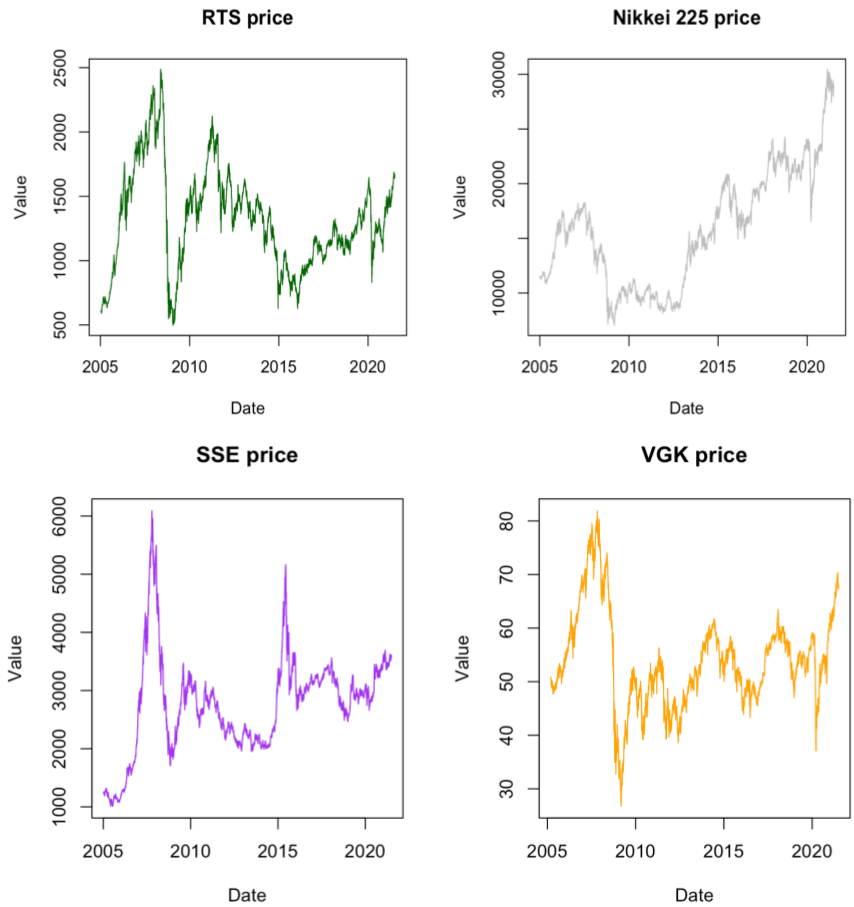
Motivation

We often hear that some assets or markets are "riskier" than others. While this makes sense from the intuitive point of view, the question is, whether we could show this numerically, on the real data using real numbers? This project will answer this question by taking several stock indices from several markets, both developed and developing, and computing the risk measures associated with these indices using parametric and non-parametric approaches. By using different approaches to computation, we would be able to implicitly answer another question: what approach to calculation is better from the risk management standpoint? While we know that non-parametric approach is free from limiting assumptions about the distribution of our data which would probably lead to more accurate risk picture, it is more computationally burdensome than the parametric approaches which could be an issue for large datasets. So, maybe parametric approaches would be able to capture the risk while staying more computationally efficient?

Moreover, in this project the overall riskiness will be evaluated not only by a single number from risk measure, but also by analyzing the impact of key parameters on that risk measure. These key parameters include the confidence level and the time horizon over which the risk measure is computed. By doing this, we hope to get a more or less clear picture of what markets are riskier and what are less.

Data

As our data we took stock market indices which represent the general picture of that particular market of 5 different markets: US, Japan, Europe, Russia and China. For the US we took SPY and QQQ ETF's (the first tracks S&P500, the second – Nasdaq 100), for Japan – Nikkei 225 which is the largest stock index in that country, for Europe – VGK which is an ETF from Vanguard which tracks performance of major European companies, for Russia – RTS which is the largest stock index on that market, and for China – SSE which is an index for Shanghai Stock exchange. The data range is from 2005 to July 2021. By taking such large data span we include the financial crisis of 2007-2009, periods of growth of 2005-2006 and 2010-2019 and the COVID crisis of 2020. Thus, we should have a more or less complete picture of the riskiness of these assets. The graph below illustrates the evolution of the values of all these indices and we can clearly see multiple boom and crisis periods on all of them.



