C.I.R.P.É.E

Centre Interuniversitaire sur le Risque,
les Politiques Économiques et l'Emploi

Cahier de recherche/Working Paper **09-48**

On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models

(*previously circulating as* **Consistent Ranking of Multivariate Volatility Models**)

Sébastien Laurent

Jeroen V.K. Rombouts

Francesco Violante

Novembre/November 2009

Laurent: Maastricht University, The Netherlands; Université Catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium. Address : Department of Quantitative Economics, Maastricht University, School of Business and Economics, P.O. Box 616, 6200 MD, The Netherlands. Tel.: +31 43 3883843; Fax: +31 43 3884874
s.laurent@maastrichtuniversity.nl

Rombouts: Institute of Applied Economics, HEC Montréal, CIRANO, CIRPÉE; Université Catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium. Address : 3000 Côte Sainte-Catherine, Montréal (QC) Canada H3T 2A7. Tel.: +1 514 3406466; Fax: +1 514 3406469
jeroen.rombouts@hec.ca

Violante: Université de Namur, CeReFim; Université Catholique de Louvain, CORE, B-1348, Louvain-la-Neuve, Belgium. Address : FUNDP-Namur, Département des Sciences Économiques, Rempart de la Vierge, 8, B-5000, Namur. Tel. +32 81 724810; Fax: +32 81 724840
francesco.violante@fundp.ac.be

We would like to thank Luc Bauwens, Raouf Boucekkine, Lars Stentoft, Mohammed Bouaddi, Giovanni Motta, the participants to the 2nd International Workshop of the ERCIM Working Group on Computing and Statistics, the 64th European Meeting of the Econometric Society, the 26th annual meeting of the Canadian Econometric Study Group and the 9th Journées du CIRPÉE for their helpful comments. Financial support from the CREF-HEC Montreal and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State Prime Minister's Office, science policy programming, is gratefully acknowledged. The usual disclaimer applies.

Abstract:
A large number of parameterizations have been proposed to model conditional variance dynamics in a multivariate framework. However, little is known about the ranking of multivariate volatility models in terms of their forecasting ability. The ranking of multivariate volatility models is inherently problematic because it requires the use of a proxy for the unobservable volatility matrix and this substitution may severely affect the ranking. We address this issue by investigating the properties of the ranking with respect to alternative statistical loss functions used to evaluate model performances. We provide conditions on the functional form of the loss function that ensure the proxy-based ranking to be consistent for the true one – i.e., the ranking that would be obtained if the true variance matrix was observable. We identify a large set of loss functions that yield a consistent ranking. In a simulation study, we sample data from a continuous time multivariate diffusion process and compare the ordering delivered by both consistent and inconsistent loss functions. We further discuss the sensitivity of the ranking to the quality of the proxy and the degree of similarity between models. An application to three foreign exchange rates, where we compare the forecasting performance of 16 multivariate GARCH specifications, is provided.

Keywords: Volatility, Multivariate GARCH, Matrix norm, Loss function, Model confidence set

JEL Classification: C10, C32, C51, C52, C53, G10

1 Introduction

A special feature of economic forecasting compared to general economic modeling is that we can measure a model's performance by comparing its forecasts to the outcomes when they become available. Generally, several forecasting models are available for the same variable and forecasting performances are evaluated by means of a loss function. Elliott and Timmermann (2008) provide an excellent survey on the state of the art of forecasting in economics. Details on volatility and correlation forecasting can be found in Andersen, Bollerslev, Christoffersen, and Diebold (2006).

The evaluation of the forecasting performance of volatility models raises the problem that the variable of interest (i.e., volatility) is unobservable and therefore the evaluation of the loss function has to rely on a proxy. However this substitution may induce a distortion with respect to the true ordering (based on the unobservable volatility). The impact on the ordering of the substitution of the true volatility by a proxy has been investigated for univariate models by Hansen and Lunde (2006a). They provide conditions, for both the loss function and the volatility proxy, under which the approximated ranking (based on the proxy) is consistent for the true ranking. Starting from this result, Patton (2009) derives necessary and sufficient conditions on the functional form of the loss function for the latter to order consistently. These results have important implications on testing procedures for superior predictive ability (see Diebold and Mariano (1995), West (1996), Clark and McCracken (2001), the reality check by White (2000) and the recent contributions of Hansen and Lunde (2005) with the superior predictive ability (SPA) test and Hansen, Lunde, and Nason (2009) with the Model Confidence Set test, among others), because when the target variable is unobservable, an unfortunate choice of the loss function may deliver unintended results even when the testing procedure is formally valid. In fact, with respect to ranking multivariate volatility model forecast performances, where conditional variance matrices are compared, little is known about the properties of the loss function. This is the first paper that addresses this issue.

In this paper, we unify and extend the results in the univariate framework to the evaluation of multivariate volatility models, that is the comparison and ordering of sequences of variance matrices. From a methodological viewpoint, we first extend to the multivariate dimension

from the CREF-HEC Montreal and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State Prime Minister's Office, science policy programming, is gratefully acknowledged. The usual disclaimer applies.

the conditions that a loss function has to satisfy to deliver the same ordering whether the evaluation is based on the true conditional variance matrix or an unbiased proxy of it. Second, similar to the univariate results in Patton (2009), we state necessary and sufficient conditions on the functional form of the loss function to order consistently in matrix and vector spaces. Third, we identify a large set of parameterizations that yield loss functions able to preserve the true ranking. Although we focus on homogeneous loss functions, unlike in the univariate case, a complete identification of the set of consistent loss functions is not available. This is because in the multivariate case there is an infinite number of possible combinations of the elements of the forecasting error matrix which yield a loss function that satisfies the necessary and sufficient conditions. We identify a number of well known vector and matrix loss functions, many of which are frequently used in practice, categorized with respect to different characteristics such as the degree of homogeneity, shape, etc. Furthermore, given the necessary and sufficient functional form, other loss functions, well suited for specific applications, can easily be derived.

Note that different loss functions may deliver different rankings depending on the characteristics of the data that each loss function is able to capture. We find that many commonly used loss functions do not satisfy the conditions for consistent ranking. However, these loss functions show desirable properties (e.g., down weighting extreme forecast errors) which can be useful in applications. We show that inconsistent loss functions are not per se inferior, and, under certain conditions they can still deliver a ranking that is insensitive under the use of a proxy. With respect to terminology, consistency of the ranking does not mean invariance of the ordering. Consistency is in fact intended only with respect to the accuracy of the proxy and for a given loss function, i.e., consistency between the true and the approximated ranking. On the other hand, invariance of the ranking means that the ordering does not change with respect to the choice of the loss function.

To make our theoretical results concrete, we focus on multivariate GARCH models to forecast the conditional variance matrix of a portfolio of financial assets. Through a comprehensive Monte Carlo simulation, we study the impact of the deterioration of the quality of the proxy on the ranking of multivariate GARCH models with respect to different choices for the loss function. The true model is a multivariate diffusion from which we compute the integrated covariance, i.e., the true daily variance matrix. The multivariate GARCH models are estimated on daily returns and used to compute 1-step ahead forecasts. The proxy of the daily variance matrix is realized covariance as defined in Andersen, Bollerslev, Diebold, and Labys (2003). The quality of this proxy is controlled through the level of aggregation of the simulated intraday data used to compute Realized Covariance. The main conclusion of our simulation is that, when ranking over a discrete set of volatility forecasts, inconsistent loss functions are not *per se* inferior to consistent ones. When the quality of the proxy is sufficiently good, consistency between the true and the approximated ranking can still be achieved. The break even point, in terms of level of accuracy of the proxy, after which the bias starts to affect the ranking, depends on the trade-off quality of the proxy vs. degree of similarity between models. That is, the closer the forecast error matrices, the higher the accuracy of the proxy needed to correctly discriminate between competing models.

We illustrate our findings using three exchange rates (Euro, UK pound and Japanese yen against US dollar). We consider 16 multivariate GARCH specifications which are frequently used in practice. The advantage of choosing a consistent loss function to evaluate model performances is striking. The ranking based on an inconsistent loss function, together with an uninformative proxy, is found to be severely biased. In fact, as the quality of the proxy deteriorates inferior models emerge and outperform models which are otherwise preferred when the comparison is based on a more accurate proxy.

The rest of the paper is organized as follows. Section 2 develops conditions for consistent ranking and derives the admissible functional form of the loss function. We discuss how to build a class of consistent loss functions and give remarks on inconsistent loss functions. Section 3 provides first a brief overview of the multivariate GARCH specifications considered in this paper and second, it introduces the realized covariance, used as a proxy for the unobserved conditional variance matrix. A detailed simulation study in Section 4 investigates the robustness of the ranking subject to consistent and inconsistent loss functions with respect to the level of accuracy of the proxy. The empirical application to three exchange rates is presented in Section 5. Section 6 concludes and discusses directions for further research. All proofs are provided in Appendix A. Supporting examples are given in Appendix B.

2 Consistent ranking and distance metrics

As explained in Andersen, Bollerslev, Christoffersen, and Diebold (2006), the problem when comparing and ranking forecasting performance of volatility models is that the true conditional variance is unobservable so that a proxy for it is required. Let us define the true, or underlying, ordering between volatility models as the ranking implied by a loss function, evaluated with respect to the unobservable conditional variance. The substitution of the latter by a proxy may introduce, because of its randomness, a ranking of volatility models that differs from the true one. Hansen and Lunde (2006a) provide a theoretical framework for the analysis of the ordering of stochastic sequences and identify conditions that a loss function has to satisfy to deliver an ordering consistent with the true ranking when a proxy for the conditional variance is used. Patton (2009) derives necessary and sufficient conditions on the functional form of the loss function for the latter to order consistently. In particular, he finds that the necessary and sufficient functional form relates to the linear exponential family of objective functions (see Gourieroux and Monfort (1995) for details).

In this section, we extend and unify these results to the case of multivariate volatility models, which requires the comparison and ordering of sequences of variance matrices. In the following subsections, we first set the notation, working assumptions and basic definitions and, as an example, we introduce a set of loss functions commonly used in a multivariate volatility context. Second, we discuss the conditions for a loss function to give a consistent ranking. Third, we characterize the functional form of a consistent loss function. Fourth, we illustrate how consistent loss functions can be constructed in practice.

2.1 Notation and definitions

We first fix the notation and make explicit what we mean by a well defined loss function and by consistent ranking. For N time series at time t we denote $R_{++}^{N\times N}$ the space of $N \times N$ positive definite matrices and $\dot{H} \subset R_{++}^{N\times N}$ a compact subset of $R_{++}^{N\times N}$. $\dot{H}$ represents the set of candidate models with typical element indexed by m, $H_{m,t}$ such that $H_{m,t} \in \dot{H}$. R_+ denotes the positive part of the real line. We define $L(\cdot,\cdot)$ an integrable loss function $L : R_{++}^{N\times N} \times \dot{H} \rightarrow R_+$ such that $L(\Sigma_t, H_{m,t})$ is the loss evaluated using the true but unobservable conditional variance matrix Σ_t with respect to model m. We refer to the ordering based on the expected loss, $E[L(\Sigma_t, H_{m,t})]$ as the true ordering. Similarly, $L(\hat{\Sigma}_t, H_{m,t})$ is the loss evaluated using $\hat{\Sigma}_t$, a proxy of Σ_t, and $E[L(\hat{\Sigma}_t, H_{m,t})]$ determines the approximated ranking. When needed, we also refer to the empirical ranking as the one based on the sample evaluation of $L(\hat{\Sigma}_t, H_{m,t})$, i.e., $T^{-1}\sum_t L(\hat{\Sigma}_t, H_t)$, where T is the length of the forecast sample. The set, $\Im_{t-1}$ denotes the information at time $t-1$ and $E_{t-1}(\cdot) \equiv E(\cdot|\Im_{t-1})$ the conditional expectation. The elements, $\sigma_{i,j,t}$, $\hat{\sigma}_{i,j,t}$ and $h_{i,j,t}$ indexed by $i, j = 1, ..., N$, refer to the elements of the matrices Σ_t, $\hat{\Sigma}_t$, H_t respectively. Furthermore, $\sigma_{k,t}$, $\hat{\sigma}_{k,t}$ and $h_{k,t}$ are the elements, indexed by $k = 1, ..., N(N+1)/2$, of the vectors $\sigma_t = vech(\Sigma_t)$, $\hat{\sigma}_t = vech(\hat{\Sigma}_t)$ and $h_t = vech(H_t)$ respectively, where $vech(\cdot)$ is the operator that stacks the lower triangular portion of a matrix into a vector. Finally, the vectorized difference between the true variance matrix and its proxy is denoted by $\xi_t = (\hat{\sigma}_t - \sigma_t)$.

The following assumptions ensure that the loss function $L(\cdot,\cdot)$ is able to correctly order with respect to the true variance matrix.

A1.1 $L(\cdot,\cdot)$ is continuous in $\dot{H}$ and it is uniquely minimized at H_t^* which represents the optimal forecast. If $H_t^* \in int(\dot{H})$, $L(\cdot,\cdot)$ is convex in $\dot{H}$.

A1.2 $L(\cdot,\cdot)$ is such that the optimal forecast equals the true conditional variance Σ_t,

$$H_t^* = \underset{H_t \in \dot{H}}{\arg\min} L(\Sigma_t, H_t) \Leftrightarrow H_t^* = \Sigma_t. \tag{1}$$

A1.3 $L(\Sigma_t, H_t) = 0 \Leftrightarrow H_t = \Sigma_t$, i.e., the loss function yields zero loss when $H_t^* = \Sigma_t$.

Definition 1 *Under assumptions A1.1 to A1.3, the loss function is well defined.*

The notion of consistency of ranking is defined as follows:

Definition 2 *Consistency between the true ranking and the ordering based on a proxy is achieved if*

$$E(L(\Sigma_t, H_{l,t})) \geq E(L(\Sigma_t, H_{m,t})) \Leftrightarrow E(L(\hat{\Sigma}_t, H_{l,t})) \geq E(L(\hat{\Sigma}_t, H_{m,t})) \tag{2}$$

is true for all $l \neq m$, where $L(\cdot,\cdot)$ is a well defined loss function in the sense of Definition 1 and $\hat{\Sigma}_t$ is some conditionally unbiased proxy of Σ_t.

By Definition 2, the ranking between any two models indexed by l and m, is consistent if it is the same whether it is based on the true conditional variance matrix or a conditionally unbiased proxy. Note that conditional unbiasedness is sufficient to ensure consistency as defined in Definition 2.

As underlined in Patton (2009) it is common practice to use several alternative measures of forecast accuracy to respond to the concern that some particular characteristics of the data may affect the result. As an example, we discuss next a selection of loss functions, listed in Table 1, which are commonly used to evaluate multivariate model performances based on forecast accuracy, or, in a more general context, to measure the distance between matrices and vectors (see Ledoit and Wolf (2003), James and Stein (1961), Bauwens, Lubrano, and Richard (1999), Koch (2007), Herdin, Czink, Ozcelik, and Bonek (2005)) and provide their classification. Although the loss function listed below are in principle well suited to measure variance forecast performances, it turns out that several are inappropriate in this setting.

Table 1: Loss functions and their classification

Matrix loss functions			
L_F	Frobenius distance	$\sum_{1\leq i,j\leq N}(\sigma_{i,j,t}-h_{i,j,t})^2$	consistent
L_S	Stein distance	$Tr[H_t^{-1}\Sigma_t]-\log\left\lvert H_t^{-1}\Sigma_t\right\rvert-N$	consistent
L_{1M}	Entrywise 1 - (matrix) norm	$\sum_{1\leq i,j\leq N}\lvert\sigma_{i,j,t}-h_{i,j,t}\rvert$	inconsistent
L_{PF}	Proportional Frobenius dist.	$Tr(\Sigma_t H_t^{-1}-I)^2$	inconsistent
$L_{LF,1}$	Log Frobenius distance (1)	$\left(\log\left\lvert\Sigma_t H_t^{-1}\right\rvert\right)^2$	inconsistent
$L_{LF,2}$	Log Frobenius distance (2)	$\left(\log\frac{Tr[\Sigma_t\Sigma_t]}{Tr[H_tH_t]}\right)^2$	inconsistent
L_{Cor}	Correlation distance	$1-\frac{Tr(\Sigma_tH_t)}{\sqrt{Tr(\Sigma_t\Sigma_t)Tr(H_tH_t)}}\in[0,1]$	inconsistent
Vector loss functions			
L_E	Euclidean distance	$\sum_{1\leq k\leq N(N+1)/2}(\sigma_{k,t}-h_{k,t})^2$	consistent
L_{WE}	Weighted Euclidean distance (with matrix of weights W)	$(\sigma_t-h_t)'W(\sigma_t-h_t)$	consistent
L_{1V}	Entrywise 1 - (vector) norm	$\sum_{1\leq k\leq N(N+1)/2}\lvert\sigma_{k,t}-h_{k,t}\rvert$	inconsistent

The first loss function, L_F, is the natural extension to matrix spaces of the mean squared error (MSE). The second, L_S, is the scale invariant loss function introduced by James and Stein (1961). L_{1M} represents the extension to matrix spaces of the mean absolute deviation (MAD) and is known as the entrywise 1 - (matrix) norm. L_{PF} is the extension of the heteroskedasticity adjusted MSE and is a quadratic loss function with the same parametric form of the Frobenius distance but which measures deviations in relative terms (see James and Stein (1961)). We refer to this loss function as proportional Frobenius distance. $L_{LF,1}$ and $L_{LF,2}$ are adaptations of the MSE logarithmic scale. In particular, the loss function in $L_{LF,2}$, alternatively defined as $\left(\log\left[\left(\sum_i \lambda^2(\Sigma_t)_i\right)\left(\sum_i \lambda^2(H_t)_i\right)^{-1}\right]\right)^2$, considers the singular values as a summary measure of a matrix. The sum of squared singular values (defined as $\sum_i \lambda^2(A)_i = Tr(AA'))$ represents the Frobenius distance of Σ_t and H_t from 0. The ratio measures the discrepancy in relative terms while the logarithm ensures that deviations are measured as factors and the squaring ensures that factors are equally weighted. We refer to this loss function as log Frobenius distance. L_{Cor} is also based on the Frobenius distance but it exploits the Cauchy-Shwartz inequality. In fact, by the inequality, the ratio is equal to one when $H_t = \Sigma_t$ and tends to 0 if H_t and Σ_t differ to a maximum extent. The ratio resembles to a correlation coefficient between the matrices H_t and Σ_t. L_E is the Euclidean distance computed on all unique elements of the forecast error matrix, while L_{WE} is a weighted version of L_E. The last function, L_{1V}, also represents an extension of the mean absolute deviation (MAD) but the distance is defined on a vector space. It differs from L_{1M} for equally weighting the unique elements of the forecast error matrix.

2.2 Conditions for consistent ranking of multivariate volatility models

We provide sufficient conditions that a loss function has to satisfy to deliver the same ordering whether the evaluation is based on the true conditional variance matrix or a proxy. To make the exposition easier, we can redefine without loss of generality the function $L(\cdot,\cdot)$ from the space $R_{++}^{N\times N} \times \dot{H}$ to R_+ as a function from $R^{N(N+1)/2} \times \dot{\mathcal{H}} \to R_+$, with $vech(H_{m,t}) \in \dot{\mathcal{H}}$ and $\dot{\mathcal{H}} \subset R^{N(N+1)/2}$, of all unique elements of the matrices Σ_t and H_t since these are variance matrices and therefore symmetric. This simplification allows to ignore $N(N-1)/2$ redundant first order conditions in the minimization problem defined in (1). We make use of the following assumptions:

A2.1 $L(\Sigma_t, H_t)$ and $L(\hat{\Sigma}_t, H_t)$ have the same parametric form $\forall H_t \in \dot{H}$ so that uncertainty depends only on $\hat{\Sigma}_t$.

A2.2 Σ_t and H_t are $\Im_{t-1}$ measurable.

A2.3 $L(\cdot,\cdot)$ is twice continuously differentiable with respect to $\hat{\sigma}_t$ and h_t.

A2.4 $\xi_t = (\hat{\sigma}_t - \sigma_t)$ is a vector martingale difference sequence with respect to $\Im_t$ with finite conditional variance matrix $V_t = E_{t-1}[\xi_t \xi_t']$.

Proposition 1 states a sufficient condition on the loss function to ensure consistent ranking.

Proposition 1 *Under assumptions A2.1 to A2.4, a well defined loss function in the sense of Definition 1 with* $\frac{\partial^2 L(\Sigma_t,H_t)}{\partial \sigma_{l,t} \partial \sigma_{m,t}}$ *finite and independent of* H_t $\forall l, m = 1, ..., N(N+1)/2$ *is consistent in the sense of Definition 2.*

The proof is given in Appendix A. Proposition 1 applies for any conditionally unbiased proxy independently of its level of accuracy. The difference between the true and the approximated ordering which is likely to occur whenever Proposition 1 is violated, is denoted as the objective bias. The bias must not be confused with sampling variability, that is the distortion between the approximated and the empirical ranking. In fact, while the latter tend to disappear asymptotically (i.e., $T^{-1}\sum_t L(\hat{\Sigma}_t, H_t) \xrightarrow{p} E\left[L(\hat{\Sigma}_t, H_t)\right]$ under ergodic stationarity of $E\left[L(\hat{\Sigma}_t, H_t)\right]$), the presence of the objective bias may induce the sample evaluation to be inconsistent for the true one irrespectively of the sample size. Note that, from the set of loss functions given in Table 1, it is straightforward to show that only L_F, L_S, L_E and L_{WE} satisfy Proposition 1.

We can further discuss the implications of Proposition 1 and elaborate on the case when Proposition 1 is violated. We show that the bias between the true and the approximated ranking depends on the accuracy of the proxy for the variance matrix: the presence of noise in the volatility proxy introduces a distortion in the approximated ordering, which tends to disappears when the accuracy of the proxy increases. More formally, consider a sequence of volatility proxies $\hat{\Sigma}_t^{(s)}$ indexed by s and denote $H_t^{*(s)}$ such that

$$H_t^{*(s)} = \underset{H_t \in int\dot{H}}{\arg\min} E_{t-1}[L(\hat{\Sigma}_t^{(s)}, H_t)]. \tag{3}$$

Furthermore, we need the following additional assumption for the next proposition:

A2.5 The volatility proxy satisfies $E_{t-1}[\xi_t^{(s)}] = 0$ $\forall s$ and $V_t^{(s)} = E_{t-1}[\xi_t^{(s)} \xi_t^{(s)\prime}] \xrightarrow{p} 0$ as $s \to \infty$.

