Evaluating multivariate volatility forecasts

Adam Clements
Mark Doolan
Stan Hurn
Ralf Becker

Evaluating multivariate volatility forecasts

A.E. Clements, M. Doolan, A.S. Hurn

School of Economics and Finance, Queensland University of Technology

R. Becker

Economics, School of Social Sciences, University of Manchester.

November 24, 2009

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 statistical-based methods, or economic loss functions based on portfolio variance are more effective in terms of identifying optimal forecasts than other indirect theory-based counterparts.

Corresponding author

Adam Clements

School of Economics and Finance

Queensland University of Technology

Brisbane, 4001

Qld, Australia

email `a.clements@qut.edu.au`

1 Introduction

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 survey of recent developments). One area of multivariate volatility modeling that is currently a fertile area of research is the evaluation of the efficacy of multivariate forecasts of conditional covariance matrices, for an overview of relevant techniques see Patton and Sheppard (2006). A number of metrics for evaluating forecasts have been developed, which fall naturally into two categories. Direct or *statistical* loss functions are based on the statistical properties of the asset returns and the multivariate volatility forecast (Anderson *et al.*, 2006). On the other hand, indirect or *economic* loss functions, derive a measure of forecast efficacy based on some underlying consideration of economic theory. Engle and Colacito (2006) compare the volatility of the minimum variance portfolio to differentiate between competing forecasts while Fleming, Kirby and Ostdiek (2001, 2003) evaluate competing forecasting methods on the basis of differences between the utility earned from selecting portfolios on the efficient frontier[1].

The central contribution of this paper is to provide an in-depth analysis of the efficacy of the metrics used to evaluate competing multivariate volatility forecasts. In so doing, the properties of loss functions are examined and a simulation study is undertaken to establish how well the proposed loss functions differentiate between competing forecasts. An empirical analysis is then undertaken based on forecasting conditional covariance matrix of a set of financial returns including U.S. based futures contracts written on the S&P 500 and NASDAQ equity indices, ten-year U.S. Treasury bonds, gold and crude oil. The general conclusion to emerge from this research is that either a likelihood based statistical, or portfolio variance based economic loss function exhibits greater power in differentiating between competing forecasts than a number of common alternatives.

The paper proceeds as follows. Section 2 provides an overview of loss functions used for evaluating volatility forecasts with Section 3 analysing 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.

---

[1]Strictly speaking most of the metrics considered are not *loss functions* but rather values of an objective function that allow forecast comparison. For consistency with the literature, the loss function terminology will be maintained in this paper.

2 Loss Functions for Evaluating Volatility Forecasts

Consider a system of N asset returns

$$r_t = \mu_t + \varepsilon_t \,, \qquad \varepsilon_t \sim \mathbf{F}(0, \Sigma_t) \,, \tag{1}$$

where r_t is an $N \times 1$ vector, μ_t the $N \times 1$ vector of conditional expected returns whose dynamics are left unspecified, ε_t is an $N \times 1$ vector of disturbances and $\mathbf{F}$ is some unspecified distribution. The conditional covariance matrix Σ_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 $\widehat{\Sigma}_t$ be an observable proxy for Σ_t such as the realized covariance matrix proposed by Andersen, Bollerslev, Diebold and Labys (2003). The MSE criterion is simply the mean squared distance between the volatility forecast H_t and the volatility proxy $\widehat{\Sigma}_t$

$$\mathcal{L}_t^{\text{MSE}}(H_t, \widehat{\Sigma}_t) = \frac{1}{N^2} \text{vec}(H_t - \widehat{\Sigma}_t)' \, \text{vec}(H_t - \widehat{\Sigma}_t) \tag{2}$$

where the vec$(\cdot)$ operator represents the column stacking operator. By convention all N^2 elements of the conditional covariance matrices H_t and $\widehat{\Sigma}_t$ 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 based on the assumption that ε_t is distributed as a multivariate normal distribution is[2]

$$\mathcal{L}_t^{\text{QLK}}(H_t) = \log |H_t| + \varepsilon_t' H_t^{-1} \varepsilon_t \,. \tag{3}$$

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

---

[2]This form for the QLK loss function is same as that shown in Patton and Sheppard (2006) up to an additive constant.

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

$$\min_{w_t} w_t' H_t w_t \qquad s.t. \quad w_t' \widehat{\mu}_t = \mu_0 \,, \tag{4}$$

where w_t is an $N \times 1$ vector of portfolio weights and μ_0 is the target return for the portfolio. As shown by Patton and Sheppard (2006), in particular cases, the volatility of the portfolio returns based on any H_t that is different from the true covariance matrix, Σ_t, will be greater than that given Σ_t. This naturally leads to economic loss functions defined on portfolios chosen on the basis of the forecast of conditional covariance, H_t.

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

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

$$w_t = \frac{H_t^{-1} \widehat{\mu}_t}{\widehat{\mu}_t' H_t^{-1} \widehat{\mu}_t} \, \mu_0 \,. \tag{5}$$

Let ι be the $N \times 1$ unit vector, then $1 - w_t' \iota$ may be invested in the riskfree 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

$$\mathcal{L}_t^{\mathrm{MVP}}(H_t) = \frac{1}{T} \sum_{t=1}^{T} w_t' r_t r_t' w_t. \tag{6}$$

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

The unique global minimum variance portfolio (GVP) is a special case of equation (5) which avoids assumptions regarding μ_t. Given H_t, the weights of this portfolio can be found by

$$w_t = \frac{H_t^{-1} \iota}{\iota' H_t^{-1} \iota}. \tag{7}$$

The variance of the returns to this portfolio is given by

$$\mathcal{L}_t^{\mathrm{GVP}}(H_t) = \frac{1}{T} \sum_{t=1}^{T} w_t' r_t r_t' w_t \,. \tag{8}$$

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 $R_{p,t}$. 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

$$\mathcal{L}^{\mathrm{UMP}}(H_t) = W_0 \left[(1 + R_f + R_{p,t}) - \frac{\gamma}{2(1+\gamma)} (1 + R_f + R_{p,t})^2 \right] \,, \tag{9}$$

4

where R_f is the risk-free rate of return, W_0 is a fixed initial level of wealth and γ is the investor's level of relative risk aversion.

