

Aleksandra Helena Pasieczna

Kozminski University, Warsaw

www.pasiecznapl@gmail.com  <https://orcid.org/0000-0001-6867-3584>**Magdalena Joanna Szydłowska**

Wroclaw University of Economics and Business

[magdalena.szydłowska@ue.wroc.pl](mailto:magdalen.szydłowska@ue.wroc.pl)  <https://orcid.org/0000-0001-6251-948X>

Estimating Model Risk of VaR under Different Approaches: Study on European Banks

Introduction

Computational finance relies on many different models to provide estimates on various financial variables, such as market risk, systemic risk, portfolio optimization, and pricing. Since modeling is at the heart of modern finance, it is important to judge these models. In literature two metrics are used for this purpose – accuracy and precision.

We define accuracy as the ‘goodness’ of the model, or equivalently the bias (the bias is the amount of inaccuracy of the average expected value of the model outputs). The bias indicates a consistent ‘shift’ in a particular direction from a benchmark. For example, if an options pricing model consistently under-evaluates an option, the bias would be the difference between the predicted and actual prices. A measure of the accuracy helps determine the model’s shortcomings with respect to the chosen benchmark, and thus reflects the hypotheses and assumptions that are used to build the model. Decision to integrate the model into a practitioner’s toolkit is usually done based on the accuracy.

Precision provides an estimate of the consistency of a model to give a reliable output on multiple simulations with slightly noisy input parameters. Unlike accuracy, there is no reference model or benchmark against which this quality can be judged, and only the repeatability (under slightly different inputs) can be checked. In essence, the precision reflects the risk of the model. If the model is the risk measure, the precision will be *the risk of the risk measure*, henceforth called *model risk*.

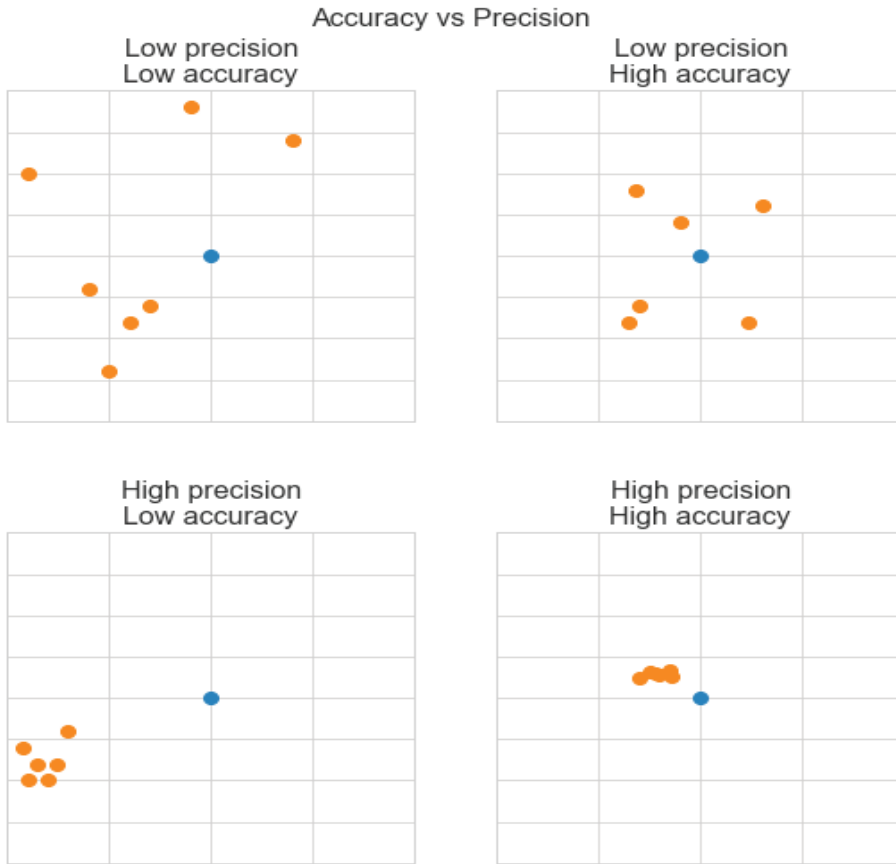


Figure 1. Accuracy and precision – comparison. The blue central dot is the benchmark, and the orange dots are the model estimates

Source: own study.

