On the efficacy of techniques for evaluating multivariate volatility forecasts

Adam Clements
Mark Doolan
Stan Hurn
Ralf Becker

Working Paper #41
February 2009
On the efficacy of techniques for evaluating multivariate volatility forecasts

A Clements, M Doolan, S Hurn
School of Economics and Finance, Queensland University of Technology

R Becker
Economic Studies, University of Manchester.

Abstract
The performance of techniques for evaluating univariate volatility forecasts are well understood. In the multivariate setting however, the efficacy of the evaluation techniques is not developed. Multivariate forecasts are often evaluated within an economic application such as portfolio optimisation context. This paper aims to evaluate the efficacy of such techniques, along with traditional statistical based methods. It is found that utility based methods perform poorly in terms of identifying optimal forecasts whereas statistical methods are more effective.

Keywords
Multivariate volatility, forecasts, forecast evaluation, Model confidence set

JEL Classification Numbers
C22, G00.

Corresponding author
Adam Clements
School of Economics and Finance
Queensland University of Technology
Brisbane, 4001
Qld, Australia
email a.clements@qut.edu.au

Acknowledgements The authors thank Andrew Patton for helpful comments regarding the preliminary results.
1 Introduction

Forecasts of volatility are important inputs into numerous financial applications. These include, derivative pricing, risk estimation and portfolio allocation among others. An extensive number of univariate volatility models have been developed from the seminal works of Engle (1982) and Bollerslev (1986). For an overview of univariate models see for instance, Campbell, Lo and MacKinlay (1997) and Gourieroux and Jasiak (2001). Poon and Granger (2003, 2005) provide a wide ranging survey of the forecast performance of various univariate techniques. For applications such as portfolio allocation, forecasts of volatility and covariances (or correlations) are required. This need led to the development of multivariate models, the first examples of which were simple extensions of univariate models. For a survey of such models and more recent developments see Andersen, Bollerslev, Christoffersen and Diebold (2006).

Evaluation of univariate volatility forecasts is relatively straightforward and relies on standard forecast evaluation techniques. The only complicating factor when dealing with volatility is that it is inherently latent. In the univariate context, Patton (2006) proves that only certain classes of statistical loss functions are appropriate in the sense that they are robust to noise in the in a volatility proxy. While familiar metrics of forecast accuracy can be utilised in the univariate setting, this is not normally the case in the multivariate setting. Given that multivariate forecasts are inputs into various economic decisions, forecasts are often compared in the context of economic applications such as portfolio allocation. Even within this particular framework, various metrics have been suggested for comparing forecasts. Fleming, Kirby and Ostdiek (2001, 2003) evaluate forecasting methods on the basis of differences between utility earned from selecting portfolios on the efficient frontier given competing forecasts. Such approaches are indirect methods for forecast evaluation and require further assumptions such as expected rates of return, the nature of the distribution of returns and the utility function employed by investors. Engle and Colacito (2006) suggest comparing the volatility of the minimum variance portfolio to differentiate between competing forecasts.

This paper provides an analysis of the efficacy of a range of methods, denoted here generally as loss functions, that have been used to evaluate multivariate volatility forecasts. Extending techniques developed in the univariate context, the ability of each of the loss functions to successfully differentiate between forecasting models is examined. To do so, the properties of various loss functions are examined and a substantial simulation study undertaken revealing the efficacy of the various loss functions to differentiate between competing forecasts. Finally, while not the main focus of this paper, an empirical analysis is also undertaken. Overall it seems as though a likelihood based measure exhibits the most power. Economic evaluation techniques
are less effective with a utility based approach having no power to statistically differentiate between competing forecasts.

The paper proceeds as follows. Section 2 provides an overview of the existing evaluation approaches. Section 3 describes the empirical data used to motivate the simulation study, and upon which the subsequent empirical analysis is based. Section 4 outlines the methodology used to address the main focus of the paper, the efficacy of the forecast evaluation techniques. This section contains a description of the simulation study and empirical analysis. Section 5 analyses important properties of the various loss functions. Sections 6 and 7 provide simulation and empirical results, respectively. Section 8 provides concluding comments.

2 Techniques for Evaluating Volatility Forecasts

While evaluation of multivariate forecasts is the focus of this paper, we begin by briefly outlining issues relating to the evaluation of univariate forecasts. This will develop the background for the current methodology.

Direct evaluation of forecast accuracy simply requires a statistical loss function, of which many exist. In the context of evaluating volatility forecasts, matters are complicated by the fact that volatility is inherently latent. We only observe proxies for the true underlying volatility such as daily squared returns $r_t^2$, realised volatility, $RV$ or daily range. Patton (2006) proved that while many loss functions exist, only certain a family of loss functions are robust to noise in the volatility proxy, denoted as $(\hat{\sigma}_t^2)$ used for evaluating a forecast, $(h_t)$. Of the commonly used loss functions, Mean Squared Error (MSE) and the Quasi-Likelihood (QLIKE) loss functions belong to this family,

$$MSE = (\hat{\sigma}_t^2 - h_t)^2$$
$$QLIKE = \log(h_t) + \frac{\hat{\sigma}_t^2}{h_t}.$$  

(1)

In the subsequent discussion, a loss function will be denoted as $L(\cdot, \cdot)$ which can be taken to be either function shown in equation 1.

A number of methods exist permitting the direct statistical comparison of forecasts. When comparing two competing forecasts, $h_t^A$ and $h_t^B$, Diebold and Mariano (1995) and West (1996) provide such pairwise tests for equal predictive accuracy (EPA),

$$H_0 : E[L(\hat{\sigma}_t^2, h_t^A)] = E[L(\hat{\sigma}_t^2, h_t^B)]$$
$$H_A : E[L(\hat{\sigma}_t^2, h_t^A)] \neq E[L(\hat{\sigma}_t^2, h_t^B)].$$  

(2)
This test is based on the computation of

$$DMW_T = \frac{\overline{d}_T}{\sqrt{\text{avar}[\overline{d}_T]}}, \quad \overline{d}_T = \frac{1}{T} \sum_{t=1}^{T} d_t,$$

where $d_t = L(\hat{\sigma}^2_t, h^A_t) - L(\hat{\sigma}^2_t, h^B_t)$ and $\text{avar}[\overline{d}_T]$ is an estimate of the asymptotic variance of the average loss differential, $\overline{d}_T$.

Often, one wished to not simply compare two forecasts, but many forecasts of volatility. There are two main approaches for achieving this. The Reality Check of White (2000) and the test for Superior Predictive Ability (SPA) of Hansen (2005) tests whether any forecast outperforms a benchmark forecast, denoted as forecast $A$ below,

$$H_0 : E[L(\hat{\sigma}^2_t, h^A_t)] \leq \min_{i \in B,C, \ldots} E[L(\hat{\sigma}^2_t, h^i_t)]$$

$$H_A : E[L(\hat{\sigma}^2_t, h^A_t)] > \min_{i \in B,C, \ldots} E[L(\hat{\sigma}^2_t, h^i_t)].$$

The Model confidence Set approach (MCS) of Hansen, Lunde and Nason (2003a) is a modified version of the SPA tests that has greater power and does not require a benchmark forecast to be chosen. This approach is discussed in more detailed as it will be utilised in this paper in the multivariate context.

