Quantitative Risk Management

Assignment II (2021): Model Validation

Kho, Kevin
(2713279)
Romero León, Pedro Carlos
(2710821)

January 19, 2022

---

1 Documentation

The documentation is clear and concise in describing the data, the portfolio and its weights, the methodology, and the different methods considered. However, some results are missing and/or are wrongly interpreted in the backtesting section. For backtesting, p-values should be computed of the difference in expected number of violations and the number of violations that are derived from the model output. This p-value tests the null hypothesis that the deviations of the model output from the expected number of violations is statistically insignificant by comparing the p-values with the significance level. Whenever the p-value is smaller than the confidence level, the deviations can be considered to be statistically significant and thus the model is appropriate for estimating the VaR and ES.

In the documentation the p-values are compared to the wrong significance level and are interpreted in an opposite manner from what is described above. First of all, the p-values are compared to a significance level of 0.05 whereas the confidence levels considered for analysis of the portfolio losses are 0.975 and 0.99. Therefore the appropriate significance levels to be considered for p-value testing would be 0.025 and 0.01, respectively. This however, under the assumption that the results in Table 5 of the documentation are correct, yields no difference in conclusion. Moreover, in the documentation it is stated that whenever the p-value is smaller than the significance level, the results are statistically insignificant. Exactly the opposite is true. As stated before, the null hypothesis claims that the deviations occurring from the models occur by random chance. Whenever the p-value attains a smaller value than the significance level, this null hypothesis will be rejected, implying statistical significance of the deviations.

Although in this case these errors do not affect the final conclusion of the report, this is a critical matter as complete wrong conclusions could be drawn if the results were slightly different.

For stress testing it is not specifically stated how the portfolio loss change is defined which would clarify matters for the validation process.

2 Theoretical Assessment

For an arbitrary portfolio, the portfolio loss as a function of its risk factor changes can be defined as:

$$
\begin{aligned}
L_{t+1} &= -(V_{t+1} - V_t) \\
&= -(f(t+1, \mathbf{Z}_t + \mathbf{X}_{t+1}) - f(t, \mathbf{Z}_t))
\end{aligned}
$$

$$
L_{t+1} = -V_t \sum_{i=1}^{N} \omega_t^i (e^{\mathbf{X}_{t+1}^{(i)}} - 1) \approx -V_t \sum_{i=1}^{N} \omega_t^i \cdot X_{t+1}^{(i)} \tag{1}
$$

where $X_{t+1}^{(i)} = \Delta S^i = \log(S_{t+1}^i) - \log(S_t^i)$ and ω_t^i represent the relative weight of each asset in the portfolio. If the mapping (f) is differentiable, we may use the following first order approximation for the loss we arrive at the following Linearised Loss Operator

$$
\begin{aligned}
L_{t+1}^{\Delta} &= -\left(f_t(t, \mathbf{Z}_t) + \sum_{i=1}^{8} f_{z_i}(t, \mathbf{Z}_t) X_{t+1,i} \right) \\
&= -V_t \Bigg[\underbrace{\omega_{1,t}(X_{1,t} + X_{5,t})}_{\text{GOLD}} + \underbrace{\omega_{2,t}(X_{2,t} + X_{6,t})}_{\text{APPLE}} + \underbrace{\omega_{3,t}(X_{3,t} + X_{7,t})}_{\text{AEX}} \\
&+ \underbrace{\omega_{5,t}(y(t,T)\Delta t - (T-t)\Delta y)}_{\text{BOND}} \Bigg]
\end{aligned}
$$

where $w_{i,t} = \alpha_i S_{i,t} / V_t$.

The documentation specifies a portfolio value of 1 million and the following asset allocation:

- $(\omega_t^{gold} = 20\%)$ Gold [USD]
- $(\omega_t^{Apple} = 20\%)$ Apple [USD]
- $(\omega_t^{AEX} = 30\%)$ AEX [EUR]
- $(\omega_t^{bond} = 30\%)$ China 10 year bond 100 bp

3 Quantitative Assessment

In the documentation of the model it is stated that the Historical Simulation method provided the best results for the 97.5% VaR whereas the Variance-Covariance method performed best for the 99% VaR with 4 degrees of freedom. In this section, a quantitative assessment will be conducted on the results of these two methods. The assessment will be two fold: First, it will be investigated whether or not the model output for the number of violations given by the documentation yields realistic results compared to the validation model. Second, the results as provided in the documentation for these two models will be investigated on their replicability.