3 Some Properties of the Loss Functions

Ideally, the loss functions described in the previous section should satisfy a minimum requirement in terms of consistency. Essentially, this requirement is that the loss function defined on the true unobservable conditional covariance matrix, Σ_t, be smaller in value than when some forecast $H_t \neq \Sigma_t$ is used. Patton and Sheppard (2006) show that MSE and QLK belong to a wider class of statistical loss functions that are robust, in this sense they reach an optimum when $H_t = \Sigma_t$. Patton (2006) further demonstrates that MSE and QLK are robust to noise in the volatility proxy, $\widehat{\Sigma}_t$ used for evaluating a forecast, H_t. Less is known about the properties of the economic loss functions. Patton and Sheppard (2006) demonstrate that the GVP loss function reaches a minimum when $H_t = \Sigma_t$. But no equivalent results are yet available for the MVP or UVP loss functions[3]. The section uses the approach of Patton and Sheppard (2006) to establish some theoretical results for the MVP and UVP cases.

Begin by defining w_t as the vector of weights generated from Σ_t, $\bar{w}_t$ as a vector of incorrect weights generated from H_t, and c_t as a vector of weighting errors $(\overline{w}_t - w_t)$ due to $H_t \neq \Sigma_t$. The impact on portfolio variance due to $H_t \neq \Sigma_t$ can be expressed as

$$
\begin{aligned}
\bar{w}_t'\Sigma_t\bar{w}_t - w_t'\Sigma_t w_t &= (w_t + c_t)'\Sigma_t(w_t + c_t) - w_t'\Sigma_t w_t \\
&= w_t'\Sigma_t w_t + 2c_t'\Sigma_t w_t + c_t'\Sigma_t c_t - w_t'\Sigma_t w_t \\
&= 2c_t'\Sigma_t \frac{\Sigma_t^{-1}\widehat{\mu}_t}{\widehat{\mu}_t'\Sigma_t^{-1}\widehat{\mu}_t}\mu_0 + c_t'\Sigma_t c_t \\
&= 2\frac{c_t'\widehat{\mu}_t}{\widehat{\mu}_t'\Sigma_t^{-1}\widehat{\mu}_t}\mu_0 + c_t'\Sigma_t c_t \\
&= c_t'\Sigma_t c_t,
\end{aligned}
\tag{10}
$$

as $w_t'\widehat{\mu}_t = \mu_0$ and $w_t'\widehat{\mu}_t + c_t'\widehat{\mu}_t = \mu_0$ hence $c_t'\widehat{\mu}_t = 0$. Given that $c_t'\Sigma_t c_t \geqslant 0$, an incorrect forecast cannot produce a smaller variance than when $H_t = \Sigma_t$. This is consistent with the GVP case but requires that all forecasts being compared assume the same vector of mean returns. This extends the result of Engle and Colacito (2006) by demonstrating this result without the need for expectations.

Extending the analysis of Patton and Sheppard (2006) 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

---

[3]Engle and Colacito (2006) and West, Edison and Cho (1993) show the weaker result that the *expected* value of the loss function is minimised when $H_t = \Sigma_t$.

on the UMP loss function may be highlighted. Using the loss function specified in equation (9), and for simplicity assuming $W_0 = 1$, the value of the function using the forecast H_t is

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

which may be subtracted from the loss obtained by using Σ_t

$$\left[(1 + R_f + w_t' r_t) - \frac{\gamma}{2(1+\gamma)} (1 + R_f + w_t' r_t)^2 \right]$$

to yield the following expression

$$\begin{aligned} \Delta \mathcal{L}^{\text{UMP}} &= c_t' r_t - \frac{\gamma}{2(1+\gamma)} \left(2 R_f c_t' r_t + 2 c_t' r_t + 2 c_t' r_t r_t' w_t + c_t' r_t r_t' c_t \right) \\ &= \left[1 - \frac{\gamma}{(1+\gamma)} (1 + R_f) \right] c_t' r_t - \frac{\gamma}{(1+\gamma)} c_t' r_t r_t' w_t - \frac{\gamma}{2(1+\gamma)} c_t' r_t r_t' c_t . \end{aligned} \quad (11)$$

The first two terms of this expression cannot be signed, that is

$$\left[1 - \frac{\gamma}{(1+\gamma)} (1 + R_f) \right] c_t' r_t \lesseqgtr 0$$

and

$$\frac{\gamma}{(1+\gamma)} c_t' r_t r_t' w_t \lesseqgtr 0 .$$

The final term, however, is always positive:

$$\frac{\gamma}{2(1+\gamma)} c_t' r_t r_t' c_t > 0 .$$

It therefore appears that $\Delta \mathcal{L}^{\text{UMP}} \lesseqgtr 0$, 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 Σ_t. Taking expectations, assuming c_t and r_t are uncorrelated, in this case yields

$$\begin{aligned} \mathbb{E} \left(\Delta \mathcal{L}^{\text{UMP}} \right) &= \left[1 - \frac{\gamma}{(1+\gamma)} (1 + R_f) \right] c_t' \mathbb{E} (r_t) - \frac{\gamma}{(1+\gamma)} c_t' \mathbb{E}(r_t r_t') w_t \\ &\quad - \frac{\gamma}{2(1+\gamma)} c_t' (r_t r_t') c_t \\ &= \left[1 - \frac{\gamma}{(1+\gamma)} (1 + R_f) \right] c_t' \mu_t - \frac{\gamma}{(1+\gamma)} c_t' \Sigma_t w_t - \frac{\gamma}{2(1+\gamma)} c_t' \Sigma_t c_t \\ &= - \frac{\gamma}{2(1+\gamma)} c_t' \Sigma_t c_t . \end{aligned} \quad (12)$$

Thus on average, UVP will identify Σ_t as the best forecast when the target return constraint is used in the portfolio optimisation problem. However, the result from equation (11) 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 which may lead to difficulties in distinguishing between competing forecasts.