Judging from the graph, RTS, SSE and VGK indices seem the most volatile, followed by Nikkei 225, while US indices seem the less volatile. Nevertheless, further analysis is needed.

Methodology

We analyze the log-returns of all the assets as they are stationary. Firstly, we plot the distributions of the returns, then obtain summary statistics and after proceed to the calculation of actual risk measures. We calculate daily **Value-at-Risk and Expected Shortfall** using the following methods:

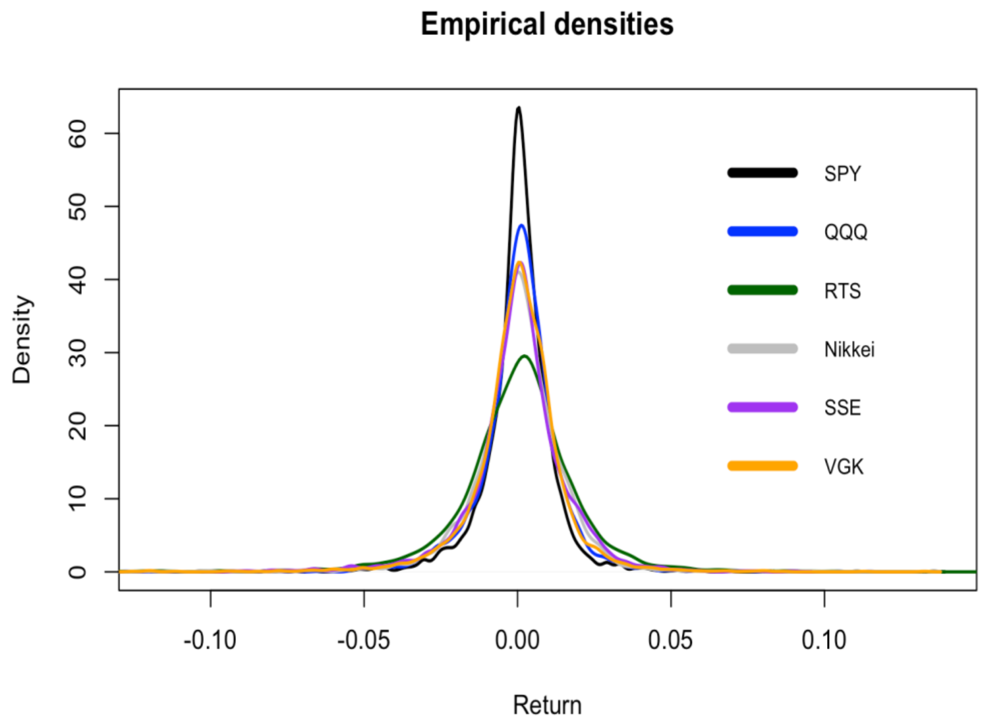
- Historical Simulation, which is a non-parametric method

  And parametric method assuming two different distributions:
- Normal: $VaR(\alpha) = -\mu - \sigma * \Phi^{-1}(\alpha)$; $ES(\alpha) = -\mu - \frac{\sigma}{1-\alpha} * \varphi(\Phi^{-1}(\alpha))$. The most standard distribution possible.
- Student's t: $VaR(\alpha) = -\mu - \sigma * \sqrt{\frac{\upsilon-2}{\upsilon}} * t^{-1}(\alpha, \upsilon)$; $ES(\alpha) = -\mu - \frac{\sigma}{1-\alpha} * \left(\frac{\upsilon + t^{-1}(\alpha)^2}{\upsilon-1}\right) * g_{\upsilon}(t^{-1}(\alpha, \upsilon))$. Has heavy tails which are common for financial data. The "fatness" of the tails depends on the value of the parameter υ -degrees of freedom. The lower it is, the heavier would be the tails. In this project we calculate VaR and ES for t distribution with 2 different df: 3 and 5. This will allow more variability and freedom to choose which distribution fits the data the most.

