A GMM procedure for combining volatility forecasts

Alessandra Amendola, Giuseppe Storti*

Dipartimento di Scienze Economiche e Statistiche, Università di Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy

Available online 7 October 2007

Abstract

A novel approach to the combination of volatility forecasts is discussed. The proposed procedure makes use of the generalized method of moments (GMM) for estimating the combination weights. The asymptotic properties of the GMM estimator are derived while its finite sample properties are assessed by means of a simulation study. The results of an application to a time series of daily returns on the S&P500 are presented.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Volatility; Forecast combination; GMM; GARCH

1. Introduction

The availability of private information and/or differences in modelling approaches leads to multiple forecasts of the level of one or more variables of interest. In this case, it has been shown how a combination of different forecasts can be used in order to increase the forecast accuracy (Clemen, 1989). The literature on the combination of forecasts dates back to the paper by Bates and Granger (1969). Given a forecast horizon $h$, the aim of any forecast combination procedure is to find an optimal vector of weights such that the new predictor, obtained as a linear combination of the candidate predictors, performs better than the single competitors, according to an appropriately chosen loss function.

In their approach, Bates and Granger imposed a convexity constraint on the combination weights. The convexity assumption is not harmless and it can be too restrictive when some of the predictors involved in the combination are biased. However, biased forecasts can still be combined to yield a forecast with zero average forecast error by resorting to a non-convex combination (Granger and Ramanathan, 1984).

Previous research on forecast combination has been mainly focused on the combination of conditional mean forecasts (for a recent review Timmermann, 2006). In this paper the interest is in the combination of volatility forecasts. The main problem is that the volatility cannot be directly observed. Hence loss functions such as the MSE (mean squared error) cannot be used unless a suitable proxy of the conditional variance is defined. A common approach is to use the squared returns as a proxy but these offer a noisy measure of the volatility and, in many settings, their use can give rise to a biased assessment of the performance of the candidate models (Hansen and Lunde, 2004). A more accurate approximation can be obtained by referring to the concept of realized volatility even if, at very high frequencies, micro-structure market frictions can distort such a measure of the unobserved volatility (Andersen et al., 2005). Also, moving from the raw (tick-by-tick) data, the generation of an ultra high frequency time series of returns usually requires the application of

* Corresponding author. Tel.: +39 089962212; fax: +39 089962049.

E-mail addresses: alamendola@unisa.it (A. Amendola), storti@unisa.it (G. Storti).

0167-9473/ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.10.001
cleaning filters which, in some cases, could have consequences on the subsequent results (Brownlees and Gallo, 2006). A further problem to be considered is that, in many applications, high frequency data on the phenomenon of interest are not available and so realized volatility measures cannot be computed. For example, this usually happens when working with macroeconomic data for which the maximum sampling frequency is monthly.

The aim of this paper is to present a novel approach to the combination of volatility forecasts in which the optimal combination weights are estimated by the generalized method of moments (GMM) imposing appropriate conditions on the standardized residuals implied by a given set of combination weights. The idea is to constrain the standardized residuals to be as close as possible to a sequence of i.i.d. zero mean, unit variance random variables. A related approach based on convex combination weights has been previously proposed (Storti, 2005).

The structure of the paper is as follows. In Section 2 the GMM procedure for combining volatility predictions is illustrated and its asymptotic properties are derived. The finite sample properties of the GMM estimator are investigated in Section 3 by means of a simulation study while Section 4 presents an application to a time series of daily stock returns. Some concluding remarks are given in the last section.

2. An unconstrained GMM approach to combining volatility forecasts

2.1. The GMM estimator

In this section a novel approach to the combination of volatility forecasts generated from different models is introduced and discussed. Let \( y_t, t = 1, \ldots, T \), be an observed time series generated by a stationary stochastic process \( \{ y_t \} \). It is assumed that the data generating process (DGP) \( y_t \) is of the form:

\[
\begin{align*}
  y_t &= x_t + u_t, \\
  u_t &= h_t z_t,
\end{align*}
\]

where \( z_t \sim (0, 1) \), \( \forall t \), is a series of i.i.d. random variables, \( (u_t|I^{t-1}) \sim (0, h_t^2) \), \( x_t = x(I^{t-1}) \), \( h_t = h(I^{t-1}) \) and \( I^{t-1} \) denotes the set of information available at time \( t - 1 \). This can potentially include lagged values of \( y_t \) as well as other regressors.

