Combining Multivariate Volatility Forecasts using Weighted Losses

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

December 11, 2018

Abstract

The ability to improve out-of-sample forecasting performance by combining forecasts is well established in the literature. This paper advances this literature in the area of multivariate volatility forecasts by developing two combination weighting schemes that are capable of placing varying emphasis on losses within the combination estimation period. A comprehensive empirical analysis of the out-of-sample forecast performance across varying dimensions, loss functions, sub-samples and forecast horizons show that new approaches significantly outperform their counterparts in terms of statistical accuracy. Within the financial applications considered, significant benefits from combination forecasts relative to the individual candidate models are observed. Although the more sophisticated combination approaches consistently rank higher relative to the equally weighted approach, their performance is statistically indistinguishable given the relatively low power of these loss functions. Finally, within the applications, further analysis highlights how combination forecasts dramatically reduce the variability in the parameter of interest, namely the portfolio weight or beta.

Keywords
Multivariate volatility, combination forecasts, forecast evaluation, model confidence set

JEL Classification Numbers
C22, G00.

Corresponding author
Mark Doolan
School of Economics and Finance
Queensland University of Technology
Brisbane, 4001
Qld, Australia
email  m.doolan@qut.edu.au

Acknowledgments The authors would like to thank the participants at the 2018 NCER Frontiers in Econometrics Workshop for their valuable feedback on the preliminary results of this paper.
1 Introduction

Combination forecasts have a long record of success in the forecasting literature. The seminal study of Bates and Granger (1969) demonstrated how even simple combination forecasts could produce lower mean squared forecast errors than the set of candidate models on which they were based. The intuition being that the combination forecast reduces forecast errors as the errors from the candidate models are not perfectly correlated. Essentially, this is the same intuition that underpins the diversification benefit that permeates modern finance theory. Subsequent research has sought to develop more sophisticated techniques for generating optimal combination forecasts, see Clemen (1989) and Timmerman (2006) for a detailed summary of this literature. Interestingly, the work on combination forecasts gave rise to the ‘Combination Puzzle’ where many of the approaches developed could not outperform a simple equally weighted average of forecasts. Smith and Wallis (2009) investigated this puzzle and highlighted how ex-ante forecasting benefits from restrictions on the estimated weights, such as equal weighting, as there is no gain to estimating optimal combination weights when the variance of forecasts errors are similar.

This paper contributes to the literature by proposing two new approaches for estimating forecast combination weights. Specifically, this paper develops a ‘time’ and a ‘state’ dependent combination approach that are capable of applying a flexible weighting emphasising different losses within the estimation window. The primary difference in these approaches is that the time dependent approach applies a strictly declining weighting scheme while the state dependent approach places more weight on losses that occurred in periods where the state variable is most similar to its current level. When applied to multivariate volatility forecasts, results show that a simple version of the time dependent approach can significantly improve forecast accuracy over candidate models and existing combination techniques.

Obviously, there are many settings in which combination forecasts can be applied. However, volatility forecasting provides a natural setting given the sheer abundance of volatility forecasting models developed to capture the characteristics of return volatility such as persistence, asymmetries and others. As an example of the number of models and their specifications that can be considered, Hansen and Lunde (2005) compare the forecast performance of 330 univariate ARCH-type models. Moreover, the forecasting problem can relate to different dimensions, time horizons and applications. Examples of the later include portfolio optimisation, asset pricing and risk management.

An extensive literature has developed in support of combination forecasts in the univariate volatility setting. For example, Becker and Clements (2008) examined combinations of univari-
ate volatility forecasting models and found that combinations of univariate volatility forecasting models significantly improve forecast accuracy. Patton and Sheppard (2009a) also found that combinations of realised volatility measures improve the accuracy of the volatility proxy. Interestingly, the Heterogeneous Autoregressive (HAR) model of Corsi (2009), which has become the benchmark realised volatility forecasting approach, is effectively a combination of volatility forecasts at the daily, weekly and monthly frequencies.

While there exists an extensive literature on the combination of univariate volatility forecasts, the research in the multivariate setting is less developed. However, those studies undertaken do provide guidance on how to combine multivariate volatility forecasts and examine their performance. For example, Pesaran, Schleicher and Zaffaroni (2009) use the Akaike Information Criteria (AIC) or Schwartz Bayesian Criteria (SBC) to create combination forecasts. Their approaches include a ‘thick’ model averaging that equally weights the best performing models while trimming all others. They also consider weightings based on the AIC and SBC to produce an ‘approximate’ Bayesian model averaging approach. Evaluated within a value-at-risk (VaR) framework, results are supportive of the ‘thick’ model averaging. However, the proposed approaches fail to consider forecast error correlations, are prone to overweight candidate models that produce almost identical forecasts and provide limited guidance on how many models to include when forming combination forecasts.

Amendola and Storti (2009) considered the performance of multivariate volatility combination forecasts in a portfolio optimisation problem. To evaluate, based on the Sharpe ratio and certain equivalence measures they found support for their proposed combination approach. A recent study by Calderi, Moura, Nogales and Santos (2017) also used the Sharpe ratio to evaluate economic performance and found that a combination weighting scheme based on the Sharpe ratio performed best. Despite the findings of these two studies, it is evident that their conclusions are drawn from sample specific results that do not necessarily relate to the quality of the forecast. For example, while Calderia et al. (2017) highlight their preferred approach generated a significantly higher Sharpe ratio of 0.14 compared to 0.06 of the benchmark, their result is driven by an excessive ex-post return of 36% p.a. relative to the target portfolio return of 12%. Moreover, given that their preferred approach generated a significantly higher variance and it has been established in the literature that variance is a consistent ranking metric for portfolio variance, it would appear that the conclusions drawn are sample specific.

---

1Corsi (2009) refers to the model as a ‘cascade’ model given the differing time horizons of the forecasts being combined.

2Bollerslev, Patton and Quaedvlieg (2016, p.2) refer to HAR as the “preferred specification for realised volatility forecasting.”

3In Appendix A, it is shown that errors in the asset weighting estimated from an incorrect forecast can lead to excessive ex-post returns that result in the incorrect forecast outranking the correct forecast.
Recently, Amendola and Storti (2015) directly applied loss functions for estimating the vector of combination weights. With 50 assets, seven candidate models and a range of statistical loss functions, their results did not generally support the use of combination forecasts. This is most evident in their first sample period, three years in the early 2000s, where the candidate models significantly outperformed all combination forecasts for three loss functions and no combination approach was found to outperform all candidate models across all six loss functions. Moreover, when the combinations did perform well, there was no consistency across the loss functions. Some evidence supporting the combinations is presented for the later samples, but no one approach consistently performs among the best forecasts and there seems to be little gain over some of the candidate models considered. Finally, although based on asset returns, there was no consideration given to potential financial applications of the forecasts.

Given the inconsistent support for multivariate volatility combination forecasts, limitations in earlier studies and the direct application of the forecasts within the context of financial applications, the multivariate volatility forecasting provides a valuable setting to apply the proposed combination approaches. In conducting a comprehensive empirical study across loss function, dimension and time horizons, results show that a simple time dependent approach produces a statistically superior forecasts across all horizons, although it is noted that the performance of the state dependent approach improves with forecast horizon. Within the financial applications of portfolio optimisation and the market model, the combination forecasts significantly outperform all the candidates models. While the more sophisticated combination approaches consistently rank higher relative to the equally weighted forecast, their performance is statistically indistinguishable given the relatively low power of these loss functions. Further analysis of the estimate of interest within the applications, the estimated portfolio weights and betas, shows that the estimates are much less variable than those based on the candidate models. Ultimately, these results highlight how the approaches proposed can significantly improve multivariate volatility forecast performance in terms of statistical accuracy. Although these approaches do not necessarily significantly outperform existing approaches within the given applications, they are consistently among the best performing forecasts.

