On the efficacy of techniques for evaluating multivariate volatility forecasts

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

R. Becker
Economics, School of Social Sciences, University of Manchester.

March 16, 2012

Abstract
The performance of techniques for evaluating multivariate volatility forecasts are not yet as well understood as their univariate counterparts. This paper aims to evaluate the efficacy of a range of traditional statistical-based methods for multivariate forecast evaluation together with methods based on underlying considerations of economic theory. It is found that a statistical-based method based on likelihood theory and an economic loss function based on portfolio variance are the most effective means of identifying optimal forecasts of conditional covariance matrices.

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

Acknowledgments This is a revised version of the paper presented at the 2009 Annual Conference of the Society for Financial Econometrics (SoFiE) in Geneva.
1 Introduction

There is little doubt surrounding the importance of forecasts of volatility for many financial applications. There is a well established literature that considers the modeling and forecasting of volatility in the univariate context, see amongst others Gourieroux and Jasiak (2001) for an overview. In parallel, a great deal of research has focused on the issue of forecast evaluation. Hansen and Lunde (2006), and Patton (2011) have examined the ability of various loss functions to consistently rank competing volatility forecasts. They find that given volatility is latent, along with the noise in volatility proxies, not all loss functions are robust to this noise.

Beyond the univariate case, providing an accurate forecast of the conditional covariance matrix of financial returns is a crucial element of optimal portfolio allocation. Consequently, there now exists a rich literature on multivariate volatility modeling (see Andersen, Bollerslev, Christoffersen and Diebold, 2006, for a recent survey). The evaluation of the forecasts from these multivariate volatility models is currently a fertile area of research. The survey of the forecasts from these multivariate volatility models is currently a fertile area of research. The survey of volatility forecast evaluation by Patton and Sheppard (2009), and an analysis of multivariate statistical loss functions undertaken by Laurent, Rombouts and Violante (2009) once again highlight that not all loss functions are robust to noise in the volatility proxy.

While statistical loss functions form the basis of many forecast evaluations, an alternative approach is to derive a measure of forecast efficacy within an underlying economic framework. In the multivariate setting, the portfolio optimisation context provides the means for this style of evaluation. Using either a measure of risk or utility from the portfolios generated from the competing forecasts, one has a measure of economic value that can be used discriminate between the forecasts. Of course, these measures may not necessarily identify the correct forecast as the best forecast. However, Patton and Sheppard (2009) have shown that the realised variance of the global minimum variance portfolio is minimised when the correct forecast is employed. Similar results relating to the variance and utility of other mean-variance efficient portfolios have been presented by Engle and Colacito (2006), and West, Edison and Cho (1993), respectively, although their findings rely on expectations rather than realisations.

In addition to ranking relative performance, these measures can be used within tests of predictive accuracy to determine if one forecast is statistically superior to others. To date, the ability of these measures to discriminate between forecasts within such tests is not well understood. Engle and Colacito (2006) have examined portfolio variance within predictive accuracy tests in simulated and empirical environments, finding that it is an effective evaluation approach. However, they did not compare this measure to others. Laurent, Rombouts and Violante (2009, 2010) do compare a range of statistical loss functions within a predictive accuracy test, but only
in an empirical setting and without comparison to economic loss functions. Furthermore, no insight has been provided on how the dimensionality of the forecasts affects forecast evaluations. Thus, the central contribution of this paper is to provide an analysis of the efficacy of the metrics commonly used to evaluate competing multivariate volatility forecasts within tests of predictive accuracy and, in doing so, examine if dimensionality impacts on the performance of such tests. Hence, the results from this study expands our understanding of the relative merits of the competing approaches. To perform the analysis, the properties of loss functions are analysed first. Second, a simulation study that employs the model confidence set (MCS) approach of Hansen, Lunde and Nason (2011) is undertaken to establish how well the proposed loss functions differentiate between competing forecasts within a test of predictive accuracy. Finally, an empirical analysis is then undertaken based on forecasting the conditional covariance matrix of a set of asset returns including equities, bonds and commodities. The purpose of the empirical analysis is not to identify the optimal forecasting model from an exhaustive list, and hence is not the focus of the paper. It is designed to examine whether the behaviour, or relative performance of the loss functions is similar in a case where the data generating process is unknown. The general conclusion to emerge from this research is that the likelihood-based statistical loss function and the portfolio variance-based economic loss function are best able to differentiate between competing forecasts.

The paper proceeds as follows. Section 2 provides an overview of loss functions used for evaluating volatility forecasts with Section 3 discussing some important properties of these loss functions. Section 4 outlines the econometric technique used for comparing the performance of competing forecasts. Section 5 reports simulation evidence relating to the ability of a number of loss functions to distinguish between competing forecasts. Section 6 provides an empirical investigation into the forecast performance of a range of multivariate models. Section 7 provides concluding comments.

2 Loss Functions for Evaluating Volatility Forecasts

Consider a system of $N$ asset returns

$$r_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim F(0, \Sigma_t),$$

where $r_t$ is an $N \times 1$ vector, $\varepsilon_t$ is an $N \times 1$ vector of disturbances and $F$ is some multivariate distribution. For the moment, the dynamics of $\mu_t$, the $N \times 1$ vector of conditional expected returns whose dynamics are left unspecified. In subsequent simulation and empirical analysis the process governing $\mu_t$ will be clearly stated. The conditional covariance matrix $\Sigma_t$ of the
disturbances is unobservable and the central problem to be addressed in this paper is how best to evaluate the accuracy of any forecast, $H_t$, of this conditional covariance matrix.

### 2.1 Statistical Loss Functions

Two statistical loss functions will be considered.

**Mean Square Error (MSE)**

Let $\hat{\Sigma}_t$ be an observable proxy for $\Sigma_t$ such as the realized covariance matrix proposed by Andersen, Bollerslev, Diebold and Labys (2003), or the simpler outer-product of observed demeaned daily returns, $\varepsilon_t \varepsilon'_t$. The MSE criterion is simply the mean squared distance between the volatility forecast $H_t$ and the volatility proxy $\hat{\Sigma}_t$

$$L_{t}^{\text{MSE}}(H_t, \hat{\Sigma}_t) = \frac{1}{N^2} \text{vec}(H_t - \hat{\Sigma}_t)' \text{vec}(H_t - \hat{\Sigma}_t),$$

where the vec(·) operator represents the column stacking operator. By convention all $N^2$ elements of the conditional covariance matrices $H_t$ and $\hat{\Sigma}_t$ are compared, notwithstanding the fact that there are only $N(N + 1)/2$ distinct elements in these matrices.