Proposition 2 *Under assumptions A2.1 to A2.5, for a well defined loss function in the sense of Definition 1, it holds:*

i) If $\frac{\partial^3 L(\Sigma_t, H_t)}{\partial \sigma_t \partial \sigma_t' \partial h_{k,t}} = 0 \; \forall k$*, then* $H_t^{*(s)} = \Sigma_t \; \forall s$,

ii) If $\frac{\partial^3 L(\Sigma_t, H_t)}{\partial \sigma_t \partial \sigma_t' \partial h_{k,t}} \neq 0$ *for some* k*, then* $H_t^{*(s)} \xrightarrow{p} \Sigma_t$ *as* $s \to \infty$.

The proof is given in Appendix A. The first statement states that, under Proposition 1, the optimal forecast is the conditional variance, and consistency is achieved regardless of the quality of the proxy. The second result in Proposition 2 shows that the distortion introduced in the ordering when using an inconsistent loss function tends to disappear as the quality of the proxy improves. Therefore, when ordering over a discrete set of models, a loss function that violates Proposition 1 may still deliver a ranking consistent to the one implied by the true conditional variance matrix, if a sufficiently accurate proxy is used in the evaluation. In other words, when the variance of the proxy is small with respect to discrepancy between any two models, the distortion induced by the proxy becomes negligible, leaving the ordering unaffected. In the simulation study in Section 4, we further investigate this issue and in particular investigate the relationship between the accuracy of the proxy (i.e., the variability of the proxy) and the degree of similarity between model performances (i.e., how close performances are). However, in practice, it may be difficult to determine ex-ante the degree of accuracy of a proxy. Since the trade off accuracy vs. similarity is difficult to quantify ex-ante, model comparison and selection based on inconsistent loss function becomes unreliable and may lead to undesired results. The empirical application in Section 5 reveals that a sufficiently accurate proxy may not be available.

2.3 Functional form of the consistent loss function

In the univariate framework, Patton (2009) identifies necessary and sufficient conditions on the functional form of the loss function to ensure consistency between the true ranking and the one based on a proxy for the variance. The set of consistent loss functions relate to the class of linear exponential densities of Gourieroux, Monfort, and Trognon (1984) and partially coincides with the subset of homogeneous loss functions associated with the most important linear exponential densities. In fact, the family of loss functions with degree of homogeneity equal to zero, one and two defined in Patton (2009), can be alternatively derived from the objective functions corresponding to the Gaussian, Poisson and Gamma densities respectively (see Gourieroux and Monfort (1995) for more details).

We propose necessary and sufficient conditions on the functional form of the loss function defined such that it is well suited to measure distances in matrix and vector spaces. Although, unlike in the univariate case, a complete identification of the set of consistent loss functions is not feasible, we are able to identify a large set of parameterizations which yield consistent loss functions. We show that several well known vector and matrix distance functions also belong to this set.

In order to proceed, we need the following assumptions:

A3.1 $\hat{\Sigma}_t|\Im_{t-1} \sim F_t \in F$ the set of absolutely continuous distribution functions of $R^{N\times N}_{++}$;

A3.2 $\exists H^*_t \in int(\dot{H})$ such that $H^*_t = E_{t-1}(\hat{\Sigma}_t)$;

A3.3 $E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right] < \infty$ for some $H \in \dot{H}$, $\left|E_{t-1}\left[\frac{\partial L(\hat{\Sigma}_t,H_t)}{\partial h_t}\Big|_{H_t=\Sigma_t}\right]\right| < \infty$ and $\left|E_{t-1}\left[\frac{\partial L(\hat{\Sigma}_t,H_t)}{\partial h_t \partial h_t'}\Big|_{H_t=\Sigma_t}\right]\right| < \infty$ for all t where the last two inequalities hold elementwise.

Note that A3.2 follows directly from A1.2 and A2.4 because $H^*_t \in int(\dot{H})$ implies $H^*_t = \Sigma_t$ by A1.2 while $E_{t-1}(\hat{\Sigma}_t) = \Sigma_t$ results from A2.4. Assumption A3.4 allows to interchange differentiation and expectation, see L'Ecuyer (1990) and L'Ecuyer (1995) for details.

Proposition 3 *Under assumptions A2.3, A2.4 and A3.1 to A3.3 a well defined loss function, in the sense of Definition 1, is consistent in the sense of Definition 2 if and only if it takes the form*

$$L(\hat{\Sigma}_t, H_t) = \tilde{C}(H_t) - \tilde{C}(\hat{\Sigma}_t) + C(H_t)'vech(\hat{\Sigma}_t - H_t), \tag{4}$$

where $\tilde{C}(\cdot)$ is a scalar valued function from the space of $N \times N$ positive definite matrices to R, three times continuously differentiable with

$$\begin{aligned} C(H_t) &= \nabla\tilde{C}(H_t) = \begin{bmatrix} \frac{\partial\tilde{C}(H_t)}{\partial h_{1,t}} \\ \vdots \\ \frac{\partial\tilde{C}(H_t)}{\partial h_{K,t}} \end{bmatrix} \\ C'(H_t) &= \nabla^2\tilde{C}(H_t) = \begin{bmatrix} \frac{\partial\tilde{C}(H_t)}{\partial h_{1,t}\partial h_{1,t}} & \cdots & \frac{\partial\tilde{C}(H_t)}{\partial h_{1,t}\partial h_{K,t}} \\ \vdots & \ddots & \\ \frac{\partial\tilde{C}(H_t)}{\partial h_{K,t}\partial h_{1,t}} & & \frac{\partial\tilde{C}(H_t)}{\partial h_{K,t}\partial h_{K,t}} \end{bmatrix} \end{aligned}$$

the gradient and the hessian of $\tilde{C}(\cdot)$ with respect to the $K = N(N+1)/2$ unique elements of H_t and $C'(H_t)$ negative definite.

The proof is given in Appendix A. An alternative expression for the loss function defined in Proposition 3 is provided in the following corollary.

Corollary 1 *Given $\hat{\Sigma}_t$ and H_t symmetric and positive definite, then the loss function specified in (4) is isometric to*

$$L(\hat{\Sigma}_t, H_t) = \tilde{C}(H_t) - \tilde{C}(\hat{\Sigma}_t) + Tr[\bar{C}(H_t)(\hat{\Sigma}_t - H_t)], \qquad (5)$$

with $\tilde{C}(\cdot)$ defined as in Proposition 3 and

$$\bar{C}(H_t) = \begin{bmatrix} \frac{\partial \tilde{C}(H)}{\partial h_{1,1,t}} & \frac{1}{2}\frac{\partial \tilde{C}(H)}{\partial h_{1,2,t}} & \cdots & \frac{1}{2}\frac{\partial \tilde{C}(H)}{\partial h_{1,N,t}} \\ \frac{1}{2}\frac{\partial \tilde{C}(H)}{\partial h_{1,2,t}} & \frac{\partial \tilde{C}(H)}{\partial h_{2,2,t}} & & \\ \vdots & & \ddots & \\ \frac{1}{2}\frac{\partial \tilde{C}(H)}{\partial h_{1,N,t}} & & & \frac{\partial \tilde{C}(H)}{\partial h_{N,N,t}} \end{bmatrix},$$

where the derivatives are taken with respect to all N^2 elements of H_t.

The proof is provided in Appendix A. Unlike in the univariate framework, the multivariate dimension offers a large flexibility in the formulation of the loss function, see Table 1 for several parameterizations. In applied work, a careful analysis of the functional form of the loss function is a crucial preliminary step to the selection based on the specific properties of a given loss function. In this respect, it is clear that Assumption A1.2 has a central role in this setting. It is interesting to elaborate on the case when A1.2 is dropped while keeping all other assumptions in place. We can show that, relaxing Proposition 1 and 2 and Definition 2 to admit loss functions badly formulated, still yields an ordering that is insensitive to the accuracy of the proxy, i.e. *apparently consistent*. However, when A1.2 is violated, such ordering is inherently invalid because the optimal forecast does not equal the true conditional variance. To illustrate this, starting from the functional form defined in Proposition 3, we consider the following generalization of (5)

$$L(\Sigma_t, H_t) = \tilde{C}(H_t) - \tilde{C}(\Sigma_t) + f[\bar{C}(H_t)(\Sigma_t - H_t)], \qquad (6)$$

assuming that there exists a linear map $f[\cdot] : R^{N \times N} \rightarrow R$ such that $L(\Sigma_t, H_t)$ satisfies second order conditions. We summarize the implications of relaxing assumption A1.2 from Proposition 1, 2 and 3 in the following remark. The proof is given in Appendix A.

Remark 1 *Define $\succ$ the true ordering between variance matrix forecasts, i.e., based on the true conditional variance matrix, and $\succ_a$ the approximated ordering, i.e., based on the volatility proxy. Under the loss function (6), if*

i) $f[\cdot] \equiv Tr[\cdot]$ (A1.2 is satisfied): $\succ$ and $\succ_a$ are equivalent, in the sense of Definition 2, and $L(\Sigma_t, H_t)$ is such that $H_t^* = E(\hat{\Sigma}_t|\Im_{t-1}) = \Sigma_t$, i.e., the loss function is well defined in the sense of Definition 1;

ii) $f[\cdot] \not\equiv Tr[\cdot]$ (A1.2 is violated): $\succ$ and $\succ_a$ are equivalent, in the sense that the substitution of the true covariance by a proxy does not affect the ordering. However, $L(\Sigma_t, H_t)$ is such that $H_t^* \neq E(\hat{\Sigma}_t|\Im_{t-1}) = \Sigma_t$, i.e., the loss function points to an optimal forecast different from the true conditional variance independently from the use of a proxy.

The first part of Remark 1 reaffirms sufficiency and necessity of the functional form defined in Proposition 3. With respect to the second part, note that, under (6), the general idea of consistency of the ranking, i.e., equivalence of true and approximated ranking, is still valid. In fact, if $f[\cdot]$ is a linear map, then $f[\bar{C}(H_t)(\Sigma_t - H_t)]$ is linear in $\sigma_{i,j,t}$ $\forall i, j = 1, ..., N$, and therefore, similarly to what stated in Proposition 1, it holds that $\partial^2 L(\Sigma_t, H_t)/\partial\sigma_t\partial\sigma_t'\partial h_{k,t} = 0$ $\forall k = 1, ..., N(N+1)/2$. This result ensures the ranking based on the volatility proxy to be apparently consistent for the one based on the true conditional variance and insensitive to the level of accuracy of the proxy. The objective bias does not represent an issue here: in absence of assumption A1.2, the underlying ordering will differ from any valid or acceptable ordering also when based on the true conditional variance. A badly defined loss function points to an optimal forecast different from the true conditional variance.

2.4 Building a class of consistent loss functions

Endowed with the functional form defined in Proposition 2 and Corollary 1, we illustrate how to recover several consistent loss functions. These loss functions can be categorized with respect to different characteristics, for instance the degree of homogeneity, the shape, the underlying family of distributions or the functional form for $\tilde{C}(\cdot)$.

We start by investigating the case of loss functions that are based only on the forecast error, that is $L(\hat{\Sigma}_t, H_t) = L(\hat{\Sigma}_t - H_t)$. Patton (2009) shows that in the univariate case the MSE loss function is the only consistent loss function that depends solely on the forecast error. The multivariate setting offers more flexibility in the functional form for a consistent loss function based on the forecast error. The following proposition defines the family of such loss functions.

Proposition 4 *A loss function based only on the forecast error $\hat{\Sigma}_t - H_t$ that is consistent in the sense of Definition 2 is defined by the quadratic form*

$$L(\hat{\Sigma}_t, H_t) = L(\hat{\Sigma}_t - H_t) = vech(\hat{\Sigma}_t - H_t)'\hat{\Lambda}vech(\hat{\Sigma}_t - H_t) \quad (7)$$

and the loss function has the following properties:

a) homogeneous of degree 2,

b) $\nabla^2 \tilde{C}(H_t) = -2\hat{\Lambda} = \Lambda$ *is a matrix of constants defined according to Proposition 3,*

c) $\hat{\Lambda}$ *defines the weights assigned to the elements of the forecast error matrix* $\hat{\Sigma}_t - H_t$*,*

d) symmetric under 180° *rotation around the origin, i.e.* $L(\hat{\Sigma}_t - H_t) = L(H_t - \hat{\Sigma}_t)$.

The proof is given in Appendix A. Proposition 4 defines a family of quadratic loss functions which depends on the choice of the matrix of weights $\hat{\Lambda}$. Formally, the quadratic polynomial in (7) defines a family of quadric surfaces, i.e., elliptic paraboloids, and $\hat{\Lambda}$ defines the shape of the surface. As described above, the loss function in (7) corresponds to the MSE in the univariate case. In that case the loss function is symmetric, i.e., equally penalizes positive and negative forecast errors. The advantage of the multivariate case is that the notion of symmetry can be analyzed from a different aspect. Although $L(.,.)$ is symmetric under 180° rotation around the origin, a particular choice of $\hat{\Lambda}$ can still generate some types of asymmetries. In the following, we derive and discuss the properties of some well known loss functions belonging to the family defined by Proposition 4.

The simplest parameterization of $\hat{\Lambda}$ yield a loss function based on the *vech*() transformation of the forecast error matrix, i.e., a loss function based on the notion of distance on a vector space rather than a matrix space. Three examples are provided (in increasing order of generality). In Appendix B, we provide a series of analytical examples.

Example 1: Euclidean distance

From (7), by setting $\hat{\Lambda} = I_K$ we obtain a loss function of the form

$$L_E = (\hat{\sigma}_t - h_t)' I_K (\hat{\sigma}_t - h_t) = \sum_{1 \leq k \leq K} (\hat{\sigma}_{k,t} - h_{k,t})^2. \tag{8}$$

The loss function defined in (8) is the square of the Euclidean norm on the *vech*() transformation of the forecast error matrix $(\hat{\Sigma}_t - H_t)$. The matrix $\hat{\Lambda}$ is such that variances and covariances forecast errors are equally weighted. The loss function has mirror symmetry about all coordinate planes, i.e., $L((\hat{\sigma}_{1,t} - h_{1,t}), ..., (\hat{\sigma}_{k,t} - h_{k,t}), ..., (\hat{\sigma}_{K,t} - h_{K,t})) = L((\hat{\sigma}_{1,t} - h_{1,t}), ..., -(\hat{\sigma}_{k,t} - h_{k,t}), ..., (\hat{\sigma}_{K,t} - h_{K,t}))$ for all k, and is also symmetric under any rotation about the origin, e.g. $L((\hat{\sigma}_t - h_t)) = L(-(\hat{\sigma}_t - h_t))$. Equal weights also imply that L_E is a symmetric polynomial, i.e., it is invariant under any permutation of the elements of $(\hat{\sigma}_t - h_t)$, that is, $L((\hat{\sigma}_{s_1,t} - h_{s_1,t}), ..., (\hat{\sigma}_{s_K,t} - h_{s_K,t})) = L((\hat{\sigma}_{1,t} - h_{1,t}), ..., (\hat{\sigma}_{K,t} - h_{K,t}))$ for some permutation s of the subscripts $1, ..., K$. The contours of the Euclidean distance are, indeed, represented by spheres centered at the origin.

Example 2: Weighted Euclidean distance

A more flexible version of (8) is the weighted Euclidean distance, where $\hat{\Lambda}$ is defined as $\hat{\lambda}_{i,i} > 0$ and $\hat{\lambda}_{i,j} = 0$, $i, j = 1, ..., K$, that is

$$L_{WE} = (\hat{\sigma}_t - h_t)'\hat{\Lambda}(\hat{\sigma}_t - h_t) = \sum_{1\leq k\leq K} \hat{\lambda}_{k,k}(\hat{\sigma}_{k,t} - h_{k,t})^2. \tag{9}$$

This loss function allows to differently weight each variance and covariance forecast error. Also L_{WE} shows mirror symmetry about all coordinate planes and it is also symmetric under a 180° rotation about the origin. However, unlike L_E, it is not invariant to permutations of the elements of $(\hat{\sigma}_t - h_t)$, unless $\hat{\lambda}_{i,i} = c$ for all i, i.e. $L_{WE} = cL_E$, where c is a constant. The first type of symmetry implies that for *each element* of the forecast error matrix over and under predictions are equally penalized. To illustrate the second type of symmetry we provide an example: For a given absolute forecast error $|\hat{\sigma}_t - h_t|$, i.e., fixing $\hat{\sigma}_t$ and h_t, consider L_{WE} evaluated at the following points in the domain: *i)* L^i_{WE}, the loss at $(\hat{\sigma}_{k,t} - h_{k,t}) > 0\ \forall k$ (all variances and covariances are under predicted) and *ii)* L^{ii}_{WE}, the loss at $(\hat{\sigma}_t - h_t) < 0\ \forall k$ (all variances and covariances are over predicted). Then it holds that $L^i_{WE} = L^{ii}_{WE}$. Furthermore, consider L^{iii}_{WE}, the loss at $(\hat{\sigma}_{k,t} - h_{k,t}) > (<)0$, for some k (some variances/covariances are under predicted, while other are over predicted), then mirror symmetry implies $L^i_{WE} = L^{ii}_{WE} = L^{iii}_{WE}$. Finally, the lack of invariance under permutations, i.e., L_{WE} is not symmetric about the bisector planes, is induced by the unequal distribution of the weights to the elements of the forecasting errors matrix. The contours of the weighted Euclidean distance are represented by ellipsoids centered at the origin and with the axes of symmetry lying on the coordinate axes.

Example 3: Pseudo Mahalanobis distance

This loss function represents a generalization of (9). It is obtained by setting $\hat{\lambda}_{i,j} > 0$, $i, j = 1, ..., K$, that is

$$L_M = (\hat{\sigma}_t - h_t)'\hat{\Lambda}(\hat{\sigma}_t - h_t) = \sum_{1\leq k,l\leq K} \hat{\lambda}_{k,l}(\hat{\sigma}_{k,t} - h_{k,t})(\hat{\sigma}_{l,t} - h_{l,t}). \tag{10}$$

with $\hat{\Lambda}$ chosen according to Proposition 4. We call the loss functions defined in (10) pseudo Mahalanobis distance because though it has the same parametric form, unlike the Mahalanobis distance, the matrix of weights $\hat{\Lambda}$ is deterministic and does not depend on $(\hat{\sigma}_t - h_t)$.

In this case, since $\hat{\Lambda}$ is non diagonal, L_M also includes the cross product of variances and covariances forecast errors. This loss function is only symmetric under a 180° rotation about the origin. The matrix $\hat{\Lambda}$ here plays a role similar to a correlation in a multivariate symmetric distribution: positive (negative) weights imply that systematic over/under predictions are penalized less (more). To illustrate this type of symmetry, as before, consider, for a given absolute forecast error $|\hat{\sigma}_t - h_t|$, L_M evaluated at the following points in the domain: *i)* L^i_M, the loss at $(\hat{\sigma}_{k,t} - h_{k,t}) > 0\ \forall k$ (all variances and covariances are under predicted), *ii)* L^{ii}_M, the loss at $(\hat{\sigma}_t - h_t) < 0\ \forall k$ (all variances and covariances are over predicted) and *iii)* L^{iii}_{WE}, the loss at $(\hat{\sigma}_{k,t} - h_{k,t}) > (<)0$, for some k (some variances/covariances are under predicted, while other are over predicted). Then it holds that $L^i_{WE} = L^{ii}_{WE}$. However, since $\lambda_{i,j} > (<)0$, $i \neq j$, then $L^i_{WE} = L^{ii}_{WE} \neq L^{iii}_{WE}$. Furthermore, L_M is not invariant to permutations of the elements of $(\hat{\sigma}_t - h_t)$, unless $\hat{\Lambda}$ is chosen such that $L_M = cL_E$, where c is a constant. The contours of the pseudo Mahalanobis distance are represented by ellipsoids centered at the origin and with the axes of symmetry, whose direction is given by the sign of the off diagonal elements of $\hat{\Lambda}$, that are rotated with respect to the coordinate axes.

The parameterization given in Proposition 4 allows to focus on the ordering implied by a subset of elements of the forecast error matrix. As an example, consider the comparison based on correlation matrices. In this case, one may want to focus on the elements of the strictly lower diagonal portion of the forecast error matrix, i.e., the $N(N-1)/2$ correlation forecast errors.

Remark 2 (Subsets of forecast errors) *The loss function in (7) can be reparameterized using $\hat{\Lambda}$ diagonal and $diag(\hat{\Lambda}) = vech(V)$ where V is symmetric with typical element, indexed by $i, j = 1, ..., N(N+1)/2$, $v_{i,j} = 0$ if $i = j$, $v_{i,j} \neq 0$ if $i \neq j$. Although in this case $\hat{\Lambda}$ does not satisfy Proposition 4, the resulting loss function represents the weighted Euclidean distance on the vector holding the strictly lower diagonal elements of the forecast error matrix with weights $v_{i,j}$, that is*

$$L_{WE} = (\hat{\sigma}_t - h_t)'\hat{\Lambda}(\hat{\sigma}_t - h_t) = \sum_{1 \leq i < j \leq N} v_{i,j}(\hat{\sigma}_{i,j,t} - h_{i,j,t})^2. \tag{11}$$

In fact, define $lvech()$ as the operator that stacks the strictly lower triangular portion of a matrix into a vector, $\hat{\sigma}^{lw}_t = lvech(\hat{\Sigma}_t)$, $h^{lw}_t = lvech(H_t)$ and $diag(\hat{\Lambda}^{lw}) = lvech(V)$, (11) is equivalent to

$$L^{lw}_{WE} = lvech(\hat{\Sigma}_t - H_t)\hat{\Lambda}^{lw}lvech(\hat{\Sigma}_t - H_t) = \sum_{1 \leq k \leq N(N-1)/2} \hat{\lambda}^{lw}_{k,k}(\hat{\sigma}^{lw}_{k,t} - h^{lw}_{k,t})^2,$$

which is well defined according to Proposition 4.

