

Application of Forecast Combination in Volatility Modelling

Adam James^a and Felix Chan^a

^aSchool of Economics and Finance, Curtin University, GPO Box U1987, Perth, Western Australia
Email: F.Chan@curtin.edu.au

Abstract: It is well known in the forecasting literature that combining forecasts from different models can often lead to superior forecast performance, at least in the Mean Squares Error (MSE) sense. It has also been noted that combining forecasts by simple averaging often performs better than more sophisticated weighting schemes, although simple averages tend to ignore correlations between forecast errors. However, it is unclear whether these stylized facts hold under different forecast criteria. This is particularly important when evaluating the performance of Value-at-Risk forecasts where MSE is not an appropriate measure. In practice the VaR performance is measured against the Back-testing procedure as outlined in Basel Accord. Given the role of VaR in risk management, it would be important to investigate if forecast combination provided any benefit in forecasting VaR. An interesting implication of this study is that, if forecast combination does in fact provide superior VaR forecasts over individual models, then it also provides a convenience way to combine qualitative forecasts (from expert opinion) and quantitative forecasts (from quantitative models). The combination of qualitative and quantitative forecasts may in fact, enhance the forecast accuracy of VaR further.

The aim of this paper is to provide an empirical evaluation of forecast combination for Value-at-Risk. Value-at-Risk forecasts based on four different volatility models, namely, EGARCH, IGARCH, Stochastic Volatility, will be constructed and combined. The forecast performance of the combined forecasts will be compared to the forecast performances of each of the individual models. Two weighting schemes are being considered in this paper, namely, simple weighted average and Quantile Regression (QR). The empirical performances of these forecasts will be based on the percentages of violation as proposed in the Basel Accord with two sets of daily data, namely FTSE and S&P 500, for the period 3rd January 1996 to 3rd August 2010.

The results show that, overall, (i) forecast combination performed better than individual models and (ii) simple weighted average performed better than QR. These results are consistent with the stylized findings in the forecast combination literature. Thus, the paper provided empirical evidence supporting the use of forecast combination in forecasting VaR thresholds.

Keywords: Forecast combination, Value-at-Risk, Quantile regression.

1 INTRODUCTION

Modelling and forecasting the distributions of assets' returns have become increasingly important over the last two decades. Specifically, the ability to accurately capture the negative tail of the return distributions will allow more accurate forecast of risk associated with the assets. Although serious efforts have been devoted in this area, the accuracy in forecasting the negative tail of the return distribution is far from satisfactory, as shown during the Global Financial Crisis. Specifically, while there have been numerous models proposed in the literature to forecast Value-at-Risk and Conditional Value-at-Risk (expected shortfalls), these models often suffer from numerical difficulties due to the curse of dimensionality or they do not perform satisfactorily consistently. Moreover, it is more difficult, if not impossible, to incorporate qualitative information into these models.

A possible solution to improve forecast accuracy is to incorporate a methodology which has shown to be exceedingly successful in a diverse range of forecasting applications. This methodology involves the aggregation of a number of independent forecasts into a forecast combination. Since the seminal works on the subject by Bates and Granger (1969) there has been an extensive body of literature developed on forecast pooling. It is the general consensus of this literature that, given the combination of forecasts from sensible models, the combination of forecasts outperforms individual forecasts. The primary focus of this paper is to investigate the potential benefit of applying forecast combination in volatility forecasting. The results from this paper will contribute to the literature in two ways. Firstly, the success of forecast combination lacks theoretical justification. There are very limited theoretical results explaining why and how do forecast combination work. Given the forecast criteria used in this literature are often restricted to the standard measures, such as Mean Square Errors (MSE), it is unclear if the superiority of forecast combination depends on the forecast criteria. Since MSE is not appropriate to evaluate the performance of VaR forecast, it provides an ideal platform to examine the performance of forecast combination with non-standard forecast criteria. Secondly, if forecast combination is shown to be beneficial for improving VaR forecast, then it will provide academics and practitioners a more flexible methods to understand the characteristics of the conditional return distribution. This increased awareness will allow for more accurate financial management as well as further developments in econometrics and modelling methodology.

The paper is organised as follows: Section 2 will introduce the forecast combination problem for VaR, Section 3 will introduce the four models used in this study and a brief discussion of the data. This is followed by the empirical findings in Section 4 and some concluding remarks can be found in Section 5.

