbb{M}(m,k)$ defined by (3.11). The surface of $\mathbb{M}(m,k)$ is plotted against m and k in Figure 2. The minimum point of the surface is attained at $(m,k) = (1,1)$, leading to an error correction model for this data set with lag order 1 and cointegration rank 1. Applying the Ljung-Box statistics to the squared residuals, we have $Q_5(1) = 63.2724$, $Q_5(5) = 305.7613$ and $Q_5(10) = 633.7103$. Based on asymptotic χ_2 distributions with degrees of freedom 11, 111 and 236,[3] the p-values of these Q statistics are all close to zero. Consequently, the portmanteau test confirms the existence of conditional heteroscedasticity. The algorithm stated in Step 2 leads to an estimator for the number of factors, and $PC(r)$ is plotted against r in Figure 3. Clearly, a two-factor structure (i.e. $\hat{r} = 2$) is determined for the residual series $\{Z_t\}$.

5.2 Comparison of Value-at-Risk forecasting results

The VaRs are computed at level 0.05 (denoted by $VaR_{0.05}$) for the last 1000 trading days of data span. We assume three models: AR-DCC, EC-DCC, EC-VF-DCC for the asset prices

---

[3]The $Q_d(l)$ statistic has asymptotically a χ^2 distribution with degree of freedom $d^2 l - n_{m,k}$ where $n_{m,k} = d + d^2(k-1) + 2dm - m^2$ is the number of free parameters in the error correction model.

$\{Y_t\}$, and four time invariant portfolios with weights $\omega^1 = (1,1,1,1,1)'/5$, $\omega^2 = (1,2,3,4,5)'/15$, $\omega^3 = (5,4,3,2,1)'/15$, $\omega^4 = (1,3,5,4,2)'/15$. To compare the VaR forecasting performances, we calculate failure rates for the different specifications. The failure rate is defined as the proportion of $r_t = \omega_t' \Delta Y_t$ smaller than the VaRs. For a correctly specified model, the empirical failure rate is supposed to be close to the true level a. Tables 4 display the results for the five percent level.

We observe from table 4 that the EC-VF-DCC performs reasonably well, while AR-DCC has a difficulty in providing failure rates close to 0.05. The empirical failure rates for AR-DCC are high, which means that it underestimates the risk. The results for the EC-DCC and EC-VF-DCC model are comparable, but the average computing time for EC-DCC is much longer, see the last column of table 4. This shows that the factor structure imposed on the residual term of an error correction model can improve the computational velocity in high-dimensional problems.

The above results show that the EC-VF model proposed in this paper is a promising tool for risk analysis. First, it incorporate the impact of cointegration which makes the VaR computation more accurate. Secondly, it deduces a high-dimensional optimization problem into a much lower-dimensional problem, thus accelerates the VaR computation to a great extent.

ACKNOWLEDGMENTS

REFERENCES

Andersen, T. G., T. Bollerslev, F. X. Diebold and P. Labys (1999). Realized volatility and correlation. *http://www.ssc.upenn.edu/ fdiebold/papers/paper29/temp.pdf*

Anderson, H. M., J. V. Issler and F. Vahid (2006). Common features. *Journal of Econometrics* **132**, 1-5.

Bai, J. S. and N. Serena (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**, 191-211.

Bauwens, L., S. Laurent and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* **21**, 79-109

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

Bollerslev, T., R. F. Engle and J. M. Wooldridge (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy* **96**, 116-131.

Duan, J. C. and S. R. Pliska (2004). Option valuation with co-integrated asset prices. *Journal of Economic Dynamics & Control* **28**, 727-754

Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* **20**, 339-350.

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

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

Engle, R. F and K. Sheppard (2001). Theoretical and empirical properties of dynamic conditionalcorrelation multivariate GARCH. *A preprint.*

Fan, J., M. Wang and Q. Yao (2005). Modelling multivariate volatilities via conditionally uncorrelated components. *J. Roy. Statist. Soc. B*, accepted.

Ghost, A. (1993). Hedging with stock index futures: estimation and forecasting with error correction model. *The Journal of Futures Markets* **13**,743-752

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 A. A. Weiss (1983). Time series analysis of error correction models, in *Studies in Econometrics, Time Series and Multivariate Statistics*, ed. by S. Karlin, T. Amemiya and L.A. Goodman. New York: Academic Press, 255-278.

Kroner, K. and J. Sultan (1993). Time-varying distributions and dynamic hedging with foreign currency futures. *Journal of Financial and Quantitative Analysis* **28**, 535-551.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* **2**, 173-84.

