Estimation of Vector Error Correction Model with GARCH Errors

Koichi Maekawa
Hiroshima University of Economics
Kusdhianto Setiawan
Hiroshima University of Economics and Gadjah Mada University

Abstract
The standard vector error correction (VEC) model assumes the *iid* normal distribution of disturbance term in the model. This paper extends this assumption to include GARCH process. We call this model as VEC-GARCH model. However as the number of parameters in a VEC-GARCH model are large, the maximum likelihood (ML) method is computationally demanding. To overcome these computational difficulties, this paper searches for alternative estimation methods and compares them by Monte Carlo simulation. As a result a feasible generalized least square (FGLS) estimator shows comparable performance to ML estimator. Furthermore an empirical study is presented to see the applicability of the FGLS.

1. Introduction

Vector Error correction (VEC) model is often used in econometric analysis and estimated by maximum likelihood (ML) method under the normality assumption. ML estimator is known as the most efficient estimator under the *iid* normality assumption. However there are disadvantages such that the normality assumption is often violated in real date, especially in financial time series, and that ML estimation is computationally demanding for a large model. Furthermore in our experience of empirical study error terns in VEC model often show a GARCH phenomenon, which violates *iid* assumption. To overcome these disadvantages and to reduce computational burden of ML estimator it may be worthwhile to reconsider the feasible generalized least square (FGLS) estimator instead of ML estimator (MLE) because FGLS method is relatively free from the distributional assumptions and ease computational burden.

The purpose of this paper is to examine the finite sample properties of FGLS estimator in VEC-GARCH model by Monte Carlo simulation and by real data analysis of the international financial time series. The paper is organized as follows: Section 2 briefly surveys the multivariate GARCH (MGARCH hereafter) model; Section 3 describes VEC representation of the vector autoregressive (VAR) model; Section 4 presents a VEC-GARCH model and shows that this model can be estimated by FGLS within the framework of the seemingly unrelated regression (SUR) model; Section 5 examines the performance of FGLS by Monte Carlo simulation; Section 6 presents an empirical application of VEC-GARCH model and shows the applicability of FGLS; finally Section 7 gives some concluding remarks.

2. Multivariate GARCH

Multivariate GARCH model has been developed and applied in financial econometrics and numerous literature were published. The recent development in this area were surveyed by Bauwens, L., S. Laurent and J. V. K. Rombouts (2006) and T. Teräsvirta (2009) . Before investigating MGARCH model in this paper we briefly introduce MGARCH model focusing on relevant MGARCH models in our study.

2.1. vech-GARCH model

The univariate GARCH model has been generalized to *N*-variable multivariate GARCH models in many ways. The most straightforward generalization is the following *vech*-GARCH model by Bollerslev, Engle, and Woodridge (1988):

$$r_t = H_t^{1/2}\eta_t \;\; with\, E(r_t) = 0,\;\; E(\eta_t\eta_t') = I \tag{1}$$

where $r_t = (r_{1t}, \cdots, r_{it}, \cdots, r_{Nt})'$, and r_t is assumed to follows a multivariate normal distribution $N(0, H_t)$. An element of the variance covariance matrix H_t is denoted by h_{ijt} :$H_t = [h_{ijt}]$. In the most general *vech*-GARCH model $vech(H_t)$ is given by

$$vech(H_t) = c + \sum_{j=1}^{q} A_j\, vech(r_{t-j}r_{t-j}') + \sum_{j=1}^{p} B_j\, vech(H_{t-j}) \tag{2}$$

where $vech(\cdot)$ is an operator that vectorizes a symmetric matrix. We briefly illustrate the 2-variable case (*N*=2) for simplicity. For *N*=2 and *p*=*q*=1 $vech(H_t)$ is written as follows:

$$vech\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{pmatrix} = (h_{11,t}, h_{21,t}, h_{22,t})',$$

and *c* is an (*N*(*N*+1)/2)×1=3×1 vector, and A_j and B_j are *N*(*N*+1)/2×*N*(*N*+1)/2 =3×3 parameter matrices. Then $vech(H_t) := h_t$ is written as

$$h_t = \begin{bmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{bmatrix}$$
$$= \begin{bmatrix} c_{01} \\ c_{02} \\ c_{03} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} r_{1,t-1}^2 \\ r_{1,t-1}r_{2,t-1} \\ r_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix} \tag{3}$$

or

$$\begin{aligned} h_{11,t} &= c_{01} + a_{11}r_{1,t-1}^2 + a_{12}r_{1,t-1}r_{2,t-1} + a_{13}r_{2,t-1}^2 \\ &\quad + b_{11}h_{11,t-1} + b_{12}h_{12,t-1} + b_{13}h_{22,t-1} \\ h_{21,t} &= c_{02} + a_{21}r_{1,t-1}^2 + a_{22}r_{1,t-1}r_{2,t-1} + a_{23}r_{2,t-1}^2 \\ &\quad + b_{21}h_{11,t-1} + b_{22}h_{12,t-1} + b_{23}h_{22,t-1} \\ h_{22,t} &= c_{03} + a_{31}r_{1,t-1}^2 + a_{32}r_{1,t-1}r_{2,t-1} + a_{33}r_{2,t-1}^2 \\ &\quad + b_{31}h_{11,t-1} + b_{32}h_{21,t-1} + b_{33}h_{22,t-1} \end{aligned}$$

This representation is very general and flexible but there is a disadvantage that only a sufficient condition for the positive definiteness of the matrix H_t is known. Furthermore the number of parameters is $(p + q)(N(N + 1)/2)^2 + N(N + 1)/2$ which is large unless *N* is small. For example, if $p = q = 1$ and $N = 2$, the number of parameters is 21, if *N*=3 it is 78. This may cause computational difficulties.

2.2. Diagonal vech model

To reduce such disadvantages mentioned above Bollerslev, Engle, and Wooldridge (1988) proposed diagonal *vech* model in which the coefficient matrices A_j and B_j are assumed diagonal. In this case the number of parameters is reduced to $(p+q+1)N(N+1)/2$. For example, if $p=q=1$ and $N=2$ then the number is 9, and if N=3 it is 8. Furthermore in this case the necessary and sufficient conditions for the positive definiteness of H_t are obtained by Bollerslev, Engle, and Nelson (1994). The variance equation (3) is simplified as follows:

$$h_t = \begin{bmatrix} h_{11,t} \\ h_{21,t} \\ h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{01} \\ c_{02} \\ c_{03} \end{bmatrix} + \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} r_{1,t-1}^2 \\ r_{1,t-1}r_{2,t-1} \\ r_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{21t-1} \\ h_{22,t-1} \end{bmatrix}$$

or

$$h_{11,t} = c_{01} + a_{11}r_{1,t-1}^2 + b_{11}h_{11,t-1}$$
$$h_{21,t} = c_{02} + a_{22}r_{1,t-1}r_{2,t-1} + b_{22}h_{21,t-1}$$
$$h_{22,t} = c_{02} + a_{22}r_{2,t-1}^2 + b_{33}h_{22,t-1}$$

2.3. BEKK model

To ensure the positive definiteness of H_t Engle and Kroner (1995) proposed a following model called as Baba-Engle-Kraft-Kroner (BEKK) model.

$$\mathrm{H_t = CC' + \sum_{j=1}^{q} \sum_{k=1}^{K} A'_{kj}\, r_{t-j}r'_{t-j}A_{kj} + \sum_{j=1}^{q} \sum_{k=1}^{K} B'_{kj}H_{t-j}B_{kj}} \tag{4}$$

where A_{kj}, B_{kj}, C are $N \times N$ coefficient matrices, C is a lower triangular matrix. Although this decomposition of the constant term can ensure the positive definiteness of H_t, which is the advantage of this model, the number of parameters is quite large. Because of this, estimation of this model is often infeasible for a large model. When *K*=1 this model is written as

$$H_t = CC' + A'r_{t-1}r'_{t-1}A + B'H_{t-1}B \tag{5}$$

In this case the number of parameters is *np*= $\mathrm{(p+q)N^2 + N(N+1)/2}$. If $p=q=1$ and *N*=2, then $np = 11$, and $np = 24$ for *N*=3. If $N \geq 4$ it may not be feasible to estimate this model. To reduce number of parameters it is common and popular to assume that the coefficient matrices *A, B* are diagonal. This model is called Diagonal BEKK model (Engle and Kroner (1995)). In this model *np*=$(p+q)N + N(N+1)/2$. If $p=q=1$ and *N*=2, then $n\mathrm{p} = 7$, and $np = 12$ for *N*=3. For small size Diagonal BEKK model the calculation is feasible. However, even in Diagonal BEKK model, *np* will be large when *N* is not small. For example, *np*=35 when $p=q=1$ and *N*=5.

We illustrate several versions of (5) for a simple case N=2 and K=1:

Unrestricted BEKK. In this case the variance covariance matrix $\mathrm{H_t} = \begin{bmatrix} \mathrm{h_{11,t}} & \mathrm{h_{12,t}} \\ \mathrm{h_{21,t}} & \mathrm{h_{22,t}} \end{bmatrix}$ is expressed as

$$\mathrm{H_t} = \begin{bmatrix} \mathrm{c_{11}} & 0 \\ \mathrm{c_{21}} & \mathrm{c_{22}} \end{bmatrix} \begin{bmatrix} \mathrm{c_{11}} & \mathrm{c_{21}} \\ 0 & \mathrm{c_{22}} \end{bmatrix} + \begin{bmatrix} \mathrm{a_{11}} & \mathrm{a_{12}} \\ \mathrm{a_{21}} & \mathrm{a_{22}} \end{bmatrix}' \begin{bmatrix} \mathrm{r_{1,t-1}^2} & \mathrm{r_{1,t-1}r_{2,t-1}} \\ \mathrm{r_{2,t-1}r_{1,t-1}} & \mathrm{r_{2,t-1}^2} \end{bmatrix} \begin{bmatrix} \mathrm{a_{11}} & \mathrm{a_{12}} \\ \mathrm{a_{21}} & \mathrm{a_{22}} \end{bmatrix}$$
$$+ \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}' H_{t-1} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

or

$$\begin{aligned} h_{11,t} &= c_{11}^2 + a_{11}^2 r_{1,t-1}^2 + 2a_{11}a_{21}r_{1,t-1}r_{2,t-1} + a_{21}^2 r_{2,t-1}^2 + b_{11}^2 h_{11,t-1} \\ &\quad +2b_{11}b_{21}h_{12,t-1} + b_{21}^2 h_{22,t-1} \\ h_{12,t} &= c_{11}c_{21} + a_{11}a_{12}r_1^2 + (a_{21}a_{12} + a_{11}a_{22})r_1 r_2 + a_{21}a_{12}r_2^2 + b_{11}b_{12}h_{11,t-1} \\ &\quad +(b_{21}b_{12} + b_{11}b_{22})h_{12,t-1} + b_{21}b_{12}h_{22,t-1} \\ h_{22,t} &= c_{12}^2 + c_{22}^2 + a_{12}^2 r_1^2 + 2a_{12}a_{22}r_1 r_2 + a_{22}^2 r_2^2 + b_{12}^2 h_{11,t-1} \\ &\quad +2b_{12}b_{22}h_{12,t-1} + b_{22}^2 r_2^2 \end{aligned}$$

where H_t is positive definite by construction.