4 Comparing Forecast Performance

Previous sections have defined generic loss functions, $\mathcal{L}(H_t)$, and discussed some of their properties. In this section, the task is to set up the procedure by which alternative forecasts of H_t may be compared. The technique to be employed is the Model Confidence Set (MCS) introduced by Hansen, Lunde and Nason (2003a). The MCS has traditionally been employed in the univariate setting, but all of the loss functions considered here generate a scalar measurement so that the MCS translates seamlessly into a multivariate setting.

When comparing two competing forecasts, H_t^a and H_t^b, Diebold and Mariano (1995) and West (1996) provide a pairwise test for equal predictive accuracy (EPA),

$$
\begin{aligned}
H_0 &\quad : \mathbb{E}[\mathcal{L}(H_t^a)] = \mathbb{E}[\mathcal{L}(H_t^b)] \\
H_A &\quad : \mathbb{E}[\mathcal{L}(H_t^a)] \neq \mathbb{E}[\mathcal{L}(H_t^b)].
\end{aligned}
\tag{13}
$$

This test is based on the computation of

$$
DMW_T = \frac{\overline{d}_T}{\sqrt{\widehat{\mathrm{avar}}[\overline{d}_T]}}, \quad \overline{d}_T = \frac{1}{T}\sum_{t=1}^{T} d_t, \, d_t = \mathcal{L}(H_t^a) - \mathcal{L}(H_t^b)
\tag{14}
$$

where $\widehat{\mathrm{avar}}[\overline{d}_T]$ is an estimate of the asymptotic variance of the average loss differential, $\overline{d}_T$. The EPA test is limited in its applicability by the fact that it can only deal with pairwise comparisons.

There are two main approaches to dealing with the common problem of comparing more than two forecasts. The Reality Check of White (2000) and the test for Superior Predictive Ability (SPA) of Hansen (2005) test whether any forecast outperforms a benchmark forecast, denoted as forecast a below,

$$
\begin{aligned}
H_0 &\quad : \mathbb{E}[\mathcal{L}(H_t^a)] \leq \min_{i \in B, C, \cdots} \mathbb{E}[\mathcal{L}(H_t^i)] \\
H_A &\quad : \mathbb{E}[\mathcal{L}(H_t^a)] > \min_{i \in B, C, \cdots} \mathbb{E}[\mathcal{L}(H_t^i)].
\end{aligned}
\tag{15}
$$

The MCS approach is essentially a modified version of the SPA test that has greater power and does not require a benchmark forecast to be chosen. It starts with a full set of candidate models $\mathcal{M}_0 = \{1, ..., m_0\}$ 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,

$$
d_{ij,t} = \mathcal{L}(H_t^i) - \mathcal{L}(H_t^j), \qquad i, j = 1, ..., m_0, \qquad t = 1, ..., T.
\tag{16}
$$

At each step, the EPA hypothesis

$$H_0 \ : \ \mathrm{E}(d_{ij,t}) = 0, \qquad \forall \ i > j \in \mathcal{M} \tag{17}$$

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 α, the worst performing model is removed and the process continues until non-rejection occurs with the set of surviving models being the MCS, $\widehat{\mathcal{M}}_\alpha^*$. If a fixed significance level α is used at each step, $\widehat{\mathcal{M}}_\alpha^*$ contains the best model from $\mathcal{M}_0$ with $(1-\alpha)$ confidence[4].

At the core of the EPA statistic is the t-statistic, along the lines of equation 14,

$$t_{ij} = \frac{\overline{d}_{ij}}{\sqrt{\widehat{\mathrm{var}}(\overline{d}_{ij})}} \tag{18}$$

where $\overline{d}_{ij} = \frac{1}{T}\sum_{t=1}^{T} d_{ij,t}$. t_{ij} provides scaled information on the average difference in the forecast quality of models i and j. The quantity $\widehat{\mathrm{var}}(\overline{d}_{ij})$ is an estimate of $var(\overline{d}_{ij})$ and is obtained from a bootstrap procedure described in Hansen et $al.$ (2003a) 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 evaluated. 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 $al.$ (2003a, 2003b) propose the use of the range statistic

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

and semi-quadratic statistic,

$$T_{SQ} = \sum_{\substack{i,j \in \mathcal{M} \\ i<j}} t_{ij}^2 = \sum_{\substack{i,j \in \mathcal{M} \\ i<j}} \frac{(\overline{d}_{ij})^2}{\widehat{\mathrm{var}}(\overline{d}_{ij})} \, , \tag{20}$$

to establish EPA. Both of these 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 $\mathcal{M}$, which effectively means that 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

$$i = \arg \max_{i \in \mathcal{M}} \frac{\overline{d}_i}{\sqrt{\widehat{\mathrm{var}}(\overline{d}_{i.})}} \, , \qquad \overline{d}_{i.} = \frac{1}{m-1} \sum_{j \in \mathcal{M}} \overline{d}_{ij} \, . \tag{21}$$

The tests for EPA are then conducted on the reduced set of models and the procedure continues to iterate until the null hypothesis of EPA is not rejected.

---

[4]Despite the testing procedure involving multiple hypothesis tests this interpretation is a statistically correct one. See Hansen et $al.$ (2003b) for details.