Li, Q., J. Pan and Q. Yao (2006). On determination of cointegration rank. *A preprint.*

Lien, D. and X. Luo (1994). Multiperiod hedging in the presence of conditional heteroscedasticity. *The Journal of Futures Markets* **14**, 927-955.

Lien, D. and Y. K. Tse (2002). Some recent developments in futures hedging. *Journal of Economic Surveys* **16**, 357-396.

Lien, D. (1996). The effect of the cointegration relationship on futures hedging: a note. *The Journal of Futures Markets* **16**, 773-780.

Pan, J. and Q. Yao (2007). Modelling multiple time series via common factors. *Biometrika*, **95**, 365-379.

Pan, J., D. Peña, W. Polonik and Q. Yao (2007). Modelling multivariate volatilities by common factors. *Preprint*

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

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

van der Weide, R. (2002). Go-GRACH: a multivariate generalized orthogonal GRACH model. *Journal of Applied Econometrics* **17**, 549-564.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

<center>APPENDIX</center>

<center>*Proofs*</center>

The first lemma shows the $\Phi_n(r, \hat{B}^r)$ defined in subsection 3.2 is a non-increasing function of the number of factors r.

Lemma A.1 If $0 \le r_1 < r_2 \le d$, then $\Phi_n(r_1, \hat{B}^{r_1}) \ge \Phi_n(r_2, \hat{B}^{r_2})$.

Proof. For $0 \le r_1 < r_2 \le d$, $\hat{B}^{r_1}$ can be written as $(\tilde{B}^{r_2}, \tilde{B}^{d-(r_2-r_1)})$ where $\tilde{B}^{r_2}$ consists of the first $d - r_2$ columns of the matrix $\hat{B}^{r_1}$. We have

$$
\begin{aligned}
\Phi_n(r_1, \hat{B}^{r_1}) &= \sup_{1 \le \tau \le \tau_0, C \in \mathcal{B}} \|(\tilde{B}^{r_2}, \tilde{B}^{d-(r_2-r_1)})' \hat{D}_{n,\tau}(C)(\tilde{B}^{r_2}, \tilde{B}^{d-(r_2-r_1)})\| \\
&= \sup_{1 \le \tau \le \tau_0, C \in \mathcal{B}} \left\| \begin{pmatrix} \tilde{B}^{r_2'} \hat{D}_{n,\tau}(C) \tilde{B}^{r_2} & \tilde{B}^{r_2'} \hat{D}_{n,\tau}(C) \tilde{B}^{d-(r_2-r_1)} \\ \tilde{B}^{d-(r_2-r_1)'} \hat{D}_{n,\tau}(C) \tilde{B}^{r_2} & \tilde{B}^{d-(r_2-r_1)'} \hat{D}_{n,\tau}(C) \tilde{B}^{d-(r_2-r_1)} \end{pmatrix} \right\| \\
&\ge \sup_{1 \le \tau \le \tau_0, C \in \mathcal{B}} \|\tilde{B}^{r_2'} \hat{D}_{n,\tau}(C) \tilde{B}^{r_2}\| = \Phi_n(r_2, \tilde{B}^{r_2}) \\
&\ge \Phi_n(r_2, \hat{B}^{r_2}).
\end{aligned}
$$

The last inequality holds because $\hat{B}^r$ is the minimizer of $\Phi_n(B)$ in the metric space $(\mathcal{H}_D^r, D)$. □

The proof of Theorem 3.2 needs the following two lemmas.

Lemma A.2 For any fixed r with $r_0 \le r \le d$, there exists a $B \in \mathcal{H}_D^r$ such that $\Phi(r, B) = 0$. For $0 \le r < r_0$, $\Phi(r, B) > 0$ holds for all $B \in \mathcal{H}_D^r$.

Proof. It is clear that $B' A_0 = 0$ implies $\Phi(r, B) = 0$ from the relation between $\Phi(r, B)$ and the factor model with true loading matrix A_0.

For $r = r_0$, there must be a matrix in $\mathcal{H}_D^{r_0}$, denoted by B^{r_0}, such that $B^{r_0'} A_0 = 0$, thus $\Phi(r_0, B^{r_0}) = 0$ and it reaches the minimum value. We have $B^{r_0} = B_0^{r_0}$ in $\mathcal{H}_D^{r_0}$ by Assumption B.

For $r_0 < r \le d$, let $B = B_0^{r_0} H$, where H is an arbitrary $(d - r_0) \times (d - r)$ matrix such that $H' H = I_{d-r}$. Then, $B \in \mathcal{H}_D^r$ and $B' A_0 = 0$. In the other words, $\Phi(r, B_0^{r_0} H) = 0$.