Assuming that a set of \( k \) candidate models for \( y_t \) is potentially available, let \( \hat{x}_{ti}, i = 1, \ldots, k \), be the one-step-ahead predictor of \( y_t \) generated by the \( i \)th model. The unconstrained combined predictor of the level of the \( y_t \) process can be defined as

\[
\tilde{x}_t = \sum_{i=1}^{k} \tilde{w}^{(x)}_i \hat{x}_{ti},
\]

(2)

with \( \tilde{w}^{(x)}_i \in \mathbb{R} \). Similarly, let \( \tilde{h}^2_{ti} \), for \( i = 1, \ldots, k \), be the (one-step-ahead) predicted volatility generated by the \( i \)th model. The unconstrained combined volatility predictor can be defined as

\[
\tilde{h}^2_t = \sum_{i=1}^{k} \tilde{w}^{(h)}_i \tilde{h}^2_{ti},
\]

(3)

where \( \tilde{w}^{(h)}_i \geq 0 \), for \( i = 1, \ldots, k \). The main advantage of adopting an unconstrained combination scheme is that it allows an unbiased combined predictor even if one or more of the candidate predictors are biased. The assumption of non-negative weights is required in order to guarantee the non-negativity of the combined volatility predictor.

To solve the problem of estimating the optimal combination weights in (3) we suggest an approach based on the application of the GMM. Moments based estimators have already been used for estimating GARCH models parameters (Kristensen and Linton, 2006; Storti, 2006). Other applications of the GMM approach in finance have been surveyed by Jagannathan et al. (2002). However, the application of these techniques to the combination of volatility forecasts still deserves investigation.

In the proposed approach, the unknown combination weights \( (\tilde{w}^{(x)}_i, \tilde{w}^{(h)}_i) \) are chosen to minimize the criterion

\[
\tilde{w} = \arg\min_w m_T(w) M_T m_T(w),
\]

(4)
where $\tilde{w} = (w^{(1)}_1, \ldots, w^{(N)}_1; \tilde{w}^{(h)}_1, \ldots, \tilde{w}^{(h)}_k)'$, $m_T(w) = (1/T)\sum_{t=1}^T \mu(w, t)$, $\mu(w, t)$ is a $(N \times 1)$ vector of moment conditions such that there exists a $w_o$ for which $E(\mu(w_o, t)) = 0$ and $MT$ is a $(N \times N)$ stochastic positive definite $Op(1)$ weighting matrix. Each element of $\mu(w, t)$ specifies a restriction on the moments structure of the process

$$z_t = \frac{y_t - \tilde{x}_t}{h_t}.$$  

It can be immediately noted how the $z_t$ can be interpreted as the standardized residuals from the combined volatility estimator.

The moment conditions which can be potentially considered are of the form

$$\mu_{t,1} = z_t, \quad \mu_{t,2} = z_t^2 - 1,$$

$$\mu_{t,i+2} = z_t z_{t-i}, \quad \mu_{t,j+g_1+2} = z_t z_{t-j} - 1,$$

for $i = 1, \ldots, g_1$, $j = 1, \ldots, g_2$ and $l = 1, \ldots, g_3$, with $N = 2 + \sum_{i=1}^3 g_i$ being the total number of moment conditions.

Under optimal forecasts, the standardized residuals are expected to be a sequence of zero mean, unit variance i.i.d. random variables. Therefore, the rationale behind the choice of the above reported moment conditions is to constrain the standardized residuals $z_t$, implied by a given set of combination weights, to be as close as possible to a sequence of i.i.d. zero mean, unit variance random variables. So, the first two conditions $\mu_{t,1}$ and $\mu_{t,2}$ control the first two moments of the marginal distribution of $z_t$ while the remaining sets of moment conditions $\mu_{t,j}$ $(j > 2)$ are aimed at imposing serial independence in the dynamic structure of the residuals $z_t$. In particular, for $2 < j \leq (g_1 + 2)$, setting $E(\mu_{t,j}) = 0$ is equivalent to impose that the residuals $z_t$ are serially uncorrelated. Differently, for $(g_1 + 2) < j \leq (g_1 + g_2 + 2)$, $E(\mu_{t,j}) = 0$ rules out any serial correlation in the squared residuals. Finally, the last set of conditions allows to account for the presence of unmodelled leverage effects in the volatility dynamics. The values of $g_1$, $g_2$ and $g_3$ can be selected on a case by case basis depending on the data collection frequency and on the nature of the problem to be analyzed. For example, when working with intraday financial returns, the volatility is typically characterized by the presence of long memory effects which means that the theoretical autocorrelation function of squared returns is hyperbolically decaying. In such a case, depending on the sampling frequency, it is advisable to choose a sufficiently large number of lags in order to allow for proper identification of the volatility dynamics. Differently, for daily returns, setting $max(g_i) \leq 10$ (10 days $\approx 2$ weeks) will in general suffice.

2.2. Statistical properties

Here the statistical properties of the proposed GMM estimator of the combination weights are investigated. Namely, we prove the consistency and asymptotic normality of the estimator. To this purpose the following assumptions are required.

(A1) If for some vector $w_o$,

$$E(\mu(w_o, t)) = 0,$$

then, for any other vector $w^* \neq w_o$,

$$E(\mu(w^*, t)) \neq 0.$$

(A2) $w \in W$ where $W$ is compact.

