

The use of loss functions in assessing the VaR measures

Uso delle funzioni di perdita nella valutazione del VaR

Alessandra Amendola and Vincenzo Candila

Abstract The identification of the best model in terms of volatility forecast accuracy is a troublesome task and many evaluation methods have been proposed on the basis of a statistical or economic approach. The aim of this work is to investigate the opportunity to use a statistical approach in a VaR framework, i.e. evaluating the VaR measures by means of a loss function. By using high-frequency data it is possible to achieve a consistent estimate of the VaR bootstrapping the intraday increments of an asset. Hence, the performances of the volatility models are compared with that employing the VaR consistent estimate. In particular, the ‘true’ VaR is used to find a threshold discriminating low from high loss function values for each volatility model. The proposed procedure is assessed by means of a Monte Carlo simulation.

Abstract *L’identificazione del miglior modello in termini di accuratezza delle previsioni di volatilità rappresenta un arduo compito e diversi metodi di valutazione sono stati proposti, sulla base di un approccio statistico o economico. L’obiettivo di questo lavoro è investigare l’opportunità di usare un approccio statistico in un VaR framework, valutando le misure del VaR con le funzioni di perdita. Una stima consistente del VaR si ottiene attraverso il bootstrap degli incrementi infragiornalieri di un generico asset. Quindi, si valutano le performance dei modelli con quella derivante dall’utilizzo del ‘vero’ VaR. Inoltre, la stima consistente del VaR è usata per ottenere una soglia che discrimini tra valori bassi e alti delle funzioni di perdita. La procedura è valutata attraverso una simulazione Monte Carlo.*

Key words: Volatility, GARCH models, Value at Risk, loss function, bootstrap

Alessandra Amendola

Dept. of Economics and Statistics, University of Salerno, Via Giovanni Paolo II, 132 84084 Fisciano (SA), Italy

e-mail: alamendola@unisa.it

Vincenzo Candila

Dept. of Economics and Statistics, University of Salerno, Via Giovanni Paolo II, 132 84084 Fisciano (SA), Italy

e-mail: vcandila@unisa.it

1 Introduction

The identification of the best model in terms of volatility forecast accuracy is a troublesome task and many evaluation methods have been proposed, on the basis of a statistical or economic approach. The statistical approach based on the use of classic loss function suffers from the latent nature of the volatility [7]. The economic approach evaluates the volatility predictions indirectly by using utility functions [3] or other risk measures like the Value at Risk (VaR) [5]. The evaluation of volatility predictions through the VaR measures concerns some coverage tests - like the Unconditional and Conditional Coverage (CC) tests [2] - that may have low statistical power, as highlighted in [6].

The aim of this work is to investigate the opportunity to use the loss functions in a VaR framework in order to evaluate the volatility predictions of a set of competing models. In other words, the statistical and economic approaches are jointly used. The loss function in a VaR framework analyses the distance between the observed daily return of an asset and the VaR. As far as we know, except the proposal of [6], there are not many loss functions employed to evaluate this distance. We propose an asymmetric loss function, where the term asymmetric means that the models with an actual number of violations larger than the expected one are over penalized.

The mean-variance approach is used to derive the VaR measures of a set of competing volatility models while, following [1], a nonparametric method is used to find the ‘true’ VaR measures, that will be used as benchmark. This method is based on the bootstrap of the intraday increments of a generic asset and the VaR measure is obtained as a quantile of the estimated distribution of the bootstrapped daily returns. Then, the loss function employing this ‘true’ VaR is used to obtain the threshold discriminating low from high loss function values. Finally, the performances of the volatility models belonging to the family of the GARCH models are compared looking at the exceeding of the threshold as well as the loss function value.

The work is organized as follows. Section 2 presents the bootstrap method used to estimate the daily VaR measure. In Section 3 the two loss functions in a VaR framework are illustrated. The results of the Monte Carlo experiment are showed in Section 4 and Section 5 concludes.

