Bocconi

Bachelor of Science in Economics,

Management and Computer Science

Bayesian Portfolio Optimization:

From Diffuse Beliefs to Economically Informed Models Beyond the Limits of the Mean-Variance Paradigm

Advisor:

Prof. ANTONIO LIJOI

Bachelor of Science thesis by:

ANDREA ALBINO VENTO

student ID no 3224274

Academic Year 2024–2025

Università Commerciale
Luigi Bocconi

Abstract

This thesis develops a Bayesian approach to the portfolio optimization problem, addressing the limitations of classical mean-variance optimization (MVO), such as its high sensitivity to estimation error and the instability of resulting portfolios. The empirical analysis implements a series of Bayesian portfolio models of increasing structural complexity: starting from diffuse priors, extending to conjugate Normal-Inverse-Wishart specifications, and concluding with the Bayesian CAPM which integrates economic structure into the prior distribution of expected returns. All models are evaluated through a rolling-window backtesting design on a fixed cross-section of major US assets from 2008 to 2022. The results indicate that Bayesian strategies, particularly those with economically informed priors and regularization terms, deliver more stable allocations, superior out-of-sample Sharpe ratios, and improved diversification compared to the classical MVO. Future research could extend this framework to include more sophisticated asset pricing models and relax some technical assumptions to better capture the complexities of financial markets.

Contents

1 Introduction

1.1 Context and Motivation

Investing is a forward-looking act, executed in the present, and guided by backward-looking information. Portfolio optimization aims to formalize this decision by striking a balance between the desired outcomes and the investor's degree of risk tolerance.

Despite being a problem that has been studied for decades, portfolio optimization remains difficult due to the necessity of choosing where to allocate capital under incomplete information, the influence of subjective beliefs, and the presence of continuously evolving market dynamics.

These challenges highlight the necessity of explicitly accounting for uncertainty and noise, learning from data over time, and continuously updating subjective beliefs as new information becomes available. In this context, Bayesian models arise naturally, introducing a mechanism face uncertainty by combining prior beliefs, external information, and observed data in a probabilistically consistent manner.

1.2 Objectives and Contributions

The modern theory of portfolio selection has its roots in the work of Harry Markowitz [4], who formalized the trade-off between risk and return introducing the mean-variance optimization (MVO) framework. Markowitz's work, introduced in Chapter 3, represented a fundamental breakthrough in investment theory, demonstrating that portfolio risk should not be evaluated by considering assets in isolation, but must instead account for the covariance structure among them. This fundamental principle is formally captured by a simple yet powerful expression that relates the risk of individual assets to the overall risk of a portfolio composed of them. Let w_1 and w_2 denote the portfolio weights assigned to assets, and let R_1 and R_2 denote their respective returns. Then, the variance of the portfolio return is given by:

$$\mathrm{Var}(w_1 R_1 + w_2 R_2) = w_1^2 \,\mathrm{Var}(R_1) + w_2^2 \,\mathrm{Var}(R_2) + 2 w_1 w_2 \,\mathrm{Cov}(R_1, R_2), \tag{1}$$

which shows that when $\mathrm{Cov}(R_1, R_2) < 0$, the total portfolio variance can be lower than the variance of either asset considered in isolation. This illustrates the diversification effect: combining imperfectly correlated assets can reduce overall portfolio risk.

This shifted the focus of portfolio construction from the selection of individual assets with the highest expected returns to the identification of combinations of assets that are imperfectly correlated. As a consequence, investors began prioritizing diversification benefits over stand-alone asset performances, as famously stated by Markowitz himself in his book *Foundations of Portfolio Theory* [5]: "diversification is the only free lunch in finance".

Despite its crucial contributions to portfolio theory, the classical mean-variance optimization framework has well-known shortcomings when it comes to practical applications. In particular, it relies on sample estimates of expected returns and covariances, which are based on limited historical data and are subject to substantial estimation error. As a result, optimal portfolios often exhibit extreme and unstable allocations, limited diversification, and high sensitivity to small variations in input estimates, as discussed by Michaud [7].

This thesis addresses these limitations by adopting a Bayesian approach to the portfolio optimization problem. In contrast to the classical approach, the Bayesian methodology treats model parameters as random variables and allows the investment choice to be informed by prior beliefs, external information, and observed data. This allows for more robust inference, particularly in settings where historical data are limited or noisy, and enables dynamic learning over time as new information becomes available. After an overview of the Bayesian framework (Chapter 4), the analysis proceeds by constructing a sequence of increasingly structured Bayesian models. The starting point is a Bayesian mean-variance optimization framework under a diffuse prior 5.2, which incorporates parameter uncertainty without introducing strong prior information. This is then extended to a conjugate setting using the Normal-Inverse-Wishart prior 5.3, which allows for analytical tractability and inclusion of external information by accurately setting the hyperparameters. Building on this foundation, the thesis introduces a Bayesian formulation of the Capital Asset Pricing Model (CAPM) (Chapter 6), where asset pricing theory is used to inform the prior distribution of parameters, incorporating economic structure in the estimation process. In order to evaluate all models in a consistent manner, the analysis is conducted on a fixed selection of the 20 largest S&P 500 constituents as of January 2008, complemented by a long-duration Treasury bond. Excess returns are computed monthly from January 2008 to December 2022 using data from WRDS and FRED (Chapter 2). Portfolio performance is assessed through a rolling-window backtesting procedure with 60-month estimation windows and 2-month holding periods, based on metrics designed to evaluate risk-adjusted performance and stability .

2 Dataset and research design

2.1 Asset Universe and Data Description

The dataset used in this study is sourced from the Wharton Research Data Services (WRDS) [11] platform and spans the period from January 2008 to December 2022, resulting in a total of 180 monthly observations. All asset returns are expressed as monthly excess returns relative to a risk-free benchmark.

The sample period is deliberately chosen to capture a wide range of economic conditions, including significant structural shifts in global financial markets. It begins in January 2008, coinciding with the the Global Financial Crisis, which marked a crucial turning point in risk perception and asset correlations. The period that follows includes the subsequent recovery, the prolonged era of low interest rates, the COVID-19 pandemic, and the return of inflationary pressures and monetary tightening in 2022. This broad temporal scope enables an evaluation of the performance of portfolio optimization methods under different and evolving market regimes.

The core equity universe comprises the 20 largest-capitalization stocks listed in the S&P 500 index as of January 3, 2008, the first trading day of the year, fixed ex-ante to eliminate survivorship and look-ahead biases. In particular, survivorship bias arises when only assets that remain in the index at the end of the sample are included in the analysis, thereby excluding firms that were delisted or experienced poor performance. This inflates historical performance and leads to unrealistic inferences. Similarly, look-ahead bias occurs when information not available at the decision point is used in model construction. By fixing the asset universe at the beginning of the sample, the analysis ensures that all investment decisions are made considering only the information available at the time the investment decision would have taken place. This approach replicates the constraints faced by real-world investors and guarantees that all backtested results reflect feasible, implementable strategies under conditions of realistic uncertainty. In addition, to complement equity exposure and enable cross-asset diversification, the dataset includes a fixed-income component represented by a long-duration United States Treasury bond.