For the second part of the project, we calculate the risk measures for different confidence levels and time horizons and then plot these relationships.

Results

Empirical distributions:



We observe different peak sizes, but long tails for all indices which is common for financial data.

Summary statistics:

INDEX	MIN	MAX	MEAN	SD	SKEWNESS	KURTOSIS
SPY	-11.59%	13.56%	0.03%	1.20%	-0.39	15.45
QQQ	-12.76%	11.48%	0.05%	1.34%	-0.35	8.67
VGK	-12.53%	13.27%	0.01%	1.54%	-0.69	10.62
RTS	-21.20%	20.20%	0.02%	2.00%	-0.55	11.98
NIKKEI	-12.11%	13.23%	0.02%	1.48%	-0.49	8.03
SSE	-9.26%	9.03%	0.02%	1.59%	0.57	4.76

We observe practically 0 mean, high kurtosis (indicator that the true distribution is far from normal), negative skewness for all indices. Moreover, we see that SPY and RTS have the largest value of kurtosis suggesting longest tails and potentially highest riskiness. However, in terms of standard deviation, RTS and SSE have the largest one. So, we need to compute real risk measures in order to find out for sure.

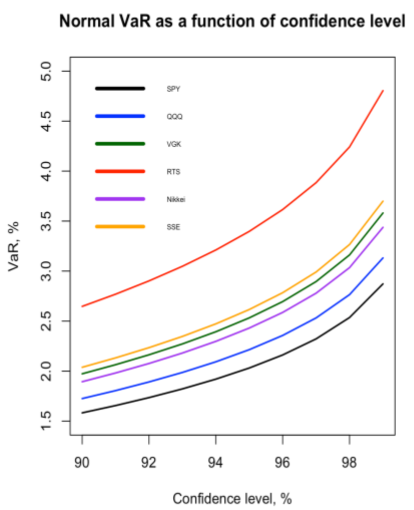
VaR and ES calculation results:

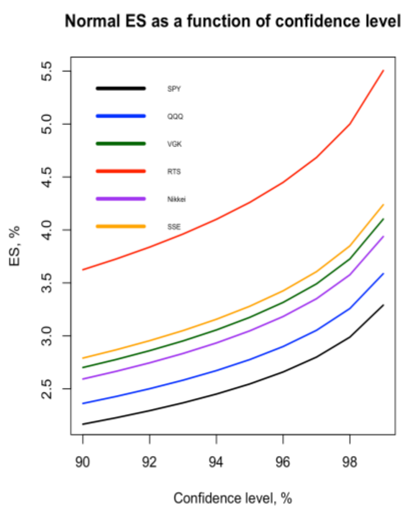
					VaR			
CL	5%					1%		
Method	HS	Normal	Student t 5 df	Student t 3 df	HS	Normal	Student t 5 df	Student t 3 df
SPY	1.8%	2.0%	1.9%	1.7%	3.7%	2.9%	3.2%	3.2%
QQQ	2.2%	2.2%	2.1%	1.8%	3.9%	3.1%	3.5%	3.5%
VGK	2.4%	2.5%	2.4%	2.0%	4.9%	3.6%	4.0%	4.0%
RTS	3%	3.4%	3.2%	2.8%	5.9%	4.8%	5.4%	5.4%
SSE	2.5%	2.6%	2.5%	2.2%	5.4%	3.7%	4.1%	4.2%