2 COMBINING FORECASTS FOR VaR

This section introduces the forecast combination problem for Value-at-Risk (VaR). In order to keep the exposition simple, the paper focuses on a specific construction of VaR based on parametric volatility models. The method of forecast performance evaluation will also be discussed.

2.1 Construction of Value-at-Risk (VaR)

Although it is often criticised in the literature as an inconsistent measure of Risk, see for example, Artzner et al. (1999), VaR continues to be an important proxy of risk for regulators. This is supported by the Basel Accord where VaR is used as a measure of market risk, which banks must appropriately model and forecast. Given banks are free to forecast VaR based on their own technology, the Basel Accord evaluates the performance of each bank's VaR forecast by utilising a back-testing procedure. In a nutshell, the VaR forecasts for the past 250 days will be compared to the actual market returns. In the case when the market returns dropped below the VaR forecasts provided by the bank, a violation is recorded. In general, for a VaR forecast with 99% confident, one would expect approximately 2 to 3 violations in 250 days. Therefore, if the percentage of violations during the back-testing period is more than the desired confident level then penalty will be imposed on the bank's capital requirement. Therefore, let p be the desired confident level, then the objective function for VaR forecasts is

$$\mathbb{L}_v(r_t - q_{i,t}) = \left| h^{-1} \sum_{t=1}^h \mathbf{1}(r_t - q_{i,t} < 0) - p \right|, \quad (1)$$

where r_t is the asset's returns at time t , $q_{i,t}$ is the VaR forecasts for time t from model i based on information available at τ with $\tau < t$ and h is the evaluation horizon. Moreover, $\mathbf{1}(A)$ is an indicator function such that $\mathbf{1}(A) = 0$ if A is false and $\mathbf{1}(A) = 1$ otherwise. Hence, the problem of seeking the optimal VaR model can be expressed as

$$q_{o,t} = \arg \min_{q_{i,t} \in Q} \mathbb{L}_v(r_t, q_{i,t}), \quad (2)$$

where $Q = \{q_{i,t} : i = 1 \cdots k\}$ is a set containing all available models for forecasting VaR.

Given the rapid development in parametric volatility models in the last two decades, it is natural to construct VaR forecasts based on these models. The method is equivalent to construct a confident intervals for the forecasts of the conditional mean (returns) accommodating potential dynamic in the conditional variance. Mathematically,

$$q_{i,t} = r_{i,t} - \kappa \sqrt{h_{i,t}} \quad (3)$$

where $r_{i,t}$ and $h_{i,t}$ are the return forecast and the conditional variance from model i for time t based on information available at time $t - 1$ (ie $\tau = t - 1$), respectively and κ is a critical value such that

$$P(\eta_{i,t} < \kappa) = p, \quad \forall t > 0$$

with

$$\eta_{i,t} = \frac{r_t - r_{i,t}}{\sqrt{h_{i,t}}}.$$

2.2 VaR Forecast Evaluations

There are two main issues concerning forecast evaluations for VaR forecasts. The first concerns with the forecast criteria such as equation (1). Essentially one is interested in testing

$$H_0 : \mathbb{E}[\mathbb{L}_v(r_t - q_{i,t})] = 0, \quad H_1 : \mathbb{E}[\mathbb{L}_v(r_t - q_{i,t})] \neq 0$$

given a confident level, p . Under the null hypothesis, the number of actual violations is not significantly different to the expected violations implied by the confident level, p . Christoffersen (1998) proposed a likelihood ratio test for this purposes. Consider the indicator sequence, $\{\mathbf{1}_t\}_{t=1}^T$ constructed from a given interval forecast, such that $\mathbf{1}_t = \mathbf{1}(r_t < q_{i,t})$. The likelihood under the null hypothesis is $L(p; \{\mathbf{1}_t\}_{t=1}^T) = (1-p)^{T_0} p^{T_1}$ and the likelihood under the alternative is $L(p'; \{\mathbf{1}_t\}_{t=1}^T) = (1-p')^{T_0} p'^{T_1}$, where $T_1 = \sum_{t=1}^T \mathbf{1}_t$ and $T_0 = T - T_1$. Therefore, the likelihood ratio test for *unconditional coverage* is

$$LR_{UC} = -2 [\log L(p; \{\mathbf{1}_t\}_{t=1}^T) - \log L(p'; \{\mathbf{1}_t\}_{t=1}^T)] \quad (4)$$

where $\hat{p}' = T_1/T$ and $LR_{UC} \stackrel{a}{\sim} \chi^2(1)$. The basic idea is to compare the expected percentage of violation with the actual percentage of violation given a sample and a set of VaR forecast. If the VaR forecasts are accurate then the actual number of violations should not be statistically different to the expected number of violations.