Figure 1. shows the differences between accuracy and precision, which highlights the use of our definitions of these two qualities. In the first case we see a case of having low precision and accuracy, where the model completely fails to be consistent and capture the correct result, thus failing to be of any theoretical or practical value. The second case shows a model of high accuracy and low precision, where the average of multiple simulations correctly captures the benchmark. This is of theoretical value, but less useful in a practical setting. In the third case we have a model with high precision and low accuracy, which indicates a consistent bias in the model from the benchmark. This model is of high use for theory and practice, since it can be confidently adjusted for implementation, and provides a hint of a shortcoming in the model for theoreticians. The final case is the best-case scenario, where the model is both, accurate and precise, thus of high value for research and practice. In this case the model itself acts as a benchmark.

In this study we focus on the accuracy and precision aspects for Value at Risk (VaR) models, extending from an earlier work (Pasiieczna, 2019). VaR, recommended by the Basel II agreement (Basel Committee on Banking Supervision, 2004), is frequently used to estimate potential losses with a certain confidence. In common usage, there are three ways to measure VaR – the historical method, the Monte Carlo method and GARCH (Holton, 2014).

This paper focuses on measuring the model risk for VaR for different configurations (confidence levels). The models used are the Monte Carlo technique (Pasiieczna, 2019), the historical simulation, and GARCH (Danielsson et al., 2016). The changes to the input parameters are made by choosing different historical periods. Our work is tested on four leading European banks.

The paper is divided as follows. The following sections give a brief overview of the theoretical and computational methods used in our work. Section 3. deals with the simulation results for the measurements of VaR risk for the different banks. We end with our conclusions.

1. Theory and Computational Methods

The dataset consisted of daily close prices between 2002 and 2019 for four banks: BNP Paribas, Credit Agricole, Commerzbank and Deutsche Bank and were obtained from the official websites of these institutions.

1.1. Value at Risk

VaR is defined as the maximum possible loss, or equivalently the most “negative” price change, whose probability is within a pre-defined confidence level over a pre-defined time horizon. The choice of a confidence level and time horizon constitutes a VaR configuration. To better understand: A portfolio with a 1-day VaR of 700 EUR with a 95% confidence implies that the portfolio has a 5% chance of losing at least 700 EUR in one day. Equivalently, there is a 95% chance that the losses over one day will be smaller than 700 EUR for the given portfolio.

This definition of VaR is non-constructive in that it specifies a property that VaR should have, but not on how to compute VaR. As a result, there are different ways proposed to compute it (Holton, 2014). For our purposes we use the historical simulation, Monte Carlo and GARCH approaches for computing VaR, with the focus on two French and two German banks. Unlike the example presented earlier, we estimate the VaR as relative price changes to make it comparable with the returns across time.

1.2. Monte Carlo Methods

We chose the Monte Carlo (MC) method (Glasserman, 2003) to simulate the bank stock relative returns for a single trading day. The main advantage of the MC method is that one can simulate the different sources of uncertainty that affect the stocks by drawing random numbers from predetermined probability distributions. As such, the model limitations are mainly due to the choice of the distributions and

the computational costs associated with the generation of statistics. MC methods have been applied in various areas of finance, such as portfolio optimization and risk analysis.

Our approach uses Monte Carlo to simulate the uncertain relative price changes and the uncertainty is described through the mean and standard deviation determined by real historical data. The VaR is then simply the quantile of these simulated price changes across multiple MC runs corresponding to the configuration (e.g. 95%). The algorithm is as follows:

1. For a given bank on every day, estimate the mean and standard deviation, based on the past historical data on relative price changes. Three historical periods were used: 125, 250, 500 trading days (approximately 0.5, 1 and 2 years).
2. Simulate the next day's relative returns by drawing random numbers from a Gaussian distribution with the precalculated mean and standard deviation. For each day 50000 numbers (MC iterations) were drawn.
3. Rank the simulated relative returns in a descending manner and choose the quantile corresponding to the confidence level as the VaR. Two confidence levels were tested, 95% and 99%.

Once the VaR was computed for the banks at the given confidence levels with the three historical periods, we estimated the VaR model risk, as described in Section 2.5.