(A3) $\{Y_t\}$ is strictly stationary and $\alpha$-mixing of size $\alpha > -r/(r - 2)$, for $r > 2$.

(A4) $M_T \xrightarrow{p} M$, where $M$ is a positive definite matrix, as $T \to \infty$.

(A5) $w_o$ is an interior point of $W$.

(A6) $E(\|\mu(w_o, t)\|^2) < \infty$.

(A7) Let $G(w, t) = \nabla_w \mu(w, t)$ be the gradient of $\mu(w, t)$ with respect to $w$. We assume that $E(\sup_{w \in W} \|G(w, t)\|) < \infty$ and that $G' MG$ is non-singular for $G = G(w_o)$. 
Our setup is as follows. To generate the combined volatility predictor at time $T + 1$, $\tilde{h}_{T+1}^2$, we use a sequence of past volatility predictions from each candidate model ($\hat{h}_{t,i}^2$), $i = 1, \ldots, k, t \leq T$. These can potentially be in-sample ($t \leq T_e$) or out-of-sample ($t > T_e$) predictions where $\{Y_1, \ldots, Y_{T_e}\}$ is the time series used for the estimation of the candidate models. In both cases, the $\hat{h}_{t,i}^2$ only depend on information up to time $T - 1$. Differently, to compute the standardized residual $\tilde{z}_T$, we need to use observation $Y_T$ as well. So the estimation sample for $w$ is given by $\{Y_1, \ldots, Y_T\}$ and the resulting estimator is indicated as $\tilde{w}_T$. The following theorems prove the consistency and asymptotic normality of $\tilde{w}_T$. The proof of both theorems follows from the application of standard results from the theory of GMM estimators.

**Theorem 1 (Consistency).** Under Assumptions 1–4, $\tilde{w}_T \xrightarrow{p} w_o$.

**Proof.** If $Y_t$ is z-mixing and stationary, it is also ergodic. Then Theorem (3.35) in White (2001) applies to show that $\mu(w, t)$ is also stationary and ergodic. The proof of the desired result then follows by Theorem 2.6 in Newey and McFadden (1994). □

**Theorem 2 (Asymptotic normality).** Under Assumptions 1–7, as $T \to \infty$, $\sqrt{T}(\tilde{w}_T - w_o) \xrightarrow{d} N(0, V)$ where

$$V = (G'MG)G'M\Omega MG(G'MG)^{-1},$$

with

$$\Omega = \lim_{T \to \infty} T E(m_T(w_o)m_T(w_o)'),$$

The asymptotic variance is minimized for $M = \Omega^{-1}$. In this case the above expression simplifies to

$$V = (G'\Omega^{-1}G)^{-1}.$$

**Proof.** The proof follows immediately from the application of Theorem 3.4 in Newey and McFadden (1994). □

Although the above assumptions are standard in the asymptotic theory of M-estimators, it is worth noting that some of them could be relaxed. The consistency of GMM estimators can be still proved even if the parameter space $W$ is not assumed to be compact. However, this requires some additional restrictions on the shape of the objective function. Newey and McFadden (1994), for example, show that consistency holds without compactness of the parameter space in cases in which the objective function is assumed to be concave.

Also, we could allow for more general forms of dependence, replacing Assumption 3 by a near epoch dependence (NED) assumption. Hansen (1991) has investigated the conditions under which GARCH (1,1) processes are NED. However, assuming strict stationarity and mixing has some important practical advantages. In the past decade much attention has been paid to the investigation of the mixing properties of conditional heteroskedastic processes leading to the derivation of readily interpretable and easy to check conditions for a wide class of GARCH type models (see, e.g. Carrasco and Chen, 2002; Francq and Zakoïan, 2006).

The weighting matrix $M_T$ plays an important role in GMM estimation. Although its choice does not affect consistency, it can have substantial effects on the efficiency of the GMM estimator. It can be shown that, in order to achieve asymptotic efficiency, it is necessary to set $M_T = \hat{\Omega}^{-1}$ where $\hat{\Omega}$ is a (weakly) consistent estimator of the asymptotic covariance matrix $\Omega$. In practice $\Omega$ can be estimated by the Newey–West heteroskedasticity and autocorrelation robust estimator (Newey and West, 1987):

$$\hat{\Omega}_{NW} = \hat{\Omega}_0 + \sum_{j=1}^q \left(1 - \frac{j}{1+q}\right) (\hat{\Omega}_j + \hat{\Omega}_j').$$ (5)
with

\[ \hat{\Omega}_j = \frac{1}{T} \sum_{t=j+1}^{T} \mu(\hat{\mathbf{w}}, t) \mu(\hat{\mathbf{w}}, t-j), \]

(6)

where \( \hat{\mathbf{w}} \) is a consistent estimator of \( \mathbf{w} \) and \( q \) is an adequately defined truncation point called the lag truncation parameter. For preserving the asymptotic properties of the GMM estimator, the value of \( q \) must increase as \( T \) does. In particular, \( q \) must go to infinity at some suitable rate as the sample size goes to infinity. The appropriate rate can be shown to be equal to \( T^{1/3} \). A procedure for selecting \( q \) automatically has been proposed (Newey and West, 1994).

