SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2009-56

On the Economic Evaluation of Volatility Forecasts

Valeri Voev

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

On the Economic Evaluation of Volatility Forecasts*

Valeri Voev†

Aarhus University
CREATES

Abstract

We analyze the applicability of economic criteria for volatility forecast evaluation based on unconditional measures of portfolio performance. The main theoretical finding is that such unconditional measures generally fail to rank conditional forecasts correctly due to the presence of a bias term driven by the variability of the conditional mean and portfolio weights. Simulations and a small empirical study suggest that the bias can be empirically substantial and lead to distortions in forecast evaluation. An important implication is that forecasting superiority of models using high frequency data is likely to be understated if unconditional criteria are used.

JEL classification: C32, C53, G11

Keywords: Forecast evaluation, Volatility forecasting, Portfolio optimization, Mean-variance analysis

---

*I would like to thank Olaf Posch, Christian Dahl, Andrew Patton and Peter R. Hansen for helpful comments. All errors are of course mine.

†School of Economics and Management, University of Aarhus, 8000 Aarhus C, Denmark. Phone +45-8942-1539, email: vvoev@creates.au.dk. Financial support by the Center for Research in Econometric Analysis of Time Series, CREATES, funded by the Danish National Research Foundation, is gratefully acknowledged.

1 Introduction

Recently, the issue of the evaluation of volatility forecasts has received increasing attention.[1] With the advent of high-frequency (HF) data, volatility measurement has undergone a substantial development to the extend that various techniques for obtaining consistent volatility estimates under general assumptions are now available.[2] Many studies have subsequently aimed to address the question of whether HF data can also bring benefits in terms of improved volatility forecasts. The fact that ex-post volatility is at best available as a proxy, has motivated Hansen & Lunde (2006) and Patton (2009) to derive necessary and sufficient conditions for loss functions to be robust to the presence of noise in the volatility measure. A number of papers employ so-called economic evaluation criteria which can potentially fully circumvent the need of having to know the true volatility ex-post. Recent studies include Fleming, Kirby & Ostdiek (2001, 2003), Liu (2009), Şerban, Brockwell, Lehoczky & Srivastava (2007), Clements, Doolan, Hurn & Becker (2009). Economic evaluation criteria involve comparing realized Sharpe ratios, tracking error variance, variance of hedging errors, etc. In this paper, we show that such criteria can be misleading in the sense that will be formalized in the theoretical section, referring to the inability of these criteria to rank forecasts correctly even asymptotically, as the number of observations on which the comparison is based tends to infinity. The reason for this failure can be found in the unconditional nature of the above-mentioned evaluation criteria. While volatility models forecast the next-period (or multi-period) conditional volatility, the ex-post comparison is based on an unconditional measure (say portfolio variance). The two are not the same, meaning that what these criteria compare, is not what we have forecast (although it is obviously related).

It might be argued that an economically meaningful loss function has more practical appeal, and should not necessarily lead to the same ranking as a statistical criterion, such as root mean squared forecast error. While this is a reasonable point, it requires the econometrician to pose the optimization problem correspondingly, e.g., as a multiperiod problem. If the forecast which minimizes the economic loss is the

[1] See e.g., Andersen & Bollerslev (1998), Andersen, Bollerslev, Diebold & Labys (2003), Andersen, Bollerslev, Christoffersen & Diebold (2006), among others.

[2] see, e.g., Hayashi & Yoshida (2005), Zhang, Mykland & Aït-Sahalia (2005), Zhang (2006*b*), Zhang (2006*a*), Griffin & Oomen (2006), Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*a*), Barndorff-Nielsen, Hansen, Lunde & Shephard (2008*b*), Barndorff-Nielsen, Hansen, Lunde & Shephard (2009), Jacod, Li, Mykland, Podolskij & Vetter (2009), Christensen, Kinnebrock & Podolskij (2009), Voev & Lunde (2007), Nolte & Voev (2007), among others.

one which would minimize a standard statistical loss function (which would be the case if the conditions in Hansen & Lunde (2006) are satisfied by the economic and the statistical loss function), the sole aim of switching to an economically motivated loss is to "translate" the differences in forecasting performance into more intuitive and meaningful terms, e.g., losses/gains in basis points as in Fleming, Kirby & Ostdiek (2003).

The paper is structured as follows: the next Section 2 looks at the theoretical aspect of the problem, Section 3 contains results from a simulation study, Section 4 provides an empirical application and Section 5 concludes.

2 Theory

We consider a risk-averse myopic investor who is maximizing expected utility on a period-by-period basis within a mean-variance framework, and therefore requires a forecast of future conditional volatility. Let us consider the global minimum variance portfolio (GMVP) problem at each t (any suitable period of time, e.g., a day):

$$\min_{w_{t+1}} w_{t+1}' \hat{\Sigma}_{t+1|t} w_{t+1} \quad \text{s.t.} \quad \iota' w_{t+1} = 1, \tag{1}$$

where $\hat{\Sigma}_{t+1|t}$ is a forecast of the covariance of the $n \times 1$ vector of returns r_{t+1}, Σ_{t+1}, ι is an $n \times 1$ vector of ones, and w_{t+1} is a portfolio allocation vector. The "$t+1|t$" notation is used to denote forecasts conditional on some information set $\mathcal{F}_t$. We analyze this simple optimization problem for several reasons: it is not affected by a forecast of the mean return and has an easy analytical solution. More importantly, it does not imply that our investor is necessarily a variance minimizer: a version of the two-fund separation result states that any mean-variance efficient portfolio can be constructed as a linear combination of the minimum variance and the so called "slope" portfolios. For the purposes of this paper, we distinguish between two forecasting "models": the oracle forecast which assumes that the investor possesses perfect foresight and knows the true conditional covariance matrix Σ_{t+1}, and a (weakly) dominated forecast $\hat{\Sigma}_{t+1|t}$.[3]

[3] In practice, the oracle forecast is not available. However, all results remain valid, if we assume that the models are not asymptotically equivalent, i.e., they provide different forecasts. This assumption can be violated, e.g., if we compare nested specifications which encompass the true data generating process. Clark & McCracken (2001) discusses testing for equal predictive ability with nested models. With volatility forecasting, recently the focus has been on comparing models using daily data to models using high-frequency data. These are obviously non-nested.

Note that volatility is not observable even ex-post, so in this case the oracle forecast is not available even after r_{t+1} has been realized. Of course, as mentioned above, various techniques based on high frequency (HF) data are available to consistently estimate Σ_{t+1}. Denote by w^*_{t+1} the portfolio weights based on the oracle forecast, i.e.,

$$w^*_{t+1} = \underset{w_{t+1}}{\text{argmin}}\, w'_{t+1}\Sigma_{t+1}w_{t+1} \quad \text{s.t.} \quad \iota' w_{t+1} = 1,$$

whose solution is given by $w^*_{t+1} = \frac{\Sigma^{-1}_{t+1}\iota}{\iota'\Sigma^{-1}_{t+1}\iota}$, i.e., it depends exclusively on (the forecast of) Σ_{t+1}. The dominated forecast $\hat{\Sigma}_{t+1|t}$ is any forecast which produces an allocation $\hat{w}_{t+1} \neq w^*_{t+1}$, that is necessarily such that $w^{*\prime}_{t+1}\Sigma_{t+1}w^*_{t+1} \leq \hat{w}'_{t+1}\Sigma_{t+1}\hat{w}_{t+1}$.
Let us now conjecture that our investor rebalances his portfolio for each $t = t_0, \ldots, T$ (out-of-sample period), where we implicitly assume that a historical sample of daily or higher frequency returns and possibly other variables of interest is available at t_0 (in-sample period) on which the investor bases his first forecast $\hat{\Sigma}_{t_0+1|t_0}$. As time evolves, the newly available data is incorporated in the information set $\mathcal{F}_t$. We denote the ex-post realized portfolio returns resulting from the oracle and the dominated forecast by $r^{p^*}_{t+1} \equiv w^{*\prime}_{t+1}r_{t+1}$ and $r^{\hat{p}}_{t+1} \equiv \hat{w}'_{t+1}r_{t+1}$, respectively.
Since Σ_{t+1} is only available as a proxy, standard loss functions should be applied with care (see Hansen & Lunde (2006), Patton (2009), and Laurent, Rombouts & Violante (2009)). To have a more intuitive measure of forecasting ability, "economic" evaluation criteria have been proposed in the literature, motivating the portfolio optimization problem above. For our simple example, a common way to evaluate volatility forecasts generated by various models, which avoids having to know Σ_{t+1}, would be to compare the sample variances of the realized portfolio returns over the out-of-sample period, defined as