Excess returns for asset i at time t are calculated as:

$$R_{i,t}^{\text{excess}} = R_{i,t} - r_{f,t}, \tag{2}$$

where $R_{i,t}$ denotes the asset's monthly total return and $r_{f,t}$ represents the risk-free rate for the corresponding month. The risk-free rate is derived from the FRED DGS1MO dataset, which reports the 1-month U.S. Treasury constant maturity yield. These annualized yields are converted to monthly total returns using the standard compounding approximation:

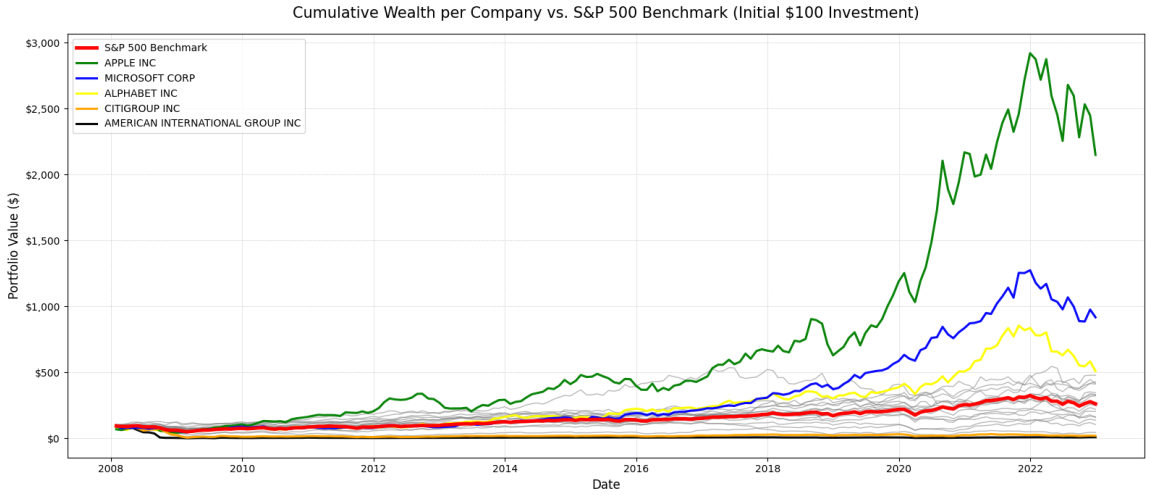
$$r_{f,t} = \left(1 + \frac{r_t^{\text{annual}}}{100}\right)^{1/12} - 1, \tag{3}$$

where r_t^{annual} is the yield expressed in percentage points. This transformation ensures consistency across return series and allows proper measurement of excess returns across equities and bonds.

To provide a meaningful external reference, portfolio performance throughout the study is benchmarked against the S&P 500 index. This benchmark represents a widely accepted proxy for passive market exposure and enables evaluation of whether active allocation strategies generate superior risk-adjusted returns across different economic conditions.

2.2 Exploratory Data Analysis

As a preliminary step, it is relevant to examine the overall behavior of each asset and the benchmark over the sample period. Figure 2.2 presents the compound growth of an initial $100 investment in every asset, alongside the S&P 500.



Clearly, there is substantial cross-sectional dispersion in realized returns: a small number of stocks, such as Apple Inc. and Microsoft Corp., achieved exceptional growth, while others, including Citigroup Inc. and American International Group Inc., were not as successful. Most

of the constituents underperformed the aggregate index, highlighting the importance of careful investment decisions.

To further characterize the dataset, the average behavior of the assets is compared to that of the S&P 500 benchmark in terms of mean monthly excess return and return volatility (Figure 1).

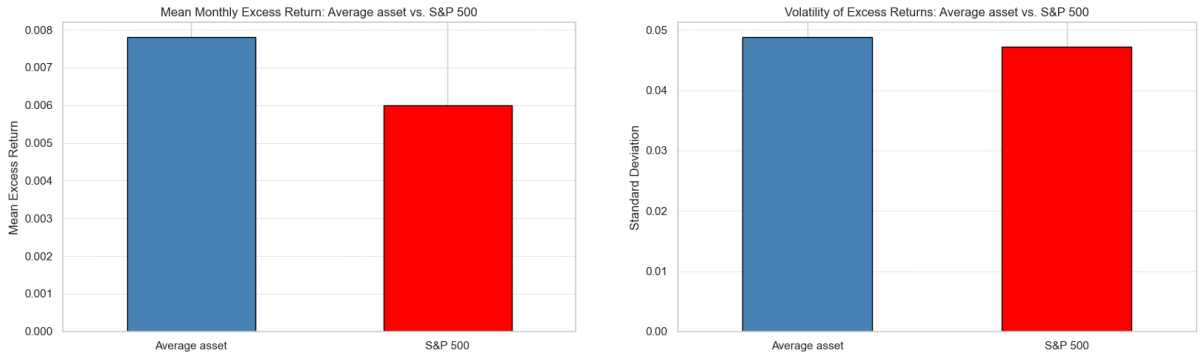


Figure 1: Mean and standard deviation of excess returns: Asset pool vs. S&P 500

While the asset pool exhibits greater average monthly returns than the benchmark, the mean volatility is slightly higher, which is something the models will have to mitigate by exploiting asset covariances.

Furthermore, a preliminary investigation of the characteristics of the distribution of asset returns is conducted to evaluate the plausibility of key modeling assumptions, to be introduced later. In particular, the classical mean-variance optimization framework assumes that asset returns are normally distributed. Notably, it is well-documented that return distributions exhibit heavier tails and greater deviation from normality at higher frequencies, such as daily data. In contrast, monthly returns tend to more closely approximate a normal distribution, with reduced skewness and thinner tails, making them more consistent with the assumptions of Gaussian-based models.

To test the initial modeling choice, a series of univariate normality tests is applied to each asset's return series. Specifically, the following tests are conducted: the Shapiro Wilk test, which is sensitive skewness and kurtosis; the Jarque Bera test, which directly tests for joint normality via skewness and excess kurtosis; the Anderson Darling test, which places greater weight on the tails of the distribution.

Each test is applied to both raw returns and log returns, resulting in a greater number of assets passing the normality tests under the raw return specification. Based on this evidence, the analysis proceeds using absolute returns rather than log returns, as they better support the Gaussian assumption.

2.3 Backtesting Methodology

A proper backtesting methodology is essential to provide an unbiased assessment of portfolio optimization models and to ensure that empirical results reflect the realistic constraints faced by investors. A rolling-window backtesting design will be implemented, explicitly constructed to avoid look-ahead bias and to facilitate the dynamic updating of model parameters and beliefs as time passes and more data is available.

An initial period of 60 months is used to observe the behavior of the assets, during which no investment decisions are made. After this preliminary phase, beginning in January 2013, the preceding 60 months of data are used to determine the optimal portfolio strategy, which is then implemented and held over the subsequent 2-month period. Once this holding period concludes, the estimation window is advanced by two months, incorporating the newly available return data while discarding the oldest observations. This updated 60-month window serves as the basis for the next portfolio strategy, which will be held for the successive 2 months, and so on until the end of the overall investment period. The choice of a 60-month window reflects a balance between statistical reliability and adaptability, as a five-year period is generally sufficient to capture different market regimes and structural shifts, while still allowing the models to respond to new information over time.

To ensure the validity of statistical estimates within each estimation window, it is important that asset returns exhibit local stationarity, that is, their mean and variance remain approximately

constant between the estimation window and the holding period. To assess this, the Augmented Dickey-Fuller (ADF) test was applied to each asset's return series during every rebalancing step. The results confirm that the local stationarity assumption is well-supported: in all estimation windows, at least 95% of assets passed the test at the 5% significance level.

This recursive backtesting process continues throughout the sample period, ensuring that every portfolio allocation decision is made strictly with data that would have been available at the corresponding decision point. This design precludes look-ahead bias by construction and provides a robust framework for evaluating the dynamic performance of both classical and Bayesian portfolio optimization models. To enable a comprehensive and consistent evaluation of all portfolio optimization approaches, key performance metrics, to be introduced later, are systematically calculated using out-of-sample returns generated by the backtesting procedure.

3 Classical Mean-Variance Optimization

3.1 Theoretical Framework

The quantitative study of portfolio selection was fundamentally shaped by the work of Harry Markowitz, who first formalized the trade-off between risk and return in a mathematical framework [4]. Markowitz recognized that investors are not solely concerned with maximizing expected returns but also with the uncertainty associated with those returns. This insight led to the foundation of the mean-variance paradigm, in which risk is identified with the variance of portfolio returns.

Within this paradigm, the fundamental link between return and risk is characterized by the idea that, ceteris paribus, investors prefer assets or portfolios that offer higher expected returns and lower variance. Specifically, for any two portfolios with the same expected return, rational investors will always prefer the portfolio with the lower variance of returns. Conversely, for a given level of risk, investors will select the portfolio with the highest expected return.

The standard setting of the Markowitz portfolio problem considers an investor allocating wealth among n risky assets. The returns of the assets are represented by the random vector $\mathbf{R} = (R_1, \ldots, R_n)^\top$, where R_i denotes the return on asset i. The expected returns are collected in the vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, where $\mu_i = \mathbb{E}[R_i]$, and the risks are summarized by the covariance matrix $\boldsymbol{\Sigma}$, where $\Sigma_{ij} = \mathrm{Cov}(R_i, R_j)$. The portfolio is defined by the vector of weights $\mathbf{w} = (w_1, \ldots, w_n)^\top$, where w_i denotes the proportion of wealth allocated to asset i.