We discuss next a particular parameterization of $\hat{\Lambda}$ which leading to a loss function known as Frobenius distance. This loss function is based on the notion of distance on a matrix space rather then a vector space. It considers the entire forecast error matrix, therefore allowing to exploit some interesting matrix properties.

Example 4: Frobenius distance

From (7), if we set $\hat{\Lambda}$ diagonal and $diag(\hat{\Lambda}) = vech(V)$ where V is symmetric with typical element, indexed by $i, j = 1, ..., N(N+1)/2$, $v_{ij} = 1$ if $i = j$, $v_{ij} = 2$ if $i \neq j$, then the resulting loss function is

$$L_F = Tr[(\hat{\Sigma}_t - H_t)'(\hat{\Sigma}_t - H_t)]. \tag{12}$$

Equation (12) represents the matrix equivalent to the MSE loss function. Alternatively, (12) can be written as

$$L_F = \sum_{1 \leq i,j \leq N} (\hat{\sigma}_{i,j,t} - h_{i,j,t})^2 = \sum_{1 \leq i \leq N} \varsigma_i(\hat{\Sigma}_t - H_t),$$

that is the sum of elementwise squared differences or equivalently the sum of the singular values, $\varsigma_i(.)$, of the forecast error matrix $(\hat{\Sigma}_t - H_t)$.

The Frobenius distance represents a special case of the weighted Euclidean distance, but the specific choice of $\hat{\Lambda}$ allows to exploit properties of symmetric matrices. Note that this loss function, by considering the entire forecast matrix, double weights the covariance forecast errors, i.e., the off diagonal elements of the forecasting error matrix. In terms of geometrical representation, it shares the same properties with L_{WE}.

In Examples 1 to 4 we denote the loss functions as a distance. Throughout the paper, we refer to the term distance to underline the different characterization, in this case the square transformation, of a loss function with respect to the underlying norm.

Remark 3 (Inconsistent loss functions:) ***Entrywise 1 - (matrix) norm.*** *The parameterization given in (8), in Example 2, closely resembles the square of the entrywise 1 - (matrix) norm, which is defined as*

$$L_{1M}^2 = \left(\sum_{1 \leq i,j \leq N} |\hat{\sigma}_{i,j,t} - h_{i,j,t}| \right)^2 .$$

but unlike the pseudo Mahalanobis distance, L_{1M}^2 is not differentiable and therefore Proposition 1 does not apply.

Frobenius norm. *The square root of (12) is a well known matrix norm called Frobenius norm, Hilbert-Schmidt norm or Schatten 2-norm and represents the matrix equivalent to the root-MSE loss function. It is straightforward to show that such loss function does not satisfy Proposition 1.*
Euclidean norm. *The square root of (8) is the Euclidean norm. Also in this case, it can be shown that Proposition 1 is violated.*

From Remark 2, it is clear that even simple transformations of a consistent loss function may cause the violation of Proposition 1. Note that, since the loss functions defined in (12) and (8) are the square of a norm, they are by definition homogeneous of degree 2.

As the class of consistent loss functions identified in Patton (2009) is related to the class of linear exponential distributions (see Gourieroux and Monfort (1995) for details), an alternative procedure to identify loss functions that belong to the family defined in Proposition 3 is to look at distributions which are defined on the support $R^{N\times N}$ or the space of $N\times N$ positive definite matrices ($R^{N\times N}_{++}$).

Remark 4 (Frobenius distance) *Consider the matrix model*

$$\hat{\Sigma}_t = H_t + \Xi,$$

where Ξ is a matrix of random errors. If $\Xi|\Im_{t-1}\sim N(H_t,\Omega,\Theta)$ then

$$P[\hat{\Sigma}_t|\Im_{t-1}] = \frac{\exp\left(-\frac{1}{2}Tr[\Omega^{-1}(\hat{\Sigma}_t - H_t)'\Theta^{-1}(\hat{\Sigma}_t - H_t)]\right)}{(2\pi)^{N^2/2}\left|\Omega\right|^{N/2}\left|\Theta\right|^{N/2}}$$

is the probability density function of $\hat{\Sigma}_t$. Since the parameter of interest is H_t, the associated objective function is the least squares loss function

$$L_F = Tr[(\hat{\Sigma}_t - H_t)'(\hat{\Sigma}_t - H_t)]. \tag{13}$$

The loss function in (13) belongs to the family of loss functions defined by (5), with $\tilde{C}(H_t) = -Tr(H_tH_t)$ and $\bar{C}(H_t) = -2H_t$.

Alternatively, if we consider the Wishart distribution we identify a loss function that is characterized by a degree of homogeneity equal to zero and that depends only on the standardized (in matrix sense) forecast error.

Example 5: Stein loss function

Assume that the conditional distribution of $\hat{\Sigma}_t$ is Wishart, with $E_{t-1}[\hat{\Sigma}_t] = H_t$, i.e. $\hat{\Sigma}_t|\Im_{t-1}\sim W_N(p^{-1}H_t, p)$. The probability density function is given by

$$P[\hat{\Sigma}_t|\Im_{t-1}] = \frac{\left|\hat{\Sigma}_t\right|^{\frac{(p-N-1)}{2}}}{2^{Np/2}p^{-1}\left|H_t\right|^{p/2}} \exp\left(-\frac{1}{2p}Tr[H_t^{-1}\hat{\Sigma}_t]\right),$$

which yields to the loss function

$$L_S = Tr[H_t^{-1}\hat{\Sigma}_t] - \log\left|H_t^{-1}\hat{\Sigma}_t\right| - N. \tag{14}$$

L_S belongs to the family defined by (5) with $\tilde{C}(H_t) = \log|H_t|$ and $\bar{C}(H_t) = H_t^{-1}$. It corresponds to the scale invariant (i.e., homogeneous of degree 0) loss function introduced by James and Stein (1961). L_S is asymmetric with respect to over/ under predictions, and, in particular, underpredictions are heavily penalized. The properties of L_S are further discussed in Appendix B using a simple example.

We have seen that even in the specific case of loss functions based only on the forecast error, the multivariate dimension allows to construct a variety of consistent loss functions sharing the same degree of homogeneity but differing in the way deviations are weighted. However, unlike the univariate case, the generalization may be computationally unfeasible when using a procedure of the type bottom-up, i.e., starting from $\nabla^2\tilde{C}(H_t) = \Lambda(H_t)$ and then integrating up to obtain the functional $\tilde{C}(H_t)$ (see proofs of Proposition 3 and 4 for details). In fact, if $L(\hat{\Sigma}_t, H_t)$ is homogeneous of degree d, then $\partial L(\hat{\Sigma}_t, H_t)/\partial h_t = \nabla^2\tilde{C}(H_t)vech(\hat{\Sigma}_t - H_t)$ is homogeneous of degree $(d-1)$, while the elements of the Hessian, $\nabla^2\tilde{C}(H_t) = \Lambda(H_t)$, are homogeneous of degree $(d-2)$. The procedure illustrated above would require to set $\Lambda(H_t)$ is such a way that its elements are *(i)* homogeneous of degree $(d-2)$, *(ii)* possibly depend on H_t and *(iii)* $\Lambda(H_t) = \nabla^2\tilde{C}(H_t)$ satisfies Proposition 3. Such generalization (e.g., L_S to a family scale invariant loss functions) is computationally cumbersome and the resulting loss function is likely to be difficult to interpret.

Alternatively, we propose, starting from (4) or (5), to choose ex-ante some functional form for $\tilde{C}(\cdot)$, possibly homogeneous, and verify on a case by case basis whether the resulting loss function satisfies Proposition 3. As an example, consider $\tilde{C}(\cdot) = Tr(A^d)$ for some $d \geq 2$ and where A is symmetric and positive definite. Since the trace is a linear operator, the resulting loss function, homogeneous of degree d, takes the form

$$L(\hat{\Sigma}_t, H_t) = Tr(\hat{\Sigma}_t^d) - Tr(H_t^d) - dTr[H_t^{d-1}(\hat{\Sigma}_t - H_t)].$$

3 Ranking multivariate GARCH models

In this section, we provide details on the set of competing models and on the volatility proxy that is used as the target in the evaluation.

3.1 Forecasting models set

We focus on the ranking of multivariate volatility models that belong to the multivariate GARCH (MGARCH) class. Consider a N-dimensional discrete time vector stochastic process r_t. Let $\mu_t = E(r_t|\Im_{t-1})$ be the conditional mean vector and $H_{m,t} = E(r_t r_t'|\Im_{t-1})$ the conditional variance matrix for model m so that we can write the model of interest as:

$$r_t = \mu_t + H_{m,t}^{1/2} z_t,$$

where $H_{m,t}^{1/2}$ is a $(N \times N)$ positive definite matrix and z_t is an independent and identically distributed random innovation vector with $E(z_t) = 0$ and $Var(z_t) = I_N$.

In the empirical application in Section 5, we consider 16 specifications for H_{mt} which are frequently used in practice. As detailed in Section 4, for the simulation study, we consider a different forecasting models set (10 models), in order to control for the degree of similarity between models. The specifications considered in this paper are: diagonal BEKK (D-BEKK) model (Engle and Kroner, 1995), multivariate RiskMetrics model (J.P.Morgan, 1996), Constant Correlation (CCC) model (Bollerslev, 1990), Dynamic Conditional Correlation (DCC) model (Engle, 2002) and Generalized Orthogonal GARCH (GOG) model (van der Weide, 2002). The univariate GARCH specifications specifications for the conditional variance equations used in the DCC, CCC and GOG are: GARCH (Bollerslev, 1986), GJR (Glosten, Jagannathan, and Runkle, 1992), Exponential GARCH (Nelson, 1991b), Asymmetric Power ARCH (Ding, Granger, and Engle, 1993), Integrated GARCH (Engle and Bollerslev, 1986), RiskMetrics (J.P.Morgan, 1996) and Hyperbolic GARCH (Davidson, 2004). Table 2 provides a summary of the forecasting models set used in the simulation and the application respectively.

In the GJR model, the impact of squared innovations on the conditional variance is different when the innovation is positive or negative. The asymmetric power ARCH model (APARCH) is a general specification which includes seven other ARCH extensions as special cases. The Exponential GARCH model (EGARCH) accommodates the asymmetric relation

Table 2: Forecasting models sets

Simulation study

BEKK-type	CCC	GOG
Diag. BEKK	GARCH	GARCH
RiskMetrics	IGARCH	IGARCH
	EGARCH	EGARCH
	RM	HYGARCH

Note: Simulation based on bivariate models. CCC and GOG specification use the same conditional variance specification for all series.

Empirical application

BEKK-type	CCC	DCC	GOG
Diag. BEKK	GARCH	GARCH	GARCH
RiskMetrics	IGARCH	IGARCH	IGARCH
	APARCH	APARCH	APARCH
	GJR	GJR	GJR
	RM	RM	

Note: Empirical application based on trivariate models. CCC, DCC and GOG specification use the same conditional variance specification for all series.

between shocks and volatility by expressing the latter as a function of both the magnitude and the sign of the shock. The Integrated GARCH (IGARCH) model is a variation of the GARCH model in which the sum of the ARCH and GARCH parameters are constrained to be equal to one, while the RiskMetrics model (RM) is basically an IGARCH model where the constant is set to zero and the ARCH and GARCH coefficients are fixed to 0.06 and 0.94 respectively. Finally, the Hyperbolic GARCH model (HYGARCH) allows to account for long run dependence in the volatility. The functional forms for H_t are briefly defined in Table 3. See Bauwens, Laurent, and Rombouts (2006) for further details. All the specifications are characterized by a constant conditional mean and the models are estimated by quasi maximum likelihood. The sample log-likelihood is given (up to a constant) by

$$-\frac{1}{2}\sum_{t=1}^{\mathcal{T}} \log \mid H_{m,t} \mid -\frac{1}{2}\sum_{t=1}^{\mathcal{T}} (r_t-\mu)' H_{m,t}^{-1}(r_t-\mu), \tag{15}$$

where $\mathcal{T}$ is the size of the estimation sample. We maximize numerically for μ and the parameters in $H_{m,t}$. All calculations and results reported in this paper are based on programs written by the authors using Ox version 6.0 (Doornik, 2002) and G@RCH version 6.0 (Laurent, 2009).

Table 3: Multivariate GARCH specifications

Model	Multivariate GARCH models for H_t $(N = 2)$	# par.
DBEKK(1,1)	$H_t = C_0^{*'} C_0^* + A^{*'} \epsilon_{t-1} \epsilon_{t-1}' A^* + G^{*'} H_{t-1} G^*$	7
RiskMetrics	$H_t = (1-\alpha)\epsilon_{t-1}\epsilon_{t-1}' + \alpha H_{t-1}$, $(\alpha = 0.96)$	0
GOG	$V^{-1/2}\epsilon_t = L f_t$ $H_t = V^{1/2} L Z_t L V^{1/2}$ $Z_t = diag(\sigma^2_{f_{1,t}}, \ldots, \sigma^2_{f_{m,t}})$ $L = P \Lambda^{1/2} U$, $U = \prod_{i<j} R_{i,j}(\delta_{i,j})$, $-\pi \leq \delta_{i,j} \leq \pi$	1+univ.
CCC	$H_t = D_t R D_t$ $D_t = diag(h^{1/2}_{1,1,t} \ldots h^{1/2}_{N,N,t})$	1+univ.
DCC(1,1)	$H_t = D_t R_t D_t$ $R_t = diag(q^{-1/2}_{1,1,t} \ldots q^{-1/2}_{N,N,t}) Q_t diag(q^{-1/2}_{1,1,t} \ldots q^{-1/2}_{N,N,t})$ $D_t = diag(h^{1/2}_{1,1,t} \ldots h^{1/2}_{NNt})$ $u_t = D_t^{-1}\epsilon_t$ $Q_t = (1-\alpha-\beta)\bar{Q} + \alpha u_{t-1} u_{t-1}' + \beta Q_{t-1}$	3+univ.
	Univariate GARCH models in Z_t and D_t $(l = 1, 2)$	
GARCH(1,1)	$h_{l,t} = \omega_l + \alpha_l \epsilon^2_{l,t-1} + \beta_l h_{l,t-1}$	6
EGARCH(1,0)	$log(h_{l,t}) = \omega_l + g(z_{l,t-1}) + \beta_l log(h_{l,t-1})$ $g(z_{l,t-1}) = \theta_{l,1} z_{l,t-1} + \theta_{l,2}(\lvert z_{l,t}\rvert - E(\lvert z_{l,t}\rvert))$	8
GJR(1,1)	$h_{l,t} = \omega_l + \alpha_l \epsilon^2_{l,t-1} + \gamma_l S^-_{l,t-1}\epsilon^2_{l,t-1} + \beta_l h_{l,t-1}$ $S^-_{l,t} = 1$ if $\epsilon_{l,t} < 0$; $S^-_{l,t} = 0$ if $\epsilon_{l,t} \geq 0$	8
APARCH(1,1)	$h^{\delta_l}_{l,t} = \omega_l + \alpha_l[\lvert\epsilon_{l,t-1}\rvert - \gamma_l \epsilon_{l,t-1}]^{\delta_l} + \beta_l h^{\delta_l}_{l,t-1}$	10
HYGARCH(1,d,1)	$h_{l,t} = \omega_l[1-\beta_l]^{-1} + \lambda(L)\epsilon^2_{l,t}$ $\lambda(L) = \left\{1 - [1-\beta_l]^{-1}\alpha_l[1 + \gamma_l(1-L)^d]\right\}$	10

3.2 A proxy for the conditional variance matrix

Following Andersen, Bollerslev, Diebold, and Labys (2003), we rely on the realized covariance ($RCov$) to proxy the ex-post variance. In the ideal case of no microstructure noise, this measure, being based on intraday observations, is characterized by a degree of accuracy that decreases as sampling frequency lowers.

Let us assume the observed return vector to be generated by a conditionally normal N-dimensional log-price diffusion $dy(u)$ and a $N(N+1)/2$-dimensional covariance diffusion $d\sigma(u)$, with $\sigma(u) = vech(\Sigma(u)) = [\sigma_{ij}(u)]$ for $i,j = 1,...,N,\ i \geq j$ and $u \in [t, t+1]$, with mean vector process $b(u)du$ and variance matrix $a(u) = s(u)s(u)'$, driven by a $N(N+3)/2$ vector of independent standard Brownian motions $W(u)$. Hence the diffusion process of the system admits the following representation

$$\begin{bmatrix} dy(u) \\ d\sigma(u) \end{bmatrix} = b(u)du + s(u)dW(u), \tag{16}$$

with $b(u)$ and $s(u)$ locally bounded and measurable. Consider now the following partition for the variance matrix of the system in (16) as

$$a(u) = s(u)s(u)' = \begin{bmatrix} \Sigma(u) & Cov(dy(u), d\sigma(u)) \\ Cov(dy(u), d\sigma(u)) & Var(d\sigma(u)) \end{bmatrix}. \tag{17}$$

Since $\Sigma(u)$ identifies the continuous time process for the variance matrix of the returns, we can define the Integrated Covariance (ICov) as (see Barndorff-Nielsen and Shephard, 2004)

$$ICov_{t+1} = \int_t^{t+1} \Sigma(u)du. \tag{18}$$

Let us now define the intraday returns as $r_{t+\Delta} = y_{t+\Delta} - y_t$ for $t = \Delta, 2\Delta, ..., T$ and where $1/\Delta$ is the number of intervals per day. In this setting $ICov_t$ can be consistently estimated by the Realized Covariance ($RCov$) (Andersen, Bollerslev, Diebold, and Labys, 2003) which is defined as

$$RCov_{t+1,\Delta} = \sum_{i=1}^{1/\Delta} r_{t+i\Delta} r'_{t+i\Delta}. \tag{19}$$

In fact, since the process defined by (16) does not allow for jumps in the returns, it holds that

$$\operatorname*{plim}_{\Delta \to 0} RCov_{t+1,\Delta} = ICov_{t+1}. \tag{20}$$

In this paper, the $RCov$ serves as a proxy for the true conditional variance matrix when evaluating the forecasting performance of the different MGARCH models. The result (20) suggests that the higher the intraday frequency used to compute $RCov$, and hence the amount of information available, the higher the accuracy of the proxy. The advantage of using $RCov$ is that it satisfies assumption A2.5 (see Barndorff-Nielsen and Shephard,2002 and Hansen and Lunde, 2006b) and therefore it ensures convergence of the approximated ordering to the true one under the inconsistent loss function (see Proposition 2). On the other hand, $RCov$ is a valid proxy even when based on very low intraday sampling frequencies. The use of RCov allows to study the behavior of the ordering as a function of the level of accuracy of the proxy for consistent and inconsistent loss functions. As noted by Andersen, Bollerslev, Diebold, and Labys (2003), positive definiteness of the variance matrix is ensured only if the number of assets is smaller then $1/\Delta$. When this condition is violated then the realized covariance matrix fails to be of full rank (i.e., $rank(RCov) = 1/\Delta < dim(RCov)$) and $RCov$ will meet only the weaker requirement to be semi-positive definite. Since the setting defined in this paper requires positive definiteness of the variance proxy, we restrict our analysis on the range of proxies that meet this requirement. Note that, other volatility proxies can be used instead, such as the multivariate realized kernels (see Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008a and Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008b, Hansen and Lunde, 2006b, Zhou, 1996) or the range based covariance estimators (Brandt and Diebold, 2006).

4 Simulation study

We investigate the ranking of the MGARCH models with respect to two dimensions: the quality of the volatility proxy and the choice of the loss function. According to Proposition 2, we find that if the quality of the proxy is sufficiently good, both consistent and inconsistent loss functions rank properly. However, when the quality of the proxy is poor, only the consistent loss functions rank correctly. Our findings also hold when the sample size in the estimation period increases.

4.1 Setup

Varying the quality of the proxy requires knowledge of the intraday sample paths for the returns. This can be obtained through the simulation of a multivariate diffusion process. In the spirit of Meddahi (2002) and Voev and Lunde (2006), we generate continuous sample paths such that the resulting $RCov$ estimators, at different time sampling frequencies, are consistent for $ICov$. Contrary to the previous literature, the diffusion approximation we introduce here, the bivariate CCC-EGARCH(1,0) model, allows to fully control for the nature and the size of the leverage effect and to preserve the correlation structure of the vector stochastic process $[y_{i,t}, ..., \sigma^2_{i,j,t}, ...]'$, $i, j = 1, ..., N$ and $i \leq j$ ensuring internal consistency of the model.

We consider the bivariate CCC-EGARCH(1,0) model (see Table 3) which admits a diffusion limit, of the type introduced by (16), defined by the continuous time vector stochastic process $[y_{1,t}, y_{2,t}, log(\sigma^2_{1,t}), log(\sigma^2_{2,t})]'$ with drift and scale given respectively by

$$b(y,\Sigma) \quad = \quad \begin{bmatrix} 0 \\ 0 \\ \omega_1 - \theta_1 \log(\sigma^2_{1,t}) \\ \omega_2 - \theta_2 \log(\sigma^2_{2,t}) \end{bmatrix} \tag{21}$$

and

$$\begin{aligned} a(y,\Sigma) &= s(y,\Sigma)s(y,\Sigma)' \\ &= \begin{bmatrix} \sigma^2_{1t} & \rho\sigma_{1,t}\sigma_{2,t} & \alpha_1\sigma_{1,t} & \rho\alpha_2\sigma_{1,t} \\ \rho\sigma_{1,t}\sigma_{2,t} & \sigma^2_{2,t} & \rho\alpha_1\sigma_{2,t} & \alpha_2\sigma_{2,t} \\ \alpha_1\sigma_{1,t} & \rho\alpha_1\sigma_{2,t} & \alpha_1^2 + \gamma_1^2(1-2/\pi) & \rho\alpha_1\alpha_2 + \gamma_1\gamma_2 C \\ \rho\alpha_2\sigma_{1,t} & \alpha_2\sigma_{2,t} & \rho\alpha_1\alpha_2 + \gamma_1\gamma_2 C & \alpha_2^2 + \gamma_2^2(1-2/\pi) \end{bmatrix}, \end{aligned} \tag{22}$$

where $C = \frac{2}{\pi}\left[\sqrt{1-\rho^2} + \rho \arcsin(\rho) - 1\right]$. The conditional variance matrix is computed, at each point in time as $\sigma_{(1,2),t} = \rho\sqrt{\sigma^2_{1,t}\sigma^2_{2,t}}$. The matrix $s(y,\Sigma)$ is computed from $a(y,\Sigma)$ by spectral decomposition. The diffusion approximation of the CCC-EGARCH model has been derived following Nelson (1991a). For details on the weak convergence of stochastic processes see Strook and Varadhan (1979), Ethier and Kurtz (1986) and Kushner (1984). Details are available upon request.

