HSBCQA Project Description

Michał Chmura, Maja Zakrzewski, Seweryn Turula,
Daniel Biernat, Raman Adamovich

January 22, 2023

Contents

1 Introduction

In this paper we will present our take on quantitative tools used for Value-at-Risk projections. Our main objective was to estimate VaR using several methods, in the form of case study on the daily returns of major stock indexes. In the following analysis, we assume the probability level at 1%. The data is sourced, with the quantmod R package, from Yahoo Finance. The analysis was done in R ver. 4.2.2, version control via Git and the reports were written in Overleaf.

2 Preliminary Data Analysis

2.1 Data description

We are examining returns from four indexes: SNP500, Dow Jones Industrial Average, NASDAQ Composite and DAX in the period from 2018-01-03 to 2022-12-30. All data comes from Yahoo Finance.

The NASDAQ Composite is a capitalization-weighted index that includes almost all stocks listed on the Nasdaq stock exchange with USD as the numeraire. The SNP500 is a capitalization-weighted stock market index tracking the stock performance of 500 largest companies listed on stock exchanges in the United States with USD as the numeraire. The Dow Jones Industrial Average is a price-weighted index that includes 30 prominent companies listed on stock exchanges in the United States. The DAX is is a stock market index consisting of the 40 major German blue chip companies trading on the Frankfurt Stock Exchange. It is a total return index with EUR as the numeraire.

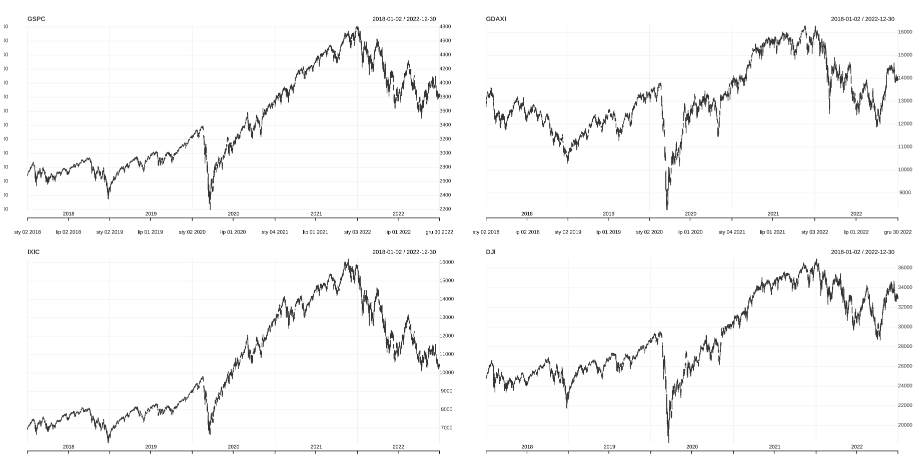

Figure 1: Historical data of indexes

Figure 1 presents the behavior of the indexes on the chosen period. All of them suffered significant loses during the march of 2020 and were growing in the period from April 2020 to the end of 2021. After that period there is significant decay in the prices of all indexes. Figure 2 presents the PnLs of examined indexes which will be the foundation of our further estimation of risk and backtesting.

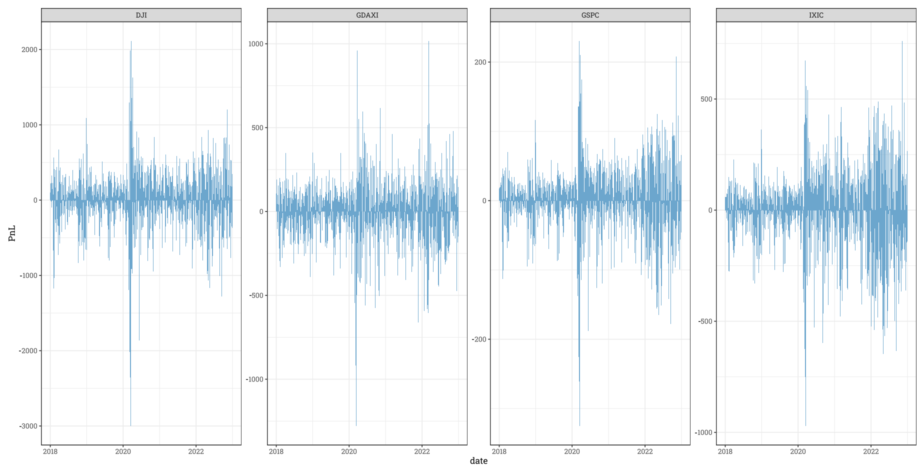

Figure 2: PnLs

2.2 Statistical properties of PnLs

This section contains tests and visualizations of the main statistical properties of the PnLs sample such as independence, homogeneity, and normality.