5 Simulation Experiments

This section describes the simulation experiments employed to highlight the efficacy with which the loss functions differentiate between competing forecasts.

5.1 Data Generation

The distribution of the system of N asset returns in equation (1) is

$$r_t \sim \Phi\left(\mu, \Sigma_t\right)$$

where $\Phi(\cdot)$ is the multivariate normal distribution with conditional covariance matrix, Σ_t, that takes the form

$$\Sigma_t = D_t R_t D_t, \tag{22}$$

where D_t is a diagonal matrix of conditional standard deviations and R_t is the conditional correlation matrix. The data generating process (DGP) selected here is the Asymmetric Dynamic Conditional Correlation (ADCC) model of Cappiello, Engle, and Sheppard (2006), which will now be described[5].

The diagonal elements of D_t, $\sigma_{i,t}$, are given by

$$\sigma_{i,t} = \varpi_i + \left(\alpha_i + \theta_i S_{i,t-1}\right) r_{i,t-1}^2 + \beta_i \sigma_{i,t-1}, \tag{23}$$

where $\varpi_i, \alpha_i, \theta_i$ and β_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

$$R_t = \operatorname{diag}(Q_t)^{-1/2} Q_t \operatorname{diag}(Q_t)^{-1/2}, \tag{24}$$

with

$$Q_t = \bar{Q}\left(1 - \alpha - \beta\right) - \theta\bar{m} + \alpha\left(z_{t-1}z_{t-1}'\right) + \theta m_{t-1}m_{t-1}' + \beta Q_{t-1}, \tag{25}$$

where α, β and ϕ are parameters, $\bar{Q}$ is the unconditional correlation matrix of the asset returns, z_t is a vector of standardized returns from $\frac{r_{i,t}}{\sqrt{\sigma_{i,t}}}$, and m_{t-1} and $\bar{m}$ are leverage effect measures. Specifically, the leverage effect measures are $m_{t-1} = \delta \odot z_{t-1}$, where δ 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.

The values of the parameters used in the simulations are set out in Table 1. To ensure that these parameter values are realistic, there are obtained from the estimation of the data set outlined below in Section 6 and used in the empirical application.

---

[5]Simulation results have also been produced assuming 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. These results are available from the authors upon request.

	Univariate GARCH					Correlation
	SP	ND	TY	GC	CL	
ϖ	0.0000	0.0000	0.0000	0.0000	0.0000	
β	0.9172	0.9268	0.9352	0.9582	0.9282	0.9615
α	0.0000	0.0259	0.0650	0.0597	0.0634	0.0219
θ	0.1414	0.0848	-0.0194	-0.0430	0.0018	0.0010

Table 1: Parameter values used to generate the conditionally correlated heteroskedastic returns. All parameter values are estimated from the entire empirical data set by maximum likelihood.

In addition to the parameters, the unconditional measures $\bar{Q}$, $\bar{m}$ from equations (23) and (25), along with an unconditional covariance matrix Σ and R are required. To provide realistic values, they are also estimated from the data set outlined below in Section 6. Finally, to mimic the empirical application, the order of the system being generated is set to five and the values of the means of the assets, μ, are $(0.03, 0.04, 0.005; 0.02; 0.025)'$ in *per annum* terms.

The simulation process can now be described as follows. Given R_t (set to the unconditional value, R at $t = 1$) a vector of correlated standardized returns is generated as $z_t = v_t \sqrt{R_t}$ where the elements of $v_t \sim N(0, 1)^6$. Using equations 25 and 24, 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 $r_{i,t} = z_{i,1} \sqrt{\sigma_{i,t}}$ (with $\sigma_{i,t}$ set to the *ith* element of Σ_t for $t = 1$). Returns are then simulated using the conditional variances for each asset which are constructed iteratively from equation 23. $2,000$ observations of five conditionally correlated heteroskedastic return series are then generated on the basis of the ADCC model with the parameters shown in Table 1.

One-step ahead volatility forecasts are then generated for each of these $2,000$ observations using seven different forecasting methods (see next Section). These seven sets of $2,000$ volatility forecast are then the basis on which the MCSs are produced. This procedure is repeated $1,000$ times.

5.2 Volatility Forecasts

The essence of the problem is that the conditional covariance matrix Σ_t is not observed and the methods used to generate one-step ahead multivariate volatility forecasts, H_t, will be utilized to compare the performance of the loss functions. A number of models have been chosen for this purpose each of which is able to generate volatility forecasts for moderately sized covariance

---

[6]A simulation study based on $t-$distributed errors has also been conducted. For the sake of brevity, only results pertaining to the normally distributed errors are reported below. Results based on the $t-$distributed errors are qualitatively similar and also indicate that QLK is the preferred loss function. These results are available from the authors upon request

matrices with the quality of their forecasts expected to vary.

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

$$H_t = \frac{1}{J} \sum_{j=1}^{J} \varepsilon_{t-j} \varepsilon'_{t-j}, \tag{26}$$

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,

$$H_t = \frac{1}{M} \sum_{m=1}^{M} \varepsilon_{t-m} \varepsilon'_{t-m}, \tag{27}$$

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,

$$H_t = (1 - \lambda) \, \varepsilon_{t-1} \varepsilon'_{t-1} + \lambda H_{t-1}, \tag{28}$$

where λ 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,

$$H_t = \alpha \exp\left(-\alpha\right) \varepsilon_{t-1} \varepsilon'_{t-1} + \exp\left(-\alpha\right) H_{t-1}, \tag{29}$$

where α 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 θ in equation 23 and the α, β and ϕ in equation 25 to zero, while DCC constrains θ and ϕ to zero. Estimation of the conditional correlation models rely on the two stage maximum likelihood procedure detailed by Engle and Sheppard (2001).