**Quasi-likelihood Function (QLK)**

Given the forecast of the conditional volatility, $H_t$, the value of the quasi log-likelihood function of the asset returns assuming a multivariate normal likelihood is

$$L_{t}^{\text{QLK}}(H_t) = \log |H_t| + \varepsilon'_t H_t^{-1} \varepsilon_t.$$  

This is not a distance measure in the vein of the MSE, but it does allow different forecasts of $\Sigma_t$ to be compared.

### 2.2 Economic Loss Functions

Three economic loss functions are considered. The underlying theory from which they derive their value is the general mean-variance portfolio optimization problem

$$\min_{w_t} w'_t \Sigma_t w_t \quad s.t. \quad w'_t \hat{\mu}_t = \mu_0,$$

where $w_t$ is an $N \times 1$ vector of portfolio weights, $\hat{\mu}_t$ is a vector of expected returns and $\mu_0$ is the target return for the portfolio. For this framework to be of practical use in terms of defining loss functions two issues need to be addressed. First, the true conditional covariance matrix, $\Sigma_t$, is unobserved so the portfolio optimization problem must be solved using the forecast of the

---

1 This form for the QLK loss function is same as that shown in Patton and Sheppard (2009) up to an additive constant.
conditional covariance matrix, $H_t$. Second, the dynamic process driving mean returns, $\mu_t$, is also unknown and is usually estimated by some parametric specification which may be denoted $\hat{\mu}_t$. With these modifications, this framework lends itself to the specification of economic loss functions defined on solution to the portfolio optimization problem.

**Variance of the Returns to the Minimum Variance Portfolio (MVP)**

The unconstrained solution to the problem posed in equation (4) is

$$ w_t = \frac{H_t^{-1} \hat{\mu}_t}{\hat{\mu}_t' H_t^{-1} \hat{\mu}_t} \mu_0. \quad (5) $$

Let $\boldsymbol{1}$ be the $N \times 1$ unit vector, then $1 - w_t' \boldsymbol{1}$ may be invested in the risk-free asset.

The MVP loss function is now defined as the variance of the return on a portfolio constructed using the weights $w_t$ and is given by

$$ \mathcal{L}_{t}^{\text{MVP}}(H_t) = w_t' r_t r_t' w_t. \quad (6) $$

**Variance of the Returns to the Global Minimum Variance Portfolio (GVP)**

The weights for the unique global minimum variance portfolio (GVP) are determined by solving

$$ \min_{w_t} w_t' \Sigma_t w_t \quad s.t. \quad w_t' \boldsymbol{1} = 1, \quad (7) $$

which avoids assumptions regarding $\mu_t$. For a given $H_t$, the analytical solution to the weights of this portfolio is

$$ w_t = \frac{H_t^{-1} \boldsymbol{1}}{\boldsymbol{1}' H_t^{-1} \boldsymbol{1}}. \quad (8) $$

The variance of the returns to this portfolio is given by

$$ \mathcal{L}_{t}^{\text{GVP}}(H_t) = w_t' r_t r_t' w_t. \quad (9) $$

**Utility from the Returns to the Minimum Variance Portfolio (UVP)**

Given a volatility forecast, $H_t$, the appropriate minimum-variance portfolio weights, $w_t$, can be computed using equation (5), and based on these weights the returns to the portfolio may be denoted $R_{p,t} = w_t' r_t$. Fleming, Kirby and Ostdiek (2001, 2003) propose that the utility of an investor with quadratic preferences based on $R_{p,t}$ can be used as an effective metric for comparison of volatility forecasts. The loss function is

$$ \mathcal{L}_{t}^{\text{UVP}}(H_t) = W_0 \left[ (1 + R_f + R_{p,t}) - \frac{\gamma}{2 (1 + \gamma)} (1 + R_f + R_{p,t})^2 \right], \quad (10) $$

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 of relative risk aversion. Consistent with the loss functions discussed earlier, equation (10) is a conditional measure. On the other hand, Fleming, Kirby and Ostdiek (2001, 2003) produce an unconditional measure of relative forecast performance from this utility function.

3 Some Properties of the Loss Functions

Ideally, the loss functions described in the previous section should be theoretically coherent in the sense that the loss function should reach a minimum when the volatility forecast, $H_t$, is equal to the true volatility, $\Sigma_t$. Patton and Sheppard (2009) show that MSE and QLK belong to a wider class of statistical loss functions that satisfy this requirement, but are also robust to noise in the volatility proxy. Laurent et al. (2009) confirm the robustness of these two loss functions.

While less is known about the properties of the economic loss functions, Patton and Sheppard (2009) demonstrate that the GVP loss function reaches a minimum when $H_t = \Sigma_t$. No equivalent results are yet available for the MVP or UVP loss functions, although Engle and Colacito (2006) and West et al. (1993) show the weaker result that the expected value of the loss function is minimized when $H_t = \Sigma_t$. The section uses the approach of Patton and Sheppard (2009) to explore the properties for the MVP and UVP cases.

Within the context of portfolio allocation, define $w_t$ as the vector of weights generated from $\Sigma_t$, $\tilde{w}_t$ as a vector of incorrect weights generated from $H_t$, and $c_t$ as a vector of weighting errors $(w_t - \tilde{w}_t)$ due to $H_t \neq \Sigma_t$. The impact on portfolio variance due to $H_t \neq \Sigma_t$ can be expressed as

$$\tilde{w}_t^\prime \Sigma_t \tilde{w}_t - w_t^\prime \Sigma_t w_t = (w_t + c_t)^\prime \Sigma_t (w_t + c_t) - w_t^\prime \Sigma_t w_t$$

$$(11)$$

$$= w_t^\prime \Sigma_t w_t + 2c_t^\prime \Sigma_t w_t + c_t^\prime \Sigma_t c_t - w_t^\prime \Sigma_t w_t$$

$$= 2c_t^\prime \Sigma_t \frac{\tilde{\mu}_t}{\tilde{\Sigma}_t} \mu_0 + c_t^\prime \Sigma_t c_t$$

$$= 2 \frac{c_t^\prime \tilde{\mu}_t}{\tilde{\mu}_t^\prime \tilde{\Sigma}_t} \mu_0 + c_t^\prime \Sigma_t c_t$$

$$= c_t^\prime \Sigma_t c_t,$$

as $w_t^\prime \tilde{\mu}_t = \mu_0$ and $w_t^\prime \tilde{\mu}_t + c_t^\prime \tilde{\mu}_t = \mu_0$ hence $c_t^\prime \tilde{\mu}_t = 0$. Given that $c_t^\prime \Sigma_t c_t \geq 0$, an incorrect forecast cannot produce a smaller variance than when $H_t = \Sigma_t$. This result holds for any given non-zero vector of expected returns.

