

TI 2017-069/III  
Tinbergen Institute Discussion Paper



# Theoretical and Empirical Differences Between Diagonal and Full BEKK for Risk Management

David Allen<sup>1</sup>  
Michael McAleer<sup>2</sup>

- 1: Department of Mathematics, University of Sydney, Australia
- 2: Department of Quantitative Finance, National Tsing Hua University, Taiwan;  
Discipline of Business Analytics, University of Sydney Business School,  
Australia; Econometric Institute, Erasmus School of Economics, Erasmus  
University Rotterdam, The Netherlands; Department of Quantitative Economics,  
Complutense University of Madrid, Spain; Institute of Advanced Sciences,  
Yokohama National University, Japan

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: [discussionpapers@tinbergen.nl](mailto:discussionpapers@tinbergen.nl)

More TI discussion papers can be downloaded at the [Tinbergen Site](#)

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam  
Gustav Mahlerplein 117  
1082 MS Amsterdam  
The Netherlands  
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam  
Burg. Oudlaan 50  
3062 PA Rotterdam  
The Netherlands  
Tel.: +31(0)10 408 8900

# Theoretical and Empirical Differences Between Diagonal and Full BEKK for Risk Management

David E. Allen<sup>a</sup>, Michael J. McAleer<sup>b</sup>,

<sup>a</sup>*School of Mathematics and Statistics, University of Sydney, Centre for Applied Finance, University of South Australia, and School of Business and Law, Edith Cowan University*

<sup>b</sup>*Department of Quantitative Finance, National Tsing Hua University, Taiwan, Discipline of Business Analytics, University of Sydney Business School, Australia, Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands, Department of Quantitative Economics, Complutense University of Madrid, Spain, and Institute of Advanced Sciences, Yokohama National University, Japan*

---

## Abstract

The purpose of the paper is to explore the relative biases in the estimation of the Full BEKK model as compared with the Diagonal BEKK model, which is used as a theoretical and empirical benchmark. Chang and McAleer [4] show that univariate GARCH is not a special case of multivariate GARCH, specifically, the Full BEKK model, and demonstrate that Full BEKK which, in practice, is estimated almost exclusively, has no underlying stochastic process, regularity conditions, or asymptotic properties. Diagonal BEKK (DBEKK) does not suffer from these limitations, and hence provides a suitable benchmark. We use simulated financial returns series to contrast estimates of the conditional variances and covariances from DBEKK and BEKK. The results of non-parametric tests suggest evidence of considerable bias in the Full BEKK estimates. The results of quantile regression analysis show there is a systematic relationship between the two sets of estimates as we move across the quantiles. Estimates of conditional variances from Full BEKK, relative to those from DBEKK, are lower in the left tail and higher in the right tail.

*Keywords:* DBEKK, BEKK, Regularity Conditions, Asymptotic Properties, Non-Parametric, Bias, Quantile regression.

*JEL:* C13, C21, C58.

---

## 1. Introduction

Conditional volatility models are the most widely estimated univariate and multivariate models of time-varying volatility (or dynamic risk) applied to financial data, in the high frequency data domains that are measured in days,

---

*Email address:* michael.mcaleer@gmail.com

hours and minutes. The stochastic processes, regularity conditions and asymptotic properties of these popular univariate conditional volatility models, such as GARCH (see Engle [7], Bollerslev [3] and GJR (see [8]) are well established in the literature. Nevertheless, McAleer and Hafner [11] raised caveats about the existence of the stochastic process and statistical properties underlying exponential GARCH (EGARCH) (see Nelson [15, 16]).

The situation with respect to multivariate conditional volatility models is considerably different. For example, the Full BEKK model, (see Baba et al. [1], and Engle and Kroner [6]) is problematic regarding the existence of its underlying stochastic processes, regularity conditions, and asymptotic properties. These properties have either not yet been established, or are simply assumed rather than derived. These conditions and properties are essential for the existence of the likelihood function, and hence valid statistical analysis of the empirical estimates.

The focus of this paper is to explore the potential and empirical biases that may exist in the estimation of the multivariate Full BEKK model, as referenced in the RATS statistical software.

We use three simulated sets of daily returns derived from ten years of daily data, from 5 March 2007 to 3 March 2017, for Google, IBM and Microsoft. The original adjusted return series were downloaded from Yahoo Finance. The random simulations, created in R, are done in blocks, including five lags so as to preserve autocorrelations and ARCH effects. We use the estimated coefficients of the conditional variances and conditional covariances derived from Diagonal BEKK (DBEKK), which has appropriate regularity conditions and statistical properties, as the benchmark.

McAleer et al. [10] showed that the QMLE of the parameters of DBEKK models are consistent and asymptotically normal, so that standard statistical inference for testing hypotheses is valid. These are compared with estimates of the same coefficients using the Full BEKK model. Non-parametric tests reveal statistically significant bias in the Full BEKK coefficient estimates for the conditional variances and covariances.

The paper is divided into four sections. The introductory section is followed by Section 2, which describes the data sets, their statistical characteristics, and the models and empirical methods used. Section 3 presents the empirical results, and Section 4 provides some concluding remarks.

## 2. Data and Research Method

### 2.1. Data

The original data series comprised ten years of daily price-adjusted data, from 5 March 2007 to 3 March 2017, for Google, IBM and Microsoft, which were downloaded from Yahoo Finance. Descriptive statistics for the three original series, each comprising 2,518 observations, are shown in Table 1. The statistics show that the three series display characteristics that are typical of financial return series, displaying skewness and excess kurtosis. Plots of the original sample of daily adjusted returns are shown in Figure 1. The QQ plots of the series shown in Figure 2 reveal that they have fat tails and do not conform to a normal distribution.

The three original series are stationary, as confirmed by the Augmented Dickey-Fuller tests using constant and trend, and also display significant ARCH effects. The results of these tests are shown in Table 2.

Table 1: Descriptive Statistics for Daily Adjusted-return Series

STATISTIC	IBM	GOOGLE	MICROSOFT
MEAN	0.000355292	0.000526422	0.000432804
MEDIAN	0.000505522	0.000267708	0.000221015
S.D.	0.0141556	0.0186157	0.0177392
SKEWNESS	-0.185381	0.608522	0.184534
KURTOSIS	5.36075	11.9078	9.48851

Table 2: Base Series Unit Root and ARCH Tests

STATISTIC	IBM	GOOGLE	MICROSOFT
ADF test with constant	-28.8918	-10.0804	-29.7647
ADF test with constant and trend	-28.8934	-10.1019	-29.7852
ARCH TEST alpha(0)	0.000145	0.000257	0.000210
ARCH TEST alpha(1)	0.290848	0.337997	0.373119

Note: All tests are significant at the 1% level.