2.2.1 Normality

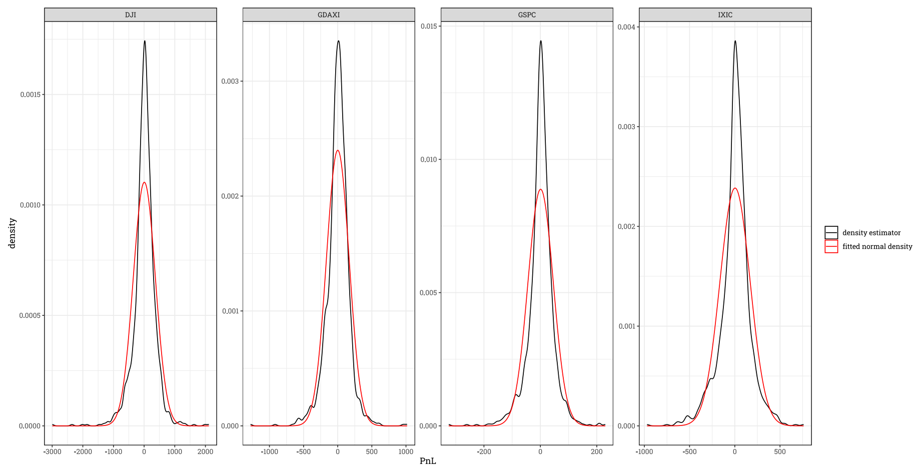

Figure 3: Distribution of PnLs

Figure 3 presents the Kernel Density Estimators of the distributions of PnLs of examined indexes and fitted normal density. This plot leads to the concusion that our data is leptokurtic and skewed in favor of the left tail. The most popular normality tests such as Shapiro-Wilk, Jarque-Bera, and Anderson-Darling are returning p-value with the order of magnitude less than $10^{-}15$, with a null hypothesis that the sample is drawn from the normal distribution. The lack of normality may impact normal and unbiased normal VaR estimators.

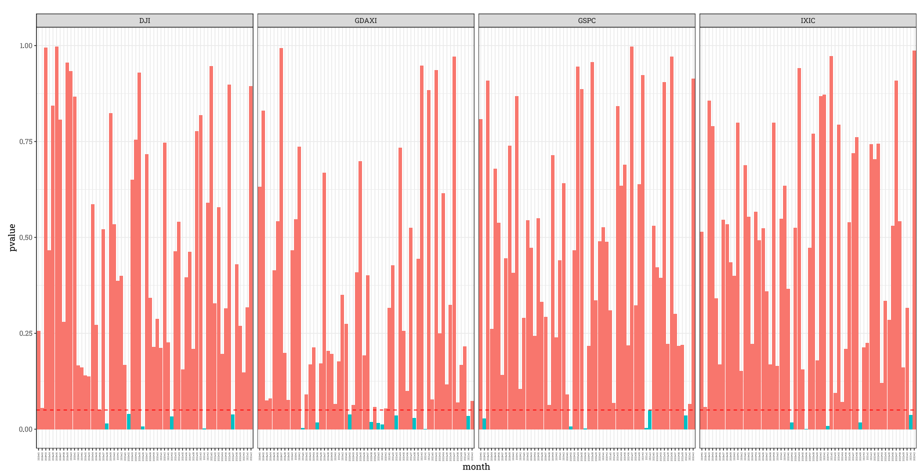

Figure 4: Normality test in monthly subsamples

However, the lack of normality on the full data may not implicate the same behavior in the shorter subsamples of the time series. In order to test that, PnL normality was tested for each month separately. The results of the Saphiro-Wilk test for subsamples are shown in figure 4. This plot shows that PnLs don't follow the normal distribution event on short periods.

2.2.2 Homogeneity

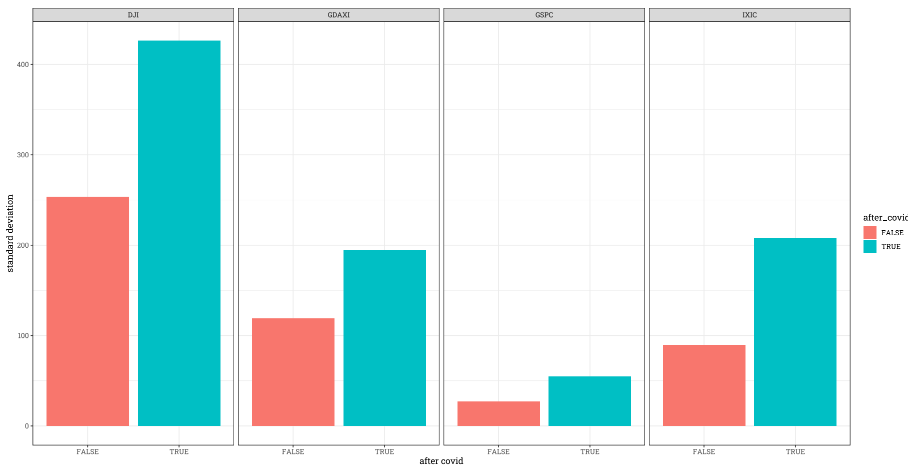

Figure 5: Standard deviation before and after covid

In order to test the homogeneity of the samples, the data was split into two subsamples. The subsample of the PnLs before the covid turmoil and the PnLs after the beginning of covid. The breaking point was chosen as the 1st of march 2020. Figure 5 shows the standard deviations of the subsamples for each PnL. One can observe that the standard deviation is almost doubled in the second subsample for every index. In order to get the quantitative result, the Fligner-Killeen test was performed. This is the test for homogeneity which is robust to a lack of normality. The null hypothesis that the samples are drawn from the same distribution is unlikely to be true due to a very low p-value with an order of magnitude less than $10^{-}12$. The Fligner-Killeen test of homogeneity for samples splited by month gives a p-value less than $10^{-}16$ for each PnL, therefore the hypothesis that the samples in short periods are homogeneous is also unlikely to be true. The lack of homogeneity may impact the VaR estimators, especially those with longer lookback periods.

