Risk Analysis Coursework Report

Basil Ibrahim, Shaan Ali Remani, José Santos, Wincy So

March 2025

Abstract

Financial risk management requires quantifying portfolio risk and understanding the trade-off between estimation risk and model risk. Guided by an 11-year dataset of six tech stocks, we backtest six VaR models, from non-parametric bootstraps to parametric Gaussian, Student's T, and EWMA. Traditional parametric approaches, reliant on normality, underestimate tail risk, while empirical methods better capture heavy tails but exhibit greater variability. Our analysis shows that non-parametric models yield 268 violations at a 90% confidence level versus 261 from the Gaussian model, and both converge to a 17.5% probability of a 5% loss over 50 days. Additionally, portfolio construction reveals that while an Equally Weighted portfolio achieves the highest cumulative return, it suffers a larger Maximum Drawdown (– 36.35%) compared to Maximum Diversification (Sharpe ratio 1.027, Maximum Drawdown –30.71%). We emphasise the necessity of truly understanding data characteristics and the implementation of correct risk management practices for risk managers to perform accurate and reliable risk calculations.

Contents

List of Figures

List of Tables

1 Introduction

When constructing a portfolio, investors often concentrate on maximising their return. When deciding which firm to invest with, individuals look at Bloomberg's Fund Performance Rankings to see which firms rank highest. However, risk plays an equally crucial role in investment rationale. Financial institutions have prescribed risk limits set by their Risk Management teams, and individuals are as much bound by how much they are willing to lose as by how much they hope to gain. This report arises precisely from the *need to quantify, manage, and control financial risks effectively.*

We begin with *backtesting the performance of six VaR models (Question 1)*, showing that different models perform differently at various confidence levels, that parametric approaches are sensitive to their assumptions, and that empirical VaR methods are a better fit for our dataset. We then analyse the impact of *portfolio construction (Question 2)* by comparing three illustrative portfolios: equally weighted, risk parity, and maximum diversification. Despite the equally weighted portfolio achieving the best cumulative returns (4.62%), it suffers from the largest maximum drawdown (-36.35%), while both the maximum diversification and risk parity portfolios mitigate risks by reducing maximum drawdown and VaR violations. Finally, we estimate the *probability of losing more than 5% at horizons from 1 day to 50 days. (Question 5)* For very short horizons (up to 5 days), the bootstrap estimate provides a higher probability of loss, whereas for longer horizons the Gaussian approach indicates a greater risk.

2 VaR Modelling (Question 1)

In this section, we perform a VaR analysis of an equally weighted portfolio of 6 equities, Apple, Microsoft, IBM, Nvidia, Alphabet, and Amazon, using their adjusted closing prices. The sample period spans from 1 January 2014 to 31 December 2024. We use adjusted closing prices because they reflect the actual returns available to investors after accounting for dividends, stock splits, and other corporate actions.

2.1 Statistical Analysis

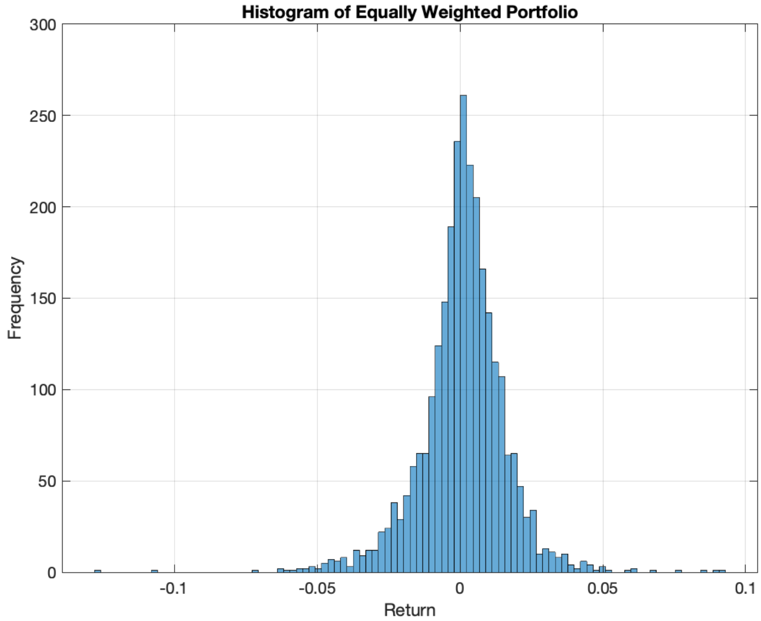


Figure 1: Histogram of daily portfolio returns.

A visual inspection of the histogram (Figure 1) reveals a heavy left tail with extreme negative values. The summary statistics (Table 1) confirm this; a normal distribution necessarily has a kurtosis of 3 and no skew, however, our data exhibit an excess kurtosis of c. 5.8 and a negative skew of -0.44. This deviation from normality foreshadows that a Gaussian VaR model is likely to underestimate the riskiness of our portfolio. In fact, Cont [2001] argues that asset and therefore portfolio returns often display heavy tails which directly cause Gaussian-models to underestimate tail-risk.

Statistic	**Mean**	**Std Dev**	**Kurtosis**	**Skewness**	**Min**	**Max**
Value	0.00098	0.01508	8.79110	-0.44361	-0.12763	0.09222

Table 1: Summary statistics of portfolio returns.

The Jarque–Bera Test for Normality (Table 2) further confirms that our return distribution is far from normal, with an overwhelmingly high Test Statistic suggesting that assuming a Gaussian distributional form may be insufficient.

Test	**Test Statistic**	**Critical Value**	**p-value**	**Hypothesis Result**
JB	3955.8	5.9717	<0.001	Reject Normality

Table 2: Jarque-Bera test results for normality.

Furthermore, an inspection of the rolling mean and rolling volatility (Figure 2) suggests persistence in returns and that periods of volatility follow previous periods of volatility (referred to as volatility clustering). The ACF autocorrelation chart (Figure 3) and the Engle ARCH test (Table 3) confirm significant serial correlation, implying that conditional estimates might be preferred in VaR models.

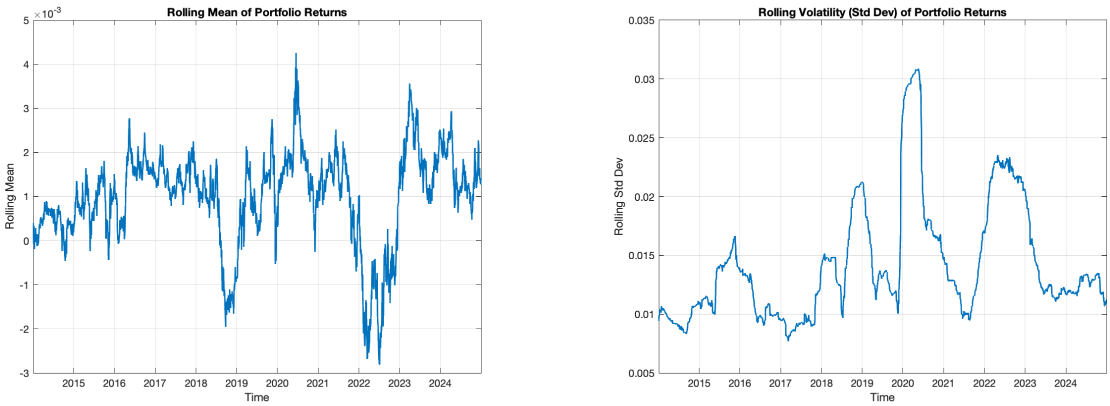


Figure 2: Rolling mean (left) and rolling volatility (right) of portfolio returns suggest conditionality.