Figure 1: Plots of Return Series

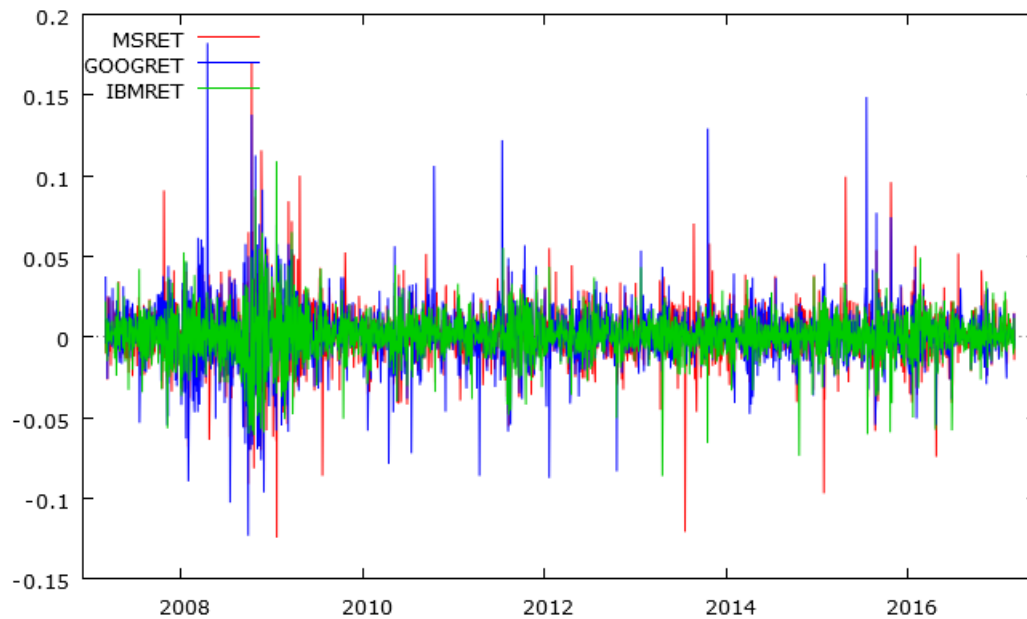
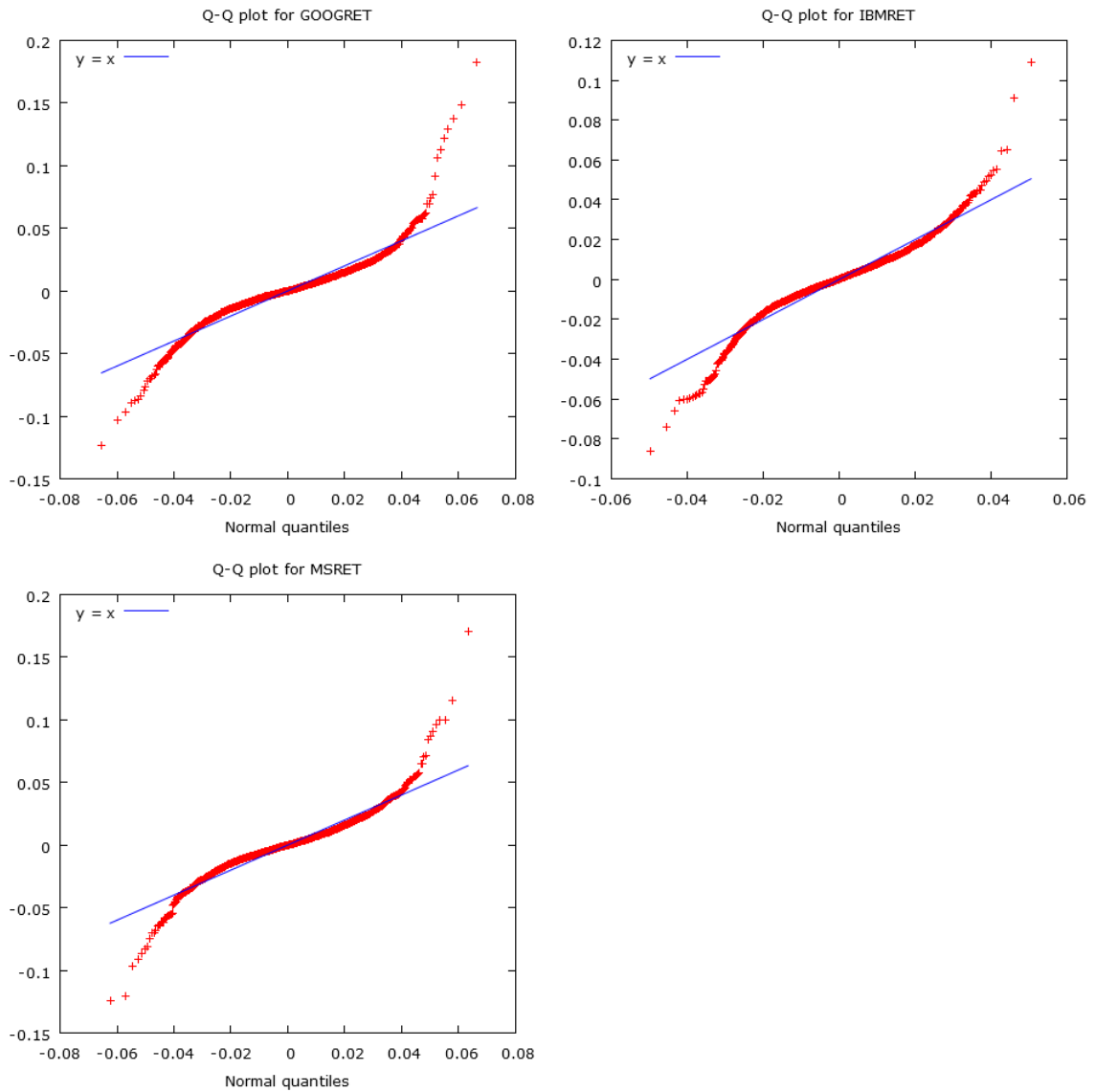


Figure 2: QQ plots of Return Series



## 2.2. Research Models, Simulations and Tests

### 2.2.1. Univariate conditional volatility models

Chang and McAleer [4] show that Full BEKK has no underlying stochastic process, regularity conditions, or asymptotic properties. They point out that, in the development of GARCH, the conditional mean of financial returns for

commodity  $i$ , in a financial portfolio of  $m$  assets, can be developed as follows:

$$y_{it} = E(y_{it} | I_{t-1}) + \varepsilon_{it}, \quad 1 = 1, 2, \dots, m. \quad (1)$$

In equation (1) above, the returns,  $y_{it}$ , represent the log difference of financial asset prices,  $I_{t-1}$  is the information set for all prices at time  $t - 1$ ,  $E(y_{it} | I_{t-1})$  is the conditional expected returns, and  $\varepsilon_{it}$  is a conditionally heteroscedastic error term. The conditional volatility specifications are based on the stochastic specification presumed to underlie the return shocks,  $\varepsilon_{it}$ . Chang and McAleer [4] consider the random coefficient autoregressive process underlying the returns shocks  $\varepsilon_{it}$ , as shown below:

$$\varepsilon_{it} = \phi_{it}\varepsilon_{it-1} + \eta_{it}, \quad (2)$$

where  $\phi_{it} \sim iid(0, \alpha_i)$ ,  $\alpha_i > 0$ ,  $\eta_{it} \sim iid(0, \omega_i)$ ,  $\omega_{ki} > 0$ ,  $\eta_{it} = \varepsilon_{kit}/\sqrt{h_{it}}$  is the standardised residual, and  $h_{it}$  is the conditional volatility of asset  $i$ . Tsay [17] suggested the following formulation for the conditional volatility of asset  $i$  as an ARCH process:

$$E(\varepsilon_{it}^2 | I_{t-1}) \equiv h_{it} = \omega_i + \alpha_i \varepsilon_{it-1}^2, \quad i = 1, 2, \dots, m, \quad m > 1. \quad (3)$$

A lagged dependent variable,  $h_{it-1}$ , is typically added to equation (3) to improve the empirical fit:

$$E(\varepsilon_{it}^2 | I_{t-1}) = \omega_i + \alpha_i \varepsilon_{it-1}^2 + \beta_i h_{it-1}, \quad \beta_i \in (-1, 1). \quad (4)$$

The specification in equation (2) suggests that  $\alpha_i$  and  $\omega_i$  should be positive because they are the unconditional variances of two different stochastic processes. Equation (4) is a GARCH(1,1) model for asset  $i$  (see Bollerslev [3]). The stability condition requires that  $\beta_i \in (-1, 1)$ . Given that the stochastic process in equation (2) follows a random coefficient autoregressive process, under normality (non-normality) of the random errors, the maximum likelihood estimators (quasi-maximum likelihood estimators, QMLE) of the parameters will be consistent and asymptotically normal.



### 2.2.2. Multivariate conditional volatility models

The multivariate extension of the univariate ARCH and GARCH models is given in Baba et al. [1] and Engle and Kroner [6]. The relationship between the returns shocks and the standardised residuals, in the multivariate case, can be written as  $\eta_{it} = \varepsilon_{it}/\sqrt{h_{it}}$ .

The multivariate extension of equation (1) can remain unchanged by assuming that the three components are now  $m \times 1$  vectors, where  $m$  is the number of financial assets. Chang and McAleer [4] consider a vector random coefficient autoregressive process of order one as:

$$\epsilon_t = \phi_t \epsilon_{t-1} + \eta_t, \quad (5)$$

where  $\epsilon_t$  and  $\eta_t$  are  $m \times 1$  vectors,  $\phi_t$  is an  $m \times m$  matrix of random coefficients,  $\phi_t \sim iid, (0, A)$ ,  $A$  is positive definite,  $\eta_t \sim iid, (0, C)$ , where  $C$  is an  $m \times m$  matrix.