The MCS procedure starts with a full set of candidate models $M_0 = \{1,...,m_0\}$. The MCS is determined by sequentially trimming models from $M_0$ therefore reducing the number of models. Prior to starting the sequential elimination procedure, all loss differentials between models $i$ and $j$ are computed,

$$d_{ij,t} = L(\hat{\sigma}^2_t, h^j_t)) - L(\hat{\sigma}^2_t, h^i_t)), \quad i, j = 1,...,m_0, \quad t = 1,...,T.$$  

At each step, the EPA hypothesis

$$H_0 : E(d_{ij,t}) = 0, \quad \forall i > j \in M$$

is tested for a set of models $M \subset M_0$, with $M = M_0$ at the initial step. If $H_0$ is rejected at the significance level $\alpha$, the worst performing model is removed and the process continued until non-rejection occurs with the set of surviving models being the MCS, $\hat{M}_\alpha^*$. If a fixed significance level $\alpha$ is used at each step, $\hat{M}_\alpha^*$ contains the best model from $M_0$ with $(1 - \alpha)$ confidence$^1$.

At the core of the EPA statistic is the $t$-statistic, along the lines of equation 3,

$$t_{ij} = \frac{\overline{d}_{ij}}{\sqrt{\text{var}(\overline{d}_{ij})}}$$

$^1$Despite the testing procedure involving multiple hypothesis tests this interpretation is a statistically correct one. See Hansen et al. 2003b for a detailed discussion of these aspects.
where $\bar{d}_{ij} = \frac{1}{T} \sum_{t=1}^{T} d_{ij,t}$. $t_{ij}$ provides scaled information on the average difference in the forecast quality of models $i$ and $j$. $\var(d_{ij})$ is an estimate of $\var(\bar{d}_{ij})$ and is obtained from a bootstrap procedure\(^2\). In order to decide at any stage, whether the MCS must be further reduced, the null hypothesis in equation 6 is to be evaluated. The difficulty being that for each set $\mathcal{M}$ the information from $(m - 1) m/2$ unique $t$-statistics needs to be distilled into one test statistic. Hansen, et al. (2003a, 2003b) propose the following range,

$$T_R = \max_{i,j \in \mathcal{M}} |t_{ij}| = \max_{i,j \in \mathcal{M}} \frac{\bar{d}_{ij}}{\sqrt{\var(\bar{d}_{ij})}}$$

(8)

and semi-quadratic statistics,

$$T_{SQ} = \sum_{i,j \in \mathcal{M}, i < j} t_{ij}^2 = \sum_{i,j \in \mathcal{M}, i < j} \frac{\bar{d}_{ij}^2}{\var(\bar{d}_{ij})}$$

(9)

as test statistics to establish EPA. Both test statistics indicate a rejection of the EPA hypothesis for large values. The actual distribution of the test statistic is complicated and depends on the covariance structure between the forecasts included in $\mathcal{M}$. Therefore $p$-values for each of these test statistics have to be obtained from the bootstrap distribution. When the null hypothesis of EPA is rejected, the worst performing model is removed from $\mathcal{M}$. The latter is identified as $\mathcal{M}_i$ where

$$i = \arg \max_{i \in \mathcal{M}} \frac{\bar{d}_i}{\sqrt{\var(\bar{d}_i)}}$$

(10)

and $\bar{d}_i = \frac{1}{m-1} \sum_{j \in \mathcal{M}} \bar{d}_{ij}$. The tests for EPA are then conducted on the reduced set of models and one continues to iterate until the null hypothesis of EPA is not rejected. While the MCS procedure has been designed to deal with univariate forecasts, it will be utilised in the multivariate context here to differentiate between a number of competing forecasts. While the MCS has not been used in the multivariate setting, Patton and Sheppard (2006) suggest it nevertheless may be useful.

Multivariate volatility forecasts often aid in economic applications. Applications such as portfolio construction, risk management, hedging and derivative pricing, among others all require forecasts of volatility. As such, indirect evaluation techniques are popular for comparing methods for volatility forecasting.

The most general mean variance optimisation problem is,

$$\min_{w_t} w_t' H_t w_t; \ s.t. \ w_t' \mu_t = \mu_0,$$

(11)

\(^2\)For specific details on the bootstrap procedure see Becker and Clements (2008) and See Hansen et al. (2003a)
where $\mu_t$ is a vector of expected excess returns and $\mu_0$ is the target return for the portfolio. The solution to this problem is

$$ w_t = \frac{H_t^{-1}\mu}{\mu'H_t^{-1}\mu} \mu_0 $$

(12)

where $1 - \sum_i w_{i,t}$ may be invested in the risk-free asset.

To avoid estimating expected returns, volatility of the global minimum variance portfolio (GMVP) can be compared. Given a forecast of the covariance matrix of returns, $H_t$, the weights of this portfolio can be found by

$$ w_t = \frac{H_t^{-1}1}{1'H_t^{-1}1}. $$

(13)

The volatility of the portfolio returns based on any $H_t$ that is different from the true covariance matrix, $\Sigma_t$ must be greater than that given $\Sigma_t$. As such, a common evaluation method has been to compare such portfolio returns generated by competing forecasts. Engle and Colacito (2006) provide a method for evaluating the benefit of covariance forecasts that takes into account the uncertainty surrounding expected returns. This however, does not extend to larger dimensions where many variance and covariance forecasts must be considered.

This general mean variance framework in equation 11 was employed by Fleming, Kirby and Ostdiek (2001, 2003) and Engle and Colacito (2006). Based on optimal weights, portfolio returns were then determined, $R_{p,t} = w_t'r_t$, where $r_t$ is the vector of observed returns. Given portfolio returns, Fleming, Kirby and Ostdiek (2003) then propose computing the utility realised by an investor with quadratic utility,

$$ U(R_{p,t}) = W_0 \left[ (1 + R_f + R_{p,t}) - \frac{\gamma}{2(1 + \gamma)} (1 + R_f + R_{p,t})^2 \right], $$

(14)

where $R_f$ is the risk-free rate of return, $W_0$ is a fixed initial level of wealth and $\gamma$ is the investor’s level relative risk aversion. Given two competing volatility forecasts, $H_1^t$ and $H_2^t$, portfolio weights can be computed using equation 12 and hence produce portfolio returns $R_{p,t}^1$ and $R_{p,t}^2$. Fleming, Kirby and Ostdiek (2001, 2003) propose to compute the utility earned using each forecast from equation 14. The economic value of using the second forecast relative to the first is determined from

$$ \sum_{t=1}^{T} U(R_{p,t}^1) = \sum_{t=1}^{T} U(R_{p,t}^2 - \Delta), $$

(15)

where $\Delta$ represents the maximum return an investor would be willing to sacrifice to capture the benefit of the second forecast, $H_2^t$.

3 Data

Empirical data utilised in this paper are returns on five U.S. based futures contracts. The contracts are written on the S&P 500 and NASDAQ equity indices, Ten year U.S. Treasury
In all cases, a roll from the nearest, to next nearest maturity contract was set at five days prior to the maturity of nearest contract. Daily returns for the period, 4 June 1996 to 7 August 2008 have been gathered, corresponding to a sample of 3024 daily return observations.

Figure 3 shows the cumulative returns from each of the contracts. Both equity returns show the increase in equity values during the late 1990’s and subsequent falls in the early part of this decade. This feature is obviously more pronounced in the NASDAQ series. Bonds on the other hand reflect their low risk and return with a relatively small increase in value with low volatility. There have been two distinct trends in gold prices, a persistently falling market followed by the more recent large increases in price. Oil prices have experienced a great deal of volatility, punctuated by a number of rapidly rising and falling prices. These patterns are borne out in the descriptive statistics in reported in Table 1.

Gold and Oil have the highest mean returns, with Oil having the greatest volatility. Both the volatility and mean return of both equity contracts are greater than those of the Bond contract. The distribution of returns are not markedly skewed with the exception of Bonds. All series exhibit some degree of excess kurtosis with all being deemed non-normal according to the Jarque-Bera test results, a common pattern found in many financial time series.