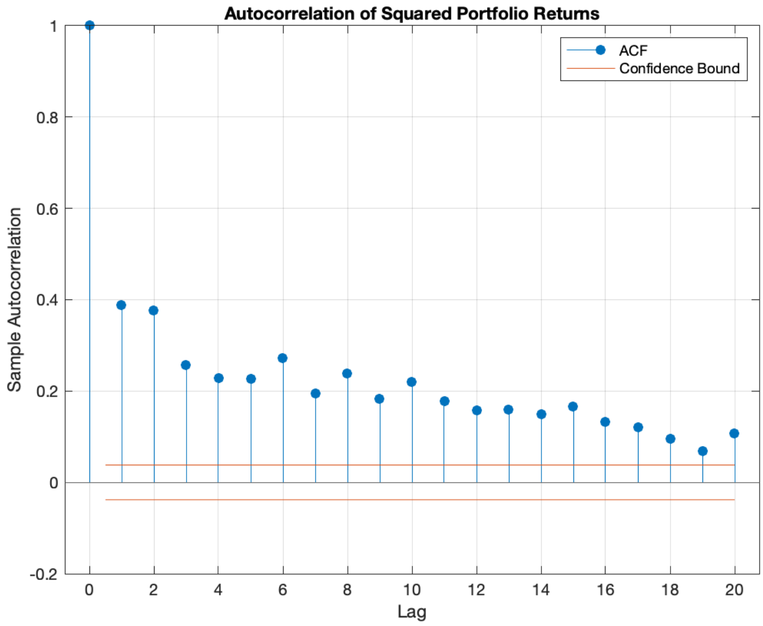


Figure 3: ACF of squared returns.

Test	Test Statistic	Critical Value	p-value	Hypothesis Result
ARCH	417.61	3.8415	0	Reject Homoscedasticity

Table 3: ARCH test results for residual heteroskedasticity.

The cumulative return profile for our portfolio is strong (Figure 4), begging the question of how much risk was taken to achieve this return.

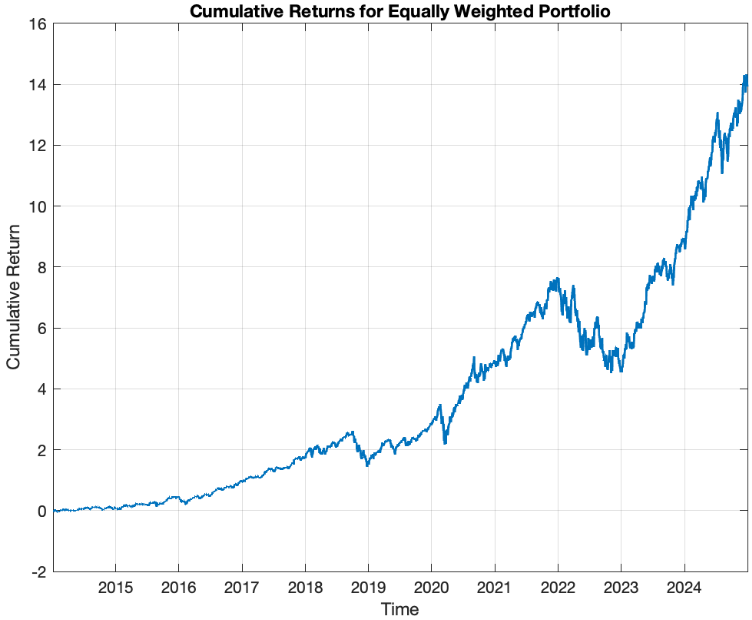


Figure 4: Cumulative returns of the equally weighted portfolio.

2.2 VaR Models

A rolling window of 6 months (starting from 1 July 2014) is used to estimate one-day VaR at 90% and 99% confidence levels. Six different VaR models are considered to balance estimation risk, suffered by empirical models, and model risk, suffered by parametric models:

1. **Bootstrap VaR (Non-Parametric):** Resamples historical returns with replacement, repeating 1000 times to reduce estimation risk. Note: while bootstrapping does not prescribe any distributional form to the data, it does assume that future return distributions mirror the past.

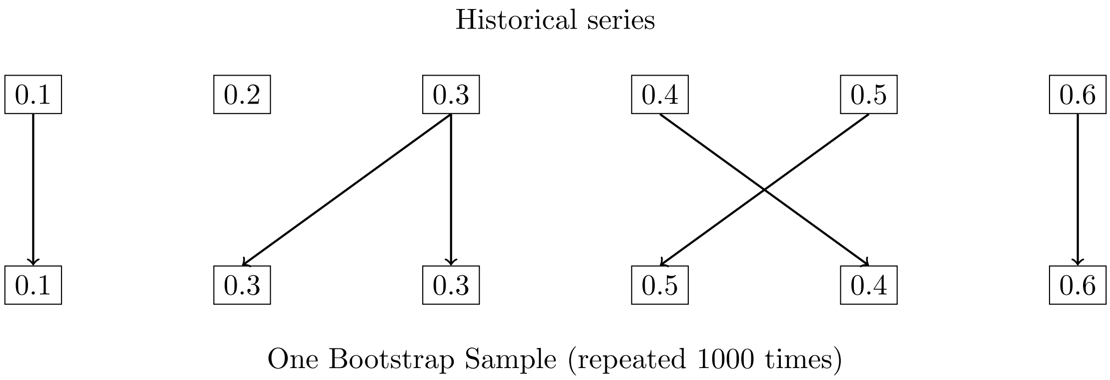


Figure 5: Standard Bootstrap: The historical series is resampled with replacement, here illustrating that one value (0.3) is selected twice while another (0.2) is omitted.

2. **Block Bootstrap VaR (Non-Parametric):** As above, but the historical return sample is split into blocks. Each block is resampled independently, preserving conditionality and volatility clustering present in the data.

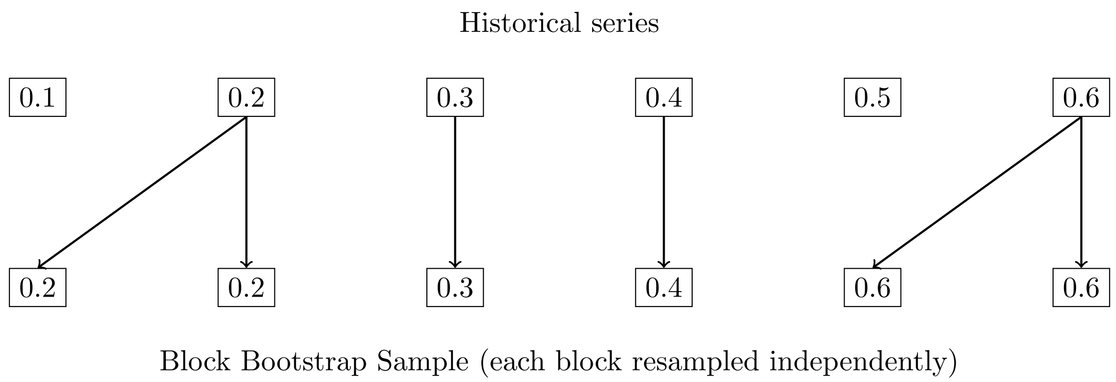


Figure 6: Block Bootstrap: The historical series is divided into blocks and each block is resampled independently. In this example, the bootstrap for Block 1 yields {0.2, 0.2, 0.3} and for Block 2 yields {0.4, 0.6, 0.6}.