In the case where  $A$  is a diagonal matrix, with  $a_{ii} > 0$  for all  $i = 1, \dots, m$ , and  $|b_{ij}| < 1$  for all  $j = 1, \dots, m$ , so that  $A$  has dimension  $m \times m$ , McAleer et al. [10] showed that the multivariate extension of GARCH(1,1) from equation (5) is given as the Diagonal BEKK model, namely:

$$Q_t = CC' + A\epsilon_{t-1}\epsilon'_{t-1}A' + BQ_{t-1}B', \quad (6)$$

where  $A$  and  $B$  are both diagonal matrices. It is essential for the matrix multiplication of  $\epsilon_{t-1}\epsilon'_{t-1}$  by  $A$  that  $A$  is diagonal and positive definite, given that the former is an  $m \times m$  matrix. If this is not the case, equation (6) could not be derived from the vector random coefficient autoregressive process in equation (5).

It was shown in McAleer et al. [10] that the QMLE of the parameters of the DBEKK model are consistent and asymptotically normal, so that standard statistical inference for testing hypotheses is valid. However, Chang and McAleer [4] demonstrate that this is not the case for the Full BEKK model. They consider element  $i$  of equation (5), which can be written as:

$$\varepsilon_{it} = \sum_{j=1}^m \phi_{ijt} \varepsilon_{ijt-1} + \eta_{it}, \quad i = 1, 2, \dots, m, \quad m > 1, \quad (7)$$

which is not equivalent to equation (2) unless  $\phi_{ijt} = 0, \forall j \neq i$ .

Chang and McAleer [4] point out that equation (7) is not a random coefficient autoregressive process because of the presence of another  $m - 1$  random coefficients. Furthermore, equation (7) is not invertible because the random processes cannot be connected to the data, which requires  $m$  equations, as shown in equation (5). This means that the stochastic process underlying univariate ARCH is not a special case of that underlying multivariate ARCH, unless  $\phi_{ijt} = 0 \forall j \neq i$ .

As a result, Chang and McAleer [4] suggest that in the case of a Full BEKK model, where there are no restrictions on the off-diagonal elements in  $\phi_t$ , and hence no restrictions in the off-diagonal elements in  $A$ , it is not possible for univariate ARCH to be a special case of its multivariate counterpart, Full BEKK. This suggests that Full BEKK does not exist, except by assumption.

Given the above result, plus the fact that Full BEKK is frequently estimated in practice and is incorporated in many commercial econometric statistical packages, the focus in this paper is to explore whether there is any evidence of bias in the coefficients estimated in Full BEKK, as compared with DBEKK. We use DBEKK as a benchmark because the mathematical and statistical conditions of DBEKK have been established.

We conduct simulations generating financial return series and use them as inputs to estimate both DBEKK and Full BEKK, from which we can compare the estimates of the conditional variances and covariances. The null hypothesis is that the two sets of estimates should not differ systematically. The method of generating the simulated financial return series is discussed below.

### 2.2.3. Simulated return series

We use the three financial return series for Google, IBM, and Microsoft, and draw on code from several packages in the R library to randomly sample the original time series in blocks of five lags to ensure that we retain the autocorrelation structures to maintain the presence of ARCH effects. The `timeSeries`, `boot` and `meboot` packages prove to be of interest.

Plots of the simulated series based on the three stocks are shown in Figure 3, and the descriptive statistics of the simulated financial return series are shown in Table 3. The simulations have similar characteristics to the base series. IBM has the lowest excess kurtosis of the set of simulations, as would be expected, given that the IBM original return series had relatively low kurtosis in relation to the other two series.

Figure 3: Simulated time series of returns

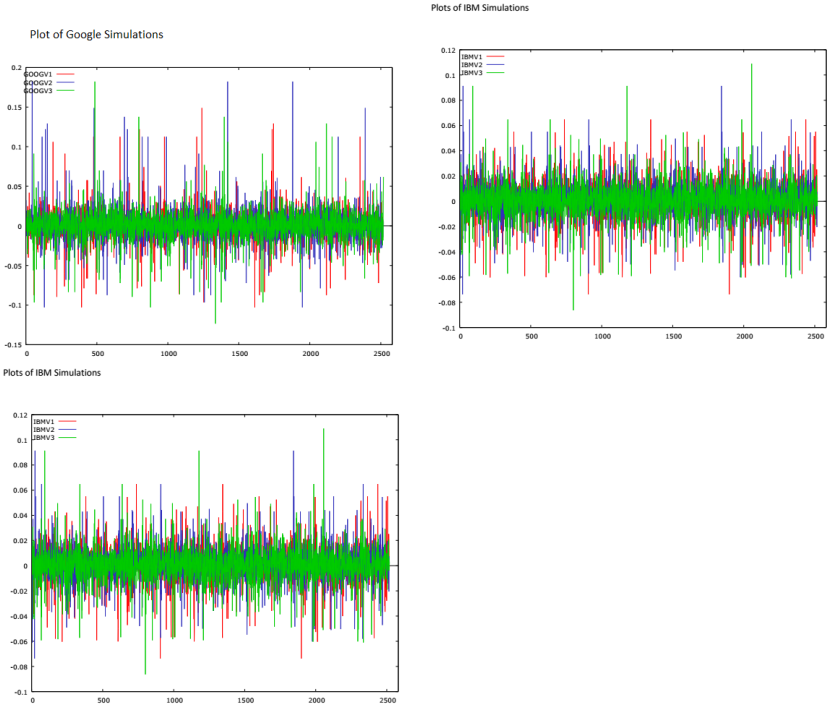


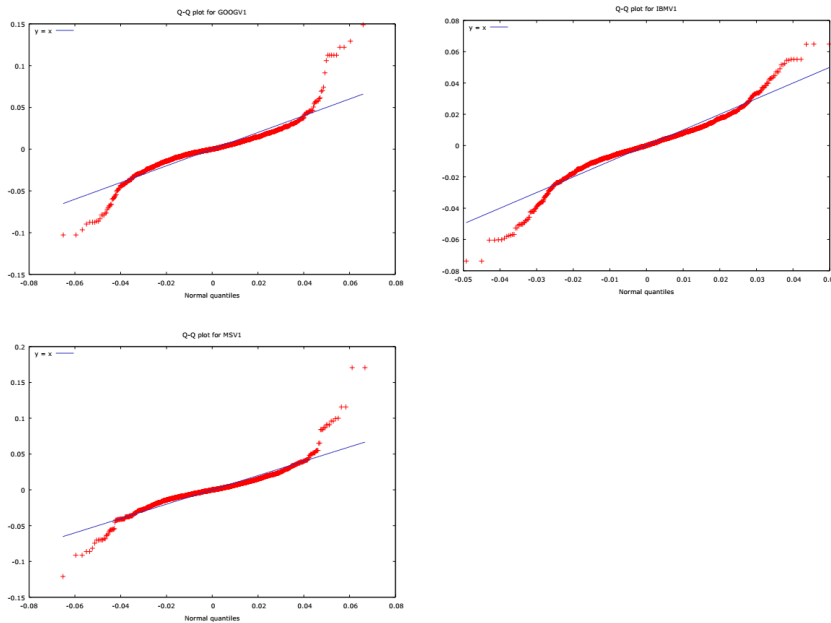
Table 3: Descriptive Statistics of the Simulated Series

	Median	Mean	St. Dev	Skewness	Ex. Kurtosis
Google Sim 1	0.000123242	0.000434784	0.0185023	0.387022	10.1534
Google Sim 2	0.000105152	0.000944741	0.0200021	1.68098	15.8985
Google Sim 3	0.000167354	0.000493751	0.0185772	0.354081	12.2140
IBM Sim 1	0.000312301	0.000332920	0.0139928	-0.308625	3.76076
IBM Sim 2	0.00103512	0.000972290	0.0136369	0.101544	4.52981
IBM Sim 3	0.000422372	0.000472666	0.0140378	-0.0305346	6.05965
Microsoft Sim 1	0.000339290	0.000732968	0.0186249	0.797044	10.8807
Microsoft Sim 2	0.000336076	0.000118718	0.0172592	-0.185154	9.83281
Microsoft Sim 3	7.86705e-005	0.000284365	0.0195287	0.312275	11.3979

Augmented Dickey Fuller tests confirmed that all the simulated series are stationary, and all displayed highly significant ARCH effects.