Extending the analysis of Patton and Sheppard (2009) to the UVP case leads to a vastly different conclusion. Once again by defining $c_t$ as a vector of weighting errors due to $H_t \neq \Sigma_t$, the impact on the UVP loss function may be highlighted. Using the loss function specified in equation (10),
and for simplicity assuming $W_0 = 1$, the value of the function using the forecast $H_t$ is
\[
\left[ (1 + R_f + \omega_t' r_t + c_t' r_t) - \frac{\gamma}{2(1 + \gamma)} (1 + R_f + \omega_t' r_t + c_t' r_t)^2 \right],
\]
which may be subtracted from the loss obtained by using $\Sigma_t$
\[
\left[ (1 + R_f + \omega_t' r_t) - \frac{\gamma}{2(1 + \gamma)} (1 + R_f + \omega_t' r_t)^2 \right],
\]
to yield the following expression
\[
\Delta L_{\text{UVP}} = c_t' r_t - \frac{\gamma}{2(1 + \gamma)} \left( 2R_f c_t' r_t + 2c_t' r_t + 2c_t' r_t' w_t + c_t' r_t' c_t \right)
\]
\[
= \left[ 1 - \frac{\gamma}{(1 + \gamma)} (1 + R_f) \right] c_t' r_t - \frac{\gamma}{(1 + \gamma)} c_t' r_t' w_t - \frac{\gamma}{2(1 + \gamma)} c_t' r_t' c_t. \tag{12}
\]

The first two terms of this expression cannot be signed, that is
\[
\left[ 1 - \frac{\gamma}{(1 + \gamma)} (1 + R_f) \right] c_t' r_t \leq 0,
\]
and
\[
\frac{\gamma}{(1 + \gamma)} c_t' r_t' w_t \leq 0.
\]
The final term, however, is always positive:
\[
\frac{\gamma}{2(1 + \gamma)} c_t' r_t' c_t > 0.
\]
It therefore appears that $\Delta L_{\text{UVP}} \leq 0$, a result which implies that $c_t' r_t > 0$ may result in $H_t \neq \Sigma_t$ being identified as a superior forecast relative to $\Sigma_t$.

Therefore, relative to the loss functions discussed previously, it is expected that the UVP loss function may experience difficulty in distinguishing between competing forecasts.

### 4 Comparing Forecast Performance

The task is now to set up the procedure by which alternative forecasts of $H_t$ may be compared. A popular method for comparing two competing forecasts, say $H_t^a$ and $H_t^b$, is the pairwise test for equal predictive accuracy (EPA), due to Diebold and Mariano (1995) and West (1996). Let $L(H_t)$ represent a generic loss function defined on a volatility forecast, then the relevant null

\footnote{Replacing the outer-product of returns, $r_t r_t'$, with the true covariance matrix, $\Sigma_t$, in order to be consistent with portfolio variance analysis, results in the same conclusion.}
and alternative hypothesis concerning \( H^a_t \) and \( H^b_t \) are

\[
H_0 : \mathbb{E}[\mathcal{L}(H^a_t)] = \mathbb{E}[\mathcal{L}(H^b_t)] \quad (13)
\]

\[
H_A : \mathbb{E}[\mathcal{L}(H^a_t)] \neq \mathbb{E}[\mathcal{L}(H^b_t)].
\]

The test is based on the computation of

\[
DMW_T = \frac{\bar{d}_T}{\sqrt{\text{var}[\bar{d}_T]}}, \quad \bar{d}_T = \frac{1}{T} \sum_{t=1}^{T} d_t, \quad d_t = \mathcal{L}(H^a_t) - \mathcal{L}(H^b_t),
\]

where \( \hat{\text{var}}[\bar{d}_T] \) is an estimate of the asymptotic variance of the average loss differential, \( \bar{d}_T \).

The EPA test is limited in its applicability by the fact that it can only deal with pairwise comparisons. There are two approaches to comparing more than two forecasts, namely, the Reality Check of White (2000) and the Superior Predictive Ability (SPA) test of Hansen (2005). Both these procedures require the specification of a benchmark forecast, say \( \tilde{H}_t \), and the test is for whether or not any of the proposed forecasts outperform the benchmark. The null and alternative hypotheses for comparing this benchmark against \( m \) alternatives, denoted \( H^i_t, i = 1, \ldots, m, \) are respectively

\[
H_0 : \mathbb{E}[\mathcal{L}(\tilde{H}_t)] \leq \min_{i} \mathbb{E}[\mathcal{L}(H^i_t)] \quad (15)
\]

\[
H_A : \mathbb{E}[\mathcal{L}(\tilde{H}_t)] > \min_{i} \mathbb{E}[\mathcal{L}(H^i_t)].
\]

The method for comparison that will be adopted in this paper is a modified version of the SPA approach, known as the Model Confidence Set (MCS) introduced by Hansen et al. (2011), in which it is not necessary to specify a benchmark forecast. The MCS has traditionally been employed in the univariate setting, but translates seamlessly into a multivariate setting when the loss function generates a scalar measure. Recent applications include Laurent et al. (2009, 2010).

The procedure starts with a full set of candidate models \( \mathcal{M}_0 = \{1, \ldots, m_0\} \) and then sequentially trims the elements of \( \mathcal{M}_0 \) thereby reducing the number of viable models.

Prior to starting the sequential elimination procedure, all loss differentials between models \( i \) and \( j \) are computed,

\[
d_{ij,t} = \mathcal{L}(H^i_t) - \mathcal{L}(H^j_t), \quad i, j = 1, \ldots, m_0, \quad t = 1, \ldots, T, \quad (16)
\]

At each step, the hypothesis

\[
H_0 : \mathbb{E}(d_{ij,t}) = 0, \quad \forall i > j \in \mathcal{M} \quad (17)
\]
is tested for a set of models $\mathcal{M} \subset \mathcal{M}_0$, with $\mathcal{M} = \mathcal{M}_0$ at the initial step. If $H_0$ is rejected at the significance level $\alpha$, the worst performing model is removed and the process continues until non-rejection occurs with the set of surviving models being the MCS, $\hat{\mathcal{M}}_\alpha$. If a fixed significance level $\alpha$ is used at each step, $\hat{\mathcal{M}}_\alpha$ contains the best model from $\mathcal{M}_0$ with $(1 - \alpha)$ confidence.\(^3\)