Table 2 reports the unconditional correlations between of the futures returns. These correla-

---

3Specifically, the crude oil contract is the light/sweet crude contract traded on NYMEX.
Table 1: Descriptive statistics for futures contract returns. *p*-values for the Jarque-Bera test for normality are shown (J-B Test).

<table>
<thead>
<tr>
<th>Contract</th>
<th>S&amp;P500</th>
<th>NASDAQ</th>
<th>Bonds</th>
<th>Gold</th>
<th>Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (×10⁻⁵)</td>
<td>8.0209</td>
<td>9.5831</td>
<td>6.2563</td>
<td>18.5724</td>
<td>55.2905</td>
</tr>
<tr>
<td>Standard Dev.</td>
<td>0.0102</td>
<td>0.0174</td>
<td>0.0040</td>
<td>0.0088</td>
<td>0.0168</td>
</tr>
<tr>
<td>Skewness</td>
<td>−0.0994</td>
<td>0.0823</td>
<td>−0.4285</td>
<td>−0.2287</td>
<td>−0.1324</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>5.8086</td>
<td>8.0169</td>
<td>4.7967</td>
<td>9.1425</td>
<td>4.7649</td>
</tr>
<tr>
<td>J-B Test</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>Min.</td>
<td>−0.0616</td>
<td>−0.0969</td>
<td>−0.0228</td>
<td>−0.0705</td>
<td>−0.0986</td>
</tr>
<tr>
<td>Max.</td>
<td>0.0504</td>
<td>0.1409</td>
<td>0.0133</td>
<td>0.0683</td>
<td>0.0907</td>
</tr>
</tbody>
</table>

Table 2: Unconditional correlation matrix of futures returns.

<table>
<thead>
<tr>
<th></th>
<th>S&amp;P500</th>
<th>NASDAQ</th>
<th>Bonds</th>
<th>Gold</th>
<th>Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P500</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASDAQ</td>
<td>0.8055</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds</td>
<td>−0.1714</td>
<td>−0.1711</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gold</td>
<td>−0.0648</td>
<td>−0.0537</td>
<td>0.0621</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>Oil</td>
<td>−0.0283</td>
<td>−0.0078</td>
<td>−0.0177</td>
<td>0.2245</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

4 Methodology

This section outlines the methodology used in this paper. Section 4.1 describes the multivariate loss functions considered here. Section 4.2 discusses how the MCS methodology is applied in the multivariate context. This focuses on how the multivariate loss functions are dealt with. Section 4.3 describes the simulation study utilised to examine the performance of the various loss functions. Finally, Section 4.4 explains the nature of the empirical study presented here.

4.1 Multivariate loss functions

The multivariate loss functions examined in this paper will now be described. First, two multivariate statistical loss function will examined. These are simple extensions of the univariate
loss functions from equation 1. A multivariate form of MSE,

\[
MSE_t = \frac{1}{b^2} \sum_{i,j} (h_{ij,t} - \hat{\sigma}_{ij,t}^2)^2
\]  

(16)

where \( h_{ij,t} \) is the forecast of the covariance between series \( i \) and \( j \), \( \hat{\sigma}_{ij,t} \) is the observed proxy for this covariance and \( b^2 \) is the total number of elements in the matrix. \( h_{ij,t} \) are elements of the forecast of the covariance matrix, \( H_t \) and \( \hat{\sigma}_{ij,t} \) elements of the proxy for the covariance matrix \( \hat{\Sigma}_t \). The multivariate version of QLIKE is once again a simple extension of the univariate case of equation 1

\[
QLIKE_t = \log |H_t| + \varepsilon_t'H_t^{-1} \varepsilon_t
\]  

(17)

where \( \varepsilon_t \) is the observed vector of unexpected returns.

The performance of three versions of the economic loss functions described in Section 2 will be examined here. The simplest is the volatility of the global minimum variance portfolio (GMVP). Based on the GMVP weights described in equation 13, the volatility of the portfolio returns can be computed as

\[
r_{GMVP,t}^2 = w_t' \mu_0 w_t.
\]  

(18)

Given a number of competing volatility forecasts, \( r_{GMVP,t}^2 \) can be found using the weights generated from each forecast. The values of \( r_{GMVP,t}^2, \forall t = 1, \cdots, T \) can be compared across forecasts. While equation 18 is in itself not a loss function directly measuring accuracy, a minimum value for the volatility will be achieved when the most accurate forecast is used.

The final two economic loss functions utilise the more general portfolio optimisation problem described in equations 11 and 12. Given the weights from equation 12, the volatility of returns can be computed using equation 18 conditional on \( \mu_0 \) and will be denoted below simply as \( MVP^{\mu_0} \). The final loss function considered, following Fleming, Kirby and Ostdiek (2001, 2003) is the utility realised from the portfolio which is determined from equation 14. Following Fleming, Kirby and Ostdiek (2001, 2003), a value of \( R_f = 6\% \text{p.a.} \) is assumed. This final loss function will be denoted below as \( U(MVP^{\mu_0}) \).

In total, the performance of five loss functions will be examined here. Two statistical, MSE and QLIKE along with three economic loss functions, GMVP, \( MVP^{\mu_0} \) and \( U(MVP^{\mu_0}) \). These economic applications are themselves not loss functions measuring forecast accuracy as MSE and QLIKE do. Their values can be compared across different forecasts in the same vein as the statistical loss functions. As shown in subsequent sections, more accurate forecasts will lead to lower values for GMVP, \( MVP^{\mu_0} \) and greater \( U(MVP^{\mu_0}) \).
4.2 Use of the MCS

To statistically distinguish between forecasts of $H_t$, the MCS framework discussed in Section 2 will be used. Recall from equation 5 that the MCS is based on testing EPA in terms of the loss differentials between models $i$ and $j$. In Section 2, the MCS approach was outlined in the context of evaluating univariate forecasts using statistical loss functions. As suggested by Patton and Sheppard (2006) it may be useful in the multivariate context.

We harness the power of the MCS methodology in the multivariate setting by defining the generic univariate loss function in equation 5 to be $L(\Sigma_t, H_t)$ with the implementation of the MCS unchanged. To examine the efficacy of the various loss functions, $L(\Sigma_t, H_t)$ will take the form of MSE, QLIKE, GMVP, $MVP_{\mu_0}$ or $U(MVP_{\mu_0})^4$. Using the MCS framework will determine whether the forecast performance of competing approaches is significantly different, given the various loss functions.

4.3 Simulation Experiment

As the purpose of the simulation experiment is to identify the ability of the loss functions to identify the forecast, data generating processes (DGP) must be selected. To do so, two DGP are selected. The most general specification is that of the Asymmetric Dynamic Conditional Correlation (ADCC) model of Cappiello, Engle, and Sheppard (2006). The Dynamic Conditional Correlation (DCC) model of Engle (2002) is also utilised. For each DGP, 1000 simulations of five conditionally correlated heteroskedastic return series, each of 2000 observations are generated.