$$\hat{\mathrm{V}}_T[r^p_{t+1}] = \frac{1}{T-t_0}\sum_{t=t_0}^{T-1}\left(r^p_{t+1} - \bar{r}^p_T\right)^2,$$

where p is a generic token ($p = p^*$ or $p = \hat{p}$, for the oracle and dominated forecast, respectively) and $\bar{r}^p_T = \frac{1}{T-t_0}\sum_{t=t_0}^{T-1} r^p_{t+1}$. Having computed the ex-post out-of-sample portfolio variances, $\hat{\mathrm{V}}_T[r^p_{t+1}]$, for the forecasting models under consideration, it is then concluded that a given model produces better volatility forecasts, if the portfolio returns it has generated have a smaller variance (possibly in some statistically significant sense). We index the variance and mean estimators by T, because we consider their asymptotic behavior as $T \to \infty$.

The question we pose, can be formulated as whether the following implication holds:

$$\begin{aligned} w^*_{t+1}, \hat{w}_{t+1} : w^{*'}_{t+1}\Sigma_{t+1}w^*_{t+1} \leq \hat{w}'_{t+1}\Sigma_{t+1}\hat{w}_{t+1}, \forall t = t_0, \ldots, T, \ &(\text{condition } \bullet) \\ \Rightarrow \operatorname*{plim}_{T\to\infty} \hat{\mathrm{V}}_T[r^{p^*}_{t+1}] \leq \operatorname*{plim}_{T\to\infty} \hat{\mathrm{V}}_T[r^{\hat{p}}_{t+1}] \text{ as } T \to \infty, \ &(\text{implication } \star) \end{aligned}$$

that is, whether a model which implies portfolio compositions with a smaller conditional variance for *every* period will also in the asymptotic (probability) limit have generated the portfolio with the smallest ex-post unconditional variance. We consider asymptotic behavior, as we want to abstract from the influence of estimation error in $\hat{\mathrm{V}}_T[r^p_{t+1}]$ and $\bar{r}^p_T$ on the results. The reason for this is that while in small samples incorrect ranking can occur due to estimation error, asymptotically a valid criterion should rank the models correctly. In the terminology of Hansen & Lunde (2006), we are interested in the *objective-bias*, while we abstract from *sampling error*. It is worth emphasizing that in Hansen & Lunde (2006) and Laurent et al. (2009) it is the use of a proxy, instead of the true volatility which is responsible for the possible inconsistency in the model ranking. In this study, it is rather the fact that the considered evaluation methods refrain from using any volatility proxy and look instead at unconditional volatility which leads to the inconsistency. For this reason, the issue we are dealing with here, does not fall into the theoretical framework of the two aforementioned papers.

Let us denote the conditional mean of r_{t+1} by μ_{t+1}, $\mathrm{E}_t[r_{t+1}] = \mu_{t+1}$. For the estimator of the unconditional variance of the returns of a generic portfolio p we have that

$$\begin{aligned} \mathrm{V}_T[r^p_{t+1}] &= \frac{1}{T-t_0}\sum_{t=t_0}^{T-1}\left(r^p_{t+1} - \bar{r}^p_T\right)^2 = \frac{1}{T-t_0}\sum_{t=t_0}^{T-1}\left(r^{p2}_{t+1} - 2r^p_{t+1}\bar{r}^p_T + \bar{r}^{p2}_T\right) \\ &= \frac{1}{T-t_0}\sum_{t=t_0}^{T-1} r^{p2}_{t+1} - 2\bar{r}^p_T\frac{1}{T-t_0}\sum_{t=t_0}^{T-1} r^p_{t+1} + \frac{1}{T-t_0}\sum_{t=t_0}^{T-1}\bar{r}^{p2}_T \\ &= \frac{1}{T-t_0}\sum_{t=t_0}^{T-1} w'_{t+1}r_{t+1}r'_{t+1}w_{t+1} - \left(\frac{1}{T-t_0}\sum_{t=t_0}^{T-1} w'_{t+1}r_{t+1}\right)^2, \end{aligned} \qquad (2)$$

since $r^{p2}_{t+1} = w'_{t+1}r_{t+1}r'_{t+1}w_{t+1}$ and $-2\bar{r}^p_T\frac{1}{T-t_0}\sum_{t=t_0}^{T-1} r^p_{t+1} + \frac{1}{T-t_0}\sum_{t=t_0}^{T-1}\bar{r}^{p2}_T = -2\bar{r}^{p2}_T +$

$\frac{T-t_0}{T-t_0}\bar{r}_T^{p2} = -\bar{r}_T^{p2}$. Using the continuous mapping theorem, it follows that[4]

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \mathrm{V}_T[r_{t+1}^p] &= \mathrm{E}[w_{t+1}' r_{t+1} r_{t+1}' w_{t+1}] - \mathrm{E}[w_{t+1}' r_{t+1}]^2 \\
&= \mathrm{E}[w_{t+1}' \mathrm{E}_t[r_{t+1} r_{t+1}'] w_{t+1}] - \mathrm{E}[w_{t+1}' \mathrm{E}_t[r_{t+1}]]^2 \\
&= \mathrm{E}[w_{t+1}' (\mu_{t+1}\mu_{t+1}' + \Sigma_{t+1}) w_{t+1}] - \mathrm{E}[w_{t+1}' \mu_{t+1}]^2 \\
&= \mathrm{E}[w_{t+1}' \mu_{t+1}\mu_{t+1}' w_{t+1}] + \mathrm{E}[w_{t+1}' \Sigma_{t+1} w_{t+1}] - \mathrm{E}[w_{t+1}' \mu_{t+1}]^2, \quad (3)
\end{aligned}$$

where we have made use of the law of iterated expectations and the fact that w_{t+1} is $\mathcal{F}_t$-measurable. Setting $v_{t+1} \equiv w_{t+1}' \mu_{t+1}$ we obtain

$$\operatorname*{plim}_{T\to\infty} \hat{\mathrm{V}}_T[r_{t+1}^p] = \mathrm{E}[w_{t+1}' \Sigma_{t+1} w_{t+1}] + \mathrm{V}[v_{t+1}]. \quad (4)$$