However, in order to attain asymptotic efficiency, a feasible solution is to estimate \( \hat{\Omega}_{NW} \) by an iterative procedure. First, a preliminary estimate of the weight vector \( \mathbf{w} \) is obtained using the identity matrix as an initial weighting matrix. In this way a consistent estimate \( \hat{\mathbf{w}}^{(1)} \) of the weights vector \( \mathbf{w} \) is obtained. These values are then plugged into (5) and a new estimate \( \hat{\mathbf{w}}^{(2)} \) of \( \mathbf{w} \) is obtained. The procedure is iterated until convergence is reached. Convergence is assessed in terms of the absolute relative variation in the GMM objective function

\[ \varepsilon^{(i)} = \frac{|Q(\mathbf{w}^{(i)}) - Q(\mathbf{w}^{(i-1)})|}{Q(\mathbf{w}^{(i-1)})}, \]

(7)

where

\[ Q(\mathbf{w}) = m_T(\mathbf{w})^\prime M_T m_T(\mathbf{w}). \]

(8)

In the above discussion, the weights have been estimated conditional on the values of the volatility estimates \( \hat{h}^2_{t,i} \) generated by the \( k \) candidate models. In practice, two different cases can arise depending on if we are interested in in-sample or out-of-sample estimation of volatility. In the first case, the parameters \( \theta_i \) of the candidate models \( (i=1, \ldots, k) \) are estimated on the same set of data used to estimate the combination weights, \((y_1, \ldots, y_T)\) leading to the estimator \( \hat{\theta}_{i,S} \) for \( i=1, \ldots, k \). Differently, in the case we are interested in out-of-sample volatility prediction, the \( \theta_i \) have been estimated on a different set of data preceding the time series used for the estimation of the combination weights \( \mathbf{w} \). More precisely, the set of available data is divided into two subseries. The first subseries \((y_1, \ldots, y_S)\) is used to obtain an estimate of the parameters of each candidate model \( \hat{\theta}_{i,S} \). For each candidate model, these estimates are used to generate a time series of volatility predictions \( \hat{h}^2_{t+1,i} \), for \( t=S, \ldots, T-1 \). The combination weights \( \mathbf{w} \) are then estimated using the second subseries \((y_{S+1}, \ldots, y_T)\).

3. A simulation experiment

In order to assess the ability of the proposed combination rule in volatility prediction, we carry out a Monte Carlo simulation experiment comparing different volatility models. As DGPs we have considered different GARCH(1,1) models (Bollerslev, 1986). In each process the parameterizations are chosen to achieve different levels of persistence \( \gamma = \alpha + \beta \). Furthermore, in order to avoid any masking effects, each of the candidate GARCH(1,1) models has been constrained to have unit unconditional variance. Namely, the chosen High (H), Medium (M) and Low (L) persistence parameterizations are

\[ h^2_{H,t} = 0.05 + 0.10u_{t-1}^2 + 0.85h^2_{1,t-1}, \]
\[ h^2_{M,t} = 0.15 + 0.15u_{t-1}^2 + 0.70h^2_{2,t-1}, \]
\[ h^2_{L,t} = 0.40 + 0.20u_{t-1}^2 + 0.40h^2_{3,t-1}. \]

These DGPs have been used as component models of three volatility forecasts combinations. Each combination includes two of the above models considering all possible pairs:

\[ h^2_{HL,t} = 0.8h^2_{H,t} + 0.5h^2_{L,t}, \]
\[ h^2_{ML,t} = 0.8h^2_{M,t} + 0.5h^2_{L,t}, \]
\[ h^2_{HM,t} = 0.8h^2_{H,t} + 0.5h^2_{M,t}. \]
The proposed GMM estimation procedure has been then applied to estimate the combination weights considering different sample sizes, $T = 1500, 3000, 4500$, (as usual the first 25% of the generated observations have been dropped out in order to avoid sensitivity to initial conditions) and different sets of moment conditions $g = \{1, 3, 5, 7, 10\}$. In all the experiments we calculate the point estimates from 1000 Monte Carlo replications. For each combination scheme, the simulated distributions of the estimated combination weights have been summarized by means of box-plots (Figs. 1–3). For the High–Low combination (1), in general, the performance of the estimation algorithm does not appear to be dramatically sensitive to the value of $g$. However, for $g \leq 5$ the variability tends to decrease with $g$ while the bias component remains fairly stable while, for $g > 5$, the bias tends to increase (in modulus) with $g$. So, overall, the choice $g = 5$ seems to offer the best compromise. A similar pattern is observed for the Medium–Low combination (2), although it must be observed how, in this case, the choice of $g$ can be critical if the estimation window used is not sufficiently long. Namely, for $g = 1$ and $T = 1500$ the estimates of the combination weights appear quite unstable even if the situation substantially improves as $g$ increases from 1 to 3. Finally, for the High–Medium combination scheme, the convergence of the simulated distributions of $\tilde{w}_i$ $(i = 1, 2)$ to their asymptotic counterparts appears to be much slower. This result can be explained by considering that the persistence gap between the two components is in this case much lower than in the other two combination schemes which is 0.10 versus 0.25 and 0.35, respectively. So, since the two
volatility components are characterized by very similar dynamical properties, more information is needed in order to properly identify their individual contributions to the volatility of the observed process.