2.2.3 Independence

In order to test the independence of the samples, the autocorrelation vector with a maximum lag of 30 was calculated, and can be seen in Figure 6. The autocorrelation doesn't seem to be significant. However, Figure 7 which shows the autocorrelation of absolute values of PnLs is significantly high and persists through the 30 days window. This property may be interpreted as evidence of the volatility clustering in the data. The lack of independence may significantly impact VaR estimators. In this case, the estimators based on time series analysis may outperform the estimators presented in this project.

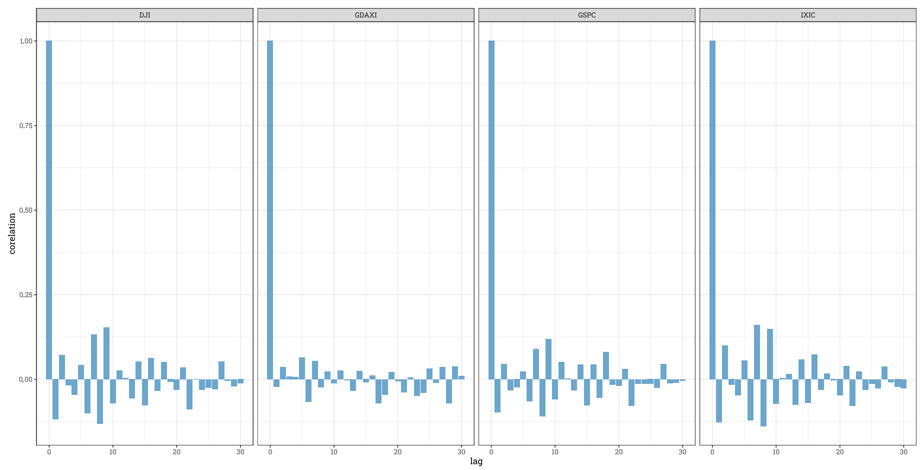

Figure 6: Autocorrelation on PnLs

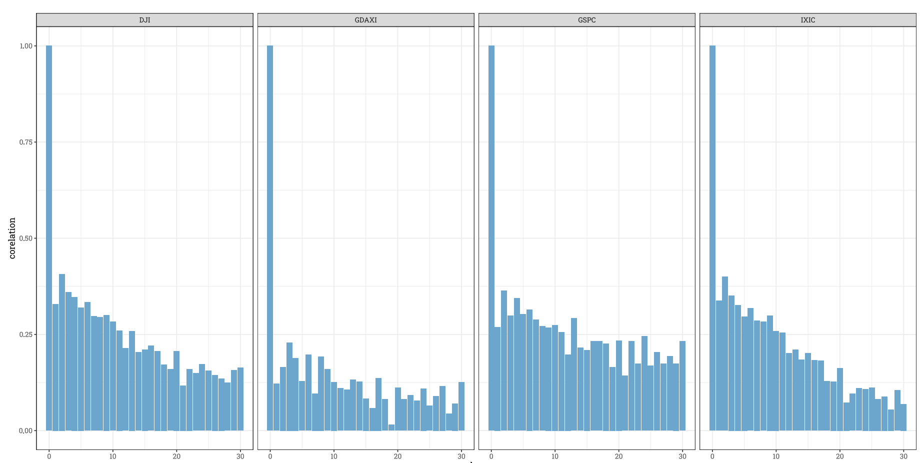

Figure 7: Autocorrelation of ablolute values of PnLs

3 VaR Estimators

We decided to use the following estimators:

- Normal Estimator:
$$-(\bar{x} + \bar{\sigma}(x)\Phi^{-1}(\alpha)) \tag{1}$$
It is based on the mean and standard deviation of the sample, also utilizes properties of standard normal distribution.

- Unbiased Normal Estimator:
$$-(\bar{x} + \bar{\sigma}(x)\sqrt{\frac{n+1}{n}}t_{n-1}^{-1}(\alpha)) \tag{2}$$
The main difference between this one and the previous one is the bias correction and the usage of the t-student distribution instead of standard normal one.

- Empirical Estimator:
$$-x_{(\lfloor n\alpha \rfloor + 1)} \tag{3}$$
This estimator is based on a sorted sample of historical results - it returns the threshold value. Simple to implement and interpret, is not based on normal distribution.