A model is deemed appropriate if the number of violations is statistically equal to the expected number of violations, dependent on the confidence level considered. For the case at hand, we tested under the null hypothesis whether the VaR models deployed are correct, p-values lower than 0.05 indicate that the observed number of violations significantly differ from the expected number of violations.

3.1 Historical Simulation

Tab. 2 and Fig. 1 summarize the violation results obtained from the validation model. Several observations can be made. First of all it must be noted that the number of expected violations as reported in the documentation does not comply with the theory underlying the violations. Recall that a violation can be defined as an indicator function that is approximately distributed as a $Be(1-\alpha)$ distribution under the null hypothesis that the VaR model is correct. For the sum of the indicator function, i.e. the number of total violations it holds that

$$\sum_{t=1}^{N} \mathbb{I}_t \sim Bin(N, 1-\alpha)$$

Therefore the number of expected violations is $\mathbb{E}\sum_{t=1}^{N} \mathbb{I}_t = N(1-\alpha)$ and its variance $\mathbb{V}ar\sum_{t=1}^{N} \mathbb{I}_t = N(1-\alpha)\alpha$. Since $N = 2351$ and $\alpha = 0.975$, this implies that the total number of expected violations is $2351(1-0.975) = 58.8$. This contradicts the value given in the documentation.

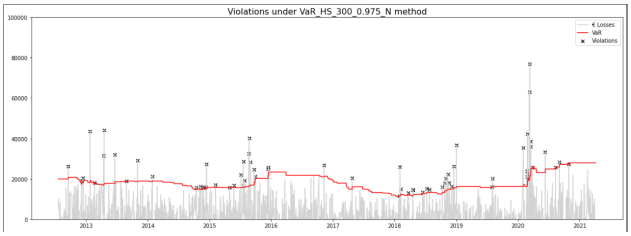
Moreover, the number of violations found under the validation model are greater than what was documented which implies that the results are not reproducible with the validation model. According to our calculations neither the number of violations, nor the significance, corresponds to that presented in the documentation. The statistical test shows the p-value obtained indicates that the difference between number of violations found under the simulation model and the expected number of violations is statistically significant. However, the statistical insignificance found under the model used to produce the documentation results could also be a result of wrong implementation of the test-statistic. Note that for computing the p-values the deviation between the model output and the expected value is tested on statistical significance by the binomial test statistic

$$T(N) = \frac{\sum_{t=1}^{N} \mathbb{I}_t - N(1-\alpha)}{\sqrt{N(1-\alpha)\alpha}} \sim \mathcal{N}(0,1)$$

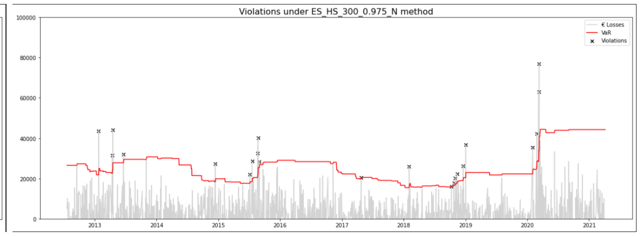
from which its p-value is computed and compared to the significance level α. From this, it can be seen that a wrong input for the expected number of violations drastically affects the test-statistic and thus also the p-value. Therefore, it is recommended to run the p-value test again using the appropriate number of expected violations, i.e. 58.8, and check if the conclusion provided by the documentation remains correct.

Table 1: Violation results as provided by the documentation compared to results obtained from model validation. Results obtain under Historical Simulation with 0.975 confidence level

	Documentation	Model Validation
0.975 Quantile		
Expected	46.2	58.8
His-sim	50 (0.55)	66 (0.037)



(a) VaR violations over time under HS model. $\alpha = 0.975$



(b) ES violations over time under HS model. $\alpha = 0.975$

Figure 1: VaR and ES violations of portfolio losses over time for 97.5% confidence level under HS model.