This test is conducted using a test statistic similar to that proposed by Diebold-Mariano-West in equation (14). The relevant $t$-statistic, $t_{ij}$, provides scaled information on the average difference in the forecast quality of models $i$ and $j$ is

$$
t_{ij} = \frac{\tilde{d}_{ij}}{\sqrt{\hat{\text{var}}(\tilde{d}_{ij})}}, \quad \tilde{d}_{ij} = \frac{1}{T} \sum_{t=1}^{T} d_{ij,t}. \tag{18}
$$

where $\hat{\text{var}}(\tilde{d}_{ij})$ is an estimate of $\text{var}(\tilde{d}_{ij})$ and is obtained from a bootstrap procedure described in Hansen, Lunde and Nason (2003), and Becker and Clements (2008). In order to decide whether the size of the MCS must be reduced at any given stage, the null hypothesis in equation (17) must be tested. The main difficulty stems from the fact 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. (2011) propose the use of the range statistic

$$
T_R = \max_{i,j \in \mathcal{M}} |t_{ij}| = \max_{i,j \in \mathcal{M}} \frac{|\tilde{d}_{ij}|}{\sqrt{\hat{\text{var}}(\tilde{d}_{ij})}}, \tag{19}
$$

or the semi-quadratic statistic,

$$
T_{SQ} = \sum_{i,j \in \mathcal{M}, i < j} t_{ij}^2 = \sum_{i,j \in \mathcal{M}, i < j} \frac{(\tilde{d}_{ij})^2}{\hat{\text{var}}(\tilde{d}_{ij})}, \tag{20}
$$

to achieve this objective. Both of these test statistics indicate a rejection of the null hypothesis in equation (17) for large values of the statistic. The theoretical distribution of the test statistic is a complex entity that depends on the covariance structure between the forecasts included in $\mathcal{M}$. In practice, the p-values for each of these test statistics must be obtained by bootstrapping. If the null hypothesis is rejected, the worst performing model is removed from $\mathcal{M}$. The latter is identified as $\mathcal{M}_i$ where

$$
i = \arg \max_{i \in \mathcal{M}} \frac{\tilde{d}_i}{\sqrt{\hat{\text{var}}(\tilde{d}_i)}}, \quad \tilde{d}_i = \frac{1}{m-1} \sum_{j \in \mathcal{M}} \tilde{d}_{ij}. \tag{21}
$$

The whole procedure is then repeated on the reduced set of models and the iterations continue until the null hypothesis cannot be rejected. The set of models which survives is then the model

\(^3\) Despite the testing procedure involving multiple hypothesis tests this interpretation is a statistically correct one. See Hansen et al. (2003) for details.
confidence set.

5 Simulation Experiments

This section describes the simulation experiments employed to determine the accuracy of the loss functions at differentiating between competing forecasts.

5.1 Data Generation

The data generating process (DGP) selected here is the Asymmetric Dynamic Conditional Correlation (ADCC) model of Cappiello, Engle and Sheppard (2006). Simulations were also produced using the simpler Dynamic Conditional Correlation model of Engle (2002) as the DGP. The results are qualitatively similar to those presented, but are omitted for the sake of brevity.

Consider again the system of $N$ asset returns

$$r_t = \mu + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma_t).$$

(22)

where $r_t$ is an $N \times 1$ vector of returns, $\mu$ is an $N \times 1$ vector of average returns and $\varepsilon_t$ is an $N \times 1$ vector of disturbances. The $N \times N$ conditional covariance matrix, $\Sigma_t$, takes the form

$$\Sigma_t = D_t R_t D_t,$$

(23)

where $D_t$ is a diagonal matrix of conditional standard deviations and $R_t$ is the conditional correlation matrix. The diagonal elements of $D_t$, $\sigma_{i,t}$, are given by

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

(24)

where $\varpi_i, \alpha_i, \theta_i$ and $\beta_i$ are parameters for the series $i$ and $S_{i,t-1}$ is an indicator variable that takes the value one if $\varepsilon_{i,t-1} < 0$ and zero otherwise. The conditional correlation matrix $R_t$ is given by

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2},$$

(25)

with

$$Q_t = Q (1 - \alpha - \beta) - \theta \bar{m} + \alpha (z_{t-1} z_{t-1}') + \theta m_{t-1} m_{t-1}' + \beta Q_{t-1}$$

(26)

and

$$z_t = \frac{\varepsilon_{i,t}}{\sqrt{\sigma_{i,t}}}.$$
where $\alpha$, $\beta$ and $\phi$ are parameters, $\bar{Q}$ is the unconditional correlation matrix of the asset returns, and $m_{t-1}$ and $\bar{m}$ are leverage effect measures. Specifically, the leverage effect measures are $m_{t-1} = \delta \odot z_{t-1}$, where $\delta$ is a dummy variable vector with elements $\delta_{i1} = 1$ if $z_{i,t-1} < 0$ and $\bar{m}$ is the sample average of the outer products of $m_t$.

To generate data consistent with the ADCC model, parameter values need to be provided. To make the exercise as realistic as possible these were obtained from the data set that is used in the empirical illustration to follow in Section 6. The data set comprises the futures contracts outlined in Table 1. In all cases, a roll from each contract to the subsequent one was set to five days prior to the maturity of the former. Daily prices for all contracts trading for the period 20 May 1998 to 5 March 2010 were collected from Bloomberg, yielding a sample of 2929 daily observations. Simulated data will be generated under the assumed ADCC DGP. The fixed level of expected returns $\mu$ is set to the unconditional mean level of returns shown in Table 1 for all series except S&P 500 (SP) and Nasdaq 100 (ND). Instead, these expected returns are set at more conventional levels of 5% p.a. and 6% p.a., respectively. Volatility parameter estimates reveal the familiar pattern found in most asset returns in that volatility is a persistent process. Throughout this paper, systems of three different dimensions, $N = 3$, 5 and 15 are considered. Specifically, the $N = 3$ case is based on SP, Ten-year U.S. Treasury bond (TY) and Gold (GC) futures contracts. For $N = 5$, ND and Crude Oil (CL) contracts are added whereas when $N = 15$ all contracts are used.

Estimates reported for correlation equations also show that conditional correlation is a persistent process. All required ADCC parameters are estimated using the entire data set, with a selection of the relevant parameters reported in Table 1. In addition to these parameters, the unconditional measures $\bar{Q}$, $\bar{m}$ from equations (24) and (26), along with an unconditional covariance matrix $\Sigma$ and $R$ are also estimated from the entire data set.

