Test Plan and Results Documentation

Jordan Boucher - jpb2232

December 2023

Contents

1 Introduction

This document serves to meet the requirements of two parts of the project, the "Test Plan" and the "Test Results."

First, we will discuss how we plan on testing our risk management model. There will be three parts to our testing. The first is component testing, which looks at the intermediate results of the model (which are not easily viewable by the end user) to see if they are reasonable. We will then perform accuracy testing. One difficult of running numerical code is that we do not normally have a particular belief about the correct answer. That is to say, we could get answer that seems reasonable even if the method is wrong. Thankfully, we have access to thorough testing by previous researchers against which we can compare our results. Finally, we will complete robustness testing where we change the inputs significantly and check if the model still performs as expected.

Second, we will produce the results from the tests outlined in the Test Plan section. We hope to show that our model is well constructed and robust. We also hope to find that the code is easy to use and modify.

2 Test Plan

2.1 Component Testing

To begin, we will test the functions that underlie the model that are not easily observable to an end user. In particular, we will test the following model capabilities:

1. Calculating a moving average of sample means for windowing and exponential weighting
2. Calculating a moving average of sample standard deviations for windowing and exponential weighting
3. Fitting geometric Brownian motion drift and volatility parameters to these moving averages

For the purposes of these tests, we will use the stocks HD and UNH.

2.2 Accuracy Testing

The second part of testing will be making sure our results are not just reasonable, but are actually correct. We have extensive VaR and ES testing data from past researchers (Stein) against which we can compare our results. Consistant with their methodology, we will once again use a portfolio of HD and UNH stock. We will run the following tests:

1. Going long and short the portfolio and calculating all four types of 5-day VaR: parametric using windowing, parametric using exponential weighting, Monte Carlo, and historical
2. Going long and short the portfolio and calculating all four types of 5-day ES
3. Calculating the number of VaR breaches using all four types of VaR for the long and short portfolio

2.3 Robustness Testing

The final section of testing will be robustness testing. This testing will include going beyond the parameters and limitations of the tests in the previous two tests. We will make modifications to portfolio construction, window sizing, option parameters, number of simulations:

1. For many different stocks (not just HD and UNH), creating portfolios that combine long, short, call, and put positions
2. Changing the window size for the purpose of parametric VaR, which will allow for more values of VaR covering a wider range of market cycles

3. Determining the trade off between time to run simulation and expected error in results stemming from the number of Monte Carlo simulations.

3 Test Results

3.1 Component Testing

The purpose of this section is to critically look at the intermediate results needed to produce VaR estimates produced by the model.

For the purposes of this section, we will look at at the stocks *HD* and *UNH* and a portfolio of the two containing 475 and 619 shares of each, respectively. This portfolio, containing only two stocks and going long both, is a relatively simple example given the power of the model. We do this testing, though, so be consistent with previous research. We will deviate from this portfolio in the third section, Robustness Testing.

Test 1: Moving Average of Sample Mean

In order to fit a stock or a portfolio to geometric Brownian motion, we need mean daily return. Note that the way that we calculate this mean is different for windowing and for exponential weighting. For windowing,

$$m_N = \frac{1}{N}\sum a_i = \frac{S_N - S_0}{S_0}$$

where $\{a_i\}$ is the set of daily percent returns and m_N is the mean return. For windowing, we have

$$m_N = (1-\lambda)a_N + \lambda m_{N-1}$$

We get the following results:

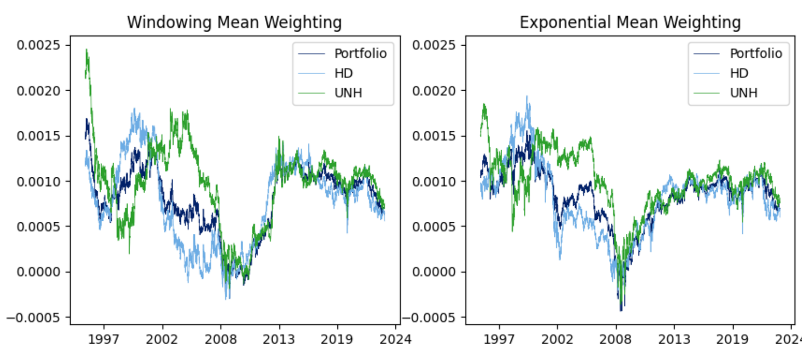

Figure 1: Sample means over time

Unfortunately, we do not have any "correct answers" against which we can compare these results. We can use heuristics, though. We know that HD and UNH have performed well of the life of the stock. For any given time, then, we would expect that the average of the sample returns be positive, which we do observe. Also, if there were to be negative returns, it is unsurprising that they would occur during the financial crisis, which again is the case. Finally, given that the portfolio is basically a moving average of the two stocks, we expect its mean returns to be between those of the two stocks. Again, this is observed. With all of this in mind, we are happy with these results.