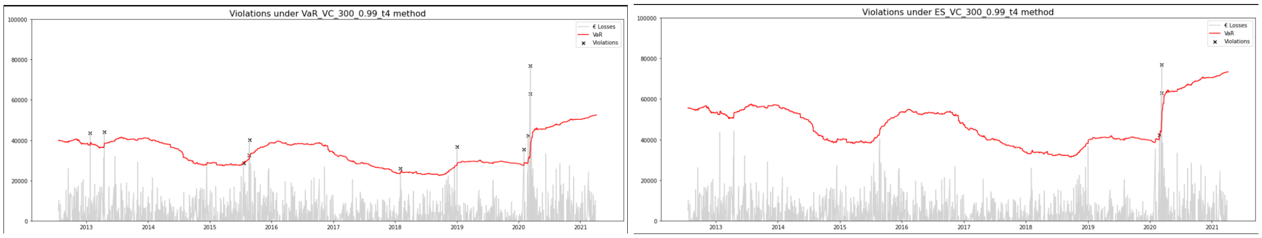
3.2 Variance-Covariance-t4

Working with sample data entails determining whether a parametric, or alternatively a non-parametric approach would better fit the data at hand. Parametric modelling often assumes the data could be described by a known distribution, which often tends to be assumed as a Gaussian distribution. If assumptions of normality are violated, one could decide to use a non-parametric approach, or assume a different distribution that best fits the data. For the case at hand, since assumptions of normally distributed returns are rather naive, we tested for non-normality, and, from the results, assumed that portfolio returns are Student-t distributed, since this distribution assigns more weight to extreme values (fatter tails).

Although it seems reasonable to choose 4 degrees of freedom to characterise the distribution, tail wise, of the risk factors included in the portfolio, we believe increasing the degrees of freedom by 1 unit would yield a better fit to the data´s distribution. Furthermore, contrary to what is stated in the documentation, our models differ significantly in the measurement of violations, in absolute terms, and their statistical significance. Differences in the models´ outputs are presented in Table 2 and Figure 2. From the results one could see that, according to our models, the difference between expected and observed violations is large in absolute terms, the VaR using the Var-Cov approach seems to overestimate the risks of holding the portfolio. In addition, according to our calculations, this difference seems to be statistically different from 0, hence we believe that this selected method does not seem to be accurate enough to correctly assess the risk exposure of holding the portfolio over the horizon considered.

Table 2: Violation results as provided by the documentation compared to results obtained from model validation. Results obtain under Variance-Covariance method with 0.99 confidence level and 4 degrees of freedom.

	Documentation	Model Validation
0.99 Quantile		
Expected	20	23.51
VC	20 (0.72)	11 (0.034)



(a) VaR violations over time under VC-t4 model. $\alpha = 0.99$

(b) ES violations over time under VC-t4 model. $\alpha = 0.99$

Figure 2: VaR and ES violations of portfolio losses over time for 99% confidence level under VC model.

4 Limitations

4.1 Limitations of Historical Simulation

Although this approach comes with the advantage of being a rather easy to compute measure, it comes with a number of disadvantages when measuring risk, as the implicit assumption that future returns will behave in a similar way to past returns might yield misleading information since forecasts of VaR will never exceed the biggest recorded loss, and future returns may show completely different trends. Moreover, the use of a one-day holding period (or ten days for regulatory capital calculations) assumes that all positions in the portfolio can be liquidated or hedged in one day. In periods of when illiquidity reigns in the market this assumption may not hold. Furthermore, the use of a 99% confidence level means that Historical VaR (HVaR) does not take into account any losses beyond this confidence level. Finally, information availability and transparency may also be a big disadvantage.

4.2 Limitations of Variance-Covariance method

The main limitation of the models selected in the documentation, the HS and Var-Cov, is their unconditional nature. Risk management techniques are typically interested in the distribution of the losses in the portfolio, the empirical distribution, from which typical measures such as ES and VaR are calculated. However, these kind of tools tend to put on the same pile returns/losses coming from stressed and normal periods. Furthremor, under volatile periods the risk estimations should change, since it is likely that over the next periods one could observe different volatility clusters, as its the case for figure 1 during the periods 2015-2016, 2018-2019 early 2020´s.

5 Conclusion

Overall, we believe the report and documentation was significantly well written, despite some minor details. However, the quantitative assessment of the models selected shows a clear lack of replicability and suitability to the portfolio. Our results show that none of the models would accurately assess the risk exposure of holding the portfolio over the considered period. For the portfolio presented we would rather recommend using a Variance-Covariance approach with 5 degrees of freedom at the 99% confidence level, since according to our models the number of expected (23.1) and observed violations (15) would not be statistically significant from 0.