The simulation of synthetic data now proceeds as follows. Given $R_t$ (set to the unconditional value, $R$ at $t = 1$) a vector of correlated standardized returns, $z_t$, is generated as $z_t = v_t \sqrt{R_t}$ where the elements of $v_t \sim N(0, 1)$. Using equations (26) and (25), 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 $r_{i,t} = \mu_i + z_{i,t} \sqrt{\sigma_{i,t}}$ (with $\sigma_{i,t}$ set to the $i$th element of $\Sigma_t$ for $t = 1$) where $\mu_i$ is the $i$th element in $\mu$ in equation (22). Returns are then simulated using the conditional variances for each asset which are constructed iteratively from equation (24). In total, 2,100 returns are generated 1,000 times for each series. While forecasts for each simulation are generated over all 2,100 observations, the first 100 forecasts are not evaluated in order to provide a run-in period for

---

$^5$Simulation have also been conducted for sample sizes of 1100 and 5100. As expected, performance of the loss functions tends to improve with increasing sample size. For the sake of brevity, all these additional results are omitted and available from the authors on request.
<table>
<thead>
<tr>
<th>Contract</th>
<th>Code</th>
<th>Return Mean</th>
<th>Stdev</th>
<th>GARCH ω</th>
<th>α</th>
<th>β</th>
<th>δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soy Beans</td>
<td>S</td>
<td>5.01</td>
<td>26.02</td>
<td>0.0000</td>
<td>0.9280</td>
<td>0.0750</td>
<td>-0.0291</td>
</tr>
<tr>
<td>Live Cattle</td>
<td>LC</td>
<td>4.47</td>
<td>15.29</td>
<td>0.0000</td>
<td>0.9572</td>
<td>0.0274</td>
<td>0.0185</td>
</tr>
<tr>
<td>Pork Bellies</td>
<td>PB</td>
<td>2.31</td>
<td>33.96</td>
<td>0.0000</td>
<td>0.9165</td>
<td>0.0512</td>
<td>0.0371</td>
</tr>
<tr>
<td>Gold</td>
<td>GC</td>
<td>9.51</td>
<td>18.63</td>
<td>0.0000</td>
<td>0.9534</td>
<td>0.0652</td>
<td>-0.0489</td>
</tr>
<tr>
<td>Silver</td>
<td>SI</td>
<td>8.19</td>
<td>30.66</td>
<td>0.0000</td>
<td>0.9521</td>
<td>0.0703</td>
<td>-0.0447</td>
</tr>
<tr>
<td>Natural Gas</td>
<td>HG</td>
<td>16.74</td>
<td>30.14</td>
<td>0.0000</td>
<td>0.9578</td>
<td>0.0379</td>
<td>0.0001</td>
</tr>
<tr>
<td>Oil</td>
<td>CL</td>
<td>10.46</td>
<td>39.22</td>
<td>0.0000</td>
<td>0.9181</td>
<td>0.0311</td>
<td>0.0482</td>
</tr>
<tr>
<td>Heating Oil</td>
<td>HO</td>
<td>13.21</td>
<td>38.67</td>
<td>0.0000</td>
<td>0.9370</td>
<td>0.0349</td>
<td>0.0165</td>
</tr>
<tr>
<td>Five-year Treasury Bond</td>
<td>FV</td>
<td>1.26</td>
<td>4.68</td>
<td>0.0000</td>
<td>0.9135</td>
<td>0.1054</td>
<td>-0.0749</td>
</tr>
<tr>
<td>Ten-year Treasury Bond</td>
<td>TY</td>
<td>1.33</td>
<td>6.88</td>
<td>0.0000</td>
<td>0.9304</td>
<td>0.0799</td>
<td>-0.0372</td>
</tr>
<tr>
<td>S&amp;P500</td>
<td>SP</td>
<td>-1.43</td>
<td>22.23</td>
<td>0.0000</td>
<td>0.9331</td>
<td>0.0000</td>
<td>0.1165</td>
</tr>
<tr>
<td>Nasdaq100</td>
<td>ND</td>
<td>1.33</td>
<td>34.49</td>
<td>0.0000</td>
<td>0.9352</td>
<td>0.0223</td>
<td>0.0795</td>
</tr>
<tr>
<td>Yen</td>
<td>JY</td>
<td>0.12</td>
<td>11.69</td>
<td>0.0000</td>
<td>0.9341</td>
<td>0.0657</td>
<td>-0.0316</td>
</tr>
<tr>
<td>Euro</td>
<td>EC</td>
<td>1.31</td>
<td>10.32</td>
<td>0.0000</td>
<td>0.9650</td>
<td>0.0315</td>
<td>-0.0021</td>
</tr>
<tr>
<td>Swiss Franc</td>
<td>SF</td>
<td>0.64</td>
<td>11.18</td>
<td>0.0000</td>
<td>0.9692</td>
<td>0.0361</td>
<td>-0.0185</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=3</td>
</tr>
<tr>
<td>N=5</td>
</tr>
<tr>
<td>N=15</td>
</tr>
</tbody>
</table>

Table 1: Descriptive statistics and selected ADCC parameter estimates. Mean and standard deviations are reported in annualized percentage terms. These are used for simulating the conditionally correlated heteroskedastic returns. All estimates are based on the full dataset. All ADCC parameter values are estimated by maximum likelihood. Cells containing - indicate that ω is not required for the correlation dynamics.

5.2 Volatility Forecasts

To examine the performance of the loss functions, a range of models for producing one-step ahead multivariate volatility forecasts, $H_t$, are required. A number of models have been chosen for this purpose. While this is clearly not an exhaustive list, each model is able to generate volatility forecasts for moderately sized covariance matrices with the quality of their forecasts expected to vary widely.

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

\[ STAT : \quad H_t = \frac{1}{J} \sum_{j=1}^{J} \varepsilon_{t-j} \varepsilon'_{t-j}, \]  \hspace{1cm} (27)

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,

$$\text{MA} : \quad H_t = \frac{1}{M} \sum_{m=1}^{M} \varepsilon_{t-m} \varepsilon'_{t-m},$$  \hspace{1cm} (28)

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,

$$\text{EWMA} : \quad H_t = (1 - \lambda) \varepsilon_{t-1} \varepsilon'_{t-1} + \lambda H_{t-1},$$  \hspace{1cm} (29)

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 utilized is the exponentially weighted model of Fleming, Kirby and Ostdiek (2001, 2003), denoted below as EXP,

$$\text{EXP} : \quad H_t = \alpha \exp (-\alpha) \varepsilon_{t-1} \varepsilon'_{t-1} + \exp (-\alpha) H_{t-1},$$  \hspace{1cm} (30)

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 model used as the DGP, the Constant Conditional Correlation (CCC) model of Bollerslev (1990) and Dynamic Conditional Correlation model of Engle (2002) are also used. The CCC model is recovered by constraining the $\theta$ in equation (24) and the $\alpha, \beta$ and $\phi$ in equation (26) 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).

Note that the model parameters used for forecasting are those estimated on the basis of the full empirical dataset and are not re-estimated at each forecasting step. This implies that ADCC model produces forecasts on the basis of the correct DGP and all other forecasts originate from misspecified models. This setup was chosen in order to focus upon the ability of the various evaluation techniques at identifying the best forecasting model.

One-step ahead volatility forecasts are then generated for each of these 2, 100 observations using
each of these seven different forecasting methods. 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 \( r_{t-1} \). As mentioned earlier, the first 100 forecasts are not used in the forecast evaluations.