The return vector is specified as

$$ r_t \sim \Phi(\mu_t, H_t) $$

(19)

where $\Phi()$ is the multivariate normal distribution and $H_t$ takes the form

$$ H_t = D_t R_t D_t, $$

(20)

where $D_t$ is a diagonal matrix of conditional standard deviations and $R_t$ is the conditional correlation matrix. Under the ADCC process, the elements of $D_t$, $\sqrt{h_{i,t}}$, have a threshold specification,

$$ h_{i,t} = \varpi_i + (\alpha_i + \theta_i S_{i,t-1}) r_{i,t-1}^2 + \beta_i h_{i,t-1}, $$

(21)

where $\varpi_i, \alpha_i, \theta_i$ and $\beta_i$ are parameters for the series $i$, while $S_{i,t-1}$ is an indicator variable that equals one if $\varepsilon_{i,t-1} < 0$. While the processes for $R_t$ and $Q_t$ is specified as

$$ Q_t = \bar{Q} (1 - \alpha - \beta) - \phi m + \alpha (z_{t-2} + m_{t-2} + z_{t-1} + m_{t-1} + \beta Q_{t-1} $$

(22)

$^4$Negative values of the utility based loss function, $U(MVP_{\mu_0})$ will be used. To be consistent with the application of the MCS, superior performance of a forecast should be reflected in lower values for any given loss function.
Table 3: Parameter values used to generate the conditionally correlated heteroskedastic returns. Panel A reports the parameter values when DCC is used as the DGP. Panel B reports the parameter values when ADCC is the DGP. All parameter values are estimated from the entire empirical data set by maximum likelihood. Constraints on the asymmetric parameters in the DCC model are noted by cells containing dashes.

\[
R_t = \text{diag} \left[ Q_t \right]^{-1/2} Q_t \text{diag} \left[ Q_t \right]^{-1/2}.
\]

where \( \alpha, \beta \) and \( \phi \) are parameters, \( \bar{Q} \) is the unconditional correlation matrix, \( z_t \) is a vector of standardised returns from \( \frac{r_{i,t}}{\sqrt{h_{i,t}}} \), and \( m_{t-1} \) and \( \bar{m} \) are leverage effect measures. Specifically, the leverage effect measures are \( m_{t-1} = \delta \odot z_t \), where \( \delta \) is a dummy variable vector with elements \( \delta_{i1} = 1 \) if \( z_{i,t-1} < 0 \), and \( \bar{m} = \frac{1}{J} \sum m_{t-j} m'_{t-j} \). To recover the DCC specification the \( \theta \) parameter in equation 21 and \( \phi \) parameter in equation 22 are constrained to zero.

Prior to the simulations, values for the parameters (\( \varpi, \alpha_i, \theta_i, \beta_i, \alpha, \beta \) and \( \phi \)) and the unconditional measures (\( \bar{Q} \) and \( \bar{m} \)) from equations 21 and 22 must be specified. In addition, an unconditional covariance matrix \( H \), \( \mu \) and \( R \) are also required. To provide realistic values, these values are estimated from the full empirical data set outlined in Section 3. The estimated parameters are presented in Table 3 while \( R \) and \( \mu \) are reported in Section 3.

The simulation process can be described as follows. Given \( R_t \) (set to the unconditional value, \( R \) at \( t = 1 \)) a vector of correlated standardised returns is generated as \( z_t = \sqrt{\bar{R}_t} \) where the elements of \( \bar{u}_t \sim N(0, 1) \). Using equations 22 and 23, a value for \( R_{t+1} \) is generated which is turn used to obtain \( z_{t+1} \). Given a value for \( z_t \), simulated returns are determined by \( r_{i,t} = z_{i,1} \sqrt{h_{i,t}} \) (with \( h_{i,t} \) set to the \( ith \) element of \( H \) for \( t = 1 \)). Returns are then simulated as the series
of conditional variances for each asset are constructed iteratively from equation 21. For each simulation, this process is repeated for 2000 time steps and 1000 simulation for each DGP.

One-step ahead multivariate volatility forecasts will be utilised to evaluate the loss functions. While a number of models have been chosen for this purpose they are not an exhaustive list. However, each model can generate volatility forecasts for moderately sized covariance matrices with the quality of their forecasts expected to vary.

The simplest model chosen is the static covariance model where the forecast is simply the unconditional covariance matrix,

\[ H_t = \frac{1}{J} \sum_{j=1}^{J} \varepsilon_{t-j} \varepsilon'_t, \tag{24} \]

where \( J \) represents the number of observations in the in-sample estimation period.

Another simplistic model is the multivariate moving average (MA) model, with forecasts based on sampling the \( M \) most recent observations,

\[ H_t = \frac{1}{M} \sum_{m=1}^{M} \varepsilon_{t-m} \varepsilon'_t, \tag{25} \]

with \( M = 100 \) used for this study.

The next model considered is the exponentially weighted moving average model (EWMA) introduced by Riskmetrics (1996). Unlike the previous models that applied an equal weight to observations within the sample period, the EWMA model applies a declining weighting scheme that places greater weight on the most recent observation. This model takes the form,

\[ H_t = (1 - \lambda) \varepsilon_{t-1} \varepsilon'_t + \lambda H_{t-1}, \tag{26} \]

where \( \lambda \) is the parameter that controls the weighting scheme. Riskmetrics (1996) specify a \( \lambda = 0.94 \) for data sampled at a daily frequency, the value used in this study.

The next model utilised is the exponentially weighted model of Fleming, Kirby and Ostdiek (2001, 2003), denoted below as FKO,

\[ H_t = \alpha \exp (-\alpha) \varepsilon_{t-1} \varepsilon'_t + \exp (-\alpha) H_{t-1}, \tag{27} \]

where \( \alpha \) is the parameter that governs the weights on lagged observations. Similar to the EWMA, a declining weighting scheme is applied to lagged observations, however this weighting parameter is estimated by maximum likelihood.

The final three models are drawn from the conditional correlation multivariate GARCH class of models. Along with the ADCC and DCC models used as DGP, the Constant Conditional Correlation (CCC) of Bollerslev (1990). The CCC model is recovered by constraining the \( \theta \) in
equation 21 and the \(\alpha, \beta\) and \(\phi\) in equation 22 to zero, while DCC constrains \(\theta\) and \(\phi\) to zero. Estimation of the conditional correlation models rely on the two stage maximum likelihood procedure detailed by Engle and Sheppard (2001).

Each model is used to make one-step-ahead volatility forecasts. The forecast for the initial time step is set to be the unconditional value, \(H\) from the empirical data. All subsequent forecasts for time \(t\) are then formed given the specification of each model and \(H_{t-1}\) and \(r_{t-1}\). Parameters and unconditional measures are not re-estimated. The performance of the MSE, QLIKE, GMVP, \(MVP_{\mu_0}\) or \(U(MVP_{\mu_0})\) loss functions will be examined, with \(\mu_0 = 8\%\).

### 4.4 An Empirical Analysis

The empirical study utilises the futures returns detailed in Section 3 and the forecasting models described in Section 4.3. For the purposes of this study returns are assumed to follow,

\[ r_t = \mu + \varepsilon_t. \]  

(28)

To begin, the period \((t = 1, 2, ..., 1000)\) is used as the initial in-sample period. From this data, a vector of means \(\hat{\mu}\), the unconditional covariance matrix \(H\) and the required values for forecasting models are estimated. A forecast of \(H_{1001}\) is then generated using each of the models, given the estimated value of \(H_{1000}\) where necessary. The in-sample period is extended to \((t = 1, 2, ..., 1001)\) and the process repeated giving a total of 2024 one step ahead forecasts. Parameter estimates for the FKO, CCC, DCC and ADCC models are obtained recursively. The empirical analysis will rely on the MCS framework to distinguish between the empirical performance of the seven competing models.

### 5 Some Properties of the Loss Functions

Patton and Sheppard (2006) show that MSE and QLIKE belong to a wider class of statistical loss functions that are robust, in this sense they reach an optimum when \(H_t = \hat{\Sigma}_t\). Patton and Sheppard (2006) also show similar arguments for the GMVP loss function in that when \(H_t = \Sigma_t\) the variance of the GMVP will reach a minimum. For any case where \(H_t \neq \Sigma_t\) one would expect that the GMVP formed from \(H_t\) will lead to a larger variance.