The CCC-EGARCH specification has been preferred to alternative MGARCH specifications - e.g., the DCC model - because it is sufficient to ensure a certain degree of dissimilarity between the true DGP and the set of competing models while keeping the limiting diffusion fairly tractable.

For the simulation study, we use the following parameter values: $\omega_i = -0.02$, $\theta_i = 1 - \beta_i = 0.03$, $\alpha_i = -0.09$, $\gamma_i = 0.4$ and $\rho = 0.9$ which ensure realistic dynamics for the return process. Our results are based on 500 replications with an estimation sample $T = 2000$ observations and a forecasting sample of 500 observations. The continuous time process (16) is approximated by generating $1/\Delta = 7200$ observations per day - i.e., 5 observations per minute. The set of MGARCH models is estimated on daily returns and recursive 1-step ahead forecasts are computed. Although, a discussion on finite sample properties of the diffusion approximation is beyond the scope of this paper, it is important to stress that, in this setting, to achieve asymptotics we must set $1/\Delta$ and T such that $\Delta T \to \infty$. In a different yet related setting, Barone-Adesi, Rasmussen, and Ravanelli (2005) showed that, in finite samples, the discrete time coefficient can be severely biased. On the other hand, our aim is to generate enough intraday observations to compute $RCov$ at frequencies sufficiently high to approximate (20), thus a sufficiently large sample size is unfeasible given the available computational power. To overcome this problem, the set of forecasting models is defined such that all competing models are expected to be inferior. Apart from the CCC-EGARCH(1,0), the set of competing models includes the diagonal BEKK, RiskMetrics, CCC-GARCH(1,1), CCC-IGARCH(1,1), CCC-RiskMetrics, GOG-GARCH(1,1), GOG-EGARCH(1,0), GOG-IGARCH(1,1) and GOG-HYGARCH(1,1) (see Table 3).

The true conditional variance matrix is measured by the integrated covariance ($ICov$) defined in (18). To proxy the daily variance matrix of day t, we use the realized covariance ($RCov_{t,\Delta}$), as defined in (19), based on equally spaced intraday returns sampled at 14 different frequencies, ranging from 1 minute (most accurate) to 24 hours (least accurate), over the forecasting horizon. In our setting, the bivariate dimension implies that the lowest available sampling frequency that ensures positive definiteness of $RCov$ is 12 hours. However, when reporting our simulation results, we also include the 24 hours frequency to (qualitatively) assess whether the evaluation based on a singular realized variance matrix has an impact on the ordering.

Since we are comparing estimated models, the underlying order, except for the best model, is unknown ex-ante and it is determined by the specific loss function used in the evaluation.

We consider the ranking implied by three consistent, L_F, L_E and L_S, and three inconsistent loss functions, L_{1M}, L_{PF} and $L_{LF,2}$, respectively.

4.2 Sample performance ranking and objective bias

We focus first on the ability of the loss function to detect the true model as the best. We compute the frequencies at which each model shows the smallest sample performance where the latter is defined as the average value of the loss function over the T forecasts.

Table 4 reports the frequencies for the consistent loss functions: the Frobenius distance, the Euclidean distance and the Stein loss function. Unsurprisingly, we find the CCC-EGARCH model ranking first most often for all consistent loss functions at all frequencies for *RCov*. When *ICov* is used, this frequency is between 50% and 54%. The remaining is distributed among the other models (from 0% to 7%) in such a way that no model dominates. One exception is the GOG-EGARCH (17%) when the Frobenius and the Euclidean distances are used. This result is not surprising since the GOG-EGARCH model is the only model in the set that allows for a leverage effect. Note that the frequency associated to the GOG-EGARCH is stable across *RCov* frequencies, that is, it only represents the ability of the GOG-EGARCH to mimic the dynamics in the variance structure generated by the DGP.

An interesting case is the CCC-RM. When the Frobenius and the Euclidean distances are used, the frequencies associated to the CCC-RM increase progressively from about 6% at *ICov* to 10% at $RCov_{12h}$ revealing a behavior that, as we will see in the following, typically suggests the presence of the objective bias. However, the set of models includes also the CCC-IGARCH, a model which shares most of the characteristics of the CCC-RM. The frequencies of the CCC-IGARCH decrease from 5% to 2% in such a way to compensate, at each *RCov* frequency, the increase in the frequency associated to the CCC-RM. The joint probability of CCC-IGARCH and CCC-RM to rank first is indeed about 12% for both L_F and L_E and is stable across *RCov*.

As expected, L_F and L_E show very similar patterns. Both loss functions share the same structure with the only difference given by the weights assigned to the covariances (i.e., $\hat{\Lambda}$ in Proposition 4), which are double weighted in the Frobenius distance. In this case, using different matrices of weights does not affect the distribution of the models.

On the other hand, the Stein loss function shows a different distribution across models.

The frequency at which the CCC-EGARCH ranks first is slightly larger than with the previous two distances and there is no evidence of a shift in the ordering between CCC-IGARCH and CCC-RM. Also, the frequency associated to the GOG-EGARCH, which is potentially the best alternative model, is again stable across frequencies but this model ranks first only about 11% of the times.

In Proposition 2, we have shown that a consistent loss function always detects the optimal forecast, if it exists, independently from the level of accuracy of the proxy. In fact, Table 4 shows that all consistent loss function point to the correct model. However, the ranking of two imperfect forecasts may differ between loss functions. In particular, it will depend on how each specific loss function penalizes deviations from the target. Consistency of the ranking is in fact intended only with respect to the accuracy of the proxy and for a given loss function and not as invariance of the ordering with respect to the choice of the loss function. The latter is only ensured between the optimal forecast and any other point in the space of the forecast matrices (see Jensen (1984)).

Apart from the issue related to finite sample properties of the diffusion approximation, the fact that the frequencies associated to the true model seem low when the loss is computed with respect to the true covariance is explained by the fact that we allow for a fairly high degree of similarity between models. The CCC-EGARCH model with a moderate leverage effect can also be accommodated by other models in the set. However, the presence of leverage effect in the DGP implies that models ignoring this feature of the data are expected to be inferior. From Table 4 we also learn that when the quality of the proxy deteriorates (the sampling frequency decreases), the relative sample performances are invariant, which implies consistency of the ranking of these loss functions across $RCov$ frequencies.

Table 5 reports the frequencies at which each model shows the smallest sample performance but for the inconsistent loss functions, i.e., L_{1M}, L_{PF} and $L_{LF,2}$. These loss functions deliver the true ranking when the target variance is $ICov$. Indeed, the CCC-EGARCH is always correctly detected as the best model, though the frequencies vary substantially across loss functions. When relying on $RCov_{1m}$ to $RCov_{1h}$, the frequencies associated to each model remain stable and there is no dominant model other than CCC-EGARCH. Hence, as shown in Proposition 2, when the proxy is nearly perfect there is no evidence of the presence of objective bias. Starting from $RCov_{2h}$, the frequency at which the CCC-EGARCH model ranks first starts declining while the performance of potentially inferior models increases rapidly as the quality of the proxy lowers. The CCC-EGARCH frequency drops from about 50% to about 38%, 36% to 22% and 26% to 12% at the 12h frequency, respectively for each loss function. In accordance with the results in Proposition 2, we find that as the quality of the proxy deteriorates inferior models seem to emerge. Although when using L_{1M} there is no model that dominates the CCC-EGARCH, the GOG-EGARCH and the CCC-RM follow closely. These models rank first in 18% and about 5% of the cases respectively when using $RCov_{1m}$ to $RCov_{30m}$ and in 29% and 20% when using $RCov_{12h}$. The relative improvement in the sample performance of inferior models, as the frequency of $RCov$ lowers, signals the presence of objective bias.

When considering $L_{LF,2}$, we find two models, namely the GOG-EGARCH and the GOG-IGARCH, that outperform the true model when the proxy is computed using returns sampled at 8h and 12h. In this case, we observe a particular behavior: the frequencies associated to the GOG-EGARCH (about 22%) and the GOG-IGARCH (20% to 23%) are fairly constant across proxies. However, the frequency associated to the CCC-EGARCH drops fast as the proxy becomes less accurate. Even if the CCC-EGARCH is found to be the best model in 35% of the times when using a proxy based on 1h returns, this frequency falls to 22% at $RCov_{12h}$.

The same remarks apply to the results obtained using L_{PF}. In this case, surprisingly we find the DBEKK to rank first more often than the true model when the proxy is computed using returns sampled at a frequency of 2h or lower.

In the first part of this simulation study, we focused on the detection of the best model in terms of sample performance. However, the analysis carried out so far, offers only a partial insight on the role of the objective bias. Indeed, in presence of a high degree of dissimilarity between the true and the competing models, the detection of the best model may not be affected. However, the objective bias may still be relevant for what concerns the other positions in the ranking. We now investigate whether the whole ordering is preserved despite the deterioration of the quality of the proxy. Since we are ranking over a set of estimated volatility models, the true ranking implied by a given loss function, except for the best model, is not known ex-ante. However, the underlying ordering implied by a given loss function, can be identified by ranking the models with respect to the true covariance, i.e.,

Table 4: Frequencies at which each model shows smallest loss (consistent loss functions)

Frobenius distance (L_F)										
	DBEKK	RM	CCCG	CCCE	CCCI	CCCRM	GOGG	GOGE	GOGI	GOGHY
$ICov$	0.002	0.008	0.058	**0.508**	0.052	0.068	0.036	0.172	0.076	0.020
$RCov_{1m}$	0.002	0.006	0.058	**0.510**	0.048	0.068	0.040	0.172	0.072	0.024
$RCov_{5m}$	0.002	0.006	0.060	**0.504**	0.048	0.076	0.042	0.166	0.070	0.026
$RCov_{10m}$	0.004	0.004	0.054	**0.512**	0.040	0.084	0.036	0.168	0.070	0.028
$RCov_{15m}$	0.002	0.008	0.056	**0.504**	0.040	0.076	0.038	0.174	0.076	0.026
$RCov_{20m}$	0.002	0.006	0.048	**0.520**	0.044	0.082	0.036	0.172	0.074	0.016
$RCov_{30m}$	0.002	0.004	0.058	**0.512**	0.042	0.084	0.038	0.170	0.072	0.018
$RCov_{1h}$	0.002	0.002	0.058	**0.522**	0.032	0.092	0.034	0.156	0.072	0.030
$RCov_{2h}$	0.006	0.004	0.048	**0.528**	0.030	0.070	0.034	0.172	0.070	0.038
$RCov_{3h}$	0.002	0.006	0.038	**0.526**	0.036	0.090	0.034	0.152	0.080	0.036
$RCov_{4h}$	0.006	0.002	0.044	**0.496**	0.040	0.098	0.026	0.170	0.074	0.044
$RCov_{6h}$	0.002	0.006	0.042	**0.524**	0.026	0.082	0.022	0.162	0.096	0.038
$RCov_{8h}$	0.006	0.006	0.040	**0.494**	0.018	0.130	0.040	0.152	0.080	0.034
$RCov_{12h}$	0.010	0.000	0.050	**0.526**	0.018	0.100	0.022	0.158	0.078	0.038
Euclidean distance (L_E)										
$ICov$	0.006	0.008	0.064	**0.512**	0.036	0.062	0.040	0.178	0.074	0.020
$RCov_{1m}$	0.006	0.006	0.060	**0.508**	0.040	0.066	0.040	0.176	0.074	0.024
$RCov_{5m}$	0.006	0.006	0.058	**0.514**	0.038	0.070	0.040	0.170	0.072	0.026
$RCov_{10m}$	0.004	0.004	0.052	**0.518**	0.036	0.074	0.034	0.176	0.080	0.022
$RCov_{15m}$	0.006	0.006	0.056	**0.514**	0.034	0.076	0.030	0.176	0.074	0.028
$RCov_{20m}$	0.004	0.004	0.046	**0.526**	0.042	0.078	0.036	0.168	0.080	0.016
$RCov_{30m}$	0.006	0.002	0.054	**0.520**	0.038	0.076	0.038	0.172	0.078	0.016
$RCov_{1h}$	0.004	0.002	0.054	**0.532**	0.030	0.086	0.030	0.162	0.070	0.030
$RCov_{2h}$	0.004	0.004	0.044	**0.526**	0.026	0.074	0.036	0.184	0.068	0.034
$RCov_{3h}$	0.004	0.006	0.038	**0.528**	0.034	0.088	0.034	0.152	0.082	0.034
$RCov_{4h}$	0.004	0.002	0.046	**0.491**	0.036	0.096	0.030	0.172	0.074	0.042
$RCov_{6h}$	0.004	0.006	0.042	**0.518**	0.026	0.080	0.020	0.166	0.100	0.038
$RCov_{8h}$	0.004	0.006	0.040	**0.516**	0.020	0.118	0.036	0.150	0.076	0.034
$RCov_{12h}$	0.008	0.000	0.050	**0.532**	0.020	0.100	0.020	0.152	0.080	0.038
Stein loss function (L_S)										
$ICov$	0.004	0.000	0.058	**0.540**	0.116	0.004	0.042	0.102	0.076	0.058
$RCov_{1m}$	0.004	0.000	0.060	**0.542**	0.114	0.004	0.038	0.098	0.080	0.060
$RCov_{5m}$	0.004	0.000	0.060	**0.550**	0.112	0.004	0.040	0.092	0.076	0.062
$RCov_{10m}$	0.004	0.000	0.058	**0.552**	0.106	0.006	0.038	0.100	0.074	0.062
$RCov_{15m}$	0.004	0.000	0.060	**0.552**	0.116	0.004	0.036	0.098	0.068	0.062
$RCov_{20m}$	0.004	0.000	0.048	**0.560**	0.112	0.004	0.040	0.108	0.066	0.058
$RCov_{30m}$	0.004	0.000	0.058	**0.558**	0.112	0.004	0.040	0.102	0.060	0.062
$RCov_{1h}$	0.004	0.000	0.056	**0.552**	0.112	0.008	0.038	0.104	0.064	0.062
$RCov_{2h}$	0.002	0.000	0.058	**0.558**	0.106	0.004	0.038	0.114	0.066	0.054
$RCov_{3h}$	0.000	0.004	0.060	**0.560**	0.094	0.006	0.032	0.106	0.078	0.060
$RCov_{4h}$	0.002	0.002	0.060	**0.540**	0.112	0.002	0.034	0.112	0.084	0.052
$RCov_{6h}$	0.000	0.004	0.064	**0.524**	0.102	0.004	0.038	0.112	0.088	0.064
$RCov_{8h}$	0.004	0.002	0.052	**0.516**	0.108	0.002	0.034	0.122	0.098	0.062
$RCov_{12h}$	0.004	0.002	0.048	**0.540**	0.106	0.006	0.034	0.108	0.090	0.062

Note: D-BEKK: Diagonal BEKK; RM: RiskMetrics; CCC-G,-E,-I,-RM: Constant Conditional Correlation with GARCH, EGARCH, IGARCH and RiskMetrics univariate conditional variances; GOG-G,-E,-I,-HY: Generalized Orthogonal GARCH with GARCH, EGARCH, IGARCH and HYGARCH univariate conditional variances

Table 5: Frequencies at which each model shows smallest loss: inconsistent loss functions

	DBEKK	RM	CCCG	CCCE	CCCI	CCCRM	GOGG	GOGE	GOGI	GOGHY
Entrywise 1 - (matrix) norm (L_{1M})										
$ICov$	0.004	0.004	0.054	**0.506**	0.024	0.060	0.030	0.186	0.110	0.022
$RCov_{1m}$	0.004	0.004	0.052	**0.504**	0.028	0.060	0.032	0.182	0.110	0.024
$RCov_{5m}$	0.004	0.004	0.046	**0.506**	0.024	0.066	0.038	0.180	0.108	0.024
$RCov_{10m}$	0.006	0.004	0.042	**0.516**	0.024	0.064	0.034	0.174	0.108	0.028
$RCov_{15m}$	0.006	0.002	0.040	**0.512**	0.022	0.068	0.030	0.180	0.114	0.026
$RCov_{20m}$	0.006	0.002	0.040	**0.506**	0.024	0.072	0.032	0.178	0.116	0.024
$RCov_{30m}$	0.006	0.002	0.044	**0.506**	0.022	0.068	0.028	0.184	0.114	0.026
$RCov_{1h}$	0.008	0.002	0.046	**0.504**	0.024	0.076	0.018	0.190	0.106	0.026
$RCov_{2h}$	0.010	0.002	0.052	**0.476**	0.022	0.096	0.010	0.218	0.090	0.024
$RCov_{3h}$	0.010	0.008	0.042	**0.474**	0.022	0.100	0.016	0.212	0.092	0.024
$RCov_{4h}$	0.012	0.008	0.030	**0.458**	0.022	0.122	0.010	0.224	0.088	0.026
$RCov_{6h}$	0.016	0.018	0.012	**0.424**	0.020	0.150	0.008	0.258	0.070	0.024
$RCov_{8h}$	0.018	0.028	0.012	**0.402**	0.010	0.178	0.008	0.260	0.068	0.016
$RCov_{12h}$	0.024	0.028	0.006	**0.376**	0.010	0.208	0.000	0.292	0.052	0.004
Frobenius distance - log ($L_{LF,2}$)										
$ICov$	0.016	0.000	0.038	**0.362**	0.032	0.030	0.036	0.220	0.204	0.062
$RCov_{1m}$	0.016	0.000	0.044	**0.356**	0.032	0.032	0.034	0.224	0.204	0.058
$RCov_{5m}$	0.022	0.000	0.040	**0.352**	0.034	0.032	0.038	0.220	0.204	0.058
$RCov_{10m}$	0.020	0.000	0.038	**0.346**	0.036	0.034	0.040	0.222	0.202	0.062
$RCov_{15m}$	0.022	0.000	0.042	**0.342**	0.032	0.032	0.032	0.222	0.210	0.066
$RCov_{20m}$	0.020	0.000	0.044	**0.348**	0.030	0.032	0.038	0.216	0.210	0.062
$RCov_{30m}$	0.024	0.000	0.036	**0.346**	0.032	0.032	0.040	0.220	0.220	0.056
$RCov_{1h}$	0.024	0.002	0.034	**0.346**	0.030	0.038	0.030	0.224	0.216	0.056
$RCov_{2h}$	0.030	0.004	0.032	**0.334**	0.034	0.044	0.030	0.218	0.232	0.042
$RCov_{3h}$	0.044	0.006	0.028	**0.318**	0.032	0.052	0.036	0.226	0.214	0.044
$RCov_{4h}$	0.046	0.006	0.020	**0.306**	0.030	0.058	0.040	0.232	0.220	0.042
$RCov_{6h}$	0.062	0.018	0.018	**0.254**	0.030	0.080	0.026	0.232	0.232	0.048
$RCov_{8h}$	0.066	0.018	0.018	0.240	0.026	0.104	0.020	0.230	**0.242**	0.036
$RCov_{12h}$	0.082	0.030	0.024	0.218	0.020	0.120	0.018	**0.224**	**0.232**	0.032
Frobenius distance - prop. (L_{PF})										
$ICov$	0.132	0.004	0.044	**0.260**	0.124	0.040	0.086	0.170	0.048	0.092
$RCov_{1m}$	0.136	0.004	0.044	**0.254**	0.126	0.040	0.086	0.170	0.050	0.090
$RCov_{5m}$	0.140	0.004	0.040	**0.256**	0.130	0.036	0.092	0.172	0.046	0.084
$RCov_{10m}$	0.134	0.004	0.042	**0.242**	0.136	0.036	0.102	0.178	0.044	0.082
$RCov_{15m}$	0.142	0.004	0.042	**0.228**	0.144	0.034	0.096	0.176	0.048	0.086
$RCov_{20m}$	0.130	0.004	0.040	**0.240**	0.134	0.036	0.106	0.178	0.046	0.086
$RCov_{30m}$	0.144	0.002	0.040	**0.228**	0.134	0.034	0.102	0.180	0.042	0.094
$RCov_{1h}$	0.158	0.004	0.034	**0.228**	0.132	0.036	0.100	0.176	0.042	0.090
$RCov_{2h}$	**0.208**	0.000	0.026	0.204	0.132	0.028	0.092	0.180	0.034	0.096
$RCov_{3h}$	**0.254**	0.002	0.028	0.180	0.134	0.030	0.082	0.176	0.028	0.086
$RCov_{4h}$	**0.276**	0.004	0.030	0.158	0.130	0.024	0.100	0.164	0.028	0.086
$RCov_{6h}$	**0.304**	0.002	0.032	0.154	0.120	0.034	0.078	0.162	0.030	0.084
$RCov_{8h}$	**0.338**	0.004	0.036	0.132	0.114	0.028	0.078	0.172	0.028	0.070
$RCov_{12h}$	**0.356**	0.000	0.032	0.120	**0.130**	0.036	0.084	0.142	0.024	0.076

Note: D-BEKK: Diagonal BEKK; RM: RiskMetrics; CCC-G,-E,-I,-RM: Constant Conditional Correlation with GARCH, EGARCH, IGARCH and RiskMetrics univariate conditional variances; GOG-G,-E,-I,-HY: Generalized Orthogonal GARCH with GARCH, EGARCH, IGARCH and HYGARCH univariate conditional variances

$ICov$.

Without loss of generality, we consider next only one consistent (Frobenius distance) and one inconsistent (entrywise 1 - (matrix) norm) loss function. Figure 1(a) shows the ranking based on the average performance (over the 500 replications) implied by the consistent loss function for various levels of proxy quality. In the following, for reference, we also report the evaluation based on $RCov_{1d}$. As pointed out in the previous section, at this sampling frequency $RCov$ does not meet the requirement of positive definiteness which may affect the evaluation. Figure 1(a) shows that the ranking is fairly stable across $RCov$ frequencies meaning that the L_F is able to consistently order models even when the quality of the proxy deteriorates. Shifts in position affect only the middle of the classification and can be justified by the extremely close average sample performances between the models, with differences at $RCov_{12h}$ smaller than 10^{-2} (Figure 1(b)). Figures 1(b) and 1(c) provide some insights to disentangle the role of the accuracy of the variance matrix proxy. Figure 1(c) reports the model average performances normalized to the average performance of the CCC-EGARCH model. Constant discrepancies between models (Figure 1(b)) confirms that not just the ordering but also the degree of similarity, i.e., the relationships between models, is preserved across $RCov$ frequencies, while Figure 1(c) suggests that the loss of accuracy only translates into a proportional increase the average sample performances for all models. Note that, the increase in the variability of the proxy also induces an increase in the variability of the loss function which, in empirical applications, may result in the impossibility to effectively discriminate between models.