Finally, while MSE, QLK and GVP loss functions rely only on the volatility forecast and volatility proxy, the MVP and UVP require additional inputs. First, they require a target portfolio return. For this analysis, an annual target return, \( \mu_0 \), is set at 4%, 6% and 8%, however, results for only \( \mu_0 = 4\% \) are presented for brevity\(^6\). Second, a vector of expected returns is required. Fleming et al. (2001, 2003), and Engle and Colacito (2006) have shown that results are affected by the selected vector of returns. While both these studies offer approaches to overcome concerns about selecting the vector of expected return, neither approach is tractible within this simulation environment. As such, to provide some evidence on how MVP and UVP results change with different expected return assumptions two vectors are employed. The first, referred to as the truth, is the annualised \( \mu \) that was used when generating the returns. The second is the less accurate alternative based on the return standard deviations reported in Table 1. Specifically, the elements of this vector are set at one-fifth of the corresponding annualised standard deviation for all series except for the currencies which are set at one-tenth. To highlight which vector is being used in the analysis, \( \text{MVP}_\mu \) and \( \text{UVP}_\mu \), and \( \text{MVP}_\sigma \) and \( \text{UVP}_\sigma \) will indicate results for the truth and scaled standard deviation vectors, respectively. Finally, consistent with Fleming et al. (2001, 2003), \( R_f \) is set at 6% p.a..

## 5.3 Results

The performance of MSE and QLK are examined first and the results are presented in Table 2. The proxy for the covariance matrix, \( \hat{\Sigma}_t \) required for the application of the MSE is set to be the outer-product of the simulated disturbances \( \varepsilon_t \varepsilon_t' \). The first result column, headed Ave. Size, reports the average MCS size across the simulations. It shows that while both loss functions regularly produce rejections of EPA, noted by average sizes smaller than seven, QLK leads to smaller MCSs on average. Furthermore, under either loss function, the average size of the MCS decreases as \( N \) increases. The next two columns under the heading DGP report the DGP model’s appearances within the MCS. They show that the DGP model is rarely excluded from the MCS under either loss function and over the three dimensions considered. In line with reported average MCS sizes, the frequency with which the MCS contains only the DGP increases as \( N \) increases with the most notable increase associated with QLK (increasing from 4.4 to 60.3). Finally, results for the non-DGP models show that the MCS under QLK rarely

---

\(^6\)Results for \( \mu_0 = 6\% \) and 8\% are similar to those for \( \mu_0 = 4\% \) and are available from the authors on request.
contains models other than DCC and ADCC, whereas under MSE the MCS contains a wider range of models. Consistent with earlier results, the frequency with which non-DGP models are excluded from the MCS increases under QLK as \( N \) increases. A similar pattern is observed under MSE loss for all non-DGP models except DCC. Overall these results appear to show that QLK exhibits more power than MSE. This result is consistent with the univariate results of Patton (2011).

<table>
<thead>
<tr>
<th>Ave. DGP Non-DGP Models</th>
<th>MSE</th>
<th>QLK</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Size</td>
<td>in only</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>3</td>
<td>3.3</td>
<td>98.1</td>
</tr>
<tr>
<td>5</td>
<td>3.0</td>
<td>99.7</td>
</tr>
<tr>
<td>15</td>
<td>2.0</td>
<td>99.6</td>
</tr>
</tbody>
</table>

Table 2: Summary results for MCS (based on the range test statistic and \( \alpha = 0.05 \)) under MSE and QLK (DGP: ADCC). The first column reports the average MCS size across the simulations. Columns 2 and 3 report the percentage of simulations in which the DGP is in the MCS and is the only model in the MCS. All remaining columns report the percentage of simulation in which the non-DGP models are reported in the MCS.

<table>
<thead>
<tr>
<th>Ave. DGP Non-DGP Models</th>
<th>GVP</th>
<th>MVP ( \mu )</th>
<th>MVP ( \sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>Size</td>
<td>in only</td>
<td>STAT</td>
</tr>
<tr>
<td>----</td>
<td>------</td>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>3</td>
<td>3.6</td>
<td>98.8</td>
<td>0.7</td>
</tr>
<tr>
<td>5</td>
<td>2.5</td>
<td>97.5</td>
<td>1.2</td>
</tr>
<tr>
<td>15</td>
<td>1.9</td>
<td>99.0</td>
<td>5.8</td>
</tr>
<tr>
<td>3</td>
<td>3.3</td>
<td>99.0</td>
<td>2.2</td>
</tr>
<tr>
<td>5</td>
<td>2.3</td>
<td>98.1</td>
<td>3.4</td>
</tr>
<tr>
<td>15</td>
<td>1.9</td>
<td>98.4</td>
<td>5.7</td>
</tr>
<tr>
<td>3</td>
<td>3.1</td>
<td>99.9</td>
<td>4.3</td>
</tr>
<tr>
<td>5</td>
<td>2.3</td>
<td>99.5</td>
<td>6.8</td>
</tr>
<tr>
<td>15</td>
<td>1.9</td>
<td>99.5</td>
<td>10.2</td>
</tr>
</tbody>
</table>

Table 3: Summary results for MCS (based on the range test statistic and \( \alpha = 0.05 \)) under GVP, MVP \( \mu \) and MVP \( \sigma \) (DGP: ADCC). The first column reports the average MCS size across the simulations. Columns 2 and 3 report the percentage of simulations in which the DGP is in the MCS and is the only model in the MCS. All remaining columns report the percentage of simulation in which the non-DGP models is reported in the MCS.

Table contains the simulation results for the GVP, MVP \( \mu \) and MVP \( \sigma \) loss functions. Consistent with MSE and QLK results, average MCS sizes show that the portfolio variance measures regularly reject EPA within the MCS. As well, the average MCS sizes decrease as \( N \) increases.
A comparison of MCS sizes reveals that the MCSs based MVP are similar to each other and marginally smaller than their GVP counterparts. Surprisingly, it is noted that for $N = 3$ more inferior models are excluded from the MCS when the standard deviation based expected return vector is used rather than the truth. Despite some differences, the similarities in MCS sizes indicate that the expected return assumption does not necessarily have a dramatic impact.

In terms of the models within the MCS, irrespective of dimension, the DGP model is rarely excluded. Of the non-DGP models, those that are most similar to the DGP, EXP and DCC, are more likely to included in the MCS. However as $N$ increases, the EXP model is excluded from almost all MCSs. Finally, compared to MSE and QLK the MCSs based on portfolio volatility tend to be marginally smaller than MSE but larger than QLK.