2 Bootstrapping the intraday increments

Let $\chi_{t,N} = \{q_{t,1}, \dots, q_{t,N}\}$ be a sequence of N intraday increments for a generic day t and a generic asset, such that $q_{t,n} = \log(P_{t,n}) - \log(P_{t,n-1})$, with $P_{t,n}$ denoting the observed intraday price. The open-to-close daily return $r_{t,N}$ is given by $r_{t,N} = \sum_{n=2}^N q_{t,n}$. Because of the dependence in the sequence $\chi_{t,N}$, the Stationary Bootstrap (SB) of [9] is used. In the SB, for each day t , B re-sampled intraday sequences are calculated, each of length N . A re-sampled intraday sequence is formed by N sampled blocks, whose average block length relies on the dependence exhibited within $\chi_{t,N}$. As [1], we use the procedure described in [8] in order to estimate the

average block length. Once the bootstrapped sequence is obtained, the resulting summation represents a re-sampled daily return, independent of the original one, but generated by the same distribution, as $\{N, B\} \rightarrow \infty$. The assumptions needed to assure that the bootstrapped daily returns and the original daily return converge in distribution are provided by [4]. Hence any moment or quantile of the original return can be now estimated by means of this B *i.i.d.* sequence. We focus on $\widehat{VaR}_t$, a consistent estimate of the 5% VaR for the day t , obtained as:

$$\widehat{VaR}_t = r_{t,N,[0.05B]}^* - \mathbb{E}^* r_{t,N}^*, \quad (1)$$

where $r_{t,N,[0.05B]}^*$ indicate the 5% quantile of the bootstrapped daily returns for the day t and $\mathbb{E}^* r_{t,N}^* \equiv r_{t,N}$.

3 Loss function in a VaR framework

A loss function (LF) compares $r_{t,N}$ to the VaR measure. In this work we consider two loss functions, illustrated in Table 1.

Table 1 Loss Functions

Magnitude Loss Function (MLF)	Asymmetric Loss Function (ALF)
$Loss_{M,t} = \begin{cases} 1 + (r_t - VaR_t)^2 & \text{if } r_t \leq VaR_t \\ 0 & \text{if } r_t > VaR_t \end{cases}$	$Loss_{A,t} = \begin{cases} 1 + P * (r_t - VaR_t)^2 & \text{if } r_t \leq VaR_t \\ 0 & \text{if } r_t > VaR_t \end{cases}$
$NS_M = T^{-1} \sum_{t=1}^T Loss_{M,t}$	$NS_A = T^{-1} \sum_{t=1}^T Loss_{A,t}$

The first LF due to [6] is called Magnitude loss function (MLF). The second is the new proposed Asymmetric loss function (ALF), that penalizes more the models with an actual number of violations ($\hat{\alpha}$) greater than the expected one (α_0). For the ALF, P is the penalizing quantity that over penalizes the model if $\hat{\alpha} > \alpha_0$. The average of the LS is called numerical score (NS) and it is denoted by NS_M and NS_A for the MLF and the ALF, respectively. Once the consistent estimate of the VaR has been obtained, it is possible to find a threshold that discriminates between low from high NS . The procedure is again based on the block bootstrap: the threshold is an empirical quantile of the bootstrapped distribution of the NS , when $\widehat{VaR}$ is used. If a NS of a volatility model lies above the threshold, that model is considered rejected.

4 The Monte Carlo Simulation

In this section the setting of the Monte Carlo simulation is illustrated. The data generating process has the following GARCH representation: $r_t = h_t z_t$, with $z_t \sim$

$\mathcal{N}(0, 1)$ and $h_t^2 \sim GARCH(1, 1)$ process. In particular:

$$h_t^2 = 0.01 + 0.10r_{t-1}^2 + 0.85h_{t-1}^2, \quad t = 1, \dots, 1099. \quad (2)$$

The first 99 observations are used as a warm-up period. The rest of the sample is used to simulate the intraday returns in order to have an estimate of the VaR for each period by means of the stationary bootstrap. For sake of simplicity, the sequence of increments is assumed to be independently drawn from a Normal distribution whose variance is constant and equal to $(1/N)h_t^2$. More specifically, we set a number of intraday increments N equal to 390 and a number of replicates R equal to 200. Once obtained the sequence of increments, the stationary bootstrap, as illustrated above, is used to have $\widehat{VaR}_t$, for $t = 100, \dots, 1099$ and $R = 1, \dots, 200$. Finally, the thresholds for each loss function and each replicate are obtained.