Several key assumptions are included in the classical formulation. Parameters such as the expected returns $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ are assumed to be fixed and known, constant across the estimation and investment period, which is clearly a major simplification. The model further presumes the absence of market frictions, with no transaction costs, taxes, or other barriers to trading. Investors are allowed to borrow or lend at a risk-free rate, and short sales of assets are permitted unless explicitly prohibited. The investment horizon is single-period, so all investment decisions are made at the beginning of the period and evaluated at its conclusion, which, in our context, corresponds to the holding window under consideration.

A critical technical assumption is that the joint distribution of asset returns is multivariate normal. Under this assumption, for each period t, where t ranges over both the estimation

window and the investment period, asset returns are assumed to be conditionally independent and identically distributed according to a multivariate normal distribution:

$$\mathbf{R}_t \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

with density

$$f(\mathbf{R}_t \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{R}_t - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{R}_t - \boldsymbol{\mu})\right) \tag{4}$$

This implies that all returns, both in the historical estimation window and in the subsequent investment period, are modeled as draws from the same distribution with fixed parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. This is a substantial simplification, as it implies that past returns are perfectly representative and predictive of future returns, ignoring the possibility of structural breaks or parameter instability over time.

Given the above assumptions, the investor's objective is to select a portfolio weight vector $\mathbf{w}$ that remains fixed throughout the holding period. For each period t, the portfolio return is given by the inner product of $\mathbf{w}$ and the realized asset returns vector $\mathbf{R}_t$:

$$R_{p,t} = \mathbf{w}^\top \mathbf{R}_t = \sum_{i=1}^{n} w_i R_{i,t} \tag{5}$$

where w_i denotes the weight assigned to asset i and $R_{i,t}$ denotes the return of asset i at time t.

Under the classical assumptions, the expected return and variance of the portfolio are independent of time and can be expressed as:

$$\mu_p = \mathbb{E}[R_{p,t}] = \mathbf{w}^\top \boldsymbol{\mu} = \sum_{i=1}^{n} w_i \mu_i \tag{6}$$

$$\sigma_p^2 = \mathrm{Var}(R_{p,t}) = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \Sigma_{ij} \tag{7}$$

Since the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are assumed to be constant and known, both the expected return μ_p and variance σ_p^2 of the portfolio remain constant for every t in the investment horizon.

Normality, or alternatively the assumption that investors have quadratic utility functions, ensures that the mean $\mathbb{E}[\mathbf{w}^\top \mathbf{R}]$ and variance $\mathrm{Var}(\mathbf{w}^\top \mathbf{R})$ of the portfolio return are sufficient statistics

for the portfolio decision problem, providing all the information needed to make investment choices.

This setting reduces the portfolio selection problem to identifying the weights $\mathbf{w}$ that optimally balance expected return and risk, in particular solving:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \sigma_p^2 = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \\ \text{subject to} \quad & \mu_p = \mathbf{w}^\top \boldsymbol{\mu} = \mu^*, \\ & \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{8}$$

or, equivalently,

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mu_p = \mathbf{w}^\top \boldsymbol{\mu} \\ \text{subject to} \quad & \sigma_p^2 = \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \leq \sigma^{*2}, \\ & \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{9}$$

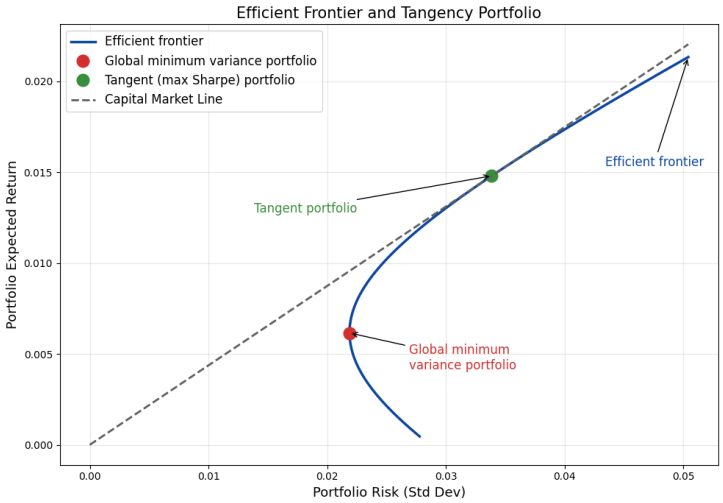
A third, equivalent formulation involves maximizing the mean-variance objective function, or equivalently, the expected utility of wealth under quadratic utility:

$$\underset{\mathbf{w}}{\text{maximize}} \quad \mathbf{w}^\top \boldsymbol{\mu} - \frac{\lambda}{2} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \tag{10}$$

where $\lambda > 0$ denotes the investor's risk aversion parameter. This formulation yields the same set of optimal portfolios as the previous two when returns are normally distributed or preferences are quadratic.

The portfolio optimization problem is fundamentally linked to individual preferences: each investor needs to select the required monthly expected return (as in 8) or the tolerated monthly risk (as in 9), obtaining a unique optimal portfolio. By varying the target return or risk, the set of optimal portfolios is obtained, referred to as the efficient frontier. This frontier consists of portfolios that deliver the highest expected return for each level of risk, or, equivalently, the lowest risk for each level of expected return. Portfolios lying below the upper boundary of the

frontier are suboptimal, as more favorable risk-return combinations are available. The efficient frontier is typically upward sloping and concave, reflecting diminishing incremental gains in expected return as risk increases.



The global minimum variance portfolio attains the lowest possible risk among all feasible portfolios; in our empirical analysis, this portfolio is characterized by $\mu = 0.00615$ and a standard deviation $\sigma = 0.0218$. The tangency portfolio (or maximum Sharpe ratio portfolio), with $\mu = 0.015$ and $\sigma = 0.033$ in our case, achieves the highest ratio of excess return to risk and, in the presence of a risk-free asset, determines the Capital Market Line. The Sharpe ratio is defined as

$$S = \frac{\mathbf{w}^\top \boldsymbol{\mu}_{rf}}{\sqrt{\mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w}}}$$

where $\boldsymbol{\mu}_{rf}$ is the vector of expected excess returns. The tangency portfolio is obtained by maximizing the Sharpe ratio with respect to the portfolio weights, and can be computed analytically using the closed-form solution:

$$\mathbf{w}^* = \frac{\boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{rf}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{rf}}$$

3.2 Practical Shortcomings and empirical evidence

Although the classical mean-variance framework provides a foundational theoretical basis for portfolio selection, there are some fundamental flaws in its practical implementation.

The main limitation comes from the fact that the model parameters, which are the expected return vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, are not known with certainty but must be estimated

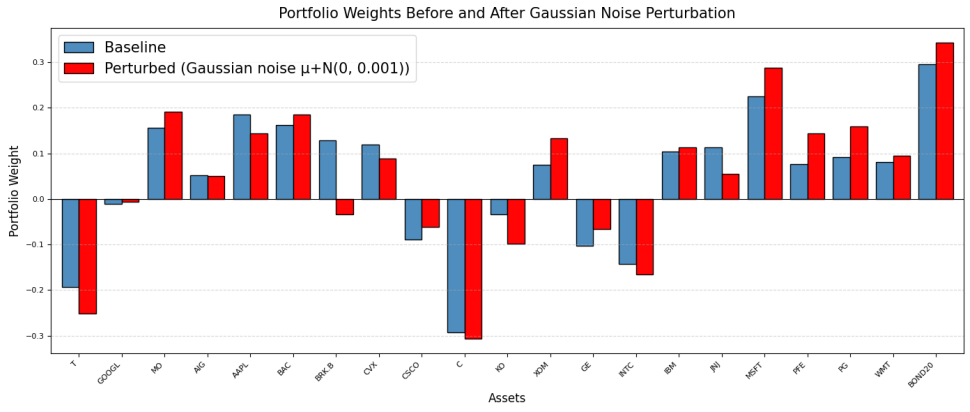
from finite samples of historical data. In the classical approach, the standard frequentist unbiased estimators are employed:

$$\hat{\boldsymbol{\mu}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}_t, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}(\mathbf{R}_t - \hat{\boldsymbol{\mu}})(\mathbf{R}_t - \hat{\boldsymbol{\mu}})^\top$$

where T is the sample size and $\mathbf{R}_t$ is the vector of observed returns at time t. However, the estimation risk is not incorporated into the optimization, possibly underestimating the variance in the returns.