The remainder of this paper is organised as follows. Section 2 outlines the generation of combination forecast while Section 3 outlines the multivariate loss functions. Sections 4 and 5 outline the data and candidate models used in the empirical study. Results for the various loss functions and forecast horizon are reported in Section 6. Finally, Section 7 concludes.
2 Combining Multivariate Volatility Forecasts

The combined, h-step ahead multivariate volatility forecast, $H_{c,T}^h$, is formed as a weighted average of $J$ well-defined multivariate volatility forecasts, $H_{j,T}^h$, from a set of candidate models:

$$H_{c,T}^h = \sum_{j=1}^{J} w_{j,T} H_{j,T}^h,$$

where $h$ is the forecast horizon and forecasts are integrated over the horizon, $w_{j,T}$ are strictly non-negative scalar weights, and all candidate forecasts and weights are estimated from the information set at time $T - 1$. A further restriction on $w_{j,T}$ is that the combination weights sum to one, $\sum_{j=1}^{J} w_{j,T} = 1$. Restricting the weights to sum to one ensures an unbiased combination forecasts if all underlying forecasts are unbiased. However, when the underlying forecasts are biased, such a restriction may not be warranted as the combination weights provide a mechanism to overcome the bias.

Amendola and Storti (2015) directly estimate the combination weights to minimise a loss function. Generally, in order to estimate the weighting vector for the combination forecast, the minimisation problem with constraints can be specified as:

$$\min_{w_T} L = \frac{1}{T} \sum_{t=1}^{T} f \left( \sum_{j=1}^{J} w_{j,T} H_{j,t}^h, \Sigma_t^h \right)$$

s.t. $w_{j,T} \geq 0 \ \forall k$

$$\sum_{j=1}^{J} w_{j,T} = 1,$$

where $\Sigma_t^h$ is the volatility proxy over the period from $t + 1$ to $t + h$. An obvious advantage of this approach is that the weights can be estimated directly within an objective function that matches the actual use of the forecasts, i.e. an economic loss function. A further advantage is the realisation of observations allows for re-estimation of the weighting vector when using either a recursive or rolling window estimation scheme.

This study proposes two alternative estimation schemes that possess the flexibility to place greater emphasis on different losses within the estimation window. The primary idea being that volatility is persistent and that either the losses from the recent past, or from periods where volatility was similar to the present provide more information on ability of the candidate forecasts within the current state. For example, if Model A has performed well recently and given that volatility is persistent, this good performance is likely to continue. Therefore, in estimating combination weights, forecast performance could improve by emphasising more recent
The first approach proposed is a ‘time’ dependent approach where past losses are weighted using an exponentially declining weighting scheme. Specifically:

$$\min_{w_T} L = \sum_{t=1}^{T} \alpha \exp\left(-\alpha(T-t)\right) f \left( \sum_{j=1}^{J} w_{j,T} H_{j,t}^h, \Sigma_t^h \right)$$  \hspace{1cm} (2)$$

where $\alpha$ is the parameter that governs the weighting scheme and the optimisation employs the same repeating constraints. As highlighted by the shape of weighting plots in Figure 1, the choice of $\alpha$ is crucial. Therefore, to examine the performance from slow to fast declining weights, this study considers fixed values of $\alpha$ of 0.001, 0.01 and 0.05. In addition, it allows for $\alpha$ to be estimated but bounded between 0.0001 and 0.1 such that the range of declining schemes can vary from almost constant weights to only weighting the most recent observations.

The second proposed approach is a state dependent approach where the losses that occurred in similar states to the present are weighted more heavily than those from periods where the
level of the selected state variable was different. To achieve this a kernel weighting scheme is employed:

$$
\min_{w_T} L = \sum_{t=1}^{T} K \left( \frac{S_T - S_t}{b} \right) f \left( \sum_{j=1}^{J} w_{j,T} H_{j,t}^h, \Sigma_t^h \right)
$$

(3)

where $K$ is the kernel density function, $b$ is the kernel’s bandwidth parameter, $S_t$ is the state variable and the optimisation is constrained. With such a general specification, multiple combinations of state variables, kernel density functions and bandwidths exist. Three values, half (under-smoothing), one and two (over-smoothing) times the baseline of Silverman’s rule-of-thumb value of $b = 1.06 \sigma_{S_t} T^{-\frac{1}{5}}$ are used, where $\sigma_{S_t}$ is the standard deviation of the state variable. For the state variable, the CBOE VIX (Chicago Board Option Exchange, Volatility Index) is employed as it represents the market’s expectation of the market volatility over the next 22 days. However, as potential extreme movements in the VIX may result in the kernel placing almost all weight on one observation, the logarithm of the VIX is used to reduce skewness.

In applying this kernel weighting approach with VIX as the state variable, it is noted that it will naturally tend to place greater weight on more recent observations due to the persistence in volatility. However, this does not mean that the approach is equivalent to a short rolling window or a declining weighting scheme. Instead, the kernel based weighting scheme is much more flexible in that it may represent a short rolling average (a recent jump in the level of volatility), a long rolling average (highly persistent period), a declining weighting scheme (recent gradual increase or decrease), or it may be a U-shaped relation (high-low-high periods within the sample).

To act as benchmarks, two additional approaches are considered. First, following from the combination puzzle, the equally weighted combination, where $w_{j,T} = 1/J, \forall j$, is employed. Second, as a somewhat objective approach, the Sharpe ratio (SR) approach of Calderia et al. (2017) is also considered. Specifically, the combination weights for the candidate models are calculated as:

$$
w_{j,T} = \frac{\left( \frac{\mu_{j,T}}{ \sigma_{j,T}} \right)^\eta}{\sum_{j=1}^{J} \left( \frac{\mu_{j,T}}{ \sigma_{j,T}} \right)^\eta},
$$

(4)

where $\mu_{j,T}$ and $\sigma_{j,T}$ are the portfolio ex-post return and standard deviation from the $j^{th}$ candidate model for the sample period up to $T$ and $\eta$ is a variable to dampen the effect of outliers and normalisation ensures weights sum to one. While Calderia et al. (2017) use a variety of values for $\eta$ and set all negative Sharpe ratios to zero to avoid negative weights, this study simply considers $\eta = 1$ and uses the absolute of the Sharpe ratio to avoid non-defined weights that
would occur if all Sharpe ratios were negative. Finally, in applying the volatility forecasts to the estimation of portfolio weight, only the global minimum-variance portfolio (GMVP) is considered as it avoids conflating issues that arise with the need to estimate the vector of expected returns.

3 Multivariate Volatility Loss Functions

Two statistical loss functions used widely in the evaluation of volatility forecasts can be used in the estimation of combination weights, they are the mean squared error (MSE) and quasi-likelihood (QLK). Specifically:

\[ L_{MSE,t}^h = \mathbf{t}' \left( \mathbf{H}_{c,t}^h - \mathbf{\Sigma}_t^h \right) \odot \left( \mathbf{H}_{c,t}^h - \mathbf{\Sigma}_t^h \right) \mathbf{t} \]

(5)

where \( \mathbf{\Sigma}_t^h \) is the volatility proxy such as squared returns or realised variance over the horizon and \( \mathbf{t} \) is a vector of ones. Alternatively:

\[ L_{QLK,t}^h = \ln |\mathbf{H}_{c,t}^h| - \mathbf{r}_t^h \mathbf{H}_{c,t}^{-1} \mathbf{r}_t^h, \]

(6)

where \( \mathbf{r}_t^h \) is a vector of returns over the horizon. If realised volatility was the proxy, an alternative functional form is easily implemented.

While the statistical loss functions measure accuracy in terms of distance, they may not necessarily relate to the application of the forecast. Therefore, alternative measures of forecast quality may be based on the intended application of the forecast, for example portfolio optimisation. The quality of the forecast is then assessed on the ex-post performance of the portfolio formed from the forecast. It is well known that the optimal asset weighting vector for GMVP of risky assets is

\[ \mathbf{x}_t = \left( \mathbf{H}_t^{-1} \right) \left( \mathbf{t}' \mathbf{H}_t^{-1} \right)^{-1}. \]

(8)

The quality of this weighting vector can then be assessed on its objective of minimising variance,

\[ L_{GMVP,t}^h = \mathbf{x}_t^h \mathbf{r}_t^h \mathbf{x}_t. \]

(9)

\(^4\)In the empirical study for \( N = 5 \) at the one-step ahead forecast horizon, there are 202 instances where all three reported Sharpe ratios are negative, which equates to approximately 5% of days in the out-of-sample period.