The simulated series have fat tails and are not Gaussian, as the QQplots for the first set of simulations of the three series, as shown in Figure 4 reveal.

Figure 4: QQplots of Series 1 of the Simulations



2.3. Tests and optimisation prodedures

Empirical estimation of the DBEKK and Full BEKK models was based on the Estima RATS (version Pro 9) econometric software (<https://www.estima.com/>).

In the empirical analysis which follows, we report the results of fitting the multivariate GARCH models, DBEKK and Full BEKK, to the simulated financial return series. This type of estimation procedure involves seeking a solution to an unconstrained minimization problem:

$$\min_{\theta} f(x, \theta), \quad (8)$$

where  $x$  are the data and  $\theta$  is a vector of models to be chosen to minimise the objective function  $f(x, \theta)$ . In the case of GARCH models, this will be a negative log-likelihood function. Typically, no closed-form expression for  $f(x, \theta)$  is available or for its partial derivatives, and so the solution minimisation process is usually achieved by the use of numerical methods. Christensen et al. [5] observe that there are two broad approaches to the construction of a minimisation algorithm, namely methods that rely on function values, or algorithms that use the derivatives of the function.

If we use an algorithm based on gradients to minimise  $f(x, \theta)$ , then we are assuming that all first and second derivatives exist. The gradient vector,  $G(\theta)$ , and the Hessian matrix,  $H(\theta)$ , of the function,  $f(x, \theta)$ , can be defined as:

$$G(\theta) = \frac{\delta f(x, \theta)}{\delta \theta}, \quad H(\theta) = \frac{\delta^2 f(x, \theta)}{\delta \theta \delta \theta'}. \quad (9)$$

The minima of the objective function occur at parameter values where the gradient is zero and the associated Hessian matrix is positive definite. The estimator,  $\hat{\theta}$ , of the parameter vector,  $\theta$ , should satisfy the condition:

$$G(\hat{\theta}) = 0. \quad (10)$$

We can start with a guess,  $\hat{\theta}_k$ , which is assumed to be near the optimal value  $\theta$  at which a minimum is attained. A Taylor series expansion of  $G(\hat{\theta})$  about  $\hat{\theta}_k$  is given as:

$$G(\hat{\theta}) = G(\hat{\theta}_k + (\hat{\theta} - \hat{\theta}_k)) = G(\hat{\theta}_k) + H(\hat{\theta}_k)(\hat{\theta} - \hat{\theta}_k) + 0(\hat{\theta} - \hat{\theta}_k)^2.$$

If we replace  $G(\hat{\theta})$  in equation (10) by the previous expression (9), and ignore all terms of order two and above, it follows that:

$$\hat{\theta} \approx \hat{\theta}_k - H(\hat{\theta}_k)^{-1}G(\hat{\theta}_k).$$

Using this result, the next guess for  $\hat{\theta}$  is:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - H(\hat{\theta}_k)^{-1}G(\hat{\theta}_k). \quad (11)$$

This is referred to as a 'full Newton step' which would be taken close to the location of the minimum. However, further away from the minimum, this step may not be guaranteed to reduce the value of the function, so the usual convention is to use a smaller step:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \alpha_k H(\hat{\theta}_k)^{-1}G(\hat{\theta}_k), \quad (12)$$

where  $\alpha_k$  is chosen to control the step size and to ensure that the function is reduced at each iteration.

All gradient-based algorithms employ the general iterative scheme set out in equation (12), and differ only in their approximation of the Hessian matrix at each iteration. Christensen et al. [5] point out that a Newton-Raphson procedure computes the Hessian matrix directly, the Method of Scoring uses the Information matrix (negative of the value of the Hessian matrix), and the BHHH algorithm (Berndt et al. [2]) approximates the Hessian by the outer product of the gradient vector. BHHH provides an approximation of the Hessian matrix that is guaranteed to be positive definite, and so is a popular choice in many econometric software packages.

The estimation process used in this paper, is BHHH, but there are several caveats. Christensen et al. [5] note that the treatment of constraints on parameters, choice of starting values, specification of termination criteria, and analytical versus numerical gradients, can materially alter the final output of a minimisation algorithm. A subsequent comment on this paper by McCullough [12] suggests that default options for a nonlinear solver are not likely to produce a correct answer, and that the answer produced by a nonlinear solver is not necessarily correct.

McCullough and Vinod [14] question the baseline accuracy of many commonly used econometric software packages, and note that there is often a trade-off between computational speed and statistical accuracy. McCullough and Renfro [13] explore the interaction between benchmarks, software standards, and econometric theory, using the GARCH model as a case study, and caution against the uncritical use of standard econometric packages.

Despite these various issues, the paper adopts a consistent framework in the estimation methods to compare DBEKK with Full BEKK. We rely on the

estimation procedures in RATS and the BHHH algorithm to fit the two models, which are then used to predict the daily condition variances and covariances generated. We then use quantile regression to analyse the two sets of estimates using the DBEKK estimates as the baseline benchmark.

#### 2.4. Quantile Regression

Quantile Regression is modelled as an extension of classical OLS (Koenker and Bassett [9]). In Quantile Regression, the conditional mean as estimated by OLS is extended to similar estimation of an ensemble of models of various conditional quantile functions for a data distribution. Therefore, Quantile Regression can quantify the conditional distribution of  $(Y|X)$  better. The central special case is the median regression estimator that minimises a sum of absolute errors.

Estimates of the remaining conditional quantile functions are obtained by minimizing an asymmetrically weighted sum of absolute errors, where the weights are functions of the quantiles of interest. This makes Quantile Regression a robust technique, even in the presence of outliers. Taken together, the ensemble of estimated conditional quantile functions of  $(Y|X)$  offers a much more complete view of the effect of covariates on the location, scale and shape of the distribution of the response variable.

For parametric estimation in Quantile Regression, quantiles as proposed by Koenker and Bassett [9] can be defined through an optimisation problem. In order to solve an OLS regression problem, a sample mean is defined as the solution of the problem of minimising the sum of squared residuals, in the same way the median quantile (0.5%) in Quantile Regression is defined through minimising the sum of absolute residuals. The symmetrical piecewise linear absolute value function assures the same number of observations above and below the median of the distribution.

The other quantile values can be obtained by minimizing a sum of asymmetrically weighted absolute residuals, giving different weights to positive and negative residuals. Solving the following:

$$\min_{\xi \in R} \sum \rho_{\tau}(y_i - \xi), \quad (13)$$

where  $\rho_{\tau}(\cdot)$  is the tilted absolute value function, which gives the  $\tau$ th sample quantile with a solution. Taking the directional derivatives of the objective

function with respect to  $\xi$  (from left to right) shows that this problem yields the sample quantile as its solution.

After defining the unconditional quantiles as an optimisation problem, it is easy to define the conditional quantiles similarly. Taking the least squares regression model as a base to proceed, for a random sample,  $y_1, y_2, \dots, y_n$ , solving:

$$\min_{\mu \in R} \sum_{i=1}^n (y_i - \mu)^2, \quad (14)$$

gives the sample mean as an estimate of the unconditional population mean,  $EY$ . Replacing the scalar,  $\mu$ , by a parametric function,  $\mu(x, \beta)$ , and then solving:

$$\min_{\mu \in R^p} \sum_{i=1}^n (y_i - \mu(x_i, \beta))^2, \quad (15)$$

gives an estimate of the conditional expectation function,  $E(Y|x)$ .

Proceeding in the same way for Quantile Regression, in order to obtain an estimate of the conditional median function, the scalar,  $\xi$ , in the first equation is replaced by the parametric function,  $\xi(x_i, \beta)$ , and  $\tau$  is set to  $1/2$ . The estimates of the other conditional quantile functions are obtained by replacing absolute values by  $\rho_\tau(\cdot)$ , and solving:

$$\min_{\mu \in R^p} \sum \rho_\tau(y_i - \xi(x_i, \beta)). \quad (16)$$

The resulting minimization problem, when  $\xi(x, \beta)$  is formulated as a linear function of parameters, can be solved efficiently by linear programming methods. We use quantile regression to compare the relative behaviour of the conditional variances across the quantiles, as predicted by the two models.