As mentioned before, we have used those three estimators considering two different lookback periods: 200 days and 500 days. We have graphically presented the results so we can understand them better. First, let us have a look at the performance of our estimators for the shorter lookback period.

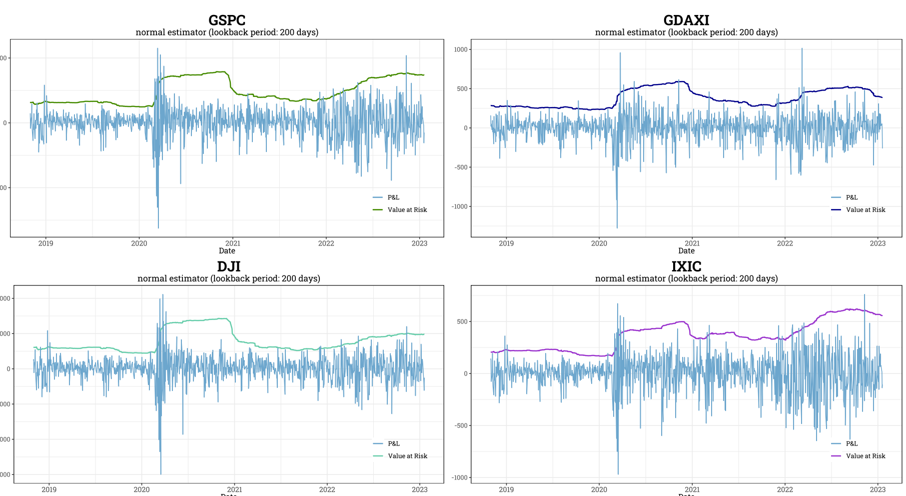

Figure 8: Normal estimator (200 days)

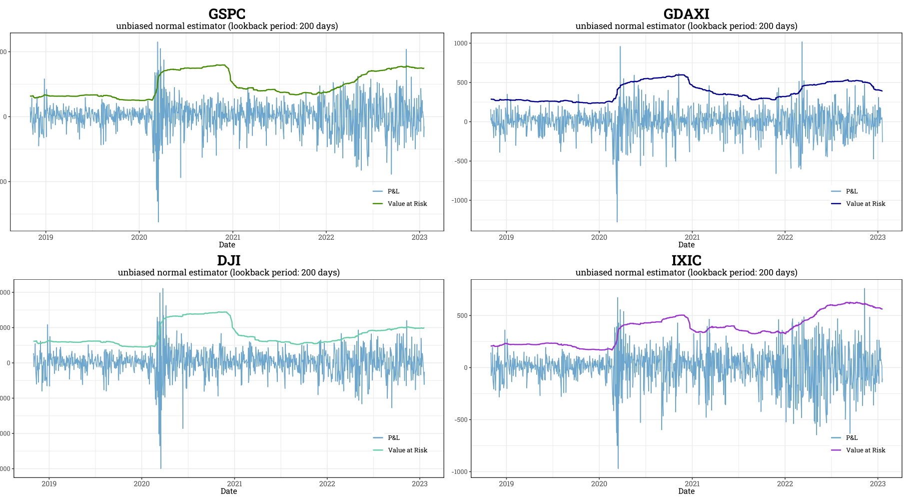

Figure 9: Unbiased normal estimator (200 days)

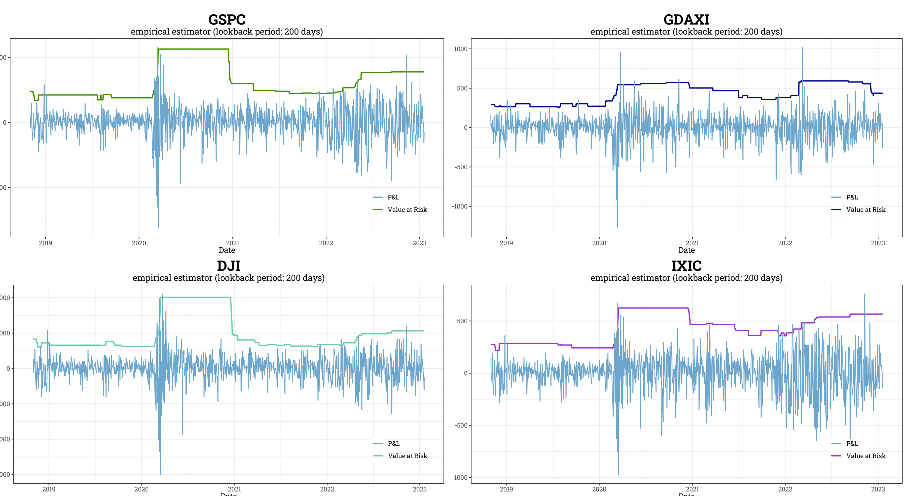

Figure 10: Empirical estimator (200 days)

Looking at those figures we can observe that all of the estimators give acceptable results for the majority of the investigated period. The only exception is year 2020, where we can observe much worse performance. We will analyze it in detail in the other part of this paper. Now let us take a look at corresponding plots for 500 days.