\(^5\)Alternatively, when the object is to minimise variance with a target return, \( \mu_0 \), and budget constraint, and there is a risk-free asset, the optimal weighting vector of assets is simply:

\[ \mathbf{x}_t = \frac{\mathbf{H}_t^{-1} \mathbf{\mu}}{\mathbf{\mu}' \mathbf{H}_t^{-1} \mathbf{\mu}} \mu_0, \]

(7)

where \( \mathbf{\mu} \) is the vector of expected returns. The obvious drawback of this approach is that it requires additional information on the vector of expected returns.
Another application of the variance-covariance matrix is in asset pricing. Although the conditional version of the Capital Asset Pricing Model (CAPM) provides the natural setting, this study uses the conditional Market Model (MM) as, for consistency with the other loss functions, it is the variance-covariance matrix of returns rather than excess returns that is evaluated. The MM is specified as:

\[ r_t^h = \beta_t x' r_t^h + e_t^h, \]  

where \( x \) does not have a time subscript as the vector of market weights is simply assumed to be \( \frac{1}{N} \), \( \beta_t = \frac{H_t x}{x' H_t x} \) and \( e_t^h \) is the vector of pricing errors. The performance of the volatility forecasts are then assessed in terms of minimising the sum of squared pricing errors from the MM,

\[ L_{MM,t}^h = e_t^h e_t^h. \]  

Following the measurement of statistical or economic loss, the relative forecast performance is easily evaluated by ranking. While ranking is simple, it does come with two caveats when applied to volatility forecasts. First, not all loss functions are consistent in that the most accurate forecast is the best forecast. For example, within the statistical loss functions, noise in the volatility proxy can lead to inconsistent rankings. While many commonly used loss functions are not robust to this noise, Patton and Sheppard (2009b) have shown that MSE and QLK are robust. Of the economic loss functions, portfolio variance has been shown to be a consistent loss function, refer Patton and Sheppard (2009b). Appendix A considers the consistency sum of squared pricing errors of the MM. Assuming that the correct variance-covariance model exists in each period, it can be shown that pricing error variance provides a consistent loss function.\(^6\)

The second caveat is that ranking does not indicate whether one forecast is significantly better than another. Therefore, to test relative forecast performance, tests of equal predictive ability (EPA) can be used. The model confidence set (MCS) procedure of Hansen, Lunde and Nason (2011) provides a testing framework ideally suited to this problem given its ability to be applied to a large number of competing models and that it does not require the specification of a benchmark approach. In summary, the procedure starts with the full set of models \( M_0 \) and trims iteratively the worst performing model, \( M_a \), when EPA is rejected at a significance level \( \alpha \). At the first instance when EPA is not rejected, remaining set of models, \( M_0^* \) are deemed to have EPA for the given level of significance. Refer Hansen, Lunde and Nason (2003, 2011) for more details.

\[^6\] Appendix A also considers the Sharpe ratio (SR):

\[ L_{SR,t}^h = \frac{x_t' r_t^h}{\sqrt{x_t' r_t^h x_t}} \]  

The major advantage of the SR over portfolio variance is that it considers both returns and risk. However, it is shown that the SR is not a consistent loss function for evaluating multivariate volatility forecasts as spurious ex-post returns can lead to inaccurate forecasts outranking their more accurate counterparts.
additional detail on the procedure.

4 Data

To analyse the empirical performance of the various approaches outlined for generating a combination forecast of multivariate volatility over a variety of dimensions, daily stock returns and VIX data is collected for the period 2 January 1990 to 29 June 2018. In total, 7,181 daily observations for each portfolio and the VIX are recorded.

The stock return data used are the 5, 12 and 30 industry portfolio returns collected from Ken French’s website. In addition, by selecting the industry portfolios, a balanced panel of assets for each dimension is assured, although it should be noted that the constituents within the portfolios can change as stocks are listed or de-listed, for example. However, relative to other portfolios such as book-to-market or momentum, the constituents within the industry portfolios offer a great degree of stability. The VIX data is collected from the CBOE website. 

Figure 2 plots the VIX and shows how the market’s expectation of future volatility has varied over the sample. It also highlights how volatility is persistent with high volatility periods noted during the dot-com period in the early 2000s, the financial crisis that peaked in 2008 and the subsequent European debt crisis.

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html
http://www.cboe.com/
5 Candidate Forecasts

To focus on the benefits of combining forecasts, a small number of distinctly different models are chosen to provide forecasts. In the first instance, each of the models is used to generate a one-step ahead forecast but are then subsequently used to generate five-step and ten-step ahead integrated forecasts.

Starting with the one-step ahead forecast, the easiest approach to generate a dynamic volatility forecast is to apply an EQually weighted Moving Average (EQMA) that samples the $M$ most recent observations. Specifically:

$$H_t = \frac{1}{M} \sum_{m=1}^{M} r_{t-m+1}r'_{t-m+1}, \quad (13)$$

where $M = 250$ for this study.

A more advanced approach is the Exponentially Weighted Moving Average (EWMA) model, which places greater emphasis on more recent observations. The Riskmetrics (1996) specification is often applied in the literature:

$$H_t = (1 - \lambda) r_{t-1}r'_{t-1} + \lambda H_{t-1}, \quad (14)$$

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

The final model considered is the Dynamic Conditional Correlation (DCC) model of Engle (2002). The DCC approach allows for the variance equations to be estimated individually and separate from the correlation equation, which enables application to moderately large dimensions. As this study uses equities data, all conditional variances, $\hat{\sigma}_{i,t}^2$, are modelled with the asymmetric specification of Glosten, Jagannathan and Runkle (1993) Generalised Autoregressive Conditional Heteroskedasticity (GARCH), GJR-GARCH(1,1):

$$\hat{\sigma}_{i,t}^2 = \omega_i + (\alpha_i + \delta_i I_{i,t-1}) r_{i,t-1}^2 + \beta_i \hat{\sigma}_{i,t-1}^2, \quad (15)$$

where $I_{i,t-1}$ is an indicator function that takes the value one if $r_{i,t-1} < 0$ and the symmetric GARCH equation is recovered when $\delta_i = 0$. Modelling the conditional correlation matrix $R_t$ involves two steps. First, the conditional covariance matrix of the standardised returns is
specified with a mean reverting GARCHII structure,

\[ Q_t = \bar{Q} (1 - \alpha - \beta) + \alpha (z_{t-1} z'_{t-1}) + \beta Q_{t-1} \]  

(16)

where \( \alpha \) and \( \beta \) are scalars, and \( z_{i,t} = \frac{r_{i,t}}{\sqrt{\hat{\sigma}^2_{i,t}}} \). The second step recovers \( R_t \) by standardising the \( Q_t \) matrix,

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

(17)

For all the models, the first 2,000 observations are set as the in-sample period from which all starting values and parameters are estimated. As such, the first one-step ahead out-of-sample forecast is \( H_{2001} \) with all remaining out-of-sample forecasts generated by rolling the sample forward. Parameters are re-estimated with each observation realisation in a rolling window of 2,000 observations.