For the oracle and dominated forecast, (4) takes the form

$$\begin{aligned}
\operatorname*{plim}_{T\to\infty} \hat{\mathrm{V}}_T[r_{t+1}^{p^*}] &= \mathrm{E}[w_{t+1}^{*'} \Sigma_{t+1} w^*{}_{t+1}] + \mathrm{V}[v_{t+1}^*] \\
\operatorname*{plim}_{T\to\infty} \hat{\mathrm{V}}_T[r_{t+1}^{\hat{p}}] &= \mathrm{E}[\hat{w}_{t+1}' \Sigma_{t+1} \hat{w}_{t+1}] + \mathrm{V}[\hat{v}_{t+1}],
\end{aligned}$$

with $v_{t+1}^* \equiv w_{t+1}^{*'} \mu_{t+1}$ and $\hat{v}_{t+1} \equiv \hat{w}_{t+1}' \mu_{t+1}$, respectively. Condition $(\bullet)$ ensures that $\mathrm{E}[\hat{w}_{t+1}' \Sigma_{t+1} \hat{w}_{t+1}] \geq \mathrm{E}[w_{t+1}^{*'} \Sigma_{t+1} w_{t+1}^*]$, but the relation between $\mathrm{V}[\hat{v}_{t+1}]$ and $\mathrm{V}[v_{t+1}^*]$ is not determined. As expected, $(\star)$ is not satisfied, unless

$$\mathrm{E}[\hat{w}_{t+1}' \Sigma_{t+1} \hat{w}_{t+1}] - \mathrm{E}[w^{*\prime}{}_{t+1} \Sigma_{t+1} w_{t+1}^*] \geq \mathrm{V}[v_{t+1}^*] - \mathrm{V}[\hat{v}_{t+1}], \quad (5)$$

i.e., unless the difference in performance is large enough to outweigh the difference in the bias terms. If the conditional mean is constant at μ, $\mathrm{V}[v_{t+1}] = \mu' \mathrm{V}[w_{t+1}]\mu$ and the ambiguity is not resolved since then we still do not know the sign and magnitude of $\mu' \mathrm{V}[\hat{w}_{t+1}]\mu - \mu' \mathrm{V}[w_{t+1}^*]\mu$. Only if $\mu = 0$, we can be certain that $(\bullet)$ implies $(\star)$. In (5), the left-hand side is always positive by assumption, while the right-hand side can be both positive or negative. In case it is negative (the truly worse model has a larger bias), it follows that the presence of a bias makes the better model look even better, which is a "positive" distortion. Obviously, if the right-hand side is negative, the truly better model has a larger bias and might eventually fare worse according to the unconditional variance criterion. In practically relevant situations, we would not

[4]Throughout the paper, we assume that returns satisfy the mild regularity conditions required for a law of large numbers to hold, see, e.g., White (2001)

know which the truly better model is, so that we would be unable to say whether we are facing the first or second distortion.

2.1 Economic evaluation criteria involving the mean

Even though the two-fund separation result motivates the interest in the global minimum variance portfolio, often the problem (1) is deemed too simple to represent what investors really do or care about. In Fleming, Kirby & Ostdiek (2001, 2003), for example, the agent minimizes volatility subject to a target return or maximizes the expected return subject to a given level of volatility. The problem can be formulated as

$$\min_{w_{t+1}} w_{t+1}'\hat{\Sigma}_{t+1|t}w_{t+1} \quad \text{s.t.} \quad \iota' w_{t+1} = 1, \quad w_{t+1}'\hat{\mu}_{t+1|t} = \mu_p, \tag{6}$$

where $\hat{\mu}_{t+1|t}$ is a mean forecast and μ_p is the target portfolio return. The constraint $\iota' w_{t+1} = 1$ is optional and if not imposed implies that there is a risk-free asset that takes the weight $1 - \iota' w_{t+1}$ as in, e.g., Fleming, Kirby & Ostdiek (2001, 2003). If the constraint is imposed then all wealth is invested in the risky assets as in, e.g., Şerban et al. (2007). A complication which arises here is the need of a mean forecast which is usually extremely noisy and from the point of view of evaluating volatility forecasts can be regarded as a nuisance parameter. Best & Grauer (1991) succinctly summarize the problem: "When only a budget constraint is imposed on the investment problem, the analytical results indicate that an MV-efficient portfolio's weights, mean, and variance can be extremely sensitive to changes in asset means." From our perspective, this implies that whether one has a good volatility forecast might not really matter for the choice of weights. One can argue, however, that all models are treated equally since the mean forecast is kept the same (e.g., at the sample mean of the data) and only the volatility forecast eventually matters in a comparison. A concern that remains is that a noisy mean forecast can clout the results to the extend that no conclusion is possible regarding the quality of the volatility forecasts. To address this issue Fleming et al. (2003) resort to Monte Carlo (MC) simulations in which the mean forecast is drawn anew at each simulation run and the portfolio performance is evaluated across the MC runs.

In this evaluation setup the investor is interested in obtaining the best mean-variance trade-off, and so the model ranking can be based on the realized ex-post Sharpe ratio of the portfolio given by

$$SR_T^p = \frac{\bar{r}_T^p - r_f}{\sqrt{\hat{\mathrm{V}}_T[r_{t+1}^p]}}, \tag{7}$$

where r_f is the return on the risk-free asset, if there is one, or zero otherwise. If the target mean return is chosen reasonably[5] then the portfolio will likely be able to match the required return and the results concerning the Sharpe ratio, will be driven by the volatility. Thus, we expect to see the same distortions in the ranking as in the GMVP case exacerbated by the uncertain mean forecast.

2.2 Implications and Remarks

An important result we obtained is that comparing unconditional portfolio variances as a tool for gauging the forecasting ability of conditional volatility models depends crucially on the behavior of the conditional mean of the return process. While at short horizons (say intradaily or daily), $\mu_t = 0$ can be a reasonable assumption, at lower frequencies the mean is not negligible. Thus the objective-bias we derived can be substantial, if we envision investors who rebalance their portfolios relatively infrequently, which is not unreasonable if they face high transaction costs.

Another factor that can drive the magnitude of the objective-bias and remained somewhat hidden in the theoretical analysis is the number of assets in the portfolio n. The objective-bias term can be written as

$$\mathrm{V}[v_{t+1}] = \mathrm{V}\left[\sum_{i=1}^{n} w_{i,t+1}\mu_{i,t+1}\right], \tag{8}$$

where the subscript i stands for the i^{th} element of the corresponding vector. Increasing n by one implies $2n+1$ additional terms in the sum: one variance and $2n$ covariances. Generally, it is not possible to determine the sign of the sum of these additional terms and while it might be negative, increasing n further cannot make the overall bias non-positive, so that the theoretical lower bound is zero. In the other extreme, the bias has no upper bound and can become arbitrarily large.

One of the reasons that the examined forecast evaluation criteria fail is that they penalize for the variability of portfolio weights. This is easily seen if we assume that μ_t is constant, implying $\mathrm{V}[v_{t+1}] = \mu'\mathrm{V}[w_{t+1}]\mu$. If the weights are constant, which would

[5] By "reasonably", we mean that it is achievable given the data; e.g., targeting a return of 20% p.a., when the stocks in the portfolio return at most 10% would be unreasonable.

be the case if we have a constant volatility forecast, then $\mathrm{V}[v_{t+1}] = 0$. The minimum of the bias is thus achieved for a constant volatility model, which is unlikely to provide a good forecast. This has interesting implications for the findings in the literature on volatility forecasting. In Fleming et al. (2003), for example, it is documented that a GARCH-type model using on HF data-based volatility estimates performs better (in terms of an unconditional Sharpe ratio criterion) than its daily-data counterpart. It is also noted that the coefficient on the innovation term is roughly twice larger for the model using HF data. The consequence of this is that the forecasts from the HF data-based model are more variable, capturing the variation in true volatility much better, and thus the implied portfolio weights change more rapidly in order to adapt to the changes in the volatility. Given our results, it is likely that the objective-bias is larger for the HF-based model and thus the documented gains understate the true gains from employing HF data. Thinking further along these lines, we can conclude that documented economic gains resulting from the use of HF data, based on unconditional portfolio performance, are generally understated. In many studies, increased turnover is explicitly taken into account and it is examined what is the maximal level of transaction costs that a given model can absorb to still realize net gains compared to a benchmark. In this aspect, our analysis suggests that these cost levels are underestimated, i.e., HF-based models are able to deliver gains at higher levels of transaction costs, than previously found.