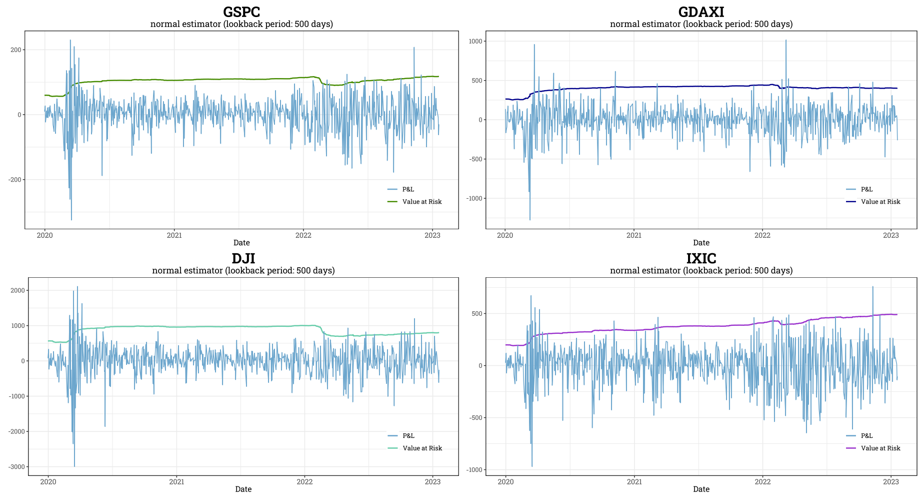

Figure 11: Normal estimator (500 days)

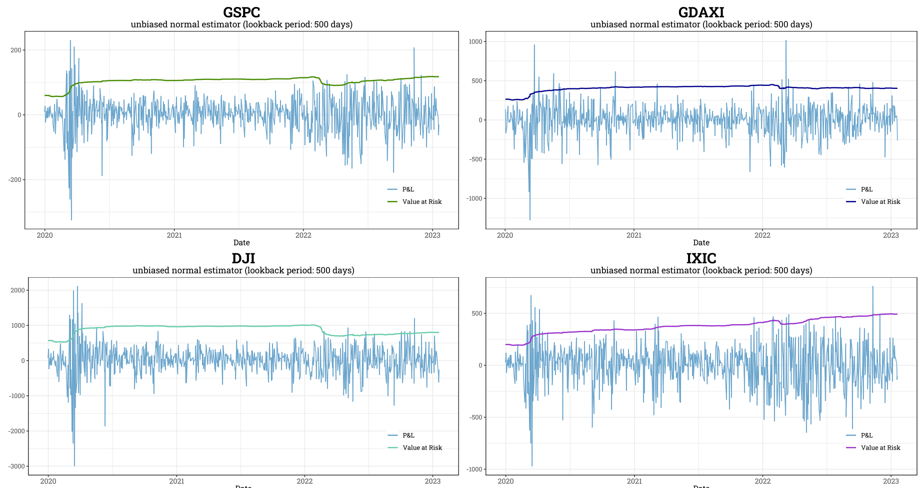

Figure 12: Unbiased normal estimator (500 days)

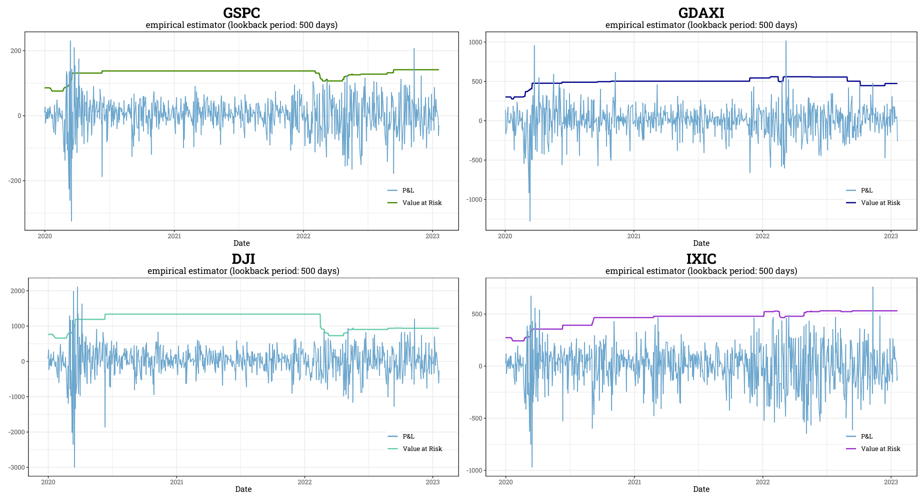

Figure 13: Empirical estimator (500 days)

We can see that our estimators with shorter lookback period were much more sensitive that the ones we look at now. The curves on the plots tend to be flat on the 500-day lookback period estimator. They also seem to cover most of the observed volatility of our PnLs. However, they do not respond to the local fluctuations of the loss distribution. Now we can focus on the plots for year 2020 only.