Each of these seven models 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}. Where required, we evaluate loss functions using a target return of $\mu_0 = 4\%$ [7].

It is important to 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.

5.3   Results

We begin by examining the performance of MSE and QLK in Table 2. Panel A shows that while both loss functions regularly produce rejections of EPA, QLK leads to smaller MCS sizes on average. Panel B shows that while both loss function rarely exclude the DGP, the instances where the MCS only contains the DGP are much more frequent under QLK. Panel C shows that under QLK, the MCS only ever contains the DCC and ADCC models, whereas MSE often cannot exclude a wider range of models from the MCS. Overall these results appear to show that QLK is the superior loss function of the two statistical loss functions.

Table 3 contains the simulation results for the GVP and MVP loss function. Beginning with the GVP results, it seems as though its performance as a loss function is somewhat better than that of MSE. Panel A shows that the average MCS size is either 2.57 or 2.30 for $\alpha = 0.05$ and 0.10 respectively, and lie between those from the MSE and QLK loss functions. Panel B shows that the DGP is hardly excluded from the MCS and the frequency with which MCS consists of only the DGP is somewhat below those of MSE and QLK. In Panel C, the incorrect models included in the MCS most frequently are the ADCC or DCC and EXP. Given these results, it appears as though the ability of GVP in distinguishing between forecasts is superior (inferior) to MSE (QLK).

Table 3 also reports the MCS simulation results for the MVP loss function with $\mu_0 = 4\%$. Panel A indicates that with the average size of the MCS being 2.14 and 1.89, MVP excludes inferior models at a rate marginally higher than the GVP. Panel B results show that, as with all loss function considered thus far, the DGP is rarely excluded from the MCS. Panel C results indicate that the MCS under this loss function is somewhat more selective than either the MSE or the GVP loss function with the simpler models, STAT and MA, surviving less often than with the afore mentioned loss functions. Thus it appears that MVP function fares better in

---

[7]Results for 2% and 6% are available upon request but are omitted for the sake of brevity.

	MSE		QLK	
α	0.05	0.10	0.05	0.10
Panel A: Average MCS size				
MCS Size	3.46	3.02	1.88	1.81
Panel B: DGP in MCS (%)				
DGP in MCS	97.4	96.4	99.5	98.7
Only DGP in MCS	5.0	10.3	11.9	17.5
Panel C: Non DGP models in MCS (%)				
STAT	23.7	18.1	0.0	0.0
MA	1.0	0.4	0.0	0.0
EWMA	0.9	0.3	0.0	0.0
EXP	76.3	63.0	0.0	0.0
CCC	52.0	34.8	0.0	0.0
DCC	94.9	89.5	88.1	82.5

Table 2: Summary results for MCS under MSE and QLK (DGP: Asymmetric DCC). 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

	GVP		MVP	
α	0.05	0.10	0.05	0.10
Panel A: Average MCS size				
MCS Size	2.57	2.30	2.14	1.89
Panel B: DGP in MCS (%)				
DGP in MCS	98.8	96.0	99.8	99.5
Only DGP in MCS	0.9	2.6	18.2	27.3
Panel C: Non DGP models in MCS (%)				
STAT	13.0	7.6	5.5	2.7
MA	2.6	1.4	0.3	0.1
EWMA	0.1	0.0	0.0	0.0
EXP	35.8	23.2	17.1	8.6
CCC	7.8	4.6	9.9	5.8
DCC	99.1	97.4	81.8	72.7

Table 3: Summary results for MCS under GVP and MVP (DGP: Asymmetric DCC). 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.

discriminating between the competing forecasts than the MSE and GVP loss functions. When compared to the QLK loss function, it becomes apparent it eliminate all but the DGP from the MCS more often than the QLK loss function, the MVP on average, leaves a greater number of, and a more varied selection of models in the MCS.

Finally, Table 4 reports the simulation results for the UVP loss function. Results are reported for $\mu_0 = 4\%$ and risk aversion, $\gamma = 1$, 10 and paint a vastly different picture from those discussed earlier. 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 remaining in the MCS. This pattern is 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 the preceding loss functions, UVP 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 3.

In summary, these simulation results indicate that the QLK and MVP loss functions, overall, are superior to the other loss functions in terms of their ability to discriminate between the forecasts. While the performance of MSE and GVP are similar in nature, the 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 major impact on the performance of the UVP 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 distinguish statistically between competing forecasts in terms of this loss function.

6 Empirical Application

The evaluation methodology described will now be applied to an empirical problem based on the returns to five financial assets. The dataset comprises U.S. based futures contracts written on the S&P 500 (SP) and NASDAQ (ND) equity indices, ten-year U.S. Treasury bonds (TY), gold (GC) and crude oil (CL)[8]. In all cases, a roll from each contract to the subsequent one was set to ten days prior to the maturity of the former. Daily returns for the period, 1 July 1997 to 30 June 2009 have been gathered, corresponding to a sample of 2983 daily return observations.

---

[8]Specifically, the crude oil contract is the light/sweet crude contract traded on NYMEX.

	UVP, $\gamma = 1$		UVP, $\gamma = 10$	
α	0.05	0.10	0.05	0.10
Panel A: Average MCS size				
MCS Size	6.89	6.76	6.88	6.72
Panel B: DGP in MCS (%)				
DGP in MCS	98.7	96.7	99.5	98.3
Only DGP in MCS	0.0	0.1	0.1	0.1
Panel C: Non DGP models in MCS (%)				
STAT	98.7	98.1	98.3	96.0
MA	98.1	95.9	96.3	92.9
EWMA	98.1	96.6	97.2	93.7
EXP	98.2	95.8	98.0	95.4
CCC	98.7	96.7	98.7	96.9
DCC	98.6	96.2	99.5	98.4

Table 4: Summary results for MCS under UVP, $\mu_0 = 4\%$. 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.