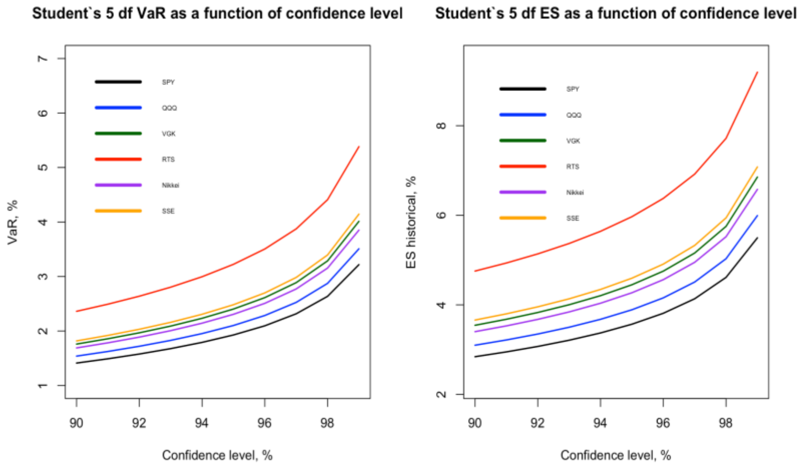
					ES			
CL	5%					1%		
Method	HS	Normal	Student 5 df	Student 3 df	HS	Normal	Student 5 df	Student 3 df
SPY	3.1%	2.5%	3.6%	4.8%	5.6%	3.3%	5.5%	8.6%
QQQ	3.3%	2.8%	3.9%	5.2%	5.4%	3.6%	6.0%	9.4%
VGK	3.9%	3.2%	4.5%	6.0%	7.0%	4.1%	6.9%	10.8%
RTS	5.1%	4.3%	5.9%	8%	9.2%	5.5%	9.2%	14.5%
SSE	4.1%	3.3%	4.6%	6.1%	6.6%	4.2%	7.0%	11.1%

We see a significant difference in computed risk measures across assets for all of the methods. The RTS index possesses the highest risk, followed by SSE, VGK, Nikkei, QQQ, SPY. This is generally consistent with the perception of what markets are riskier. Moreover, the normal parametric method slightly overestimates 95% VaR compared to historical simulation and significantly underestimates all other measures, especially Expected Shortfall. On the contrary, Student's t- distribution with 3 df tends to overestimate risk, especially in the tail. Thus, the best parametric method (which gives the closest results to historical simulation) is Student's t-distribution with 5 df.

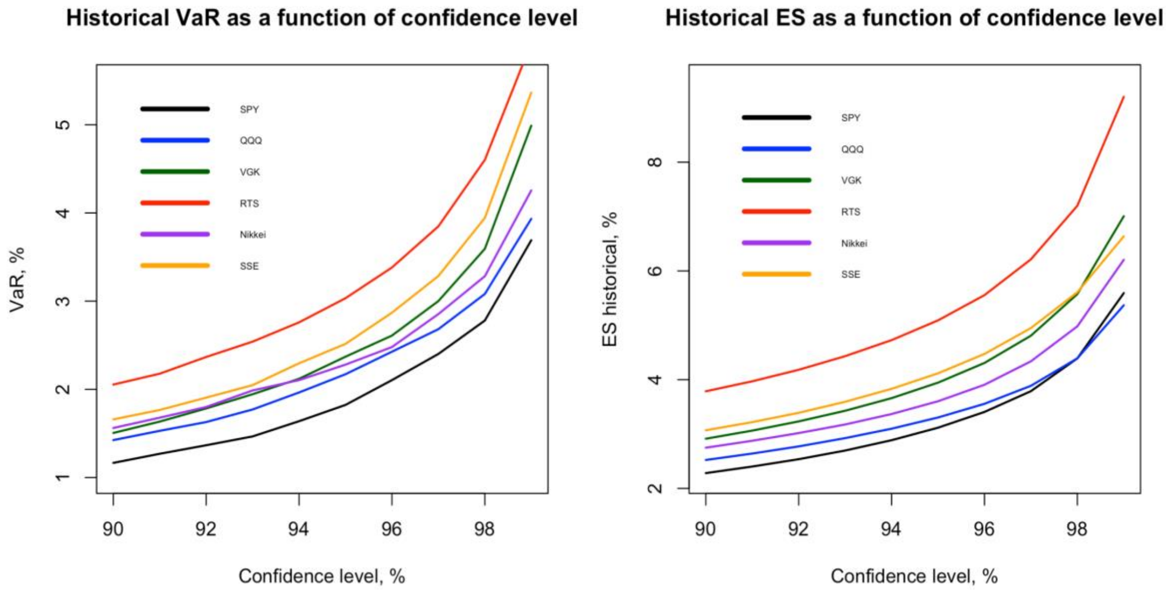
Impact of the confidence level:







Firstly we plot the results for two parametric methods. Both produce smooth, non-intersecting curves. The difference is in levels: t distribution produces higher values of risk measures than Normal for high confidence levels and lower values for low confidence levels. The lines being parallel implies that under parametric methods for all confidence levels one asset is necessarily either more or less risky than another. Next, we proceed to the HS:

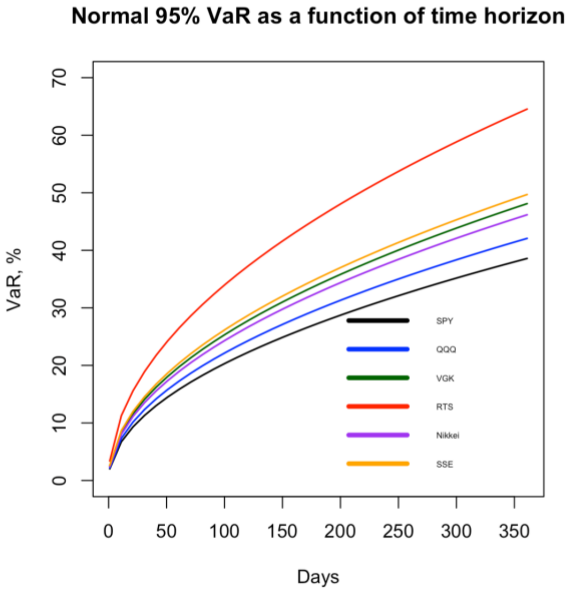


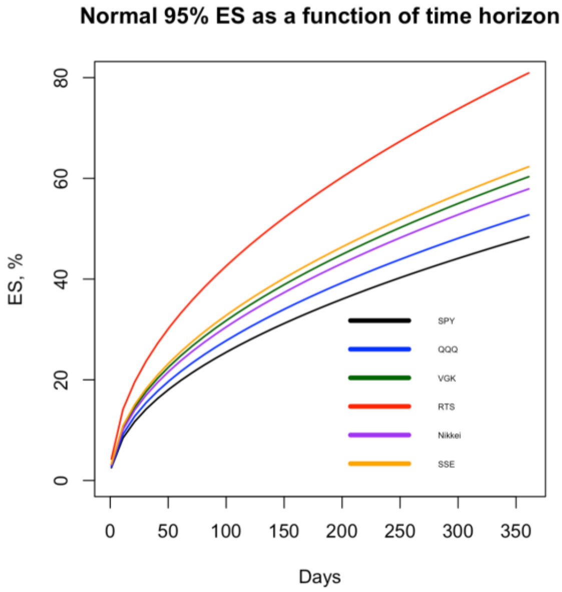
The picture is more interesting here: the lines are not necessarily parallel and may even intersect each other. Thus, on some confidence levels an asset may be less “risky” than another asset, while on a higher confidence level it becomes riskier. Especially this phenomenon is pronounced for the Expected Shortfall: at 98 and 99% CL ES of SPY crosses ES of QQQ and ES of VGK crosses ES of SSE. Still Russian RTS is the unambiguous “leader” in terms of risk: the red line is much higher than all other lines.

We also see that in all cases the relationship between risk measure and confidence level is convex. Thus, as we increase confidence level, the asset becomes increasingly riskier.

Impact of time horizon:

Here we performed analysis only under the assumption of normality and for 95% CL. Under Normal distribution, the n-day volatility is obtained by scaling 1d volatility by square root of n





As expected, we got concave relationships.

Conclusions

In this project it was found out that indeed riskiness highly depend on the market. In this study Russian RTS index showed the highest VaR and ES numbers for all methods of calculation which is in line with the general perception of Russian market being riskier than European or American. Moreover, it was shown that usage of Normal distribution significantly underestimates risk numbers, especially for tail risk: ES numbers were much smaller than those computed historical simulation.

Finally, the graphical analysis revealed that risk measures computed via Historical Simulation approach do not necessarily need to have same “ranking” for different confidence levels: if a certain asset had lower risk at 95% confidence level than another asset, it doesn’t have to have lower risk at 99% confidence level. And the further we move to the tail, the more surprising the results can be, the more irregular the shape of the curves we could obtain.

This gives a thought that if we seek to apply parametric method to computation of risk, we may need to assume different distributions for different parts of data.