In addition to the actual number of violations. It is also important to examine if the violations are correlated. Theoretically, the violations should occur independently from each other. Under the assumption of a first-order Markov chain. Diebold and Lopez (1996) and Christoffersen (1998) developed a likelihood ratio test of independence.

Consider a first order Markov chain for the indicator sequence, $\{\mathbf{1}_t\}_{t=1}^T$, the unrestricted transition matrix is

$$\Pi' = \begin{pmatrix} 1 - p_{01} & p_{01} \\ 1 - p_{11} & p_{11} \end{pmatrix} \quad (5)$$

where $p_{ij} = P(\mathbf{1}_{t+1} = j | \mathbf{1}_t = i)$. The associated likelihood function is $L(\Pi' | \{\mathbf{1}_t\}_{t=1}^T) = p_{11}^{T_{11}} p_{01}^{T_{01}} (1 - p_{11})^{T_{10}} (1 - p_{01})^{T_{00}}$ where T_{ij} is the number of observations such that $\mathbf{1}_t = j$ and $\mathbf{1}_{t-1} = i$. Thus, the unrestricted transition matrix can be estimated by $p_{ij} = T_{ij}/T$. Moreover, imposing independence is equivalent to applying the restriction $p_{10} = p_{01}$ and therefore the restricted transition matrix is

$$\Pi = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix} \quad (6)$$

which leads to the likelihood function, $L(\Pi | \{\mathbf{1}_t\}_{t=1}^T) = (1-p)^{T_0} p^{T_1}$ where p can be estimated by $\hat{p} = T_1/T$. Thus, the likelihood ratio test for independence is

$$LR_{IND} = -2 [\log L(\Pi | \{\mathbf{1}_t\}_{t=1}^T) - \log L(\Pi' | \{\mathbf{1}_t\}_{t=1}^T)] \quad (7)$$

with $LR_{IND} \stackrel{a}{\sim} \chi^2(1)$. Notice this test does not concern if \hat{p} is different to the given statistical confidence level p , the sole objective of this test is to examine the independence between violations. However, it is possible to combine the tests of unconditional coverage and independence by simply combining the two test statistics, resulting in the *conditional coverage* test, namely

$$LR_{CC} = LR_{UC} + LR_{IND} \stackrel{a}{\sim} \chi^2(2) \quad (8)$$

for the following hypothesis:

$$H_0 : \mathbb{E}(\mathbf{1}_t - p) = 0, \quad H_1 : \mathbb{E}(\mathbf{1}_t - p) \neq 0.$$

2.3 Forecast Combination of VaR

Let y_t denotes the variable of interest, or the targeted variable, and $y_{i,t}$ denotes the forecast of y_t by model i at time τ with $\tau < t$. Moreover, assume the forecasts from each model is combined following the specification of a continuous function $g(y_{1,t}, \dots, y_{k,t}; w)$ with a vector of parameters, w , then the forecast combination problem given a particular forecast criterion at time τ is:

$$\hat{w} = \arg \min_{w \in W} \mathbb{L}[y_t - g(y_{1,t}, \dots, y_{k,t}; w)] \quad (9)$$

The problem is to seek the vector \hat{w} so that the forecast criterion, \mathbb{L} , is minimised with a given combination function g . Obviously, the determination of \hat{w} depends on the choice of the forecast criterion, \mathbb{L} , and the specification of g . \mathbb{L} is typically chosen to be Mean-Squared Errors (MSE), Mean Absolute Deviations (MAD) or Mean Percentage Absolute Errors (MAPE). The general forecast combination is an aggregate of information from the various forecasts based on the different models resulting in a summary measure. Bates and Granger (1969), Clemen (1989), Hogarth (1989), Makridakis (1989), Winkler (1989), Hibon and Makridakis (2000), Clements and Hendry (2001), Timmermann (2005), Guidolin and Timmermann (2007) among others, have established that combining forecasts implies incorrect specification of the underlying models. These studies proposed that each model captured a different aspect of the data generating process. Therefore, forecasts from these models will each contain some level of independent information. This information is due to the assumptions made within the specific models about the inter-variable relationships and the importance placed on individuals variables. Combining multiple forecasts captures numerous partially overlapping interpretations of the underlying data generating process, as argued by Clements and Hendry (2001).