The set of competing models are showed in Table 2. We consider a RiskMetrics (RM), three GARCH(1,1), noted as G(1,1), and three GJR-GARCH(1,1) models, indicated as GJR(1,1). All the models are misspecified.

Table 2 The configuration of the models used in the simulation

	M1	M2	M3	M4	M5	M6	M7
Model	RM	G(1,1)	G(1,1)	G(1,1)	GJR(1,1)	GJR(1,1)	GJR(1,1)
α_1	-	0.07	0.05	0.25	0.05	0.01	0.15
β_1	-	0.88	0.90	0.70	0.88	0.90	0.70
γ	-	-	-	-	0.02	0.04	0.10

Because we aim to compare the MSE function within the Monte Carlo framework, the frequency at which each model has the smallest MSE is calculated. As regards to the economic approach, the Conditional Coverage (CC) test is reported. In particular, the null hypothesis of the CC test jointly checks if the violations are independent distributed over time and the actual number of violations are coherent with the expected one. We calculate the frequency at which the CC test is not rejected. With reference to the loss function approach in a VaR, we calculate a global accuracy measure for each loss function, varying between zero and one. More specifically, the global accuracy measure, AM , is given by the linear combination of three accuracy measures. Recall that the benchmark of the loss function values is represented by the numerical score obtained when $\widehat{VaR}_t$ is used. The first accuracy measure, $Acc_{1,i}$, checks the performance of the model i compared to that of the process that employs $\widehat{VaR}_t$. The second accuracy measure, $Acc_{2,i}$, reports the frequency of not exceeding the thresholds. The third, $Acc_{3,i}$, is the frequency at which each model i has the smallest numerical score. Formally, let

$$v = R^{-1} \sum_{r=1}^R \mathbf{1}_{(NS_{0,r} > \widehat{TR}_r)} \text{ and } \{v\} = \{\cup_{r=1}^R r \cdot \mathbf{1}_{(NS_{0,r} > \widehat{TR}_r)}\};$$

$$J_i = R^{-1} \sum_{r=1}^R \mathbf{1}_{(NS_{i,r} > \widehat{TR}_r)} \text{ and } \{J_i\} = \{\cup_{r=1}^R r \cdot \mathbf{1}_{(NS_{i,r} > \widehat{TR}_r)}\},$$

where $NS_{0,r}$ and $NS_{i,r}$ are the r^{th} numerical scores when $\widehat{VaR}_t$ and the model i are used, respectively; $\widehat{TR}_r$ is the threshold for the r^{th} replicate when $\widehat{VaR}_t$ is plugged into the loss function and $\mathbf{1}_{(\cdot)}$ is the indicator function that is equal to one when the argument is true. Hence, the formalization of the accuracy measures is in Table 3.

Table 3 The specifications of the accuracy measures

$Acc_{1,i}$	$Acc_{2,i}$	$Acc_{3,i}$
$card[\{v\} \cap \{J_i\}] / v^1$	$1 - J_i / R$	$RF_i = R^{-1} \sum_{r=1}^R CF_{i,r},$ where $CF_{i,r} = \begin{cases} 1 & \text{if } NS_{i,r}^* : \nexists NS_{l,r} < NS_{i,r}^*, \forall l \neq i \\ 0 & \text{otherwise} \end{cases}$

¹ $card[\cdot]$ stands for the cardinality of the argument.

By construction, each accuracy measure lies in the interval $[0,1]$. The global accuracy measure for the model i is $AM_i = 3^{-1} [Acc_{1,i} + Acc_{2,i} + Acc_{3,i}]$. The higher AM_i is, the better the performance of the model i is, because it has all the replicates below the threshold, it has the same performance of the process that uses a consistent estimate of the VaR and its numerical scores are low compared to those of the other models.