To generate each of the forecast combinations, the optimisation approaches outlined in Section 2 are implemented with the loss functions, \( L \) described in Section 3. This includes the equally weighted approach (EQ), the Calderi et al. (2017) Sharpe ratio approach (SR), the standard optimisation approach where a constant weight is applied to all lagged observations (\( C_L \)), the fixed declining weighting scheme (\( D_{\alpha,L} \)), the declining weighting scheme where \( \alpha \) is estimated (\( \hat{D}_{\alpha,L} \)), and the state dependent kernel based approach (\( K_{b,L} \)). All approaches that require estimation use a rolling window of 1,000 out-of-sample forecasts. As such, the underlying candidate forecasts of \( H_{2001} \) to \( H_{3000} \) provide the information required to estimate the initial vector of combination weights, \( w_{3001} \), which is used to form the first out-of-sample combination forecast, \( H_{c,3001} \). All subsequent combination forecasts are then generated by rolling the sample period forward one observation such that combination weights are re-estimated at the daily frequency. Despite the loss of 3,000 observations for estimation, the out-of-sample evaluation period still contains 4,181 observations in which to evaluate forecast performance.

Finally, to generate five-step and ten-step ahead forecasts, the candidate models are estimated from daily data as per above. The forecast is then integrated forward over the h-steps. For the EQMA and EWMA this is simply a case of scaling the one-step ahead forecast by the \( h \). For the DCC, the h-step ahead conditional variances and correlations are calculated within the specification and then integrated by summing. To ensure that data periods for the evaluation do not overlap, which is done to ensure that the application matches the longer horizon forecasts, the h-step ahead forecast is only generated at each h-days. This means that number of out-of-sample observations for combination estimation and forecast evaluations are h-times smaller than the one-step ahead forecasts. Therefore, to ensure sufficient observations to estimate
combination weights, the estimation period is extended to include the first third of the one-step ahead evaluation period. In the case of the ten-step ahead forecasts, there are 239 observations to estimate combination weights and 279 observations to evaluate forecast performance.

6 Results

This section presents out-of-sample forecast results across the dimensions of 5, 12 and 30 assets for the four loss functions \((L)\); MSE, QLK, GMVP and MM. Although different loss functions can be used with each of the approaches to estimate combinations weights, constant \((C_L)\), fixed declining \((D_{\alpha,L})\), estimated declining \((D_{\hat{\alpha},L})\) and kernel \((K_{b,L})\), which equates to 32 combination approaches in total. Here for the sake of brevity, for each loss function used for evaluation, only the results from forecast combinations formed on the basis of that loss function are reported. For example, results for QLK will only consider the combination estimation schemes \(C_{QLK}\), \(D_{\alpha,QLK}\), \(D_{\hat{\alpha},QLK}\) and \(K_{b,QLK}\). In total, each forecast evaluation will involve thirteen forecasts (the three candidate models and ten combinations) that are ranked from best performing \((Rank = 1)\) to worst performing \((Rank = 13)\). Forecasts for a given dimension and loss function are also tested for EPA. Following the literature, the MCS confidence intervals are reported at the 10% and 25% levels, refer Hansen, Lunde and Nason (2003). The initial results are for one-step ahead forecasts over the full-sample period. However, a sub-sample analysis is also performed to examine the robustness of the results over different volatility periods. Further, forecasts horizons of five-step (one week) and ten-step (two weeks) ahead are also examined.

6.1 One-step Ahead Forecasts

Table 1 presents the model ranks and MCS results for the MSE, QLK, GMVP and MM loss functions based on one-step ahead forecasts. The first results of note are those for MSE. Across the three dimensions, almost all forecast models display equal predictive ability (EPA) under MSE, with EQMA being the only model capable of rejection. This result is consistent with the literature in that the multivariate MSE is a robust loss function but it lacks the power to discriminate between competing forecasts. The ranking results indicate some variation across the dimensions. For example, the DCC model records ranks between first and ninth while the kernel approach with larger bandwidths displays similar rank variation. However, there are consistencies of note. The EQMA is always the worst performer, and the EQ, SR and kernel with smaller bandwidth, \(K_{0.5b,MSE}\), combinations also tend to perform poorly. While there is

---

12 Results based on the full evaluation of all 37 forecasting models (three candidate models, EQ, SR and 32 additional combinations) are available from the corresponding author on request. In summary, these results support the analysis of the smaller set of competing forecasts as the best combination forecasts tend to be generated from the loss function from which the forecasts are evaluated.
no clear ‘best’ performer among the combination approaches that estimate the weight using MSE, the constant weighting of past losses, \( C_{MSE} \), and those that apply a declining weight, \( D_{\alpha, MSE} \) and \( D_{\hat{\alpha}, MSE} \), tend to be among the highest ranked approaches.

The QLK results in the second column stand in stark contrast to those for MSE. Under QLK, the MCS excludes all forecasts bar the fixed declining weight combination that applies the fastest declining weight, \( D_{0.05,QLK} \). However, it is noted that its performance appears to be related to dimension, as it improves as the dimension increases. The \( D_{0.01,QLK} \) also performs well as it is reported in the MCS at \( N = 5 \) and 12 and, despite being excluded from the MCS at \( N = 30 \), it does rank second. The consistent performance of the declining weight combination approaches indicates the significant gain to forecast performance by combining forecasts, estimating combination weights and placing more emphasis on recent performance. Of those models excluded, the combination approaches always out rank the candidate models and those that estimate their combination weight outrank EQ and SR based approaches in most instances. While the dramatic decline in the DCC’s ranks highlights the sensitivity of rankings to the loss function, it is noted that the MCS based on QLK is a subset of those reported in the MCS based on MSE.

The final two loss functions are the economic loss functions, GMVP and MM, that are based on the forecasts financial application. The size of the reported MCS sits between that of QLK and MSE with the MCS varying in size from 3 to 10 models when using the 25% level of confidence. However, even if the confidence level is reduced to 10%, it is clearly evident that the original candidate models are excluded from the MCS in all but one instance. Moreover, the performance of the DCC model has declined further with it ranking worst in four instances and second worst in the remaining two instances. Of the combination approaches, the ranks do indicate differing levels of performance, with \( C_{L} \), \( D_{0.001,L} \) and \( D_{0.01,L} \) tending to rank well across the loss functions and dimensions. Interestingly, even though the weight applied by \( D_{0.001,L} \) is close to \( C_{L} \), the small adjustment in weight improves performance in five of six instances. However, this improved ranking performance noted across the combinations tends to be insignificant, as almost all combination approaches are reported in the MCS at the 10% level. As such, within the given application, there are gains to using combination forecasts. However, the gains to using more complex combination approaches within the applications tend not to be significant relative to their simpler counterparts. Finally, it is evident that the kernel based approach performs best with the larger bandwidth.

To investigate the forecast performance results further, Figure 3 plots the weights for selected combination forecasts for \( N = 12 \). As a benchmark, Panel A plots MA, EWMA and DCC weights for \( C_{QLK} \) against VIX. It is evident from the plot that all models receive a reasonable
ing weight (DCC), and combinations based on equal weighting (EQ), Sharpe ratio (SR), constant weighting of losses (C_L), fixed declining weight (D_{\alpha,L}), where \( \alpha = 0.001, 0.01 \) and 0.05 estimated, declining weight (D_{\alpha,L}) and kernel weighting (K_{b,L}) where the bandwidth \( b \) is adjusted by a factor of 0.5, 1 and 2. Ranks are reported from best performing (1) to worst performing (13). Inclusion in the MCS at the 10% and 25% confidence levels is indicated by * and **, respectively. Model parameters and combination weights are estimated daily with combination weight estimates based on evaluating loss function. Volatility proxy is daily squared returns for MSE and QLK, and MCS uses block bootstrap with block length of 22 observations.

### Table 1: Full-sample ranking and MCS results for one-step ahead forecasts based on the loss functions \((L)\) MSE, QLK, GMVP and MM loss functions for the industry portfolios of dimensions of 5, 12 and 30. Forecasts are from the three candidate models (EQMA, EWMA and DCC), and combinations based on equal weighting (EQ), Sharpe ratio (SR), constant weighting of losses \((C_L)\), fixed declining weight \((D_{\alpha,L})\), where \( \alpha = 0.001, 0.01 \) and 0.05 estimated, declining weight \((D_{\alpha,L})\) and kernel weighting \((K_{b,L})\) where the bandwidth \( b \) is adjusted by a factor of 0.5, 1 and 2. Ranks are reported from best performing (1) to worst performing (13). Inclusion in the MCS at the 10% and 25% confidence levels is indicated by * and **, respectively. Model parameters and combination weights are estimated daily with combination weight estimates based on evaluating loss function. Volatility proxy is daily squared returns for MSE and QLK, and MCS uses block bootstrap with block length of 22 observations.