A different picture emerges when considering the inconsistent loss function (Figure 2(a)). In this case, the ranking is preserved up to the $1h$ sampling frequency. Due to the presence of the objective bias, we observe major shifts at lower frequencies at most levels of the classification. The impact of the objective bias is amplified by the fact that except for the first two positions, i.e., CCC-EGARCH and GOG-EGARCH, all the other models exhibit very close average sample performances (Figure 2(b)), with differences smaller than 10^{-2} at $RCov_{12h}$. Inferior models like RiskMetrics and CCC-RM, 10th and 9th respectively according to $ICov$, improve up to the 3rd and 2nd positions respectively. The CCC-EGARCH is classified as the best forecasting model at all frequencies, followed by the GOG-EGARCH. This result is due to the fact that these two models are sufficiently different from the others (they are the only models in the set allowing for leverage effect), with the CCC-EGARCH clearly dominating the GOG-EGARCH (Figure 2(b)). Although the objective bias does not become an issue when ordering between these two models, Figure 2(b) shows that, as the frequency for *RCov* lowers, the average sample performance of the latter gets closer to the CCC-EGARCH performance. Since, as underlined above, the variability of the loss function increases along with the variability of the proxy, the probability to rank the GOG-EGARCH first increases at low frequencies. This conclusion is consistent with the results reported in Table 5.

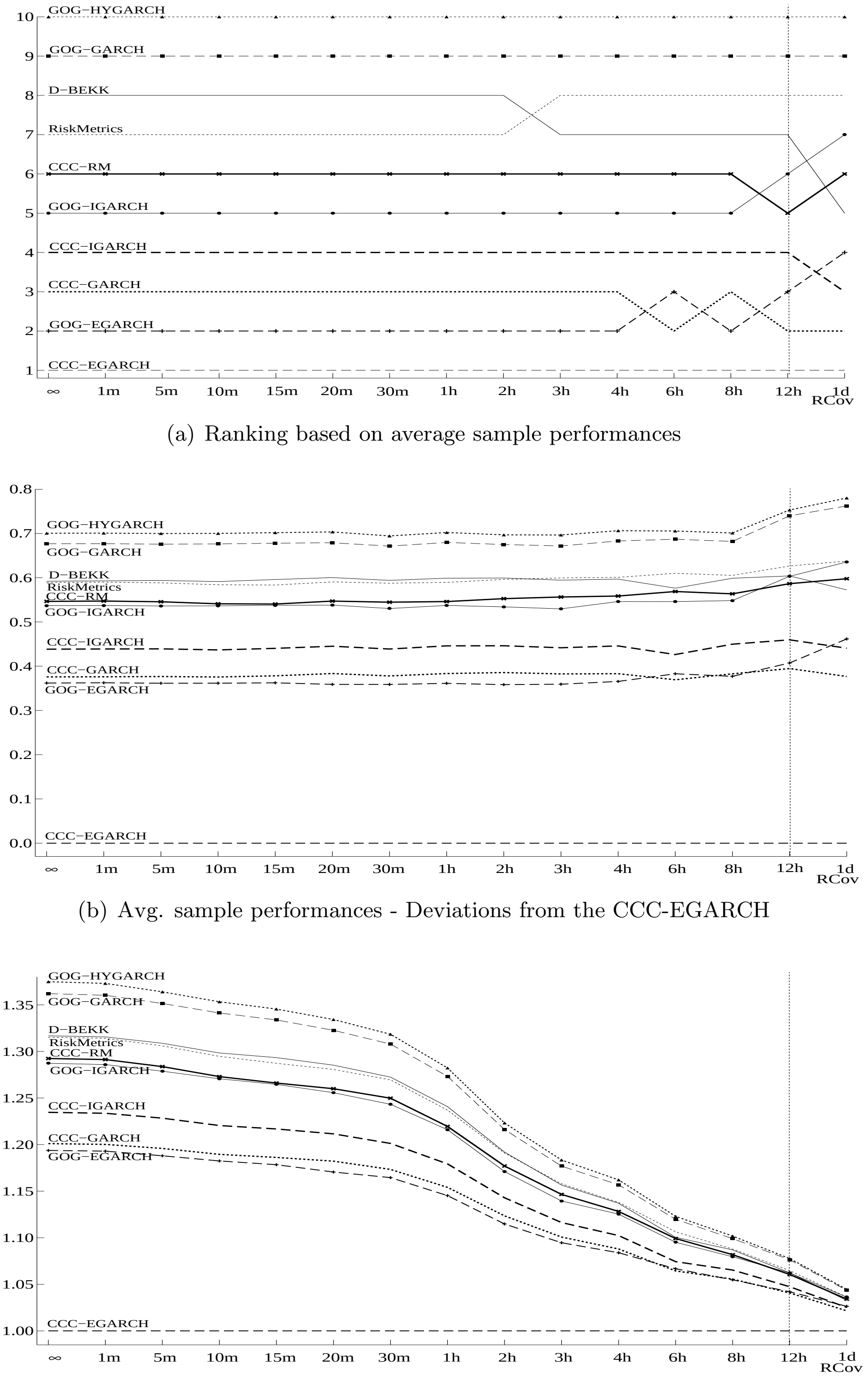
Besides varying the quality of the proxy and studying several loss functions we also investigate the impact of the estimation sample size on the rankings. Increasing the sample size to 3000 observations gives qualitatively similar results (results are available upon request).

5 Empirical application

5.1 Data description and estimation results

The empirical application is based on the Euro, British Pound and the Japanese Yen exchange rates expressed in US dollars (EUR, GBP and JPY). The sample period goes from January 6, 1987 to June 30, 2004 (i.e., 4287 trading days). Intraday returns and realized covariances are computed from five-minutes intervals last mid-quotes, implying 288 intraday observations per day. The data have been provided by Olsen & Associates. Missing values are replaced by linearly interpolating 5-minute price. The dataset has been cleaned from weekends, holidays and early closing days. Days with too many missing values and/or constant prices are also removed. Five-minute returns are computed as the first difference of the logarithmic prices. The estimation sample ranges from January 6, 1987 to December 28, 2001 (3666 trading days), while the remaining observations (621 trading days) are used for the out-of-sample forecasts evaluation. Table 6 reports descriptive statistics for the estimation sample and the forecasting sample. With respect to the daily frequency, the EUR and GBP exchange rates share similar data characteristics and are relatively highly correlated. JPY has quite a higher kurtosis and a more pronounced skewness. The 5-minute realized variances and correlations are quite dispersed. For example the correlations vary between -0.12 and 0.85. We also remark that the variances are positively skewed and the correlations negatively skewed.

The proxy for the conditional variance matrix is realized covariance (*RCov*) as defined

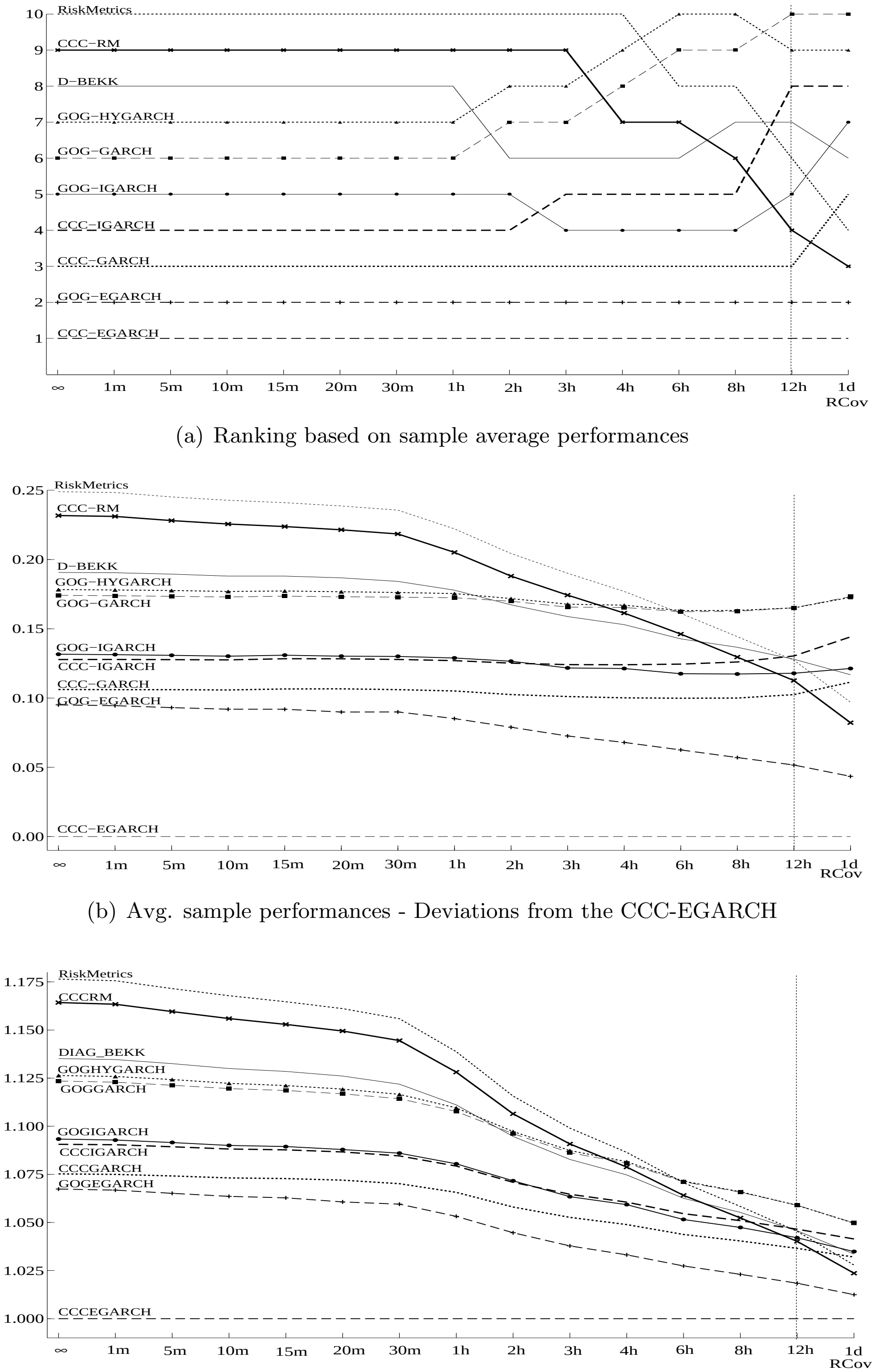


(a) Ranking based on average sample performances

(b) Avg. sample performances - Deviations from the CCC-EGARCH

(c) Normalized average sample performances

Figure 1: Simulation results for L_F (consistent)



(a) Ranking based on sample average performances

(b) Avg. sample performances - Deviations from the CCC-EGARCH

(c) Normalized average sample performances

Figure 2: Simulation results for L_{1M} (inconsistent)

Table 6: Descriptive statistics

Series	min	mean	max	std.dev	skew.	kurt.
Estimation sample: January 6, 1987 to December 28, 2001 (3666 obs)						
EUR	−3.557	−0.003	3.419	0.683	0.043	4.939
GBP	−4.168	−0.002	3.425	0.623	−0.161	6.140
JPY	−4.207	0.003	7.724	0.729	0.619	9.503
Forecasting sample: January 3, 2002 to June 30, 2004 (621 obs)						
EUR	−2.001	0.051	1.837	0.647	−0.227	3.270
GBP	−1.756	0.035	2.051	0.524	−0.221	3.873
JPY	−2.203	0.033	2.686	0.595	−0.129	4.260
$RCov_{5m,EUR}$	0.122	0.457	2.526	0.200	3.024	24.52
$RCov_{5m,GBP}$	0.079	0.315	1.564	0.156	2.410	14.02
$RCov_{5m,JPY}$	0.041	0.413	2.385	0.235	3.221	20.52
$RCor_{5m,EUR,GBP}$	0.012	0.550	0.852	0.120	−0.303	3.359
$RCor_{5m,EUR,JPY}$	−0.035	0.410	0.800	0.147	−0.343	2.639
$RCor_{5m,GBP,JPY}$	−0.122	0.279	0.653	0.127	−0.131	2.885

Notes: The estimated correlations for the estimation sample are $\rho_{EUR,GBP} = 0.720$, $\rho_{EUR,JPY} = 0.493$ and $\rho_{GBP,JPY} = 0.415$. The estimated correlations for the forecasting sample are $\rho_{EUR,GBP} = 0.721$, $\rho_{EUR,JPY} = 0.490$ and $\rho_{GBP,JPY} = 0.416$.

in (19) computed at 14 different frequencies ranging from 5 min. to 24 h. We stress again, like in the simulation study, that we should stop at the 8h frequency if we want to have a positive definite realized variance matrix at each point in time. We include the results until the 24h frequency to illustrate what happens when the realized variance matrix is not positive definite. One-step-ahead forecasts are computed from 4:05 pm to 4:00 pm Eastern Time (ET) and are compared to the $RCov$ using one consistent (Frobenius distance) and one inconsistent (Entrywise 1 - (matrix) norm) loss function. Estimation results for the 16 MGARCH models are reported in Table 7. Note that estimates for the RiskMetrics and CCC-RM models are not reported in Table 7 since they do not require parameter estimation (the sample correlation is used to estimate the constant correlation in the CCC-RM). Generally speaking, we observe that the parameters estimates for the conditional variance, covariance and correlations imply highly persistent processes. Furthermore, in almost all cases, the null of no leverage effect cannot be rejected at standard significance levels.

5.2 Model comparison

The empirical ranking of the 16 MGARCH models, as a function of the level of aggregation of the data used to compute $RCov$, is reported in Figures 3 and 4. The consistent loss function in Figure 3(a) points to the CCC-GARCH as the best forecasting model at almost all frequencies. More generally, we can conclude that the subset given by the CCC and the DCC, both with GARCH and GJR specifications for the variances, outperform all the other models. These models exhibit particularly stable and extremely close sample performances (Figure 3(b)). The overall ranking is well preserved across all frequencies.

The GOG model is always largely dominated by all other models regardless of the conditional variance specification. There is no clear dominance between the CCC and the DCC models and their ranking position depends on the model chosen for the conditional variance. The GARCH/GJR represents the best combination, followed by the APARCH, RiskMetrics and finally the IGARCH. The three models based on the RiskMetrics approach, (i.e., RiskMetrics, CCC-RM and DCC-RM) are positioned in the middle of the classification. The overall ranking is found to be particularly stable when $RCov$ is computed using $5m$ to $1h$ returns (Figure 3(a)). When the frequency gets lower, the ranking is apparently more volatile. In fact, such range strikes a good compromise between the loss of accuracy (low frequencies)

Table 7: Estimation results

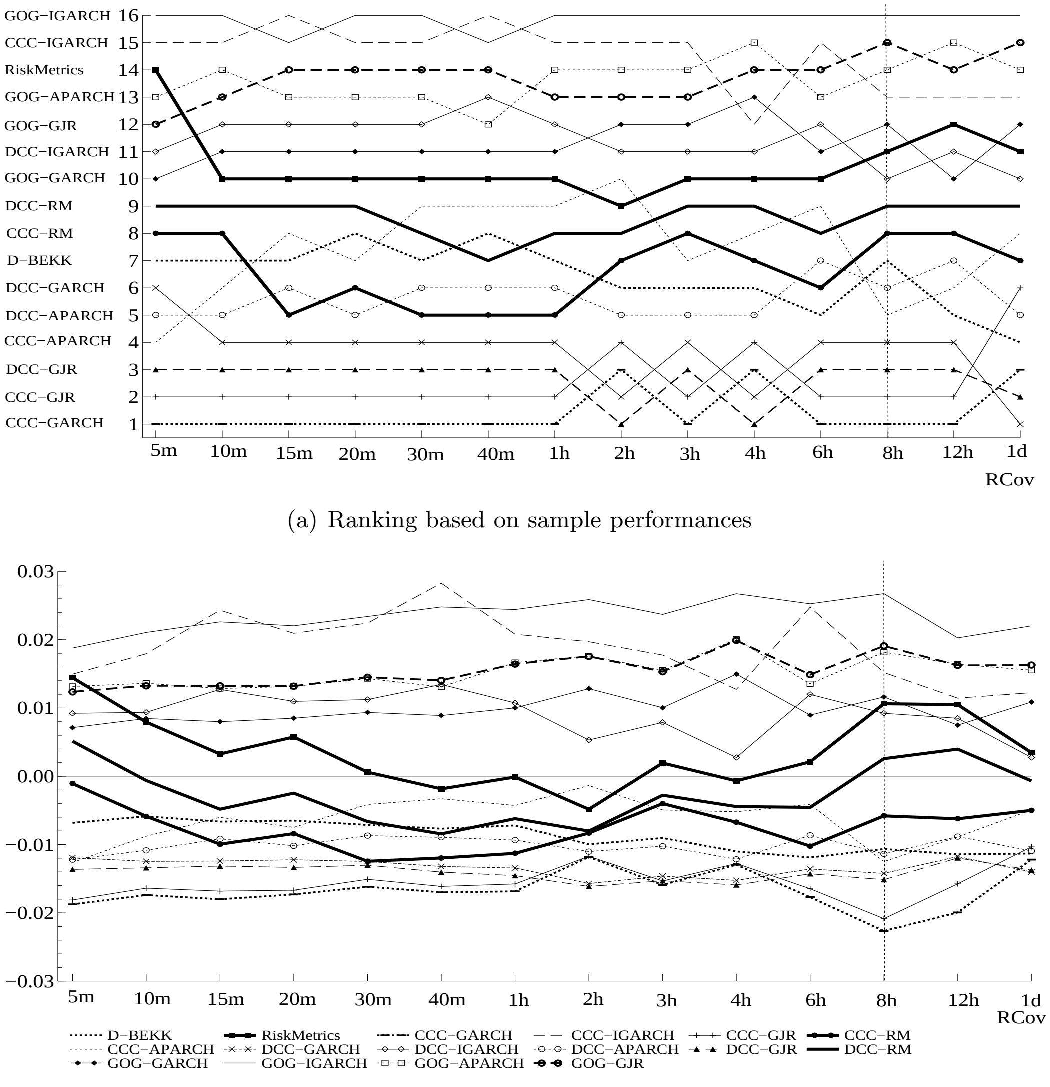
	DBEKK		*CCC*				*DCC*					*GOG*			
			G	*I*	*A*	*J*	*G*	*I*	*A*	*J*		*G*	*I*	*A*	*J*
$\hat{c}_{11}$	0.060 (0.008)	$\hat{\rho}_{21}$	0.710 (0.011)	0.729 (0.010)	0.710 (0.011)	0.710 (0.011)	0.706 (0.055)	0.758 (0.057)	0.707 (0.053)	0.708 (0.054)	$\hat{\varphi}_1$	0.086 (0.058)	0.084 (0.055)	0.111 (0.005)	0.090 (0.059)
$\hat{c}_{21}$	0.044 (0.007)	$\hat{\rho}_{31}$	0.511 (0.016)	0.545 (0.015)	0.518 (0.015)	0.511 (0.016)	0.643 (0.064)	0.732 (0.059)	0.613 (0.067)	0.645 (0.064)	$\hat{\varphi}_2$	0.179 (0.044)	0.185 (0.050)	0.167 (0.012)	0.182 (0.044)
$\hat{c}_{13}$	0.039 (0.007)	$\hat{\rho}_{32}$	0.430 (0.017)	0.462 (0.017)	0.432 (0.017)	0.430 (0.017)	0.511 (0.076)	0.608 (0.073)	0.480 (0.077)	0.511 (0.076)	$\hat{\varphi}_3$	0.417 (0.086)	0.431 (0.088)	0.376 (0.008)	0.416 (0.087)
$\hat{c}_{22}$	0.041 (0.006)	$\hat{\vartheta}_p$					0.026 (0.003)	0.026 (0.003)	0.025 (0.003)	0.025 (0.003)	$\hat{c}$	0.006 (0.003)	0.002 (0.001)	0.006 (0.004)	0.007 (0.004)
$\hat{c}_{23}$	0.006 (0.004)	$\hat{\vartheta}_q$					0.971 (0.004)	0.971 (0.004)	0.971 (0.004)	0.971 (0.004)	$\hat{\alpha}$	0.033 (0.009)	0.035 (0.008)	0.033 (0.011)	0.041 (0.013)
$\hat{c}_{33}$	0.050 (0.008)	$\hat{c}$	0.013 (0.004)	0.003 (0.001)	0.018 (0.005)	0.014 (0.004)	0.007 (0.002)	0.002 (0.001)	0.008 (0.002)	0.007 (0.002)	$\hat{\beta}$	0.961 (0.010)		0.959 (0.011)	0.959 (0.012)
$\hat{a}_{11}$	0.192 (0.011)	$\hat{\alpha}$	0.034 (0.006)	0.044 (0.006)	0.045 (0.010)	0.033 (0.006)	0.040 (0.005)	0.044 (0.006)	0.048 (0.007)	0.045 (0.006)	$\hat{\gamma}$			−0.102 (0.087)	−0.013 (0.011)
$\hat{a}_{22}$	0.201 (0.017)	$\hat{\beta}$	0.936 (0.011)		0.933 (0.013)	0.934 (0.012)	0.945 (0.008)		0.943 (0.059)	0.945 (0.009)	$\hat{\delta}$			2.078 (0.419)	
$\hat{a}_{22}$	0.188 (0.014)	$\hat{\gamma}$			0.032 (0.083)	0.005 (0.009)			−0.074 (0.059)	−0.009 (0.007)	$\hat{c}$	0.007 (0.004)	0.005 (0.003)	0.012 (0.005)	0.008 (0.005)
		$\hat{\delta}$			1.314 (0.171)				1.586 (0.214)		$\hat{\alpha}$	0.065 (0.020)	0.068 (0.022)	0.077 (0.018)	0.060 (0.018)
$\hat{g}_{11}$	0.978 (0.003)	$\hat{c}$	0.010 (0.004)	0.003 (0.001)	0.012 (0.007)	0.012 (0.005)	0.004 (0.002)	0.002 (0.001)	0.003 (0.002)	0.004 (0.002)	$\hat{\beta}$	0.930 (0.022)		0.928 (0.076)	0.927 (0.023)
$\hat{g}_{22}$	0.976 (0.004)	$\hat{\alpha}$	0.037 (0.010)	0.045 (0.010)	0.037 (0.014)	0.022 (0.009)	0.040 (0.009)	0.042 (0.009)	0.031 (0.013)	0.032 (0.009)	$\hat{\gamma}$			0.160 (0.076)	0.014 (0.016)
$\hat{g}_{22}$	0.979 (0.003)	$\hat{\beta}$	0.933 (0.019)		0.930 (0.023)	0.930 (0.021)	0.952 (0.012)		0.954 (0.012)	0.953 (0.012)	$\hat{\delta}$			0.909 (0.217)	
		$\hat{\gamma}$			0.194 (0.082)	0.028 (0.013)			0.068 (0.062)	0.011 (0.010)	$\hat{c}$	0.009 (0.004)	0.005 (0.002)	0.007 (0.004)	0.008 (0.004)
		$\hat{\delta}$			1.900 (0.311)				2.306 (0.391)		$\hat{\alpha}$	0.050 (0.012)	0.055 (0.013)	0.040 (0.013)	0.038 (0.011)
		$\hat{c}$	0.009 (0.004)	0.003 (0.001)	0.011 (0.006)	0.009 (0.004)	0.009 (0.004)	0.003 (0.001)	0.012 (0.006)	0.010 (0.005)	$\hat{\beta}$	0.941 (0.014)		0.945 (0.013)	0.944 (0.014)
		$\hat{\alpha}$	0.045 (0.012)	0.049 (0.012)	0.054 (0.013)	0.050 (0.016)	0.045 (0.010)	0.047 (0.009)	0.058 (0.013)	0.052 (0.015)	$\hat{\gamma}$			0.096 (0.058)	0.019 (0.013)
		$\hat{\beta}$	0.939 (0.018)		0.946 (0.015)	0.938 (0.019)	0.937 (0.004)		0.937 (0.018)	0.936 (0.019)	$\hat{\delta}$			2.309 (0.378)	
		$\hat{\gamma}$			−0.225 (0.116)	−0.009 (0.013)			−0.128 (0.086)	−0.013 (0.011)					
		$\hat{\delta}$			0.797 (0.186)				1.269 (0.268)						
L	−8481		−9007	−9085	−8975	−8999	−8469	−8498	−8452	−8463		−8643	−8653	−8623	−8637