Test 2: Moving Average of Sample Standard Deviation

Another data point we need to fit our historical data to GBM is sample standard deviation. Once again, it is calculated differently for windowing and exponential weighting. For windowing,

$$\sigma_N = \sqrt{\frac{1}{N}\sum a_i^2 - \bar{a}^2}$$

where, again, $\{a_i\}$ is the set of daily percent returns. For exponential weighting,

$$\sigma_N = \sqrt{r_N - m_N^2}$$
$$r_N = (1-\lambda)a_N^2 + \lambda r_{N-1}$$

We get the following results:

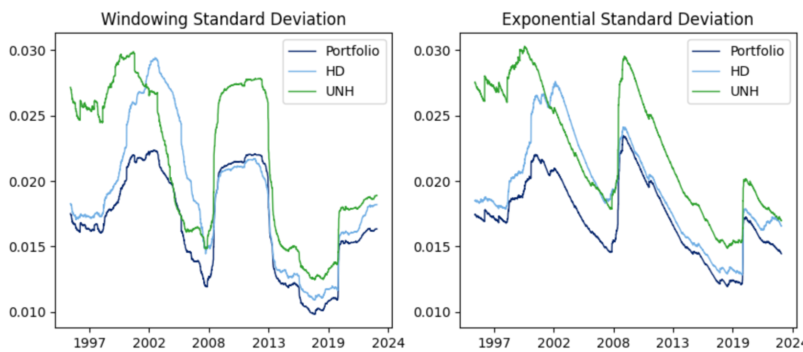

Figure 2: Sample standard deviations over time

Again, we do not have any "correct results" against which we can compare these graphs. We do know that standard deviation is positive by definition, so the fact that all of our sample values are greater than zero is encouraging. Also, we can see that the periods of higher volatility are during the dot-com bubble, the Great Recession, and Covid, which is consistent with our expectations. Also, we can clearly see the "sliding" effect that exponential weighting should have on the data. Once again, we are happy with our results.

Test 3: Geometric Brownian Motions Parameters

Now that we have samples for mean and standard deviation, we now need to estimate our GBM parameters. We make the conversions in the following way:

$$\mu_{GBM} \approx 252\bar{\mu} + \frac{\sigma^2}{2}$$
$$\sigma_{GBM} \approx \bar{\sigma}\sqrt{252}$$

As discussed in our Model Documentation, we know that fitting both a portfolio of stocks and the stocks themselves to GBM does not make sense mathematically. If the stocks follow GBM, then the portfolio cannot and vice versa. Given that we do not have a strong belief as to which one does follow GBM, it is acceptable for our purposes to do both. We start with the μ_{GBM} parameter:

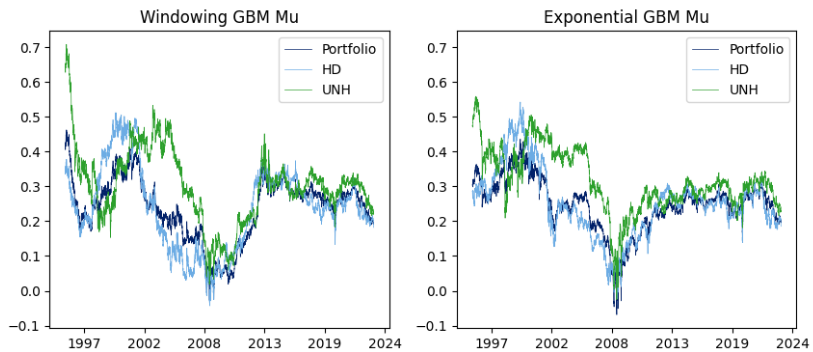

Figure 3: Geometric Brownian motion μ parameter estimate

The parameter is nearly nearly a linear transformation of the sample mean, so it is unsurprising that the shape is similar. There is a volatility term that should change the shape slightly, but this change is small enough that we cannot easily observe. The scale of the parameters is as expected. Now, for the σ_{GBM} parameter:

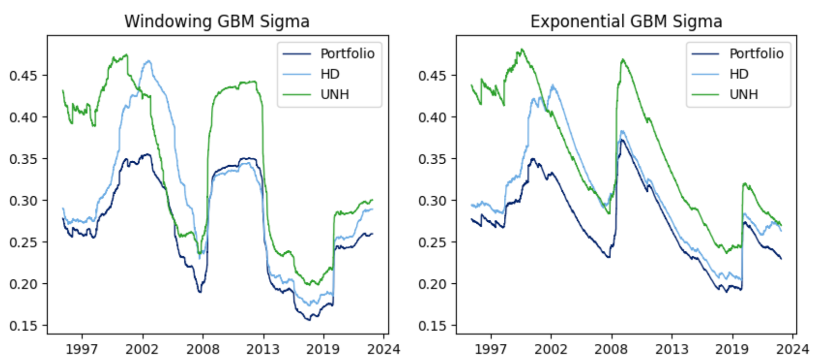

Figure 4: Geometric Brownian motion σ parameter estimate

Indeed, σ_{GBM} is a direct linear transformation of the above sample standard deviation. Likewise, the shape is exactly the same as expected. Additionally, the scale is consistent with our expectations. We are content with the results of this test.

3.2 Accuracy Testing

Now that we are happy with the construction and accuracy of the hidden functions (from the perspective of the end user), we can now analyze the VaR and ES produced. Once again, we will be using a simple portfolio of HD and UNH described above. We do this because the "correct" answers are known from previous researchers (Stein). While these tests do not cover the entirety of the programs functionality, they will give us a good idea about the accuracy of our code and our model.

Long and Short VaR

We begin by looking at the VaR of the portfolio for the four types of VaR: parametric VaR using windowing, parametric VaR using exponential weighting, Monte Carlo VaR, and historical VaR. Both the long and short portfolio are included.

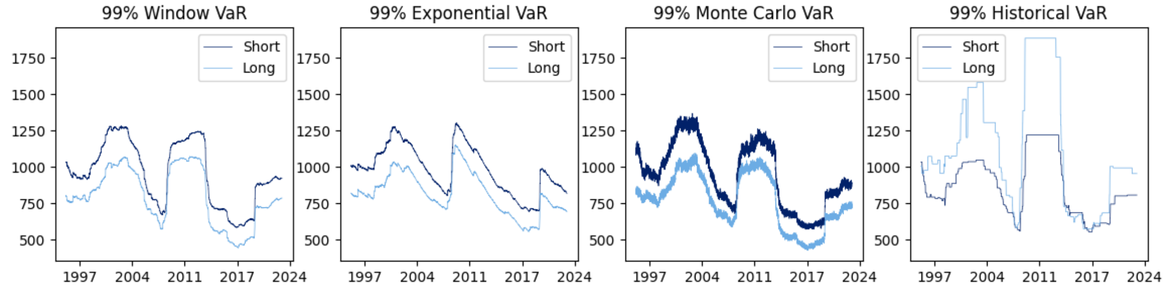

Figure 5: Estimates of VaR over time using various methods

These results are consistant with those observed in previous work by Stein. We have serveral observations:

1. The parametric (window and exponential) VaR and Monte Carlo VaR are almost identical. Given that the parametric VaRs fit GBM directly to the portfolio while the Monte Carlo fit GBM to the individual stocks, this is slightly surprising. In a way, this supports our notion of robustness.

2. For parametric VaR and Monte Carlo VaR, the short portfolio has greater potential for losses than the long portfolio. Thinking logically, this makes sense given that a short portfolio has the potential for infinite losses whereas the long portfolio can only lose the initial investment.

3. The Monte Carlo VaR tells a different story. Consistantly, it shows larger VaRs for the long positon. This is due to the fact that much of the data comes from the Great Recession. Indeed, the market suffered sharp losses in 2008 and 2009, and these are the defining VaR values for the 5 years afterwards. Naturally, the short positions performed better during this time.