3. **Hybrid Block-Filtered Bootstrap VaR (Semi-Parametric):** As above, but applies a GARCH(1,1) filter first. A GARCH(1,1) model estimates conditional mean and conditional variance for each rolling window, then computes standardised residuals. These residuals are 'block bootstrapped' to maintain short-run dynamics, and finally rescaled using the GARCH estimate for conditional volatility.

4. **Gaussian VaR (Parametric):** Assumes that returns follow a normal distribution: $R_t \sim \mathcal{N}(\mu, \sigma^2)$, where μ denotes expected return and σ^2 denotes variance of returns, and calculates VaR using the normal cumulative distribution function.[1]

5. **Student's T VaR via Method of Moments (Parametric):** Assumes that returns follow a Student's T distribution, calibrated with sample moments, to capture heavy tails. The degrees of freedom are estimated as

$$\nu = 4 + \frac{6}{\text{Kurtosis} - 3},$$

and a scaling parameter is adjusted proportionately. Student's T distributions seek to better model heavy tailed distributions which Gaussian VaR fails to capture.

6. **EWMA VaR (RiskMetrics, Parametric):** Captures conditional volatility through exponentially weighted moving averages. Setting a decay factor of $\lambda = 0.94$, we place greater emphasis on recent returns to amplify current market movements and soften long-term behaviour.

Justification of VaR Estimation Models: The six models above span both parametric and non-parametric approaches, with the aim of identifying which model balances the trade-off between estimation risk and model risk most effectively for our dataset. It is worth noting here that there is no singular 'best' VaR model for all portfolios, and modelling VaR depends heavily on the observed time series of our portfolio returns. The standard bootstrap method relies solely on empirical data and does not impose a specific distribution, capturing non-normal features (e.g. heavy tails, excess kurtosis, skew etc.). However, by resampling each individual observation, it can lose the temporal dependence inherent in financial returns. The block bootstrap, introduced by Politis and Romano [1994], addresses this by resampling contiguous

[1] In our case, expected return is modelled simply using the historical average. Other models include the famous Capital Asset Pricing Model, Fama-French Three Factor Model and Dividend Discount Model, see Cuthbertson and Nitzsche [2008]. Alternatives in the last 10 years include Machine Learning, Deep Learning, and Sentiment Analysis models. See de Prado [2018] for an introduction to Financial Machine Learning.

blocks of data, to preserve volatility clustering. However, choosing the optimal block length remains an active area of research. There is no clear rule governing conditionality in either the market or individual stocks; if such a rule did exist, profiting would be straightforward.[2] The hybrid block-filtered bootstrap serves an interesting 'middle case', where the problem of arbitrariness is replaced by a conditional variance estimate first, a process refereed to as filtering, before block bootstrapping is applied. For this reason, we refer to this approach as Semi-Parametric.

In contrast, the parametric methods (Gaussian, Student's T, and EWMA) impose distributional assumptions; while the Gaussian model is simple, it fails to capture fat tails, and the Student's T model, although better at accounting for extreme events, still imposes a fixed form. The EWMA model, which weights recent returns more heavily, is responsive to changing market conditions but depends on the assumption that past volatility is indicative of the future. This assumption can be devastating, as seen during the 2008 financial crisis, and 2020 COVID-19 pandemic: models based on normality significantly underestimated tail risk, leading to widespread underestimation of potential losses.

2.3 VaR Results and Analysis

Figures 7 and 8 display the daily rolling VaR estimates at 90% and 99% confidence levels for both empirical and parametric approaches. Notably, the bootstrapped methods produces higher VaR estimates than parametric models, aligning with our statistical analysis that the heavy left tail of the empirical distribution is not well captured by normality. Further, while Gaussian and Student's T models exhibit similar patterns, the Hybrid Block-Filtered and EWMA models capture more extreme volatility, causing larger swings over time. Interestingly, the model with the most violations at 90% (Hybrid Block-Filtered) differs from the one with the most violations at 99% (Gaussian), suggesting that violation counts alone are not sufficient to judge model effectiveness (Figure 9). Instead, we require formal VaR backtests.

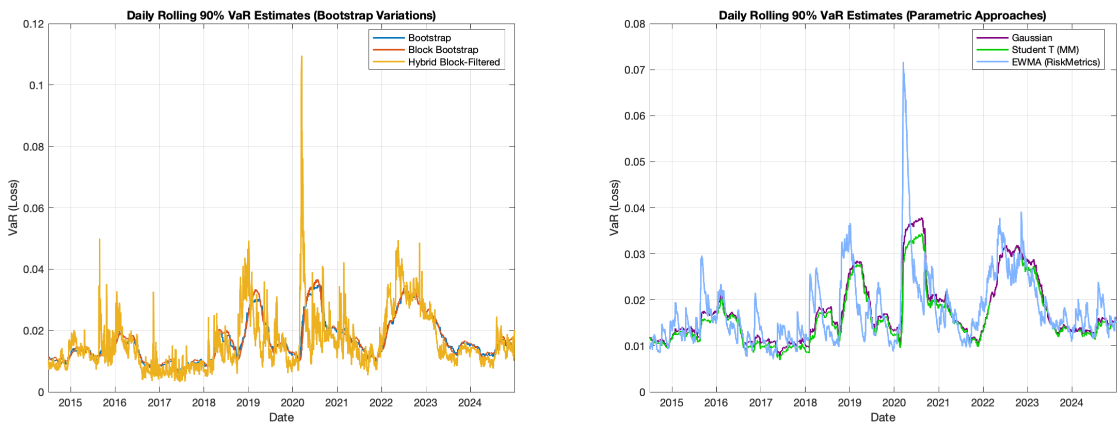


Figure 7: Daily Rolling 90% VaR estimates using Bootstrap (left) versus Parametric (right).

[2]For an introduction to block length selection and its implications for sensitivity analysis, see Lahiri [2003].

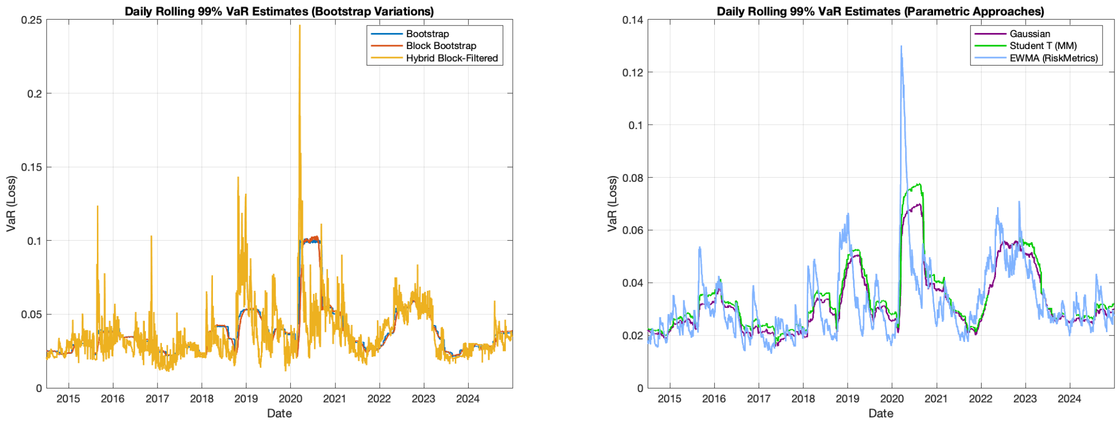
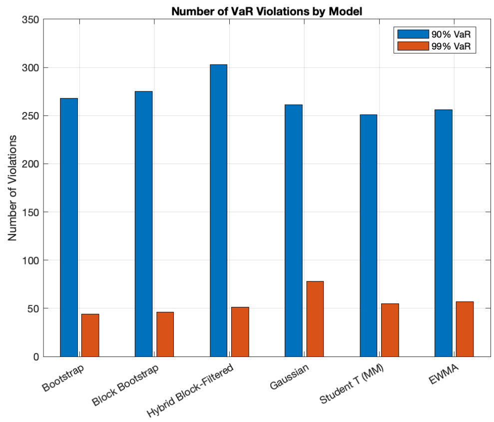