### 3. Empirical Results

We estimated both DBEKK and Full BEKK using the simulated financial return series. The estimates from DBEKK are used as the benchmark, given that it has established statistical regularity conditions. We decided to keep the comparison tests as simple as possible, and first estimated a two-variable version of the DBEKK and Full BEKK models using the the three sets of simulated return series in pairs. This was then followed by a single three-variable set of



estimates, to verify that the same pattern of results exists. The null hypothesis is that Diagonal and Full BEKK are equivalent when the off-diagonal coefficients in Full BEKK are zero, so the asymptotic tests are statistically valid. We proceeded by estimating the coefficients for the conditional variances and the conditional covariances for the two models, and then used non-parametric sign tests on the differences between the two sets of estimates.

The estimates of the constants, ARCH effects and conditional variances for the two models are shown in Table 4. DBEKK and Full BEKK fitted to the pairs of simulated series were highly significant (bearing in mind that asymptotic theory is not valid for the QMLE of Full BEKK), and all but three pairs of the fifty-four coefficients estimated in the models, and presented in Table 4, were significant at the 1% level. (The insignificant coefficients are marked with an asterisk (\*) in Table 4.) The coefficients of the conditional covariances are shown in Table 5. The majority of these estimates are insignificant, so we concentrated our analysis on the conditional variances.

We then undertook a set of non-parametric sign tests on the values of the estimated coefficients, reported in Table 4. We ran the tests in a number of different formats, both on the full set of coefficients reported in Table 4, and the full set minus the three pairs of insignificant estimates. The results of the sign tests are reported in Tables 6 and 7, which suggest that there are no significant differences in the values of the coefficients for the constants, ARCH effects and the conditional variances estimated for the two variables. However, these tests treat the coefficients in isolation, and regard them as being independent, which is not the case when they are incorporated into a DBEKK or Full BEKK model.

Table 4: Variable Coefficient Estimates for DBEKK and Full BEKK

	DBEKK						Full BEKK							
	C(1,1)	C(2,2)	A(1)	A(2)	B(1)	B(2)	C(1,1)	C(2,2)	A(1,1)	A(1,2)	A(2,1)	A(2,2)	B(1,1)	B(2,2)
IBMV1, GOOGV1	0.00893	0.01681	0.3773	0.46698	0.67011	0.01026*	0.00894	0.01681	0.37976	0.02095*	0.00392*	0.46380	0.66628	0.01287*
IBMV2, GOOGV2	0.00739	0.01747	0.32571	-0.59969	0.77078	0.02727*	0.00746	0.01570	0.30726	0.13081	-0.01782*	0.59406	0.77047	-0.36849
IBMV3, GOOGV3	0.008325	0.01185	0.47073	0.32044	0.65976	0.70551	0.01269*	0.00002*	0.46101	-0.04771*	0.01276*	0.18860	0.67095	0.81205
GOOGV1, MSV1	0.013884	0.01253	0.36731	0.47581	0.056591	0.58379	0.01665	0.01116	0.43057	0.07513	-0.10471	0.43840	0.10125*	0.646048
GOOGV2, MSV2	0.01651	0.011449	-0.47029	0.53336	-0.39279	0.53336	0.01366	0.00003	0.041332	0.07426	0.09744*	0.76597	0.84413	0.32680
GOOGV3, MSV3	0.00835	0.011846	0.47073	0.32044	0.65976	0.70551	0.00752	-0.01070	0.44772	0.06484*	0.02899*	0.30716	0.70504	0.73385
IBMV1, MSV1	0.00906	0.00906	0.35582	0.55170	0.67042	0.49239	0.00909	0.01282	0.36532	0.06623	-0.03269*	0.53945	0.66093	0.48406
IBMV2, MSV2	0.00760	0.01170	0.32290	0.57413	0.76189	0.48309	0.00738	0.01145	0.31417	0.07554	-0.03147	0.52961	0.76531	0.50461
IBMV3, MSV3	0.00830	0.01263	0.421447	0.52233	0.69487	0.585204	0.00797	0.01253	0.42744	-0.06357	0.01946*	0.53773	0.69503	0.56645

NB: All the estimated coefficients are significant at the 1% level except those indicated by \*.

Table 5: Conditional Covariance Estimates for DBEKK and Full BEKK

	DBEKK	Full BEKK				
	C(2,1)	C(2,1)	A(1,2)	A(2,1)	B(1,2)	B(2,1)
IBMV1, GOOGV1	-0.00011*	-0.00052*	0.02095*	0.00392*	0.02271*	-0.04056*
IBMV2, GOOGV2	0.00153	0.00003*	0.13081	-0.01782*	0.04062*	0.05223*
IBMV3, GOOGV3	0.00043*	0.02107*	-0.04771*	0.01276*	0.10412*	-0.28249*
GOOGV1, MSV1	-0.00024*	-0.00168	0.07513	-0.10471	-0.16666	0.13603
GOOGV2, MSV2	-0.00019*	-0.01597*	0.07426*	0.09744	0.34576*	-0.33889
GOOGV3, MSV3	0.00043*	0.00337*	-0.05325*	0.00331*	0.09699*	-0.12799*
IBMV1, MSV1	0.00019*	0.00226	0.06623	-0.03270*	-0.20763	0.04362*
IBMV2, MSV2	-0.00005*	-0.00018*	0.07555	-0.03148	-0.08764*	0.09885
IBMV3, MSV3	0.000208*	-0.00223*	-0.06358	0.01947	-0.02675*	-0.09230

NB: All the estimated coefficients are significant at the 1% level except those indicated by \*.

Table 6: Sign Tests of the Differences in Estimates for DBEKK and FULL BEKK: All Coefficients

Sign Test for difference between DBEKK and BEKK
Number of differences $n = 53$
Number of cases $DBEKK > BEKK$ : $W=28$ (52.83%)
Under the null hypothesis of no difference, $W$ follows $Bin(53, 0.5)$
Probability ( $W \leq 28$ ) = 0.708434
Probability ( $W \geq 28$ ) = 0.391923
Sign Test for difference between DBEKK and BEKK, Wilcoxon Signed-Rank Test
Number of differences $n = 53$
Null hypothesis: median difference is zero
$W_+ = 739$ , $W_- = 745$
(0 differences: 1, non-zero ties: 0)
Expected value = 742
Variance = 13488.5
$Z = -0.0215257$
$P(Z < -0.0215257) = 0.491413$
Two-tailed p-value = 0.982826

Table 7: Sign Tests of Differences in Estimates for DBEKK and FULL BEKK: Significant Coefficients

Sign Test for difference between DBEKK and BEKK
Number of differences $n = 48$
Number of cases $DBEKK > BEKK$ : $W=26$ (54.17%)
Under the null hypothesis of no difference, $W$ follows $Bin(48, 0.5)$
Probability ( $W \leq 26$ ) = 0.76456
Probability ( $W \geq 26$ ) = 0.332733
Sign Test for difference between DBEKK and BEKK, Wilcoxon Signed-Rank Test
Number of differences $n = 48$
Null hypothesis: median difference is zero
$W_+ = 618$ , $W_- = 606$
(0 differences: 1, non-zero ties: 0)
Expected value = 612
Variance = 10106
$Z = 0.0547108$
$P(Z < 0.0547108) = 0.478184$
Two-tailed p-value = 0.956369

Table 8: Sign Tests of Differences in the Predicted Conditional Variances of DBEKK and FULL BEKK