1.3. Historical Simulation

We also use the standard historical simulation (HS) to compute the VaR for different banks. HS is a less sophisticated approach that assumes that the distribution of the returns is completely given by past returns (Holton, 2014). It simply considers the quantile of the past relative returns corresponding to the VaR confidence level as the VaR. As in the MC approach, we compute VaR for three historical periods, so as to estimate the accuracy and precision metrics.

1.4. GARCH

The third model we studied is Generalized Autoregressive Conditional Heteroskedastic (GARCH) model, with $p = q = 1$, i.e. the GARCH(1, 1) process, and the AR(1) process applied to the relative returns (Ferenstein, Gąsowski, 2004). The AR(1) residuals passed to the GARCH model are assumed to be normal. This is an extension of the models originally proposed by Engle (Engle, 1982) and Bollerslev (Bollerslev, 1986). In our study we use this approach to compute the conditional volatility of the next day, from which we estimate the VaR. As with the other models, the VaR is computed daily using data from past rolling historical periods.

1.5. Model Risk Measures

In this work we use the accuracy as a measure of the performance of the model on average and the precision as the basis for the model risk measures. There is an

inherent risk in using an inaccurate model, namely in the form of ‘known unknowns’ and ‘unknown unknowns’, however we do not focus on accuracy as a foundation of a model risk measure, but more as a performance evaluation metric. For our purposes, the accuracy metric is simply the percentage crossing, i.e. the fraction of the number of days when the realized loss is higher than the average VaR for a given configuration (averaged estimates across different historical periods). We also compute the Kupiec Proportion of Failures test to judge the percentage crossings. A Kupiec POF value of 0 is considered ideal.

Two measures were developed from the precision – the spread across VaR estimates, and the ratio of the highest to lowest VaR estimates. Note that the estimates used in these measures always correspond to the same configuration (prediction time period and confidence level), but different historical periods. Consider the case of computing the spread for a 1-day VaR with a 95% confidence; the spread will be the difference between the maximum and minimum of the three VaR estimates obtained from the three historical periods (125, 250, 500 trading days).

Using the spread as a model risk measure is analogous to how the risk of a stock investment strategy is its volatility. By having the historical periods varying by a factor of two, we essentially determine the response of the model when subject to half or twice the amount of data. The spread is represented in the units of the VaR itself (in our case, relative price change).

The other model risk measure, the ratio, is defined as the highest to the lowest VaR estimates for a given confidence level and prediction time period. This measure has already been used to study various systemic risk measures and VaR, while comparing different implementations of VaR (Danielsson et al., 2016). However, in our case, we refrain from combining the results of the two implementations and stick to analyzing the ratios from a single approach only. Our justification for this comes from the fact that comparing ratios across different implementations might lead to comparing estimates derived from different hypotheses.

The four metrics presented here are used to judge the MC, HS and GARCH approaches for estimating VaR, and we present a brief comparison in the next section.

2. Results and Discussion

We present here our simulation results for the 4 European banks, where the 1-day VaR is computed at 95% and 99% confidence levels. Estimates from three historical periods (125, 250, 500 days) for given confidence levels were used to judge the accuracy and precision.

Table 1. Accuracy and precision metrics for the different VaR configurations based on our calculations.

Ideal percentage crossing values for VaR at 95% and 99% are 5% and 1% respectively. Ideal Kupiec POF values should be 0. Ideal ratio and spread values should be 1 and 0 respectively

		Accuracy (% crossings)			Accuracy (Kupiec POF)			Precision (avg ratio)			Precision (avg spread)		
		HS	MC	G	HS	MC	G	HS	MC	G	HS	MC	G
VAR 95	<i>BNP Paribas</i>	5.531	4.352	4.621	2.338	3.752	1.258	1.407	1.409	1.169	0.011	0.012	0.005
	<i>Credit Agricole</i>	5.654	3.762	3.859	3.52	14.299	12.051	1.395	1.353	1.171	0.012	0.012	0.006
	<i>Deutsche Bank</i>	5.654	4.696	4.671	3.52	0.805	0.949	1.397	1.383	1.164	0.012	0.012	0.005
	<i>Commerzbank</i>	5.261	4.082	4.4	0.572	7.684	3.205	1.423	1.418	1.192	0.014	0.015	0.007
VAR 99	<i>BNP Paribas</i>	1.475	1.451	1.45	8.086	7.325	7.316	1.431	1.406	1.169	0.017	0.017	0.008
	<i>Credit Agricole</i>	1.426	1.279	1.008	6.581	2.93	0.003	1.474	1.346	1.171	0.019	0.016	0.008
	<i>Deutsche Bank</i>	1.5	1.623	1.524	8.889	13.409	9.726	1.407	1.381	1.164	0.017	0.018	0.008
	<i>Commerzbank</i>	1.622	1.402	1.278	13.397	5.888	2.925	1.512	1.411	1.192	0.022	0.021	0.01