Figure 8: Daily Rolling 99% VaR estimates using Bootstrap (left) versus Parametric (right).



Method	**Violations**	
	90%	**99%**
Bootstrap	268	44
Block Bootstrap	275	46
Hybrid Block-Filtered	303	51
Gaussian	261	78
Student T (MM)	251	55
EWMA	256	57

Figure 9: Number of VaR violations at 90% and 99% confidence levels across different models.

2.4 Backtesting Methodology and Empirical Discussion

Before presenting the backtest results, we introduce the key tests used to assess our VaR models. These are taken from Chapter 13 of Christoffersen [2003]'s *Elements of Financial Risk Management*.

Let I_t denote the hit sequence,

$$I_t = \begin{cases} 1, & \text{if the loss at time } t \text{ exceeds the VaR forecast,} \\ 0, & \text{otherwise.} \end{cases}$$

Under a correctly specified model, I_t should follow an i.i.d. Bernoulli(p) process, *where p is the nominal coverage probability.*

2.4.1 Unconditional Coverage Test

The unconditional coverage test, also called the Kupiec test, checks if the observed violation frequency,

$$\hat{\pi} = \frac{1}{T}\sum_{t=1}^{T} I_t,$$

matches the promised rate p. The likelihood ratio test statistic is defined as

$$LR_{uc} = -2\ln\left(\frac{L(p)}{L(\hat{\pi})}\right),$$

which, under $H_0 : \hat{\pi} = p$, follows a χ^2 distribution with 1 degree of freedom. A low p-value indicates that the observed frequency of violations is significantly different from the expected rate, suggesting that the model may be misspecifying risk.

2.4.2 Independence Test

To determine if violations occur randomly over time, the hit sequence is modelled as a first-order Markov process. Let π_{01} and π_{11} denote the probabilities of a violation following a non-violation and a violation, respectively. Under the independence assumption, both equal p. A likelihood ratio test compares this to an alternative model where π_{01} and π_{11} can differ. Low p-values here suggest that violations are clustered, indicating that the model does not capture the temporal dynamics of risk adequately. We do not quote independence tests directly, but use these to form the *Conditional Convergence Test.*

2.4.3 Conditional Coverage Test

The conditional coverage test combines the unconditional coverage and independence tests,

$$LR_{cc} = LR_{uc} + LR_{ind},$$

which is asymptotically χ^2 distributed with 2 degrees of freedom. Significant values of LR_{cc} imply that the model either misestimates the average violation rate or fails to capture the independence of violations, or both.

2.4.4 Backtest Results

Model	LR_{uc}		**Kupiec p-value**		LR_{cc}		**Cond. Coverage p-value**	
	90%	99%	90%	99%	90%	99%	90%	99%
Bootstrap	0.07104	9.8550	0.78983	**0.0016937**	4.0984	14.0200	0.12884	**0.00090289**
Block Bootstrap	0.91689	14.3490	0.33829	**0.00015188**	10.0360	28.4020	**0.0066178**	**6.8021e-07**
Hybrid Block-Filtered	5.4862	22.3476	**0.019167**	**2.275e-06**	5.9799	22.9630	0.05029	**1.0319e-05**
Gaussian	0.053788	66.7581	0.81660	**3.3307e-16**	3.7762	76.1772	0.15136	**0**
Student T (MM)	0.75448	23.8173	0.38506	**1.0593e-06**	5.5290	31.4163	0.063009	**1.5068e-07**
EWMA	0.29599	26.8675	0.58641	**2.1789e-07**	0.78126	31.0436	0.67663	**1.8154e-07**

Table 4: Backtesting results for various models at 90% and 99% confidence levels.

Our backtest results in Table 4 indicate that almost all models perform well at the 90% confidence level, but all models are remarkably inadequate at the 99% level. Isolating one example, the Bootstrap model exhibits a very low LR_{uc} at 90%, p = 0.79, suggesting correct average coverage, but, at 99%, the same model yields $LR_{uc} = 9.8550$ with p = 0.0017, strongly rejecting the null hypothesis of correct average coverage; suggesting this model fails to capture extreme losses. The same analysis applies to Block Bootstrap, Gaussian, Student T, and EWMA. Notably, this confirms that the normality assumption does, indeed, struggle in the left-tail. In contrast, the Hybrid Block-Filtered model underestimates risk even at moderate risk levels this model may underestimate risk slightly.

The conditional coverage test for the Gaussian and EWMA models at 90%, p = 0.1514 and 0.6766, respectively, indicates that while the overall number of violations might be acceptable, there is some serial dependence in the hit sequence that these models fail to capture.

While the hit sequence and its related tests may provide the the correct number of hits on average, they offer limited insight into the magnitude of losses when violations occur. This is magnified at higher confidence levels, i.e., 99%, where the number of VaR violations is lower. That is, a model might register the correct number of hits on average, but underestimate the severity of extreme losses. As a result, we compute a tail-specific test, the Kolmogorov–Smirnov test, in the final section. These observations reinforce Artzner et al. [1999] call for considering coherent risk measures, such as Expected Shortfall, to better capture extreme tail risk.[3]

2.5 Distributional Tests

Since VaR models are primarily concerned with extreme losses, we further examine the distributional assumptions using the Probability Integral Transform (PIT). For each day, we compute

$$u_t = F(r_t),$$

where F is the forecasted cumulative distribution function and r_t is the observed return. Under a correctly specified model, the sequence u_t should be i.i.d. Uniform(0,1).

To perform formal testing, the uniform variables are transformed to standard normals:

$$z_t = \Phi^{-1}(u_t),$$

and the Kolmogorov–Smirnov (KS) test is applied. Our results reveal that while the full PIT distribution for bootstrapped models is acceptable, all models fail to capture the left tail successfully (Table 5. Further refinements may be necessary to better forecast extreme losses.

Model	**Full**		**Left-Tail**	
	KS (h)	**p-value**	**KS (h)**	**p-value**
Bootstrap	0	0.6293	1	**1.27e-190**
Block Bootstrap	0	0.6293	1	**1.27e-190**
Hybrid Block-Filtered	0	0.6293	1	**1.27e-190**
Gaussian	1	**1.63e-07**	1	**1.64e-186**
Student T (MM)	1	**0.0002**	1	**1.61e-179**
EWMA	1	**4.03e-05**	1	**7.51e-198**

Table 5: Kolmogorov–Smirnov (KS) test results for the Probability Integral Transform (PIT). The full test assesses the entire distribution, while the left-tail test focuses on the lower tail. Statistically significant p-values are in bold.