The results of the volatility evaluation are showed in Table 4.

Table 4 The results of the volatility evaluation

	M1	M2	M3	M4	M5	M6	M7
MSE ^a	0.1050	0.2700	<u>0.6250</u>	0.0000	0.0000	0.0000	0.0000
CC test ^b	0.8950	<u>0.9550</u>	0.8950	0.6050	0.8600	0.6550	0.2750
AM_M ^c	0.3883	0.4567	<u>0.5767</u>	0.3667	0.3900	0.3800	0.3400
AM_A	0.3917	0.4533	<u>0.5833</u>	0.3700	0.3867	0.3800	0.3367

^a The first row shows the frequency at which the MSE is the smallest for each model. ^b The second row shows the frequency at which the Conditional Coverage test is not rejected. ^c The third and fourth rows show the global accuracy measures for the Magnitude and Asymmetric loss functions, respectively.

The first row shows the statistical approach. For the model i , the MSE is equal to $\mathbb{E}[h_t^2 - \hat{h}_{t,i}^2]^2$, where h_t^2 comes from (2) and $\hat{h}_{t,i}^2$ from the formulation of the conditional variance as expressed by the model i . In the table it is clear that the statistical approach does not prefer any GJR models. Instead, it awards about 63% of times

the model 3. The CC test does not bring any advantage to the analysis, given that four models are awarded: the models 1-3 and the model 5 reach the highest percentages of not rejections of the Conditional Coverage test. Interestingly, the loss function approach in a VaR framework awards model 3. This means that M3 has a same behaviour compared to that of the process that uses $\widehat{VaR}_t$, its numerical scores are below the thresholds and the smallest among all the models. Moreover, the global accuracy measure when the asymmetric loss function is used reaches a slightly greater value than that of Magnitude loss function (0.5833 against 0.5767). In situations where more than one model have the same performance, this could be an advantage.

To conclude, we argue that the using of loss function in a VaR framework helps the research of the best model, when the economic approach does not lead to a clear decision. The results of this approach are consistent with those obtained from a robust statistical loss function.

5 Conclusions

The evaluation of volatility forecasts by means of statistical or economic approaches may lead to ambiguous conclusions. We aimed to investigate the opportunity to use the loss function in a VaR framework. To do this, a consistent estimate of the VaR measures has been provided by using the stationary bootstrap. Then these VaR measures have been used to find the threshold discriminating low from high loss function values. The analysis has been conducted with two loss functions, of which one is new. It has emerged that this method helps the model selection in situations in which the traditional approaches do not clearly determine the best model.

References

1. Bowers, C., Heaton, C.: Bootstrapping daily returns (2013). Available at SSRN <http://dx.doi.org/10.2139/ssrn.2361393>
2. Christoffersen, P. F.: Evaluating interval forecasts. *Int. Econ. Rev.* **39**(4), 841–862 (1998)
3. Fleming, J., Kirby, C., Ostdiek, B.: The economic value of volatility timing using “realized” volatility. *J. of Financ. Econ.* **67**(3), 473–509 (2003)
4. Gonçalves, S., de Jong, R.: Consistency of the stationary bootstrap under weak moment conditions. *Econ. Lett.* **81**(2), 273–278 (2003)
5. Jorion, P.: *Value at risk*. McGraw-Hill, New York (2007)
6. Lopez, J.A.: Methods for evaluating value-at-risk estimates. *Econ. Policy Rev.* **4**(3), 119–124 (1998)
7. Patton, A.: Volatility Forecast Comparison using Imperfect Volatility Proxies. *J. of Econom.* **160**(1), 246–256 (2011)
8. Patton, A., Politis, D. N., White H.: Correction to automatic block-length selection for the dependent bootstrap. *Econom. Rev.* **28**(4), 372–375 (2009)
9. Politis, D.N., Romano, J. P.: The stationary bootstrap. *J. of the Am. Stat. Assoc.* **89**, 1303–1313 (1994)