Following from Patton and Sheppard (2006), we can show that \(MVP_{\mu_0}\) loss function exhibits the same properties. Begin by defining \(w_t\) as the vector of weights generated from \(\Sigma_t\), \(\hat{w}_t\) as a vector of incorrect weights generated from \(H_t\), and \(c_t\) as a vector of weighting errors due to

\(^5\text{Results for 4\% and 12\% are available upon request but are omitted for the sake of brevity. Results reflect the same patterns discussed in subsequent sections.}\)
\( H_t \neq \Sigma_t \). The impact on portfolio variance due to \( H_t \neq \Sigma_t \) can be expressed as

\[
\begin{aligned}
\bar{w}_t' \Sigma_t w_t - w_t' \Sigma_t w_t &= (w_t + c_t)' \Sigma_t (w_t + c_t) - w_t' \Sigma_t w_t \\
&= w_t' \Sigma_t w_t + 2c_t' \Sigma_t w_t + c_t' \Sigma_t c_t - w_t' \Sigma_t w_t \\
&= 2c_t' \Sigma_t - \frac{\Sigma_t^{-1} \mu_t}{\mu_t' \Sigma_t^{-1} \mu_t} \mu_0 + c_t' \Sigma_t c_t \\
&= 2 \frac{c_t' \mu_t}{\mu_t' \Sigma_t^{-1} \mu_t} \mu_0 + c_t' \Sigma_t c_t \\
&= c_t' \Sigma_t c_t,
\end{aligned}
\]

as \( w_t' \mu_t = \mu_0 \) and \( w_t' \mu_t + c_t' \mu_t = \mu_0 \) hence \( c_t' \mu_t = 0 \). Given that \( c_t' \Sigma_t c_t \geq 0 \), an incorrect forecast cannot produce a smaller variance than the \( H_t = \Sigma_t \) case. This is consistent with the result of Engle and Colacito (2006) without the use of expectations.

Extending the analysis of Patton and Sheppard (2006) to the \( U(MVP^{\mu_0}) \) case leads to a vastly different conclusion. Once again by defining \( c_t \) as a vector of weighting errors due to \( H_t \neq \Sigma_t \), we can highlight the impact on the \( U(MVP^{\mu_0}) \) loss function. Using the utility function specified in equation 14 the difference in utility due to \( c_t \) can be expressed as

\[
\Delta U_t = \left[ (1 + R_f + w_t' r_t + c_t' r_t) - \frac{\gamma}{2(1 + \gamma)} (1 + R_f + w_t' r_t + c_t' r_t)^2 \right] - \left[ (1 + R_f + w_t' r_t) - \frac{\gamma}{2(1 + \gamma)} (1 + R_f + w_t' r_t)^2 \right] = c_t' r_t - \frac{\gamma}{2(1 + \gamma)} (2R_f c_t' r_t + 2c_t' r_t + 2c_t' r_t' w_t + c_t' r_t' c_t) = \left[ 1 - \frac{\gamma}{2(1 + \gamma)} (1 + R_f) \right] c_t' r_t - \frac{\gamma}{2(1 + \gamma)} c_t' r_t' w_t - \frac{\gamma}{2(1 + \gamma)} c_t' r_t' c_t.
\]

It is obvious that \( \Delta U_t \leq 0 \), as \( \left[ 1 - \frac{\gamma}{2(1 + \gamma)} (1 + R_f) \right] c_t' r_t \leq 0 \) and \( \frac{\gamma}{2(1 + \gamma)} c_t' r_t' w_t \leq 0 \). However, \( \frac{\gamma}{2(1 + \gamma)} c_t' r_t' c_t > 0 \). Clearly, this result demonstrates that \( U_t \) does not consistently identify the true volatility matrix, as instances of \( c_t' r_t > 0 \) may result in \( \Delta U_t > 0 \), which means that the \( H_t \neq \Sigma_t \) is identified as a superior forecast relative to \( \Sigma_t \). Taking expectations in this case yields

\[
E(\Delta U_t) = \left[ 1 - \frac{\gamma}{2(1 + \gamma)} (1 + R_f) \right] c_t' E(r_t) - \frac{\gamma}{2(1 + \gamma)} c_t' E(r_t' r_t') w_t
\]

Thus on average, \( U(MVP^{\mu_0}) \) will identify \( \Sigma_t \) as the best forecast when the target return constraint is used in the portfolio optimisation problem. However, the result from equation 30 shows that observed returns will directly influence the rankings of forecasts at each point in time, therefore we expect the differences in utility to be highly variable.

14
6 Simulation Results

This section presents the results of the simulation experiment described in Section 4.3 designed to examine the efficacy of the various loss functions. Results from the application of the MCS are presented in the following form in all cases. Panel A of each table provides a summary of the MCS by reporting average size of the MCS itself across all simulations. Panel B reports how effective the loss functions are at identifying the DGP model by reporting the occurrence of DGP in the MCS. Finally, Panel C reports the occurrences of the other forecasting models in the MCS. In all cases, results are presented at four levels of significance ($\alpha = 0.01, 0.05, 0.1, 0.2$).

We begin by examining the performance of MSE and QLIKE in the case where DCC is the DGP, with the results presented in Table 4. The results in Panel A clearly show that both loss functions do reject EPA across the models in that the average size of the MCS (at all $\alpha$) is always less than the maximum MCS size of seven. However, when the loss functions are compared, it is obvious that QLIKE produces more rejections of EPA. The MCS under QLIKE consists of approximately 1 model on average, whereas the MSE leads to between 2 and 4 models in the MCS. While these results appear to favour QLIKE, it is now necessary to examine which models are contained in the MCS.

Results in Panel B lend further support to QLIKE. Neither loss function virtually ever excludes the DGP, as indicated by the results of the first row. However, when combined with the result in the second row, showing that in no less than 99.5% of cases QLIKE excludes all other models where as MSE achieves this in 36.3% at most, QLIKE appears to be more effective at identifying the DGP. The last panel of Table 4 shows that MSE exhibits difficulty in distinguishing between the DGP and relatively similar models, FKO, CCC and ADCC. Clearly the performance of QLIKE does not suffer in this way.

Results in Table 5 where ADCC is the DGP present a similar picture. Once again, Panel A shows that while both loss functions are clearly excluding inferior models, QLIKE leads to smaller MCS sizes on average. Panel B again confirms the superiority of QLIKE. While both loss function rarely exclude the DGP, the instances where the MCS only contains the DGP are much more frequent under QLIKE. Though the frequency is lower than observed in Table 4 when DCC was the DGP. Panel C shows that under QLIKE, the MCS only ever contains the DCC model, whereas once again MSE often cannot exclude a wider range of models from the MCS. Overall these results appear to show that QLIKE is the superior loss function of the two statistical loss functions. These results also lend support to the use of the MCS in the multivariate setting, as suggested by Patton and Sheppard (2006). It now remains to compare the performance of the economic loss functions to these statistical loss functions.
Table 4: Summary results for MCS when MSE and QLIKE and DCC is the DGP. Panel A details the average size of the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS and is the only model in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>QLIKE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01 0.05 0.10 0.20</td>
<td>0.01 0.05 0.10 0.20</td>
</tr>
<tr>
<td><strong>Panel A: Average MCS size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average MCS size</td>
<td>4.0  3.1  2.6  2.1</td>
<td>≈ 1.0  ≈ 1.0  ≈ 1.0  1.0</td>
</tr>
<tr>
<td><strong>Panel B: % where MCS contains DGP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCS includes DCC</td>
<td>100.0 100.0 100.0 99.9</td>
<td>100.0 100.0 100.0 100.0</td>
</tr>
<tr>
<td>MCS includes only DCC</td>
<td>0.7  7.5  18.4 36.3</td>
<td>99.5 99.9 99.9 100.0</td>
</tr>
<tr>
<td><strong>Panel C: % where MCS contains non-DGP models</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>44.6  26.5 18.6 11.8</td>
<td>0.1  0.0  0.0  0.0</td>
</tr>
<tr>
<td>MA</td>
<td>11.8  2.3  0.5  0.2</td>
<td>0.0  0.0  0.0  0.0</td>
</tr>
<tr>
<td>EWMA</td>
<td>2.2  0.2  0.1  0.0</td>
<td>0.0  0.0  0.0  0.0</td>
</tr>
<tr>
<td>FKO</td>
<td>98.3  88.2 75.4 56.4</td>
<td>0.3  0.1  0.1  0.0</td>
</tr>
<tr>
<td>CCC</td>
<td>63.1  31.6 18.2 9.0</td>
<td>0.0  0.0  0.0  0.0</td>
</tr>
<tr>
<td>ADCC</td>
<td>83.8  57.3 42.7 36.2</td>
<td>0.1  0.0  0.0  0.0</td>
</tr>
</tbody>
</table>