Discussion of Limitations and Further Improvements: The parametric models, particularly Gaussian VaR, struggle to capture the heavy-tailed behaviour observed in the data as predicted. Although the Student's T VaR improves slightly by incorporating tail risk, it still relies on a fixed functional form. Non-parametric bootstrapping methods overcome these issues, but assume that historical patterns persist into the future. Future work may explore alternative volatility structures, such as EGARCH or GJR-GARCH, to improve the PIT fit in the left-tail. Alternatively, tail models based on Extreme Value Theory, e.g., Generalised Pareto Distribution [Embrechts et al., 1997]. Further, sensitivity analyses on the optimal bootstrap block length for our dataset, as explored by Politis and Romano [1994] and Lahiri [2003], may identify a more accurate VaR model that preserves our left-tail.

Finally, note that alternative risk measures may provide better risk estimates. Ballotta and Fusai [2017, p. 9–10] note that 'it is possible to present examples for which... the VaR measure is violating the subadditivity property.' That is, the combined risk of a portfolio can be greater than the sum of the individual assets, contradictory to diversification. In contrast, Expected Shortfall provides a more comprehensive picture of tail risk and serves as a *coherent risk measure.*

[3]This idea is further explored in *Discussion of Limitations and Further Improvements*

3 The Risk Parity Portfolio (Question 2)

We have constructed 4 portfolios based on different portfolio strategies. These strategies include: the Risk Parity portfolios (both parametric and non-parametric approaches), the Maximum Diversification portfolio, and the Equally-Weighted portfolio. Using unique optimisation objectives and risk measures, each strategy represents a different approach to constructing a portfolio that balances risk and return.

The **Risk Parity** strategy aims to allocate capital such that each asset contributes equally to the portfolio's overall risk. This is measured by the Component Value at Risk (CVaR). The strategy can be implemented using different two approaches: (1) the **parametric approach**, which relies on the assumption of normal distribution and uses the sample covariance matrix to calculate risk contributions; and (2) the **non-parametric approach**, which is based on empirical distributions and historial data to determine asset risk contributions. In contrast, the **Maximum Diversification** strategy seeks to maximize the diversification ratio:

$$\max_{w} \frac{w'\sigma}{\sqrt{w' \sum w}},$$

with the objective of improving portfolio efficiency by spreading risk more effectively across assets. Finally, the **Equally-Weighted** portfolio simply assigns equal capital weights to all assets, offering a straightforward baseline for comparison.

The table below presents the portfolio weights for the six assets using each of these strategies. These weights were determined using the in-sample data and form the basis for subsequent evaluation.

	Risk parity (Parametric)	**Risk parity (Non-parametric)**	**Maximum diversification**	**Equally-weighted**
AAPL	0.17565	0.17191	0.19818	0.16667
MSFT	0.16822	0.19731	0.03716	0.16667
IBM	0.23576	0.19743	0.35240	0.16667
NVDA	0.11374	0.11036	0.15413	0.16667
GOOGL	0.16865	0.20613	0.09455	0.16667
AMZN	0.13798	0.11686	0.16358	0.16667

Table 6: Portfolio Weights by Method

The Risk Parity portfolios, both parametric and non-parametric, allocated weights relatively evenly across assets, as their primary objective is to equalise each asset's contribution to the overall portfolio risk rather than to maximise returns or diversification. However, there is a noticeable difference in the allocation to GOOGL and IBM between the two approaches, where the non-parametric Risk Parity portfolio assigns the highest weight to GOOGL (20.61%), whereas GOOGL holds only the third-largest weight in the parametric Risk Parity portfolio. This is because the parametric approach assigns higher weights to assets with lower volatility and lower correlation under the assumption of normality, while the non-parametric Risk Parity portfolio, which relies on empirical data, adjusts allocations based on observed historical distributions, allowing it to better capture tail risks.

The Maximum Diversification portfolio demonstrates a more concentrated allocation, with a substantial weight assigned to IBM (35.24%) and reduced exposure to MSFT (3.72%) and GOOGL (9.46%). This allocation pattern reflects the underlying objective of maximising the diversification ratio, which favours assets that provide the best trade-off between their individual volatility and correlation with the rest of the portfolio. Having the largest concentration on IBM suggests that IBM contributes significantly to improving the diversification ratio by providing relatively high individual volatility while maintaining low correlations with other assets. Next, we will examine the risk contributions of each strategy captured by the Component VaR analysis.

The parametric Risk Parity portfolio achieves its objective by equalising the risk contributions of each asset to the total portfolio risk. As shown in Table 7, each asset contributes precisely 16.67% to the total Component Value at Risk (CVaR). Despite the differences in asset volatilities and correlations, the optimisation process adjusts weights to ensure that the marginal contribution of each asset aligns perfectly with the target risk parity condition. For example, even though NVDA exhibits a high marginal VaR (MVaR) of 0.028520 compared to IBM's 0.013759, the weight adjustment ensures their CVaR contributions are equalized.

Asset	**Weight**	**MVaR**	**CVaR**	**CVaR (%)**
AAPL	0.17565	0.018467	0.0032438	16.667
MSFT	0.16822	0.019283	0.0032438	16.667
IBM	0.23576	0.013759	0.0032438	16.667
NVDA	0.11374	0.028520	0.0032438	16.667
GOOGL	0.16865	0.019233	0.0032438	16.667
AMZN	0.13798	0.023510	0.0032438	16.667

Table 7: Component VaR for Risk Parity Portfolio (Parametric)

In the non-parametric Risk Parity portfolio, risk contributions are approximately equal but display slight variations due to the use of empirical data rather than parametric assumptions. As shown in Table 8, the CVaR contributions range from 15.66% for AAPL to 17.31% for MSFT. The slight deviation from perfect equality are the result from relying on historical return distributions, which in theory better capture asset-specific tail risks and skewness. IBM has a noticeably lower CVaR percentage contribution, reflecting its relatively lower historical tail risk as captured by the empirical method. Although the CVaR for each asset is not perfectly balanced, the non-parametric approach provides a more data-driven allocation of risk contributions.

Asset	**Weight**	**CVaR**	**CVaR (%)**
AAPL	0.17191	0.0030412	15.656
MSFT	0.19731	0.0033618	17.307
IBM	0.19743	0.0030465	15.684
NVDA	0.11036	0.0033050	17.014
GOOGL	0.20613	0.0032418	16.689
AMZN	0.11686	0.0032627	16.796

Table 8: Component VaR for Risk Parity Portfolio (Non-Parametric)

The Maximum Diversification portfolio, shown in Table 9, presents a considerably different distribution of risk contributions compared to the Risk Parity strategies. CVaR contributions are highly concentrated, with IBM accounting for 26.98% of the total portfolio risk and NVDA for 23.73%, while MSFT contributes only 3.26%. This uneven distribution results from the optimisation objective of maximising the diversification ratio, which focuses on improving the trade-off between weighted asset volatilities and overall portfolio volatility rather than balancing risk contributions. As a result, assets like IBM, which offer a high diversification benefit due to their correlation structure and individual volatility, receive a much larger allocation and subsequently dominate the portfolio's risk contribution. Conversely, MSFT's low allocation and minimal contribution to CVaR reflect its limited impact on improving diversification under this methodology.