<table>
<thead>
<tr>
<th>Model</th>
<th>MSE ( N = 5 )</th>
<th>MSE ( N = 12 )</th>
<th>MSE ( N = 30 )</th>
<th>QLK ( N = 5 )</th>
<th>QLK ( N = 12 )</th>
<th>QLK ( N = 30 )</th>
<th>GMVP ( N = 5 )</th>
<th>GMVP ( N = 12 )</th>
<th>GMVP ( N = 30 )</th>
<th>MM ( N = 5 )</th>
<th>MM ( N = 12 )</th>
<th>MM ( N = 30 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQMA</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>EWMA</td>
<td>10*</td>
<td>9*</td>
<td>8*</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>DCC</td>
<td>1**</td>
<td>1**</td>
<td>9**</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>EQ</td>
<td>11**</td>
<td>11**</td>
<td>10**</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>SR</td>
<td>8**</td>
<td>10**</td>
<td>12**</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>9**</td>
<td>10**</td>
<td>10**</td>
<td>6</td>
<td>9</td>
<td>10**</td>
</tr>
<tr>
<td>C_L</td>
<td>3**</td>
<td>3**</td>
<td>6**</td>
<td>6</td>
<td>9</td>
<td>8</td>
<td>1**</td>
<td>5**</td>
<td>7**</td>
<td>2**</td>
<td>4**</td>
<td>2**</td>
</tr>
<tr>
<td>D_{0.001,L}</td>
<td>2**</td>
<td>4**</td>
<td>5**</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>2**</td>
<td>3**</td>
<td>6**</td>
<td>1**</td>
<td>1**</td>
<td>1**</td>
</tr>
<tr>
<td>D_{0.01,L}</td>
<td>4**</td>
<td>5**</td>
<td>7**</td>
<td>1**</td>
<td>2**</td>
<td>2**</td>
<td>8</td>
<td>2**</td>
<td>2**</td>
<td>2**</td>
<td>4**</td>
<td>4**</td>
</tr>
<tr>
<td>D_{0.05,L}</td>
<td>6**</td>
<td>6**</td>
<td>3**</td>
<td>3**</td>
<td>1**</td>
<td>1**</td>
<td>10**</td>
<td>1**</td>
<td>1**</td>
<td>10**</td>
<td>6**</td>
<td>6**</td>
</tr>
<tr>
<td>D_{0.1,L}</td>
<td>5**</td>
<td>2**</td>
<td>4**</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>3**</td>
<td>6**</td>
<td>4**</td>
<td>4**</td>
<td>4**</td>
<td>3**</td>
</tr>
<tr>
<td>K_{0.05,L}</td>
<td>12**</td>
<td>12**</td>
<td>11**</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>4**</td>
<td>9</td>
<td>9**</td>
<td>9**</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>K_{0.1,L}</td>
<td>9**</td>
<td>8**</td>
<td>2**</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>7**</td>
<td>4**</td>
<td>8**</td>
<td>8**</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>K_{0.2,L}</td>
<td>7**</td>
<td>7**</td>
<td>1**</td>
<td>2**</td>
<td>4</td>
<td>5</td>
<td>5**</td>
<td>8**</td>
<td>5**</td>
<td>5**</td>
<td>5**</td>
<td>5**</td>
</tr>
</tbody>
</table>

Full out-of-sample period: Nov. 2001 to Jun. 2018

weight across the entire sample. For example, no model tends to weight less than 10% or more than 60%. Interestingly, only in the first part of the sample does the DCC model receive the highest weight of any model. These weights clearly indicate the value of the simple EQMA and EWMA models in improving forecast performance. There is some evidence that the EQMA weight increases after periods of relative calm, for example prior to the global financial crisis (pre-2008) and post Euro crisis (post-2012). Such a change is in-line with combination estimation using a rolling window of 1,000 observations. It is also evident that no model performed particular well during the heights of the financial crisis as noted by the convergence of weights during this period, which is consistent with Smith and Wallis (2009) observation on the existence of the equal weighting combination puzzle.

Panels B of Figure 3 plots the DCC weights under C_{QLK} against those from D_{0.01,QLK}. The panel shows that the constant weight and fixed declining weight combinations place similar weights on DCC. However, the declining weight approach is a little more variable and tends to lead. For example, in the first third of the out-of-sample period, the D_{0.01,QLK} weight is lower with the C_{QLK} tracking down towards it. When the D_{0.01,QLK} weight then moves up, the C_{QLK} weight also moves up but at a much slower rate. Obviously, the greater emphasis on more recent observations by D_{0.01,QLK} means that it reacts quicker than its constant counterpart. When compared to the VIX, this emphasis on more recent observations sees the DCC weight generally
increase as volatility increases, with the exception being the extreme increase in volatility during the financial crisis.

The final plots in Panels C and D are for the DCC weights for $D_{\alpha, QLK}$ and $K_{QLK}$. The weights from both panels tend to follow those of the $D_{\alpha, QLK}$ weight plotted in Panel B, although the weights are much more variable for these more flexible approaches. The increased variability directly follows from limited observations being weighted in the estimation. For example, the highly variability in weights for $D_{\alpha, QLK}$ tends to be associated with lower levels of volatility, which correspond to periods where the $\hat{\alpha}$ is close to 0.1. There are also periods where it is less volatile and aligns to $C_{QLK}$ weight. These periods correspond to $\hat{\alpha} = 0.0001$, which is where the declining weighting scheme is effectively flat. Finally, for $K_{b, QLK}$, it is evident that its volatile period corresponds to when volatility has moved to levels not present within the rolling window, which is where the kernel will place most weight on the most recent observations.
To further highlight the performance of DCC within the applications, Figure 4 plots the GMVP weight and MM beta from the DCC and D\textsubscript{0.01,L} for the ‘Retail’ sector portfolio in the case of N = 30\textsuperscript{13}. The GMVP weight, Panel A, shows that both approaches estimate weights with similar levels, but the D\textsubscript{0.01,GMVP} weight is substantially less volatile than that of DCC. This is confirmed by standard deviation of weight changes calculated over the full sample, with the combination approach found to have 39% less variation than the DCC model. Although it cannot be specifically concluded that less volatility in the portfolio weights directly results in improved performance, it is clear from this graph and the results presented in Table 1 that the less volatile weights have improved performance as measured by GMVP in this instance. A more pronounced results is evident for the MM betas, Panel B, where the standard deviation in the beta change is 62% smaller for the combination relative to DCC, and this is despite D\textsubscript{0.01,MM} only ranking fourth in the full-sample evaluation.

6.2 Sub-sample results

Table 2 presents the sub-sample results for the one-step ahead forecasts across the three sub-sample periods that corresponds to one-third of the out-of-sample observations, which is approximately 1,390 observations each. Although the periods are of equal length, they have also been selected to align to the low, high and moderate volatility states evident in Figure 3. Consistent with earlier results, the MCSs are generally large for MSE, small for QLK and somewhere in between for GMVP and MM. However, some variation is noted as the MCS is somewhat smaller for MSE in the final sub-sample (moderate volatility sample) and larger for QLK in the first (low volatility) and final sub-samples.

In terms of forecast performance, the combination forecasts outperform their candidate models in most instances. The exceptions are DCC ranking well in the first and second sub-samples under MSE, and the odd instance of EQMA and EWMA being reported in either the GMVP or MM MCSs. However, in terms of consistency across all the loss functions, no approach appears in all MCSs but the fixed declining weight approach, D\textsubscript{0.01,L}, offers the best performance. It appears in the MCS under all financial applications, when excluded from the QLK MCS it still ranks second across all dimensions. This compares to D\textsubscript{0.05,L}, which although reported all MCSs based on QLK, performs poorly in the financial applications where it ranks among the worst performing models and is excluded from the MCS based on MM in the first and finally sub-samples. As such, it is evident there is value in placing some weight on moderately older observations when estimating combination weights.