Source: own study.

Our results for the accuracy and precision are summarized in Table 1. We found that the percentage crossings for the VaR at 95% are higher for HS (between 5.3 and 5.7%), whereas there are lower for the MC and GARCH models (between 3.8 and 4.7%). The corresponding Kupiec POF values range between 0.5 and 3.5 for HS, 0.8 and 14.3 for MC, and 0.9 and 12.1 for GARCH. For this VaR configuration HS simulations seemed slightly more accurate in terms of Kupiec POF than MC and GARCH. For VaR at 99%, all models had more than 1% crossings. The corresponding Kupiec POF values were between 6.6 and 13.4 for HS, 2.9 and 13.4 for MC, and 0 and 9.7 for GARCH. GARCH models gave more accurate results (Kupiec POF) than HS and MC for VaR at 99%.

Overall, for the accuracy, it is less clear as to which model outperforms the others independent of the VaR configuration. More detailed studies with more banks might be useful to judge the significance of these remarks. On the other hand, for both VaR configurations, GARCH simulation gave consistently better precision metrics than HS and MC. This was seen for both precision metrics. No discernible difference was observed between HS and MC precision results.

To compare the temporal evolution of precision metrics, we plot the individual bank's VaR estimates (in units of relative price change) along with the daily spreads and ratios for the 1-day VaR at 99% (Figures 2–5). These plots also allow us to study how the various quantities evolve during periods of crises (e.g. 2008) and post-crises (before 2020).

In the following figures, we notice that the VaR curves follow closely the shape of the envelope of the returns as expected. We also highlight the observation that the GARCH models seemed to react quicker than HS and MC, which can be understood as an effect of the GARCH model giving more weight to more recent events. During the 2008 crisis, we observe more crossings than during periods of relative calm, indicating a degradation of the accuracy during volatile periods. This behavior has been reported earlier (Danielsson et al., 2016).

In terms of the precision metrics, the model is considered precise when the spread and ratio are as close as possible to 0 and 1 respectively, which would imply that the model is not highly dependent on the input parameters (here, historical time period). However, in our study we observe that the ratio is never 1, and the spread is always greater than 0. This means that the model risk is never completely nullified.

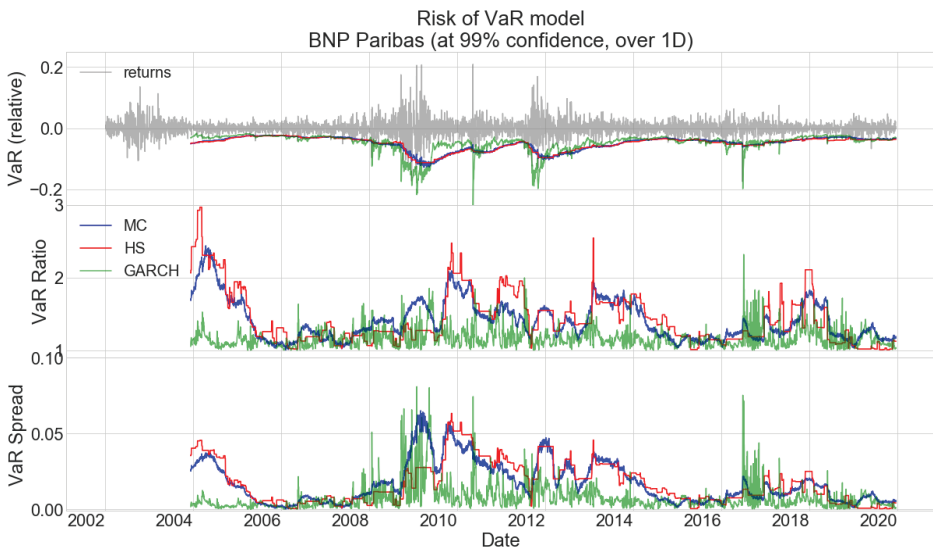


Figure 2. Results for BNP Paribas with 1-Day VaR computed at 99% confidence level. The blue, red and green curves refer to the MC, HS and GARCH results. The first subplot shows the average VaR along with the relative returns (grey). The second subplot shows the ratios between the highest and lowest VaR estimates. The third subplot shows the spread (difference between the highest and lowest estimates)

Source: own study.