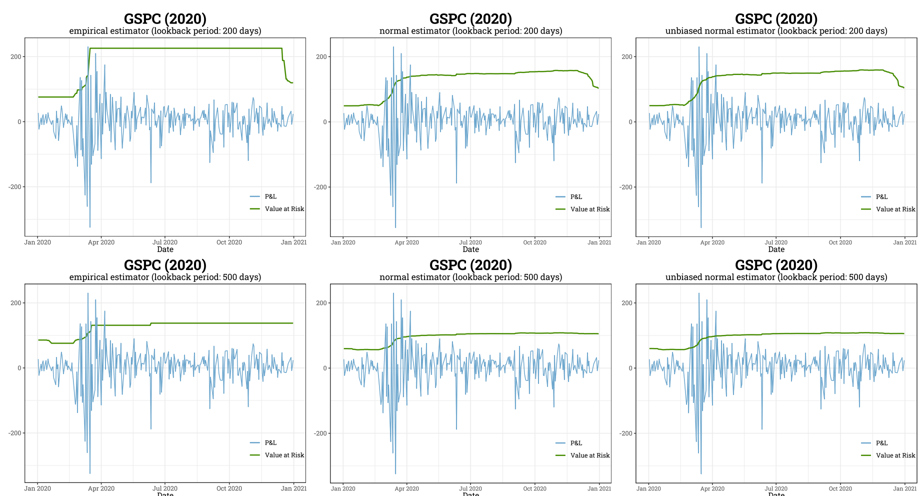

Figure 14: GSPC in 2020

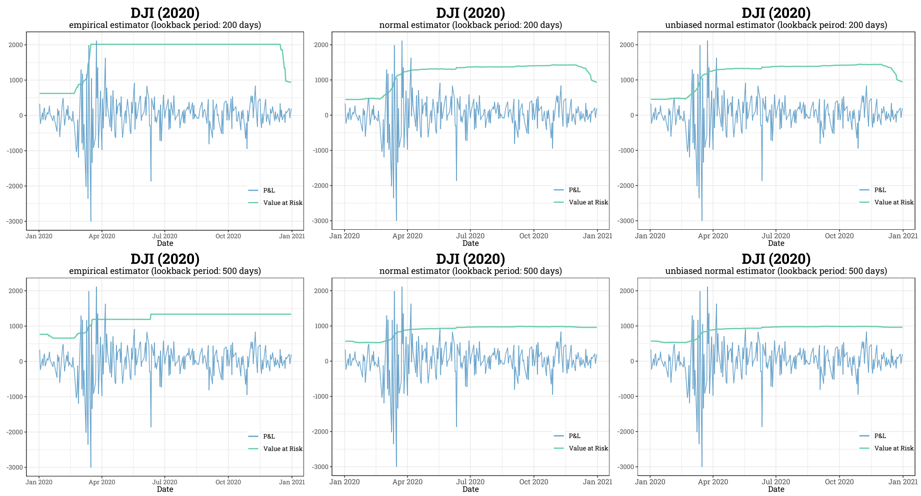

Figure 15: DJI in 2020

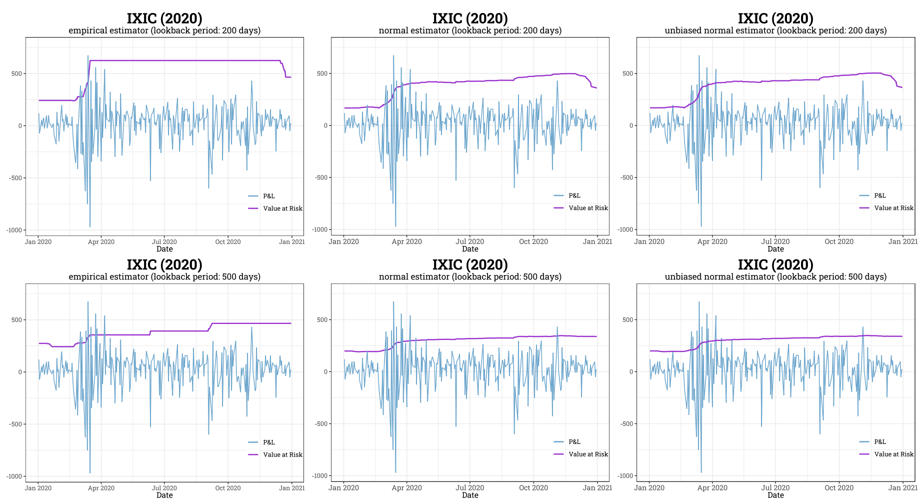

Figure 16: IXIC in 2020

Seeing more precise plots for the first three indices we can try to compare the examined estimators. We can see that the empirical estimator is the most sensitive one and that it increased as soon as the loss distribution started increasing. On the other hand, it is visibly overestimating VaR in the period after the first covid shock.