\textsuperscript{13}Of the 30 industry portfolios provided by Ken French, Retail is one of the 29 named portfolios, with ‘Other’ being the final portfolio.
Figure 4: Panels A and B plot the GMVP weight and MM Beta, respectively, for the ‘Retail’ portfolio in the industry $N = 30$ portfolios. The estimates are based on the DCC (black line) in both panels while $D_{0.01, L}$ combination (red line) is for $L_{GMVP}$ in Panel A and $L_{MM}$ in Panel B.
In specific reference to the applications and based on inclusion in the MCS, the combination based on the constant weight, $C_L$ and the marginally declining weight approach $D_{0.001,L}$ offer the best performance. They are never excluded from the MCS at the 25% level, which indicates they have EPA, and they rank in the top three forecasts on 25 of 36 occasions. When the ranks are directly compared, it is evident that there is a gain to the declining weight scheme in the periods of high and moderate volatility; the second and final sub-samples respectively. Finally, it is noted that the EQ approach is only excluded from one application based MCS. Although this approach generally does not tend to produce top ranked forecasts, the MCS results indicate that superior forecast performance can be obtained with a simple weighting of multivariate volatility forecasts.

### 6.3 h-step Ahead Forecasts

Table 3 present the MCS results for the five-step and ten-step forecast horizons in Panels A and B, respectively. As the analysis is conducted on non-overlapping forecasts, which limits the availability of data for estimation, the first third of the out-of-sample period used in the one-step ahead forecasts is now used for the estimation of combination weights. This leaves the final two thirds of the out-of-sample period for evaluating the longer horizon forecasts, which equates to 557 and 279 observations for the five-step and ten-step ahead forecasts, respectively.

It is evident from the Table 3 there are patterns consistent with the earlier results. For example, the MCS is smallest under QLK and largest under MSE. The EQMA model performs poorly under MSE and QLK while the DCC results decline from MSE to the other loss functions. In regards to forecast accuracy, the combination approaches with estimated weights outperform the candidate models and simpler combination approaches. Moreover, across both panels, the only model found in all MCSs is $D_{0.05,L}$. Given the time between longer horizon forecasts due to the non-overlapping sample, it is not overly surprising to find that the declining approach that places more weight on more recent observations has displayed the strongest performance. Interestingly, the relative accuracy of the state dependent approach, $K_{b,L}$, has improved with the forecast horizon. Most notably, this is seen at the ten-step ahead horizon where $K_{0.5b,L}$ is reported in all MCSs and ranks first 4 of 6 times and second on the remaining instances. Given that the state variable is the VIX, which is a 22-day ahead forecast, it is then interesting to note that the forecast performance of the state dependent approach improves as the forecast horizon align with the information within the state variable. Moreover, the performance of the narrower bandwidth that would tend to weight more recent observations is consistent with the performance of $D_{0.05,L}$.

Despite some consistencies with earlier results, it is evident from Table 3 that the forecast
### Table 2: Sub-sample ranking and MCS results for one-step ahead forecasts based on MSE, QLK, GMVP and MM loss functions for the industry portfolios of dimensions of 5, 12 and 30.

Panels A, B and C report results for the first, second and final third of sample, respectively, with each sub-sample corresponding to approximately 1,390 observations. Forecasts are from the three candidate models (EQMA, EWMA and DCC), and combinations based on equal weighting (EQ). Sharpe ratio (SR), constant weighting of losses ($C_L$), fixed declining weight ($D_{a,L}$), where $\alpha = 0.001, 0.01$ and 0.05 estimated, declining weight ($D_{a,L}$) and kernel weighting ($K_{b,L}$) where the bandwidth $b$ is adjusted by a factor of 0.5, 1 and 2. Ranks are reported from best performing (1) to worst performing (13). Inclusion in the MCS at the 10% and 25% confidence levels is indicated by * and **, respectively. Model parameters and combination weights are estimated daily with combination weight estimates based on evaluating loss function. Volatility proxy is daily squared returns for MSE and QLK, and MCS uses block bootstrap with block length of 22 observations.

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>QLK</th>
<th>GMVP</th>
<th>MM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 5</td>
<td>N = 12</td>
<td>N = 30</td>
<td>N = 5</td>
</tr>
<tr>
<td><strong>Panel A: First third of out-of-sample period; Nov. 2001 to Jun. 2007</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQMA</td>
<td>13**</td>
<td>13*</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>EWMA</td>
<td>12*</td>
<td>12*</td>
<td>12*</td>
<td>12</td>
</tr>
<tr>
<td>DCC</td>
<td>2**</td>
<td>6**</td>
<td>6**</td>
<td>11</td>
</tr>
<tr>
<td>EQ</td>
<td>1**</td>
<td>10*</td>
<td>9*</td>
<td>8*</td>
</tr>
<tr>
<td>SR</td>
<td>10*</td>
<td>11*</td>
<td>10**</td>
<td>10*</td>
</tr>
<tr>
<td>$C_L$</td>
<td>6*</td>
<td>7*</td>
<td>2*</td>
<td>5*</td>
</tr>
<tr>
<td>$D_{0.001,L}$</td>
<td>7*</td>
<td>8*</td>
<td>3**</td>
<td>4**</td>
</tr>
<tr>
<td>$D_{0.01,L}$</td>
<td>4*</td>
<td>3**</td>
<td>4**</td>
<td>2**</td>
</tr>
<tr>
<td>$D_{0.05.L}$</td>
<td>5*</td>
<td>5**</td>
<td>8*</td>
<td>6*</td>
</tr>
<tr>
<td>$D_{0.1,L}$</td>
<td>8*</td>
<td>4**</td>
<td>5**</td>
<td>9*</td>
</tr>
<tr>
<td>$K_{0.5a,L}$</td>
<td>9*</td>
<td>9*</td>
<td>11**</td>
<td>7*</td>
</tr>
<tr>
<td>$K_{b,L}$</td>
<td>3**</td>
<td>6**</td>
<td>7**</td>
<td>3**</td>
</tr>
<tr>
<td>$K_{26,L}$</td>
<td>1**</td>
<td>2**</td>
<td>1**</td>
<td>1**</td>
</tr>
</tbody>
</table>

| EQMA   | 13 | 13 | 13 | 13 | 12 | 12 | 12 | 11 | 11 | 13 | 12 | 12 |
| EWMA   | 10** | 9* | 8* | 12 | 13 | 13 | 13 | 11** | 13 | 13 | 2** | 6* | 9* |
| DCC    | 1* | 1* | 9* | 11 | 11 | 11 | 13 | 12** | 12* | 12 | 13 | 13 |
| EQ     | 11** | 11* | 10** | 10 | 7 | 7 | 7** | 8** | 5* | 11** | 11* | 8* |
| SR     | 8** | 10** | 12** | 9 | 3 | 9 | 8* | 7** | 10* | 10* | 10* | 11* |
| $C_L$  | 3** | 3** | 6** | 7 | 10 | 10 | 3** | 5** | 9* | 4** | 7** | 4* |
| $D_{0.001,L}$ | 2** | 4** | 5** | 5** | 8 | 8 | 4** | 6** | 8* | 3** | 4** | 2* |
| $D_{0.01,L}$ | 4** | 5** | 7** | 2** | 2** | 2 | 9** | 2** | 2* | 6** | 2** | 3* |
| $D_{0.05.L}$ | 6** | 7** | 3** | 1** | 1** | 1** | 10* | 1** | 1* | 9** | 1** | 6* |
| $D_{0.1,L}$ | 5** | 2** | 4** | 6 | 4 | 4 | 2** | 5** | 7* | 5** | 3* | 1** |
| $K_{0.5a,L}$ | 12** | 12** | 11** | 8 | 9 | 5 | 1** | 4** | 4* | 1* | 9** | 10 |
| $K_{b,L}$ | 9** | 8* | 1** | 4** | 6 | 3 | 6* | 3** | 3* | 8* | 8* | 7* |
| $K_{26,L}$ | 7** | 6** | 2** | 3** | 5 | 6 | 5* | 10** | 6* | 7** | 5** | 5** |