The practical consequences of estimation risk cannot be ignored, as they can substantially affect allocation decisions. Small errors in the estimates of expected returns can have disproportionately large effects on the optimal portfolio weights. Chopra and Ziemba [1] examine the relative impact of estimation errors and find that errors in the means can be up to ten or eleven times more impactful than errors in variances. Moreover, Merton [6] notes that variances and covariances of returns are more stable over time than expected returns.

As a result, classical mean-variance optimized portfolios are often unstable, with weight allocations that change dramatically in response to even minor changes in input estimates. This issue is empirically evident in our analysis: when the expected returns are estimated over the entire 180 month sample and then subjected to a small Gaussian perturbation $\mathcal{N}(0, 0.0015)$, the tangency portfolio weights differ on average by 44.07% compared to the baseline solution, as shown in the figure below.



Moreover, this estimation risk is further amplified when the estimation window is reduced. While the above demonstration uses 180 monthly observations, the potential instability becomes even more pronounced in realistic rolling strategies where the estimation sample is only 60 months.

This instability leads directly to high portfolio turnover, as substantial rebalancing is frequently required, increasing potential transaction costs and reducing the reliability of portfolio decisions.

For these reasons, classical mean-variance optimization is prone to overfitting, often delivering promising results on the training or estimation data, while performing poorly out of sample. Unconstrained optimization also tends to produce concentrated portfolios that reflect overfitting to noise, resulting in insufficient diversification and excessive exposure to individual assets.

To address these challenges, practitioners commonly introduce constraints to improve stability, limit turnover, and enhance robustness. However, it is important to note that imposing too many constraints may transform the optimization problem into one of mere feasibility, leaving little space for effective portfolio selection or risk-return trade-off.

3.3 Evaluation Metrics

To assess the empirical performance and robustness of portfolio optimization strategies, several evaluation metrics are employed.

The primary measure of investment performance is the mean monthly return, computed as

$$\bar{r}_p = \frac{1}{T} \sum_{t=1}^{T} R_{p,t}$$

where $R_{p,t}$ denotes the realized out-of-sample return of the portfolio at time t and T is the number of evaluation periods. For interpretability, annualized returns can be computed as: $r_{\text{ann}} = (1 + \bar{r}_p)^{12} - 1$.

Portfolio risk is evaluated by the standard deviation of realized monthly returns,

$$\hat{\sigma}_p = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (R_{p,t} - \bar{r}_p)^2}$$

which reflects the ex post volatility of the portfolio.

Risk-adjusted performance is measured using the annualized Sharpe ratio, given by

$$\text{Sharpe}_{\text{ann}} = \frac{\bar{r}_p}{\hat{\sigma}_p} \sqrt{12}$$

which measures the trade-off between risk and reward.

In order to quantify the stability of investment choices, portfolio turnover is measured at each rebalancing period:

$$\text{Turnover}_t = \sum_{i=1}^{n} |w_{i,t} - w_{i,t-1}|$$

where $w_{i,t}$ is the weight of asset i in the portfolio at time t. Hence, the average turnover across the backtesting period is given by

$$\text{Average Turnover} = \frac{1}{T-1} \sum_{t=2}^{T} \text{Turnover}_t$$

To assess diversification, the Herfindahl-Hirschman Index (HHI) is computed at each period as

$$\text{HHI}_t = \sum_{i=1}^{n} w_{i,t}^2$$

with lower values indicating better diversification and higher values indicating concentration of the portfolio in a few assets.

3.4 Empirical implementation and results

3.4.1 Strategies

In implementing the classical Mean-Variance Optimization (MVO), three different strategies are used. In all cases, the optimization seeks to maximize the expected return subject to a target monthly volatility of 0.03, which corresponds to an annualized volatility of approximately 10.39%, reflecting the typical risk profile of diversified equity portfolios and provides a realistic benchmark for portfolio construction.

The unconstrained mean-variance optimization problem is formulated as

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mathbf{w}^\top \boldsymbol{\mu} \\ \text{subject to} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \leq 0.03^2, \\ & \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{11}$$

with no additional constraints.

The ridge-regularized optimization introduces an additional penalty term in the objective to control instability and prevent extreme allocations:

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mathbf{w}^\top \boldsymbol{\mu} - \lambda \|\mathbf{w}\|_2^2 \\ \text{subject to} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \leq 0.03^2, \\ & \sum_{i=1}^{n} w_i = 1 \end{aligned} \tag{12}$$

where $\lambda > 0$ (set to 0.1) is the regularization parameter and $\|\mathbf{w}\|_2^2 = \sum_{i=1}^{n} w_i^2$ is the squared ℓ_2 norm of the weight vector.

The long-only strategy imposes non-negativity constraints on all asset weights:

$$\begin{aligned} \underset{\mathbf{w}}{\text{maximize}} \quad & \mathbf{w}^\top \boldsymbol{\mu} \\ \text{subject to} \quad & \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \leq 0.03^2, \\ & \sum_{i=1}^{n} w_i = 1, \\ & w_i \geq 0 \quad \text{for all } i = 1, \ldots, n \end{aligned} \tag{13}$$

where the additional constraint is used to avoid shorting positions, which can not be taken by some investors.

3.4.2 Preliminary Results

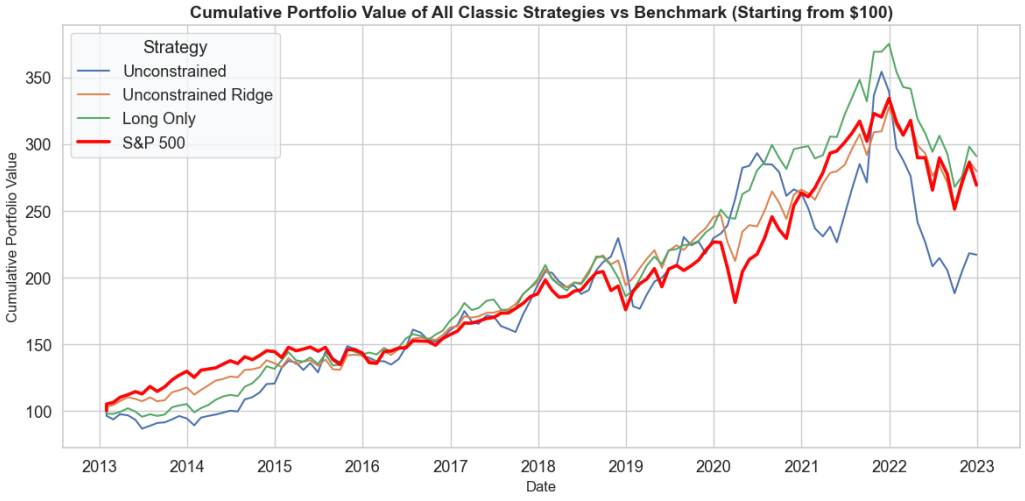


Figure 2: Cumulative portfolio value of all strategies and S&P 500 index, 2013–2023.

The mean monthly return is highest for the long-only and ridge-regularized strategies, 0.009 (11.36% annualized), marginally higher than the unconstrained portfolio, 0.007 (8.70%). As expected, the application of regularization or a long-only constraint enhances out-of-sample returns relative to the unconstrained case, which is prone to estimation error and instability.

Table 1: Mean Monthly and Annualized Returns

Strategy	**Mean Monthly Return**	**Annualized Return**
Unconstrained	0.0074	9.24%
Unconstrained Ridge	0.009	11.36%
Long Only	0.009	11.36%
S&P 500	0.006	7.43%

In terms of risk, the unconstrained strategy exhibits the highest volatility at 0.055, while the ridge-regularized and long-only portfolios achieve substantially lower volatilities of 0.036 and 0.037, respectively. This noticeable risk reduction observed in the regularized and long-only cases highlights the instability of the unconstrained strategy, the volatility of which falls far from the target of 0.03.

Table 2: Monthly and Annualized Volatility

Strategy	**Volatility (Monthly Std)**	**Annualized Volatility**
Unconstrained	0.055	0.190
Unconstrained Ridge	0.036	0.124
Long Only	0.037	0.128
S&P 500	0.047	0.163

It should be noted that although the unconstrained mean-variance portfolio displays a higher mean monthly return than the benchmark, its cumulative return over the sample period is lower. This apparent contradiction is explained by the high volatility and significant drawdowns experienced by the unconstrained portfolio. Since cumulative return is determined by the geometric mean, which is adversely affected by volatility, the elevated risk of the unconstrained strategy leads to inferior compounded performance despite its higher arithmetic mean return.