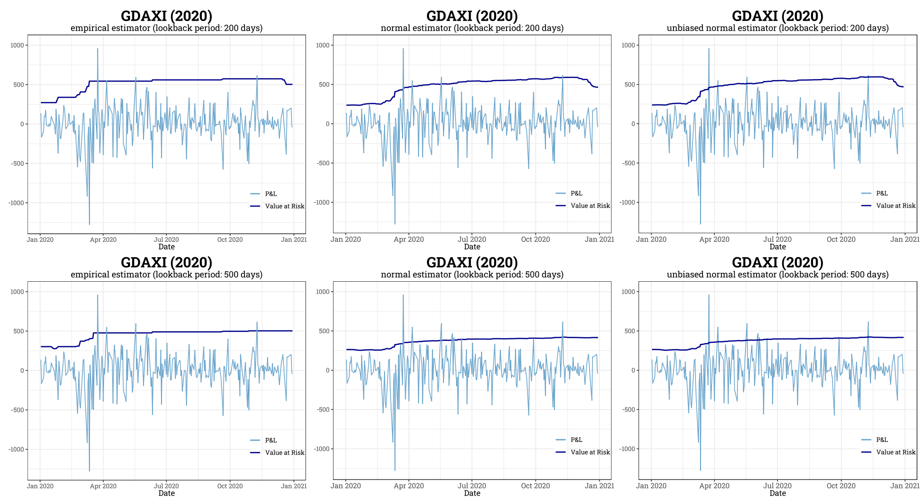

Figure 17: GDAXI in 2020

For the last index we can see that the difference between estimators is blurring, which is probably caused by the shape of the loss distribution curve. In general, we can see that in the covid period the empirical estimator worked best and that the lookback period of 200 days was significantly better than 500 days.

4 Standard Regulatory Backtest

This section contains Standard Regulatory backtest of 3 VaR estimators. We use 250 days lookback period for 99 % VaR.

4.1 Methodology

- We use i to denote i-th backtesting day. Defining x_i as i-th day Profit and Loss (PnL), and $\hat{p}_i = \hat{p}_i(x_{i-1}, \ldots, x_{i-k})$ as i-th day VaR.
- Then defining y_i as a sum of i-th day PnL and i-th day VaR.

$$T_k(y) := \frac{1}{k} \sum_{i=0}^{k} 1_{\{y_i < 0\}} \tag{4}$$

- By using Formula 1 we count the number of capital breaches (exceptions) divided by lookback period ($k = 250$) receiving exception rate.

- For VaR at level 1 % the regulator is using $k = 250$ and the model said to be in:

    1. **Green zone**, if there are **less than 5 breaches**: Under the correct model, this is expected to happen in around 90 % of all cases and corresponds to $T_k \in [0.00, 0.02)$;
    (*This corresponds results that do not themselves suggest a problem with the quality or accuracy of a bank's model.*)
    2. **Yellow zone**, if there are **between 5 and 10 breaches**: Under the correct model, this is expected to happen in around 10 % of all cases and corresponds to $T_k \in [0.02, 0.04)$;
    (*This encompasses results that do raise questions in this regard, for which such a conclusion is not definitive.*)
    3. **Red zone**, of there are **more than 10 breaches**: Under the correct model, this is expected to happen in around 0.01 % of all cases and corresponds to $T_k \in [0.04, 1.00)$;
    (*This indicates a result that almost certainly indicates a problem with a bank's risk model.*)

- Also, Regulatory backtest assuming the sample is not an i.i.d. sample

4.2 Normal and Normal Unbiased VaR backtest

Normal VaR (1) and Normal Unbiased VaR (2) assumes the sample to be an i.i.d. where in practice, as shown in first section, we can't expect our sample to be one. Such, Regulatory backtesting is a good choice in the case of not an i.i.d. sample. As we see at the graphs, Figure 19 and Figure 20, Normal and Normal Unbiased estimators start to perform worse during the Covid period. We see a quick raise and the red zone in the middle of year 2020. Green zone is expected to be in 90 % of all cases given the model is correct. Our results give us the ground to reject models for Normal VaR and Normal Unbiased VaR.