Interestingly, even if volatility is indeed constant, but we have variation in the conditional mean, the bias is generally positive and correct ranking of the models is not guaranteed. This reveals another source of failure, namely that the unconditional criterion involves the variability of the conditional mean, which is not a factor playing a role in the optimization problem (1) and thus should not affect the evaluation.

The question which arises given the failure of the unconditional criteria to rank conditional volatility forecasts, is what we should do instead. If one is purely interested in the ranking of the models, the answer is fairly straightforward: use a loss function, satisfying the conditions in Hansen & Lunde (2006) and Laurent et al. (2009) in conjunction with a good HF measure of realized volatility. What if we are interested in some economic measure of improved performance, however? The suggested strategy in this case would be to base the evaluation on conditional measures of portfolio performance as in Chiriac & Voev (2009) and in the simulation and empirical study in this paper. This, of course, implies that one still needs an ex-post measure of realized volatility, which the unconditional criteria do not require. To our opinion, the use of an ex-post volatility measure is unavoidable, if we want to ensure correct model ranking.

3 Simulation Study

In this simulation study, we analyze empirically the consequences of using the evaluation criteria described in the previous section on the ranking of volatility forecasting models. We simulate a 5-variate ($n = 5$) daily return series, with time varying covariance driven by a DCC (Engle (2002)) model. The mean equation is specified as

$$r_t = \mu_t + \varepsilon_t$$

where μ_t is the conditional mean and

$$\varepsilon_t = \Sigma_t^{1/2} z_t,$$

with z_t a multivariate standard normal variable. $\Sigma_t^{1/2}$ is the Cholesky decomposition of the time varying conditional variance of r_t. The conditional variance series evolve as GARCH(1,1) processes

$$\sigma_{i,t}^2 = \omega_i + \alpha\varepsilon_{i,t-1}^2 + \beta\sigma_{i,t-1}^2, \quad i = 1, \ldots, n,$$

where $\sigma_{i,t}^2$ is the i-th diagonal element of Σ_t. The conditional correlation R_t from the decomposition $D_t R_t D_t$ with $D_t = diag(\sigma_{1,t}, \ldots, \sigma_{n,t})$ is given by

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

where Q_t is an $n \times n$ symmetric and positive definite matrix given by

$$Q_t \;\; = \;\; (1 - \theta_1 - \theta_2)\bar{Q} + \theta_1 u_{t-1} u_{t-1}' + \theta_2 Q_{t-1},$$

where $u_t = D_t^{-1}\varepsilon_t$ and $\bar{Q}$ is the unconditional covariance of u_t. The parameters of the data generating process (DGP) are $(\omega', \alpha, \beta, \theta_1, \theta_2)' = (0.1, 0.15, 0.2, 0.25, 0.3, 0.05, 0.9, 0.05, 0.9)$ and $vech(\bar{Q}) = (1, 0.5, 0.3, 0.4, 0.1, 1, 0.2, 0.35, 0.4, 1, 0.15, 0.6, 1, 0.25, 1)'$. All correlations are positive as typical with financial data (e.g., stocks) and the population values of the unconditional volatility of $\varepsilon_{i,t}$ are 22.36, 27.39, 31.62, 35.35 and 38.73 % p.a., $i = 1, \ldots, 5$, respectively. The specification of the conditional mean function

μ_t plays a crucial role in the analysis, as it is the decisive factor in the objective-bias. Often financial returns are assumed to have a constant conditional mean which is justified from a theoretical perspective, since returns behave as semi-martingales if markets are arbitrage-free. In our simulations, we initially experiment with two values for μ_t: $\mu_t = 0$ and $\mu_t = \bar{\mu} = (0.02, 0.028, 0.04, 0.052, 0.06)'$, corresponding to expected returns of 5, 7, 10, 13 and 15 % p.a. In the third specification, we let μ_t follow a VAR(1) process:

$$\mu_t \;=\; c + \Phi r_{t-1},$$

with $c = (I - \Phi)\bar{\mu}$ where the unconditional mean $\bar{\mu} = (0.02, 0.028, 0.04, 0.052, 0.06)'$ is as in the case with non-zero constant mean. The elements of the autoregressive matrix Φ are drawn from a uniform distribution on $[-0.2, 0.2]$, allowing for a low degree of persistence.[6]

The oracle forecast $\Sigma^*_{t+1|t} = \Sigma_{t+1}$ implies portfolio weights w^*_{t+1} which minimize the objective function in (1). We compare these weights to a forecast of a misspecified DCC model with the correct unconditional correlation $\bar{Q}$ but with $\hat{\alpha} = \hat{\theta}_1 = 0.02$ and $\hat{\beta} = \hat{\theta}_2 = 0.96$. The value of $\hat{\omega}$ is set to match the unconditional variance of r_t: $\hat{\omega} = \hat{\mathrm{V}}_T[r_t](1 - \hat{\alpha} - \hat{\beta})$. The misspecified model is "smoother" than the true DGP in the sense that it produces less variable time series covariance matrices, due to a smaller α and lager β, resulting in smoother portfolio weights, denoted by $\hat{w}_{t+1}$. The reason to choose a smoother misspecified model is that the true DGP is supposed to mimic a series of covariance matrices constructed with high-frequency data. It is well known, that such series are typically much more variable then GARCH processes with small α and large β, as often obtained empirically (see Shephard & Sheppard (2009)).

In order to analyze how the variability of portfolio weights affects the evaluation in a more controlled way, we include in the comparison 10 smoothed versions of the oracle forecast given by $\tilde{\Sigma}_{t+1|t} = (1 - \lambda_s)\Sigma_{t+1} + \lambda_s\bar{\Sigma}$, $s = 1, \ldots, 10$, where $\lambda_s = s/10$ controls the smoothness and $\bar{\Sigma} = \hat{\mathrm{V}}_T[r_t]$ is the unconditional covariance of returns. The smoothed forecasts can be regarded as an application of shrinkage, whereby the forecast is shrunk towards a constant matrix.[7] The value $s = 10$ represents the extreme case of constant volatility and thus constant portfolio weights. To summarize, we

[6] We note that the existence of transaction costs and other market imperfections can lead to a small degree of return forecastability without violating absence of arbitrage.

[7] Note that while shrinkage is usually beneficial in realistic settings (see, e.g., Ledoit & Wolf (2004)), here we shrink the oracle forecast and thus it represents a perturbation.

consider 12 "models": the oracle forecast, 10 smoothed versions of it, and a forecast from a misspecified model. The results for the GMVP problem are collected in Table 1. The table is separated in three blocks corresponding to the different conditional mean specifications. In each block, the first column is the realized portfolio variance (the forecast evaluation criterion), the second column is the true model performance in terms of expected conditional variance, the third column is the term representing the objective-bias, and the last column is the sum of columns two and three which given our theoretical analysis, should equal the value in the first column, apart from MC variation. In the zero mean scenario, $\mathrm{V}[v_{t+1}]$ is zero and thus not surprisingly, the model ranking is correct. In the constant mean scenario, we see that the objective bias term is increasing as the weight variability increases, but the difference in true performance is large enough to offset this effect. As a result, while $\mathrm{V}[r^p]$ is not a valid evaluation criterion, it still ranks the oracle forecast as the best one. In the last block, we observe that the portfolio variance fails dramatically to rank the models correctly. The oracle forecast ranks 9^{th} out of 12 due to the massive objective-bias. The misspecified GARCH forecast turns out to be smooth enough and not too bad in terms of real performance to rank higher than the oracle.