Risk-adjusted performance, as measured by the annualized Sharpe ratio, is maximized by the ridge-regularized and long-only strategies, which achieve Sharpe ratios of 0.832 and 0.842, respectively, almost doubling the value observed for the unconstrained approach (0.463) and substantially exceeding the S&P 500 (0.439). This improvement is attributable to the joint enhancement in mean return and reduction in volatility afforded by these methods. In contrast, the unconstrained MVO yields poor risk-adjusted outcomes, underscoring the negative impact of estimation risk.

Table 3: Annualized Sharpe Ratio.

Strategy	**Annualized Sharpe Ratio**
Unconstrained	0.463
Unconstrained Ridge	0.832
Long Only	0.842
S&P 500	0.439

Portfolio turnover reflects the stability and implementability of the strategies. The unconstrained strategy is characterized by extremely high turnover, with an average of 1.278, indicating nearly complete rebalancing at each period. Ridge-regularization and the long-only constraint reduce average turnover dramatically, to 0.139 and 0.297, respectively. The S&P 500 benchmark, as a buy-and-hold index, naturally exhibits zero turnover. High turnover in the unconstrained approach directly reflects estimation instability, leading to excessive transaction costs and lack of practical feasibility.

Table 4: Average Turnover.

Strategy	**Average Turnover**
Unconstrained	1.278
Unconstrained Ridge	0.139
Long Only	0.297

Finally, portfolio concentration is assessed via the average Herfindahl-Hirschman Index (HHI). The unconstrained portfolio is highly concentrated, with an HHI of 1.127, indicating significant exposure to a small subset of assets and limited diversification. The ridge-regularized portfolio

achieves a markedly lower HHI of 0.075, consistent with near-uniform diversification, while the long-only strategy results in an intermediate concentration level (HHI = 0.287). These results demonstrate the essential role of regularization and constraints in achieving effective diversification and risk control in portfolio construction.

Table 5: Average Herfindahl-Hirschman Index (HHI).

Strategy	**Average HHI**
Unconstrained	1.127
Unconstrained Ridge	0.075
Long Only	0.287

In summary, unconstrained mean-variance optimization suffers from low stability and inferior risk-adjusted performance. Regularization and long-only constraints substantially improve all empirical metrics, producing portfolios with superior return, lower risk, greater stability, and enhanced diversification compared to both the unconstrained strategy and the S&P 500 benchmark.

4 The Bayesian Paradigm

4.1 Probability as Belief

Bayesian and Frequentist statistics are two fundamental paradigms in the study of uncertainty, differing both from a methodological and a philosophical standpoint. Bayesian thinking treats probability as a degree of belief about uncertain outcomes, whereas the frequentist approach defines probability as the long-run relative frequency of events. When flipping a coin ten times and obtaining seven heads, a frequentist would estimate the fixed, but unknown probability p of heads that most likely explains the observed outcome, to find this true, objective parameter. In this setting, the data points are assumed to be independent and identically distributed from a common distribution with fixed parameter:

$$X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} f_\theta, \quad \theta \in \Theta.$$

By contrast, a Bayesian statistician begins with a prior belief regarding the probability of heads, revising or updating such belief as new information becomes available, balancing convictions and evidence. In the Bayesian framework, the parameter θ is itself treated as a random variable, and the initial belief about its value is described through a prior distribution $p(\theta)$. The data are assumed to be conditionally independent given the parameter:

$$X_1, \dots, X_n \mid \theta \overset{\text{i.i.d.}}{\sim} f_\theta, \quad \theta \sim p(\theta).$$

4.2 Bayes' Theorem and Sequential Learning

The Bayesian framework relies on Bayes Theorem to update the beliefs about the unknown parameters, incorporating new knowledge coming from observing a process.

Let $X_1, \dots, X_n$ denote observed data and θ the parameter of interest. The relationship between prior knowledge and observed evidence is expressed through the posterior distribution, given by

$$p(\theta \mid X_1, \dots, X_n) = \frac{p(X_1, \dots, X_n \mid \theta)\, p(\theta)}{p(X_1, \dots, X_n)},$$

where $p(\theta)$ is the prior distribution expressing beliefs about θ before observing the data,

$p(X_1, \ldots, X_n \mid \theta)$ is the likelihood function summarizing the information provided by the data, and $p(\theta \mid X_1, \ldots, X_n)$ is the posterior distribution representing updated beliefs after observing $X_1, \ldots, X_n$. The denominator $p(X_1, \ldots, X_n)$ serves as a normalizing constant ensuring that the posterior is a valid probability distribution:

$$p(X_1, \ldots, X_n) = \int_\Theta p(X_1, \ldots, X_n \mid \theta)\, p(\theta)\, d\theta.$$

which does not depend on θ, leading to:

$$p(\theta \mid X_1, \ldots, X_n) \propto p(\theta)\; p(X_1, \ldots, X_n \mid \theta).$$

An important consequence of the Bayesian approach is that inference does not end with the estimation of the posterior distribution of parameters but can be used to make probabilistic statements about future, unobserved data through the posterior predictive distribution.

Given observed data $X_1, \ldots, X_n$, the posterior predictive distribution for a new observation X_{n+1} is defined as

$$p(X_{n+1} \mid X_1, \ldots, X_n) = \int_\Theta p(X_{n+1} \mid \theta)\, p(\theta \mid X_1, \ldots, X_n)\, d\theta.$$

This expression reflects the Bayesian principle that predictions about new data should be averaged over the uncertainty in the parameter θ, as quantified by the posterior distribution.

In practice, posterior predictive samples are typically generated using Monte Carlo methods. Since direct iid sampling from the posterior distribution $p(\theta \mid X_1, \ldots, X_n)$ is rarely feasible, approximate samples are obtained using Markov Chain Monte Carlo (MCMC) techniques, such as Gibbs sampling or the Metropolis-Hastings algorithm. For each sampled parameter value $\theta^{(s)}$, a draw $X_{n+1}^{(s)}$ is simulated from the conditional distribution $p(X_{n+1} \mid \theta^{(s)})$. Repeating this procedure over many iterations yields an empirical approximation of the posterior predictive distribution, enabling probabilistic forecasting and risk quantification under full parameter uncertainty.

4.3 Prior Distributions

As clear from Bayes' theorem, a crucial component of the Bayesian framework is the prior distribution $p(\theta)$, which encodes all information or beliefs about the unknown parameter θ before any data are observed. The choice of prior has important implications, particularly in situations with limited data, and reflects both subjective judgments and available external information.

An informative prior provides substantial information about the likely values of θ and typically concentrates probability mass in a narrow region, often based on previous knowledge or studies. A noninformative or diffuse prior represents vague or weak prior beliefs, and its influence on the posterior distribution is limited. Examples include the uniform prior $p(\theta) \propto 1$ and the Jeffreys prior $p(\theta) \propto \sqrt{I(\theta)}$ [3], which will later be used in the diffuse Bayesian MVO framework due to its property of being invariant under reparametrization.

Moreover, a prior is said to be proper if it is a valid probability distribution, integrating to one over the parameter space,

$$\int_{\Theta} p(\theta)\, d\theta = 1.$$

If this condition is not satisfied, the prior is improper. Improper priors, such as the uniform prior $p(\theta) \propto 1$ on the real line, are frequently employed to express diffuse beliefs. While the prior itself may not correspond to a probability distribution, it is possible for the resulting posterior to be proper, that is,

$$\int_{\Theta} p(\theta \mid X_1, \ldots, X_n)\, d\theta = 1.$$

In many models, conjugate priors are employed because they lead to posteriors of the same functional form as the prior, simplifying both analysis and computation. For example, in the multivariate normal setting with unknown mean vector and covariance matrix, just like the Mean-Variance optimization one, the conjugate prior is the Normal-Inverse-Wishart distribution, which allows for analytical updating of both parameters.

The selection of an appropriate prior depends on the context, the amount of available prior information, and the desired properties of the posterior distribution.

5 Bayesian Portfolio Optimization

5.1 Bayesian setting and assumptions for portfolio selection

As discussed in Chapter 3, classical mean-variance optimization assumes that the mean vector and covariance matrix of asset returns are known and fixed. In practice, both parameters are unknown and must be estimated from historical data, which introduces estimation risk. The standard approach ignores this aspect, underestimating risk and leading to unstable allocations, especially when the sample size is limited.