4. Empirical evidence on financial data

4.1. Forecasting the volatility of the S&P 500

The results of an application to a time series of daily returns on the S&P500 stock market index are presented. We focus on percentage returns calculated as the first difference of the log-transformed price series ($\times 100$)

$$r_t = 100 \times \log \left( \frac{P_t}{P_{t-1}} \right),$$

where $P_t$ is the index value at time $t$. The series (Fig. 4) covers the period from October 24, 1988 to December 29, 2006, for a total of 4586 observations.

As expected the data are characterized by a high kurtosis and a moderate negative skewness (Table 1). Also, the correlogram of the raw data reveals a significant Ljung-Box $Q$ statistic at lags 5 and 10. In order to account for this we tentatively specified low order AR and MA models. However, the estimated coefficients did not turn out to
Fig. 3. High–Medium persistence: simulated sampling distribution of $\tilde{w}_i$ ($i = 1, 2$) for different sample sizes $T = \{1500, 3000, 4500\}$ and sets of moment conditions $g = \{1, 3, 5, 7, 10\}$.

Fig. 4. Daily returns on the S&P 500 from October 24, 1988 to December 29, 2006.
be significantly different from zero at the usual significance levels. So, for the sake of parsimony, we assume the conditional mean model to be given by a simple constant term

\[ r_t = \mu + u_t. \]

differently the squared returns are characterized by a persistent autocorrelation function decreasing very slowly as the lag increases (Fig. 5).

The volatility of the mean corrected data \( u_t \) has been modelled using three different parameterizations, a GARCH, a threshold GARCH (TGARCH) model (Rabemananjara and Zakoïan, 1993) and a two-component GARCH (CGARCH) model (Ding and Granger, 1996). All the models used are assumed to be of order (1,1). For the TGARCH(1,1) model, the updating equation for the conditional variance is given by

\[
h_{t}^2 = (\omega + a_1 |u_{t-1}| + b_1 h_{t-1} + c_1 I(u_{t-1} < 0)|u_{t-1}|)^2,
\]

with \( I(\cdot) \) being the indicator function. Note that, differently from standard GARCH, this model focuses on the dynamic structure of the conditional standard deviation rather than the conditional variance. Also, it allows to account for the presence of leverage effects in the conditional variance dynamics.

The CGARCH(1,1) model is defined as a mixture of two GARCH(1,1) models. Namely, the resulting conditional variance equation is given by

\[
h_{t}^2 = \eta_1 h_{t,1}^2 + (1 - \eta_1) h_{t,2}^2,
\]

where

\[
h_{t,1}^2 = \omega_1 + a_{1,1}u_{t-1}^2 + b_{1,2}h_{t-1,1},
\]

\[
h_{t,2}^2 = a_{1,2}u_{t-1}^2 + (1 - a_{1,2}) h_{t-1,2}^2.
\]

The second volatility component is restricted to be integrated GARCH in order to guarantee the identifiability of the model parameters. The reason for which we have considered the CGARCH model as candidate model is that it allows to reproduce long-memory effects in the volatility dynamics. This issue was first investigated in the paper by Ding and Granger (1996). Furthermore, Maheu (2005) has documented, by means of a simulation study and a forecasting exercise, how a modified CGARCH model (Engle and Lee, 1999) is able to capture long-range dependence in the volatility process. Finally, on a theoretical ground, the ability of CGARCH models to reproduce a hyperbolically decaying volatility autocorrelation function has been recently investigated (Haas, 2007). All the models have been

---

Table 1

Descriptive statistics and \( p \)-Values of the Ljung-Box \( Q \) statistic at lags 1, 5, 10 for the S&P 500 returns

<table>
<thead>
<tr>
<th>Mean</th>
<th>s.d.</th>
<th>min.</th>
<th>max.</th>
<th>skew.</th>
<th>kur.</th>
<th>( Q(1) )</th>
<th>( Q(5) )</th>
<th>( Q(10) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.035</td>
<td>0.984</td>
<td>-7.113</td>
<td>5.574</td>
<td>-0.162</td>
<td>7.207</td>
<td>0.735</td>
<td>0.047</td>
<td>0.014</td>
</tr>
</tbody>
</table>

---

Fig. 5. Autocorrelation function of S&P 500 returns (left) and squared returns (right) from October 24, 1988 to December 29, 2006.
estimated by maximum likelihood under the assumption of errors distributed according to a Student’s $t$ with $v$ degrees of freedom. For both model coefficients and combination weights we have used a rolling sample technique based on a moving data window of fixed length $n$ where the value of $n$ has been set equal to 3000. This choice was due to the fact that a sufficiently high number of data points are required for reliably identifying the conditional variance components in the estimation of CGARCH models.