Security	Number of differences	Z score	Probability
IBMV1/GOOGV1	2518 (IBMV1CV)	$z = 8.66177$	$P(Z > 8.66177) = 0$
IBMV1/GOOGV1	2518(GOOGV1CV)	$z = 8.66177$	$P(Z > 8.66177) = 0$
IBMV2/GOOGV2	2518(IBMV2CV)	$z = -70.7228$	$P(Z < -70.7228) = 0$
IBMV2/GOOGV2	2518(GOOGV2CV)	$z = -17.8438$	$P(Z < -17.8438) = 1.61378e-071$
IBMV3/GOOGV3	2518(IBMV3CV)	$z = -178.968$	$P(Z < -178.968) = 0$
IBMV3/GOOGV3	2518(GOOGV3CV)	$z = -178.682$	$P(Z < -178.682) = 0$
IBMV1/MSV1	2518(IBMV1CV)	$z = 10.0413$	$P(Z > 10.0413) = 0$
IBMV1/MSV1	2518(MSV1CV)	$z = 21.2183$	$P(Z > 21.2183) = 0$
IBMV2/MSV2	2518(IBMV2CV)	$z = 27.6135$	$P(Z > 27.6135) = 0$
IBMV2/MSV2	2518(MSV2CV)	$z = 9.78013$	$P(Z > 9.78013) = 0$
IBMV3/MSV3	2518(IBMV3CV)	$z = 25.524$	$P(Z > 25.524) = 0$
IBMV3/MSV3	2518(MCV3CV)	$z = 41.8977$	$P(Z > 41.8977) = 0$
GOOGV1/MSV1	2518(GOOGV1CV)	$z = 4.39668$	$P(Z > 4.39668) = 5.49587e-006$
GOOGV1/MSV1	2518(MSV1CV)	$z = -21.4004$	$P(Z < -21.4004) = 6.62768e-102$
GOOGV2/MSV2	2518(GOOGV1CV)	$z = -169.726$	$P(Z < -169.726) = 0$
GOOGV2/MSV2	2518(MSV2CV)	$z = -178.718$	$P(Z < -178.718) = 0$
GOOGV3/MSV3	2518(GOOGV3CV)	$z = 36.0083$	$P(Z > 36.0083) = 0$
GOOGV3/MSV3	2518(MSV3CV)	$z = -17.9135$	$P(Z < -17.9135) = 4.62876e-072$

DBEKK and Full BEKK are multivariate GARCH models which are used for forecasting conditional volatility. The crucial issue for purposes of risk management is how the forecasts of conditional volatility derived from the two models are compared. These are vital components for assessing risk, and might be used to compute the Value-at-Risk (VaR) of a portfolio of financial assets, for example.

The simulated financial return samples for the nine variables contain ten years of daily data, or 2581 data points. We filter these through the DBEKK and Full BEKK models, and obtain corresponding estimates of the conditional variance projections, for each simulated security, from the two models. These forecasts of the conditional variances are then compared using non-parametric sign tests. The results for each simulated security are shown in Table 8.

The sign tests in Table 8 are based on the null hypothesis that the median difference in the conditional variances produced by the two models, DBEKK and Full BEKK, for the simulated securities, is zero. The null hypothesis is strongly rejected in all cases, and the differences are highly significant. We also ran sign tests, not reported, based on the null hypothesis that there was no difference in the conditional variances predicted by the two models. These results also strongly rejected the null hypothesis in all cases.

While it is valuable to know that the two models produce different predictions of the conditional variances, it is also of interest to determine if there are *systematic differences* in the predictions of the conditional variances. We

Table 9: Quantile Regression of Predicted Conditional Variances for Full BEKK

Security	Quantile slope coefficients				
	0.05	0.25	0.50	0.75	0.95
IBMV1/GOOGV1(IBMVCV)	0.970853	0.995444	1.00706	1.01274	1.02605
IBMV1/GOOGV1(GOOGCV)	0.952430	0.974888	0.987685	0.998475	1.01715
IBMV2/GOOGV2(IBMVCV)	0.862003	0.894723	0.913300	0.930500	0.966278
IBMV2/GOOGV2(GOOGV2CV)	0.860421	0.962226	0.993431	1.02805	1.15073
IBMV3/GOOGV3(IBMVCV) <sup>1</sup>	0.328928	0.328928	0.416581	0.462995	0.532234
IBMV3/GOOGV3(GOOGV3CD)	0.922288	0.956421	0.985430	1.01335	1.01855
IBMV1/MSV1(IBMVCV)	0.949942	1.01037	1.03986	1.06483	1.12862
IBMV1/MSV1	0.854267	0.931472	0.959137	0.989454	1.00981
IBMV2/MSV2(IBMVCV)	0.861948	0.951856	0.987461	1.03175	1.13855
IBMV2/MSV2(MSV2CDV)	-49.6363	-5.29685	-1.97048	9.20840	18.2802
IBMV3/MSV3(IBMVCV)	1.00633	1.05196	1.07648	1.09911	1.15981
IBMV3/MSV3(MSV3CV)	0.940177	0.998115	1.03939	1.06727	1.12067
GOOGV1/MSV1(GOOGV1CV)	0.0831902	0.660919	1.22251	1.36371	1.52337
GOOGV1/MSV1(MSV1CV)	0.753483	0.837598	0.900926	1.06852	1.20022
GOOGV2/MSV2(GOOGV2CV) <sup>1</sup>	0.00744599	0.00762628	0.00729433	0.0188172	-0.0147150
GOOGV2/MSV2(MSV2CV)	0.977758	1.58318	1.92646	2.03505	2.17220
GOOGV3/MSV3(GOOGV3CV) <sup>2</sup>	0.911108	0.921100	0.954618	1.03009	1.09428
GOOGV3/MSV3(MSV3CV)	0.867499	0.917009	0.955980	0.988218	1.01098

Notes: All the slope coefficients across the quantiles estimated using robust errors are significant at the 1% level

1. Problems with convergence encountered in Full BEKK model.

2. Estimation of the Full BEKK model in this case failed to converge using BHHH, so we switched to BFGS

explore this issue by means of quantile regression. The advantage of quantile regression is that we can explore the relationship between the two sets of predictions from DBEKK and Full BEKK at particular quantiles. We regress the predicted conditional variances from the Full BEKK model on the corresponding predictions from DBEKK for each of the simulated securities, in the pairs of securities modelled. We treat the predictions of conditional variances from the Full BEKK model as the dependent variable. The results of these quantile regressions are shown in Table 9.

Table 9 reveals a distinct pattern of an increase in the slope coefficients as we move across the quantiles from the lowest 0.05 quantile to the highest 0.95 quantile. The most extreme case is the prediction of the conditional variances for the relationship between IBMV2 and MSV2. In the 0.05 quantile, the conditional variance prediction by Full BEKK is 50 times lower than for DBEKK, and in the 0.95 quantile, it is 18 times higher (though there were convergence issues encountered in Full BEKK estimation in this case). Even so, the difference across these two extreme quantiles usually varies by between 10 and 20 percent. This is still very large if we intend to use the models to predict a portfolio VAR.

If we use the predictions of DBEKK as the benchmark, then application of the Full BEKK model to the same data set may underestimate risk in the lowest quantile and overestimate risk in the highest quantile. In the 18 examples, 7 of the total estimates suggest risk in the 0.05 quantile will be lower by 10 percent or more when estimated by Full BEKK, as opposed to DBEKK. Similarly, in 9 of the total cases, the estimate of risk in the 0.095 quantile is 10 per cent or more, when estimated by Full BEKK, as compared with DBEKK. Thus, there are considerable discrepancies in the predictions of conditional volatility based on these applications of the two models.

These discrepancies in the regression slope coefficients are apparent in Figures 5 and 6, which present graphs of the estimated slope coefficients across the quantiles for the pairs of simulated securities considered. The quantile regression bounds estimated at the 0.95 level are shown around the quantile slope estimates in each figure. The horizontal lines, in the centre of the figures, show the ordinary least squares regression slopes for the regressions of conditional variances for each security, regressed on the conditional variances for the same security, when considered in the same pairwise estimates produced by DBEKK. The ordinary least squares slope coefficients are not very informative, and merely suggest whether the predicted conditional variances from Full BEKK are relatively above or below those from DBEKK. There is considerable variation in the figures, but most of them are slightly below one.

The quantile regression analyses are much more informative. The lines in Figures 5 and 6 link the slope coefficients estimated at the 0.05, 0.25, 0.50, 0.75, and 0.95 quantiles, when the predictions of the conditional variances from Full BEKK are regressed on those from DBEKK. In all cases, bar one, shown in Figures 5 and 6, the estimates at the lowest 0.05 quantile reveal a relationship between the two sets of estimates that is markedly different from that suggested by ordinary least squares, which captures the average relationship. The relationship is markedly different, at this quantile, frequently by ten to twenty percent.