Table 5: Summary results for MCS when MSE and QLIKE and ADCC is the DGP. Panel A details the average size of the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS and is the only model in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>QLIKE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01 0.05 0.10 0.20</td>
<td>0.01 0.05 0.10 0.20</td>
</tr>
<tr>
<td><strong>Panel A: Average MCS size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average MCS size</td>
<td>3.9  3.1  2.7  2.3</td>
<td>1.8  1.7  1.5  1.4</td>
</tr>
<tr>
<td><strong>Panel B: % where MCS contains DGP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCS includes ADCC</td>
<td>99.9 99.4 98.5 96.9</td>
<td>100.0 100.0 99.9 99.7</td>
</tr>
<tr>
<td>MCS includes only ADCC</td>
<td>1.7  6.8 11.1 20.3</td>
<td>17.3 32.9 45.6 59.9</td>
</tr>
<tr>
<td><strong>Panel C: % where MCS contains non-DGP</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>33.8  18.5 11.6 5.3</td>
<td>0.0  0.0  0.0  0.0</td>
</tr>
<tr>
<td>MA</td>
<td>6.6  0.8  0.1  0.0</td>
<td>0.0  0.0  0.0  0.0</td>
</tr>
<tr>
<td>EWMA</td>
<td>1.8  0.3  0.1  0.0</td>
<td>0.0  0.0  0.0  0.0</td>
</tr>
<tr>
<td>FKO</td>
<td>95.2  78.1 61.5 40.3</td>
<td>0.1  0.0  0.0  0.0</td>
</tr>
<tr>
<td>CCC</td>
<td>55.5  24.1 13.1 5.2</td>
<td>0.0  0.0  0.0  0.0</td>
</tr>
<tr>
<td>DCC</td>
<td>98.3  93.2 88.9 79.7</td>
<td>82.7 67.1 54.4 40.1</td>
</tr>
</tbody>
</table>
Table 6: Summary results for MCS when GMVP is loss function and DCC is the DGP. Panel A details the average size of the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS and is the only model in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.

Tables 6 and 7 contain the simulation results for the GMVP loss function given DCC and ADCC DGP respectively. Overall, it seems as though its performance as a loss function is somewhat better than that of MSE. For both DGP, Panel A of Tables 6 and 7 show that the average MCS size ranges between 1.2 and 2.8, these values are between those produced by the statistical loss functions. As with the statistical loss functions, Panel B of each tables show that the DGP is hardly excluded given either DGP. Panel B in each case, also indicates that the frequency with which MCS consists of only the DGP is similar to MSE. That pattern extends to the results in Panel C of each table. The incorrect models included in the MCS most frequently are the ADCC or DCC and FKO. Given these results, it appears as though the performance of GMVP in distinguishing between forecasts is similar in nature to MSE.

Tables 8 and 9 report the MCS simulation results for the $MVP^{\mu_0}$ loss function with $\mu_0 = 8\%$. Panel A of each table indicates that with an average size of the MCS ranging between 1.2 and 2.9, the $MVP^{\mu_0}$ excludes inferior at a similar rate to the GMVP. Panel B results show that, as with all loss function considered thus far, the DGP is rarely excluded from the MCS. However, the rate at which the MCS is restricted to the DGP is noticeably lower than the GMVP and is similar to MSE. Panel C results indicate that the MCS under this loss function often contain the much simpler models, Static and MA along with DCC or ADCC and FKO. It would appear
Table 7: Summary results for MCS when GMVP is loss function and ADCC the is DGP. Panel A details the average size of the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS and is the only model in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.

that the $MVP^{\mu_0}$ loss function has somewhat lower power than the preceding loss functions in discriminating between the competing forecasts.

Tables 10 and 11 report the MCS simulation results for the utility based, $U(MVP^{\mu_0})$ loss function. Results are reported for $\mu_0 = 8\%$ and risk aversion, $\gamma = 1, 10$. These results paint a vastly different picture from those discussed earlier. Given either DGP, Panel A results show that the average MCS size is often close to the original seven models under consideration. It is clear there are few rejections of EPA occurring. Panel B results show that while the DGP is excluded from the MCS relatively infrequently, it is never the sole model. This pattern is obviously consistent with results in Panel A in that few rejections of EPA occur. These results are reflected in Panel C where the frequency with which non-DGP models are contained in the MCS is extremely high. It is clear these results stand irrespective of the level of risk aversion. The simulation results show that in comparison to a number of loss functions, mainly QLIKE, MSE and GMVP, $U(MVP^{\mu_0})$ has virtually no power to distinguish between the forecasts. It is conjectured that this lack of power is due to the impact of the variability associated with realised returns as discussed in Section 5. The differential between the performance of the $MVP^{\mu_0}$ and $U(MVP^{\mu_0})$ would be due to the fact that the realised portfolio returns directly enter the utility function.
Table 8: Summary results for MCS when $MV^P \mu_0$, $\mu_0 = 8\%$ is the loss function and DCC is the DGP. Panel A details the average size of the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS and is the only model in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.

<table>
<thead>
<tr>
<th>Volatility</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Average MCS size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average MCS size</td>
<td>2.9</td>
<td>1.9</td>
<td>1.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Panel B: % where MCS contains DGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCS includes DCC</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>MCS includes only DCC</td>
<td>16.7</td>
<td>47.8</td>
<td>67.0</td>
<td>82.7</td>
</tr>
<tr>
<td>Panel C: % where MCS contains non-DGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>17.8</td>
<td>7.5</td>
<td>3.4</td>
<td>1.1</td>
</tr>
<tr>
<td>MA</td>
<td>18.8</td>
<td>6.1</td>
<td>3.0</td>
<td>1.1</td>
</tr>
<tr>
<td>EWMA</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>FKO</td>
<td>63.1</td>
<td>34.0</td>
<td>20.2</td>
<td>9.4</td>
</tr>
<tr>
<td>CCC</td>
<td>16.1</td>
<td>5.1</td>
<td>1.4</td>
<td>0.5</td>
</tr>
<tr>
<td>ADCC</td>
<td>71.0</td>
<td>37.2</td>
<td>22.1</td>
<td>10.8</td>
</tr>
</tbody>
</table>

Table 9: Summary results for MCS when $MV^P \mu_0$, $\mu_0 = 8\%$ is the loss function and ADCC is the DGP. Panel A details the average size of the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS and is the only model in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.