The Bayesian framework addresses this limitation by treating the parameters μ and Σ as random variables, rather than as fixed but unknown quantities [8]. Before observing the data, beliefs about μ and Σ are encoded in a prior distribution $p(\mu, \Sigma)$. After observing a sample of returns $\mathbf{R} = \{R_1, \ldots, R_T\}$, Bayes' theorem updates these beliefs. The resulting posterior distribution is given by

$$p(\mu, \Sigma \mid \mathbf{R}) \propto p(\mathbf{R} \mid \mu, \Sigma)\, p(\mu, \Sigma),$$

where $p(\mathbf{R} \mid \mu, \Sigma)$ is the likelihood under the multivariate normal model.

This approach allows the use of the posterior distribution to obtain a posterior predictive distribution of future returns, so that both estimation and portfolio optimization can explicitly take into account estimation risk:

$$p(R_{T+1} \mid \mathbf{R}) = \int p(R_{T+1} \mid \mu, \Sigma)\, p(\mu, \Sigma \mid \mathbf{R})\, d\mu\, d\Sigma,$$

where the predictive distribution averages over all plausible values of the parameters given the observed data.

Incorporating estimation risk in this manner regularizes the optimization, producing more robust allocations, reducing overfitting, thus improving out of sample performance. It should be noted that when the sample size is large, the Bayesian solution naturally converges to the classical one as external data dominate the posterior distribution.

However, with limited or noisy data, just like in our case with 60 observations per estimation window, the Bayesian method is expected to show significant in stability and reliability.

The posterior predictive distribution of future returns provides the necessary inputs for portfolio

optimization under uncertainty. Let $\tilde{\mu}$ and $\tilde{\Sigma}$ denote the predictive mean vector and covariance matrix of next-period returns, that is,

$$\tilde{\mu} = \mathbb{E}[R_{T+1} \mid \mathbf{R}], \qquad \tilde{\Sigma} = \mathrm{Cov}[R_{T+1} \mid \mathbf{R}].$$

Portfolio selection proceeds by substituting these Bayesian predictive moments in place of their classical counterparts. The mean-variance optimization problem is reformulated as

$$\begin{aligned} \underset{\omega}{\text{minimize}} \quad & \omega^\top \tilde{\Sigma} \omega \\ \text{subject to} \quad & \omega^\top \tilde{\mu} \geq \mu^* \\ & \omega^\top \mathbf{1} = 1 \end{aligned} \tag{14}$$

where ω is the vector of portfolio weights and μ^* is the target expected return.

The set of solutions to this problem accounts for the estimation risk, reshaping the efficient frontier to reflect the additional information, as will be shown in the empirical part.

For example, in the case of maximizing the Sharpe ratio (the tangency portfolio), the optimal weights are given by

$$\omega^* = \frac{\tilde{\Sigma}^{-1}\tilde{\mu}}{\mathbf{1}^\top \tilde{\Sigma}^{-1}\tilde{\mu}} \tag{15}$$

which directly matches the classical result but uses Bayesian predictive estimates for both the mean and covariance.

5.2 Diffuse Prior

In the absence of strong prior information about the parameters μ and Σ, it is natural to adopt a diffuse (noninformative) prior. This choice reflects the principle that all acceptable parameter values are a priori equally likely, allowing the data to drive inference. The standard diffuse prior in the multivariate normal setting assumes independence between the mean and the covariance matrix, such that

$$p(\mu, \Sigma) = p(\mu)\, p(\Sigma) \propto |\Sigma|^{-(N+1)/2},$$

which corresponds to the Jeffreys prior for the covariance matrix. Despite the prior being improper, it results in a proper posterior distribution as long as the sample size T exceeds N, which in our case translates to having more observations than assets in the investible universe.

Given observed returns $\mathbf{R} = \{R_1, \ldots, R_T\}$, the normal likelihood (see equation (4)) is:

$$p(\mathbf{R} \mid \mu, \Sigma) = \prod_{t=1}^{T} \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(R_t - \mu)^\top \Sigma^{-1}(R_t - \mu)\right)$$

which can be equivalently written in compact form as

$$p(\mathbf{R} \mid \mu, \Sigma) = (2\pi)^{-NT/2}|\Sigma|^{-T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(R_t - \mu)^\top \Sigma^{-1}(R_t - \mu)\right).$$

Combining prior and likelihood, the joint posterior for (μ, Σ) is proportional to

$$p(\mu, \Sigma \mid \mathbf{R}) \propto |\Sigma|^{-(T+N+1)/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(R_t - \mu)^\top \Sigma^{-1}(R_t - \mu)\right).$$

Conditional on Σ, the posterior for μ is

$$\mu \mid \Sigma, \mathbf{R} \sim \mathcal{N}\left(\hat{\mu}, \frac{1}{T}\Sigma\right), \quad \text{where} \quad \hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} R_t \quad \text{is the sample mean.}$$

while the marginal posterior for Σ is an inverse Wishart distribution

$$\Sigma \mid \mathbf{R} \sim \mathcal{W}^{-1}\left(T - 1, \hat{\Sigma}\right), \quad \text{where} \quad \hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}(R_t - \hat{\mu})(R_t - \hat{\mu})^\top \quad \text{is the sample covariance.}$$

Integrating over the joint posterior, the posterior predictive distribution for R_{T+1} can be shown to be a multivariate Student's t distribution,

$$p(R_{T+1} \mid \mathbf{R}) = \int p(R_{T+1} \mid \mu, \Sigma)\, p(\mu \mid \Sigma, \mathbf{R})\, p(\Sigma \mid \mathbf{R})\, d\mu\, d\Sigma,$$

and thus,

$$R_{T+1} \mid \mathbf{R} \sim t_{T-N}\left(\hat{\mu}, \tilde{\Sigma}\right),$$

where t_{T-N} denotes the multivariate t distribution with $T - N$ degrees of freedom, and

$$\tilde{\mu} = \hat{\mu} \qquad \text{as the location parameter}$$

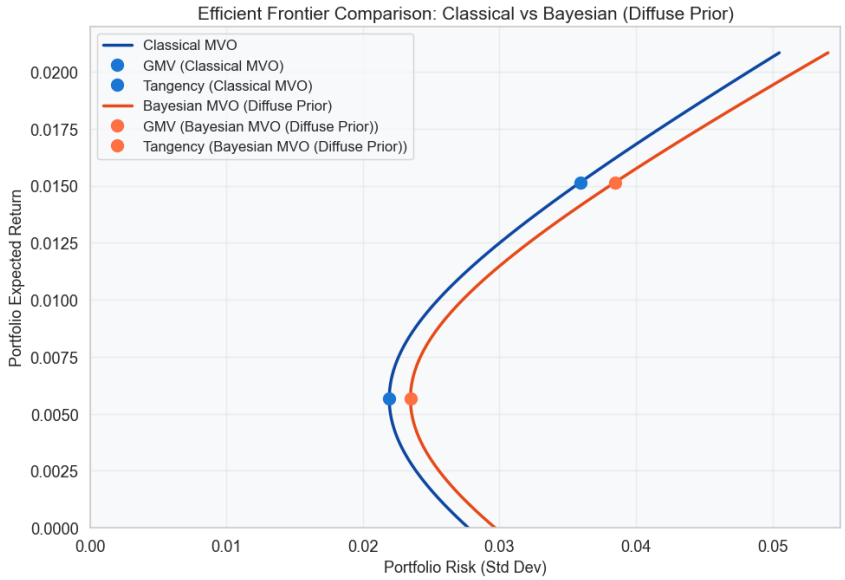
$$\tilde{\Sigma} = \frac{(1 + \frac{1}{T})(T - 1)}{T - N - 2}\hat{\Sigma} \qquad \text{as the scale matrix}$$

As a result of the goal of including estimation risk without introducing meaningful prior information, the Bayesian estimates obtained under the diffuse prior are intentionally close to the classical ones. Specifically, the predictive mean $\tilde{\mu}$ coincides with the sample mean, while the predictive covariance $\tilde{\Sigma}$ is the sample covariance $\hat{\Sigma}$ inflated by a factor

$$\frac{(1 + \frac{1}{T})(T - 1)}{T - N - 2},$$

which explicitly accounts for parameter uncertainty. This adjustment is especially relevant in small samples, where the risk of underestimating portfolio variance is greatest. Notably, as the sample size T increases, this factor approaches one and the Bayesian predictive moments converge to the classical sample estimates, reflecting the fact that data increasingly dominates prior uncertainty in large samples.