| **Panel C: Final third of out-of-sample period; Dec. 2012 to Jun. 2018** |
| EQMA   | 13 | 13 | 13 | 13 | 12 | 12 | 12 | 11 | 11 | 13 | 12 | 12 |
| EWMA   | 12* | 12 | 12 | 13 | 13 | 13 | 12 | 13 | 13 | 12 | 12 | 13 |
| DCC    | 8** | 8* | 9* | 11 | 12 | 11 | 13 | 12 | 12 | 13 | 13 | 12 |
| EQ     | 1** | 2** | 2** | 2** | 9 | 10 | 3** | 6** | 6* | 11** | 10 | 4** |
| SR     | 3** | 1** | 1** | 3** | 10 | 9 | 4** | 7** | 5* | 10** | 9 | 3** |
| $C_L$  | 6** | 7 | 5 | 6 | 8 | 1** | 2** | 4* | 3** | 3* | 2** |
| $D_{0.001,L}$ | 7* | 6 | 6 | 4** | 4 | 6 | 2* | 1** | 3* | 2* | 2* | 1** |
| $D_{0.01,L}$ | 4* | 5** | 3** | 1** | 1** | 1** | 6** | 1* | 2** | 1** | 4** | 8* |
| $D_{0.05.L}$ | 5** | 3** | 4** | 8** | 3** | 2** | 9** | 10* | 7* | 9 | 11 | 10 |
| $D_{0.1,L}$ | 2** | 4** | 5** | 6** | 2** | 3** | 5** | 3** | 1** | 7** | 6 | 5* |
| $K_{0.5a,L}$ | 9* | 10 | 11 | 10 | 8 | 7 | 11** | 11** | 10** | 8 | 8 | 9* |
| $K_{b,L}$ | 11* | 11 | 10 | 9 | 7 | 5 | 10* | 9** | 11** | 6* | 5 | 7** |
| $K_{26,L}$ | 10** | 9* | 8** | 7* | 5 | 4** | 7** | 8* | 9** | 4** | 1** | 6** |
and combination weights are estimated every h-step. Volatility proxy is daily squared returns (D

given applications can be achieved through a simple equally weighted average of the combination

where the bandwidth \( b \) is adjusted by a factor of 0.5, 1 and 2. Ranks are reported from best performing (1) to worst performing (13). Inclusion in the MCS at the 10% and 25% confidence levels is indicated by * and **, respectively. Model parameters and combination weights are estimated every h-step. Volatility proxy is daily squared returns summed over the h-steps for MSE and QLK, and MCS uses block bootstrap with block length of 4 for \( h = 5 \) and 2 for \( h = 10 \) observations.

Table 3: Ranking and MCS results for five-step (Panel A) and ten-step (Panel B) ahead forecasts based on MSE, QLK, GMVP and MM loss functions for the industry portfolios of dimensions of 5, 12 and 30. Forecasts are from the three candidate models (EQMA, EWMA and DCC), and combinations based on equal weighting (EQ), Sharpe ratio (SR), constant weighting of losses (\( C_L \)), fixed declining weight (\( D_{1.001.L} \)), where \( \alpha = 0.001 \), 0.01 and 0.05 estimated, declining weight (\( D_{b,L} \)) and kernel weighting (\( K_{b,L} \)) where \( b \) is adjusted by a factor of 0.5, 1 and 2. Ranks are reported from best performing (1) to worst performing (13). Inclusion in the MCS at the 10% and 25% confidence levels is indicated by * and **, respectively. Model parameters and combination weights are estimated every h-step. Volatility proxy is daily squared returns summed over the h-steps for MSE and QLK, and MCS uses block bootstrap with block length of 4 for \( h = 5 \) and 2 for \( h = 10 \) observations.


<table>
<thead>
<tr>
<th></th>
<th>MSE N = 5</th>
<th>MSE N = 12</th>
<th>MSE N = 30</th>
<th>QLK N = 5</th>
<th>QLK N = 12</th>
<th>QLK N = 30</th>
<th>GMVP N = 5</th>
<th>GMVP N = 12</th>
<th>GMVP N = 30</th>
<th>MM N = 5</th>
<th>MM N = 12</th>
<th>MM N = 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQMA</td>
<td>13^*</td>
<td>13</td>
<td>13^*</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>8^*</td>
<td>6^*</td>
<td>11^*</td>
<td>2^*</td>
<td>6^*</td>
<td>11^*</td>
</tr>
<tr>
<td>EWMA</td>
<td>11^*</td>
<td>10^*</td>
<td>10^*</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>12^*</td>
<td>12^*</td>
<td>13</td>
</tr>
<tr>
<td>DCC</td>
<td>1^*</td>
<td>1^*</td>
<td>9^*</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>9^*</td>
<td>10^*</td>
<td>12</td>
<td>13</td>
<td>13^*</td>
<td>12^*</td>
</tr>
<tr>
<td>EQ</td>
<td>10^*</td>
<td>11^*</td>
<td>11^*</td>
<td>2^*</td>
<td>9</td>
<td>10</td>
<td>2^*</td>
<td>1^*</td>
<td>2^*</td>
<td>2^*</td>
<td>10^*</td>
<td>4^*</td>
</tr>
<tr>
<td>SR</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>4^*</td>
<td>10</td>
<td>9</td>
<td>1^*</td>
<td>2^*</td>
<td>7^*</td>
<td>6^*</td>
<td>2^*</td>
<td>1^*</td>
</tr>
<tr>
<td>( C_L )</td>
<td>8^*</td>
<td>3^*</td>
<td>3^*</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>3^*</td>
<td>7^*</td>
<td>9^*</td>
<td>8</td>
<td>10^*</td>
<td>3^*</td>
</tr>
<tr>
<td>( D_{0.001.L} )</td>
<td>6^*</td>
<td>4^*</td>
<td>4^*</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>4^*</td>
<td>8</td>
<td>8^*</td>
<td>9^*</td>
<td>9</td>
<td>4^*</td>
</tr>
<tr>
<td>( D_{0.01.L} )</td>
<td>3^*</td>
<td>7^*</td>
<td>7^*</td>
<td>1^*</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>12</td>
<td>3^*</td>
<td>3^*</td>
<td>8^*</td>
<td>8^*</td>
</tr>
<tr>
<td>( D_{0.05.L} )</td>
<td>4^*</td>
<td>6^*</td>
<td>6^*</td>
<td>5^*</td>
<td>5^*</td>
<td>5^*</td>
<td>5^*</td>
<td>5^*</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>5^*</td>
</tr>
<tr>
<td>( D_{b,L} )</td>
<td>7^*</td>
<td>12</td>
<td>12</td>
<td>6^*</td>
<td>4</td>
<td>2^*</td>
<td>6</td>
<td>12</td>
<td>11</td>
<td>6^*</td>
<td>5^*</td>
<td>4^*</td>
</tr>
<tr>
<td>( K_{b,L} )</td>
<td>5^*</td>
<td>7^*</td>
<td>6^*</td>
<td>10</td>
<td>5</td>
<td>3^*</td>
<td>10</td>
<td>3^*</td>
<td>1</td>
<td>1^*</td>
<td>1</td>
<td>9^*</td>
</tr>
<tr>
<td>( K_{2b,L} )</td>
<td>2^*</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>4^*</td>
<td>6^*</td>
<td>5^*</td>
<td>4^*</td>
<td>4^*</td>
<td>7^*</td>
<td>7^*</td>
</tr>
</tbody>
</table>

