Assignment 1 model validation

MeiFang Li(2719570), Yuhao Qian(2684098)

1. Introduction

Model validation is an essential part in the model development life cycle. The risk model might have significant losses if not appropriately validated. This report validates the Value-at-Risk(VaR) model provided by Luuk Oudshoorn and Willem-Jan de Voogd. Since they propose Filter Historical Simulation(FHS) as the best VaR model, we will solely focus on this. First, we will check the replicability of their model according to the documentation and the given data. Then we will assess the model to see whether it is adequate and reliable. We will try to interpret the limitations and further provide some discussions in the last section.

2. Documentation

Luuk and Willem-Jan set up a complicated portfolio consisting of commodities, stocks and bonds. In detail, the commodity is CRYTR(in USD); equity indices are MSCI Europe, AEX, Bist, Nasdaq(in USD) and Emerging equity(in USD); the bonds are France government bonds and US Mortgage Backed Loans. They assume they are Euro investors, so all the returns have been converted into EUR already. They provide us with a data set with daily log returns over ten years (from 2011-01-03 to 2021-01-29). The missing values in the historical data are replaced by 0 during synchronisation. Intriguingly, the assets weights and the initial investment of the portfolio are missing in the documentation; thus, we could only assume these assets are equally weighted and do not include capital during validation. Although the maturity time of the bonds are also not clear, we could still make good use of the data they provided.

In their risk measurement, they mention four different methods: VC, HS, FHS-EWMA and CCC-GARCH. They find that VC method with student-t distribution performs well in the backtesting. Despite this, they prefer FHS-EWMA because this conditional method has relatively more even VaR exceedings. They investigate a lot how their portfolio would react to the dramatic changes, but somehow they seem to omit some details of their portfolio. Hence, we will solely focus on VaR with FHS-EWMA in these assumptions: the λ is 0.94; the initial σ^2 is the variance of the first 50 days; the time horizon of VaR is 1 day.

3. Assessment

3.1. Theoretical Assessment

Luuk and Willem-Jan do not seem to clarify the loss operator in the report. The exchange rate is one of the risk factors in the portfolio, but this column is missing from the data provided. The logarithmic price is the risk factor for stocks and indices in this case. France government bonds and US Mortgage Backed Loans are the two bonds in their portfolio, but the interpretations regarding the maturity time and how they compute the yield change seem omitted. Hence, it makes difficulties for us to identify their loss operator. Hereby, we assume the loss operator as:

$$l_{[t]}(\mathbf{x}) = -V_t \sum_{i=1}^{d} \omega_{t,i} \left(e^{x_i} - 1\right) \tag{1}$$

where $\omega_{t,i} = \alpha_i S_{t,i}/V_t$ is the relative weight at time t, x_i is the risk factor change(log return).

Figure 1 clearly shows that the portfolio loss has heavy tails on both sides, and standard normal distribution can not fit it well. Figure 2 indicates that the student distribution with 4.5 degrees of freedom may better represent the loss.

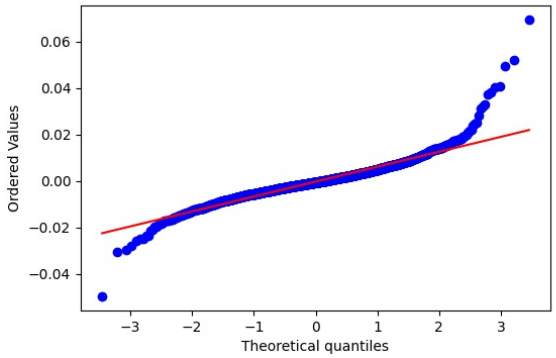

Fig. 1: QQ plot for loss(normal)

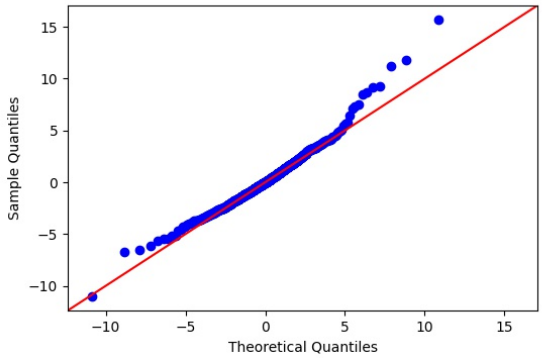

Fig. 2: QQ plot for loss(student 4.5 degrees of freedom)

3.2. Quantitative Assessment

When it comes to the quantitative assessment, we refer to the results of FHS model they mentioned in Table 2 and Table 7 of their report, where the number of violations per year and the p-value for statistical testing are presented. They apply a rolling window with two-year length so the scoring period is from 2013 to 2021, 9 years(2018 days) in total. Additionally, they perform their FHS-EWMA model using $\lambda = 0.94$ and the initial variance is obtained by the variance of the first 50 days. Hence, We run our FHS-EWMA model in the same way with all the details they mentioned.

According to the Table 1, we do notice a significant difference in both violation numbers and p-value between their and our results. For the 97.5% VaR, the total number of violations they get over the years is 51 but we get 75 instead. To be specific, the numbers of our result are almost twice as theirs in several years, for instance 2020. And the p-value of theirs is 0.46, larger than 0.05, indicating that their violations are not significantly different from expected number while our p-value of 0.0016 cannot draw the same conclusion.