	$\mu_t = 0$				$\mu_t = \bar{\mu}$				μ_t =VAR(1)			
	$\mathrm{V}[r^p]$	$\mathrm{E}[w'\Sigma w]$	$\mathrm{V}[v]$	Sum	$\mathrm{V}[r^p]$	$\mathrm{E}[w'\Sigma w]$	$\mathrm{V}[v]$	Sum	$\mathrm{V}[r^p]$	$\mathrm{E}[w'\Sigma w]$	$\mathrm{V}[v]$	Sum
p_0	15.5626	15.5651	0	15.5651	15.5626	15.5651	0.0374	15.5652	16.1075	15.5651	4.1514	16.1092
p_1	15.5664	15.5686	0	15.5686	15.5665	15.5686	0.0334	15.5686	16.0807	15.5686	4.0325	16.0824
p_2	15.5768	15.5786	0	15.5786	15.5768	15.5786	0.0295	15.5786	16.0636	15.5786	3.9237	16.0651
p_3	15.5932	15.5947	0	15.5947	15.5933	15.5947	0.0258	15.5947	16.0552	15.5947	3.8236	16.0566
p_4	15.6157	15.6168	0	15.6168	15.6157	15.6168	0.0222	15.6168	16.0551	15.6168	3.7313	16.0564
p_5	15.6441	15.6449	0	15.6449	15.6441	15.6449	0.0186	15.6450	16.0631	15.6449	3.6461	16.0642
p_6	15.6789	15.6794	0	15.6794	15.6789	15.6794	0.0150	15.6794	16.0793	15.6794	3.5678	16.0802
p_7	15.7204	15.7207	0	15.7207	15.7205	15.7207	0.0113	15.7207	16.1040	15.7207	3.4959	16.1048
p_8	15.7697	15.7698	0	15.7698	15.7698	15.7698	0.0077	15.7698	16.1381	15.7698	3.4307	16.1386
p_9	15.8280	15.8278	0	15.8278	15.8280	15.8278	0.0039	15.8278	16.1827	15.8278	3.3724	16.1831
p_{10}	15.8972	15.8966	0	15.8966	15.8972	15.8966	0.0000	15.8966	16.2399	15.8966	3.3217	16.2400
p_{11}	15.6528	15.6550	0	15.6550	15.6529	15.6550	0.0343	15.6550	16.0986	15.6549	3.7635	16.1009

Table 1: Performance results of GMVP's. $\mathrm{V}[r^p]$, $\mathrm{E}[w'\Sigma w]$ and $\mathrm{V}[v]$ are shorthand notations for $\hat{\mathrm{V}}_T[r^p_{t+1}]$, $\mathrm{E}[w'_{t+1}\Sigma_{t+1}w_{t+1}]$, and $\mathrm{V}[v_{t+1}]$, respectively, with $v_{t+1} \equiv w'_{t+1}\mu_{t+1}$. "Sum" is the sum of $\mathrm{E}[w'\Sigma w]$ and $\mathrm{V}[v]$. p_0 is the portfolio based on the oracle forecast, $p_1 - p_{10}$ are the smoothed versions for $s = 1, \ldots, 10$ (thus, p_{10} is the constant-weight portfolio), and p_{11} is the portfolio based on the misspecified DCC model. The simulations are based on $T = 800000$. All numbers are in annualized volatility terms, i.e., the reported values are $\sqrt{250 \times \text{daily value}}$. Thus, the annualized value of "Sum" is not equal to the sum of the annualized values of $\mathrm{E}[w'\Sigma w]$ and $\mathrm{V}[v]$.

In the following, we present the results for the evaluation criteria involving the mean, referring to the portfolio optimization problem defined in (6). Here, apart from a volatility forecast, the investor requires as an input a mean return forecast. We set this equal to $\bar{\mu} = \sum_{t=1}^{T} r_t$, that is the sample average over the whole sample. The reason for using the mean over the whole sample is that we want to minimize the impact of estimation error in the mean on the optimization problem.[8] The target return, μ_p, is set to the cross-sectional average of $\bar{\mu}$. Tables 2 and 3 refer to the case when portfolio is fully invested in the risky assets, i.e., imposing the constraint $\iota' w_{t+1} = 1$, while Tables 4 and 5 refer to the case when the latter constraint is not imposed and the amount $1 - \iota' w_{t+1}$ is invested in the risk free asset. In this last case we set the return on the risk-free asset to be $r_f = \mu_p - 0.02$, i.e., the risk-free asset returns 5% less on an annualized basis than the average return of the risky assets. Tables 2 and 4 are structured in the same way as Table 1, while in Tables 3 and 5 we look at the realized portfolio mean return and Sharpe ratio. The Sharpe ratio could now be used as a ranking criterion, since here we are not purely interested in minimizing the portfolio variance, but rather in an optimal mean-variance trade-off.

[8]Note that this estimates the correct conditional mean in the constant mean scenarios. As for the VAR(1) case, attempting to estimate Φ is rather hopeless, given its relatively small magnitude and would most probably lead to rather imprecise forecasts, due to the large estimation uncertainty.

	$\mu_t = 0$				$\mu_t = \bar{\mu}$				μ_t =VAR(1)			
	V$[r^p]$	E$[w'\Sigma w]$	V$[v]$	Sum	V$[r^p]$	E$[w'\Sigma w]$	V$[v]$	Sum	V$[r^p]$	E$[w'\Sigma w]$	V$[v]$	Sum
p_0	16.2426	16.2544	0	16.2544	16.5087	16.5072	0.0029	16.5072	16.6432	16.3975	2.8569	16.6446
p_1	16.2462	16.2576	0	16.2576	16.5115	16.5098	0.0026	16.5098	16.6255	16.4001	2.7348	16.6266
p_2	16.2559	16.2669	0	16.2669	16.5193	16.5174	0.0023	16.5174	16.6150	16.4077	2.6217	16.6158
p_3	16.2715	16.2823	0	16.2823	16.5319	16.5298	0.0020	16.5298	16.6111	16.4199	2.5169	16.6117
p_4	16.2931	16.3035	0	16.3035	16.5491	16.5469	0.0017	16.5469	16.6135	16.4367	2.4197	16.6138
p_5	16.3208	16.3309	0	16.3309	16.5710	16.5686	0.0015	16.5686	16.6221	16.4581	2.3301	16.6222
p_6	16.3548	16.3646	0	16.3646	16.5977	16.5951	0.0012	16.5951	16.6370	16.4842	2.2479	16.6368
p_7	16.3956	16.4053	0	16.4053	16.6296	16.6268	0.0009	16.6268	16.6583	16.5155	2.1735	16.6579
p_8	16.4442	16.4536	0	16.4536	16.6671	16.6641	0.0006	16.6641	16.6866	16.5523	2.1075	16.6859
p_9	16.5014	16.5106	0	16.5106	16.7111	16.7079	0.0003	16.7079	16.7227	16.5954	2.0510	16.7217
p_{10}	16.5689	16.5779	0	16.5779	16.7626	16.7592	0.0000	16.7592	16.7676	16.6460	2.0054	16.7663
p_{11}	16.3171	16.3291	0	16.3291	16.5833	16.5842	0.0027	16.5842	16.6621	16.4726	2.5270	16.6653

Table 2: Performance results of return targeting portfolios in the absence of a risk-free asset (weights sum up to one). V$[r^p]$, E$[w'\Sigma w]$ and V$[v]$ are shorthand notations for $\hat{\mathrm{V}}_T[r^p_{t+1}]$, E$[w'_{t+1}\Sigma_{t+1}w_{t+1}]$, and V$[v_{t+1}]$, respectively, with $v_{t+1} \equiv w'_{t+1}\mu_{t+1}$. "Sum" is the sum of E$[w'\Sigma w]$ and V$[v]$. p_0 is the portfolio based on the oracle forecast, $p_1 - p_{10}$ are the smoothed versions for $s = 1, \ldots, 10$ (thus, p_{10} is the constant-weight portfolio), and p_{11} is the portfolio based on the misspecified DCC model. The simulations are based on $T = 800000$. All numbers are in annualized volatility terms, i.e., the reported values are $\sqrt{250 \times \text{daily value}}$. Thus, the annualized value of "Sum" is not equal to the sum of the annualized values of E$[w'\Sigma w]$ and V$[v]$.