<table>
<thead>
<tr>
<th>Volatility</th>
<th>0.01</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel A: Average MCS size</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average MCS size</td>
<td>2.9</td>
<td>2.3</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Panel B: % where MCS contains DGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCS includes ADCC</td>
<td>99.9</td>
<td>99.4</td>
<td>98.9</td>
<td>96.8</td>
</tr>
<tr>
<td>MCS includes only ADCC</td>
<td>1.5</td>
<td>6.8</td>
<td>12.8</td>
<td>22.8</td>
</tr>
<tr>
<td>Panel C: % where MCS contains non-DGP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static</td>
<td>12.3</td>
<td>4.2</td>
<td>2.6</td>
<td>1.3</td>
</tr>
<tr>
<td>MA</td>
<td>11.5</td>
<td>3.5</td>
<td>1.6</td>
<td>0.4</td>
</tr>
<tr>
<td>EWMA</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>FKO</td>
<td>53.9</td>
<td>27.4</td>
<td>16.9</td>
<td>8.3</td>
</tr>
<tr>
<td>CCC</td>
<td>12.0</td>
<td>3.3</td>
<td>1.6</td>
<td>0.5</td>
</tr>
<tr>
<td>DCC</td>
<td>98.5</td>
<td>93.2</td>
<td>87.2</td>
<td>77.2</td>
</tr>
</tbody>
</table>
\[
\gamma = 1 \quad \text{and} \quad \gamma = 10
\]

Panel A: Average MCS size
Average MCS size \( \approx 7.0 \quad 6.9 \quad 6.8 \quad 6.5 \quad \approx 7.0 \quad 6.9 \quad 6.8 \quad 6.5 \)

Panel B: % where MCS contains DGP
MCS Includes DCC \( 99.9 \quad 98.8 \quad 96.0 \quad 91.2 \quad 99.9 \quad 99.4 \quad 98.2 \quad 94.3 \)
MCS Includes only DCC \( 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \quad 0.0 \)

Panel C: % where MCS contains non-DGP
Static \( 100.0 \quad 99.2 \quad 98.0 \quad 95.8 \quad 99.9 \quad 99.2 \quad 97.8 \quad 93.8 \)
MA \( 99.8 \quad 98.2 \quad 95.7 \quad 91.3 \quad 99.6 \quad 97.7 \quad 94.6 \quad 90.3 \)
EWMA \( 99.8 \quad 98.5 \quad 96.8 \quad 93.7 \quad 99.6 \quad 97.5 \quad 95.8 \quad 91.7 \)
FKO \( 99.8 \quad 98.2 \quad 95.6 \quad 90.3 \quad 99.8 \quad 98.6 \quad 96.7 \quad 92.4 \)
CCC \( 100.0 \quad 98.6 \quad 96.9 \quad 93.1 \quad 99.9 \quad 98.4 \quad 96.6 \quad 93.6 \)
ADCC \( 100.0 \quad 98.0 \quad 96.1 \quad 90.7 \quad 99.9 \quad 98.6 \quad 96.2 \quad 92.6 \)

Table 10: Summary results for MCS when \( U(MVP^{\mu_0}) \), \( \mu_0 = 8\% \) is the loss function and DCC is the DGP. Results are reported for risk aversion, \( \gamma = 1, 10 \). Panel A details the average size of the MCS and is the only model in the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.

In summary, the simulation results indicate that statistical loss functions, overall are superior to the economic loss functions. While the performance of MSE, GMVP and \( MVP^{\mu_0} \) being similar in nature, QLIKE exhibits the most power with and \( U(MVP^{\mu_0}) \) the least. The results would appear consistent with the properties of the loss functions outlined in Section 5. It was shown that observed returns have a major impact on the performance of the \( U(MVP^{\mu_0}) \) loss function. These simulation results indicate that the variability in returns leads to such a high degree of variability in utility differentials, that it is almost impossible to statistically distinguish between competing forecasts.

7 Empirical Results

This section presents the results of the empirical analysis described in Section 4.4 and will be interpreted in light of the simulation results of the previous section. Table 12 contains the MCS results for the statistical loss functions. Given either loss function, ADCC is the best performing forecast. Under QLIKE it is significantly superior to all competing models. MSE indicates that at \( \alpha = 0.2 \) one would draw the same conclusion, however for relatively small reductions in \( \alpha \)
Table 11: Summary results for MCS when $U(MVP^\mu_0)$, $\mu_0 = 8\%$ is the loss function and ADCC is the DGP. Results are reported for risk aversion, $\gamma = 1, 10$. Panel A details the average size of the MCS. Panel B reports the percentage of simulations where the DGP is included in the MCS and is the only model in the MCS. Panel C reports the percentage of simulations where the MCS contains a non-DGP model.
Table 12: Empirical MCS results. Reading left to right, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported. Panel A presents the MSE results. Panel B presents the QLIKE results.

<table>
<thead>
<tr>
<th>MCS Size</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: Empirical MCS results with MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>Static</td>
<td>MA</td>
<td>CCC</td>
<td>EWMA</td>
<td>FKO</td>
<td>DCC</td>
<td>ADCC</td>
</tr>
<tr>
<td>P-value</td>
<td>0.06</td>
<td>0.13</td>
<td>0.13</td>
<td>0.15</td>
<td>0.17</td>
<td>0.17</td>
<td>1.00</td>
</tr>
</tbody>
</table>

| Panel B: Empirical MCS results with QLIKE |
| QLIKE | Static | CCC | EWMA | MA | FKO | DCC | ADCC |
| P-value | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.00 |

Table 13: Empirical MCS results when GMVP is the loss function. Reading left to right, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported.

<table>
<thead>
<tr>
<th>MCS Size</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: MCS with Volatility of GMVP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>Static</td>
<td>CCC</td>
<td>EWMA</td>
<td>DCC</td>
<td>MA</td>
<td>ADCC</td>
<td>FKO</td>
</tr>
<tr>
<td>P-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.16</td>
<td>0.16</td>
<td>0.31</td>
<td>1.00</td>
</tr>
</tbody>
</table>

other models will remain in the MCS. These results are consistent with the simulation results in that QLIKE exhibits more power in distinguishing between forecasts and leads to smaller MCS.

Table 13 reports the MCS results for the GMVP loss function. At any reasonable level of significance, ADCC and FKO are contained in the MCS. If one were to reduce $\alpha = 0.31$ to $\alpha = 0.16$, the MCS would also contain DCC and MA. Given this drop in $\alpha$, one would likely conclude that the ADCC and FKO models were superior to the others. Relatively similar results are found when using the $MVP^{\mu_0}$ loss function, the results of which are reported in Table 14. At $\alpha = 0.43$ the MCS contains FKO, DCC and ADCC.

The MCS results given the $U(MVP^{\mu_0})$ loss function in Table 15 once again show a different pattern to the preceding results. Given either $\gamma$, $U(MVP^{\mu_0})$ cannot distinguish between any of the competing forecasts with the MCS in both cases containing all forecasts. This results is once again a reflection of the significant impact that observed returns have upon this loss function.

Given these results a number of interesting conclusions arise. First, in terms of direct forecast accuracy as measured by QLIKE, ADCC is the superior forecast. While the MCS based on
Table 14: Empirical MCS results when $MV^\mu_0$, $\mu_0 = 8\%$ is the loss function. Reading left to right, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported.

<table>
<thead>
<tr>
<th>MCS Size</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A:</strong> MCS with Volatility, $\mu_0 = 8%$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>Static</td>
<td>CCC</td>
<td>EWMA</td>
<td>MA</td>
<td>DCC</td>
<td>ADCC</td>
<td>FKO</td>
</tr>
<tr>
<td>P-value</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.43</td>
<td>0.60</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 15: Empirical MCS results when $U(MVP^\mu_0)$, $\mu_0 = 8\%$ is the loss function. Reading left to right, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported. Results are presented for risk aversion, $\gamma = 1, 10$.