The last 2886 data points have been kept for out of sample forecast evaluation while the first 3000 observations have been used for computing an initial set of estimates for the parameters of the candidate volatility models. Furthermore, as in the previous section, we have explored the sensitivity to the choice of the moments conditions used by the GMM estimator considering different values of $g$ and, namely, $g = 1, 3, 5, 7, 10$.

To initialize the procedure, the series of the first $n$ in-sample volatility predictions from each candidate model is used to generate an initial estimate of the vector of combination weights ($\hat{\mathbf{w}}_n$). For $t > n$, the weights and model coefficients are then re-estimated each $m$ observations over a moving window of size $n$ including observations from time $t - n + 1$ to $t$. In practice, the estimates are updated at all time points $t$ such that $t = n + (h \times m)$, where $h = 0, \ldots, s$ and $s$ is the maximum integer such that $n + sm \leq T$. Setting $m = 1$ implies that the weights and model coefficients are re-estimated each time a new observation becomes available. In this application the value of $m$ has been set equal to 100. Even if this choice was mainly aimed at limiting the computational burden it must also be considered that, in order to observe a relevant variation in the combination weights from one estimation step to the following one, the value of $m$ must be chosen far away from 1.

Letting $t = n + hm$, for some integer $h \geq 0$, at each re-estimation, we

1. Form the time series of past volatility predictions from each candidate model $\tilde{h}_{t-j,i}^2$, for $i = 1, \ldots, k$, $j = 0, \ldots, n - 1$.
2. Define $\tilde{z}_{t-j} = u_{t-j} / \sum_{j'=1}^k \tilde{w}_{i} \tilde{h}_{t-j,i}^2$, $j = 0, \ldots, n - 1$.
3. Use $\tilde{z}_{t-j}$ to define an objective function of the type in (4). Then estimate $\hat{\mathbf{w}}_t$ minimizing with respect to $\mathbf{w}$.
4. Predict volatility at time $t+j$ as $\tilde{h}_{t+j,i}^2 = \sum_{i=1}^k \tilde{w}_{i} \tilde{h}_{t+j,i}^2$, for $j = 1, \ldots, m$ with

$$\tilde{h}_{t+j,i}^2 = f(\hat{\Theta}_{t,j}; u^{t+j-1}),$$

where $\hat{\Theta}_{t,j}$ is the vector of model $i$ coefficients estimated at time $t$, $u^{t+j-1} = \{u_1, \ldots, u_{t+j-1}\}$, for $i = 1, \ldots, k$.
5. Repeat steps 1–4 at time $n + (h + 1)m$ and go on until the end of the series is reached.

It is important to note that the combined volatility predictor $\tilde{h}_{n+1,i}^2 (A > 0)$ as well as the $k$ predictors associated to the candidate models $\tilde{h}_{n+1,i}^2$ will depend on the same set of information, which is $u^{n+1}$–1. Furthermore, the estimated model coefficients and combination weights used to generate volatility predictions for any arbitrarily chosen time point $n + A$ are also based on the same data window. It follows that we can fairly compare the forecasting performance of the combined volatility predictor with that of the single candidate models.

To initialize the procedure, the $\tilde{\mathbf{w}}_t$ are estimated using in-sample volatility estimates. As we move forward in the future, these are gradually replaced by genuine out-of-sample volatility forecasts. So, in general, the weights are estimated over a data window including, in different proportions depending on the time of prediction and on the value of $m$, in-sample, in the first part, and out-of-sample volatility predictions, in the last part. Table 2 reports the computing times and number of GMM iterations needed to estimate the combination weights over the set of in-sample volatility predictions. As we can see the computing time remains reasonably low even if it tends to increase exponentially with the value of $g$. In Fig. 6 the estimated combination weights for the case $g = 1$ have been reported (the results for other values of $g$ are available from the authors). The constrained combination scheme assigns weights equal to one to...
the TGARCH model excluding the GARCH and CGARCH models from the combination. The same situation arises for higher values $g$. A similar behavior is observed for the unconstrained predictor even if the weight assigned to the TGARCH model is now systematically lower than one (dotted horizontal line in the plot). This is expected to account for potential biases in the volatility predictions generated by the TGARCH model. In order to assess the predictive performance of the volatility estimators considered, we use the squared returns as a proxy of volatility and then refer to the following well-known loss functions: the MSE, the QLIKE, the Mean Absolute Error (MAE) and its equivalent formulation in terms of standard deviations (MAEsd)