Asset	Weight	MVaR	CVaR	CVaR (%)
AAPL	0.19818	0.018312	0.0036290	18.533
MSFT	0.037162	0.017191	0.00063885	3.263
IBM	0.35240	0.014993	0.00528350	26.982
NVDA	0.15413	0.030148	0.00464670	23.730
GOOGL	0.094551	0.017488	0.00165350	8.444
AMZN	0.16358	0.022802	0.00373000	19.049

Table 9: Component VaR for Maximum Diversification Portfolio (Parametric)

The Equally-Weighted portfolio assigns an identical weight of 16.67% to each asset, regardless of their individual risk profiles, volatilities, or correlations. While this approach offers simplicity, it leads to an uneven distribution of risk contributions, as illustrated by the CVaR figures in Table 10. Under this portfolio, NVDA accounts for 25.09% of the portfolio's total risk, whereas IBM contributes the least at 10.07%. This disparity highlights a key limitation of the equally-weighted strategy: it disregards asset-specific risk characteristics, resulting in a concentration of risk in more volatile assets. Consequently, the equally-weighted portfolio may expose investors to higher tail risk from more volatile assets.

Asset	Weight	MVaR	CVaR	CVaR (%)
AAPL	0.16667	0.01808	0.0030134	14.638
MSFT	0.16667	0.01896	0.0031599	15.350
IBM	0.16667	0.012433	0.0020722	10.066
NVDA	0.16667	0.030983	0.0051638	25.085
GOOGL	0.16667	0.019051	0.0031752	15.425
AMZN	0.16667	0.024006	0.0040010	19.436

Table 10: Component VaR for Equally-Weighted Portfolio (Parametric)

The Component VaR analysis highlights the trade-offs inherent in each portfolio construction approach, particularly in terms of how risk is distributed among assets. We then evaluate the portfolio performance with the second half of the samples. The key performance metrics, including cumulative return, Sharpe ratio, maximum drawdown, and VaR violation, are calculated to provide a comprehensive comparison of the relative strengths and weaknesses of different strategies.

	Sharpe Ratio	Maximum Drawdown	VaR Violations	Cumulative Return
Risk parity (Parametric)	0.98511	-0.32250	59	4.0145
Risk parity (Non-parametric)	0.98251	-0.33578	69	4.0814
Maximum diversification	1.02690	-0.30713	60	4.1839
Equally-weighted	1.01680	-0.36351	62	4.6155

Table 11: Performance Metrics Comparison

The performance metrics of each portfolio strategy are summarized in Table 11. The Equally Weighted portfolio achieved the highest cumulative return (4.6155) and a Sharpe Ratio of 1.0168 over the out-of-sample period. However, it also has the largest maximum drawdown (-36.35%), indicating greater exposure to downside risk. Both Risk Parity portfolios demonstrated lower drawdowns and fewer VaR violations compared to the Equally Weighted portfolio. Among the two approaches, the parametric approach delivered slightly better risk control with a maximum drawdown of -32.25% and 59 VaR violations.

The Maximum Diversification portfolio recorded the lowest maximum drawdown (-30.71%), suggesting excellent downside protection achieved through maximising diversification. In addition, it outperformed both Risk Parity strategies in terms of risk-adjusted return, achieving the highest Sharpe Ratio (1.0269) among all strategies.

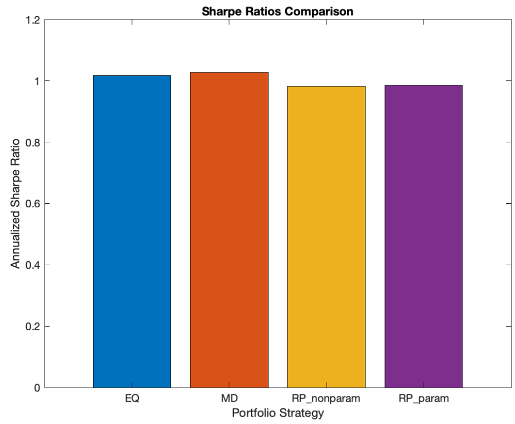


Figure 10: Comparison of Sharpe Ratios Across Portfolio Strategies

The Risk Parity and Maximum Diversification strategies are theoretically designed to deliver higher risk-adjusted returns. Specifically, Risk Parity aims to balance risk contributions across assets, leading to a more stable and efficient risk-return profile, while Maximum Diversification seeks to maximise the diversification ratio, enhancing return potential by spreading risk more effectively. In this analysis, the Maximum Diversification portfolio delivered the highest **Sharpe Ratio** (1.0269), closely followed by the Equally Weighted portfolio (1.0168). Both Risk Parity portfolios underperformed in terms of Sharpe Ratio relative to the Maximum Diversification and Equally Weighted portfolios, with Sharpe Ratios of 0.98511 (parametric) and 0.98251 (non-parametric).

The parametric Risk Parity portfolio shows an even spread of weights across assets, which closely resembles the allocation of the Equally Weighted strategy. This is largely due to the limited variation in the Marginal VaR (MVaR) across the six assets. Since these assets exhibit relatively similar risk characteristics (volatility and correlations), the optimization process in Risk Parity does not produce significantly different weights from the simple equal-weight approach. As a result, the risk allocation benefits that typically distinguish Risk Parity are less pronounced in this specific case, and its performance ends up being similar-but not superior-to the Equally Weighted portfolio.

Another considerable factor limiting the effectiveness of the Risk Parity is their reliance on the sample covariance matrix to estimate asset volatilities and correlations. While the sample covariance matrix is an unbiased estimator, it is prone to high estimation error, specifically when there is limited historical data or during periods of structural market shifts. This can lead to unstable, ill-conditioned, or even positive semi-definite matrices, resulting in noisy portfolio weight allocations, overfitting, and poor out-of-sample performance. Moreover, small changes in input data can cause large fluctuations in optimal portfolio weights, undermining the stability and robustness of risk-based allocation strategies. There are several techniques that can be employed to mitigate these issues, including Shrinkage Estimators, Factor Models and Exponentially Weighted Moving Average (EWMA). These methods can enhance the stability of covariance matrix estimates and reduce sensitivity to estimation errors.

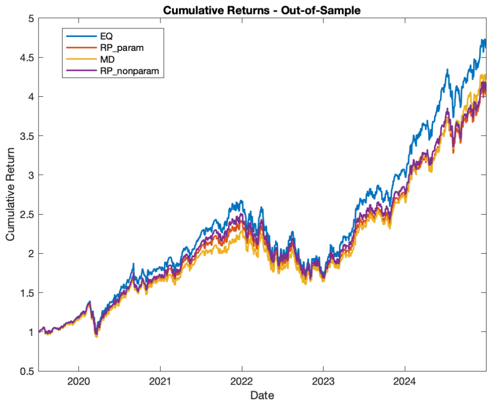


Figure 11: Comparison of Cumulative Returns Across Portfolio Strategies

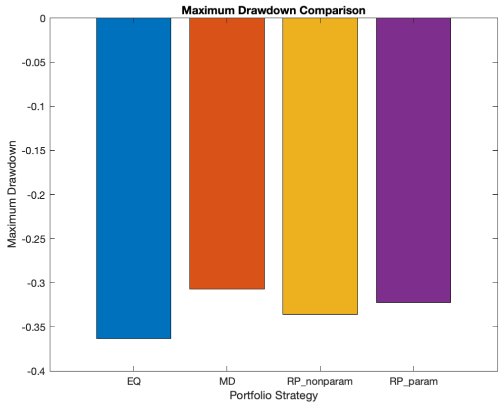


Figure 12: Comparison of Maximum Drawdowns Across Portfolio Strategies

Maximum drawdown measures the downside risk, capturing the largest peak-to-trough decline in portfolio value during the investment period. It provides insight into the potential worst-case scenario for capital loss. In this analysis, the Maximum Diversification portfolio has the lowest maximum drawdown (-30.71%), highlighting its effectiveness in downside protection. By maximising the diversification ratio, this strategy spreads risk across assets with low correlations, helping to cushion the impact of adverse market movements. While Maximum Diversification offered the best risk-adjusted returns and downside protection, it is noteworthy to take into account that the concentration in IBM highlights the trade-off between maximising diversification and exposure to individual asset risks.