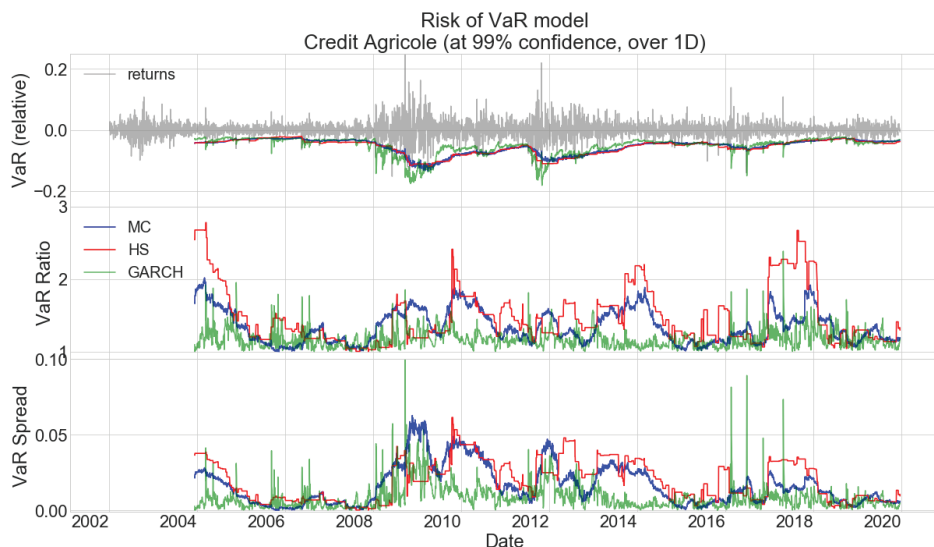


Figure 3. Results for Credit Agricole with 1-Day VaR computed at 99% confidence level. The blue, red and green curves refer to the MC, HS and GARCH results. The first subplot shows the average VaR along with the relative returns (grey). The second subplot shows the ratios between the highest and lowest VaR estimates. The third subplot shows the spread (difference between the highest and lowest estimates)

Source: own study.

We see an increase in the spreads and ratios *after* the 2008 crisis. We interpret this as a delay in capturing information immediately as well as an interplay between different historical periods. In other words, the VaR estimate from the 125-day historical period will react faster than the estimate with the 500-day historical period, leading to an increase in the spread and the ratio. The delay factor is due to the requirement of having at least 125 days, which would include days before the crisis. We expect the GARCH models to react significantly faster due to their *larger* weight on more recent data. This explains the outperformance of these models over HS and MC in terms of precision metrics. Furthermore, we observe that the HS approach leads the MC, the reason for which is attributed to the inclusion of actual returns (tail events as well) in HS. The MC approach relies on collecting data for building the standard deviation and means, which requires much longer waiting periods.

We also observe a ‘bumpy’ structure for the HS curves, specifically prominent in the ratio. These features are due to the nature how the HS approach works, by inclusion or exclusion of a single point which might change the quantile suddenly. On the other hand, the MC approach does not depend on a single point and is more robust to sudden changes. The disadvantage of MC is of course the introduction of a much longer reaction time. Finally, the GARCH structure seems to be the most volatile (fluctuates the most), which is a result of the quick adjustment.