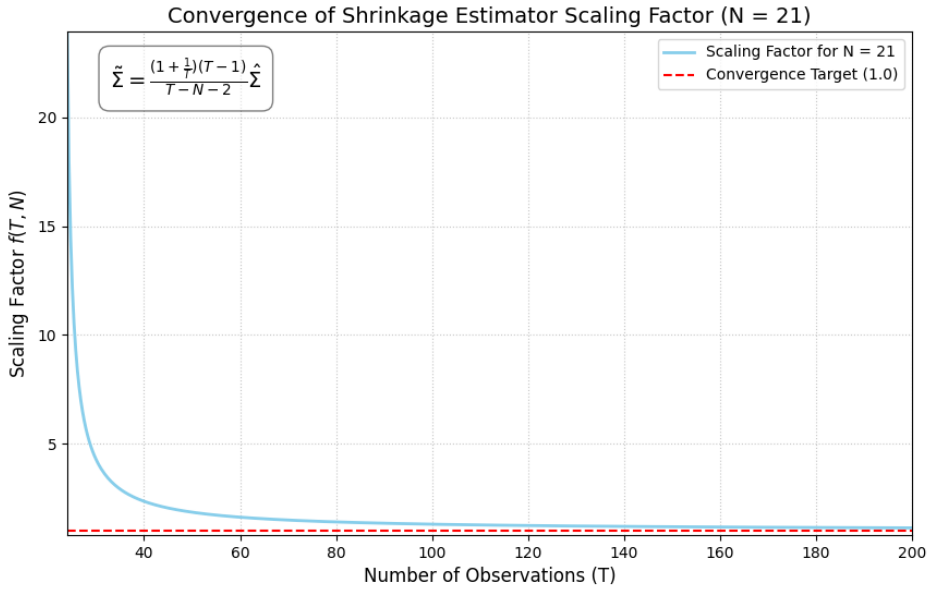
The effect of accounting for uncertainty is evident when observing the efficient frontier, which is shifted towards the right.



This shift arises because the inflated Bayesian predictive covariance matrix $\tilde{\Sigma}$ increases the

expected risk of all portfolios for a fixed expected return. Consequently, every portfolio on the Bayesian frontier exhibits higher volatility than its classical counterpart with the same target return, clearly shifting the maximum Sharpe ratio (tangency) and global minimum variance portfolios as well.

This adjustment is most pronounced when the number of observations, T, is low relative to the quantity of assets, N. As the sample size increases, the Bayesian and classical frontiers converge, since the inflation factor approaches one, as shown below



5.3 Conjugate Prior: Normal-Inverse Wishart

In order to allow for the possibility of introducing meaningful prior information, the prior beliefs can be represented by the conjugate prior for (μ, Σ), the Normal-Inverse Wishart model, defined as:

$$p(\mu, \Sigma) = p(\mu \mid \Sigma)\, p(\Sigma),$$

where

$$\mu \mid \Sigma \sim \mathcal{N}(\mu_0, \frac{1}{\kappa_0}\Sigma), \qquad \Sigma \sim \mathcal{W}^{-1}(\nu_0, S_0),$$

with μ_0 the prior mean vector, κ_0 the strength of belief in μ_0, $\nu_0 > N - 1$ the degrees of freedom controlling the overall influence of the prior on Σ, and S_0 the scale matrix determining the prior's central value for the covariance encoding both the structure and magnitude of prior beliefs regarding Σ.

In this case, the initial beliefs assume prior dependence between the mean and the covariance, reflecting the link between risk and returns.

Given observed returns $\mathbf{R} = \{R_1, \ldots, R_T\}$, Bayes' theorem yields a Normal-Inverse Wishart posterior, as expected because of conjugacy.

The conditional posterior distributions are

$$\mu \mid \Sigma, \mathbf{R} \sim \mathcal{N}\left(\mu_n, \frac{1}{\kappa_n}\Sigma\right), \qquad \Sigma \mid \mathbf{R} \sim \mathcal{W}^{-1}(\nu_n, S_n).$$

with the updated hyperparameters:

$$\kappa_n = \kappa_0 + T, \qquad \mu_n = \frac{\kappa_0}{\kappa_0 + T}\mu_0 + \frac{T}{\kappa_0 + T}\hat{\mu}, \qquad \nu_n = \nu_0 + T,$$

$$S_n = S_0 + (T-1)\hat{\Sigma} + \frac{\kappa_0 T}{\kappa_0 + T}(\hat{\mu} - \mu_0)(\hat{\mu} - \mu_0)^\top,$$

Marginalizing over the posterior Normal Inverse-Wishart, the predictive distribution for a new observation R_{T+1} is a multivariate Student's t, just like in the diffuse case:

$$R_{T+1} \mid \mathbf{R} \sim t_{\nu_n - N + 1}\left(\widetilde{\mu}, \widetilde{\Sigma}\right),$$

where

$$\widetilde{\mu} = \frac{\kappa_0}{\kappa_0 + T}\mu_0 + \frac{T}{\kappa_0 + T}\hat{\mu},$$

$$\widetilde{\Sigma} = \frac{\kappa_0 + T + 1}{(\kappa_0 + T)(\nu_0 + T - N + 1)} \cdot [S_0 + (T-1)\hat{\Sigma} + \frac{\kappa_0 T}{\kappa_0 + T}(\hat{\mu} - \mu_0)(\hat{\mu} - \mu_0)^\top],$$

and $\hat{\mu}$, $\hat{\Sigma}$ are the sample mean and sample covariance of the observed returns.

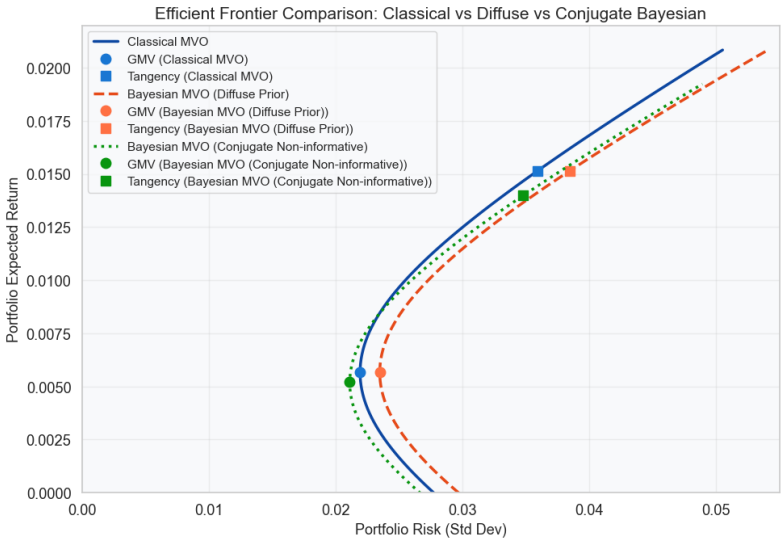
Clearly, $\widetilde{\mu}$ is a weighted average of the prior mean and the sample mean, with the weights depending on κ_0, the prior confidence about the mean. As for $\widetilde{\Sigma}$, it is no longer proportional to the sample covariance, but is influenced by three terms: S_0, $\hat{\Sigma}$, and $(\hat{\mu} - \mu_0)(\hat{\mu} - \mu_0)^\top$, each with its own degree of inflation.

As the amount of available data T increases, both $\widetilde{\mu}$ and $\widetilde{\Sigma}$ converge to their classical sample estimates, as they must:

$$\lim_{T \to \infty} \widetilde{\mu} = \hat{\mu}, \qquad \lim_{T \to \infty} \widetilde{\Sigma} = \hat{\Sigma}$$

If one still wishes to include no prior information, κ_0 should be set close to zero, which ensures that the predictive mean $\widetilde{\mu}$ is almost fully determined by the sample mean $\hat{\mu}$. Similarly, choosing ν_0 just above $N-1$ ensures that the prior is proper and minimizes the effective prior sample size for the covariance matrix, reducing the impact of S_0. However, if the prior mean of Σ is required to exist, a stricter bound $\nu_0 > N+1$ must be imposed, since the expected value of Σ under the Inverse Wishart prior is $\mathbb{E}[\Sigma] = \frac{S_0}{\nu_0 - N - 1}$, which is only defined for $\nu_0 > N+1$. The scale matrix S_0 should be specified as a small multiple of the identity matrix so as to introduce minimal structural information about Σ. In the limit as $\kappa_0 \to 0$ and $\nu_0 \to N$, the Normal-Inverse Wishart prior becomes essentially noninformative, with the efficient frontier converging to the classical one.