Diagonal BEKK is expressed as

$$\mathrm{H_t} = \begin{bmatrix} \mathrm{c_{11}} & 0 \\ 0 & \mathrm{c_{22}} \end{bmatrix} \begin{bmatrix} \mathrm{c_{11}} & 0 \\ 0 & \mathrm{c_{22}} \end{bmatrix} + \begin{bmatrix} \mathrm{a_{11}} & 0 \\ 0 & \mathrm{a_{22}} \end{bmatrix}' \begin{bmatrix} \mathrm{r_{1,t-1}^2} & \mathrm{r_{1,t-1}r_{2,t-1}} \\ \mathrm{r_{2,t-1}r_{1,t-1}} & \mathrm{r_{2,t-1}^2} \end{bmatrix} \begin{bmatrix} \mathrm{a_{11}} & 0 \\ 0 & \mathrm{a_{22}} \end{bmatrix}$$
$$+ \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}' H_{t-1} \begin{bmatrix} b_{11} & 0 \\ 0 & b_{22} \end{bmatrix}$$

or

$$\begin{aligned} h_{11,t} &= c_{11}^2 + a_{11}^2 r_{1,t-1}^2 + b_{11}^2 h_{11,t-1} \\ h_{12,t} &= a_{11}a_{22}r_{1,t-1}r_{2,t-1} + b_{11}b_{22}h_{12,t-1} \\ h_{22,t} &= c_{22}^2 + a_{22}^2 r_{2,t-1}^2 + b_{22}^2 h_{2,t-1} \end{aligned}$$

where $h_{ij,t}$ in these variance covariance equations only depend on their own lagged values $h_{ij,t-1}$.

Engle and Kroner (1995) shows that the diagonal *vech* and the diagonal BEKK are equivalent as follows: By stacking the diagonal elements of A and B of the diagonal vech model, i.e.,

$$\alpha = (a_{11}, a_{22}, a_{33})', \qquad \beta = (b_{11}, b_{22}, b_{33})'$$

and write

$$\Sigma_\mathrm{t} = \mathrm{M} + \alpha\alpha' \odot \mathrm{r_{t-1}r_{t-1}'} + \beta\beta' \odot \Sigma_{\mathrm{t-1}}$$

then it is easy to see that $vech(\Sigma_\mathrm{t})$ is identical to the diagonal *vech*.

There are many other types of multivariate GARCH model. They are surveyed, for example, in Bauwens, L., S. Laurent and J. V. K. Rombouts (2006) and Silvennoinen and Terasvirta (2009).

Bollerslev, Engle, and Wooldridge (1988) introduced a restricted version of the general multivariate *vec* model of GARCH with following representation:

$$\mathrm{H_t} = \Omega + \mathrm{A} \odot \mathrm{r_{t-1}r'_{t-1}} + \mathrm{B} \otimes \mathrm{H_{t-1}}$$

where the operator $\odot$ is the Hadamard product and $\otimes$ is Kronecker Product. To ensure the positive semi-definiteness (PSD) there are several ways for specifying coefficient matrices. One example is to specify Ω, A, and B as products of Cholesky factorized triangular matrices. Such parameterization will be used in the latter section in this paper.

2.4. Log-likelihood function of vech-GARCH

If the distribution of errors η_t is a multivariate normal, then the log-likelihood function of (1) is given by

$$\sum_{t=1}^{T} l_t(\theta) = c - \frac{1}{2}\sum_{t=1}^{T} ln|H_t| - \frac{1}{2}\sum_{t=1}^{T} r'_t H_t^{-1} r_t \tag{6}$$

In calculating MLE we have to invert H_t at every time *t*. This is computationally tedious when *T* and *N* are not small. Furthermore H_t is often noninvertible.

3. VEC representation of VAR model

We consider *M*-variate and *k*-th order vector autoregressive time series Y_t $=[y_{1,t} \cdots y_{i,t} \cdots y_{M,t}]'$

$$\mathrm{Y_t} = \Pi_1 \mathrm{Y_{t-1}} + \cdots + \Pi_k \mathrm{Y_{t-k}} + \epsilon_t \tag{7}$$

This model is called Vector Autoregressive (VAR) Model. The subscript *t* denotes time: $t = 1,2,\cdots,n$. The errors ϵ_t are assumed to follows *iid* *M*-dimensional multivariate normal distribution $N(0, \Sigma)$. Note that Σ does not depend on time *t*. Later in this paper we consider the time dependent case, i.e., Σ_t. Now by introducing a $M \times M$ matrix Π defined by

$$\Pi = I_p - \Pi_1 - \cdots - \Pi_k$$

We can rewrite (7) as

$$\Delta Y_t = C^0 + \Pi Y_{t-1} + \Phi \Delta Y_{t-1} + \varepsilon_t \tag{8}$$

where,

Y_{t-1} $= [y_{1,t-1} \dots y_{i,t-1} \dots y_{M,t-1}]'$: a vector of first order lagged of Y_t.

ΔY_t $= [\Delta y_{1,t}\ \Delta y_{2,t} \dots \Delta y_{i,t} \dots \Delta y_{M,t}]'$: a vector of first difference of Y_t at time *t*.

C^0 $= [c_1^0\ c_2^0 \dots c_i^0 \dots c_M^0]'$: a vector of constant terms.

ε_t $= [\varepsilon_1\ \varepsilon_2\ ...\ \varepsilon_M]'$: a vector of disturbance errors which is assumed *iid* *M*-dimensional multivariate normal distribution *N*(0, Σ).

In what follows we consider a case in which all elements in Y_t are I(1). In this case as the left hand side variables $\Delta \mathrm{Y_t}$ are sationary I(0) the right hand side of (8) should be also stationary. To ensure the stationarity of the right hand side of (8), the rank of the coefficient matrix Π is less than *M* or rank(Π)<*M*. The reason is as follows: if rank(Π)=*M* then there exists Π^{-1} and the equation (8) can be solved for I(1) variable Y_{t-1} as a linear combination of stationary variables $\Delta \mathrm{Y_t}$ and $\Delta \mathrm{Y_{t-1}}$. This is a contradiction. This is because why rank(Π)<M. Under this rank condition Π can be decomposed as follows: $\Pi = AB$

where

A $= [a_1\ a_2\ ...\ a_i\ ...\ a_M]'$: vector of coefficients in cointegrating equation (loading matrix that contains adjustment parameters) and,

B $= [b_1\ b_2\ ...\ b_i\ ...\ b_M]$: a vector of cointegrating coefficient,

Φ $= \begin{bmatrix} \varphi_{11} & \cdots & \varphi_{1M} \\ \vdots & \ddots & \vdots \\ \varphi_{M1} & \cdots & \varphi_{MM} \end{bmatrix}$: a *M* by *M* matrix,

where BY_{t-1} is assured to be stationary (Granger's representation theorem). The stationarity of BY_{t-1} means that a linear combination of elements in Y_{t-1} is stationary, in such elements are called as co-integrated and B is called as co-integration vector. The coefficient matrix *A* is called as loading vector because *A* conveys cointegrated variables to the system.

4. Vector Error correction with GARCH errors (VEC-GARCH model)

4.1. VECM with BEKK errors

So far we have considered the standard Vector Error Correction Model (VECM), where a set of time series is nonstationary at level, but stationary at their first differences and $\varepsilon_t \sim iid\ N(0, \Sigma)$. Matrix Π represents the long run relationship between the variables in Equation (8) and Johansen (1988) proposed a maximum likelihood estimation of (8) for the case of the rank of matrix $\Pi = r$, where $0 < r < M$.

In what follows, we relaxed the assumption of homoscedasticity of the errors. Instead, we assume that ε_t has zero mean and time dependent variance-covariance matrix of H_t that has the BEKK GARCH structure as given by (6):

$$H_t = CC' + A' r_{t-1} r'_{t-1} A + B H_{t-1} B'.$$

4.2. SUR representation

VEC model with GARCH errors can be represented by Seemingly Unrelated Regression (SUR) model as follows. SUR representation of VEC model seems to be worthwhile to consider. For simplicity we consider three-equation VEC model such as:

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

or

$$\begin{bmatrix}\Delta Y_{1,t}\\ \Delta Y_{2,t}\\ \Delta Y_{3,t}\end{bmatrix} = \begin{bmatrix}\pi_{11} & \pi_{12} & \pi_{13}\\ \pi_{21} & \pi_{22} & \pi_{23}\\ \pi_{31} & \pi_{32} & \pi_{33}\end{bmatrix}\begin{bmatrix}Y_{1,t-1}\\ Y_{2,t-1}\\ Y_{3,t-1}\end{bmatrix} + \begin{bmatrix}\phi_{11} & \phi_{12} & \phi_{13}\\ \phi_{21} & \phi_{22} & \phi_{23}\\ \phi_{31} & \phi_{32} & \phi_{33}\end{bmatrix}\begin{bmatrix}\Delta Y_{1,t-1}\\ \Delta Y_{2,t-1}\\ \Delta Y_{3,t-1}\end{bmatrix} + \begin{bmatrix}\varepsilon_{1,t-1}\\ \varepsilon_{2,t-1}\\ \varepsilon_{3,t-1}\end{bmatrix},$$

for t=1, 2, ..., n.