<table>
<thead>
<tr>
<th></th>
<th>Ave. DGP Non-DGP Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>UVP $\mu,\gamma=1$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>UVP $\sigma,\gamma=1$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>UVP $\mu,\gamma=10$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>UVP $\sigma,\gamma=10$</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

Table 4: Summary results for MCS (based on the range test statistic) under UVP, $\mu_0 = 4\%$ (DGP: ADCC). Results are reported for risk aversion, $\gamma = 1, 10$. The first column reports the average MCS size across the simulations. Columns 2 and 3 report the percentage of simulations in which the DGP is in the MCS and is the only model in the MCS. All remaining columns report the percentage of simulation in which the non-DGP models are reported in the MCS.

The final simulation results reported are for the UVP loss function and they are presented in Table 4. Results are reported for two values of the risk aversion parameter, $\gamma = 1$ and $\gamma = 10$, and for the two expected return vectors. These results paint a considerably different picture to those discussed earlier. For either risk aversion coefficient or vector of expected returns, the average MCS size for all utility based measures is often close to the original seven models under consideration and it is quite clear that the null hypothesis of EPA is very rarely rejected. Results relating to the DGP being in the MCS show that while the DGP is excluded from the MCS relatively infrequently, it is rarely, if ever, the sole model remaining in the MCS. This pattern is
consistent with the few rejections of EPA identified in the first column. These conclusions are reinforced by the extremely high frequency with which non-DGP models are contained in the MCS. The simulation results show that in comparison to the preceding loss functions, UVP has virtually no power to distinguish between the forecasts. It is conjectured that this lack of power is primarily due to the leading term in equation (12) which reflects the possibility that errors in the portfolio weights can lead to additional unexpected portfolio return, \( c_t r_t \). This clearly increases utility even though the errors in the weights derive from poor volatility forecasts, thus limiting the ability of utility-based measures to distinguish between good and bad forecasts of volatility.

In summary, these simulation results indicate that the QLK is superior to the other loss functions in terms of its ability to discriminate between the forecasts. While MSE, GVP and MVP are not as powerful as QLK, they are still capable of discriminating between superior and inferior forecasts. Finally, UVP is clearly unable to differentiate between different forecasts. The results would appear consistent with the properties of the loss functions outlined in Section 3 where it was shown that observed returns have a direct impact on the performance of the UVP loss function.

6 Empirical Application

The evaluation methodology described above will now be applied to an empirical problem based on the futures returns described in Section 5.1. The purpose of this empirical study is not to attempt to identify the optimal forecasting model, but to consider whether the relative behaviour of the loss functions remains unchanged when the DGP is unknown. The empirical analysis will also be conducted assuming \( N = 3, 5 \) and 15, with the same combinations of contracts used in the previous simulation study.

The empirical study uses the volatility forecasting models described in Section 5. Where necessary, the proxy for \( \Sigma_t \) is the outer-product of observed demeaned returns \( \varepsilon_t \varepsilon_t' \). To begin, the observations corresponding to \( (t = 1, 2, ..., 1000) \) are used as the initial in-sample period. From this data, an estimate of the vector of expected returns \( \hat{\mu} \), 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 1,928 one step ahead forecasts. Parameter estimates for the EXP, 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.
Table 5: Empirical MCS using statistical loss functions, MSE and QLK. Range MCS p-values are reported (see equation (19)). * indicates the model is included in the MCS with 95% confidence whereas ** indicates 90%.

<table>
<thead>
<tr>
<th>N</th>
<th>STAT</th>
<th>MA</th>
<th>EWMA</th>
<th>EXP</th>
<th>CCC</th>
<th>DCC</th>
<th>ADCC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>0.241**</td>
<td>0.165**</td>
<td>0.286**</td>
<td>0.286**</td>
<td>0.406**</td>
<td>0.406**</td>
</tr>
<tr>
<td>MSE</td>
<td>5</td>
<td>0.008</td>
<td>0.176**</td>
<td>0.498**</td>
<td>0.265**</td>
<td>0.265**</td>
<td>0.498**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.048</td>
<td>0.124**</td>
<td>0.826**</td>
<td>0.124**</td>
<td>0.124**</td>
<td>0.556**</td>
</tr>
<tr>
<td>QLK</td>
<td>3</td>
<td>0.005</td>
<td>0.005</td>
<td>0.021</td>
<td>0.126**</td>
<td>0.005</td>
<td>0.126**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.117**</td>
<td>0.000</td>
<td>0.117**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.026</td>
<td>0.000</td>
<td>1.000**</td>
</tr>
</tbody>
</table>

Once again, the MVP and UVP rely on auxiliary information relating to a target return and vector of expected returns. Consistent with the simulation study, portfolios are formed with \( \mu_0 = 4\% \), 6\% and 8\%, but only results for \( \mu_0 = 4\% \) are reported. Three vector of expected returns are used in the analysis. They are the truth, \( \mu \), which is based on the full unconditional vector of expected returns, an in-sample mean that is recursively estimated, \( \hat{\mu}_t \), and a vector of scaled in-sample standard deviations, \( \hat{\sigma} \). Consistent the simulation study, the currency standard deviations are scaled by one-tenth and all others are scaled by one-fifth. Subscripts \( \mu, \hat{\mu}_t \) and \( \hat{\sigma} \) are added to the MVP and UVP titles to indicate which vector is being used in the analysis. Finally, \( R_f = 6\% \).

Table 5 contains the MCS results given the MSE and QLK loss functions. There is little change in the MCS as \( N \) increases under MSE. Only the simplistic STAT is eliminated when moving from \( N = 3 \) to \( N = 5 \) or \( N = 15 \). All other models continue to remain in the MCS. The results for QLK paint a very different picture with only DCC and ADCC remaining in the MCS across all \( N \), these being the most sophisticated models under consideration. Apart from EXP being included when \( N = 3 \) and \( N = 5 \), all other models are always excluded. Differences in the empirical results for MSE and QLK are consistent with the simulation results reported earlier in that QLK appears to have more power in distinguishing between competing models.

Table 6 reports the equivalent results for the GVP and MVP loss functions. Consistent with MSE and QLK, the number of models within the MCSs tends to decline as \( N \) increases. Furthermore, when EPA is rejected, the simple models are those excluded from the MCS. In terms of relative performance of the GVP and MVP measures, at \( N = 3 \) neither shows a greater ability than the other, but for \( N = 5 \) and \( N = 15 \) MVP has tended to produce smaller MCSs. Within the MVP results, variations in MCSs sizes demonstrate that the vector of expected returns significantly affects the reported MCS. For example at \( \alpha = 0.05 \) and \( N = 5 \), the MCS sizes of three, four and five are reported with the smallest associated with the true vector of expected returns. In fact, along with the GVP results, these findings show that better forecasts of the vector of
expected returns tend to produce smaller MCSs. Despite this finding, which is supported by results of Engle and Colacito (2006), variations in MCS size generally appear small and it can be noted that the models within the smallest portfolio variance MCSs are always included in the larger ones. Thus it is concluded that while better forecasts of the vector of expected returns would be preferred when evaluating volatility forecasts, as they tend to produce smaller MCSs, they are not crucial because similar MCSs are produced across a variety of expected returns. When compared to MSE and QLK, the MCS based portfolio variance are generally smaller than those from MSE and, surprisingly, are similar to QLK for the larger dimensions. As well, portfolio variance results indicate that the EXP model is the superior forecasting model rather than DCC or ADCC.