Panel B: Ten-step ahead forecasts; Jun. 2007 to Jun. 2018

<table>
<thead>
<tr>
<th></th>
<th>MSE N = 5</th>
<th>MSE N = 12</th>
<th>MSE N = 30</th>
<th>QLK N = 5</th>
<th>QLK N = 12</th>
<th>QLK N = 30</th>
<th>GMVP N = 5</th>
<th>GMVP N = 12</th>
<th>GMVP N = 30</th>
<th>MM N = 5</th>
<th>MM N = 12</th>
<th>MM N = 30</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQMA</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>EWMA</td>
<td>4^*</td>
<td>4^*</td>
<td>8</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>7</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>DCC</td>
<td>6^*</td>
<td>11^*</td>
<td>10^*</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>5^*</td>
<td>4^*</td>
<td>7^*</td>
<td>7^*</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>EQ</td>
<td>5^*</td>
<td>10^*</td>
<td>11^*</td>
<td>5^*</td>
<td>9</td>
<td>9</td>
<td>2^*</td>
<td>3^*</td>
<td>1^*</td>
<td>10*</td>
<td>6</td>
<td>2^*</td>
</tr>
<tr>
<td>SR</td>
<td>3^*</td>
<td>12</td>
<td>12</td>
<td>3^*</td>
<td>10</td>
<td>10</td>
<td>1^*</td>
<td>2^*</td>
<td>3^*</td>
<td>11</td>
<td>8</td>
<td>7^*</td>
</tr>
<tr>
<td>( C_L )</td>
<td>12^*</td>
<td>9^*</td>
<td>9^*</td>
<td>9^*</td>
<td>8^*</td>
<td>7^*</td>
<td>3^*</td>
<td>5^*</td>
<td>6^*</td>
<td>5^*</td>
<td>5^*</td>
<td>5^*</td>
</tr>
<tr>
<td>( D_{0.001.L} )</td>
<td>10^*</td>
<td>8^*</td>
<td>7^*</td>
<td>8^*</td>
<td>7^*</td>
<td>8^*</td>
<td>4^*</td>
<td>7^*</td>
<td>5^*</td>
<td>6^*</td>
<td>7</td>
<td>6^*</td>
</tr>
<tr>
<td>( D_{0.01.L} )</td>
<td>8^*</td>
<td>4^*</td>
<td>5^*</td>
<td>6^*</td>
<td>6^*</td>
<td>6^*</td>
<td>6^*</td>
<td>4^*</td>
<td>8^*</td>
<td>9^*</td>
<td>4^*</td>
<td></td>
</tr>
<tr>
<td>( D_{0.05.L} )</td>
<td>11^*</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>1^*</td>
<td>10^*</td>
<td>5^*</td>
<td>2^*</td>
<td>9</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>( D_{b,L} )</td>
<td>9^*</td>
<td>6^*</td>
<td>6^*</td>
<td>10</td>
<td>3</td>
<td>5^*</td>
<td>7</td>
<td>4</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>3^*</td>
</tr>
<tr>
<td>( K_{b,L} )</td>
<td>5^*</td>
<td>4^*</td>
<td>4^*</td>
<td>4^*</td>
<td>4^*</td>
<td>4</td>
<td>4</td>
<td>4^*</td>
<td>4^*</td>
<td>9^*</td>
<td>10</td>
<td>10^*</td>
</tr>
<tr>
<td>( K_{2b,L} )</td>
<td>2^*</td>
<td>2^*</td>
<td>2^*</td>
<td>1^*</td>
<td>4^*</td>
<td>3^*</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>2^*</td>
<td>3^*</td>
<td>8^*</td>
</tr>
</tbody>
</table>

performance is more variable. This in part may reflect difficulties with differentiating between models with a smaller out-of-sample period and longer forecast horizons where the relative performance has declined. This decline is most evident for the financial applications where the simple equally weighted approach, EQ, and Sharpe ratio, SR, approaches generally record better ranks than their more complex counterparts and are included in all the MCS based on the financial application. Therefore, while these approaches are not necessarily the most accurate given the MSE and QLK results, they show that the gains to combination forecasting within the given applications can be achieved through a simple equally weighted average of the combination of the candidate models. As such, these results support the Smith and Wallis (2009) conjecture that equally weighted approach performs well when the forecast error variances are similar.
7 Conclusion

This study demonstrates that combination forecasts of multivariate volatility outperform the candidate forecasts on which they formed across both statistical and economic loss measures. This result is consistent across dimension, sub-samples and forecast horizon. Of the combination approaches examined here, those that place greater emphasis on more recent observations and estimate weights based on the evaluating loss function display the best performance in general. However, within the given applications, it is noted that a simple equally weighted combination tends to perform reasonably well. Finally, while the state dependent approach was not necessarily among the best performers in the case of the one-step ahead forecasts, its performance improved as the forecast horizon came closer to the 22-day ahead forecast horizon of the VIX, which was used as the state variable. Additional directions for future research include examining different state variables within the combination approach, applications to realised volatility and using the combination forecasts for the testing of conditional asset pricing models.
8 Appendix A

Following the approach of Patton and Sheppard (2009), it can be shown that the incorrect forecast cannot generate a smaller variance of the pricing errors of the market model (MM). Specifically, if \( e_t = r_t - \beta_t w_t^r r_t = 0 \) when the correct \( \beta_t \) is used, then \( e_t' e_t = 0 \). Therefore:

\[
e_t' e_t \leq \hat{e}_t' \hat{e}_t
\]

\[
(r_t - \beta_t w_t^r r_t)'(r_t - \beta_t w_t^r r_t) \leq (r_t - (\beta_t + c_t) w_t^r r_t)'(r_t - (\beta_t + c_t) w_t^r r_t)
\]

\[
-2w_t^r r_t' \beta_t r_t + \beta_t w_t^r r_t' w_t r_t \leq -2w_t^r r_t (\beta_t + c_t)' r_t + (\beta_t + c_t)' w_t^r r_t' w_t (\beta_t + c_t)
\]

\[
0 \leq -2w_t^r r_t c_t' r_t + c_t' w_t^r r_t' w_t c_t
\]

Given that \( c_t' w_t^r r_t' w_t c_t \geq 0 \), it must be that \( \hat{e}_t' \hat{e}_t - e_t' e_t \geq 0 \).

For the Sharpe ratio (SR), it can be shown that an incorrect forecast can be superior. Specifically, where \( x_t \) is portfolio weight of the minimum variance portfolio based on the correct volatility forecast and \( c_t \) is the error in weights from the incorrect volatility forecast, then:

\[
SR_t - \hat{SR} = \frac{x_t r_t}{\sqrt{x_t' \Sigma x_t}} - \frac{(x_t + c_t)' r_t}{\sqrt{(x_t + c_t)' \Sigma (x_t + c_t)}}
\]

\[
= \frac{x_t r_t}{\sqrt{x_t' \Sigma x_t}} - \frac{(x_t + c_t)' r_t}{\sqrt{x_t' \Sigma x_t} + 2x_t' \Sigma c_t + c_t' \Sigma c_t}
\]

\[
= \frac{x_t r_t}{\sqrt{x_t' \Sigma x_t}} - \frac{(x_t + c_t)' r_t}{\sqrt{x_t' \Sigma x_t} + c_t' \Sigma c_t}
\]

\[
= \frac{x_t r_t}{A} - \frac{(x_t + c_t)' r_t}{A(1 + B)}
\]

\[
= \frac{x_t r_t}{A} - \frac{x_t r_t}{A(1 + B)} - \frac{c_t r_t}{A(1 + B)}
\]

\[
= \frac{x_t r_t}{A} \left[ 1 - \frac{1}{(1 + B)} \right] - \frac{c_t r_t}{A(1 + B)}
\]

\[
\leq 0,
\]

where \( B \geq 0 \) as \( c_t' \Sigma c_t \geq 0 \), which implies \( \frac{x_t r_t}{A} \left[ 1 - \frac{1}{(1 + B)} \right] > 0 \) only when \( x_t r_t > 0 \). Even then, it may be that \( \frac{c_t r_t}{A(1 + B)} > 0 \) such that the incorrect forecast generates a higher Sharpe ratio at a point in time.
References