Alternatively this system can be written as

$$\begin{aligned}\Delta Y_{1,\cdot} &= Y_{-1}\Pi_1' + \Delta Y_{-1}\Phi_1' + \varepsilon_1\\ \Delta Y_{2,\cdot} &= Y_{-1}\Pi_2' + \Delta Y_{-1}\Phi_2' + \varepsilon_2\\ \Delta Y_{3,\cdot} &= Y_{-1}\Pi_3' + \Delta Y_{-1}\Phi_3' + \varepsilon_3\end{aligned} \tag{9}$$

where Π_i and Φ_i are the ith row of Π and Φ respectively, i.e.,

$$\Pi = \begin{bmatrix}\pi_{11} & \pi_{12} & \pi_{13}\\ \pi_{21} & \pi_{22} & \pi_{23}\\ \pi_{31} & \pi_{32} & \pi_{33}\end{bmatrix} = \begin{bmatrix}\Pi_1\\ \Pi_2\\ \Pi_3\end{bmatrix},\ \Phi = \begin{bmatrix}\phi_{11} & \phi_{12} & \phi_{13}\\ \phi_{21} & \phi_{22} & \phi_{23}\\ \phi_{31} & \phi_{32} & \phi_{33}\end{bmatrix} = \begin{bmatrix}\Phi_1\\ \Phi_2\\ \Phi_3\end{bmatrix},$$

$$Y_{-1} = \begin{pmatrix}y_{11} & y_{21} & y_{31}\\ \vdots & \vdots & \vdots\\ y_{1,n-1} & y_{2,n-1} & y_{3,n-1}\end{pmatrix}, \quad \Delta Y_{-1} = \begin{pmatrix}\Delta y_{11} & \Delta y_{21} & \Delta y_{31}\\ \vdots & \vdots & \vdots\\ \Delta y_{1,n-1} & \Delta y_{2,n-1} & \Delta y_{3,n-1}\end{pmatrix},$$

$\Delta Y_{i,\cdot} = [\Delta Y_{i,2}, \Delta Y_{i,3}, \dots, \Delta Y_{i,t}, \dots, \Delta Y_{i,n}]'$ and,
$\varepsilon_i = [\varepsilon_{i,2}, \varepsilon_{i,3}, \dots, \varepsilon_{i,t}, \dots, \varepsilon_{i,n}]'$.

Defining new matrices X and β by

$$X = [Y_{-1}, \Delta Y_{-1}] \text{ and } \beta' = [\beta_1', \beta_2', \beta_3'],$$

the 3-equation VEC model (8) can be written as SUR model as follows:

$$\begin{bmatrix}\Delta Y_1\\ \Delta Y_2\\ \Delta Y_3\end{bmatrix} = \begin{bmatrix}X & 0 & 0\\ 0 & X & 0\\ 0 & 0 & X\end{bmatrix}\beta + \begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix}.$$

We assume that $E(\varepsilon) = 0$, $E(\varepsilon_{is}, \varepsilon_{it}) = 0$ for $s \neq t$, and the variance and covariance $E(\varepsilon_{it}^2) = h_{iit}$ and $E(\varepsilon_{it}, \varepsilon_{jt}) = h_{ijt}$ follow MGARCH(1,1). Let us define $\Omega = E(\varepsilon\varepsilon')$, or in the complete form:

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

where, Ω_{ij} is a $n \times n$ diagonal matrix where its main diagonal elements are elements of n-vector of $h_{ij,t}$ and zeros on the off diagonal elements and, $\Omega_{ij} = \Omega_{ji}$, i.e.,

$$
\Omega_{ij} = \begin{bmatrix} h_{ij,1} & & & & \\ & \ddots & & 0 & \\ & & h_{ij,t} & & \\ & 0 & & \ddots & \\ & & & & h_{ij,n} \end{bmatrix}
$$

Thus we have

$$
\Omega = \left[\begin{array}{ccccc|ccccc|ccccc}
h_{11,1} & & & & & h_{12,1} & & & & & h_{13,1} & & & & \\
& \ddots & & 0 & & & \ddots & & 0 & & & \ddots & & 0 & \\
& & h_{11,t} & & & & & h_{12,t} & & & & & h_{13,t} & & \\
& 0 & & \ddots & & & 0 & & \ddots & & & 0 & & \ddots & \\
& & & & h_{11,n} & & & & & h_{12,n} & & & & & h_{13,n} \\
\hline
h_{21,1} & & & & & h_{22,1} & & & & & h_{23,1} & & & & \\
& \ddots & & 0 & & & \ddots & & 0 & & & \ddots & & 0 & \\
& & h_{21,t} & & & & & h_{22,t} & & & & & h_{23,t} & & \\
& 0 & & \ddots & & & 0 & & \ddots & & & 0 & & \ddots & \\
& & & & h_{21,n} & & & & & h_{22,n} & & & & & h_{23,n} \\
\hline
h_{31,1} & & & & & h_{32,1} & & & & & h_{33,1} & & & & \\
& \ddots & & 0 & & & \ddots & & 0 & & & \ddots & & 0 & \\
& & h_{31,t} & & & & & h_{32,t} & & & & & h_{33,t} & & \\
& 0 & & \ddots & & & 0 & & \ddots & & & 0 & & \ddots & \\
& & & & h_{31,n} & & & & & h_{32,n} & & & & & h_{33,n}
\end{array}\right]
$$

where $h_{ij,t}$ follow multivariate MGARCH(1,1) process.

After obtaining an estimate $\widehat{\Omega}$, we have FGLS,

$$
\hat{\beta} = [X'\widehat{\Omega}^{-1}X]^{-1}X'\widehat{\Omega}^{-1}\Delta Y.
$$

Note that inverting a large and sparse matrix Ω often causes computational problems such as memory size, computer time, and inaccurate numerical results. To avoid those problems we propose the following algorithm: After estimating MGARCH process we construct a relatively small matrix $\widehat{H}_t$ and its inverse $\widehat{H}_t^{-1}$ at each time t such that,

$$
\widehat{\mathrm{H}}_t = \begin{bmatrix} \hat{h}_{11,t} & \hat{h}_{12,t} & \hat{h}_{13,t} \\ \hat{h}_{21,t} & \hat{h}_{22,t} & \hat{h}_{23,t} \\ \hat{h}_{31,t} & \hat{h}_{32,t} & \hat{h}_{33,t} \end{bmatrix}, \text{and } \widehat{\mathrm{H}}_t^{-1} = \begin{bmatrix} \hat{h}_t^{11} & \hat{h}_t^{12} & \hat{h}_t^{13} \\ \hat{h}_t^{21} & \hat{h}_t^{22} & \hat{h}_t^{23} \\ \hat{h}_t^{31} & \hat{h}_t^{32} & \hat{h}_t^{33} \end{bmatrix} \tag{10}
$$

where $\widehat{H}_t$ and $\hat{h}_{ij,t}$ are estimated variance covariance of MGARCH.

Replacing $\hat{h}_{ij,t}$ with $\hat{h}_t^{ij}$ in $\hat{\Omega}$ we have easily obtain $\hat{\Omega}^{-1}$ without inverting a large matrix Ω.

5. Monte Carlo Simulation

5.1 Data generating Process (DGP)

Monte Carlo simulation is carried out by generating artificial data of three series. The data generating process (DGP) is repeated for 1000 times. We run the simulation for the number of observations n : 100, 300 and 500. For removing the initial value effect, we generate $2n$ observations for each series and remove the first half of the generated data in each simulation run. The true model for generating the data is specified as follows:

$$Y_t = PY_{t-1} + QY_{t-2} + U_t \tag{11}$$

or in stacked model it can be restated as,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{bmatrix} + \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix} \begin{bmatrix} y_{1t-2} \\ y_{2t-2} \\ y_{3t-2} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

where U_t follow GARCH process, $U_t \sim N(0, H_t)$ and H_t follows the diagonal BEKK:

$$H_t = M^* + \alpha^* \odot \varepsilon'_{t-1}\varepsilon_{t-1} + \beta^* \odot H_{t-1}$$

with

$$\alpha^* = \begin{bmatrix} \alpha_{11}^2 & \alpha_{11}\alpha_{22} & \alpha_{11}\alpha_{33} \\ \alpha_{11}\alpha_{22} & \alpha_{22}^2 & \alpha_{22}\alpha_{33} \\ \alpha_{11}\alpha_{33} & \alpha_{22}\alpha_{33} & \alpha_{33}^2 \end{bmatrix} = \begin{bmatrix} 0.090 & 0.180 & 0.060 \\ 0.180 & 0.360 & 0.120 \\ 0.060 & 0.120 & 0.040 \end{bmatrix}$$

$$\beta^* = \begin{bmatrix} \beta_{11}^2 & \beta_{11}\beta_{22} & \beta_{11}\beta_{33} \\ \beta_{11}\beta_{22} & \beta_{22}^2 & \beta_{22}\beta_{33} \\ \beta_{11}\beta_{33} & \beta_{22}\beta_{33} & \beta_{33}^2 \end{bmatrix} = \begin{bmatrix} 0.090 & 0.150 & 0.120 \\ 0.150 & 0.250 & 0.200 \\ 0.120 & 0.200 & 0.160 \end{bmatrix}$$

$$M^* = \begin{bmatrix} m_{11}^2 & 0 & 0 \\ 0 & m_{22}^2 & 0 \\ 0 & 0 & m_{33}^2 \end{bmatrix} = \begin{bmatrix} 0.025 & 0 & 0 \\ 0 & 0.090 & 0 \\ 0 & 0 & 0.049 \end{bmatrix}$$

α^*, β^* are transformed matrices of $\alpha'\alpha$ and $\beta'\beta$ where where α and β are [0.3,0.6,0.2], [0.3,0.5,0.4] respectively. M* is a transformed matrix of M'M where M is a diagonal matrix with its diagonal elements are [0.5,0.3,0.7]. Equivalently, the variance-covariance equations are as follow:

$$h_{11t} = 0.025 + 0.09u_{1t}^2 + 0.09h_{11,t-1}$$
$$h_{21t} = 0.18u_{1t}u_{2t} + 0.15h_{21,t-1}$$
$$h_{31t} = 0.06u_{3t}u_{1t} + 0.12h_{31,t-1}$$
$$h_{22t} = 0.09 + 0.36u_{2t}^2 + 0.25h_{22,t-1}$$
$$h_{32t} = 0.12u_{3t}u_{2t} + 0.2h_{32,t-1}$$
$$h_{33t} = 0.049 + 0.04u_{3t}^2 + 0.16h_{33,t-1}$$

Equation (11) can be rewritten as Vector Error Correction Model (VECM):

$$\Delta Y_t = \Pi Y_{t-1} + \phi \Delta Y_{t-1} + U_t \tag{12}$$

where $\Pi = P + Q - I$ and $\phi = -Q$. The true values of P and Q are set as follow:

$$P = \begin{bmatrix} 0.7 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0.5 & 0 & 0.4 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0.5 & 1 & 0.1 \end{bmatrix}$$

thus $\Pi = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & -0.5 \end{bmatrix}$ which can be decomposed into loading vector $[0 \quad 0 \quad 1]'$ and cointegrating vector $[1 \quad 1 \quad -0.5]$.

Before we generate Y_t, we have to generate $U_t \sim \mathrm{N}(0, H_t)$ as follows.

Step 1. Generate $\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \end{bmatrix} \sim N(0, I)$
Step 2. Generate H_t using Diagonal BEKK model from ε_t
Step 3. Transform ε_t to U_t by applying Cholesky Decomposition: $U_t = L_t \varepsilon_t$, where L_t is lower triangular matrix obtained from decomposing $H_t = L_t {L_t}'$.

By construction, the positive definiteness (PD) of H_t is assured.