<table>
<thead>
<tr>
<th>MCS Size</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A:</strong> MCS with Utility, $\mu_0 = 8%, \gamma = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>EWMA</td>
<td>ADCC</td>
<td>DCC</td>
<td>CCC</td>
<td>FKO</td>
<td>Static</td>
<td>MA</td>
</tr>
<tr>
<td>P-value</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.59</td>
<td>1.00</td>
</tr>
</tbody>
</table>

| **Panel B:** MCS with Utility, $\mu_0 = 8\%, \gamma = 10$ | | | | | | | |
| Model | Static | ADCC | DCC | CCC | EWMA | FKO | MA |
| P-value | 0.58 | 0.58 | 0.70 | 0.73 | 0.99 | 0.99 | 1.00 |

QLIKE only contains ADCC, it appears in the MCS in all other cases at any reasonable level of significance. Following from the simulation results, the utility based loss function is not able to distinguish between any of the forecasts. These results would indicate that if an investor was a quadratic utility maximiser, the choice of volatility model has no significant impact.

### 8 Conclusion

Techniques for evaluating univariate volatility forecasts are well understood and often rely on traditional statistical measure of accuracy. Literature regarding the evaluation of multivariate volatility forecasts is less developed in terms of our understanding of the performance of the techniques used.

In contrast to the univariate setting, multivariate forecasts are often compared in terms of their performance in the context of an economic application. The most common application being portfolio allocation. This paper has attempted to address the efficacy of these techniques, along with statistical measures, for evaluating multivariate forecasts.
Simulation results presented here indicate that overall, statistical loss functions are the dominant approaches for evaluating forecasts. Economic applications that rely on expected asset, and realised portfolio returns, specifically utility based measures have the weakest power to distinguish between competing forecasts. Thus if the goal is to determine a superior forecasts from a set of competing forecasts, these results would indicate to rely on statistical loss functions and not a utility based measure. If one were truly utility maximising portfolio allocator then the volatility forecasting model matters little. A range of models are statistically indistinguishable in terms of their performance within this framework. As a corollary, the MCS has been found to be quite useful in the multivariate context as has been suggested.
References


List of NCER Working Papers

No. 40  (Download full text)  
Lawrence M. Kahn  
The Economics of Discrimination: Evidence from Basketball

No. 39  (Download full text)  
Don Harding and Adrian Pagan  
An Econometric Analysis of Some Models for Constructed Binary Time Series

No. 38  (Download full text)  
Richard Dennis  
Timeless Perspective Policymaking: When is Discretion Superior?

No. 37  (Download full text)  
Paul Frijters, Amy Y.C. Liu and Xin Meng  
Are optimistic expectations keeping the Chinese happy?

No. 36  (Download full text)  
Benno Torgler, Markus Schaffner, Bruno S. Frey, Sascha L. Schmidt and Uwe Dulleck  
Inequality Aversion and Performance in and on the Field

No. 35  (Download full text)  
T M Christensen, A. S. Hurn and K A Lindsay  
Discrete time-series models when counts are unobservable

No. 34  (Download full text)  
Adam Clements, A S Hurn and K A Lindsay  
Developing analytical distributions for temperature indices for the purposes of pricing temperature-based weather derivatives

No. 33  (Download full text)  
Adam Clements, A S Hurn and K A Lindsay  
Estimating the Payoffs of Temperature-based Weather Derivatives

No. 32  (Download full text)  
T M Christensen, A S Hurn and K A Lindsay  
The Devil is in the Detail: Hints for Practical Optimisation

No. 31  (Download full text)  
Uwe Dulleck, Franz Hackl, Bernhard Weiss and Rudolf Winter-Ebmer  
Buying Online: Sequential Decision Making by Shopbot Visitors

No. 30  (Download full text)  
Richard Dennis  
Model Uncertainty and Monetary Policy

No. 29  (Download full text)  
Richard Dennis  
The Frequency of Price Adjustment and New Keynesian Business Cycle Dynamics
No. 28  (Download full text)
Paul Frijters and Aydogan Ulker
Robustness in Health Research: Do differences in health measures, techniques, and time frame matter?

No. 27  (Download full text)
Paul Frijters, David W. Johnston, Manisha Shah and Michael A. Shields

No. 26  (Download full text)
Paul Frijters and Tony Beatton
The mystery of the U-shaped relationship between happiness and age.

No. 25  (Download full text)
T M Christensen, A S Hurn and K A Lindsay
It never rains but it pours: Modelling the persistence of spikes in electricity prices

No. 24  (Download full text)
Ralf Becker, Adam Clements and Andrew McClelland
The Jump component of S&P 500 volatility and the VIX index

No. 23  (Download full text)
A. S. Hurn and V. Pavlov
Momentum in Australian Stock Returns: An Update

No. 22  (Download full text)
Mardi Dungey, George Milunovich and Susan Thorp
Unobservable Shocks as Carriers of Contagion: A Dynamic Analysis Using Identified Structural GARCH

No. 21  (Download full text) (forthcoming)
Mardi Dungey and Adrian Pagan
Extending an SVAR Model of the Australian Economy

No. 20  (Download full text)
Benno Torgler, Nemanja Antic and Uwe Dulleck
Mirror, Mirror on the Wall, who is the Happiest of Them All?

No. 19  (Download full text)
Justina AV Fischer and Benno Torgler
Social Capital And Relative Income Concerns: Evidence From 26 Countries

No. 18  (Download full text)
Ralf Becker and Adam Clements
Forecasting stock market volatility conditional on macroeconomic conditions.

No. 17  (Download full text)
Ralf Becker and Adam Clements
Are combination forecasts of S&P 500 volatility statistically superior?
No. 16  (Download full text)
Uwe Dulleck and Neil Foster
Imported Equipment, Human Capital and Economic Growth in Developing Countries

No. 15  (Download full text)
Ralf Becker, Adam Clements and James Curchin
Does implied volatility reflect a wider information set than econometric forecasts?

No. 14  (Download full text)
Renee Fry and Adrian Pagan
Some Issues in Using Sign Restrictions for Identifying Structural VARs

No. 13  (Download full text)
Adrian Pagan
Weak Instruments: A Guide to the Literature

No. 12  (Download full text)
Ronald G. Cummings, Jorge Martinez-Vazquez, Michael McKee and Benno Torgler
Effects of Tax Morale on Tax Compliance: Experimental and Survey Evidence

No. 11  (Download full text)
Benno Torgler, Sascha L. Schmidt and Bruno S. Frey
The Power of Positional Concerns: A Panel Analysis

No. 10  (Download full text)
Ralf Becker, Stan Hurn and Vlad Pavlov
Modelling Spikes in Electricity Prices

No. 9  (Download full text)
A. Hurn, J. Jeisman and K. Lindsay
Teaching an Old Dog New Tricks: Improved Estimation of the Parameters of Stochastic Differential Equations by Numerical Solution of the Fokker-Planck Equation

No. 8  (Download full text)
Stan Hurn and Ralf Becker
Testing for nonlinearity in mean in the presence of heteroskedasticity.

No. 7  (Download full text) (published)
Adrian Pagan and Hashem Pesaran
On Econometric Analysis of Structural Systems with Permanent and Transitory Shocks and Exogenous Variables.

No. 6  (Download full text) (published)
Martin Fukac and Adrian Pagan
Limited Information Estimation and Evaluation of DSGE Models.

No. 5  (Download full text)
Andrew E. Clark, Paul Frijters and Michael A. Shields
Income and Happiness: Evidence, Explanations and Economic Implications.
No. 4  (Download full text)
Louis J. Maccini and Adrian Pagan
Inventories, Fluctuations and Business Cycles.

No. 3  (Download full text)
Adam Clements, Stan Hurn and Scott White

No. 2  (Download full text)
Stan Hurn, J.Jeisman and K.A. Lindsay

No. 1  (Download full text)
Adrian Pagan and Don Harding
The Econometric Analysis of Constructed Binary Time Series.