The Risk Parity portfolios exhibited moderate drawdowns, with the parametric version performing slightly better (-32.25%) than the non-parametric approach (-33.58%). These outcomes reflect the strategy's focus on equalizing risk contributions, providing reasonable resilience to extreme losses, but may not offer as much downside protection as maximising diversification.

The Equally Weighted portfolio recorded the largest maximum drawdown (-36.35%), highlighting its greater vulnerability to market downturns. This result is consistent with its uneven risk distribution in the Component VaR analysis, where more volatile assets (e.g., NVDA and AMZN) contributed disproportionately to total portfolio risk. Despite this, the Equally Weighted portfolio achieved the highest cumulative return and Sharpe Ratio, suggesting that investors seeking higher returns must be willing to accept higher downside risk.

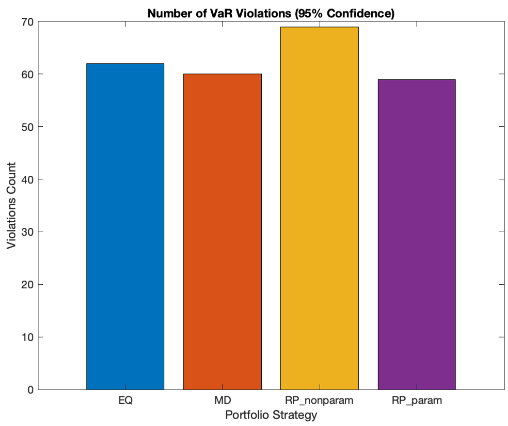


Figure 13: Comparison of VaR Violations Across Portfolio Strategies

The number of **VaR violations** assess how well a portfolio's risk model predicts potential losses. A high number of VaR violations suggests that the portfolio's risk estimates are either underestimating risk or not capturing tail events effectively. On the other hand, a lower number of violations indicates that the risk model is more conservative or better calibrated to historical risk events.

In this analysis, the Risk Parity (Non-Parametric) presented the highest number of VaR violations, with 69 breaches during the out-of-sample period. Although the non-parametric approach uses empirical return distributions, which theoretically better capture tail risk and skewness, the results show the opposite behaviour. This can be caused by overfitting to in-sample data or greater sensitivity to extreme events that were not well accounted for in the historical sample, resulting in a higher frequency of breaches when tested out-of-sample. The Risk Parity (Parametric) and Maximum Diversification portfolios recorded 59 and 60 VaR violations respectively, representing better alignment between their predicted risk estimates and actual portfolio losses. These strategies benefit from systematic risk allocation and diversification principles, which can mitigate extreme losses and reduce downside risk. The Equally-Weighted portfolio has 62 VaR violations. Since it does not account for the risk characteristics, this exposed the portfolio to riskier and more volatile assets, causing greater number of breaches.

In summary, the results illustrate the inherent trade-offs among these strategies. The Risk Parity portfolios achieved more balanced risk contributions, though their performance was constrained by reliance on unstable covariance estimates. The Maximum Diversification portfolio provided excellent downside protection through risk spreading, hence delivering the highest Sharpe ratio. Meanwhile, the Equally Weighted portfolio delivered the cumulative return over the out-of-sample period, it did so at the cost of higher maximum drawdown and uneven risk contributions.

4 Estimating the Probability of Losing (Question 5)

We performed an analysis of the probability that a portfolio will lose more than 5% of its value over horizons ranging from 1 to 50 days. Two methods are used to obtain these estimates. The first method uses a non-parametric bootstrap simulation, which resamples historical daily log-returns in order to simulate multi-day cumulative returns. The second method assumes that daily log-returns follow a Gaussian distribution, allowing for a closed-form calculation of the probability of a 5% loss.

All data used in this analysis comes from the historical adjusted close prices of the set of technology stocks mentioned in Question 1 and 2. The portfolio is assumed to be equally weighted, and daily log-returns are computed by transforming the adjusted closing prices.

Because the portfolio under consideration is equally weighted, the daily portfolio log-return on day t is simply the average of the daily log-returns of all stocks under analysis. Letting $w_i = \frac{1}{N_{\text{Assets}}}$ denote the weight of each asset, the portfolio log-return on day t is:

$$r_t^{(p)} = \sum_{i=1}^{N_{\text{Assets}}} w_i \, r_{t,i}.$$

The result is a time series of daily log-returns of length T, where T is the total number of trading days in the dataset.

Considering that we are working with log-returns, we used a continuous-compounding approach to set the threshold at $\log(0.95)$, which is approximately -0.0513. A cumulative log-return below this value implies that the portfolio has lost more than 5% of its value.

As previously mentioned, the first method for estimating the probability of losing more than 5% uses a non-parametric bootstrap approach, which relies on the historical distribution of daily log-returns rather than assuming any particular distribution form.

In the bootstrap procedure, we created 1000 return paths by resampling from the historical daily log-returns with replacement. Each simulated path has a length equal to the maximum horizon of interest (between 1 and 50 days). The daily log-returns are then accumulated sequentially over the desired time horizon to produce a series of multi-day cumulative returns, which collectively form the empirical distribution used in the analysis.

Once the cumulative returns are calculated, we compare them with the loss threshold $\log(0.95)$. For each day h in the 50-day horizon, the number of simulated paths whose cumulative returns fall below the threshold is recorded. This is then averaged over all bootstrap iterations to yield a final probability estimate of losing more than 5% on day h. By repeating this process for each h from 1 to 50, we obtain a sequence of probabilities that reflect the empirical risk of a 5% loss at various horizons.

The second method for estimating the probability of a 5% loss relies on the assumption that daily log-returns are independent and identically distributed as a normal random variable with mean μ and standard deviation σ. Under this assumption, the sum of h daily log-returns follows a normal distribution with mean $h\mu$ and standard deviation $\sqrt{h}\,\sigma$.

To identify whether a 5% loss has occurred over an h-day period, we again use the threshold $\log(0.95)$. The probability of a cumulative log-return being below this threshold is computed using the standard normal cumulative distribution function (CDF). Concretely, if $X_h = \sum_{t=1}^{h} r_t^{(p)}$ denotes the h-day cumulative log-return, then

$$\mathbb{P}[X_h < \log(0.95)] = \Phi\left(\frac{\log(0.95) - h\mu}{\sqrt{h}\,\sigma}\right),$$

where $\Phi(\cdot)$ is the standard normal CDF. Since μ and σ can be estimated from the sample mean and sample standard deviation of the daily log-returns, this formula provides a direct way to calculate the theoretical probability of losing more than 5% for each horizon h.