Another startling feature is that all the slopes depicted in Figures 5 and 6 are strongly positive, in that the estimated slope coefficients all increase, with one exception, from the lowest to the highest quantile. Thus, the conditional variances estimated from the Full BEKK model are much higher, at the 0.95 quantile, often by 20 per cent or more, than the conditional variance estimated by the DBEKK model.

These results have strong implications if we try use the two multivariate models to estimate portfolio risk. The analysis reported in the paper, on these simulated financial return series, suggests that the use of the Full BEKK model will underestimate conditional variances in the left-hand tail of the portfolio return distribution, relative to DBEKK, and overestimate it in the right-hand tail of the distribution.

These results are subject to certain caveats. We have estimated the models using the Estima RATS econometric package, and used the default settings when fitting the models. We have not changed any of the tolerances in the algorithms used to fit the models, or changed the settings for the initialization

Figure 5: Quantile Regression Slope Coefficients of Conditional Variances from Full BEKK Regressed on DBEKK Estimates (first 10 pairs)

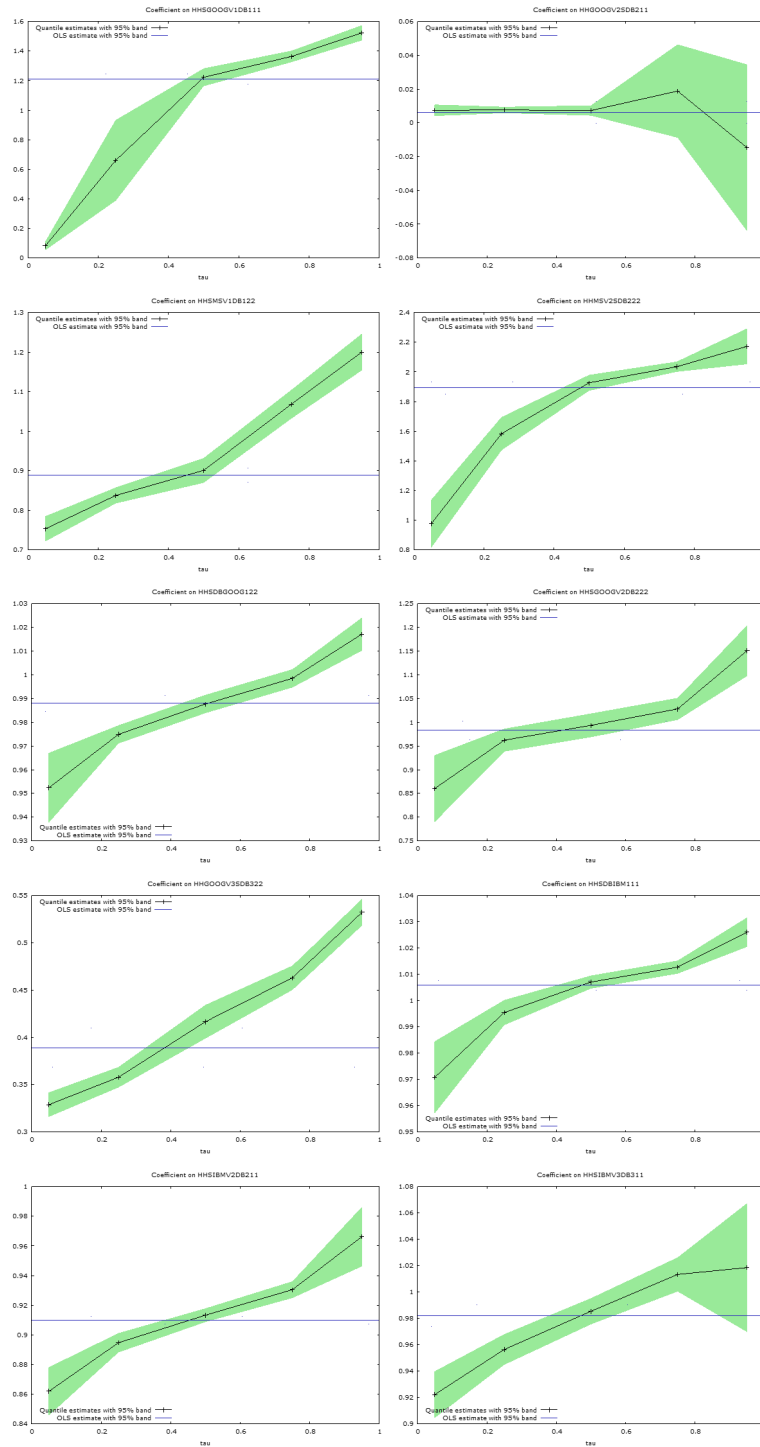




Figure 6: Quantile Regression Slope Coefficients of Conditional Variances from Full BEKK Regressed on DBEKK Estimates (remaining pairs)

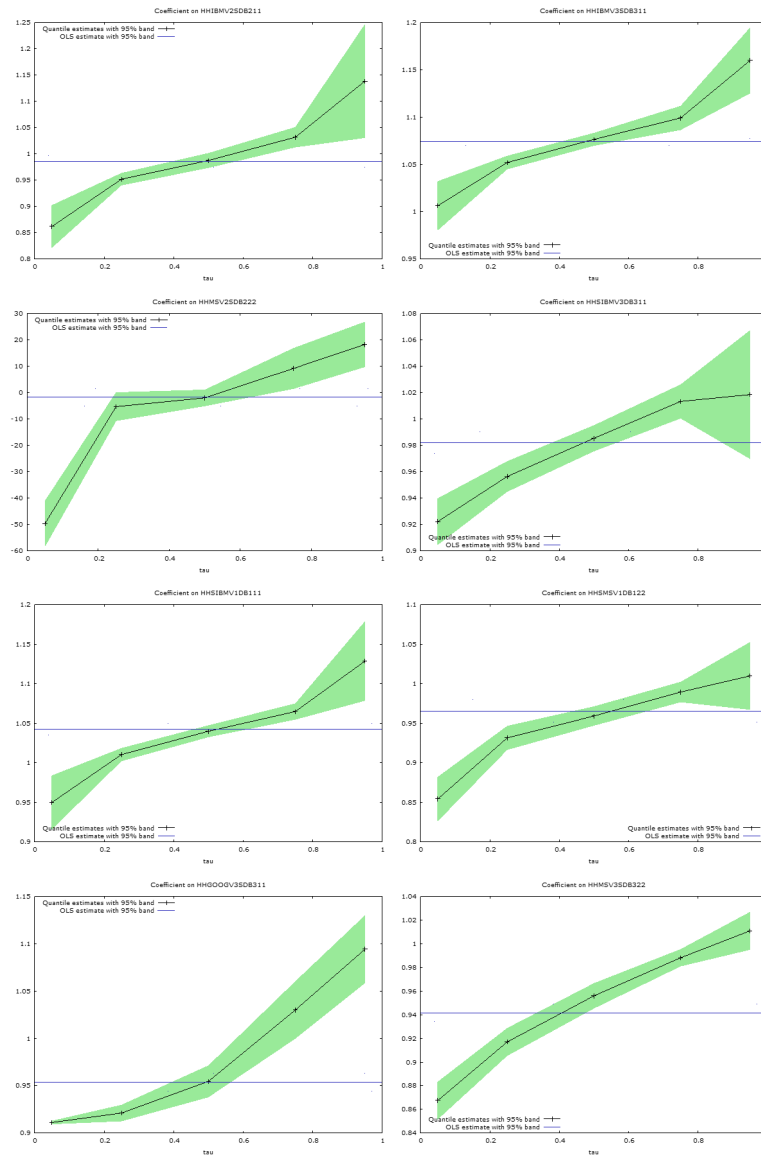


Table 10: Variable Coefficient Estimates for DBEKK and Full BEKK: GOOGV1, IBMV1, MSV1

Coefficients	DBEKK	Full BEKK
C(1,1)	0.01226	0.02926
C(2,1)	-0.00029*	-0.00119*
C(2,2)	0.01312	0.00083*
C(3,1)	-0.000001*	0.00364*
C(3,2)	0.00011*	0.00785*
C(3,3)	0.00902	0.00747*
A(1)	0.28346	0.47788
A(2)	0.48530	0.48004
A(3)	0.34639	0.36256
B(1)	0.69899	0.28255*
B(2)	0.54312	0.80864
B(3)	0.67893	0.78892
A(1,2)	-	0.26978
A(1,3)	-	0.02999*
A(2,1)	-	-0.03014*
A(2,3)	-	-0.01886*
A(3,1)	-	0.11740*
A(3,2)	-	0.17046*
B(1,2)	-	-0.42607*
B(1,3)	-	-0.09707*
B(2,1)	-	0.00375*
B(2,3)	-	0.01354*
B(3,1)	-	-0.15853*
B(3,2)	-	-0.16545*

NB: All the coefficients are significant at the 1% level except those indicated by \*.

of the algorithms used to commence the models. We have also instructed the program to use the BHHH optimization procedure to fit the models. All the models have been fitted using a Gaussian distribution, and the estimates would be different if we used a t-distribution. (We also did some analysis using the t-distribution, which are not reported in the paper, that revealed a virtually identical pattern of relationships across the quantiles, to those reported in the paper). The intention was to use a consistent approach to the fitting of the models, and then to explore the consistency of the results.