	$\mu_t = 0$			$\mu_t = \bar{\mu}$			μ_t =VAR(1)		
	E$[r^p]$	V$[r^p]$	SR^p	E$[r^p]$	V$[r^p]$	SR^p	E$[r^p]$	V$[r^p]$	SR^p
p_0	0.4396	16.2426	0.0271	10.5357	16.5087	0.6382	10.3883	16.6432	0.6242
p_1	0.4535	16.2462	0.0279	10.5387	16.5115	0.6383	10.3884	16.6255	0.6249
p_2	0.4680	16.2559	0.0288	10.5424	16.5193	0.6382	10.3895	16.6150	0.6253
p_3	0.4832	16.2715	0.0297	10.5469	16.5319	0.6380	10.3913	16.6111	0.6256
p_4	0.4991	16.2931	0.0306	10.5522	16.5491	0.6376	10.3940	16.6135	0.6256
p_5	0.5158	16.3208	0.0316	10.5583	16.5710	0.6372	10.3974	16.6221	0.6255
p_6	0.5334	16.3548	0.0326	10.5652	16.5977	0.6365	10.4015	16.6370	0.6252
p_7	0.5520	16.3956	0.0337	10.5729	16.6296	0.6358	10.4065	16.6583	0.6247
p_8	0.5716	16.4442	0.0348	10.5816	16.6671	0.6349	10.4122	16.6866	0.6240
p_9	0.5926	16.5014	0.0359	10.5913	16.7111	0.6338	10.4188	16.7227	0.6230
p_{10}	0.6149	16.5689	0.0371	10.6023	16.7626	0.6325	10.4265	16.7676	0.6218
p_{11}	0.4819	16.3171	0.0295	10.5254	16.5833	0.6347	10.3768	16.6621	0.6228

Table 3: Annualized realized ex-post means, standard deviations and Sharpe ratios of return targeting portfolios in the absence of a risk-free asset (weights sum up to one). E$[r^p]$, V$[r^p]$ and SR^p are shorthand notations for $\hat{\mathrm{E}}_T[r^p_{t+1}]$, $\hat{\mathrm{V}}_T[r^p_{t+1}]$, and SR^p_T, respectively. The simulations are based on $T = 800000$.

	$\mu_t=0$				$\mu_t=\bar{\mu}$				$\mu_t=$VAR(1)			
	V[r^p]	E[$w'\Sigma w$]	V[v]	Sum	V[r^p]	E[$w'\Sigma w$]	V[v]	Sum	V[r^p]	E[$w'\Sigma w$]	V[v]	Sum
p_0	14.0618	14.0775	0	14.0775	15.6595	15.6630	0.0030	15.6630	15.6499	15.4682	2.4215	15.6566
p_1	14.0674	14.0825	0	14.0825	15.6639	15.6669	0.0027	15.6669	15.6346	15.4721	2.2892	15.6405
p_2	14.0826	14.0971	0	14.0971	15.6757	15.6783	0.0024	15.6783	15.6294	15.4833	2.1708	15.6347
p_3	14.1074	14.1213	0	14.1213	15.6946	15.6968	0.0021	15.6968	15.6338	15.5013	2.0654	15.6383
p_4	14.1418	14.1551	0	14.1551	15.7203	15.7222	0.0018	15.7222	15.6471	15.5262	1.9725	15.6510
p_5	14.1862	14.1988	0	14.1988	15.7531	15.7545	0.0015	15.7545	15.6692	15.5579	1.8922	15.6725
p_6	14.2413	14.2532	0	14.2532	15.7930	15.7940	0.0012	15.7940	15.7003	15.5966	1.8250	15.7030
p_7	14.3080	14.3193	0	14.3193	15.8408	15.8414	0.0009	15.8414	15.7409	15.6430	1.7716	15.7430
p_8	14.3878	14.3984	0	14.3984	15.8973	15.8974	0.0006	15.8974	15.7918	15.6980	1.7335	15.7934
p_9	14.4828	14.4927	0	14.4927	15.9637	15.9635	0.0003	15.9635	15.8544	15.7627	1.7123	15.8554
p_{10}	14.5959	14.6051	0	14.6051	16.0422	16.0415	0.0000	16.0415	15.9307	15.8392	1.7105	15.9313
p_{11}	14.1746	14.1885	0	14.1885	15.7664	15.7720	0.0028	15.7720	15.7144	15.5747	2.1473	15.7221

Table 4: Performance results of return targeting portfolios in the presence of a risk-free asset (unrestricted weights). V[r^p], E[$w'\Sigma w$] and V[v] are shorthand notations for $\hat{\mathrm{V}}_T[r^p_{t+1}]$, E[$w'_{t+1}\Sigma_{t+1}w_{t+1}$], and V[$v_{t+1}$], respectively, with $v_{t+1} \equiv w'_{t+1}\mu_{t+1}$. "Sum" is the sum of E[$w'\Sigma w$] and V[$v$]. p_0 is the portfolio based on the oracle forecast, $p_1 - p_{10}$ are the smoothed versions for $s = 1, \ldots, 10$ (thus, p_{10} is the constant-weight portfolio), and p_{11} is the portfolio based on the misspecified DCC model. The simulations are based on $T = 800000$. All numbers are in annualized volatility terms, i.e., the reported values are $\sqrt{250 \times \text{daily value}}$. Thus, the annualized value of "Sum" is not equal to the sum of the annualized values of E[$w'\Sigma w$] and V[v].

	$\mu_t=0$			$\mu_t=\bar{\mu}$			$\mu_t=$VAR(1)		
	E[r^p]	V[r^p]	SR^p	E[r^p]	V[r^p]	SR^p	E[r^p]	V[r^p]	SR^p
p_0	3.2110	14.0618	0.2284	5.9843	15.6595	0.3821	6.0386	15.6499	0.3859
p_1	3.2423	14.0674	0.2305	5.9770	15.6639	0.3816	6.0281	15.6346	0.3856
p_2	3.2727	14.0826	0.2324	5.9715	15.6757	0.3809	6.0195	15.6294	0.3851
p_3	3.3024	14.1074	0.2341	5.9677	15.6946	0.3802	6.0125	15.6338	0.3846
p_4	3.3314	14.1418	0.2356	5.9655	15.7203	0.3795	6.0071	15.6471	0.3839
p_5	3.3598	14.1862	0.2368	5.9649	15.7531	0.3787	6.0031	15.6692	0.3831
p_6	3.3878	14.2413	0.2379	5.9659	15.7930	0.3778	6.0006	15.7003	0.3822
p_7	3.4154	14.3080	0.2387	5.9684	15.8408	0.3768	5.9995	15.7409	0.3811
p_8	3.4425	14.3878	0.2393	5.9725	15.8973	0.3757	5.9998	15.7918	0.3799
p_9	3.4694	14.4828	0.2396	5.9783	15.9637	0.3745	6.0015	15.8544	0.3785
p_{10}	3.4960	14.5959	0.2395	5.9858	16.0422	0.3731	6.0048	15.9307	0.3769
p_{11}	3.3158	14.1746	0.2339	5.9368	15.7664	0.3765	6.0121	15.7144	0.3826

Table 5: Annualized realized ex-post means, standard deviations and Sharpe ratios of return targeting portfolios in the presence of a risk-free asset (unrestricted weights). E[r^p], V[r^p] and SR^p are shorthand notations for $\hat{\mathrm{E}}_T[r^p_{t+1}]$, $\hat{\mathrm{V}}_T[r^p_{t+1}]$, and SR^p_T, respectively. The simulations are based on $T = 800000$.