Notes: DCC-RM: $\hat{\rho}_{21} = 0.576$ (0.061), $\hat{\rho}_{31} = 0.461$ (0.075), $\hat{\rho}_{32} = 0.300$ (0.094), $\hat{\vartheta}_p = 0.028$ (0.002) and $\hat{\vartheta}_q = 0.969$ (0.002). L=loglikelihood value at the maximum likelihood estimates. Standard errors Standard errors in brackets. Definitions of the models are in Table 3.

and the presence of microstructure noise (high frequencies). It is clear that the accuracy of the volatility proxy plays an important role here. As pointed out by Hansen and Lunde (2006a), we can observe discrepancies between the empirical and the approximated ranking in finite samples (i.e., sampling error). Indeed, as the accuracy of the proxy deteriorates, the loss function becomes less informative. As a result, it is more difficult to identify superior models. This effect becomes more severe when there is a high degree of similarity between models under evaluation.

Figure 4(a) illustrates how the presence of the objective bias can affect the ranking when an inconsistent loss function is used. The overall ordering between models is generally preserved and stable across frequencies with three striking exceptions. The CCC and the DCC models with RM conditional variances rank 5th and 8th respectively at $RCov_{5m}$, but they rapidly climb towards the top of the classification as the frequency for $RCov$ lowers. Starting from $10m$ frequency for $RCov$ they reach the top of the classification, ranking first and second. Interestingly, (see Figure 4(b)), the sample performances of these two models are extremely close, with discrepancies at each frequency ranging between 0 and 0.02. Similarly, the RiskMetrics model, ranking 10th when $RCov_{5m}$ is used, joins the top of the ranking at a relatively high frequency. When $RCov$ is computed using data sampled at a frequency equal or lower then $40m$, the RiskMetrics model ranks 3rd, behind the CCC-RM and DCC-RM models. Given that these models are characterized by a dynamic in the variance structure imposed ex-ante and independent from the data (with the only exception of the DCC-RM for which the parameters of the dynamic correlation are estimated), it is unlikely that such models are the best forecasting models. The presence of a biased ordering is therefore striking. The ranking obtained at low frequencies is in no way compatible with the one obtained when a more accurate proxy is used. Since model performances are close (see Figure 4(b)), the objective bias severely affects the ranking even when the proxy used in the evaluation is based on rather high frequency data.

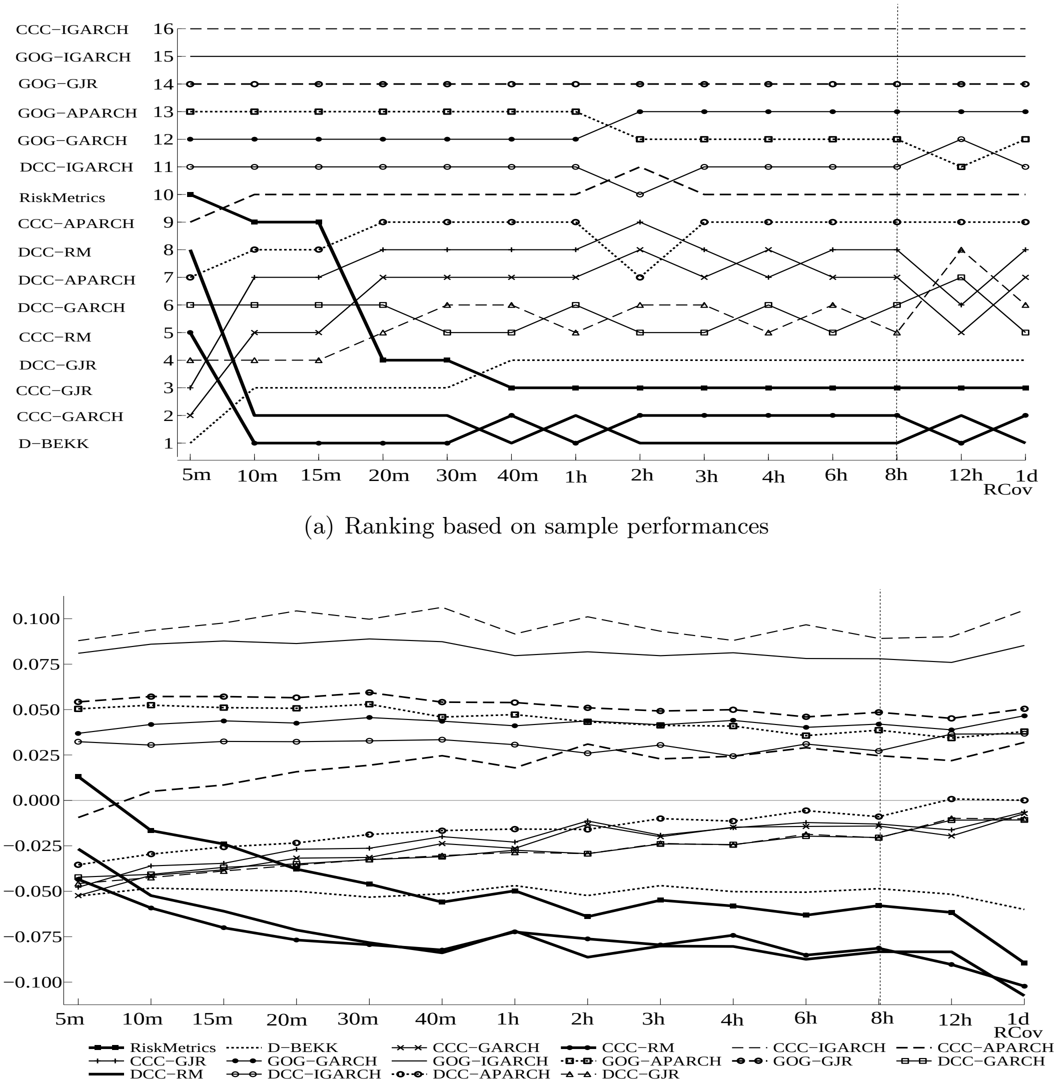
Since the CCC and the DCC models are extremely close in terms of sample performances, in Figures 5 and 6, we concentrate the analysis on a reduced set of models (CCC is excluded). Since we consider models characterized by a lower degree of similarity, the impact of sampling error is now reduced. The ranking implied by the consistent loss function is highly stable for a larger range of frequencies. Again, when the inconsistent loss function is used, the appearance



(a) Ranking based on sample performances

(b) Deviations from the average across models

Figure 3: Estimation results for GBP, EUR and CHF - L_F (consistent)



(a) Ranking based on sample performances

(b) Deviations from the average across models

Figure 4: Estimation results for GBP, EUR and CHF - L_{1M} (inconsistent)

of the objective bias clearly affects the ordering. In Figure 6(b), we observe the relative improvement in terms of sample performances of the DCC-RM and the RiskMetrics models with respect to the other models in the set, with a striking dominance of the DCC-RM.

5.3 Model confidence set

To illustrate the crucial role of an adequate choice of the loss function for model selection based on forecasting ability, we apply the Model Confidence Set (MCS) test of Hansen, Lunde, and Nason (2009) to the reduced set considered in Figures 5 and 6. The MCS test allows to identify a subset of equivalent models in terms of predictive ability which are superior to the others. Note that the MCS depends on the orderings implied by a loss function (e.g., the ranking given in Figures 5(a) and 6(a)). An unfortunate choice of the loss function can result in an incorrect identification of the set of superior models even if the testing procedure is formally valid. Table 8 reports the MCS obtained under L_F (consistent) and L_{1M} (inconsistent) and three choices of the volatility proxy ($RCov_{5m}$, $RCov_{20m}$, $RCov_{8h}$).

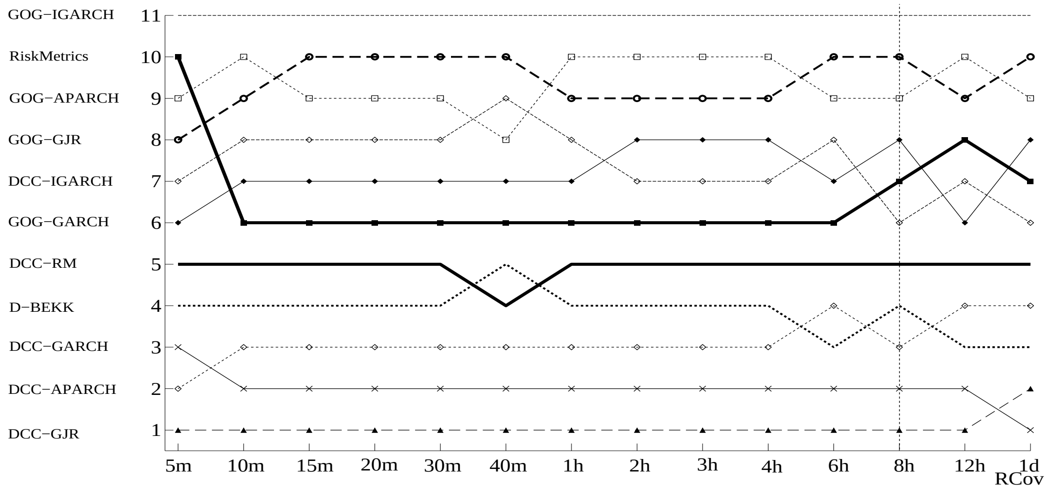
We apply first the MCS test using the L_F. The results reported in Table 8 are consistent across frequencies. Furthermore, the set of equally good models gets larger as the sampling frequency for $RCov$ lowers. This result is due to the loss of accuracy of the proxy which translates into a higher variability of the sample evaluation of each model. Since, at a given confidence level, it is more difficult to discriminate between models, the number of equally good models increases.

Table 8: Model Confidence Set test.

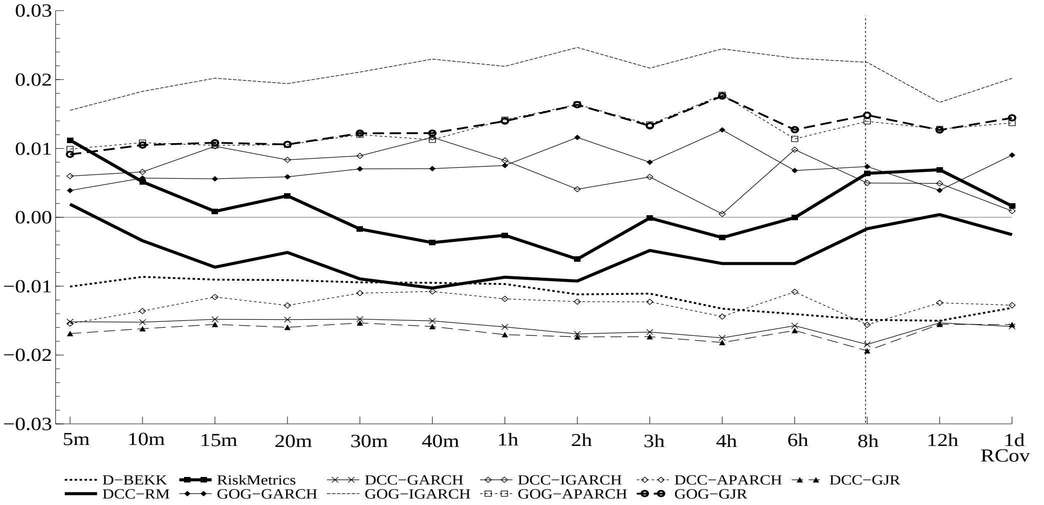
Loss function	$RCov_{5m}$	$RCov_{20m}$	$RCov_{8h}$
Frobenius distance (L_F)	DCC-APARCH DCC-GARCH DCC-GJR	DCC-GARCH DCC-GJR	DCC-APARCH DCC-GARCH DCC-GJR D-BEKK
Entrywise 1 - norm (L_{1M})	DCC-GARCH DCC-GJR D-BEKK	DCC-RM D-BEKK	DCC-RM

Notes: The initial set contains 11 models. Significance level $\alpha = 0.05$.
Sample size 621 obs. Standard errors based on 1000 bootstrap resamples.

Results based on the inconsistent L_{1M} loss function suggest the presence of the objective bias. Indeed, the MCS gets smaller and its composition changes as the frequency for $RCov$ lowers. At $RCov_{8h}$ the set is made up only of the DCC-RM model, which corroborates the

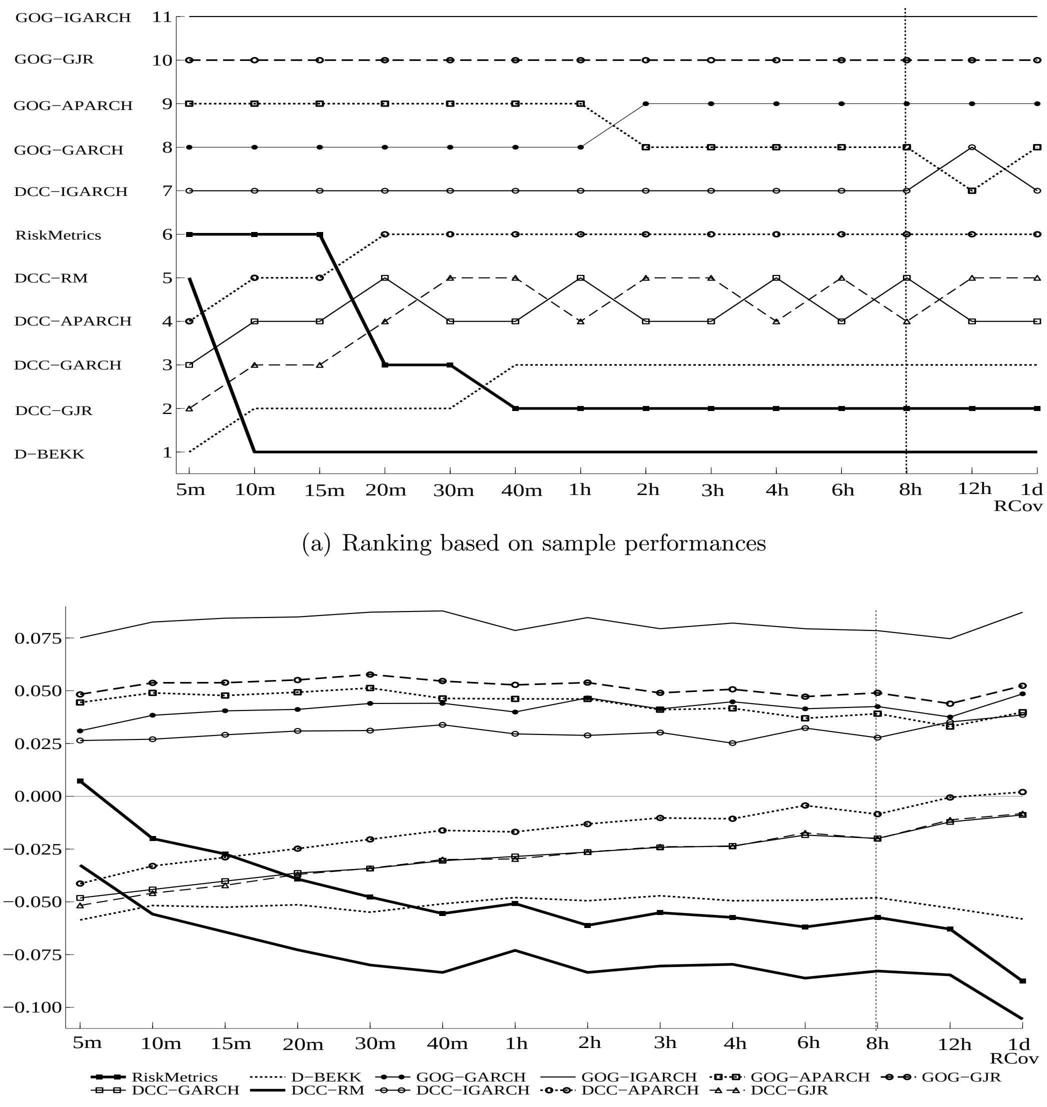


(a) Ranking based on sample performances



(b) Deviations from the average across models

Figure 5: Estimation results for GBP, EUR and CHF (reduced set) - L_F (consistent)



(a) Ranking based on sample performances

(b) Deviations from the average across models

Figure 6: Estimation results for GBP, EUR and CHF (reduced set) - L_{1M} (consistent)

findings in the previous subsection.

6 Conclusion

Two important issues arise when we want to rank several multivariate volatility models with respect to their forecasting performance. First, there is the choice of the loss function (how can we compare predicted variance matrices) and second the choice of a proxy of the unobservable volatility measure used to evaluate models forecasts. In fact, when the unobservable volatility is substituted by a proxy, the ordering implied by a loss function may be biased with respect to the intended one.

In this paper, we first define sufficient conditions for a loss function to satisfy to ensure consistency between the true, but unobservable, ranking - based on the true conditional variance matrix - and the approximated one - based on a proxy. Second, we identify a necessary and sufficient functional form for the loss function to ensure consistent ordering, under the use of a proxy, in matrix and vector spaces. Finally, we provide a large set of consistent parameterization that yield loss functions with different characteristics such as the degree of homogeneity, shape, etc.

In the simulation study, we sample from a continuous time multivariate diffusion process and estimate discrete time multivariate GARCH models to illustrate the sensitivity of the ranking to different choices of the loss functions and to the quality of the proxy. We observe that if the quality of the proxy is sufficiently good, both consistent and inconsistent loss functions rank properly. However, when the quality of the proxy is poor, only the consistent loss functions rank properly. Our findings also hold when the sample size in the estimation period increases. This is an important message for the applied econometrician.

The application to three foreign exchange rates nicely illustrates, in an out-of-sample forecast comparison among 16 multivariate GARCH models, to what extent the ranking and the Model Confidence Set test are affected when we combine an uninformative proxy with an inconsistent loss function.

There are several extensions for future research. First, this paper ranks multivariate volatility models based on statistical loss functions and focuses on conditions for consistent ranking from a more theoretical viewpoint. At some point an economic loss function has to be introduced when the forecasted volatility matrices are actually used in financial applications such as portfolio management and option pricing. It is clear that the model with the smallest statistical loss is always preferred but it may happen that other models with small statistical losses become indistinguishable in terms of economic loss. This issue has not been addressed in this paper. Second, multivariate volatility forecast comparison for higher horizons than one day is not studied yet. Third, other proxies than realized covariance that enter the loss functions should be further investigated.

Appendix A: Proofs

Proof of Proposition 1. To illustrate the validity of Proposition 1, consider the second order Taylor expansion of $L(\hat{\Sigma}_t, H_t)$ around the true value Σ_t:

$$L(\hat{\Sigma}_t, H_t) \cong L(\Sigma_t, H_t) + \left(\frac{\partial L(\Sigma_t, H_t)}{\partial \sigma_t}\right)' (\hat{\sigma}_t - \sigma_t) + \frac{1}{2}\left[(\hat{\sigma}_t - \sigma_t)' \frac{\partial^2 L(\Sigma_t, H_t)}{\partial \sigma_t \partial \sigma_t'} (\hat{\sigma}_t - \sigma_t)\right]. \quad (23)$$

Taking conditional expectations with respect to $\Im_{t-1}$ we get

$$E_{t-1}[L(\hat{\Sigma}_t, H_t)] \cong L(\Sigma_t, H_t) + \frac{1}{2}\left[E_{t-1}\left(\xi_t' \frac{\partial^2 L(\Sigma_t, H_t)}{\partial \sigma_t \partial \sigma_t'} \xi_t\right)\right], \quad (24)$$

because, under A2.2 and A2.4 and when Proposition 1 is satisfied, we have:

(a) $E_{t-1}\left[\left(\frac{\partial L(\Sigma_t,H_t)}{\partial \sigma_t}\right)' \xi_t\right] = \left(\frac{\partial L(\Sigma_t,H_t)}{\partial \sigma_t}\right)' E_{t-1}(\xi_t) = 0$, i.e., $\hat{\sigma}_t$ is conditionally unbiased with respect to σ_t;

(b) for all m, $\left(\frac{\partial^2 L(\Sigma_t,H_{m,t})}{\partial \sigma_t \partial \sigma_t'}\right) = \Psi(\sigma_t^2, .)$, i.e., the last term in (24) does not depend on model m.