Figure 1 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 the current decade. This feature is obviously more pronounced in the NASDAQ series. Both equity series exhibit the falls associated with the recent global financial turmoil. 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 slowly 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 periods of rapidly rising and falling prices. These patterns are also evident in the descriptive statistics in reported in Table 5.

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 6 reports the unconditional correlations between of the futures returns. These correlations will vary across time, but examining the unconditional values is nevertheless useful. Not surprisingly, the equity indices exhibit very strong correlation. Apart from weak negative cor-
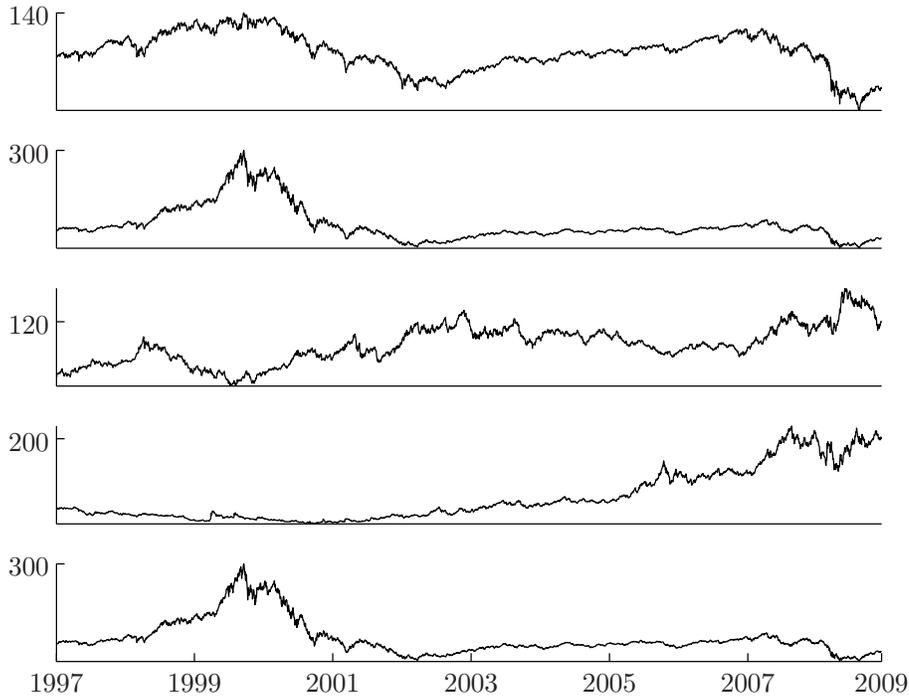
Figure 1: Cumulative returns from the five futures contracts SP, ND, TY, GC and CL are shown in the top to bottom panels respectively.

relation between equities and bonds, and weak positive correlation between Gold and Oil, there is little relationship between the other combinations.

The empirical study utilizes the models described in Section 5 with returns given by

$$r_t = \mu + \varepsilon_t. \tag{30}$$

To begin, the observations corresponding to $(t = 1, 2, ..., 1000)$ are 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

	SP	ND	TY	GC	CL
Mean (% p.a.)	-2.67	-1.54	1.54	5.83	1.44
Standard Deviation (% p.a.)	20.69	29.57	6.84	16.26	22.27
Skewness	0.03	0.09	-0.20	0.19	-0.30
Kurtosis	14.99	8.18	6.64	10.22	10.55
Jarque Bera (p-value)	0.00	0.00	0.00	0.00	0.00
Minimum Daily Return (%)	-10.48	-10.57	-2.50	-6.90	-9.61
Maximum Daily Return (%)	13.31	13.40	3.57	8.57	9.48

Table 5: Descriptive statistics for futures contract returns.

16

	SP	ND	TY	GC	CL
SP	1.00				
ND	0.82	1.00			
TY	−0.23	−0.21	1.00		
GC	−0.06	−0.07	0.08	1.00	
CL	0.18	0.10	−0.13	0.25	1.00

Table 6: Unconditional correlation matrix of futures returns.

extended to $(t = 1, 2, ..., 1001)$ and the process repeated giving a total of $1,983$ one step ahead forecasts. Results are also reported for two sub-sample capturing the first and second halves of this period. 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. MCS results will be presented using each loss function for the full out-of-sample period along with two sub-samples. Presenting the results in this manner should indicate whether the turmoil due to the recent financial crisis impacts upon the performance of any of the models and loss functions.

Table 7 contains the MCS results given the MSE loss function. Results clearly indicate that the MSE loss function has difficulty distinguishing between the models. In the first sub-sample, it appears as though only the MA and STAT forecasts are eliminated from the MCS. In comparison, the results given the QLK loss function shown in Table 8 show a much stronger distinction between the models. In the first sub-sample and full out-of-sample period the ADCC forecast is the sole model in the MCS and hence is statistically superior to all competing forecasts. Only the EXP and DCC forecasts are included in the first sub-sample. Overall, these results are consistent with the simulation results reported earlier in the MSE has relatively less power than QLK in distinguishing between the competing forecasts. Based on these statistical loss functions, it appears as though the relative performance of the forecasts remains quite stable across the different sub-samples.

Moving to the GVP loss function results in Table 9, its performance does seem to be affected by the large variations in volatility in the second sub-sample. In the first sub-sample only the EXP and DCC models remain in the MCS while in the second sub- and full samples all models with the exception of EWMA and STAT remain. While the simulation results reported earlier show that GVP does not exhibit the power of QLK, this loss function finds it more difficult to distinguish between forecasts in this particular period of high volatility. In contrast, results based on the MVP loss function indicate that it is more effective at eliminating clearly inferior models with the MCS remaining largely unchanged across the full and respective sub-samples.