The conclusions we draw from Table 2 are very similar to the findings we obtained for the GMVP case. In the constant mean scenarios, the minimum realized portfolio variance is obtained for the oracle forecast. Again, we emphasize that for the case $\mu_t = \bar{\mu}$ results can easily change, if we increase the value of μ, because this could lead to the objective-bias more than offsetting the difference in true performance. In the case of time-varying mean, portfolio p_3 has the smallest variance, indicating again the failure of the realized ex-post portfolio variance to consistently rank the models. If we use the Sharpe ratio to rank our models, the randomness of the realized return affects the comparison gravely: the oracle forecast is never ranked first for any of the conditional mean specifications. For the unconstrained portfolio, the Sharpe ratio results in Table 5 are surprising: the Sharpe ratio is actually highest for the oracle forecast exactly in the scenarios where the volatility fails as a comparison criterion. In the zero mean case, where the volatility ranks the model correctly, the Sharpe ratio ranks the oracle forecast last. These results reinforce the detrimental effect of including the realized portfolio return as an input in criteria supposed to rank volatility forecasts.

4 Empirical Study

To illustrate how volatility forecasting model evaluation is affected by using conditional vs. an unconditional measures of portfolio performance in a real-world setting, we use a dataset and a range of models used in Chiriac & Voev (2009). The data consists of tick-by-tick bid and ask quotes from the NYSE Trade and Quotations (TAQ) database sampled from 9:30 until 16:00 for the period 01.01.2000 – 30.07.2008 ($T = 2156$ trading days) on six highly liquid stocks: American Express Inc. (AXP), Citigroup (C), General Electric (GE), Home Depot Inc. (HD), International Business Machines (IBM) and JPMorgan Chase & Co (JPM).[9] We employ the previous-tick interpolation method, described in Dacorogna, Gençay, Müller, Olsen & Pictet (2001) and obtain 78 intraday returns by sampling every 5 minutes. The daily series of realized covariance (RC) matrices computed using the 5-minute returns and subsampling are used as an input for the HF-data based models.[10] The "benchmark" model is a

[9] We are grateful to Asger Lunde for providing us with the data.

[10] For details on the subsampled realized covariance estimator, we refer the reader to Chiriac & Voev (2009), which also contains detailed description of the implementation of the forecasting models which we very briefly sketch below. A full description of the models is not central to the issue at

DCC specification, which only requires daily returns as an input. The four HF-based models considered here are succinctly described in the following. To make use of the long memory of volatility, Chiriac & Voev (2009) propose a multivariate ARFIMA model (see Sowell (1989)), which is applied to the series of Cholesky factors of the original RC series. At each step, the forecast of the Cholesky factors is "squared" to ensure that the out-of-sample covariance forecast is positive definite. As an alternative to the ARFIMA process, the second model uses HAR dynamics (see Corsi (2009)) on the Cholesky factors. The positivity of the matrix forecast can also be guaranteed by the Wishart Autoregressive (WAR) model of Gourieroux, Jasiak & Sufana (2009), which constitutes the third model in our comparison. A HAR-type extension to the original WAR specification completes the list of HF-data based models (see Bonato, Caporin & Ranaldo (2009)). We use the following acronyms for the four HF-based models: ARFIMA, HAR, WAR, and WAR-HAR, respectively.

For the purposes of forecast evaluation, we split the data into an in-sample period (01.01.2000 – 31.12.2005, $t_0 = 1508$) and an out-of-sample period (01.01.2006 – 30.07.2008) so that the model evaluation is based on $T - t_0 = 648$ observations. Obviously, sampling error will have to be accounted for, whenever we attempt to recognize whether a given model is *significantly better* than another one. Nevertheless, as we shall see, the role of the objective bias is also rather pronounced. In order to have some kind of benchmark loss function, we also rank the models with respect to their root mean squared forecast error (RMSFE) based on the Frobenius norm[11] of the matrix forecast error

$$e_{t+1} \equiv \Sigma_{t+1} - \hat{\Sigma}_{t+1|t}, \quad t = t_0, \ldots, T-1, \tag{9}$$

where $\hat{\Sigma}_{t+1|t}$ is a forecast from one of the five volatility models and the true volatility Σ_t is proxied by the subsampled RC matrix for day t. We note that the RMSFE as a loss function is robust to the use of volatility proxy in the place of the true unobservable volatility. In order to have a notion of statistically significant differences in performance we rely on a recent methodology proposed by Hansen, Lunde & Nason (2009), the model confidence set (MCS), which determines the set of models containing the best one given a user-defined loss function and a level of confidence. Using the MCS method[12], we can attempt to distinguish whether a given model (or models)

hand here and would necessarily involve replications of large portions of the above mentioned study.

[11] The Frobenius norm of a real $m \times n$ matrix A is defined as $||A|| = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$.

[12] In terms of implementation, we use the Ox package MulCom v1.00 provided by the authors.

significantly outperforms the rest given a RMSFE or a GMVP loss criterion. The results are summarized in Table 6. Unfortunately, the true conditional mean of the return series is unknown, so we cannot report the value of the bias term. Considering the RMSFE criterion, the DCC is outperformed substantially by the HF-based models, which is most likely due to the fact that it uses much less data, but could also be attributed to its lack of long memory dynamics. The ARFIMA model is the only model in the 95% MCS, although it seems to only marginally improve over the HAR specification. The reasons for the differences in performance in the models is not something we want to address here and is discussed at length in Chiriac & Voev (2009). Focusing on the last two columns of the table, we recognize that using a conditional portfolio variance criterion, the ARFIMA model is again the only model in the 95% MCS. In comparison with the DCC, the ARFIMA model enables the investor to construct a portfolio with a 0.7 % lower volatility on average. According to the unconditional criterion, this difference is less than 0.05 % due to the presence of objective-bias. A further drawback of the unconditional criteria in general, is that they are not amenable to standard testing procedures for significance as there is no underlying series of forecast errors. Thus, the MCS approach, or any other test for equal or superior predictive ability, is not applicable.

It is instructive to compare the results of this small empirical study to the findings in Fleming et al. (2003) related to volatility timing strategies (Table 4, Panel A in their paper). In their study, the HF data-based model outperforms the model using daily data in terms of unconditional Sharpe ratio, and still does so after accounting for the randomness in the mean return.

Model	RMSFE	$\mathrm{V}[r^p]$	$\mathrm{E}[w'\Sigma w]$
DCC	5.1991	13.9740	13.2236
ARFIMA	3.8971*	13.9282	12.5221*
HAR	3.9398	13.9284	12.5299
WAR	4.9904	14.6427	12.8018
WAR-HAR	4.5967	14.3905	12.6770

Table 6: Root mean squared forecast error and annualized unconditional volatility and mean of conditional volatility of GMVP's based on covariance forecasts from the five conditional covariance models in the first column. $\mathrm{V}[r^p]$ and $\mathrm{E}[w'\Sigma w]$ are shorthand notations for $\hat{\mathrm{V}}_T[r^p_{t+1}]$ and $\mathrm{E}[w'_{t+1}\Sigma_{t+1}w_{t+1}]$, respectively. An asterisk (*) signifies that the model belongs to the 95% MCS of Hansen et al. (2009).

Comparing the unconditional volatilities of portfolios generated by both models, however, reveals that they are indistinguishable at least up to one decimal point which is the precision of the reported results. Given our analysis, we believe that not only is the HF-data based model better, but it is potentially much better than found, which would imply considerably higher fees that investors would be willing to pay to switch from the daily to the HF model.

5 Conclusion

In this paper, we study the suitability of so-called economic criteria for evaluation of volatility forecasts. The common feature of these criteria is that they are based on the unconditional performance (volatility, mean, tracking error, etc.) of some optimally chosen portfolios for which a forecast of the conditional covariance matrix of the components is needed.