The MCS results given the UVP loss function in Table 7 once again reveal a different pattern to the preceding results. Given any of the three vectors of expected returns or either γ, UVP can rarely distinguish between any of the competing forecasts. The MCS in most cases contains all forecasts with only exception is when MA is excluded. This result is once again a reflection of the significant impact that observed returns have upon the performance of this loss function.

Given these results a number of interesting conclusions arise. Overall, the empirical performance of the MSE, QLK, GVP and UVP within the MCS is consistent with the simulation study. That is, QLK displays the greatest power among these loss functions, while UVP has little or no power. The performance MVP has not been as expected. Instead, it has displayed a power similar to QLK in the larger dimensions. While this result maybe sample specific, it highlights...
95% confidence whereas p-values are reported (see equation (19)).

<table>
<thead>
<tr>
<th>Model</th>
<th>N</th>
<th>STAT</th>
<th>MA</th>
<th>EWMA</th>
<th>EXP</th>
<th>CCC</th>
<th>DCC</th>
<th>ADCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$UVP_{\mu,\gamma=1}$</td>
<td>3</td>
<td>1.00**</td>
<td>0.673**</td>
<td>0.673**</td>
<td>0.673**</td>
<td>0.847**</td>
<td>0.673**</td>
<td>0.673**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.00**</td>
<td>0.144**</td>
<td>0.643**</td>
<td>0.286**</td>
<td>0.643**</td>
<td>0.643**</td>
<td>0.643**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1.00**</td>
<td>0.025</td>
<td>0.924**</td>
<td>0.924**</td>
<td>0.879**</td>
<td>0.892**</td>
<td>0.879**</td>
</tr>
<tr>
<td>$UVP_{\hat{\mu},\gamma=1}$</td>
<td>3</td>
<td>1.00**</td>
<td>0.055*</td>
<td>0.147**</td>
<td>0.147**</td>
<td>0.147**</td>
<td>0.147**</td>
<td>0.147**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.00**</td>
<td>0.477**</td>
<td>0.621**</td>
<td>0.621**</td>
<td>0.621**</td>
<td>0.621**</td>
<td>0.621**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1.00**</td>
<td>0.088*</td>
<td>0.898**</td>
<td>0.878**</td>
<td>0.878**</td>
<td>0.878**</td>
<td>0.878**</td>
</tr>
<tr>
<td>$UVP_{\hat{\sigma},\gamma=1}$</td>
<td>3</td>
<td>1.00**</td>
<td>0.602**</td>
<td>0.240**</td>
<td>0.452**</td>
<td>0.602**</td>
<td>0.310**</td>
<td>0.452**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.00**</td>
<td>0.837**</td>
<td>0.891**</td>
<td>0.891**</td>
<td>0.917**</td>
<td>0.891**</td>
<td>0.891**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1.00**</td>
<td>0.636**</td>
<td>0.636**</td>
<td>0.636**</td>
<td>0.636**</td>
<td>0.636**</td>
<td>0.636**</td>
</tr>
<tr>
<td>$UVP_{\mu,\gamma=10}$</td>
<td>3</td>
<td>0.998**</td>
<td>0.763**</td>
<td>0.763**</td>
<td>0.763**</td>
<td>1.00**</td>
<td>0.763**</td>
<td>0.763**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.00**</td>
<td>0.227**</td>
<td>0.776**</td>
<td>0.313**</td>
<td>0.958**</td>
<td>0.815**</td>
<td>0.776**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.981**</td>
<td>0.004</td>
<td>0.981**</td>
<td>1.00**</td>
<td>0.914**</td>
<td>0.981**</td>
<td>0.981**</td>
</tr>
<tr>
<td>$UVP_{\hat{\mu},\gamma=10}$</td>
<td>3</td>
<td>1.00**</td>
<td>0.129**</td>
<td>0.258**</td>
<td>0.250**</td>
<td>0.258**</td>
<td>0.258**</td>
<td>0.258**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.00**</td>
<td>0.592**</td>
<td>0.687**</td>
<td>0.701**</td>
<td>0.701**</td>
<td>0.701**</td>
<td>0.701**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.997**</td>
<td>0.050*</td>
<td>1.00**</td>
<td>0.955**</td>
<td>0.955**</td>
<td>0.955**</td>
<td>0.955**</td>
</tr>
<tr>
<td>$UVP_{\hat{\sigma},\gamma=10}$</td>
<td>3</td>
<td>1.00**</td>
<td>0.974**</td>
<td>0.315**</td>
<td>0.764**</td>
<td>0.974**</td>
<td>0.635**</td>
<td>0.764**</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.872**</td>
<td>0.877**</td>
<td>0.966**</td>
<td>0.940**</td>
<td>1.00**</td>
<td>0.966**</td>
<td>0.966**</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1.00**</td>
<td>0.939**</td>
<td>0.939**</td>
<td>0.939**</td>
<td>0.939**</td>
<td>0.939**</td>
<td>0.939**</td>
</tr>
</tbody>
</table>

Table 7: Empirical MCS using the economic loss function, UVP for $\gamma = 1, 10$. Range MCS p-values are reported (see equation (19)). * indicates the model is included in the MCS with 95% confidence whereas ** indicates 90%.
that economic loss functions can be as effective as statistical loss functions for discriminating between forecasts. Finally, in regards to the forecasting models, no one forecasting model is reported in all MCSs, but DCC, ADCC and EXP are generally among the superior models.

7 Conclusion

Techniques for evaluating univariate volatility forecasts are well understood and often rely on traditional statistical measures of accuracy. By contrast, the evaluation of multivariate volatility forecasts, where comparisons are often made in terms of an economic application such as portfolio allocation, is a less well developed strand of the literature. Often, disparate techniques are used by different studies for evaluation purposes. This paper has sought to contribute to our understanding in this area by evaluating a variety measures often used for distinguishing between competing multivariate volatility forecasts. Simulation results presented here indicate that the likelihood based statistical loss function, or variance based economic loss functions exhibit the most power and are the dominant approaches for evaluating multivariate volatility forecasts. Economic loss functions that rely on expected asset and realized portfolio returns and on investor utility have little or no power to distinguish between competing forecasts. The empirical study conducted generally supports these findings. Thus, if the goal is to determine which forecast from a set of competing forecasts is superior, these results would suggest to rely on either a statistical or portfolio variance based loss functions and to avoid measures based on investor utility.
References