\[
MSE = \frac{1}{n} \sum_{j=1}^{n} (r_j^2 - h_j^2)^2, \quad QLIKE = \frac{1}{n} \sum_{j=1}^{n} \left[ \log(h_j^2) - \frac{r_j^2}{h_j^2} \right],
\]

\[
MAE = \frac{1}{n} \sum_{j=1}^{n} |r_j^2 - h_j^2|, \quad MAE_{sd} = \frac{1}{n} \sum_{j=1}^{n} ||r_j| - h_j|,
\]

where $n$ is the number of periods to be forecast and $h_j^2 (j = 1, \ldots, n)$ is the series of conditional variance predictions generated by some given model. The results are reported in Table 3. Except for $uc(3)$ and $uc(5)$, which return MSE values higher than those yielded by the TGARCH model, the uncostrained combination always performs better than competing predictors.

The issue now is to assess the significance of these differences by means of a formal test. The problem of comparing different non-nested conditional heteroskedastic models has been recently addressed by Chen et al. (2006). However, in this case each of the three candidate forecasting models (TGARCH, CGARCH, GARCH) is naturally nested within the combined model. This situation rules out traditional procedures for testing the hypothesis of equal predictive accuracy such as the Diebold–Mariano test (Diebold and Mariano, 1995), as discussed by Corradi and Swanson (2004) and Clark and McCracken (2001), and motivates the need for alternative testing procedures. Here we make use of the conditional predictive ability (CPA) test (Giacomini and White, 2006). The reasons for our choice can be summarized as follows: the test (i) allows to compare non-nested as well as nested forecasting models, (ii) takes into account estimation uncertainty and (iii) is relatively easy to implement and the computing time required is negligible. The results of the test for each of the four loss functions considered have been reported in Table 4.
Table 3
Forecast loss functions evaluated for the TGARCH (TG), CGARCH (CG), GARCH (G) models versus unconstrained (uc(g)), constrained (c(g)) and equally weighted (ew) combined predictors computed assuming different sets of moment conditions (g = 1, 3, 5, 7, 10)

<table>
<thead>
<tr>
<th></th>
<th>MSE</th>
<th>QLIKE</th>
<th>MAE</th>
<th>MAEsd</th>
</tr>
</thead>
<tbody>
<tr>
<td>uc(1)</td>
<td>4.9819</td>
<td>0.8240</td>
<td>1.1812</td>
<td>0.5584</td>
</tr>
<tr>
<td>uc(3)</td>
<td>4.9905</td>
<td>0.8234</td>
<td>1.1640</td>
<td>0.5520</td>
</tr>
<tr>
<td>uc(5)</td>
<td>4.9889</td>
<td>0.8242</td>
<td>1.1695</td>
<td>0.5545</td>
</tr>
<tr>
<td>uc(7)</td>
<td>4.9846</td>
<td>0.8224</td>
<td>1.1707</td>
<td>0.5541</td>
</tr>
<tr>
<td>uc(10)</td>
<td>4.9811</td>
<td>0.8215</td>
<td>1.1724</td>
<td>0.5540</td>
</tr>
<tr>
<td>c(1)</td>
<td>4.9854</td>
<td>0.8264</td>
<td>1.1929</td>
<td>0.5634</td>
</tr>
<tr>
<td>c(3)</td>
<td>4.9854</td>
<td>0.8264</td>
<td>1.1929</td>
<td>0.5634</td>
</tr>
<tr>
<td>c(5)</td>
<td>4.9854</td>
<td>0.8264</td>
<td>1.1929</td>
<td>0.5634</td>
</tr>
<tr>
<td>c(7)</td>
<td>4.9854</td>
<td>0.8264</td>
<td>1.1929</td>
<td>0.5634</td>
</tr>
<tr>
<td>c(10)</td>
<td>4.9854</td>
<td>0.8264</td>
<td>1.1929</td>
<td>0.5634</td>
</tr>
<tr>
<td>ew</td>
<td>5.1708</td>
<td>0.8322</td>
<td>1.1967</td>
<td>0.5647</td>
</tr>
<tr>
<td>TG</td>
<td>4.9854</td>
<td>0.8264</td>
<td>1.1929</td>
<td>0.5634</td>
</tr>
<tr>
<td>CG</td>
<td>5.6341</td>
<td>0.8713</td>
<td>1.2315</td>
<td>0.5765</td>
</tr>
<tr>
<td>G</td>
<td>5.1296</td>
<td>0.8290</td>
<td>1.1875</td>
<td>0.5591</td>
</tr>
</tbody>
</table>