MCS Size	1	2	3	4	5	6	7
Panel A: Full out-of-sample period							
Model	ADCC	DCC	EWMA	EXP	CCC	MA	STAT
p_R	1.00	0.29	0.29	0.29	0.27	0.20	0.15
Panel B: First half of out-of-sample period							
Model	EXP	ADCC	EWMA	DCC	CCC	MA	STAT
p_R	1.00	0.80	0.67	0.46	0.28	0.08	0.00
Panel C: Second half of out-of-sample period							
Model	ADCC	DCC	EWMA	CCC	EXP	MA	STAT
p_R	1.00	0.28	0.28	0.28	0.28	0.22	0.22

Table 7: Empirical MCS with MSE. Reading right to left, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported (see equation 19). Panel A, B and C presents the MCS for the full, first half and second half of the out-of-sample period.

MCS Size	1	2	3	4	5	6	7
Panel A: Full out-of-sample period							
Model	ADCC	DCC	EXP	CCC	MA	EWMA	STAT
p_R	1.00	0.05	0.05	0.00	0.00	0.00	0.00
Panel B: First half of out-of-sample period							
Model	EXP	ADCC	DCC	MA	CCC	EWMA	STAT
p_R	1.00	0.89	0.32	0.00	0.00	0.00	0.00
Panel C: Second half of out-of-sample period							
Model	ADCC	DCC	EXP	CCC	MA	EWMA	STAT
p_R	1.00	0.09	0.04	0.00	0.00	0.00	0.00

Table 8: Empirical MCS with QLK. Reading right to left, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported (see equation 19). Panel A, B and C presents the MCS for the full, first half and second half of the out-of-sample period.

This is consistent with results of the earlier simulation study.

MCS Size	1	2	3	4	5	6	7
Panel A: Full out-of-sample period							
Model	EXP	ADCC	DCC	MA	CCC	EWMA	STAT
p_R	1.00	0.95	0.95	0.22	0.22	0.06	0.00
Panel B: First half of out-of-sample period							
Model	EXP	DCC	MA	ADCC	EWMA	CCC	STAT
p_R	1.00	0.25	0.02	0.02	0.02	0.01	0.00
Panel C: Second half of out-of-sample period							
Model	ADCC	CCC	DCC	EXP	MA	EWMA	STAT
p_R	1.00	0.90	0.90	0.89	0.53	0.03	0.00

Table 9: Empirical MCS with GVP. Reading right to left, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported (see equation 19). Panel A, B and C presents the MCS for the full, first half and second half of the out-of-sample period.

MCS Size	1	2	3	4	5	6	7
Panel A: Full out-of-sample period							
Model	DCC	ADCC	EXP	MA	CCC	EWMA	STAT
p_R	1.00	0.96	0.96	0.03	0.03	0.01	0.00
Panel B: First half of out-of-sample period							
Model	EXP	ADCC	DCC	CCC	MA	EWMA	STAT
p_R	1.00	0.94	0.93	0.11	0.07	0.07	0.00
Panel C: Second half of out-of-sample period							
Model	DCC	ADCC	EXP	MA	CCC	EWMA	STAT
p_R	1.00	0.91	0.91	0.08	0.08	0.06	0.00

Table 10: Empirical MCS with MVP. Reading right to left, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported (see equation 19). Panel A, B and C presents the MCS for the full, first half and second half of the out-of-sample period.

The MCS results given the UVP loss function in Tables 11 and 12 once again reveal a different pattern to the preceding results. Given either γ, $U(MVP^{\mu_0})$ cannot distinguish between any of the competing forecasts with the MCS in both cases containing all forecasts. This result is once again a reflection of the significant impact that observed returns have upon the performance of

this loss function.

MCS Size	1	2	3	4	5	6	7
Panel A: Full out-of-sample period							
Model	STAT	CCC	EWMA	DCC	ADCC	EXP	MA
p_R	1.00	0.71	0.71	0.71	0.61	0.54	0.39
Panel B: First half of out-of-sample period							
Model	EWMA	STAT	ADCC	DCC	EXP	MA	CCC
p_R	1.00	0.91	0.91	0.91	0.76	0.76	0.52
Panel C: Second half of out-of-sample period							
Model	STAT	CCC	EWMA	DCC	ADCC	EXP	MA
p_R	1.00	0.74	0.38	0.24	0.23	0.23	0.23

Table 11: Empirical MCS with UVP, $\gamma = 1$. Reading right to left, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported. Panel A, B and C presents the MCS for the full, first half and second half of the out-of-sample period.

MCS Size	1	2	3	4	5	6	7
Panel A: Full out-of-sample period							
Model	STAT	CCC	DCC	EWMA	ADCC	EXP	MA
p_R	1.00	0.93	0.89	0.89	0.67	0.60	0.41
Panel B: First half of out-of-sample period							
Model	EWMA	DCC	ADCC	STAT	EXP	MA	CCC
p_R	1.00	0.96	0.96	0.96	0.86	0.86	0.47
Panel C: Second half of out-of-sample period							
Model	CCC	STAT	DCC	EWMA	ADCC	EXP	MA
p_R	1.00	0.83	0.35	0.35	0.35	0.35	0.35

Table 12: Empirical MCS with UVP, $\gamma = 10$. Reading right to left, each panel presents the order that the models are removed from the MCS. Range MCS p-values are reported. Panel A, B and C presents the MCS for the full, first half and second half of the out-of-sample period.

Given these results a number of interesting conclusions arise. In light of the earlier simulation results, it appears as though ADCC is the superior model from the forecasting perspective. It is essentially the only forecast in the MCS under the QLK loss function and remains in the MCS under the other loss functions in virtually every instance. Consistent with the results reported in the simulation study, the utility based loss function is not able to distinguish between any of

the forecasts. Counter-intuitively, what this result suggests is that if an investor is a quadratic utility maximizer, the choice of volatility model has no significant impact on utility.