<center>16</center>

For any $B \in \mathcal{H}_D^r$ with $r < r_0$, $B'A_0 \neq 0$. If $\Phi(r, B) = 0$, which means that for any $1 \leq \tau \leq \tau_0$ and any $C \in \mathcal{B}$, $B'D_\tau(C)B = 0$, by choosing $C = \mathcal{R}^d$, we have $B'A_0 E(F_t F_t') A_0' B = 0$. This is impossible because $E(F_t F_t')$ is a positive definite matrix. $\qquad\square$

Lemma A.3 For any $0 \leq r < r_0$, there exists a $\kappa_r > 0$ such that

$$\mathrm{p}\lim_{n\to\infty} [\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0})] \geq \kappa_r,$$

where $\mathrm{p}\lim$ denotes the limit in probability. For any $r_0 \leq r < d$, it holds that

$$\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0}) = O_p(\frac{1}{\sqrt{n}}).$$

Proof. It follows from the definition of $\hat{B}$ that

$$\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0}) \geq \Phi_n(r, \hat{B}^r) - \Phi_n(r_0, B_0^{r_0}).$$

Recall that $\Phi(r_0, B_0^{r_0}) = 0$ by Lemma A.2. Hence,

$$\begin{aligned}
&\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, B_0^{r_0}) \\
=\quad & [\Phi_n(r, \hat{B}^r) - \Phi(r, \hat{B}^r)] - [\Phi_n(r_0, B_0^{r_0}) - \Phi(r_0, B_0^{r_0})] + \Phi(r, \hat{B}^r) \\
=\quad & O_p(\frac{1}{\sqrt{n}}) + \Phi(r, \hat{B}^r) \geq O_p(\frac{1}{\sqrt{n}}) + \Phi(r, B_0^r).
\end{aligned} \tag{A.1}$$

The second equality holds by the similar way to (3.6) with a slight modification that $\hat{B}^r$ is related to n. The last inequality is from the definition of B_0. These imply that, for any $0 \leq r < r_0$,

$$\mathrm{p}\lim_{n\to\infty} [\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0})] \geq \kappa_r := \Phi(r, B_0^r),$$

and from Lemma A.2, $\kappa_r > 0$.

For the second part, since

$$|\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0})| \leq |\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, B_0^{r_0})| + |\Phi_n(r_0, B_0^{r_0}) - \Phi_n(r_0, \hat{B}^{r_0})|$$

$$\leq 2 \max_{r_0 \leq r \leq d} |\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, B_0^{r_0})|,$$

it is sufficient to prove that for any $r_0 \leq r \leq d$,

$$\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, B_0^{r_0}) = O_p(\frac{1}{\sqrt{n}}).$$

Notice that, from (A.1), $\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, B_0^{r_0}) = O_p(\frac{1}{\sqrt{n}}) + \Phi(r, \hat{B}^r)$. Thus we need to prove $\Phi(r, \hat{B}^r) = O_p(\frac{1}{\sqrt{n}})$ for any $r_0 \leq r \leq d$, where

$$\Phi(r, \hat{B}^r) = \sup_{1 \leq \tau \leq \tau_0, C \in \mathcal{B}} \|\hat{B}^{r'} D_\tau(C) \hat{B}^r\|.$$

For an arbitrary $(d - r_0) \times (d - r)$ matrix H such that $H'H = I_{d-r}$, we have

$$\|\hat{B}^{r'} D_\tau(C) \hat{B}^r\|$$
$$= \|(\hat{B}^r - B_0^{r_0} HH' B_0^{r_0'} \hat{B}^r + B_0^{r_0} HH' B_0^{r_0'} \hat{B}^r)' D_\tau(C)(\hat{B}^r - B_0^{r_0} HH' B_0^{r_0'} \hat{B}^r + B_0^{r_0} HH' B_0^{r_0'} \hat{B}^r)\|$$
$$= \|[(I_d - B_0^{r_0} HH' B_0^{r_0'}) \hat{B}^r]' D_\tau(C) \hat{B}^r + (B_0^{r_0} HH' B_0^{r_0'} \hat{B}^r)' D_\tau(C)(I_d - B_0^{r_0} HH' B_0^{r_0'}) \hat{B}^r\|$$

where the last equality holds because the relation $B_0^{r_0'} A_0 = 0$ implies that $B_0^{r_0'} D_\tau(C) B_0^{r_0} = 0$ for any $\tau \geq 1$ and $C \in \mathcal{B}$. Hence,

$$\|\hat{B}^{r'} D_\tau(C) \hat{B}^r\| \leq \|(I_d - B_0^{r_0} HH' B_0^{r_0'}) \hat{B}^r\| \|D_\tau(C)\| (\|\hat{B}^r\| + \|B_0^{r_0} HH' B_0^{r_0'} \hat{B}^r\|)$$
$$= D(\hat{B}^r, B_0^{r_0} H) \|D_\tau(C)\| (\sqrt{d-r} + \|B_0^{r_0} HH' B_0^{r_0'} \hat{B}^r\|)$$
$$\leq D(\hat{B}^r, B_0^{r_0} H) \|D_\tau(C)\| (\sqrt{d-r}(1 + d - r)).$$