Hence $E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]$ and $L(\Sigma_t, H_t)$ induce the same ordering over m.
To conclude, (24) implies that in order to achieve consistency between the approximated and the true ranking, the equivalence between $E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]$ and $L(\Sigma_t, H_t)$ is not required, but it suffices that the discrepancy, $E_{t-1}\left(\xi_t' \Psi(\sigma_t^2, .)\xi_t\right)$, is constant across models, thus not affecting the ranking. ■

Proof of Proposition 2. Under assumptions A2.1 to A2.4, considering (24), the first order conditions in (3) are

$$\frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t^{(s)}, H_t)\right]}{\partial h_{k,t}} - \frac{\partial L(\Sigma_t, H_t)}{\partial h_{k,t}} \cong \frac{1}{2}\left[\frac{\partial}{\partial h_{k,t}} E_{t-1}\left(\xi_t^{(s)'} \Psi(\sigma_t^2, h_t)\xi_t^{(s)}\right)\right] \quad (25)$$

$$\begin{aligned}
&\cong \frac{1}{2}\frac{\partial}{\partial h_{k,t}} E_{t-1}\left[\sum_{l,m} \xi^{(s)}_{l,t}\xi^{(s)}_{m,t}\Psi(\sigma^2_t, h_t)_{l,m}\right] && (26)\\
&\cong \frac{1}{2}\sum_{l,m}\frac{\partial \Psi(\sigma^2_t, h_t)_{lm}}{\partial h_{k,t}} E_{t-1}[\xi^{(s)}_{l,t}\xi^{(s)}_{m,t}] && (27)\\
&\cong \frac{1}{2}\sum_{l,m}\frac{\partial \Psi(\sigma^2_t, h_t)_{lm}}{\partial h_{k,t}} V^{(s)}_{l,m,t} && (28)
\end{aligned}$$

for all s, with $l, m = 1, ..., N(N+1)/2$, $k = 1, ..., N(N+1)/2$ and where $V^{(s)}_{l,m,t} = E_{t-1}[\xi^{(s)}_{l,t}\xi^{(s)}_{m,t}]$ and $\Psi(\sigma^2_t, h_t)_{l,m}$ represent respectively the element $[l, m]$ of the variance matrix of the proxy $V^{(s)}_t = E_{t-1}[\xi^{(s)}_t \xi^{(s)\prime}_t]$ and of $\Psi(\sigma^2_t, h_t)$, the matrix of second derivatives of $L(.,.)$ with respect to σ^2_t.

The first order conditions imply that $H^{*(s)}_t$ is the solution of

$$\frac{\partial E_{t-1}\left[L(\hat{\Sigma}^{(s)}_t, H^{*(s)}_t)\right]}{\partial h_{k,t}} = 0 \ \forall k$$

and, under A2.3, A1.1 ensures that second order conditions are satisfied. Then, we have that

$$-\frac{\partial L(\Sigma_t, H^{*(s)}_t)}{\partial h_{k,t}} \cong \frac{1}{2}\sum_{l,m}\frac{\partial \Psi(\sigma^2_t, .)_{lm}}{\partial h_{k,t}} V^{(s)}_{l,m,t}. \qquad (29)$$

Under $i)$, i.e., $\frac{\partial \Psi(\sigma^2_t,.)_{lm}}{\partial h_{k,t}} = 0 \ \forall k$, the first order conditions of the loss function based on the proxy lead to the same optimal forecast as if the true variance matrix was observable, even in presence of a noisy volatility proxy. From A1.2 it follows that

$$\frac{\partial L(\Sigma_t, H^{*(s)}_t)}{\partial h_{k,t}} = 0 \Leftrightarrow H^{*(s)}_t = \Sigma_t \ \forall s,$$

that is the optimal forecast equals the conditional variance. By assumption A1.2, A2.2 and A2.4, we also have that $H^*_t = \Sigma_t = E_{t-1}(\hat{\Sigma}_t)$.

Under $ii)$, i.e., $\frac{\partial \Psi(\sigma^2_t,h_t)_{lm}}{\partial h_{k,t}} \neq 0$ for some k, then as $s \to \infty$, by A2.5 and (29) we have

$$\frac{\partial L(\Sigma_t, H^{*(s)}_t)}{\partial h_{k,t}} \xrightarrow{p} 0 \Leftrightarrow H^{*(s)}_t \xrightarrow{p} \Sigma_t \ \forall s$$

which concludes the proof. ■

Proof of Proposition 3. To prove the proposition, we proceed as in Patton (2009). We show the equivalence of the following statements:

-S1: the loss function takes the form in the proposition;

-S2: the loss function is consistent in the sense of Definition 2;

-S3: the optimal forecast under the loss function is the conditional variance matrix.

Step 1: S1⇒S2. The result follows directly form Proposition 1, in fact:

$$\frac{\partial^2 L(\Sigma_t, H_t)}{\partial \sigma_t \partial \sigma_t'} = \nabla^2 \tilde{C}(\Sigma_t) = \Psi(\sigma_t^2, .)$$

since $\frac{\partial^2 (C(H_t)'\sigma_t)}{\partial \sigma_t \partial \sigma_t'} = 0$, and does not depend on H_t.

Step 2: S2⇒S3. By assumption A3.2, there exists an H_t^* in the support of $L(\hat{\Sigma}_t, H_t)$ such that $H_t^* = E_{t-1}(\hat{\Sigma}_t)$. This implies that $\forall H_t \in int(\dot{H}) \setminus \{H_t^*\}$:

$$E_{t-1}\left[L(\hat{\Sigma}_t, H_t^*)\right] \quad \leq \quad E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]$$

and therefore by the law of iterated expectations:

$$E\left[L(\hat{\Sigma}_t, H_t^*)\right] \quad \leq \quad E\left[L(\hat{\Sigma}_t, H_t)\right].$$

Then by Definition 2, under S2, we can write

$$E(L(\hat{\Sigma}_t, H_t^*)) \leq E(L(\hat{\Sigma}_t, H_t)) \Leftrightarrow E(L(\Sigma_t, H_t^*)) \leq E(L(\Sigma_t, H_t))$$

if we set $H_t = \Sigma_t$, then by assumptions A1.1 to A1.3, $E(L(\Sigma_t, \Sigma_t)) = 0 \Rightarrow E(L(\Sigma_t, H_t^*)) = 0$ and therefore $H_t^* = \Sigma_t$.

Step 3: S1⇔S3. The last step uses the arguments of Gourieroux and Monfort (1995), which prove sufficiency and necessity of the linear exponential functional form for the pseudo true density to prove consistency of the pseudo maximum likelihood estimator.

First, we prove sufficiency (S1⇒S3). Consider the first order conditions evaluated at the optimum ($H_t = H_t^*$), that is

$$\begin{aligned}\frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]}{\partial h_t} &= C(H_t^*) + \nabla^2 \tilde{C}(H_t) vech(E_{t-1}(\hat{\Sigma}_t) - H_t^*) - C(H_t^*) = 0 \\ &= \nabla^2 \tilde{C}(H_t) vech(E_{t-1}(\hat{\Sigma}_t) - H_t^*) = 0 \\ &\Leftrightarrow E_{t-1}(\hat{\Sigma}_t) = H_t^*.\end{aligned}$$

Second, to prove necessity (S3⇒S1), consider that at the optimum we must have $E_{t-1}(\hat{\Sigma}_t) = H_t^*$, and consequently

$$E_{t-1}\left(\frac{\partial L(\hat{\Sigma}_t, H_t^*)}{\partial h_t}\right) = 0,$$

for any conditional distribution $F_t \in F$.

Applying Lemma 8.1 in Gourieroux and Monfort (1995), there exists a square matrix Λ of size $k = N(N+1)/2$ which is only function of H_t^* such that

$$\frac{\partial L(\hat{\Sigma}_t, H_t^*)}{\partial h_t} = \Lambda(H_t^*) vech(\hat{\Sigma}_t - H_t^*). \tag{30}$$

Since we want to ensure that H_t^* is the minimizer of $L(\hat{\Sigma}_t, H_t^*)$ then we must have $\frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]}{\partial h_t \partial h_t'}$ satisfying second order necessary or sufficient conditions. Using assumption A3.3 we can interchange differentiation and expectation (see L'Ecuyer (1990) and L'Ecuyer (1995) for details) to obtain

$$\begin{aligned}
E_{t-1}\left(\frac{\partial L(\hat{\Sigma}_t, H_t^*)}{\partial h_t \partial h_t'}\right) &= E_{t-1}\left(\frac{\partial \Lambda(H_t^*) vech(\hat{\Sigma}_t - H_t^*)}{\partial h_t}\right) \\
&= E_{t-1}\left(\begin{bmatrix} \sum_{i=1}^{K} \frac{\partial \Lambda(H_t^*)_{1i}}{\partial h_1}(\sigma_i - h_i^*) & \dots & \sum_{i=1}^{K} \frac{\partial \Lambda(H_t^*)_{1i}}{\partial h_k}(\sigma_i - h_i^*) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{K} \frac{\partial \Lambda(H_t^*)_{ki}}{\partial h_1}(\sigma_i - h_i^*) & \dots & \sum_{i=1}^{K} \frac{\partial \Lambda(H_t^*)_{ki}}{\partial h_k}(\sigma_i - h_i^*) \end{bmatrix}\right) - \Lambda(H_t^*) \\
&= -\Lambda(H_t^*),
\end{aligned}$$

with $K = N(N+1)/2$.

Now, it suffices to integrate (30) (up to a constant and/or a term that solely depends on $\hat{\Sigma}_t$) to recover the loss function of the form stated in the proposition. In fact, if we define

$$\Lambda(H_t) = \nabla^2 \tilde{C}(H_t) = C'(H_t),$$

and rewrite (30) as

$$C'(H_t) vech(\hat{\Sigma}_t) - C'(H_t) vech(H_t),$$

we have that

$$\begin{aligned}
C'(H_t) vech(\hat{\Sigma}_t) &= \frac{\partial C(H_t)' vech(\hat{\Sigma}_t)}{\partial h_t} \\
C'(H_t) vech(H_t) &= \frac{\partial C(H_t)' vech(H_t)}{\partial h_t} - C(H_t) \\
&= \frac{\partial C(H_t)' vech(H_t)}{\partial h_t} - \frac{\partial \tilde{C}(H_t)}{\partial h_t}.
\end{aligned}$$

Therefore (30) admits as primitive

$$C(H_t)' vech(\hat{\Sigma}_t) - C(H_t)' vech(H_t) + \tilde{C}(H_t).$$

Rearranging and allowing for a term that depends on $\hat{\Sigma}_t$, we obtain

$$L(\hat{\Sigma}_t, H_t) = \tilde{C}(H_t) + \tilde{C}(\hat{\Sigma}_t) + C'(H_t)vech(\hat{\Sigma}_t - H_t),$$

where $\frac{\partial \tilde{C}(\hat{\Sigma}_t)}{\partial h_t} = 0$, which concludes the proof. ■

Proof of Corollary 1. Since $\hat{\Sigma}_t$ and H_t are symmetric, then

$$\begin{aligned} Tr[\bar{C}(H_t)(\hat{\Sigma}_t - H_t)] &= \sum_i \bar{c}_{i,i}(H_t)(\hat{\sigma}_{i,i,t} - h_{i,i,t}) + 2\sum_{i<j} \bar{c}_{i,j}(H_t)(\hat{\sigma}_{i,j,t} - h_{i,j,t}) \text{ for } i,j = 1,...,N \\ &= \sum_i \frac{\partial \tilde{C}(H_t)}{\partial h_{i,i,t}}(\hat{\sigma}_{i,i,t} - h_{i,i,t}) + 2\sum_{i<j} \frac{1}{2}\frac{\partial \tilde{C}(H_t)}{\partial h_{i,j,t}}(\hat{\sigma}_{i,j,t} - h_{i,j,t}) \\ &= C(H_t)'vech(\hat{\Sigma}_t - H_t), \end{aligned}$$

with $C(H_t)'$ as defined in Proposition 2. ■

Proof of Remark 1. The proof of part *i)* of the Remark follows from Proposition 3.

For the second part, notice that

$$\frac{\partial^2 L(\Sigma_t, H_t)}{\partial \sigma_t \partial \sigma_t'} = -\tilde{C}''_{\sigma_t}(\Sigma_t),$$

since if $f[\cdot]$ is a linear map, then $f[\bar{C}(H_t)(\Sigma_t - H_t)]$ is linear in $\sigma_{i,j,t}$ $\forall i,j = 1,...,N$. Hence, the general conclusion of Proposition 1 holds even under violation of A1.2: the ordering implied by $E_{t-1}[L(\hat{\Sigma}_t, H_t)]$ is apparently consistent for the one based on $L(\Sigma_t, H_t)$ in the sense that is insensitive to the substitution of the true variance matrix by a proxy (by the same reasoning provided in the proof of Proposition 1), i.e., $\arg\min_{H_t \in \dot{H}} L(\Sigma_t, H_t) = \arg\min_{H_t \in \dot{H}} E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]$.

We now show that, though apparently consistent, the ordering obtained when $f[\cdot] \not\equiv Tr[\cdot]$ is not a valid one, that is it differs from any valid or acceptable ordering and in particular it holds $H_t^* \neq E_{t-1}(\hat{\Sigma}_t) = \Sigma_t$.

Consider the first order conditions of (6) evaluated at the optimum H_t^*, that is

$$\frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]}{\partial h_t} = C(H_t^*) + f'_{h_t}[\bar{C}(H_t^*)(E_{t-1}(\hat{\Sigma}_t) - H_t^*)] = 0. \tag{31}$$

Recall that $C(H_t) = \nabla \tilde{C}(H_t)$ and f'_{h_t} is the gradient of f with respect to h_t. Using the fact that f is a linear map, the typical element of the gradient of $E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]$, indexed by $i,j = 1,...,N$, $i \leq j$ is (we omit the time index to simplify notation)

$$\frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]}{\partial h_{i,j}} = \tilde{C}'_{h_{i,j}}(H_t^*) + f\left[\frac{\partial \bar{C}(H_t^*)}{\partial h_{i,j}}(E_{t-1}(\hat{\Sigma}_t) - H_t^*)\right] - f\left[\bar{C}(H_t^*)\frac{\partial H_t^*}{\partial h_{i,j}}\right] = 0. \tag{32}$$

To deliver an appropriate ordering, the loss function must be such that it is uniquely minimized at $H_t^* = E_{t-1}(\hat{\Sigma}_t) = \Sigma_t$, that is optimal forecast is the true conditional variance. Therefore, it must be the case that

$$\frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t^*, H_t)\right]}{\partial h_{i,j}} = f\left[\frac{\partial \bar{C}(H_t^*)}{\partial h_{i,j}}(E_{t-1}(\hat{\Sigma}_t - H_t^*)\right] = 0.$$

Therefore, in (6), it must hold

$$f\left[\bar{C}(H_t^*)\frac{\partial H_t^*}{\partial h_{i,j}}\right] = \tilde{C}'_{h_{i,j}}(H_t^*). \tag{33}$$

Since $\frac{\partial H_t^*}{\partial h_{i,j}}$, for all $i, j = 1, ..., N$ $i \leq j$, is a $N \times N$ symmetric matrix with elements indexed by $[i, j]$ and $[j, i]$ equal to 1 and zero elsewhere, (33) holds if and only if $f(.) = Tr(.)$. In fact, from (33)

$$\begin{aligned} i &= j \Longrightarrow Tr\left[\bar{C}(H_t^*)\frac{\partial H_t^*}{\partial h_{i,i}}\right] = \bar{c}_{i,i}(H_t^*) = \tilde{C}'_{h_{i,i}}(H_t^*) \\ i &\neq j \Longrightarrow Tr\left[\bar{C}(H^*)\frac{\partial H^*}{\partial h_{i,j}}\right] = 2\bar{c}_{i,j}(H_t^*) = \tilde{C}'_{h_{i,j}}(H_t^*). \end{aligned} \tag{34}$$

Substituting (34) in (32), we obtain

$$\begin{aligned} \frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t^*, H_t)\right]}{\partial h_{i,j}} &= \tilde{C}'_{h_{i,j}}(H_t^*) + Tr\left[\frac{\partial C(H_t^*)}{\partial h_{ij}}(E_{t-1}(\hat{\Sigma}_t) - H_t^*)\right] - \tilde{C}'_{h_{ij}}(H_t^*) \\ &= Tr\left[\frac{\partial C(H_t^*)}{\partial h_{i,j}}(\hat{\Sigma}_t - H_t^*)\right], \end{aligned}$$

and finally

$$\begin{aligned} \frac{\partial E_{t-1}\left[L(\hat{\Sigma}_t, H_t)\right]}{\partial h_t} &= \begin{bmatrix} Tr\left[\frac{\partial \bar{C}(H_t^*)}{\partial h_{1,1}}(E_{t-1}(\hat{\Sigma}_t) - H_t^*)\right] \\ ... \\ Tr\left[\frac{\partial \bar{C}(H_t^*)}{\partial h_{i,j}}(E_{t-1}(\hat{\Sigma}_t) - H_t^*)\right] \\ ... \end{bmatrix} = 0 \\ &\Leftrightarrow \quad H_t^* = E_{t-1}(\hat{\Sigma}_t), \end{aligned}$$

which concludes the proof. ■

Proof of Proposition 4. By Proposition 2, a consistent loss functions based on the forecast error must have the form

$$L(\hat{\Sigma}_t - H_t) = \tilde{C}(H_t) - \tilde{C}(\hat{\Sigma}_t) + C(H_t)'vech(\hat{\Sigma}_t - H_t). \tag{35}$$

Consider

$$\begin{aligned} \frac{\partial L(\hat{\Sigma}_t - H_t)}{\partial h_t} &= \nabla^2 \tilde{C}(H_t) vech(\hat{\Sigma}_t - H_t) \qquad (36) \\ \frac{\partial L(\hat{\Sigma}_t - H_t)}{\partial \sigma_t} &= C(H_t) - C(\hat{\Sigma}_t). \qquad (37) \end{aligned}$$

Note that since the loss function is only based on the forecast error then $L(\hat{\Sigma}_t - H_t)$ then $L(\hat{\Sigma}_t - H_t) = L(H_t - \hat{\Sigma}_t)$, i.e., $L(.,.)$ is symmetric under 180° rotation around the origin and, which implies

$$-\frac{\partial L(\hat{\Sigma}_t - H_t)}{\partial h_t} = \frac{\partial L(\hat{\Sigma}_t - H_t)}{\partial \sigma_t}, \qquad (38)$$

and therefore

$$\nabla^2 \tilde{C}(H_t) vech(\hat{\Sigma}_t - H_t) = C(H_t) - C(\hat{\Sigma}_t),$$

for all $\hat{\Sigma}_t$ and H_t. Differentiating both sides of (38) with respect to σ_t we obtain

$$\nabla^2 \tilde{C}(H_t) = \nabla^2 \tilde{C}(\hat{\Sigma}_t),$$

which implies

$$\nabla^2 \tilde{C}(H_t) = \Lambda, \qquad (39)$$

where Λ is a matrix of constants.

Equation (39) implies that $C(H_t) = \nabla^2 \tilde{C}(H_t) vech(H_t)$ is homogeneous of degree 1, and hence $\tilde{C}(\cdot)$ is homogeneous of degree 2 then so is $L(\hat{\Sigma}_t - H_t)$. Applying Euler theorem for homogeneous functions we have that $2\tilde{C}(H_t) = C(H_t)' vech(H_t)$. The loss function in (35) can be rewritten as

$$L(\hat{\Sigma}_t - H_t) = -\tilde{C}(H_t) - \tilde{C}(\hat{\Sigma}_t) + C(H_t)' vech(\hat{\Sigma}_t). \qquad (40)$$

In order to satisfy second order conditions Λ must be negative definite, according to Proposition 3. Since $L(\hat{\Sigma}_t, H_t)$ is homogeneous of degree 2, starting from (39), we can apply Euler theorem for homogeneous functions and obtain

$$\begin{aligned} C(H_t) &= \Lambda vech(H_t) \qquad (41) \\ \tilde{C}(H_t) &= \frac{1}{2} vech(H_t)' \Lambda vech(H_t). \end{aligned}$$

Substituting the expression for $\tilde{C}(.)$ in (40) and rearranging we obtain the quadratic loss

$$\begin{aligned} L(\hat{\Sigma}_t - H_t) &= -\frac{1}{2} vech(\hat{\Sigma}_t - H_t)' \Lambda vech(\hat{\Sigma}_t - H_t) \\ &= vech(\hat{\Sigma}_t - H_t)' \hat{\Lambda} vech(\hat{\Sigma}_t - H_t). \end{aligned}$$

with $\hat{\Lambda} = -\frac{1}{2}\Lambda$. $\blacksquare$

Appendix B: Examples for Section 2.4

In the following examples, for ease of exposition, we consider a forecast error matrix of dimension $N = 2$.

In the first three, examples we investigate the properties of loss functions belonging to the family of quadratic loss functions defined in Proposition 4. The vector of forecast errors of interest is therefore

$$vech(\Sigma_t - H_t) = \begin{pmatrix} \sigma_{1,1,t} - h_{1,1,t} \\ \sigma_{1,2,t} - h_{1,2,t} \\ \sigma_{2,2,t} - h_{2,2,t} \end{pmatrix},$$

which allows to plot contours of the loss function.

The first loss function that we consider is the Euclidean distance, which corresponds to a choice of $\hat{\Lambda} = I_K$ and can be expressed as

$$L_E = (\sigma_{1,1,t} - h_{1,1,t})^2 + (\sigma_{1,2,t} - h_{1,2,t})^2 + (\sigma_{2,2,t} - h_{2,2,t})^2.$$

Figure 7 reports the contour of $L_E = 1$.