5.2. Estimation Strategy

Under the above DGP we carried out Monte Carlo simulation for the following five cases:

Case 1 (OLS): We estimate parameters equation by equation in equation (9) by OLS without considering GARCH error structure and obtain the followings:

$$\Delta Y_{1,\cdot} = Y_{-1}\widehat{\Pi}_1' + \Delta Y_{-1}\widehat{\Phi}_1' + \hat{u}_1$$
$$\Delta Y_{2,\cdot} = Y_{-1}\widehat{\Pi}_2' + \Delta Y_{-1}\widehat{\Phi}_2' + \hat{u}_2$$
$$\Delta Y_{3,\cdot} = Y_{-1}\widehat{\Pi}_3' + \Delta Y_{-1}\widehat{\Phi}_3' + \hat{u}_3$$

Case 2 (VECM): We estimate parameters in equation (12) by VECM system equation without considering GARCH error structure and obtain the followings:

$$\Delta Y_t = \widehat{\Pi} Y_{t-1} + \widehat{\Phi} \Delta Y_{t-1} + \widehat{U}_t$$

Case 3 (FGLS-OLS-GARCH/FOLSH): First we calculate OLS residuals $\hat{u}_i$ for each equations without considering GARCH error structure as in Case 1. Next, we use $\hat{u}_i$ for obtaining variance covariance matrix $\widehat{H}_t$ and $\widehat{H}_t^{-1}$ in the diagonal BEKK model.

Having $\widehat{H}_t$ and $\widehat{H}_t^{-1}$ in hand we can construct $\widehat{\Omega}$ and $\widehat{\Omega}^{-1}$ to have feasible generalized least square (FGLS) estimator.

Case 4 (FGLS-VECM-GARCH/FVECH): We use VECM system equations as in Case 2 for estimating $\widehat{\Omega}$. First we obtain each residual $\tilde{u}_i$ from VECM in Case 2. Next, we use $\tilde{u}_i$ for obtaining variance covariance matrix $\widehat{H}_t$ and $\widehat{H}_t^{-1}$ in the diagonal BEKK model. Having $\widehat{H}_t$ and $\widehat{H}_t^{-1}$ in hand we can construct $\widehat{\Omega}$ and $\widehat{\Omega}^{-1}$ to have feasible generalized least squre (FGLS) estimator.

Case 5 (MLE): We estimate all parameters in the mean equation (12) and the diagonal BEKK variance equation (5) by MLE and obtain the estimated system as follows:

Mean equation: $\Delta \mathrm{Y}_{\mathrm{t}} = \widehat{\Pi}\mathrm{Y}_{\mathrm{t}-1} + \widehat{\Phi}\Delta\mathrm{Y}_{\mathrm{t}-1} + \widehat{U}_{\mathrm{t}}$

Variance equation: $\widehat{H}_t = \hat{C}\hat{C}' + \hat{A}'\widehat{U}_{t-1}\widehat{U}_{t-1}'\hat{A} + \hat{B}'\widehat{H}_{t-1}\hat{B}$

or equivalently the variance-covariance equations are as follow:

$$
\begin{aligned}
\hat{h}_{11\mathrm{t}} &= \widehat{\mathrm{m}}_{11}^2 + \hat{a}_{11}^2\hat{u}_{1\mathrm{t}-1}^2 + \hat{b}_{11}^2\hat{h}_{11,t-1} \\
\hat{h}_{21\mathrm{t}} &= \hat{a}_{22}\hat{a}_{11}\hat{u}_{2t}\hat{u}_{1t-1} + \hat{b}_{22}\hat{b}_{11}\hat{h}_{21,t-1} \\
\hat{h}_{31\mathrm{t}} &= \hat{a}_{33}\hat{a}_{11}\hat{u}_{3t}\hat{u}_{1t-1} + \hat{b}_{33}\hat{b}_{11}\hat{h}_{31,t-1} \\
\hat{h}_{22\mathrm{t}} &= \widehat{\mathrm{m}}_{22}^2 + \hat{a}_{22}^2\hat{u}_{2\mathrm{t}-1}^2 + \hat{b}_{22}^2\hat{h}_{22,t-1} \\
\hat{h}_{32\mathrm{t}} &= \hat{a}_{33}\hat{a}_{22}\hat{u}_{3t}\hat{u}_{2t-1} + \hat{b}_{33}\hat{b}_{22}\hat{h}_{32,t-1} \\
\hat{h}_{33\mathrm{t}} &= \widehat{\mathrm{m}}_{33}^2 + \hat{a}_{33}^2\hat{u}_{3\mathrm{t}-1}^2 + \hat{b}_{33}^2\hat{h}_{33,t-1}
\end{aligned}
$$

In estimating the parameters we maximize log likelihood function as specified in Equation (6). We run the simulation in Eviews program (version 7.2). For Case 5, in order to starting the iteration, the initial values of VECM parameters (the mean equation) were set based on single OLS equations as in Case 1. Meanwhile, the initial values for MGARCH parameters in the variance equations were set based on univariate GARCH.

5.2. Simulation Results

The main estimation methods under investigation in this paper are FGLS-based estimator (FOLSH and FVECH) and Maximum Likelihood Estimator (MLE). These strategies are taking into account the presence of MGARCH error structure. Presumably, the strategies are expected to outperform the other strategies that neglect the MGARCH error structure (OLS and VECM). Summary of simulation results is presented in Table 1. From the table, we observed that estimation methods FOLSH, FVECH, and MLE seem to outperform the other methods (OLS and VECM); the mean of the estimated parameter from 1000 times simulation run tends to be closer to its true value in most cases.

OLS and VECM under the heteroscedasticity condition still provide us an unbiased estimator, but their standard deviations are larger than the methods that assume MGARCH error structure. Table 2 shows that MLE, FOLSH, and FVECH are more efficient than OLS and VECM. It shows that methods ignoring the MGARCH error structure would result in less efficient estimator. All methods are consistent estimator and the efficiency measured by the Mean Squared Error (MSE) are improving when larger sample size is used.

Table 1 – Parameter Estimates from Monte Carlo Simulation

n=100											
		OLS		VECM		FOLSH		FVECH		MLE	
Parameters	True Value	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
$\hat{\pi}_{11}$	0.000	-0.048	0.082	-0.011	0.074	-0.043	0.082	-0.042	0.081	-0.038	0.078
$\hat{\pi}_{12}$	0.000	-0.010	0.083	-0.011	0.074	-0.008	0.083	-0.008	0.084	-0.007	0.079
$\hat{\pi}_{13}$	0.000	0.005	0.040	0.005	0.037	0.003	0.040	0.003	0.040	0.003	0.037
$\hat{\varphi}_{11}$	-0.300	-0.272	0.127	-0.279	0.127	-0.275	0.129	-0.275	0.131	-0.282	0.122
$\hat{\varphi}_{12}$	0.000	-0.001	0.082	0.018	0.132	-0.002	0.085	-0.001	0.084	-0.002	0.079
$\hat{\varphi}_{13}$	0.000	0.001	0.049	-0.520	0.149	0.001	0.052	0.001	0.050	0.001	0.048
$\hat{\pi}_{21}$	0.000	-0.017	0.099	-0.019	0.090	-0.009	0.092	-0.009	0.094	-0.010	0.076
$\hat{\pi}_{22}$	0.000	-0.072	0.109	-0.020	0.091	-0.051	0.098	-0.047	0.102	-0.040	0.084
$\hat{\pi}_{23}$	0.000	0.009	0.047	0.010	0.045	0.004	0.043	0.004	0.049	0.005	0.036
$\hat{\varphi}_{21}$	0.000	0.016	0.133	0.000	0.080	0.010	0.127	0.010	0.132	0.009	0.103
$\hat{\varphi}_{22}$	-0.700	-0.647	0.101	-0.668	0.100	-0.656	0.095	-0.660	0.095	-0.670	0.087
$\hat{\varphi}_{23}$	0.000	-0.003	0.057	-1.018	0.105	-0.004	0.052	-0.003	0.053	-0.001	0.045
$\hat{\pi}_{31}$	1.000	1.026	0.101	1.025	0.102	1.027	0.109	1.026	0.109	1.026	0.109
$\hat{\pi}_{32}$	1.000	1.025	0.101	1.026	0.103	1.027	0.110	1.026	0.112	1.025	0.114
$\hat{\pi}_{33}$	-0.500	-0.512	0.048	-0.512	0.049	-0.513	0.052	-0.513	0.053	-0.513	0.052
$\hat{\varphi}_{31}$	-0.500	-0.520	0.149	0.000	0.049	-0.521	0.158	-0.523	0.157	-0.522	0.161
$\hat{\varphi}_{32}$	-1.000	-1.017	0.106	-0.003	0.056	-1.018	0.114	-1.016	0.115	-1.018	0.117
$\hat{\varphi}_{33}$	-0.100	-0.094	0.094	-0.093	0.066	-0.095	0.069	-0.093	0.071	-0.095	0.070
n=300											
		OLS		VECM		FOLSH		FVECH		MLE	
Parameters	True Value	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
$\hat{\pi}_{11}$	0.000	-0.021	0.042	-0.007	0.039	-0.019	0.043	-0.019	0.043	-0.016	0.037
$\hat{\pi}_{12}$	0.000	-0.006	0.042	-0.007	0.039	-0.005	0.042	-0.005	0.042	-0.004	0.038
$\hat{\pi}_{13}$	0.000	0.004	0.020	0.003	0.020	0.003	0.021	0.003	0.020	0.002	0.019
$\hat{\varphi}_{11}$	-0.300	-0.281	0.073	-0.285	0.074	-0.284	0.074	-0.285	0.073	-0.287	0.073
$\hat{\varphi}_{12}$	0.000	0.004	0.043	0.005	0.078	0.003	0.043	0.003	0.044	0.001	0.041
$\hat{\varphi}_{13}$	0.000	0.000	0.028	-0.502	0.088	0.000	0.028	0.000	0.028	0.001	0.025
$\hat{\pi}_{21}$	0.000	-0.004	0.054	-0.004	0.052	0.000	0.045	-0.001	0.045	-0.001	0.035
$\hat{\pi}_{22}$	0.000	-0.021	0.055	-0.004	0.052	-0.011	0.046	-0.011	0.046	-0.008	0.035
$\hat{\pi}_{23}$	0.000	0.002	0.026	0.002	0.026	0.000	0.022	0.000	0.022	0.000	0.017
$\hat{\varphi}_{21}$	0.000	0.005	0.078	0.004	0.043	0.001	0.070	0.002	0.069	0.002	0.054
$\hat{\varphi}_{22}$	-0.700	-0.683	0.057	-0.690	0.057	-0.688	0.052	-0.689	0.051	-0.692	0.040
$\hat{\varphi}_{23}$	0.000	-0.001	0.033	-1.002	0.059	-0.001	0.028	0.000	0.028	0.000	0.021
$\hat{\pi}_{31}$	1.000	1.003	0.056	1.003	0.056	1.002	0.055	1.002	0.056	1.002	0.057
$\hat{\pi}_{32}$	1.000	1.003	0.056	1.004	0.056	1.003	0.056	1.003	0.056	1.002	0.057
$\hat{\pi}_{33}$	-0.500	-0.502	0.027	-0.502	0.027	-0.502	0.027	-0.502	0.027	-0.501	0.028
$\hat{\varphi}_{31}$	-0.500	-0.502	0.088	0.000	0.028	-0.501	0.089	-0.501	0.089	-0.501	0.088
$\hat{\varphi}_{32}$	-1.000	-1.002	0.059	-0.001	0.032	-1.001	0.060	-1.002	0.059	-1.000	0.060
$\hat{\varphi}_{33}$	-0.100	-0.096	0.037	-0.096	0.037	-0.096	0.038	-0.096	0.038	-0.096	0.038