Table 4
P-values of the CPA test for MSE, QLIKE, MAE, MAEsd

<table>
<thead>
<tr>
<th></th>
<th>ew</th>
<th>TG</th>
<th>CG</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>uc(1)</td>
<td>0.0392</td>
<td>0.3736</td>
<td>0.0020</td>
<td>0.0286</td>
</tr>
<tr>
<td>uc(3)</td>
<td>0.0149</td>
<td>0.3854</td>
<td>0.0007</td>
<td>0.0072</td>
</tr>
<tr>
<td>uc(5)</td>
<td>0.0180</td>
<td>0.3788</td>
<td>0.0009</td>
<td>0.0102</td>
</tr>
<tr>
<td>uc(7)</td>
<td>0.0192</td>
<td>0.3761</td>
<td>0.0010</td>
<td>0.0106</td>
</tr>
<tr>
<td>uc(10)</td>
<td>0.0205</td>
<td>0.3409</td>
<td>0.0012</td>
<td>0.0119</td>
</tr>
</tbody>
</table>

| QLIKE  |        |        |        |        |
| ac(1)  | 0.0010 | 0.0003 | 0.0001 | 0.1619 |
| ac(3)  | 0.0002 | 0.0233 | 0.0000 | 0.0167 |
| ac(5)  | 0.0004 | 0.0318 | 0.0001 | 0.0573 |
| ac(7)  | 0.0001 | 0.0000 | 0.0000 | 0.0462 |
| ac(10) | 0.0002 | 0.0001 | 0.0000 | 0.0687 |

| MAE    |        |        |        |        |
| ac(1)  | 0.0767 | 0.0000 | 0.0093 | 0.2656 |
| ac(3)  | 0.0000 | 0.0000 | 0.0000 | 0.0004 |
| ac(5)  | 0.0002 | 0.0000 | 0.0003 | 0.0090 |
| ac(7)  | 0.0003 | 0.0000 | 0.0004 | 0.0149 |
| ac(10) | 0.0009 | 0.0000 | 0.0007 | 0.0326 |

| MAEsd  |        |        |        |        |
| ac(1)  | 0.0017 | 0.0000 | 0.0011 | 0.0113 |
| ac(3)  | 0.0000 | 0.0000 | 0.0000 | 0.0002 |
| ac(5)  | 0.0000 | 0.0000 | 0.0000 | 0.0030 |
| ac(7)  | 0.0000 | 0.0000 | 0.0000 | 0.0032 |
| ac(10) | 0.0000 | 0.0000 | 0.0000 | 0.0041 |

Here, the predictive accuracy of the unconstrained volatility predictors, for different values of g, is compared to that of the equally weighted combination, TGARCH, CGARCH and GARCH models, respectively. We omit considering the constrained combination since, as previously discussed, it coincides with the TGARCH model.

For the MSE loss function, it turns out that the combined predictors outperform all the other competitors except the TGARCH model for which the CPA test results do not allow to reject the null hypothesis of equal predictive accuracy.
Differently, if we refer to any of the other loss functions considered, exception made for a few cases, the CPA test suggests the (conditional) predictive ability of each of the unconstrained combined volatility predictors, measured in terms of that loss function, is significantly higher than that of the other competitors. This becomes particularly evident if we focus on more resistant loss functions such as the MAE and MAEs. However, a caveat applies in the interpretation of the results obtained for MAE and MAEs. For the MSE and QLIKE criterion, the expected loss is minimized in correspondence of the true conditional variance, while, for the MAE and MAEs, the minimum is achieved in correspondence of the conditional median of squared returns given past information. This does not mean that these loss functions are invalid or inappropriate. It just means that the user is assuming a different measure of volatility (e.g. the conditional median) and this leads him to focus on the prediction of a statistic different from the conditional variance of returns (Patton, 2006).

5. Conclusions

In volatility modelling, model identification is not an easy task due to the noisy nature of financial returns. In this framework, combining volatility forecasts from different models offers a simple and practical solution for dealing with model uncertainty avoiding the risks related to the use of a single candidate model.

The combination method which has been proposed in this paper has well-defined theoretical properties, such as consistency and asymptotic normality. Also it is computationally fast and easy to implement on a personal computer using standard software, such as Matlab. Finally, the evidence provided by simulated as well as real data suggests that our combination algorithm can be considered as an useful tool for risk management applications and, in general, financial modelling. Although, in this paper, our analysis has been focused on the one-step-ahead prediction of volatility, in many applications, such as the estimation of Value at Risk and Expected Shortfall, the focus is on the generation of sequential multi-step ahead forecasts. Sequential k-steps ahead forecast errors are, however, not guaranteed to be i.i.d. since it can be only shown that, even for optimal forecasts, they will be at most (k − 1)-dependent. So, application of our framework to multi-step ahead forecasts is potentially feasible but not immediate. Hence, we leave the investigation of this point to future research.

Acknowledgments

The authors would like to thank the editor and the three anonymous referees for their valuable comments and suggestions. The usual disclaimer applies.

References