The contours of L_E are spheres centered at the origin. The loss function has mirror symmetry about all coordinate planes. It is also symmetric under any rotation about the origin and, being a symmetric polynomial, it is symmetric about the bisector planes.

The second loss function is the weighted Euclidean distance with

$$\hat{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

which implies

$$L_{WE} = (\sigma_{1,1,t} - h_{1,1,t})^2 + 4(\sigma_{1,2,t} - h_{1,2,t})^2 + 2(\sigma_{2,2,t} - h_{2,2,t})^2,$$

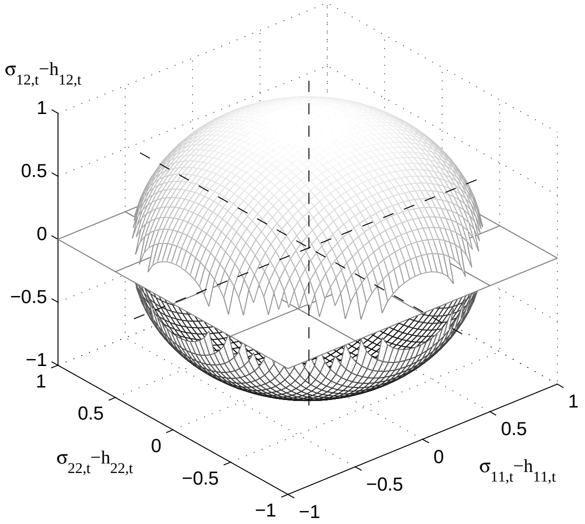


Figure 7: Euclidean distance - $L_E = 1$

which implies that $(\sigma_{2,2,t} - h_{2,2,t})$ is penalized twice with respect to $(\sigma_{1,1,t} - h_{1,1,t})$, while the covariance forecast error is penalized four times more. The reason behind such a particular choice of $\hat{\Lambda}$ is to emphasize the role of each weight and to show how they affect the shape of the loss function. The contour of $L_{WE} = 1$ is an ellipsoid centered at the origin, see Figure 8. The contour is squeezed around the $(\sigma_{1,1,t} - h_{1,1,t})$ axis due to the unequal weighting. The loss function in symmetric about all coordinate planes and it is also symmetric under a 180° rotation around the origin, i.e., considering the absolute forecast error vector $|\sigma_t - h_t| = (0.2, 0.4, 0.8)$, we have

$$\begin{aligned}
L_{WE}(0.2, 0.4, 0.8) &= L_{WE}(-0.2, -0.4, -0.8) = 1.96 \\
L_{WE}(0.2, 0.4, -0.8) &= L_{WE}(0.2, 0.4, -0.8) = 1.96 \\
L_{WE}(0.2, -0.4, 0.8) &= L_{WE}(0.2, -0.4, 0.8) = 1.96 \\
&\ldots
\end{aligned}$$

However, L_{WE} is not symmetric about the bisector planes, i.e.

$$\begin{aligned}
L_{WE}(0.2, 0.4, 0.8) &= 1.96 \neq 2.92 = L_{WE}(0.2, 0.8, 0.4) \\
&\neq 1.12 = L_{WE}(0.8, 0.2, 0.4) \\
&\ldots
\end{aligned}$$

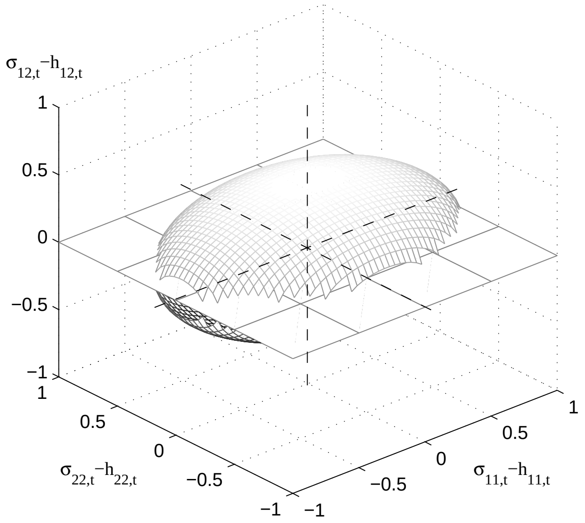


Figure 8: Weighted Euclidean distance - $L_{WE} = 1$

The third loss function is the pseudo Mahalanobis distance with

$$\hat{\Lambda} = \begin{bmatrix} 1 & 0 & 0.6 \\ 0 & 4 & 0 \\ 0.6 & 0 & 2 \end{bmatrix},$$

that is

$$L_M = (\sigma_{1,1,t} - h_{1,1,t})^2 + 4(\sigma_{1,2,t} - h_{1,2,t})^2 + 2(\sigma_{2,2,t} - h_{2,2,t})^2 + 1.2(\sigma_{1,1,t} - h_{1,1,t})(\sigma_{2,2,t} - h_{2,2,t}).$$

For illustrative purposes, we set only one off diagonal element of the matrix of weights different from 0. As in the previous case, the contour of $L_M = 1$ is an ellipsoid centered at the origin (Figure 9). It is clear that L_M is only symmetric under a 180° around the origin. Furthermore, the axes of symmetry (dashed lines in Figure 9), whose directions depend on the sign of the off diagonal elements of $\hat{\Lambda}$, are rotated with respect to the coordinate axes (e.g. in Figure 9, $\hat{\Lambda}$ implies an horizontal rotation). In this regard, since the loss function also includes the cross product of the elements of $(\sigma_t - h_t)$ weighted by the off diagonal elements of $\hat{\Lambda}$ (which can be positive and/or negative provided $\hat{\Lambda}$ satisfies Proposition 4), a positive weight means that, for given absolute forecast errors $|\sigma_t - h_t|$, L_M will penalize more the outcomes where both variances are over/under predicted. In fact, consider L_M evaluated at $|\sigma_t - h_t| = (0.8, 0, 0.4)$, then

$$\begin{aligned} L_M(0.8, 0, -0.4) &= L_M(-0.8, 0, 0.4) = 0.576 \\ L_M(0.8, 0, 0.4) &= L_M(-0.8, 0, -0.4) = 1.344. \end{aligned}$$

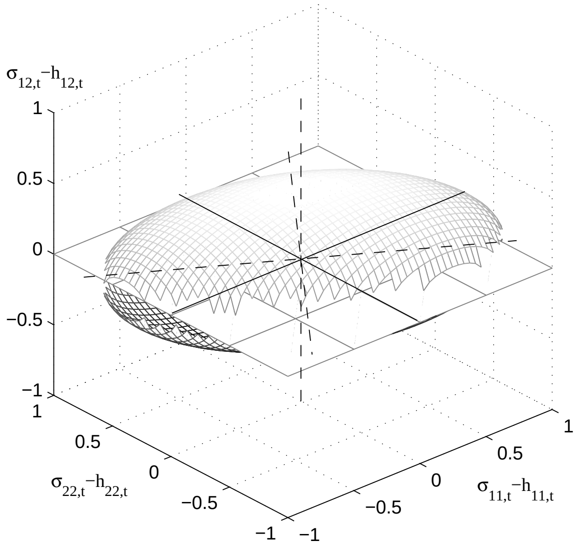


Figure 9: Pseudo Mahalanobis distance - $L_M = 1$

In a final example, we focus on the Stein loss function. Note that a comprehensive illustration of the geometric properties of L_S is more complex then in the previous cases. We have shown that quadratic loss functions are defined on the forecast error matrix $\Sigma_t - H_t$ which implies that $L(\Sigma_t - H_t) : R^{N\times N} \to R_+$ even if Σ_t and $H_t \in R^{N\times N}_{++}$ (the space of positive definite matrices). This allows for a graphical representation of the forecast error vector, i.e., the vector of unique elements of $\Sigma_t - H_t$, in the space $R^{N(N+1)/2}$. On the other hand, L_S is defined on the standardized (in matrix sense) forecast error $\Sigma_t H_t^{-1}$ which is positive definite. Since the domain of $L(\Sigma_t H_t^{-1})$ is $R^{N\times N}_{++} \subset R^{N\times N}$, the graphical representation of the contours in the Euclidean space is difficult. Furthermore, unlike the loss functions based on the forecast error matrix, L_S cannot be expressed as a combination of functions of the elementwise forecast errors, i.e., $L(\Sigma_t, H_t) = L(l(\sigma_{1,t}, h_{1,t}), ..., l(\sigma_{K,t}, h_{K,t}))$, except in the trivial case when Σ_t and H_t are diagonal. Therefore, to illustrate the properties of the Stein loss function, we rely on some numerical examples and the analysis of conditional loss.

Consider a standardized forecast error matrix of dimension $N = 2$.

$$\Sigma_t H_t^{-1} = \begin{bmatrix} \frac{\sigma_{1,2,t}h_{1,2,t}-\sigma_{1,1,t}h_{2,2,t}}{h_{1,2,t}^2-h_{1,1,t}h_{2,2,t}} & \frac{\sigma_{1,1,t}h_{1,2,t}-\sigma_{1,2,t}h_{1,1,t}}{h_{1,2,t}^2-h_{1,1,t}h_{2,2,t}} \\ \frac{\sigma_{2,2,t}h_{1,2,t}-\sigma_{1,2,t}h_{2,2,t}}{h_{1,2,t}^2-h_{1,1,t}h_{2,2,t}} & \frac{\sigma_{1,2,t}h_{1,2,t}-\sigma_{2,2,t}h_{1,1,t}}{h_{1,2,t}^2-h_{1,1,t}h_{2,2,t}} \end{bmatrix}.$$

The Stein loss function defined in (14) is therefore

$$L_S = \frac{\sigma_{1,1,t}h_{2,2,t} + \sigma_{2,2,t}h_{1,1,t} - 2\sigma_{1,2,t}h_{1,2,t}}{h_{1,1,t}h_{2,2,t} - h_{1,2,t}^2} - \ln\left(\frac{\sigma_{1,1,t}\sigma_{2,2,t} - \sigma_{1,2,t}^2}{h_{1,1,t}h_{2,2,t} - h_{1,2,t}^2}\right) - 2$$

For ease of exposition, we set Σ_t to some arbitrary values, say

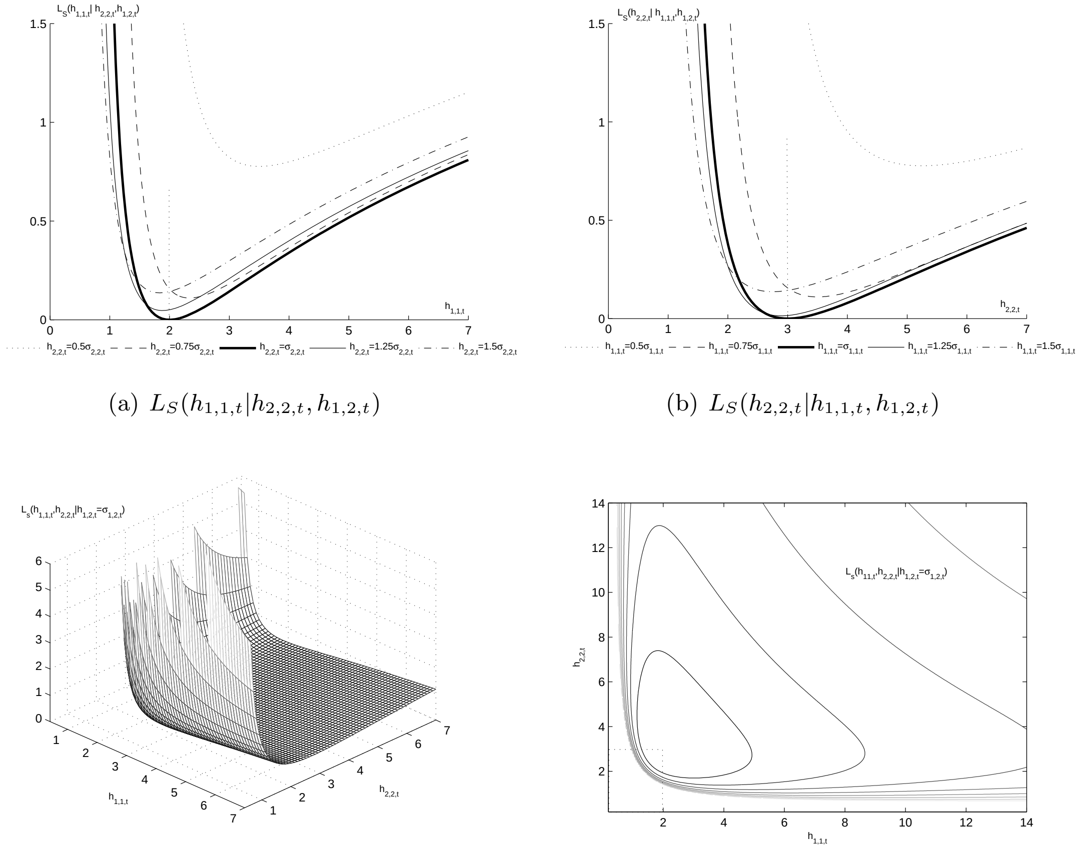
$$\Sigma_t = \begin{bmatrix} 2 & 1.5 \\ 1.5 & 3 \end{bmatrix}.$$

Since the loss function is expressed in terms of standardized forecast errors, we first assess the case of over/under prediction of size $\pm 0.5\Sigma_t$. The loss when each element of H_t over/under predicts the corresponding element of Σ_t (setting the others at their optimal values), is

	$(-)$	$(+)$
$L_S(h_{1,1,t} = 2 \pm 1)$	2.390	0.143
$L_S(h_{2,2,t} = 3 \pm 1.5)$	2.390	0.143
$L_S(h_{1,2,t} = 1.5 \pm 0.75)$	2.213	0.164
$L_S(H_t = (1 \pm 0.5)\Sigma_t)$	0.613	0.144

The Stein loss function is therefore asymmetric with respect to over/under predictions, and, in particular, underpredictions are heavily penalized. However, the conditional losses with respect to the variances are symmetric up to a proportionality constant. Figure 10(a) and 10(b) report L_S as a function of $h_{1,1,t}$ for several values of $h_{2,2,t}$ (with $h_{1,2,t} = \sigma_{1,2,t}$). Figure 10(c) reports L_S as a function of both $h_{1,1,t}$ and $h_{2,2,t}$, given $h_{1,2,t} = \sigma_{1,2,t}$, while Figure 10(d) reports the contours of the representation in Figure 10(c).

Of particular interest is the representation of L_S as a function of the covariance. Note that, if for any given $h_{1,1,t}$ and $h_{2,2,t}$, then $h_{1,2,t} = \rho\sqrt{h_{1,1,t}h_{2,2,t}}$ with $\rho \in (-1, 1)$. The domain of the conditional loss of $h_{1,2,t}$ is therefore centered at 0. Its representation is given in Figure 11 using the values suggested above. Finally, note that the conditional loss in Figure 11 is symmetric about the vertical axis only in the trivial case where Σ_t is diagonal.



(a) $L_S(h_{1,1,t}|h_{2,2,t},h_{1,2,t})$

(b) $L_S(h_{2,2,t}|h_{1,1,t},h_{1,2,t})$

(c) $L_s(h_{1,1,t},h_{2,2,t}|h_{1,2,t}=\sigma_{1,2,t})$ (loss)

(d) $L_s(h_{1,1,t},h_{2,2,t}|h_{1,2,t}=\sigma_{1,2,t})$ (contours)

Figure 10: Stein loss function

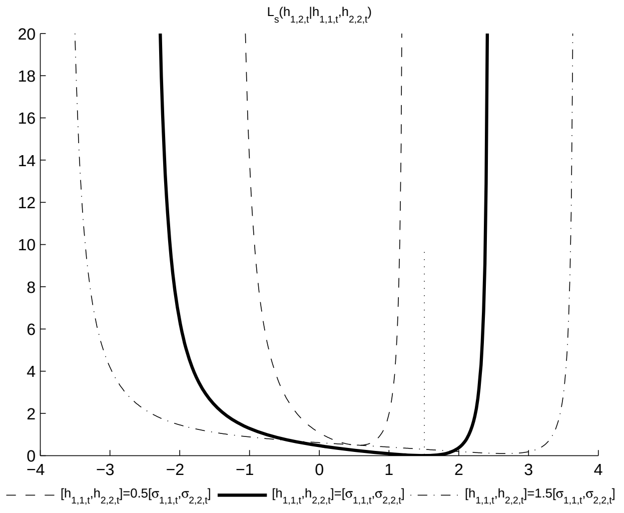


Figure 11: $L_s(h_{1,2,t}|h_{1,1,t},h_{2,2,t})$

References

Andersen, T., T. Bollerslev, P. Christoffersen, and F. Diebold (2006): *Volatility and Correlation Forecasting* chap. 15, pp. 777–877, Handbook of Economic Forecasting. North Holland.

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 579–625.

Barndorff-Nielsen, O., P. Hansen, A. Lunde, and N. Shephard (2008a): “Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76(6), 1481–1536.

——— (2008b): “Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading,” *DP, Oxford University.*

Barndorff-Nielsen, O., and N. Shephard (2002): “Econometric Analysis of Realized Covariation: High Frequency Covariance, Regression and Correlation in Financial Economics,” *DP, Nuffield College, Oxford University.*

——— (2004): “Measuring the Impact of Jumps in Multivariate Price Processes Using Bipower Covariation,” *DP, Nuffield College, Oxford University.*

Barone-Adesi, G., H. Rasmussen, and C. Ravanelli (2005): “An Option Pricing Formula for the GARCH Diffusion Model,” *Computational Statistics and Data Analysis*, 49-2, 287–310.

Bauwens, L., S. Laurent, and J. Rombouts (2006): “Multivariate GARCH Models: A Survey,” *Journal of Applied Econometrics*, 21, 79–109.

Bauwens, L., M. Lubrano, and F. Richard (1999): *Bayesian Inference in Dynamic Econometric Models.* Oxford University Press.

Bollerslev, T. (1986): “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, 31, 307–327.

——— (1990): "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH model," *Review of Economics and Statistics*, 72, 498–505.

Brandt, M., and F. Diebold (2006): "A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations," *Journal of Business*, 79, 61–74.

Clark, T., and M. McCracken (2001): "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, 105, 85–110.

Davidson, J. (2004): "Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model," *Journal of Business and Economic Statistics*, 22, 16–29.

Diebold, F., and R. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993): "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83–106.

Doornik, J. (2002): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press.

Elliott, G., and A. Timmermann (2008): "Economic Forecasting," *Journal of Economic Literature*, 46, 3–56.

Engle, R. (2002): "Dynamic Conditional Correlation - a Simple Class of Multivariate GARCH Models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R., and T. Bollerslev (1986): "Modelling the Persistence of Conditional Variances," *Econometric Reviews*, 5, 1–50.

Engle, R., and F. Kroner (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

Ethier, S., and T. Kurtz (1986): *Markov processes: Characterization and convergence*. John Wiley and Sons.

Glosten, L., R. Jagannathan, and D. Runkle (1992): "On the Relation Between the Expected Value and Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 46, 1779–1801.

Gourieroux, C., and A. Monfort (1995): *Statistics and Econometric Models.* Cambridge University Press.

Gourieroux, C., A. Monfort, and A. Trognon (1984): "Pseudo Maximum Likelihood Methods Theory," *Econometrica*, 52, 681–700.

Hansen, P., and A. Lunde (2005): "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)," *Journal of Applied Econometrics*, 20, 873–889.

——— (2006a): "Consistent Ranking of Volatility Models," *Journal of Econometrics*, 131, 97–121.

——— (2006b): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–218.

Hansen, P., A. Lunde, and J. Nason (2009): "Model Confidence Sets," Federal Reserve of Atlanta Working Paper.

Herdin, M., N. Czink, H. Ozcelik, and E. Bonek (2005): "Correlation Matrix Distance, a Meaningful Measure for Evaluation of Non-stationary MIMO Channels," IEEE-VCT.

James, W., and C. Stein (1961): "Estimation with Quadratic Loss," *Proc. Fourth Berkley Symp. on Math. Statist. and Prob.*, 1, 361–379.

Jensen, D. (1984): "Invariant Ordering and Order Preservation," *Inequalities in Statistics and Probability*, 5, 26–34.

J.P.Morgan (1996): *Riskmetrics Technical Document, 4th ed.* J.P.Morgan, New York.

Koch, K. (2007): *Introduction to Bayesian Statistics.* Springer Verlag Berlin.

Kushner, H. (1984): *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory.* The MIT Press.

Laurent, S. (2009): *G@RCH 6. Estimating and Forecasting Garch Models.* Timberlake Consultants Ltd.

L'Ecuyer, P. (1990): "A Unified View of the IPA, SF and LR Gradient Estimation Techniques," *Management Science*, 36.

——— (1995): "On the Interchange of Derivative and Expectation for Likelihood Ratio Derivative Estimators," *Management Science*, 41.

Ledoit, O., and M. Wolf (2003): "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance*, 10.

Meddahi, N. (2002): "A Theoretical Comparison between Integrated and Realized Volatility," Université de Montréal Working paper.

Nelson, D. (1991a): "ARCH Models as a Diffusion Approximation," *Journal of Econometrics*, 45, 7–38.

——— (1991b): "Conditional Heteroskedasticity in Asset Returns: a New Approach," *Econometrica*, 59, 349–370.

Patton, A. (2009): "Volatility Forecast Comparison Using Imperfect Volatility Proxies," *Forthcoming in Journal Econometrics.*

Strook, D., and S. Varadhan (1979): *Multidimensional diffusion processes.* Springer-Verlag.

van der Weide, R. (2002): "GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model," *Journal of Applied Econometrics*, 17, 549–564.

Voev, V., and A. Lunde (2006): "Integrated Covariance Estimation Using High Frequency Data in Presence of Noise," *Journal of Financial Econometrics*, 5.

West, K. (1996): "Asymptotic Inference About Predictive Ability," *Econometrica*, 64, 1067–1084.

White, H. (2000): "Reality Check for Data Snooping," *Econometrica*, 68, 1097–1126.

Zhou, B. (1996): "High-frequency Data and Volatility in Foreign Exchange Rates," *Journal of Business & Economic Statistics*, 14.