Table 1: The expected and actual number of VaR(97.5%) violations per year and p-value of binomial test

year	2013	2014	2015	2016	2017	2018	2019	2020	2021	p-value
Trading days	250	250	250	250	250	249	250	251	19	
expected number	6	6	6	6	6	6	6	6	0	-
peers	3	4	6	5	3	13	7	9	1	0.46
our result	8	7	8	7	2	12	11	17	2	0.0016

In Figure 5 in their report, we can not clearly identify the violations because their plot records the violations for both confidence levels. But it is distinguishable that violation clustering occurs in 2015-2016 and 2020 Q1. This observation contradicts the annual number they gave in Table7, where 2018 has the most violations. Furthermore, Figure 3 reveals that VaR violations happen more frequently between 2018 and 2020 afterwards, which means violation clustering occurs. This observation coherent with the violation counts in Table 1. Our statistical testing also confirms the presence of clustering. Thus, their given model does not seem to perform well during the whole process(including the stressed period in the pandemic).

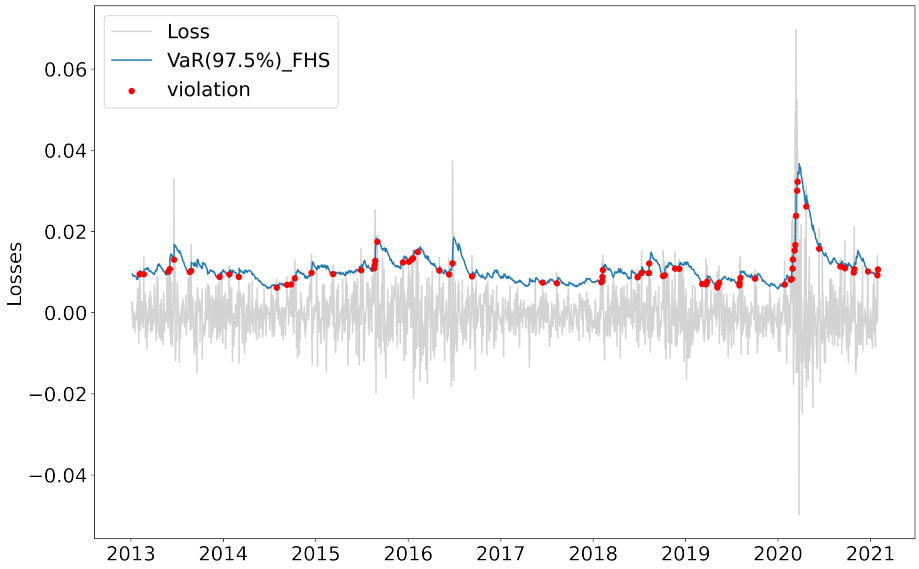

Fig. 3: The VaR(97.5%) violation results for FHS

4. Limitation

The limitation of a model usually means it only applies to specific scenarios. For instance, the given report tests their FHS-EWMA on the whole period(including the stressed period in the pandemic). However, the model does not perform well and they do not replicate the testing excluding the stressed period. The FHS-EWMA model should perform better concerning a more stable market without 2020 since the market during the pandemic is extremely volatile.

The parameter calibration in the FHS-EWMA process is another limitation. In the process of EWMA, they directly set λ as 0.94 without calibrating. Although 0.94 is commonly used in the industry, this value is not necessarily suitable in every specific case. Besides, the initial σ^2 might have a particular effect on the simulations. They treat the variance in the first 50 days as their initial σ^2, but the definition of 'first 50 days' seems ambiguous when taking the rolling window into account and other options of initial σ^2 are worthy of investigation. Further, the window size during FHS is another critical parameter. Different lengths of the window would undoubtedly lead to different performance since the VaR at next time stamp is determined by quantile of losses with number of scenarios equal to the length. Last but not least, the VaR model itself has limitations. VaR has been fundamentally criticized as a risk measure on the grounds that is has poor aggregation properties(Non-subadditivity) [1]. Moreover, the interpretation of VaR neglects any problems related to market liquidity [1].

5. Discussion

The abovementioned result differs from Luuk's result both quantitively and qualitatively. A possible explanation is that there is some missing information regarding the portfolio. We can not identify their loss operator due to lack of the weights of the assets and unclear risk factors. Certain assumptions(equal weights etc.) have to be clarified before the validation. It is impossible that validator and the model provider would make the exact same assumptions without any communication.

They provide violation number of the whole period under 99% and 97.5% confidence level in Table 2, whereas they only offer annual 97.5% result in Table 7. Therefore, we can not compare the result under 99% confidence level further.

Moreover, their model might still have some room for improvement. The parameters(window size and λ etc.) in the FHS-EWMA model are not fully investigated before decisions. Some techniques like the maximum likelihood method could apply to the parameter calibration.

6. Conclusion

This report validates the recommended FHS-EWMA by our peer student Luuk and Willem-Jan. Because of the lacking of important information in the portfolio and loss operator, we have to make certain assumptions before testing replicability. However, our result differs from their actual result significantly. From our perspective, the model is not replicable and the final validation result would be red if we must draw the conclusion.

References

1. P. Embrechts, R. Frey, and A. McNeil, "Quantitative risk management.," 2011.