To facilitate comparison, both the bootstrap-based probabilities and the Gaussian-based probabilities are plotted on figure 14, with the horizontal axis representing the time horizon (from 1 to 50 days) and the vertical axis showing the probability of exceeding the 5% loss threshold. As we can see from the figure, both the bootstrap-based (green) and Gaussian-based (red) curves exhibit a rising trend as the investment horizon extends from 1 to 50 days, with the Gaussian approach ultimately plateauing around 17.5%. This increase reflects the effect of uncertainty over time. Each daily return contributes its own risk, so as additional days are accumulated, the overall variability of the cumulative return increases. As the time horizon extends, the variance of cumulative returns grows. This increase in variance causes the distribution of returns to widen over time. Consequently, even if the average return remains constant, the probability of extreme outcomes - such as incurring a loss greater than 5% - increases. This effect arises directly from the scaling property of volatility.

The bootstrap line shows more variability, particularly at longer horizons, which is expected given that it directly samples from historical data. Empirical methods are more prone to capture volatility clustering and heavier tails, therefore introducing fluctuations in the estimated probabilities. In contrast, the Gaussian curve is noticeably smoother because it assumes that returns follow a normal distribution. Under this assumption, the distribution of cumulative returns is entirely characterised by the sample mean and variance, leading to a more uniform progression in probability estimates.

Despite these differences, both approaches converge to a similar level of risk, suggesting that the normality assumption does not dramatically misrepresent the data for this set of returns. However, the slightly higher variability in the bootstrap curve indicates that actual returns may deviate from normality, showing that parametric methods can sometimes overlook important characteristics of return behaviour.

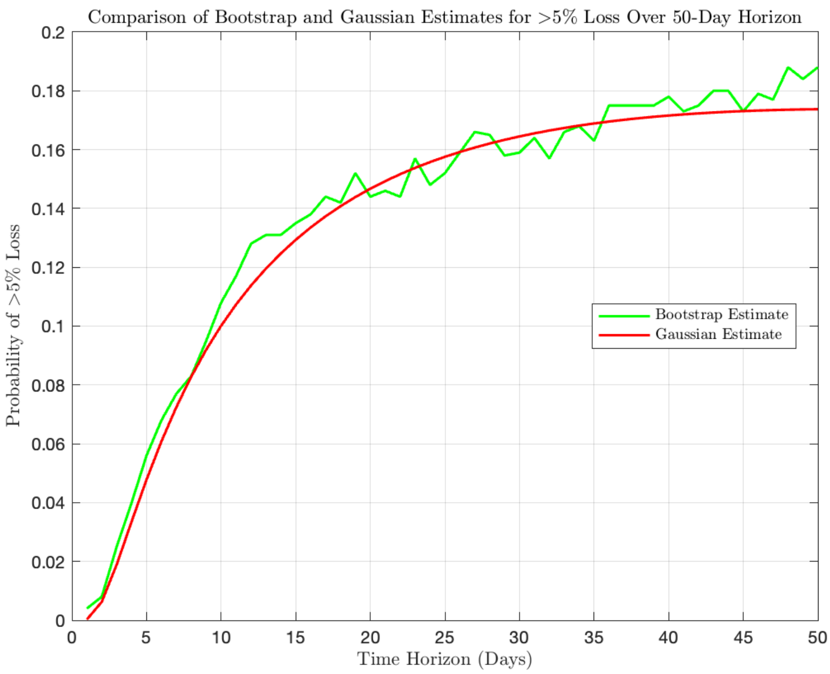


Figure 14: Bootstrap vs Gaussian Probability of a 5% loss over an horizon of 50 days.

As a final note we would like to emphasise that even though our analysis focuses on estimating the probability of losing more than 5% over various horizons, it is important to recognise that this measure of tail risk is closely related to the concept of Value at Risk (VaR). In essence, the "probability of losing more than 5%" answers the question, "Given my return distribution, how likely is it that I suffer at least a 5% loss?" Here, the loss threshold is fixed at 5%, and the goal lies on trying to predict how likely it is to exceed that threshold. In contrast, the VaR at 5% addresses a different question: "What is the smallest loss level, L, such that there is only a 5% chance of experiencing a loss greater than L?" Mathematically, while evaluating

$$P(\text{Loss} > 5\%)$$

computes the probability at a given loss level of 5%, determining $\text{VaR}_{5\%}$ involves finding the quantile α that satisfies

$$P(\text{Loss} > \alpha) = 5\%.$$

These two approaches provide complementary perspectives: one fixes the loss threshold and evaluates the probability of exceeding it, while the other fixes the tail probability and determines the corresponding loss threshold.

In summary, our investigation into the probability of losing more than 5% over horizons of up to 50 days aligns with broader themes from earlier VaR analyses. As with other non-parametric approaches, the bootstrap offers a flexible reflection of historical return distributions - capturing heavier tails and non-normal features - but assumes that past patterns remain predictive. Meanwhile, the Gaussian approach provides a more direct, closed-form solution but may understate extreme outcomes if returns deviate substantially from normality. Although both methods yield rising probabilities that converge near 17.5% by day 50, refining these techniques can improve tail-risk estimation. A block or hybrid block-filtered bootstrap approach can better preserve time-dependence and volatility clustering. Incorporating GARCH-based volatility forecasting or EWMA may enhance responsiveness to evolving market conditions. Ultimately, combining empirical and analytical strategies - and periodically revisiting model assumptions - provide deeper insights into risk management.

5 Conclusion

Investors, unaware of risk, may be tempted to chase high returns. Our analysis demonstrates that a correct understanding of risk is essential to portfolio construction.

Simple parametric models fail to capture heavy-tailed distributions and volatility clustering observed in our dataset. Empirical models go some way towards addressing this, but further analysis is required to capture tail behaviour.

By comparing three portfolio constructions, we demonstrate that the largest cumulative returns often comes with significantly higher risk, and a larger Maximum Drawdown, than its risk-adjusted counterparts.

Finally, while Gaussian models produce smoother term structures of probability of loss, there is no guarantee that this comprehensively captures the empirical distribution, particularly when data exhibits non-normal properties.

References

Philippe Artzner, Freddy Delbaen, Jean-Michel Eber, and David Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999. doi: 10.1111/1467-9965.00068.

Laura Ballotta and Gianluca Fusai. A gentle introduction to value at risk. *SSRN Electronic Journal*, March 28 2017. doi: 10.2139/ssrn.2942138. URL `https://ssrn.com/abstract=2942138`.

Peter Christoffersen. *Elements of Financial Risk Management*. Academic Press, San Diego, 2003.

Rama Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

Keith Cuthbertson and Dirk Nitzsche. *Investments*. Wiley, 2nd edition, 2008. ISBN 978-0470519561.

Marcos López de Prado. *Advances in Financial Machine Learning*. John Wiley & Sons, Hoboken, NJ, 2018. ISBN 978-1-119-48208-6.

Paul Embrechts, Claudia Klüppelberg, and Thomas Mikosch. *Modelling Extremal Events for Insurance and Finance*, volume 33 of *Applications of Mathematics*. Springer, Berlin, Heidelberg, 1997. doi: 10.1007/978-3-642-33483-2.

Soumendra N. Lahiri. *Resampling Methods for Dependent Data*. Springer, New York, 2003.

Dimitris N. Politis and Joseph P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994.