For the sake of comparison, we also include the VaR statistics of a portfolio of just HD stock:

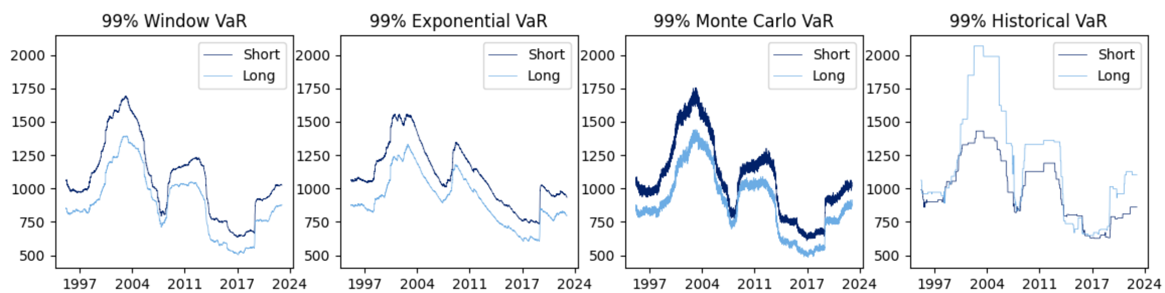

Figure 6: Estimates of VaR for just HD stock

Of note, the VaRs across the methods and across time are higher for the stock than for the portfolio. This fact demonstrates one of the most important aspects of portfolio management: diversification.

Long and Short ES

We now analyze the expected shortfall of the same portfolio. We get the following results:

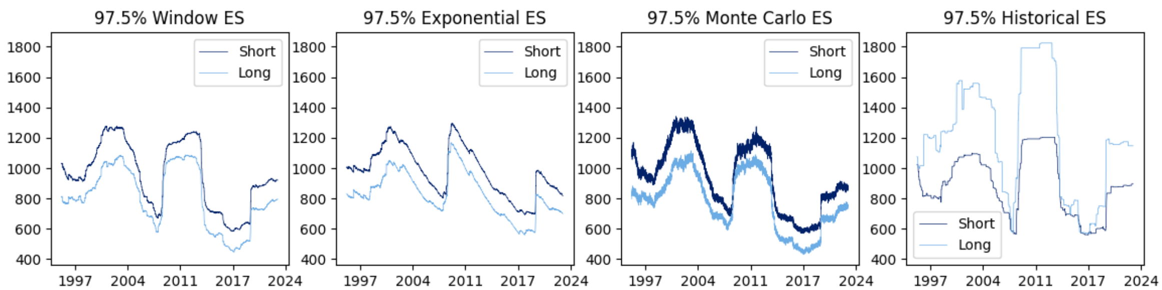

Figure 7: Estimates of ES over time of the portfolio

We have two observations:

1. The parametric and Monte Carlo ES is almost identical to the parametric and Monte Carlo VaR, which is unsurprising. Given the simplifying assumptions under parametric VaR, it can be mathematically proven that the parametric VaR results should be almost identical. Likewise, given that Monte Carlo also uses implicit normality assumptions, the results being similar are also not surprising.

2. There are differences (albeit small) between the historical VaR and the historical ES. Specifically, the historical ES is larger than the historical VaR. This is evidence of the fact that extreme tail outcomes are under-predicted by normal assumptions.

VaR Backtesting

Now that we have our VaR estimates over time, it is time to see how well they perform. We do this by looking at a moving window of number of VaR days, where a VaR day is defined as a day in which the loss to the portfolio was greater than what was predicted by the VaR. The window we will use is 1 year (252 days), so we expect there to be $252 * 1\% = 2.52$ VaR days in the past year at any point in time. We get the following results, which are consistant with the "right" answers:

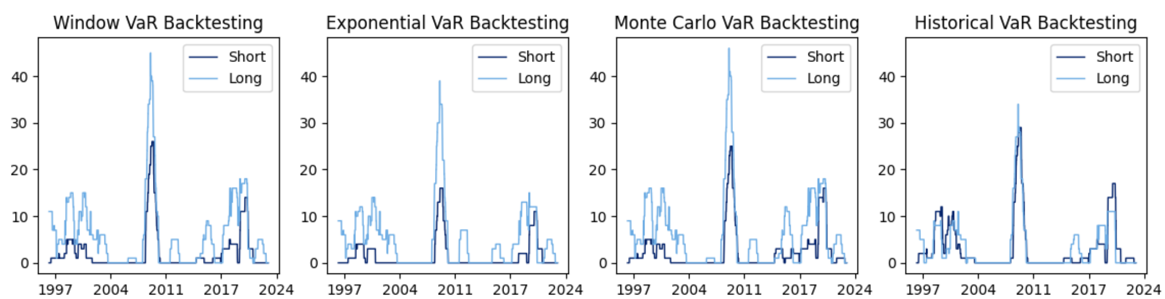

Figure 8: Number of VaR days over the past year for various VaR methods

We reach the following conclusions:

1. There is tremendous variation in the number of VaR days. During the dot com bubble, the Great Recession, and the Covid, we can see that the number of VaR days increases dramatically, meaning there were many VaR days over a short span of time. This outcome speaks to a major weakness of the model for calculating VaR: if we know past information about VaR days, the model is no longer unbiased. That is to say, if we know that there were several VaR days last week, the 99% VaR calculated using any of the four methods will not produce a loss estimate that has a 1% chance of being broken.[1]

[1] This summer, I interned at Bank of America in risk management, and my main project for the summer was trying

2. For parametric VaR and Monte Carlo VaR, there are more VaR days for the long portfolio than the short portfolio. Remember, the short portfolio for these methods had a larger VaR, indicating that it had more risk. In reality, though, the extreme losses to both portfolio constructions were similar, hence the fewer VaR days of the portfolio with the higher VaR.

3.3 Robustness Testing

We are content with both our component testing and our accuracy testing. It appears as though our underlying functions produce the expected results, and our VaR calculations are "correct," even if they do produce highly correlated VaR days. We will now spend a significant amount of time testing the limits of the model by changing many of the parameters.

Portfolio Construction

In the previous tests, we set our portfolio to be fixed and compared various methods of VaR. Here, we set our VaR method to be fixed (Monte Carlo[2]), and change our portfolio.

To begin, we no longer limit ourselves to HD and UNH as our equities. We instead now use AAPL, JNJ, and GE. The convenience of using the *yfinance* package is that we do not have to manually pull the data. Also, we no longer limit ourselves to a portfolio of only going long a portfolio of stocks or going short a portfolio of stocks. Instead, we will compare and contrast the risk profiles of three different portfolio constructions.

1. For simplicity, we will consider the portfolio of holding the three stocks and going long, holding 100 shares of AAPL, 150 shares of JNJ, and 200 shares of GE. We choose this ratio so that the proportion of our holdings is roughly equal in the 3 stocks. As before, we will normalize our VaR to portfolio holdings of $10,000

2. We will also consider the portfolio of shorting 200 shares of AAPL, going long 150 shares of JNJ, and going long 300 GE calls with the following characteristics[3]:

$$\sigma_{imp} = 40\%$$
$$T = \frac{1}{2}\ year$$
$$r = 0.5\%$$

   Note that the proportion of holdings changes as asset prices change over time. For instance, we already know that AAPL outperformed GE and JNJ over the past several decades, so we expect that AAPL's 200 shares will have a greater impact on VaR over time. This will become relevant for our call options, whose prices grow more slowly than stocks.

3. The final portfolio construction will be the most chaotic, going long 100 calls on AAPL, 100 puts on JNJ, and 100 puts on GE. We will use the following parameters for our options:

	AAPL	GE	JNJ
σ_{imp}	40%	25%	20%
T (years)	$\frac{1}{12}$	1	$\frac{1}{6}$
r	1%	1%	1%

   Note that we do have the software limitation that the time to maturity of the options must be longer than the length of VaR projection (5 days).

Side by side, we get the following Monte Carlo Var of the three portfolios:

to find a solution to this problem: how do you make VaR days independent? We found that historical VaR with anchor adjustments did the best job.

[2] We choose Monte Carlo because of its ability to easily generalize to complex portfolios which have a combination of short and long positions and contain options. Parametric VaR does not have easy formulas to use for such portfolios, and historical VaR can be challenging with options.