Table 1 (Continued) – Parameter Estimates from Monte Carlo Simulation

		n=500									
		OLS		VECM		FOLSH		FVECH		MLE	
Parameters	True Value	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
$\hat{\pi}_{11}$	0.000	-0.010	0.033	-0.002	0.032	-0.010	0.033	-0.009	0.033	-0.007	0.029
$\hat{\pi}_{12}$	0.000	-0.001	0.033	-0.002	0.031	-0.001	0.034	-0.001	0.034	0.000	0.029
$\hat{\pi}_{13}$	0.000	0.001	0.016	0.001	0.016	0.001	0.016	0.000	0.016	0.000	0.014
$\hat{\varphi}_{11}$	-0.300	-0.297	0.056	-0.299	0.056	-0.296	0.059	-0.297	0.057	-0.297	0.050
$\hat{\varphi}_{12}$	0.000	0.000	0.032	0.003	0.060	0.000	0.032	0.000	0.032	0.000	0.029
$\hat{\varphi}_{13}$	0.000	-0.001	0.022	-0.506	0.066	-0.001	0.022	-0.001	0.022	0.000	0.019
$\hat{\pi}_{21}$	0.000	-0.004	0.040	-0.004	0.040	-0.002	0.035	-0.002	0.035	-0.001	0.026
$\hat{\pi}_{22}$	0.000	-0.015	0.041	-0.004	0.040	-0.009	0.036	-0.009	0.036	-0.005	0.027
$\hat{\pi}_{23}$	0.000	0.002	0.020	0.002	0.020	0.001	0.017	0.001	0.017	0.000	0.013
$\hat{\varphi}_{21}$	0.000	0.003	0.060	0.001	0.031	0.001	0.053	0.001	0.052	0.002	0.043
$\hat{\varphi}_{22}$	-0.700	-0.691	0.044	-0.695	0.044	-0.694	0.038	-0.695	0.037	-0.696	0.029
$\hat{\varphi}_{23}$	0.000	0.001	0.025	-1.002	0.042	0.001	0.022	0.000	0.021	0.000	0.016
$\hat{\pi}_{31}$	1.000	1.004	0.039	1.004	0.039	1.004	0.040	1.003	0.040	1.005	0.044
$\hat{\pi}_{32}$	1.000	1.003	0.040	1.003	0.040	1.003	0.041	1.002	0.041	1.004	0.043
$\hat{\pi}_{33}$	-0.500	-0.502	0.020	-0.502	0.020	-0.502	0.020	-0.501	0.020	-0.502	0.022
$\hat{\varphi}_{31}$	-0.500	-0.506	0.066	-0.001	0.022	-0.504	0.066	-0.504	0.066	-0.506	0.072
$\hat{\varphi}_{32}$	-1.000	-1.002	0.042	0.001	0.025	-1.001	0.043	-1.001	0.043	-1.001	0.045
$\hat{\varphi}_{33}$	-0.100	-0.099	0.028	-0.099	0.028	-0.099	0.028	-0.099	0.028	-0.099	0.031

MLE is still the most efficient estimator as shown by the least average MSE in every sample size. However, MLE become computationally demanding when number of parameter is large. Table 2 shows that FGLS-based estimator (FOLSH and FVECH) perform better than OLS and VECM and only slightly inferior to MLE. It suggests that FGLS-based estimator could be useful in overcoming computation burden of the MLE. FGLS-based estimator needs to compute inverse of $\widehat{\Omega}$ which is a very large and sparse matrix, but the inversion of that matrix may cause computational problems as mentioned above in Section 4. Such problems can be solved by the suggested method in that section. The algorithm for matrix inversion in most statistical software is still limited only for matrix in small dimension. We already tried to compute $\widehat{\Omega}^{-1}$ using standard command in EViews and MATLAB in our simulation, while n<100 FGLS-based estimators perform fairly good that comparable to MLE. However, when n becomes larger (i.e. n=300 and n=500), the FGLS-based estimator become poorly inefficient since it produces extreme values for the estimated parameters. All estimated parameters from FGLS-based estimators presented in this paper are based on our matrix inversion procedure. The results based on standard matrix inversion in statistical software are not presented to save space.

Table 2 – Average of Mean Squared Error

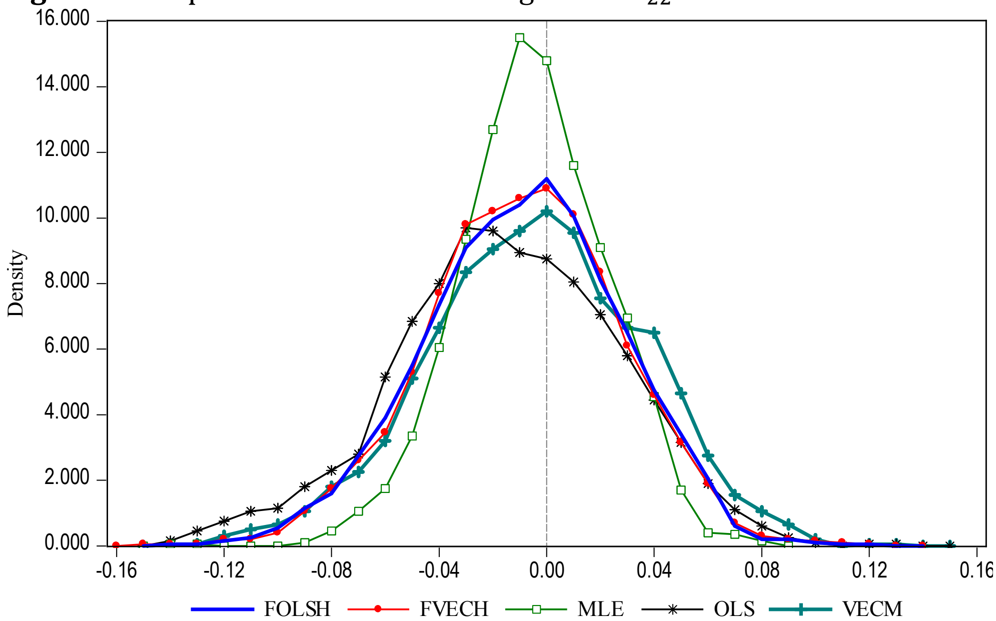
	OLS	VECM	FOLSH	FVECH	MLE
n=100	0.00970	0.15012	0.00936**	0.00958*	0.00834***
n=300	0.00280	0.14180	0.00255*	0.00253**	0.00218***
n=500	0.00154	0.14104	0.00144*	0.00141**	0.00127***

Note: *** The best estimator, ** 2nd best estimator, * 3rd best estimator

Figures 1 compares the distribution of $\hat{\pi}_{22}$ with n=500. The figures show that MLE is the most efficient estimator. FOLSH and FVECH have very similar efficiency as shown by the empirical distribution histogram and relatively are more efficient than OLS and VECM. The figure also shows that OLS estimator is biased to the left although the sample size is large (n=500).

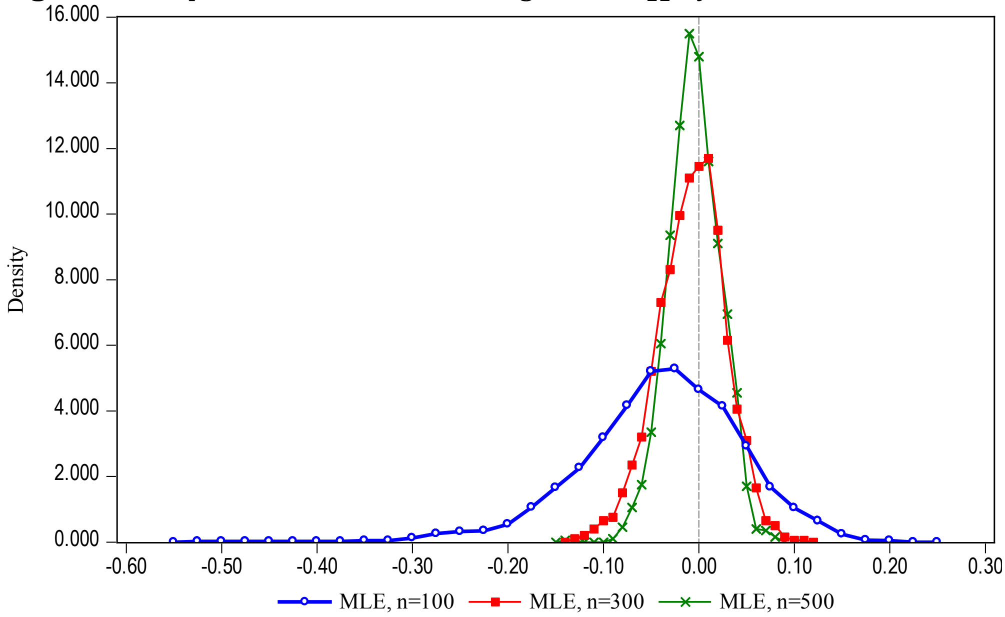
Figure 2 and 3 show example of empirical distribution of the estimated parameter $\hat{\pi}_{22}$ by MLE and FOLSH respectively, for n=100, 300, and 500. Those figures suggest that both ML estimator and FOLSH are consistent estimators as the estimated parameter more converge to the true value when the sample size is larger. Both MLE and FOLSH tend to be unbiased when sample size is large.