We show that generally such evaluation approaches do not ensure correct ranking, due to their unconditional nature, which generates a source of objective-bias. This bias is positively related to the variability of implied portfolio weights and conditional mean of returns. The consequence is that a truly better conditional covariance forecast implying more variable portfolio compositions can be ranked as inferior to a worse forecast, if the penalty (bias) is large enough to outweigh the difference in true performance.

Our theoretical results and simulations indicate that the behavior of the conditional mean is a key driver of the objective-bias. Furthermore, we show that using more sophisticated decision problems involving the mean return adds an additional source of distortion in the evaluation, which is related to the sensitivity of mean-variance efficient portfolios to changes in the mean of the assets.

Finally, the empirical study shows that the presence of objective-bias can significantly affect the evaluation of volatility forecasting models in a realistic situation. Furthermore, the fact that the unconditional evaluation is only based on a single number and not on a series of forecast errors, implies that statistical tests for equal or superior predictive ability are not applicable.

References

Andersen, T., Bollerslev, T., Christoffersen, P. F. & Diebold, F. X. (2006), Volatility forecasting, *in* G. Elliott, C. Granger & A. Timmermann, eds, 'Handbook of Economic Forecasting', 1 edn, Elsevier, chapter 15.

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review* **39**, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica* **71**, 579–625.

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (2008*a*), 'Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise', *Econometrica* **76**, 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (2008*b*), Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. CREATES Working Paper 2008-63.

Barndorff-Nielsen, O. E., Hansen, P., Lunde, A. & Shephard, N. (2009), 'Realized kernels in practice: Trades and quotes', *Econometrics Journal* . forthcoming.

Best, M. & Grauer, R. (1991), 'Sensitivity analysis for mean-variance portfolio problems', *Management Science* **37**, 980–989.

Bonato, M., Caporin, M. & Ranaldo, A. (2009), Forecasting realized (co)variances with a block wishart autoregressive model. Working paper.

Chiriac, R. & Voev, V. (2009), 'Modelling and forecasting multivariate realized volatility', *Journal of Applied Econometrics* . forthcoming.

Christensen, K., Kinnebrock, S. & Podolskij, M. (2009), Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. CREATES Research Paper 2009-45, Aarhus University.

Clark, T. E. & McCracken, M. W. (2001), 'Tests of equal forecast accuracy and encompassing for nested models', *Journal of Econometrics* **105**(1), 85 – 110.

Clements, A., Doolan, M., Hurn, S. & Becker, R. (2009), On the efficacy of techniques for evaluating multivariate volatility forecasts, NCER Working Paper Series 41, National Centre for Econometric Research.

Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**, 174–196.

Şerban, M., Brockwell, A., Lehoczky, J. & Srivastava, S. (2007), 'Modelling the dynamic dependence structure in multivariate financial time series', *Journal of Time Series Analysis* **28**, 763–782.

Dacorogna, M. M., Gençay, R., Müller, U. A., Olsen, R. B. & Pictet, O. V. (2001), *An Introduction to High-Frequency Finance*, San Diego Academic Press.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroscedasticity models', *Journal of Business and Economic Statistics* **20**, 339–350.

Fleming, J., Kirby, C. & Ostdiek, B. (2001), 'The economic value of volatility timing', *Journal of Finance* **56**(1), 329–352.

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

Gourieroux, C., Jasiak, J. & Sufana, R. (2009), 'The wishart autoregressive process of multivariate stochastic volatility', *Journal of Econometrics* **150**, 167–181.

Griffin, J. E. & Oomen, R. C. A. (2006), Covariance measurement in the presence of non-synchronous trading and market microstructure noise. Working Paper, University of Warwick.

Hansen, P. R. & Lunde, A. (2006), 'Consistent ranking of volatility models', *Journal of Econometrics* **131**, 97–121.

Hansen, P. R., Lunde, A. & Nason, J. M. (2009), The model confidence set. Working Paper, Aarhus University.

Hayashi, T. & Yoshida, N. (2005), 'On covariance estimation of non-synchronously observed diffusion processes', *Bernoulli* **11**, 359–379.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009), 'Microstructure noise in the continuous case: The pre-averaging approach', *Stochastic Processes and their Applications* **119**(7), 2249 – 2276.

Laurent, S., Rombouts, J. & Violante, F. (2009), Consistent ranking of multivariate volatility models. ECORE Discussion Paper 2009/15.

Ledoit, O. & Wolf, M. (2004), 'Honey, I shrunk the sample covariance matrix', *Journal of Portfolio Management* **31**, 110–119.

Liu, Q. (2009), 'On portfolio optimization: How and when do we benefit from high-frequency data?', *Journal of Applied Econometrics* **24**, 560–582.

Nolte, I. & Voev, V. (2007), Estimating high-frequency based (co-) variances: A unified approach. Working paper, University of Konstanz.

Patton, A. (2009), 'Volatility forecast comparison using imperfect volatility proxies', *Journal of Econometrics* . forthcoming.

Shephard, N. & Sheppard, K. (2009), Realising the future: Forecasting with high frequency based volatility (HEAVY) models. Oxford-Man Institute, University of Oxford.

Sowell, F. (1989), Maximum likelihood estimation of fractionally integrated time series models. Carnegie Mellon University.

Voev, V. & Lunde, A. (2007), 'Integrated covariance estimation using high-frequency data in the presence of noise', *Journal of Financial Econometrics* **5**, 68–104.

White, H. (2001), *Asymptotic Theory for Econometricians*, Academic Press, San Diego.

Zhang, L. (2006*a*), 'Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach', *Bernoulli* **12**, 1019–1043.

Zhang, L. (2006*b*), Estimating covariation: Epps effect and microstructure noise. Working Paper.

Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005), 'A tale of two time scales: Determining integrated volatility with noisy high frequency data', *Journal of the American Statistical Association* **100**, 1394–1411.

Research Papers 2009

CREATES
Center for Research in Econometric Analysis of Time Series

2009-43: Lasse Bork, Hans Dewachter and Romain Houssa: Identification of Macroeconomic Factors in Large Panels

2009-44: Dennis Kristensen: Semiparametric Modelling and Estimation: A Selective Overview

2009-45: Kim Christensen, Silja Kinnebrock and Mark Podolskij: Pre-averaging estimators of the ex-post covariance matrix

2009-46: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Robust Data-Driven Inference for Density-Weighted Average Derivatives

2009-47: Mark Podolskij and Mathias Vetter: Understanding limit theorems for semimartingales: a short survey

2009-48: Isabel Casas and Irene Gijbels: Unstable volatility functions: the break preserving local linear estimator

2009-49: Torben G. Andersen and Viktor Todorov: Realized Volatility and Multipower Variation

2009-50: Robinson Kruse, Michael Frömmel, Lukas Menkhoff and Philipp Sibbertsen: What do we know about real exchange rate non-linearities?

2009-51: Tue Gørgens, Christopher L. Skeels and Allan H. Würtz: Efficient Estimation of Non-Linear Dynamic Panel Data Models with Application to Smooth Transition Models

2009-52: Torben G. Andersen, Dobrislav Dobrev and Ernst Schaumburg: Jump-Robust Volatility Estimation using Nearest Neighbor Truncation

2009-53: Florian Heinen, Philipp Sibbertsen and Robinson Kruse: Forecasting long memory time series under a break in persistence

2009-54: Tue Gørgens and Allan Würtz: Testing a parametric function against a nonparametric alternative in IV and GMM settings

2009-55: Michael Jansson and Morten Ørregaard Nielsen: Nearly Efficient Likelihood Ratio Tests for Seasonal Unit Roots

2009-56: Valeri Voev: On the Economic Evaluation of Volatility Forecasts