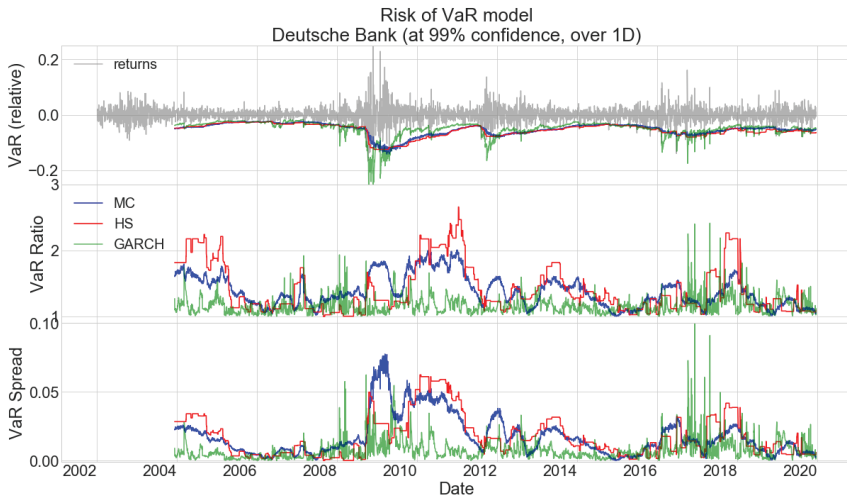


Figure 4. Results for Deutsche Bank with 1-Day VaR computed at 99% confidence level. The blue, red and green curves refer to the MC, HS and GARCH results. The first subplot shows the average VaR along with the relative returns (grey). The second subplot shows the ratios between the highest and lowest VaR estimates. The third subplot shows the spread (difference between the highest and lowest estimates)

Source: own study.

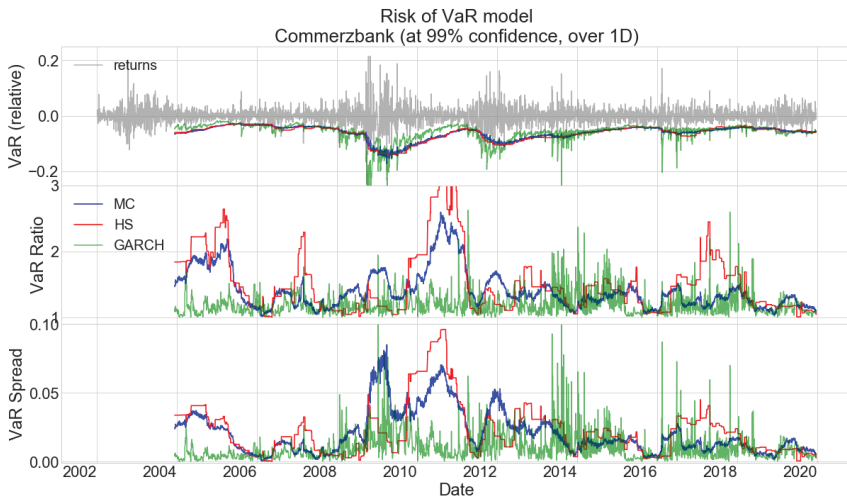


Figure 5. Results for Commerzbank with 1-Day VaR computed at 99% confidence level. The blue, red and green curves refer to the MC, HS and GARCH results. The first subplot shows the average VaR along with the relative returns (grey). The second subplot shows the ratios between the highest and lowest VaR estimates. The third subplot shows the spread (difference between the highest and lowest estimates)

Source: own study.

Another important point is to notice that the spread and ratio for the three approaches are not constant in time. This is most visible in Figure 5. (Commerzbank), where the HS ratio can take values as high as 3 or above, indicating that any two VaR estimates might differ by a factor of 3 or more. It implies that if a VaR estimate has a value of 10 Euros for the upcoming day, another estimate might predict 30 Euros or 3.33 Euros. These high values can be reached for all three models. Finally, the highest value for the spread or ratio is reached when the returns volatility is small post-crises, indicating that the risk is overestimated just after the crisis has passed.

Conclusions

We presented an approach to judge a risk model based on accuracy and precision metrics. We proposed using two accuracy metrics as performance evaluators – percentage crossings and Kupiec POF, and two precision metrics as model risk measures – spread and ratio of the VaR estimates under three different historical periods. We tested these on 4 European banks and two VaR configurations under three different methods. Our results indicate that all three models performed similarly in terms of overall accuracy. With regards to the precision metrics, the GARCH outperformed MC and HS models. All models were found to be less precise during the crises and just after. The approach presented here is general and can be applied to any kind of risk models, such as Expected Shortfall or even systemic risk measures. As perspectives, the comparisons between different VaR methods can be performed on more banks and the approach.