Figure 1 –Empirical Distribution Histogram of $\hat{\pi}_{22}$ when n=500



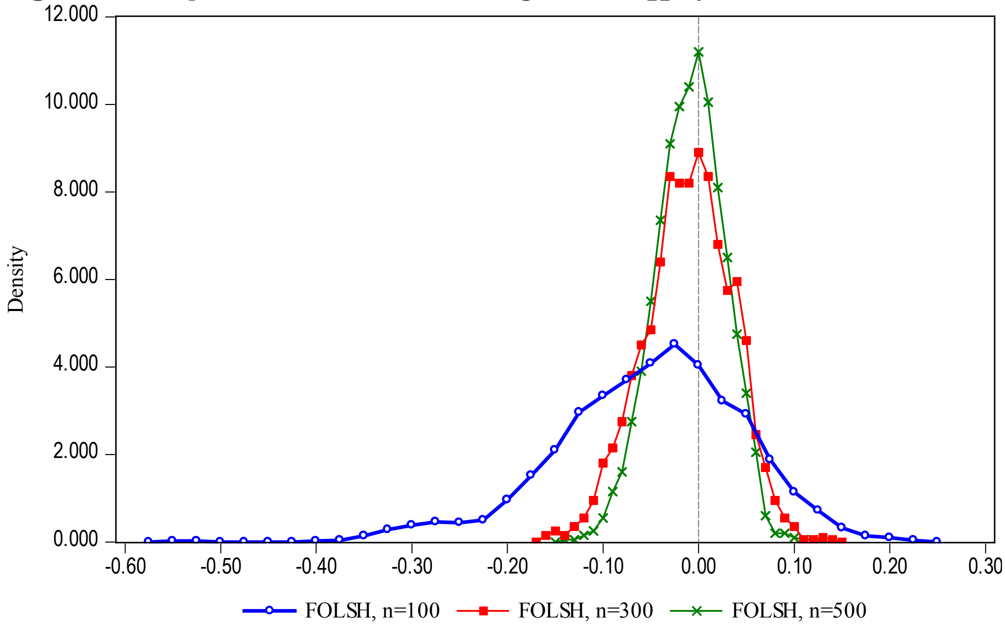
Note: The true value for $\hat{\pi}_{22}$ is 0 as shown by the vertical dashed line

Figure 2 –Empirical Distribution Histogram of $\hat{\pi}_{22}$ by MLE



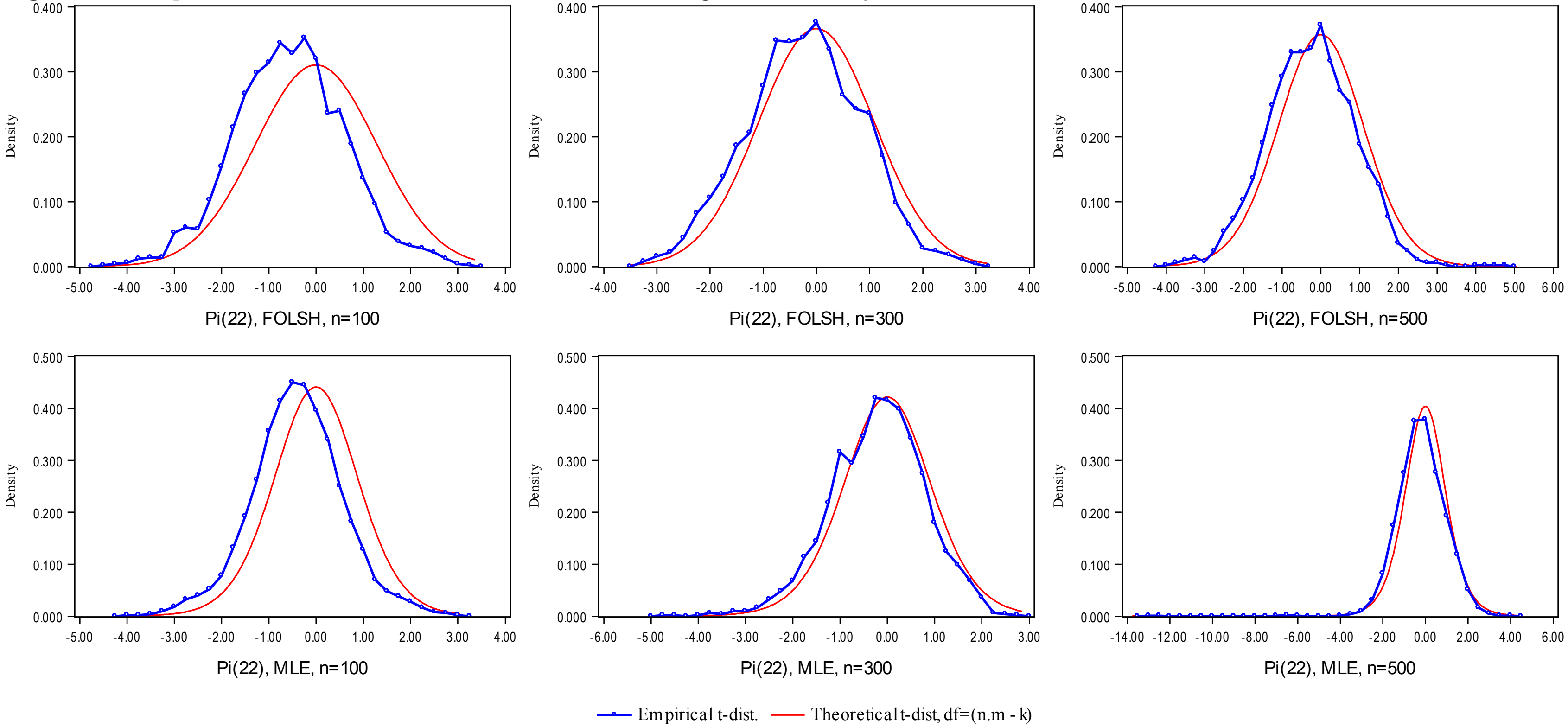
Note: The vertical dashed line indicates the true value of the parameter

Figure 3 –Empirical Distribution Histogram of $\hat{\pi}_{22}$ by FOLSH



Note: The vertical dashed line indicates the true value of the parameter.

Figure 4 –Empirical and Theoretical t-Distribution Histogram of $\hat{\pi}_{22}$ by FOLSH and MLE



Note: degree of freedom = nm-k, where n=number observation, m=number of equation (3), and k=number of parameter (18)

Table 3 – Average of Rejection Rate of Null Hypothesis* Test at 5 Percent Significance Level

n=100	$\hat{\pi}_{11}$	$\hat{\pi}_{12}$	$\hat{\pi}_{13}$	$\hat{\varphi}_{11}$	$\hat{\varphi}_{12}$	$\hat{\varphi}_{13}$	$\hat{\pi}_{21}$	$\hat{\pi}_{22}$	$\hat{\pi}_{23}$	$\hat{\varphi}_{21}$	$\hat{\varphi}_{22}$	$\hat{\varphi}_{23}$	$\hat{\pi}_{31}$	$\hat{\pi}_{32}$	$\hat{\pi}_{33}$	$\hat{\varphi}_{31}$	$\hat{\varphi}_{32}$	$\hat{\varphi}_{33}$	Average
OLS	0.151	0.127	0.127	0.142	0.119	0.099	0.135	0.203	0.128	0.118	0.171	0.102	0.107	0.100	0.105	0.095	0.113	0.104	0.125
VECM	0.111	0.107	0.111	0.142	0.110	0.969	0.136	0.131	0.134	0.117	0.155	1.000	0.139	0.086	0.100	1.000	1.000	0.113	0.315
FOLSH	0.024	0.071	0.072	0.097	0.080	0.071	0.061	0.032	0.086	0.088	0.138	0.059	0.080	0.077	0.028	0.035	0.035	0.060	0.066
FVECH	0.021	0.070	0.079	0.099	0.081	0.062	0.052	0.027	0.073	0.092	0.126	0.062	0.085	0.082	0.028	0.040	0.035	0.063	0.065
MLE	0.008	0.031	0.033	0.037	0.028	0.026	0.033	0.022	0.043	0.043	0.070	0.037	0.049	0.043	0.015	0.018	0.012	0.023	0.032
n=300	$\hat{\pi}_{11}$	$\hat{\pi}_{12}$	$\hat{\pi}_{13}$	$\hat{\varphi}_{11}$	$\hat{\varphi}_{12}$	$\hat{\varphi}_{13}$	$\hat{\pi}_{21}$	$\hat{\pi}_{22}$	$\hat{\pi}_{23}$	$\hat{\varphi}_{21}$	$\hat{\varphi}_{22}$	$\hat{\varphi}_{23}$	$\hat{\pi}_{31}$	$\hat{\pi}_{32}$	$\hat{\pi}_{33}$	$\hat{\varphi}_{31}$	$\hat{\varphi}_{32}$	$\hat{\varphi}_{33}$	Average
OLS	0.143	0.111	0.097	0.154	0.093	0.098	0.146	0.165	0.140	0.110	0.151	0.100	0.103	0.113	0.111	0.122	0.112	0.108	0.121
VECM	0.099	0.099	0.099	0.157	0.111	1.000	0.138	0.137	0.138	0.090	0.147	1.000	0.119	0.105	0.104	1.000	1.000	0.115	0.314
FOLSH	0.030	0.054	0.072	0.105	0.068	0.055	0.061	0.030	0.064	0.069	0.085	0.054	0.049	0.056	0.053	0.056	0.057	0.060	0.060
FVECH	0.026	0.048	0.075	0.102	0.070	0.062	0.063	0.031	0.062	0.075	0.087	0.053	0.049	0.055	0.051	0.057	0.054	0.062	0.060
MLE	0.015	0.030	0.041	0.075	0.043	0.038	0.045	0.026	0.053	0.071	0.077	0.042	0.038	0.042	0.048	0.049	0.053	0.046	0.046
n=500	$\hat{\pi}_{11}$	$\hat{\pi}_{12}$	$\hat{\pi}_{13}$	$\hat{\varphi}_{11}$	$\hat{\varphi}_{12}$	$\hat{\varphi}_{13}$	$\hat{\pi}_{21}$	$\hat{\pi}_{22}$	$\hat{\pi}_{23}$	$\hat{\varphi}_{21}$	$\hat{\varphi}_{22}$	$\hat{\varphi}_{23}$	$\hat{\pi}_{31}$	$\hat{\pi}_{32}$	$\hat{\pi}_{33}$	$\hat{\varphi}_{31}$	$\hat{\varphi}_{32}$	$\hat{\varphi}_{33}$	Average
OLS	0.122	0.122	0.110	0.142	0.083	0.101	0.135	0.153	0.134	0.111	0.148	0.109	0.096	0.095	0.094	0.108	0.096	0.104	0.115
VECM	0.110	0.110	0.110	0.147	0.114	1.000	0.137	0.137	0.137	0.089	0.156	1.000	0.106	0.087	0.091	1.000	1.000	0.106	0.313
FOLSH	0.031	0.073	0.072	0.074	0.045	0.055	0.061	0.037	0.076	0.075	0.072	0.069	0.052	0.050	0.036	0.044	0.041	0.040	0.056
FVECH	0.038	0.077	0.067	0.073	0.042	0.053	0.065	0.040	0.070	0.071	0.065	0.063	0.052	0.049	0.036	0.045	0.044	0.041	0.055
MLE	0.023	0.047	0.051	0.062	0.046	0.048	0.061	0.040	0.059	0.086	0.054	0.055	0.047	0.044	0.038	0.039	0.039	0.043	0.049

*The null hypothesis: the estimated parameter = its true value, the alternative hypothesis: the estimated parameter ≠ its true value

Figure 4 shows example of empirical t-statistic distribution for $\hat{\pi}_{22}$. From the figures, both FOLSH and MLE tend to conform to student-t distribution when larger sample size is used. The empirical distribution for $\hat{\pi}_{22}$ estimated by FVECH is very similar to that by FOLSH. Table 3 shows that rejection rate of null hypothesis that each parameter is equal to its true value is also close to the significance level (0.05) for parameter estimated by FOLSH, FVECH, and MLE. From the table it is also apparent that estimators that do not consider multivariate GARCH error structure (OLS and VECM) has higher rejection rate compares to those of estimators that consider the error structure (FOLSH, FVECH, and MLE). These findings show us that neglecting the presence of multivariate GARCH error structure will increase the rejection rate or the type I error.