7   Conclusion

Techniques for evaluating univariate volatility forecasts are well understood and often rely on traditional statistical measure 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. This paper has sought to contribute to understanding in this area by undertaking a substantial evaluation of a variety of methods 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 are the dominant approach for evaluating multivariate volatility forecasts. Economic loss functions that rely on expected asset and realised portfolio returns and on investor utility have weaker power to distinguish between competing forecasts. 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

Andersen T.G., Bollerslev T., Diebold F.X. and Labys P. (2003). Modeling and forecasting realized volatility. *Econometrica.* 71, 579-625.

Andersen, T.G., Bollerslev, T., Christoffersen, P.F. and Diebold, F.X. (2006), Volatility and Correlation Forecasting, in *Handbook of Economic Forecasting*, eds. Elliot, G., Granger, C.W.J. and Timmerman, A., Elsevier, Burlington.

Becker, R., and Clements, A. (2008), "Are combination forecasts of S&P 500 volatility statistically superior?", *The International Journal of Forecasting*, **24**, 122-133.

Bollerslev, T. (1990), "Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model", *Review of Economics and Statistics*, **72**, 498-505.

Cappiello, L., Engle, R.F. and Sheppard, K. (2004). "Asymmetric dynamics in the correlations of global equity and bond returns", *Journal of Financial Econometrics*, **4**, 537-572.

Campbell, J.Y., Lo, A.W. and MacKinlay, A.G. (1997), *The Econometrics of Financial Markets*, Princeton University Press, Princeton NJ.

Diebold, F.X. and Mariano, R.S. (1995), "Comparing predictive accuracy", *Journal of Business and Economics Statistics*, **13**, 253-263.

Engle, R.F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica,* **50**, 987-1007.

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

Engle, R. and Colacito, R. (2006), "Testing and valuing dynamic correlations for asset allocation", *Journal of Business and Economic Statisitcs*, **24**, 238-253.

Engle, R. F. and Sheppard, K. (2001), "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH", Working Paper, University of California San Diego.

Fleming, J., C. Kirby and B. Ostdiek. (2001), "The economic value of volatility timing", *The Journal of Finance*, **56**, 329-352.

Fleming, J., C. Kirby and B. Ostdiek. (2003). "The economic value of volatility timing using realized volatility", *Journal of Financial Economics*, **67**, 473-509.

Gourieroux, C. and Jasiak, J. (2001), *Financial Econometrics*. Princeton University Press, Princeton NJ.

Hansen, P. (2005), "A test for superior predictive ability", *Journal of Bsuiness and Economic Statistics*, **23**, 365-380.

Hansen, P., Lunde, A. and Nason, J. (2003), "Choosing the best volatility models: the model confidence set approach", *Oxford Bulletin of Economics and Statistics*, **65**, 839-861.

Hansen, P., Lunde, A. and Nason, J., (2003), *Model confidence sets for forecasting models approach*, Working Paper 2003-5, Brown University.

Patton, A. (2006), Volatility Forecast Comparison Using Imperfect Volatility Proxies, Quantitative Finance Research Centre, University of Technology, Sydney, research Papaer Series, n. 175.

Patton, A. and Sheppard, K. (2007),. Evaluating Volatility Forecasts, in *Handbook of Financial Time Series*, Andersen, T.G., Davis, R.A., Kreiss, J.P. and Mikosch, T. eds., Springer-Verlag.

Poon, S-H. and Granger, C.W.J. (2003), "Forecasting volatility in financial markets: a review", *Journal of Economic Literature,* **41**, 478-539.

Poon, S-H. and Granger, C.W.J. (2005), "Practical Issues in forecasting volatility", *Financial Analysts Journal*, **61**, 45-56.

Riskmetrics. (1996), Technical Document, New York, JP Morgan/Reuters.

West, K.D. (1996), "Asymptotic inference about predictive ability", *Econometrica*, **64**, 1067-1084.

West, K.D., Edison, H.J. and Cho, D. (1993), "A utility-based comparison of some models of exchange rate volatility ", *Journal of International Economics*, **35**, 23-45.

White, H. (2000), "A reality check for data snooping", *Econometrica*, **68**, 1097-1125.

List of NCER Working Papers

No. 48   (Download full text)
Egon Franck, Erwin Verbeek and Stephan Nüesch
Inter-market Arbitrage in Sports Betting

No. 47   (Download full text)
Raul Caruso
Relational Good at Work! Crime and Sport Participation in Italy. Evidence from Panel Data Regional Analysis over the Period 1997-2003.

No. 46   (Download full text) (Accepted)
Peter Dawson and Stephen Dobson
The Influence of Social Pressure and Nationality on Individual Decisions: Evidence from the Behaviour of Referees

No. 45   (Download full text)
Ralf Becker, Adam Clements and Christopher Coleman-Fenn
Forecast performance of implied volatility and the impact of the volatility risk premium

No. 44   (Download full text)
Adam Clements and Annastiina Silvennoinen
On the economic benefit of utility based estimation of a volatility model

No. 43   (Download full text)
Adam Clements and Ralf Becker
A nonparametric approach to forecasting realized volatility

No. 42   (Download full text)
Uwe Dulleck, Rudolf Kerschbamer and Matthias Sutter
The Economics of Credence Goods: On the Role of Liability, Verifiability, Reputation and Competition

No. 41   (Download full text)
Adam Clements, Mark Doolan, Stan Hurn and Ralf Becker
On the efficacy of techniques for evaluating multivariate volatility forecasts

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
Early Child Development and Maternal Labor Force Participation: Using Handedness as an Instrument

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
Estimating Stochastic Volatility Models Using a Discrete Non-linear Filter.

No. 2   (Download full text)
Stan Hurn, J.Jeisman and K.A. Lindsay
Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations.

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