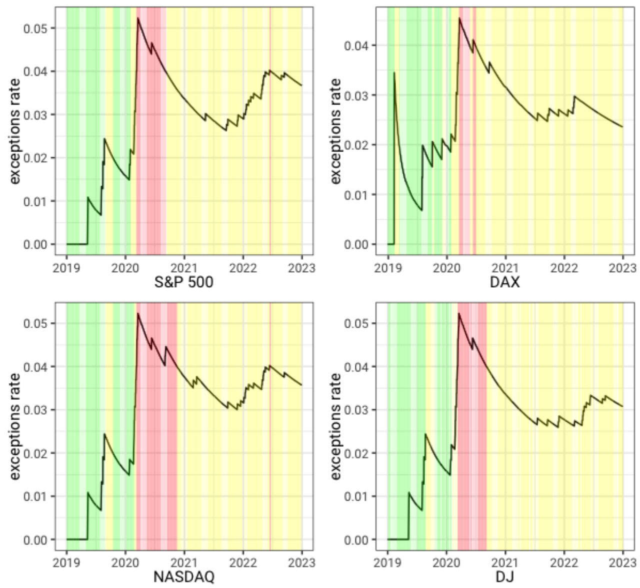

Figure 18: Normal VaR backtest

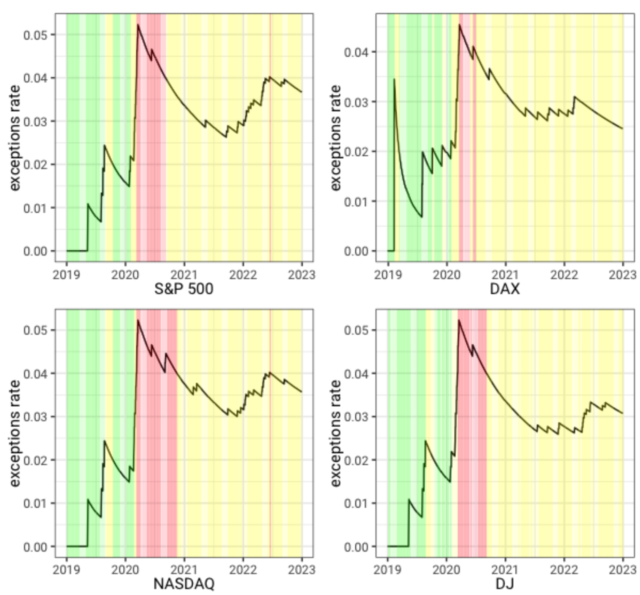

Figure 19: Normal Unbiased VaR backtest

4.3 Emperical VaR backtest

For the Emperical VaR estimator we observe in Figure 21 the green zone is almost everywhere and quick raise in the middle of 2020 (Covid). Such, based on the regulatory backtesting we accept the Emperical VaR model.

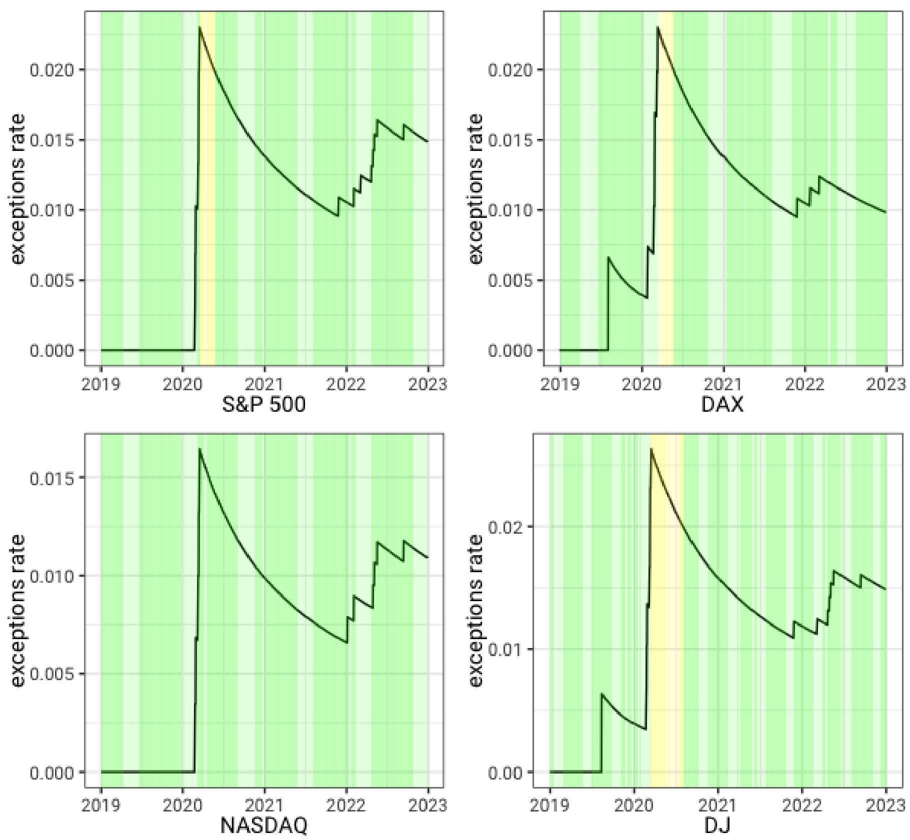

Figure 20: Emperical VaR backtest

4.4 Backtest conclusions

In conclusion, we are rejecting Normal and Normal Unbiased VaR estimators, and accepting Emperical VaR as our final estimator. Normal and Normal Unbiased VaR assumes sample to be a normally distributed one, but in practice we do not meet neither of those assumptions. As we see comparing graphs, Emperical VaR (which does not assume anything) works better for the given period of 5 years for our 4 indices.

5 Conclusion

Based on the analysis of plots, backtesting and model assumptions we conclude that the empirical estimator seems to be the best.
The following actions can be done to improve the result:

- The clustering of breaches, lack of homogeneity, and dependence observed in the data may imply that time series based estimators would be more appropriate for this analysis.
- We can use another backtests (e.g. *PIT-based* or *mean quantile score evaluation*). For example, comparing Regulatory backtest to PIT-based. Regulatory does not assume a sample to be i.i.d.. Then (in the case of i.i.d. sample) PIT-based backtest will provide us with more accurate picture of backtesting for Normal and Normal Unbiased VaR estimators.