5. Empirical Application

Weekly data from July 1997 until July 2011 of US S&P500, Japan Nikkei225 and Malaysia KLSE composite index are collected as a dataset for our model (n=732). The indexes are stated in logarithmic and are measured in US Dollar. Since they are in log index, their first order differences can be regarded as stock market return of the respective markets.

Unit root test indicates that the three time series are non-stationary at level, but they are stationary at their first difference. The Augmented Dickey Fuller (ADF) statistic (τ-*stat.*) for data in level indicates the null hypothesis that the series has unit root cannot be rejected at 10 percent significance level or less. Meanwhile, the τ-stat. for the respective series in the first order difference significantly rejects the null hypothesis of unit root at one percent significance level.

Table 4 - Unit Root Test

	Level		1st Differences	
Unit Root Test	ADF τ-*stat.*	P-Value	ADF τ-*stat.*	P-Value
S&P500	-2.4618	0.1254	-29.7881	0.0000
Nikkei225	-2.4258	0.1349	-27.8684	0.0000
KLSE	-0.8080	0.8158	-28.1092	0.0000

Null Hypothesis: Series has unit root

Johansen's cointegration test was performed for the dataset, the results, as presented in Table 5, show that one cointegrating equation is found from tests based on both Trace and Maximum Eigenvalue method.

Estimation of VECM with one cointegrating equation is shown in Table 6, where Y_1, Y_2, and Y_3 correspond to log of S&P500, Nikkei225, and KLSE index respectively. From the table, it shows that coefficients of error correction for cointegrating equation are all significant to show that the stock markets have long run price relationship. In the VAR part, lagged S&P500 return has significant effect to itself and to both Japanese and Malaysian stock market returns. The results indicate that US stock market is still a very dominant market that shares its greater influence to other markets.

Table 5 - Johansen Cointegration Test

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized		Trace	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.0436	38.0006	29.7971	0.0046
At most 1	0.0058	5.4141	15.4947	0.7634
At most 2	0.0016	1.1569	3.8415	0.2821
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized		Max-Eigen	0.05	
No. of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None *	0.0436	32.5865	21.1316	0.0008
At most 1	0.0058	4.2572	14.2646	0.8313
At most 2	0.0016	1.1569	3.8415	0.2821

Trace and Max-eigenvalue test indicates 1 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Table 6 - Vector Error Correction Model (VECM)

Coint.Eq.	Coef.		
Y_{1t-1}	1.000		
Y_{2t-1}	-0.682		
	(0.090)		
Y_{3t-1}	-0.024		
	(0.054)		
C	-3.700		
E.C. Eq.	ΔY_{1t}	ΔY_{2t}	ΔY_{3t}
Coint.Eq.	-0.024	0.027	0.040
	(0.009)	(0.012)	(0.015)
ΔY_{1t-1}	-0.095	0.214	0.174
	(0.041)	(0.053)	(0.069)
ΔY_{2t-1}	-0.007	-0.076	0.032
	(0.032)	(0.042)	(0.055)
ΔY_{3t-1}	0.012	-0.047	-0.076
	(0.024)	(0.031)	(0.040)
C	0.000	-0.001	0.000
	(0.001)	(0.001)	(0.002)

Standard Error in Parenthesis

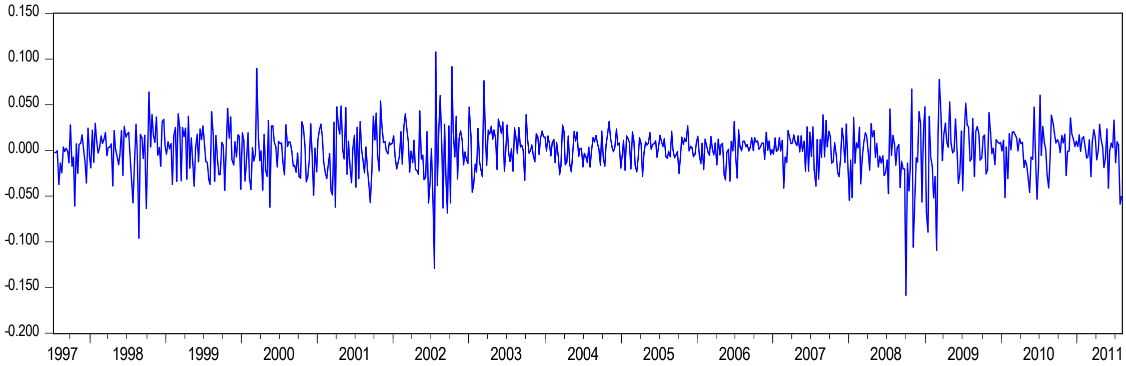
In addition, the significant VECM coefficients also indicate that past information (lagged variables of both price and return) can explain the present stock market returns. It implies that the stock markets are neither informationally efficient nor perfectly integrated. The importance of past information may be used for setting arbitrage strategies in the markets to exploit the market inefficiency.

The residuals of estimated VECM show a non-homoscedastic structure as it is shown in Figure 5. The residual of VECM can be regarded as a market shock or the unexpected return, and from the figure we can observed that during period of 1999-2002 and 2008-2009 the volatility of the US residuals were higher compared to that in the other periods. The two sub-periods are known as the burst of dot-com bubble and the collapse of financial institutions in the US market. The pattern of the Japan residuals is less clear to be connected with some events; however, it is clear that the residuals are also not homoscedastic. Meanwhile, the residuals plot of Malaysian stock market returns show that higher volatility is detected during the Asian financial crisis in 1997-1998 and also during the US financial turmoil in late 2008 until 2009.

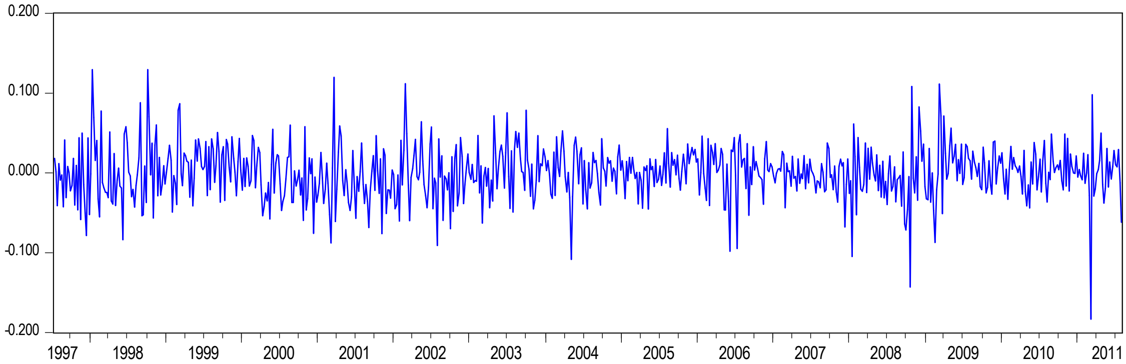
The similar pattern of residuals during a crisis period, i.e. during the collapse of Lehman Brothers in US, indicates the presence of volatility spillover from US to other markets, and thus it become evidence of the correlated structure of the residuals. This phenomenon is often seen in financial market. The latter property of the residuals becomes a motivation to apply SUR type model.

Residuals from each single OLS model are also computed, the results are similar to those of VECM's residuals that they indicate that the residuals are heteroscedastic. The residuals are then used in estimating $\widehat{H}_t$ by Diagonal BEKK. Having the variance-covariance series, we proceed to the next step for constructing matrix $\widehat{\Omega}$ and used it to obtain FGLS estimators.

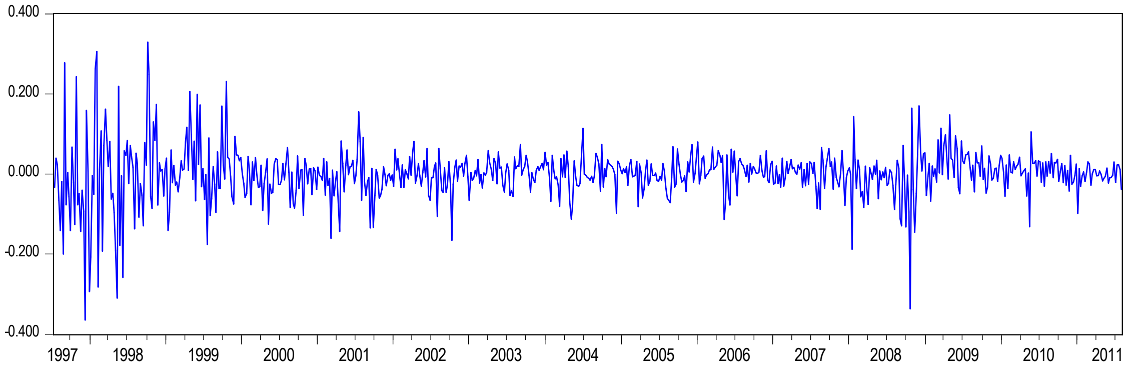
Figure 5 - Residuals of VECM



Residuals of D(Ln(S&P500))