Clearly, the forecast combination problem can be modified for VaR. Define the actual VaR, q_t , such that $P(r_t - q_t < 0) = p$ for some level of confident, p . The forecast criterion, in this case, is \mathbb{L}_v , as defined in equation (1). Therefore, the forecast combination problem for VaR is

$$\hat{w} = \arg \min_{w \in W} \mathbb{L}_v[r_t - g(q_{1,t}, \dots, q_{k,t})] \quad (10)$$

Similar to the standard problem, the problem above cannot be solved in practice due to the unavailability of r_t at time τ . Therefore, weighting schemes are proposed to approximate \hat{w} . In the case of simple averaging,

$$g(q_t, w) = w'q_t \quad (11)$$

with $w = i/k$ where $q_t = (q_{1,t}, \dots, q_{k,t})'$ and i is a column vector of 1's. An alternate weighting schemes is based on Quantile Regression (QR) as proposed originally by Bassett and Koenker (1978). The idea is to seek the linear combination that best fitted the $(1-p)^{th}$ quantile of existing return. Formally,

$$Q_\nu(r_t|q_t) = w_0 + w'q_t \quad (12)$$

where $Q_\nu(r_t|q_t)$ denotes the ν^{th} quantile of r_t conditional on q_t . The weight vector can be found by solving

$$(\tilde{w}_0, \tilde{w}')' = \arg \min_{(w_0, w')' \in \mathbb{R}^{k+1}} \sum_{\varepsilon_t < 0} p |\varepsilon_t| + \sum_{\varepsilon_t \geq 0} (1-p) |\varepsilon_t| \quad (13)$$

where $\varepsilon_t = r_t - w_0 - w'q_t$.

An objective of this paper is to compare the forecast performances between the two weighting schemes. VaR forecasts from four different models will be constructed and combined based on the two weighting schemes and the forecast performances will be compared using the forecast criterion as defined in equation (1). The forecast performances of each of the individual models will also be discussed. The details of model considered in the paper as well as the data used for the empirical sections will be discussed in the next section.

3 MODELS AND DATA

This paper considered four different volatility models to construct four different VaR forecasts. The four models considered are the Exponential GARCH model of Nelson (1991), the Stochastic Volatility originated in Taylor (1986) and popularized by Harvey et al. (1994) and Harvey and Shephard (1996), Integrated GARCH model of Engle and Bollerslev (1986) and Implied Volatility Model.

These models are chosen because they assume different dynamics on the conditional variance of the asset returns and they are not nested within each other. This satisfies the condition of information independence as discussed in previous sections. Given the rapid theoretical developments in estimating and testing these models, the readers are referred to the aforementioned references for technical details.

The data used in this paper are the daily returns of S&P 500 and FTSE price index from 3 January 1996 to 3 August 2010 provided by DataStream. Let P_t denotes the price at time t , the returns are calculated as: $r_t = 100 * \log(P_t/P_{t-1})$.

4 EMPIRICAL RESULTS

The p-values of the conditional coverage tests for the VaR forecasts from the six different models for both S&P and FTSE can be found in Tables (1), respectively. Interestingly, forecasts from both Simple Weighted Average and IGARCH passed the conditional coverage test for both series. That is, both forecast series did not have significant different number of violations at the confidence levels and the violations appeared to be independent from each other. This is also true for EGARCH for FTSE but not for S&P while the SV model failed the conditional coverage test in both cases. The performance of Quantile regression is also a little disappointing. It passed the conditional coverage test at 5% level but not at 1% or 2% for both indexes. This is not overly surprising as the expected numbers of violations are small at the 1% and 2% levels. This implies the number of observations available for estimating the lower quantile is very small and therefore producing less reliable results. However, it passed the conditional coverage test at 5% where the number of observations available for estimating the quantiles is sufficient to produce more reliable estimates.

The forecast performance of each model for both indexes as measured by comparing the percentage of actual violations with the expected percentage of violations, namely the specified confidence level can be found in Tables (2). Interestingly, the Simple Weighted Average was the most robust in terms of forecast performance as being consistently the best or second best across all confidence levels and indexes. More importantly, QR performed really well at 5% but not at 1% or 2%. This is consistent with the results of the conditional coverage test and provided further evidence to support the aforementioned reasons for the poor results from QR. The performance of IGARCH was also robust and consistent with the results from the conditional coverage test.