Note that $\Phi(r, B_0^{r_0} H) = 0$ by Lemma A.2, that is $D(B_0^{r_0} H, B_0^r) = 0$. Thus $D(\hat{B}^r, B_0^{r_0} H) = O_p(\frac{1}{\sqrt{n}})$. It is easy to see that $\sup_{1 \leq \tau \leq \tau_0, C \in \mathcal{B}} \|D_\tau(C)\| = O_p(1)$. Therefore, $\Phi(r, \hat{B}^r) = O_p(\frac{1}{\sqrt{n}})$. This completes the proof. $\qquad\square$

Proof of Theorem 3.2. The objective is to verify that $\lim_{n \to \infty} P(PC(r) - PC(r_0) < 0) = 0$ for all $0 \leq r \leq d$ and $r \neq r_0$, where

$$PC(r) - PC(r_0) = \Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0}) - (r_0 - r)g(n).$$

For $r < r_0$, if $g(n) \to 0$ as $n \to \infty$,

$$P(PC(r) - PC(r_0) < 0) = P(\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0}) < (r_0 - r)g(n)) \to 0$$

because, by Lemma A.3, $\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0})$ has a positive limit in probability.

For $r > r_0$, Lemma A.3 implies that $\Phi_n(r, \hat{B}^r) - \Phi_n(r_0, \hat{B}^{r_0}) = O_p(\frac{1}{\sqrt{n}})$. Thus, if $\sqrt{n} g(n) \to \infty$ as $n \to \infty$, we have

$$P(PC(r) - PC(r_0) < 0) = P(\Phi_n(r_0, \hat{B}^{r_0}) - \Phi_n(r, \hat{B}^r) > (r - r_0)g(n))$$
$$= P(\sqrt{n}[\Phi_n(r_0, \hat{B}^{r_0}) - \Phi_n(r, \hat{B}^r)] > (r - r_0)\sqrt{n} g(n)) \to 0.$$

The proof of Theorem 3.2 is completed. $\qquad\square$

Table 1: Simulation results: summary statistics of estimation errors

		$D_1(\hat{A}, A)$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
	Mean	0.0563	0.0179	0.0894	0.7414
	Median	0.0438	0.0183	0.0827	0.7521
n=500	STD	0.0601	0.0022	0.0403	0.0935
	Bias	-	-0.0021	-0.0106	-0.0186
	RMSE	-	0.0029	0.0454	0.0958
	Mean	0.0477	0.0193	0.0922	0.7481
	Median	0.0390	0.0199	0.0897	0.7543
n=1000	STD	0.0426	0.0010	0.0276	0.0724
	Bias	-	-0.0007	-0.0078	-0.0119
	RMSE	-	0.0013	0.0295	0.0766

Table 2: Relative frequencies for $\hat{r}$ taking different values, when $r = 1$.

$\hat{r}$	0	1	2	3	4	5	6
n=500	.0120	.8425	.1310	.0105	.0040	0	0
n=1000	.0090	.9765	.0100	.0045	0	0	0

Table 3: Summary statistics of the log-returns

n=2263	CSCO	DELL	INTC	MSFT	ORCL
Mean	0.000423	0.000523	1.95×10^{-5}	0.0002	0.000418
Stdev	0.031847	0.03027	0.030313	0.023074	0.0364
Min	-0.145	-0.20984	-0.24868	-0.16976	-0.34615
Max	0.218239	0.163532	0.183319	0.178983	0.270416
Skewness	0.149215	-0.11826	-0.39156	-0.17347	-0.22637
Kurtosis	4.55802	3.690575	5.63186	5.955046	8.51963

Table 4: Comparison of $VaR_{0.05}$

	ω^1		ω^2		ω^3		ω^4		t (Min)
AR-DCC	0.067	(0.001)	0.071	(0.000)	0.065	(0.005)	0.062	(0.032)	287.3
EC-DCC	0.052	(0.659)	0.059	(0.061)	0.051	(0.713)	0.053	(0.268)	294.7
EC-VF-DCC	0.049	(0.713)	0.056	(0.308)	0.053	(0.268)	0.055	(0.312)	41.5

Figures in parentheses are p-values for the Kupiec likelihood ratio test used to compare the empirical failure rate with its theoretical value, see Kupiec (1995). The average computing time in minute for each model is recorded in the last column.