[3] A basic understanding of vanilla options is required, which was given in the Model Documentation

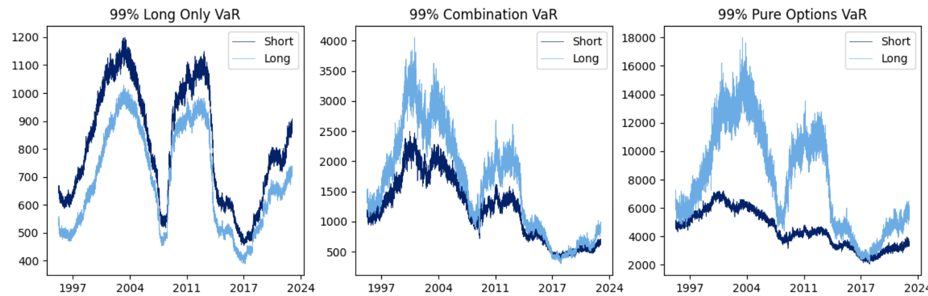

Figure 9: Monte Carlo VaR across portfolio constructions

Looking at the scales of these three plots, it is clear that VaR grows exponentially as the percent holdings in options increase. This is expected. We can also see that in the pure options graphs, we have VaRs that are greater than 10,000. This seems like it shouldn't be possible, but it indeed is: going short call options or short stock directly exposes the trader to infinite downside risk. We should note, too, that the vast majority of the risk from these options come from AAPL due to those options having a shorter time to maturity. Indeed, as time to maturity of at the money options decrease, the leverage increases.

We can also see or ES as follows:

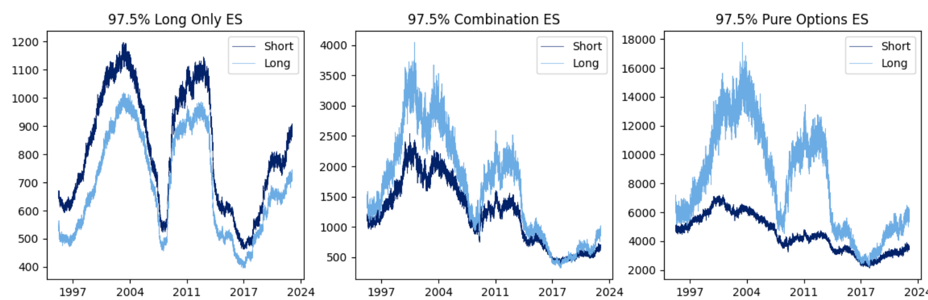

Figure 10: Monte Carlo ES across portfolio constructions

As before, the results of the ES and VaR are nearly identical. Finally, we see that our backtesting produces:

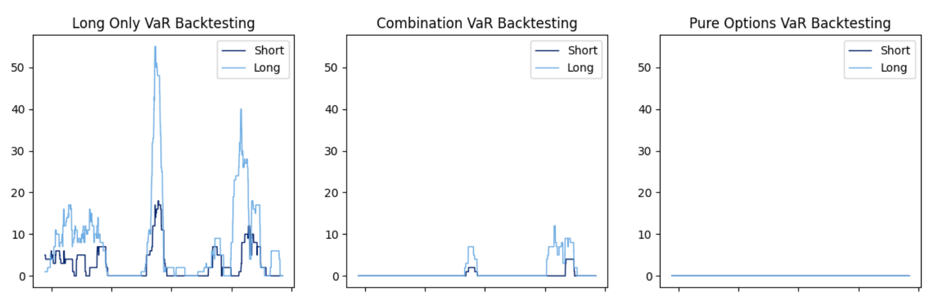

Figure 11: Number of VaR days over the past year for various portfolio constructions

We can see that the backtesting of the long only portfolio is similar to what we observed in the Accuracy Testing section. This is to be expected given that both portfolios considered were a long only. Interestingly, for the combination portfolio and the pure option portfolio, there were almost no VaR days. It appears as though one weakness of the model is that the risk estimate scales too quickly with the addition of options and short positions.

Window Sizing

Now, we will investigate the effect of window size on parametric VaR. Here, we will use a static portfolio: long 100 shares of AAPL and going long 200 shares of GE. We will compare parametric VaRs with windows of 1 year, 3 years, and 8 years. We get the following for VaR:

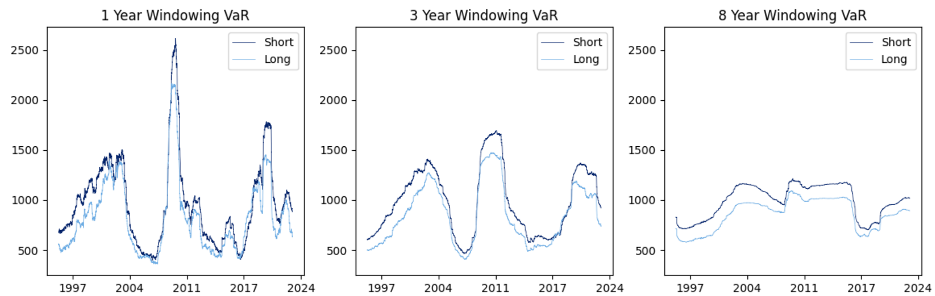

Figure 12: Parametric VaR over time using various window lengths

We can see that the VaR is far more volatile and reactive when using the 1 year windowing than the 8 year windowing. This is unsurprising given that each day has roughly 8 times the influence in the first VaR over the second. The 3 year windowing is similar to what was observed in previous sections which used 5-year windowing. Note that the 97.5% ES looks identical. Now, looking at the backtesting:

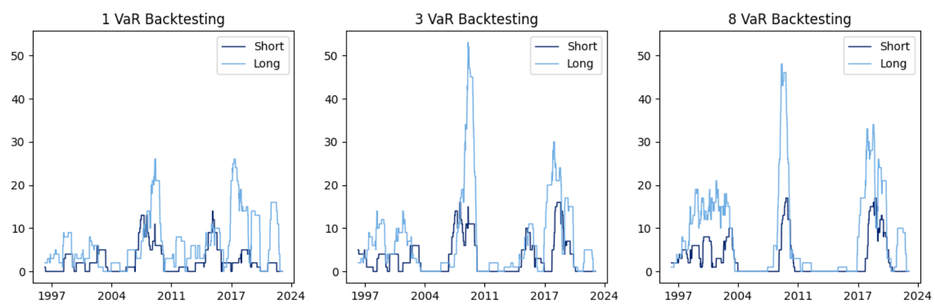

Figure 13: Number of VaR days over the past year for parametric VaR window lengths

Surprisingly, it can be argued that the 1 year windowing produces the strongest results, assuming our goal is to minimize the peak of VaR days. Thinking logically, it makes sense that the 1 year windowing would have the smallest peak because it is the quickest to change in the face of increased volatility. What is surprising, however, is the fact that the 8 year window outperformed the 3 year window. This is due to the fact that the 3 year was volatile enough to decrease significantly after the dot-com bubble but too slow at adjusting to the heightened risk during the financial crisis. This analysis demonstrates the importance of hyperparameter tuning when running these VaR models.

Monte Carlo Simulations

We conclude our robustness testing by changing the number of simulations on a fixed portfolio and investigating the time to run the simulation and the error of the VaR estimate. We will go long 200 shares of JNJ and short 150 shares of GE. We will consider a Monte Carlo simulations of length 100, 10,000, and 1,000,000. We get the following results:

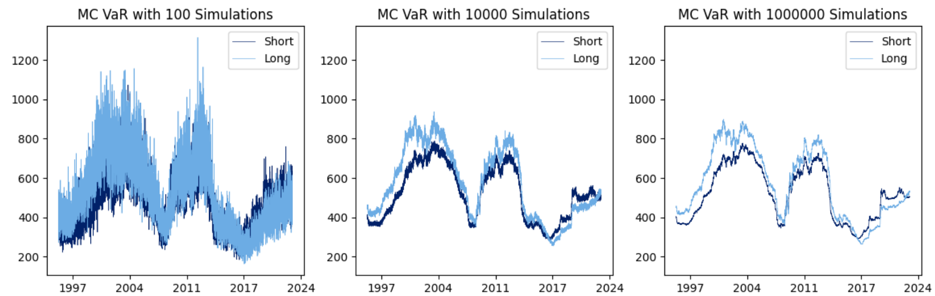

Figure 14: Monte Carlo VaR over time using various simulation lengths

The first thing we notice is the difference in the stability of the VaRs. Indeed, shorter simulation lengths result in greater variability in the VaR estimates, making it more sensitive to the specific random paths generated in a given period. On the other hand, longer simulation lengths tend to smooth out some of this variability. Given that ES results were similar, we move on to the backtesting:

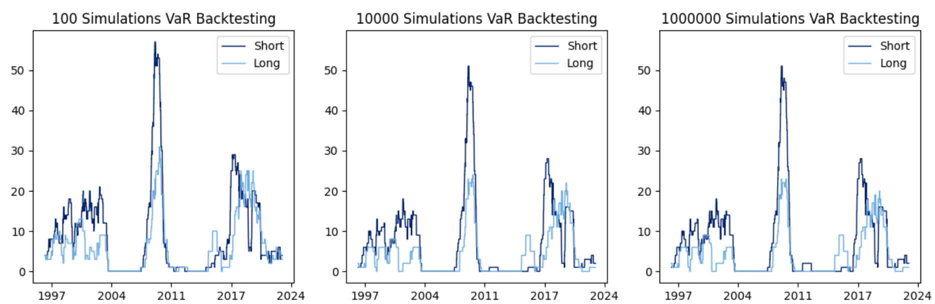

Figure 15: Number of VaR days over the past year using MC VaR

We can see that the longer simulation lengths generally provide more accurate estimates of risk, allowing for a better understanding of tail events and extreme market conditions. Shorter simulations may not capture the full range of potential outcomes simple due to random chance, potentially underestimating tail risks. We do notice that there is a decreasing marginal benefit in the number of simulations: the difference in doing 100 simulations and 10000 was much larger than the difference between 10000 and 1000000, even though the latter took multiple hours to calculate.

4 Conclusion

In this document, we tested the intermediate results of our software as well as the final results for which we had known "correct" answers. We found that our results were consistent and that software performed well. We then extended those results to more generalized portfolios containing options, parametric VaR with various window sizes, and Monte Carlo with different simulation sizes.

4.1 Weaknesses

We discovered the following weaknesses in the model:

1. VaR becomes systematically too high when options are included in the portfolio. Indeed, for portfolios containing some options and portfolios entirely constructed using options, the VaR backtesting revealed that the estimate risk was greater than the actual risk. That is to say, we have evidence that the model may be biased for such portfolios.

2. Across portfolios and VaR methods, there is a large amount of year-over-year variation in the number of VaR days. During stress periods like the Great Recession, we witnessed the number of VaR days spike, meaning there were many VaR days over a short span of time. This outcome speaks to a fundamental weakness of this model for calculating VaR: if we know past information about VaR days, the model is no longer unbiased. For example, if we know that there were several VaR days last week, the 99% VaR calculated using any of the methods will not produce a loss estimate that has a 1% chance of being broken. It stands to reason, too, that if there hasn't been a VaR day in many months, the likelihood of a 99% VaR days is less than 1%.

3. The model makes many simplifying assumptions about the nature of options contracts and about stock dynamics, namely that they follow Black-Scholes and geometric Brownian motion, respectively. To begin, geometric Brownian motion assumes constant volatility and independent, identically distributed returns, which rarely captures the dynamics of financial markets characterized by changing volatility regimes and non-normal distributions. Additionally, the Black-Scholes model assumes constant volatility, risk-free rates, and a log-normal distribution of stock prices, overlooking factors such as market jumps and skewness. In reality, financial markets exhibit time-varying volatility and non-stationary behavior, making these assumptions simplifications that may lead to mispricing and an underestimation of portfolio risk, particularly during periods of extreme market events or structural shifts. Therefore, users of this model should exercise caution and consider incorporating more sophisticated models or adjusting assumptions to enhance the accuracy of value at risk calculations.

4.2 Strengths

Now, we discuss the strengths:

1. The model has flexible portfolio construction capabilities, enabling the calculation of Value at Risk (VaR) for a diverse range of investment strategies. The model accommodates both long and short positions in stocks, as well as vanilla options, allowing for the incorporation of various asset classes and hedging strategies within a single framework. By allowing for the inclusion of a wide array of financial instruments, the model facilitates a more nuanced and realistic representation of portfolio dynamics, empowering users to assess and manage risk across a spectrum of investment scenarios.

2. Similarly, unlike some financial models that may struggle with scalability as the number of assets increases, this model remains viable regardless of portfolio size. The lack of constraints on the number of assets enhances the model's applicability to institutional and diversified portfolios, providing a valuable tool for risk management across a broad spectrum of investment scenarios and asset allocations. One specific example of this feature being useful is if a modeler was trying to hedge their pre-existing portfolio with options contracts. Instead of having to totally change the construction of their portfolio for the purposes of calculating VaR, they could simply add the position and run the software again.

3. The inclusion of VaR backtesting in the model represents a significant strength by providing a robust mechanism to assess the reliability and accuracy of the VaR estimates. The model generates a moving window of VaR breaches over the past year, allowing for a dynamic validation of the VaR time series in the context of historical market conditions. This contextualization of VaR numbers through backtesting not only enhances the transparency of risk assessments but also provides a valuable tool for refining and adjusting the model parameters, ultimately improving its predictive power.