Overall, the paper obtained evidence that forecast combination can be beneficial in forecasting VaR thresholds. Similar to the case of forecasting conditional mean, the Simple Weighted Average seemed to be more robust and superior than the more sophisticated QR technique. However, this might be due to limited number of observations at one lower quantile as confidence level decrease. At the confidence level where all quantiles have sufficient observations, QR could in fact produce better results than the Simple Weighted Average.

Models	S&P			FTSE		
	1%	2%	5%	1%	2%	5%
QR	< 0.001	0.009	0.731	< 0.001	< 0.001	0.734
SWC	0.951	0.743	0.606	0.613	0.815	0.631
EGARCH	< 0.001	0.001	0.017	0.223	0.323	0.038
IGARCH	0.407	0.252	0.481	0.632	0.815	0.366
SV	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
IV	0.007	< 0.001	< 0.001	0.007	< 0.001	< 0.001

Table 1. LR Test for Conditional Coverage

Models	S&P			FTSE		
QR	6	3	1	6	4	1
SWC	1	1	2	2	2	2
EGARCH	4	4	4	3	3	4
IGARCH	2	2	3	1	2	3
SV	6	6	6	6	6	6
IV	3	5	6	4	5	6

Table 2. Model Performance (Rank)

5 CONCLUSION

This paper investigated the potential benefit of forecast combination in forecasting VaR threshold. It provided a theoretical framework for analysing forecast combination in the second moment and it compared the forecast performances in VaR thresholds between two different forecast combination methods along with 4 different volatility models. The results suggested that forecast combination could in fact, improve forecast accuracy. Moreover, similar to the forecast combination problem for conditional mean, the Simple Weighted Average is the most robust method of combining forecast. However, a technique based on quantile regression can be superior, especially when the number of observations in each quantile are sufficient to ensure reliable estimates.

It is important to note that the construction of VaR forecasts and the choice of volatility models were very limited in the study. No comparison was made between different distributional assumptions and there was no comparison between non-parametric volatility, semi-parametric and parametric models. However, even under such limited set of models, at least one type of forecast combination technique can produce superior forecast that satisfied both unconditional coverage (number of actual violations is not significantly different to the expected number of violations) and independence (violations are independent) criteria.

ACKNOWLEDGEMENT

The author would like to thank Harry Bloch, Pat Fraser, Laurent Pauwels, Marco Reale and Hiroaki Suenaga for insightful discussions and the participants of the 2011 Conference on Computational and Financial Econometrics, London. The financial support from the Australian Research Council is gratefully acknowledged.

REFERENCES

- Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9, 203–228.
- Bassett, G. J. and R. Koenker (1978). Regression quantiles. *Econometrica* 46, 33–50.
- Bates, J. and C. Granger (1969). The combination of forecasts. *Operation Research Quarterly* 20, 451–468.
- Christoffersen, P. (1998). Evaluating interval forecast. *International Economic Review* 39, 841–862.
- Clemen, R. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5, 559–583.
- Clements, M. and D. Hendry (2001). Pooling of forecast. *Econometrics Journal* 5, 1–26.
- Diebold, F. and J. Lopez (1996). Forecast evaluation in combination. Technical Report Research Paper 9525, Federal Reserve Bank of New York.
- Engle, R. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1–50.
- Guidolin, M. and A. Timmermann (2007). Forecasts of U.S. short-term interest rates: A flexible forecast combination approach. *Federal Reserve Bank of St. Louis Working Papers No*, 2005–059.
- Harvey, A., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance model. *Review of Economic Studies* 61, 247–264.
- Harvey, A. and N. Shephard (1996). Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics* 14, 429–434.
- Hibon, M. and S. Makridakis (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting* 16, 451–476.
- Hogarth, R. (1989). On combining diagnostic ‘forecasts’: Thoughts and some evidence. *International Journal of Forecasting* 5, 593–597.
- Makridakis, S. (1989). Why combining works? *International Journal of Forecasting* 5, 601–603.
- Nelson, D. (1991). Conditional heteroskedasticity in asset returns - a new approach. *Econometrica* 59(2), 347–370.
- Taylor, S. (1986). *Modelling Financial Time Series*. Chichester: Wiley.
- Timmermann, A. (2005). Forecast combination. *CEPR Discussion Papers CEPR Discussion Paper*, 5361.
- Winkler, R. (1989). Combining forecasts: A philosophical basis and some current issues. *International Journal of Forecasting* 5, 605–609.