Figure 1: Plots of daily log-returns (%) of (a)CSCO, (b)DELL, (c)INTC, (d)MSFT and (e)ORCL. Time span is from June 19, 1997 to June 16, 2006 with 2263 observations.
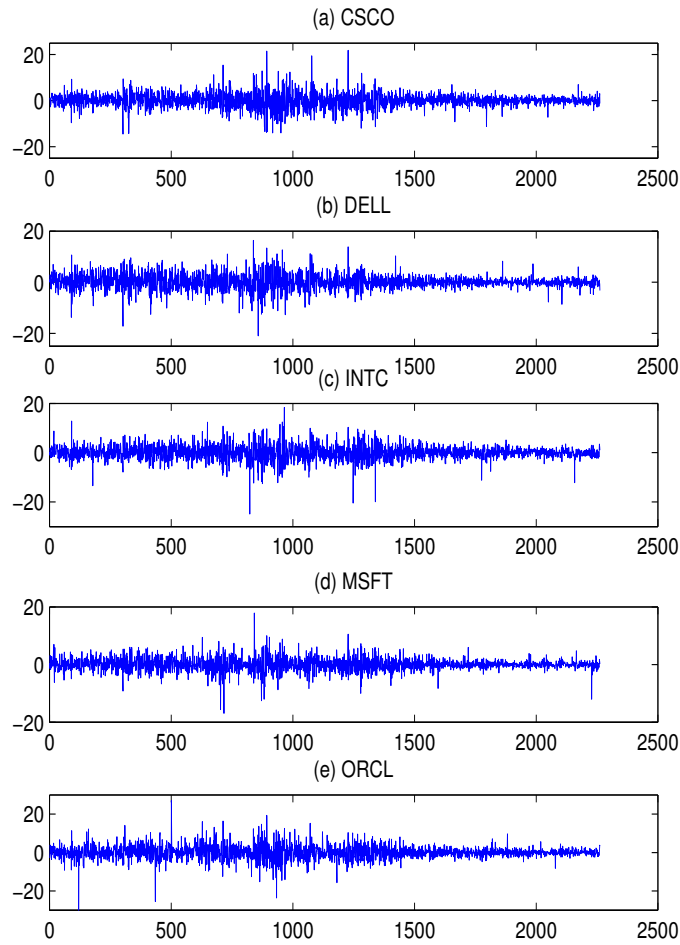
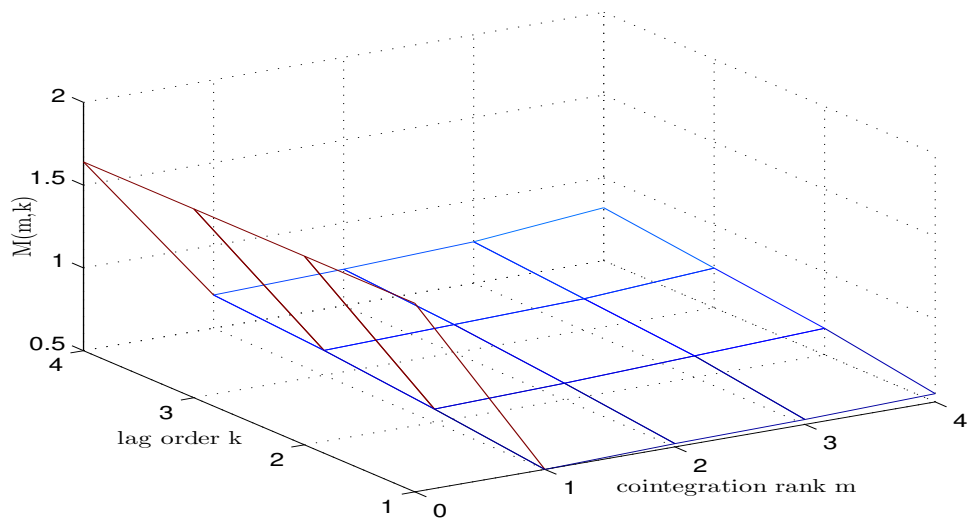Figure 2: Plot of $\mathbb{M}(m, k)$ against the cointegration rank m and the lag order k

Figure 3: Plot of $PC(r)$ against the number of factors r