We also estimated DBEKK and Full BEKK using three variables jointly, namely GOOGV1, IBMV1, and MSV1, to check that similar behaviour was displayed when we used a trivariate estimation procedure. The results are shown in Table 10.

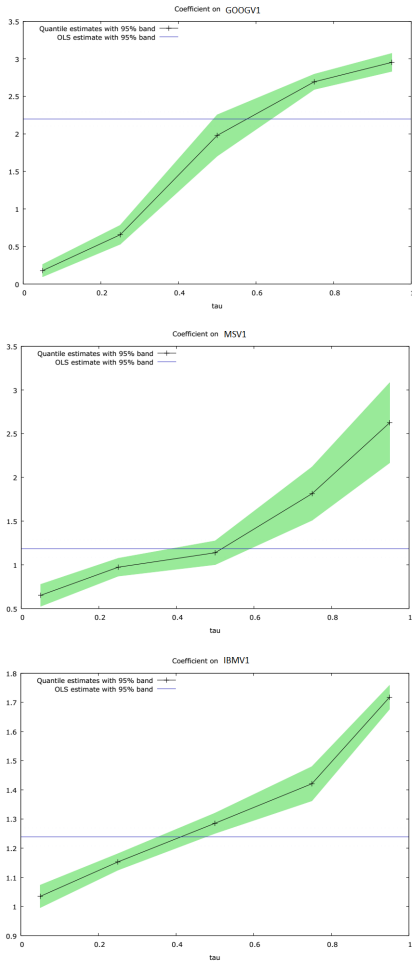
It is evident from Table 10 that many of the additional terms included in the Full BEKK model are not statistically significant (bearing in mind that asymptotic theory is not valid for the QMLE of Full BEKK), at least in this

Table 11: Quantile Regression Results, Predicted Conditional Variances from Full BEKK as the dependent variable: GOOGV1, MSV1, IBMV1

Security	Quantile				
	0.05	0.25	0.5	0.75	0.95
GOOGV1 Conditional Variance	0.178998	0.656654	1.97967	2.69374	2.95327
MSV1 Conditional Variance	0.651730	0.973131	1.13849	1.81488	2.62494
IBMV1 Conditional Variance	1.03471	1.15279	1.28522	1.42100	1.71742

NB: All the coefficients are significant at the 1% level

Figure 7: Quantile regression slope coefficient estimates of conditional variance from Full BEKK regressed on DBEKK estimates for GOOGV1, MSV1, IBMV1.



simulated data set . We also ran a quantile regression analysis of the condi-

tional variances produced by Full BEKK, regressed on the conditional variances produced by DBEKK, using these three securities. The results are shown in Table 11. Plots of the quantile regression slope coefficients are shown in Figure 7.

It can be seen in Table 11 and from Figure 7, that the same pattern of results emerges when we estimate DBEKK and Full BEKK using three securities jointly, as in the previous case with pairs of securities. The conditional variance estimates for Full BEKK, relative to DBEKK, are comparatively lower in the 0.05 quantile, increase across the quantiles, and are relatively higher in the 0.95 quantile.

We used the DBEKK model as a benchmark, given that the mathematical and statistical properties of this model have been established. The results, using this benchmark suggest that there is an observable and relative bias in the predictions of the conditional variances from the Full BEKK model. This has serious practical implications about the use of Full BEKK for risk management and modelling purposes.

#### 4. Conclusion

This paper explored the relative biases in the estimation of the Full BEKK model, as compared with the Diagonal BEKK model, which is used as an empirical benchmark. Chang and McAleer [4] showed that univariate GARCH is not a special case of multivariate GARCH, specifically the Full BEKK model, and demonstrate that Full BEKK which, in practice, is estimated almost exclusively in the literature, has no underlying stochastic process, regularity conditions, or asymptotic properties. Diagonal BEKK (DBEKK) does not suffer from these limitations, and hence provides a suitable benchmark.

We used simulated financial returns series to contrast the estimates of the conditional variances from DBEKK and Full BEKK. The results of the non-parametric tests on their values shows evidence of considerable bias in the Full BEKK estimates relative to those of DBEKK. The results of quantile regression analysis showed there was a systematic relationship between the two sets of estimates as we moved across the quantiles. Estimates of conditional variances from Full BEKK, relative to those from DBEKK, are lower in the left tail and higher in the right tail. The phenomenon appears to be all-pervasive in estimates reported in the simulated financial return series. This result has serious practical

implications for the use of Full BEKK as a risk management tool.

### Acknowledgements

For financial support, the first author acknowledges the Australian Research Council, and the second author wishes to thank the Australian Research Council, and the National Science Council, Ministry of Science and Technology (MOST), Taiwan. The authors are grateful to A.K. Singh for R coding suggestions.

### References

- [1] Baba, Y., R.F. Engle D. Kraft, and K.F. Kroner, (1985) Multivariate simultaneous generalized ARCH, unpublished manuscript, Department of Economics, University of California, San Diego, CA, USA.
- [2] Berndt, E., B. Hall, R. Hall, and J Hausman, (1974) Estimation and inference in nonlinear structural models. *Annals of Social Measurement*, 3, 653-665.
- [3] Bollerslev, T., (1986) Generalised autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, 307-327.
- [4] Chang, C.L., and M. McAleer, (2017) The fiction of full BEKK, working paper, TI 2017-015/III Tinbergen Institute Discussion Paper.
- [5] Christensen, T.M., A.S. Hurn, and K.A. Lindsay, (2008) The devil is in the detail: Hints for practical optimisation, *Economic Analysis and Policy*, 38, 345-368.
- [6] Engle, R.F., and K.F. Kroner, (1995) Multivariate simultaneous generalized ARCH, *Econometric Theory*, 11(1), 122-150.
- [7] Engle, R.J., (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, 987-1007.
- [8] Glosten, L.R., R. Jagannathan, and D.E. Runkle, (1993) On the relation between the expected value and volatility of nominal excess return on stocks, *Journal of Finance*, 48(5), 1779-1801.
- [9] Koenker, R., and G. Bassett, (1978) Regression quantiles, *Econometrica*, 46, 33-50.
- [10] McAleer, M., F. Chan, S. Hoti, and O. Lieberman, (2008) Generalized autoregressive conditional correlation, *Econometric Theory*, 24(6), 1554-1583.

- [11] McAleer, M., and C. Hafner, (2014) A one line derivation of EGARCH, *Econometrics*, 2(2), 92- 97.
- [12] McCullough, B.D., (2009) The devil is in the detail: Hints for practical optimisation: A comment, *Economic Analysis and Policy*, 39: 155-157.
- [13] McCullough, B.D., and C.G. Renfro, (2000) Some numerical aspects of non-linear estimation, *Journal of Economic and Social Measurement*, 26 63-77.28
- [14] McCullough, B.D., and H. Vinod, (1999) The numerical reliability of econometric software, *The Journal of Economic Literature*, 37, 633- 665.
- [15] Nelson, D.B., (1990) ARCH models as diffusion approximations, *Journal of Econometrics*, 45(1- 2), 7-38.
- [16] Nelson, D.B., (1991) Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, 59(2), 347-370.
- [17] Tsay, R., (1987) Conditional heteroscedastic time series models, *Journal of the American Statistical Association*, 82(398), 590-604.