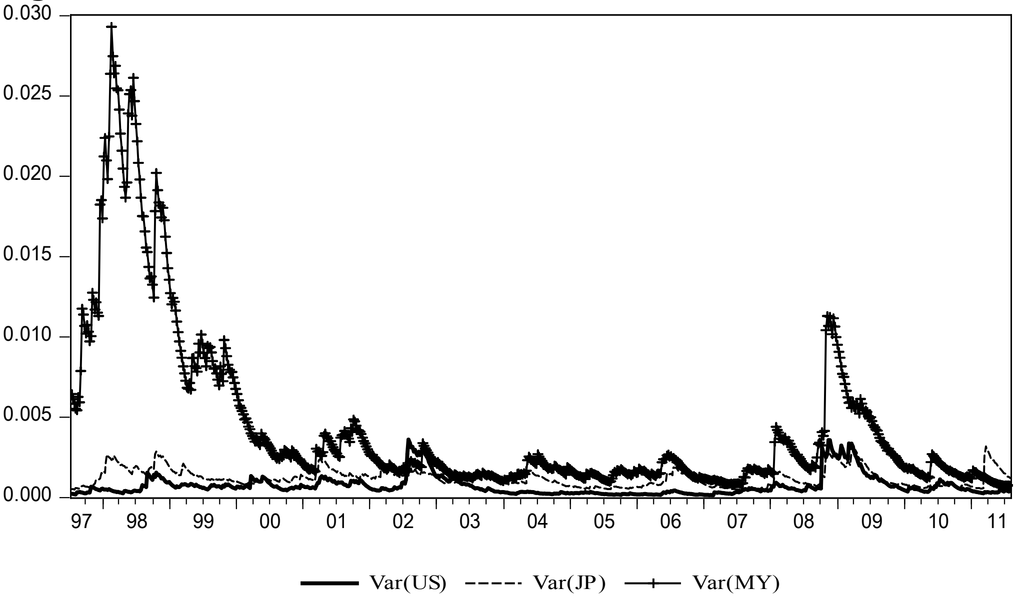


Residuals of D(Ln(Nikkei225))

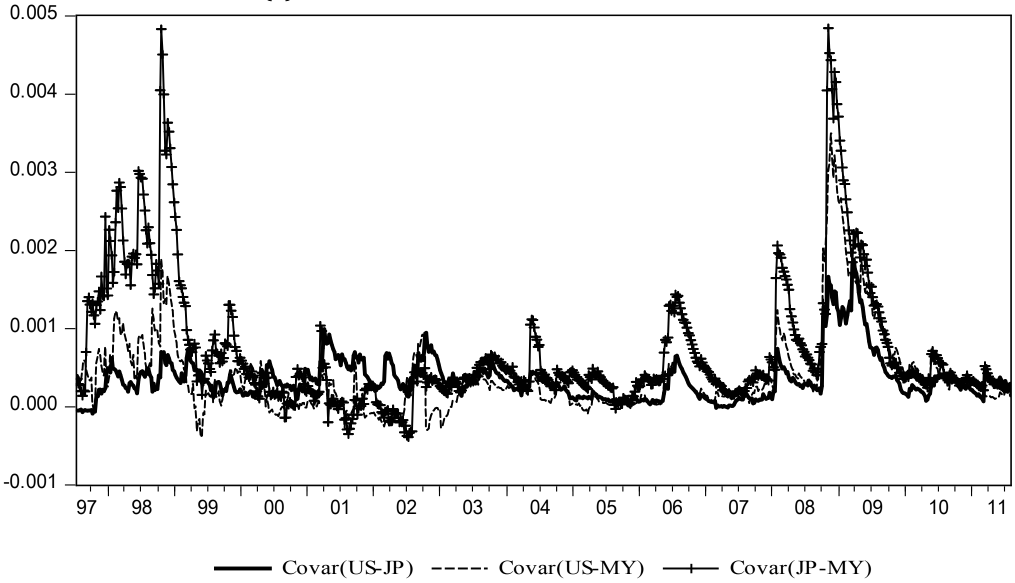


Residuals of D(Ln(KLSE))

Figure 6 - Estimated Conditional Variance-Covariance



(a) Conditional Variance



(b) Conditional Covariance

The FGLS estimators, the restated VECM (without GARCH), OLS, and MLE estimated parameters are shown in Table 7. As shown in the table, although the sign and value of the estimated parameters are very similar among the various estimation methods, but the probability of significance are sometime different. Based on the data properties shown in Figure 5 and 6, the GARCH error structure does exists. And based on the simulation results, estimation methods that take into account the GARCH structure are more efficient than those that ignore the structure. Therefore, in the empirical example, the use of such methods (OLS and VECM) might produce wrong conclusion regarding the significance of the estimated parameters.

For example, $\hat{\pi}_{32}$ estimated by OLS (and VECM) is significantly different from zero, but it is not significant when it is estimated by FOLSH, FVECH, and MLE. It means that when we estimate the parameter using method that neglecting the MGARCH error structure we would conclude that lagged of Nikkei225 Index (Japanese stock prices) affects Malaysia KLSE returns (Malaysian stock returns), while we should not.

Table 6 - Estimated Parameters of OLS, VECM, FOLSH, FVECH, and MLE

Estimated Parameter	OLS		VECM		FOLSH		FVECH		MLE	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
$\hat{C}_1^0$	0.139	0.047**	0.089	#	0.138	0.045**	0.195	0.045**	0.139	0.038**
$\hat{\pi}_{11}$	-0.028	0.009**	-0.024	#	-0.029	0.010**	-0.040	0.010**	-0.024	0.007**
$\hat{\pi}_{12}$	0.014	0.007*	0.016	#	0.014	0.004**	0.019	0.004**	0.008	0.004
$\hat{\pi}_{13}$	-0.001	0.003	0.001	#	0.000	0.000**	0.000	0.000**	-0.001	0.002
$\hat{\varphi}_{11}$	-0.093	0.041*	-0.095	0.041*	-0.091	0.159	-0.090	0.061	-0.106	0.040**
$\hat{\varphi}_{12}$	-0.006	0.032	-0.007	0.032	0.007	0.010	0.008	0.005	-0.004	0.024
$\hat{\varphi}_{13}$	0.013	0.024	0.012	0.024	0.038	0.028	0.038	0.017*	0.035	0.015*
$\hat{C}_2^0$	-0.023	0.061	-0.102	#	-0.017	0.025	0.028	0.016*	-0.006	0.048
$\hat{\pi}_{21}$	0.020	0.012	0.027	#	0.015	0.016	0.005	0.003*	0.019	0.009*
$\hat{\pi}_{22}$	-0.024	0.009**	-0.019	#	-0.020	0.031	-0.015	0.009*	-0.023	0.006**
$\hat{\pi}_{23}$	-0.001	0.003	-0.001	#	0.001	0.002	0.001	0.000*	-0.003	0.003
$\hat{\varphi}_{21}$	0.218	0.053**	0.214	0.053**	0.259	0.086**	0.261	0.066**	0.200	0.042**
$\hat{\varphi}_{22}$	-0.075	0.042	-0.076	0.042	-0.094	0.045*	-0.095	0.031**	-0.050	0.038
$\hat{\varphi}_{23}$	-0.046	0.031	-0.047	0.031	-0.056	0.037	-0.057	0.023**	-0.026	0.022
$\hat{C}_3^0$	-0.129	0.079	-0.147	#	-0.109	0.272	-0.052	0.035	-0.017	0.050
$\hat{\pi}_{31}$	0.040	0.016*	0.040	#	0.026	0.041	0.011	0.007*	0.007	0.011
$\hat{\pi}_{32}$	-0.026	0.011*	-0.027	#	-0.014	0.029	-0.004	0.003	-0.003	0.008
$\hat{\pi}_{33}$	-0.005	0.004	-0.001	#	-0.001	0.001	-0.001	0.001	-0.003	0.004
$\hat{\varphi}_{31}$	0.173	0.069*	0.174	0.069*	0.246	0.119*	0.251	0.082**	0.223	0.036**
$\hat{\varphi}_{32}$	0.031	0.055	0.032	0.055	0.006	0.003*	0.002	0.001**	0.011	0.030
$\hat{\varphi}_{33}$	-0.073	0.040	-0.076	0.040	-0.067	0.041	-0.061	0.023**	-0.026	0.037

** significant at 0.01
* significant at 0.05
The Standard error marked by # indicates that the coefficient is computed from loading vector and adjustment vector in the error correction equations, the respective standard error for these parameters are shown in Table 5.

6. Concluding Remarks

The standard Vector Error correction model (VECM), which is based on normality assumption of error term, is often applied to analyze the real financial time series. However, as shown in the section 5 it is often seen that residuals of this model seem to follow GARCH errors process. From this experience we extend the standard VECM to include GARCH error process. We call such model as VEC-GARCH model.

Although the maximum likelihood (ML) estimator is known as the most efficient estimator under the normality assumption, ML estimation is computationally demanding when a model to be estimated is not small. To overcome these disadvantages and to reduce computational burden of ML estimator we consider the generalized least square estimator (GLS) instead of ML estimator. GLS is relatively free from the distributional assumptions.

In this paper we mainly concerns with the GLS representation, the algorithm of it, and the properties of it, we have examined the performance of GLS and MLE in VEC-GARCH model by Monte Carlo simulation and the applicability of it by real data analysis of the financial time series. The Monte Carlo simulation naturally has shown that MLE is still better than the FGLS. However FGLS-based estimators that also consider GARCH error structure are also more efficient than estimators that neglect the error structure. The performance of MLE and FGLS-based estimator in our simulation are only slightly different, yet both are better estimators compare to the OLS and VECM. Thus, the suggested FGLS-based estimator may overcome the disadvantages of MLE, especially in reducing the computational burden.

Our suggested method for the large matrix inversion successfully overcomes the computational problem such as memory size, computer time, and innacurate numerical results. The estimated parameters from the FGLS-based estimator performed in the simulation is as good as the MLE.

There, however, remain several problems in estimating VECM with GARCH errors for the future research as follows: (1) to use realized volatility (RV) instead of multivariate GARCH model, (2) to compare the GLS and MLE under non-normality by Monte Carlo simulation, (3) to carry out theoretical comparisons of asymptotic properties of the GLS and MLE, under normality and non-normality, (4) to examine the performance of VEC model with GARCH errors when it is applied to empirical analysis of financial time series. We have a plan to attack these problems in future.

REFERENCES

Bauwens, L., S. Laurent and J. V. K. Rombouts (2006). Multivaritae GARCH Models: a survey, Journal of applied Econometrics, 21: 79-109.

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.

Engle, R. F. and K. F. Kroner (1995). Multivariate Simultaneous Generalized ARCH. *Econometric theory,* 11, 122-150.

Johansen, S. (1995). *Likelihood-based Inference in Cointegrated Vector autoregressive Models*, Oxford: Oxford University Press.

Johnston, J. and J. DiNardo (2007). *Econometric Methods*, McGraw-Hill Co, Inc.

Silvennoinenn, A. and T. Teräsvirta (2009). Multivariate GARCH models, Handbook of Financial Time Series, ed. by T. G. Andersen, R. A. Davis, J. P. Kreiss and T. Mikosch, New Yor: Springer.

Zellner, A. (1962). An Efficient Method of Estimating Seemingly Unrelated Regressions and Test of Aggregation Bias, *Journal of the American Statistical Association,* 57, 500-509.