Abstrakt

Szacowanie ryzyka modelu VaR w różnych podejściach: Badanie na przykładzie banków europejskich

Celem badania było oszacowanie ryzyka modelu wartości zagrożonej ryzykiem (VaR), rozumianego jako precyzja, oraz dokładności tego modelu, za pomocą trzech metod: symulacji historycznej (HS), Monte Carlo (MC) i uogólnionego modelu ARCH (GARCH). W tej pracy do analizy modelu VaR wykorzystano dokładność i precyzję. Oszacowania dokładności i precyzji dokonano w ramach trzech podejść dla czterech banków europejskich, przy poziomach ufności 95 i 99%. Procentowy udział przekroczenia wartości VaR oraz miara POF Kupca zostały użyte do oceny dokładności modelu; natomiast stosunek szacunkowych wartości maksymalnych i minimalnych VaR oraz rozstęp między tymi wartościami zastosowano do oszacowania ryzyka modelu. Wykonano to poprzez zmianę parametrów wejściowych, a dokładniej przedziałów estymacji (125, 250, 500 dni). Dokładność sama w sobie nie jest wystarczająca do oceny modelu i wymagana jest również precyzja. Ewolucja miary precyzji w czasie wykazała niespójność metod szacowania VaR w różnych warunkach rynkowych. W tym artykule skupiono się na koncepcjach dokładności i precyzji stosowanych do oszacowania ryzyka modelu wartości zagrożonej (VaR).

VaR stanowi podstawę zaawansowanych wskaźników ryzyka, w tym miar ryzyka systemowego, takich jak MES i Delta CoVaR, dlatego też zrozumienie ryzyka związanego z VaR ma kluczowe znaczenie dla praktyków finansów.

Słowa kluczowe: VaR, Monte Carlo, ryzyko modelu, precyzja, symulacja historyczna, GARCH

Abstract

Estimating Model Risk of VaR under Different Approaches: Study on European Banks

The objective of this research is to estimate the model risk, represented as precision, and the accuracy of the Value at Risk (VaR) measure, under three different approaches: historical simulation (HS), Monte Carlo (MC), and generalized ARCH (GARCH). In this work, to analyze the VaR model, the accuracy and precision were used. Estimation of the accuracy and precision was done under the three approaches for four European banks at 95 and 99% confidence levels. The percentage crossings and Kupiec POF were used to judge the model accuracy, whereas the ratio of the maximum and minimum VaR estimates, and the spread between the maximum and minimum VaR estimates were used to estimate the model risk. This was achieved by changing input parameters, specifically, the estimation time window (125, 250, 500 days). Implications/Recommendations: The accuracy alone is not sufficient to evaluate a model and precision is also required. The temporal evolution of the precision metrics showed that the VaR approaches were inconsistent under different market conditions. This article focuses on the accuracy and precision concepts applied to estimate model risk of the Value at Risk (VaR). VaR is the foundation for sophisticated risk metrics, including systemic risk measures like Marginal Expected Shortfall and Delta Conditional Value at Risk. Thus, understanding the risk associated with the use of VaR is crucial for finance practitioners.

Keywords: VaR, Monte Carlo, model risk, precision, historical simulation, GARCH

References

- Basel Committee on Banking Supervision. (2004). *Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Bank for International Settlements.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Danielsson, J., James, K. R., Valenzuela, M., Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, 23, 79–91. <https://doi.org/10.1016/j.jfs.2016.02.002>

- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4), 987–1007. <https://doi.org/10.2307/1912773>
- Ferenstein, E., Gąsowski, M. (2004). Modelling stock returns with AR-GARCH processes. *Statistics and Operations Research Transactions*, 28(1), 55–68.
- Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer Science+Business Media.
- Holton, G. A. (2014). *Value-at-Risk: Theory and Practice, Second Edition*. Value-at-risk. <https://www.value-at-risk.net> (accessed: 17th November 2021).
- Pasieczna, A. H. (2019). Monte Carlo Simulation Approach to Calculate Value at Risk: Application to WIG20 and MWIG40. *Financial Sciences*, 24(2), 61–75. <https://doi.org/10.15611/fins.2019.2.05>