To show the effect of weak beliefs on the efficient frontier, the hyperparameters are calibrated based on a pre-sample of historical returns from January 2002 to December 2007. This six-year window precedes the main evaluation period and is used to encode minimal structural information into the prior. The prior mean vector is set to $\mu_0 = 0$, reflecting the belief that historical returns may not be informative about future performance, especially across regimes. The weak degree of confidence in this prior mean is encoded via $\kappa_0 = 15$, which implies that the posterior mean will give a relative weight of $\kappa_0/(\kappa_0 + T) = \frac{1}{13}$ to the prior, effectively shrinking the sample estimates toward zero. For the covariance matrix, the degrees of freedom parameter is fixed at $\nu_0 = N+2$, the minimum value that ensures the existence of the prior expectation $\mathbb{E}[\Sigma]$. The scale matrix is set to $S_0 = \hat{\Sigma}_{\text{pre}} \cdot (\nu_0 - N - 1)$, where $\hat{\Sigma}_{\text{pre}}$ denotes the sample covariance matrix estimated from the 2002–2007 pre-sample, implying $\mathbb{E}[\Sigma] = \hat{\Sigma}_{\text{pre}}$, as covariances and variances tend to be more stable over time (Merton [6]).



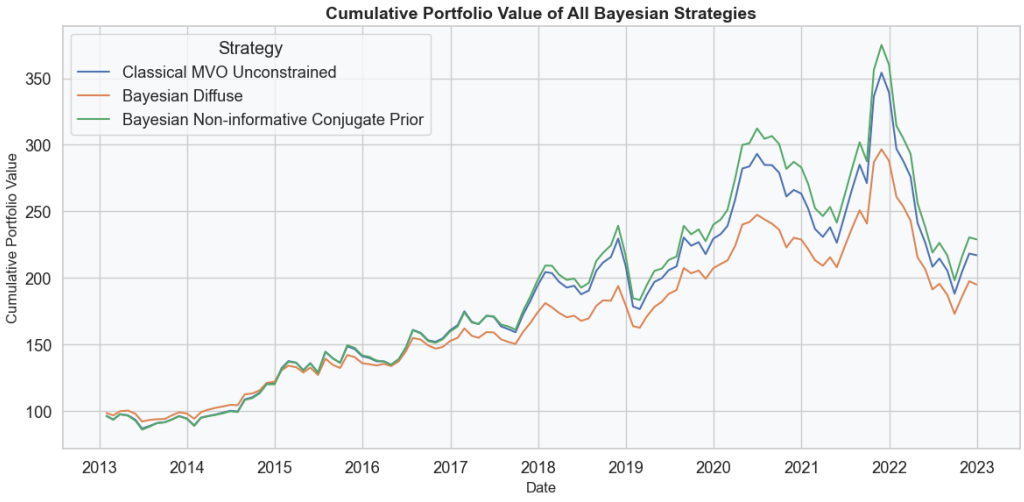
As shown in the figure above, the resulting Bayesian efficient frontier reflects the regularizing effect of the prior, without deviating substantially for the classical frontier, which is coherent with weak beliefs.

5.4 Empirical Results

Following the same rolling-window backtesting procedure employed in previous Chapters, the Bayesian strategies are implemented to sequentially allocate portfolio weights across multiple holding periods. Their performance is then evaluated against the classical unconstrained MVO benchmark, both with and without regularization.

5.4.1 Non-Regularised Strategies

The figure below compares the cumulative portfolio performance of the classical unconstrained mean-variance optimiser with the diffuse and conjugate Bayesian strategies.



As expected when employing non-informative priors, the Bayesian trajectories are highly correlated with the classical one. Although Bayesian inference does not significantly alter portfolio behavior relative to the baseline, it still introduces implicit regularization. This leads to improved stability in the portfolio weights and reduced variability in allocation, particularly under the diffuse prior, which achieves the lowest realized volatility.

Table 6: Volatility (Monthly Standard Deviation)

Strategy	**Volatility**
Unconstrained	0.0554
Bayesian Diffuse	0.0439
Bayesian Non-informative Conjugate Prior	0.0562

Table 7: Portfolio Concentration: Herfindahl-Hirschman Index (HHI)

Strategy	**Average HHI**	**HHI Std Dev**
Unconstrained	1.1266	0.3897
Bayesian Diffuse	0.7591	0.2894
Bayesian Non-informative Conjugate Prior	1.1419	0.3527

The classical unconstrained MVO exhibits significant concentration risk, while the Bayesian diffuse model yields more diversified allocations, as reflected in the lower average HHI and its reduced variability. The conjugate prior specification does not lead to improved diversification, suggesting limited regularization benefits in this setting.

Table 8: Turnover Statistics

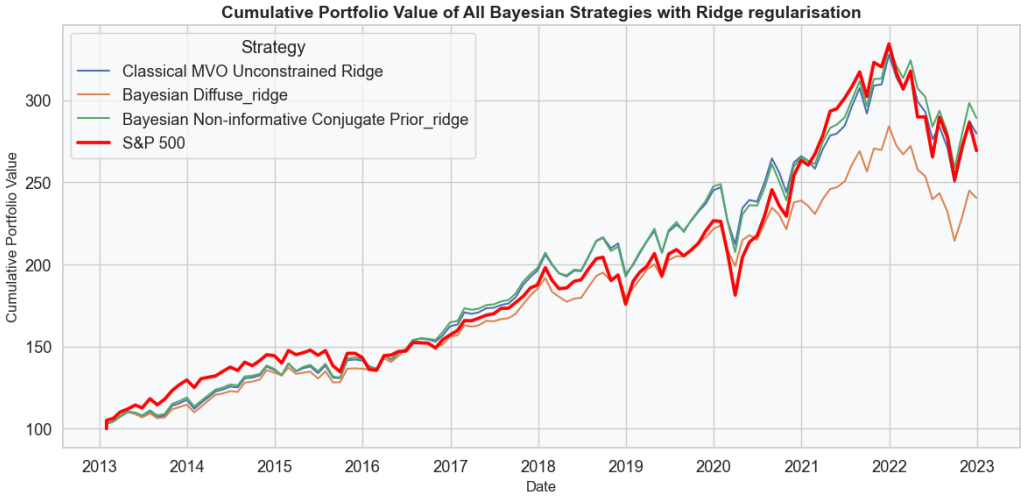
Strategy	**Average Turnover**	**Turnover Std Dev**
Unconstrained	1.2776	0.4703
Bayesian Diffuse	0.9692	0.3680
Bayesian Non-informative Conjugate Prior	1.2654	0.4429

The diffuse prior results in a moderate reduction in turnover, improving the stability of weights. In contrast, the conjugate Bayesian formulation behaves similarly to the classical approach, both in average trading activity and variability. This further supports the suitability of the diffuse specification when prior information is limited or deliberately vague.

5.4.2 Ridge-Regularised Strategies

The figure below illustrates the cumulative performance of the classical and Bayesian strategies under ridge regularization. Compared to the non-regularised case, the benefits of stabilisation

through shrinkage are more pronounced, especially in the Bayesian conjugate case.



The general pattern observed in the unregularised models persists. The diffuse prior introduces more risk aversion through inflated covariance estimates, whereas the conjugate prior remains more data-driven. However, under ridge penalisation, both Bayesian specifications deliver improved out-of-sample stability and risk control. The conjugate prior, in particular, achieves the highest cumulative return.

Ridge regularization significantly reduces portfolio volatility across all strategies. The diffuse Bayesian model reaches the lowest realized volatility as seen in the base case, while the classical and conjugate models exhibit comparable volatility levels.

Table 9: Volatility (Monthly Standard Deviation)

Strategy	**Volatility**
Unconstrained Ridge	0.0361
Bayesian Diffuse_ridge	0.0316
Bayesian Non-informative Conjugate Prior_ridge	0.0378

Table 10: Portfolio Concentration: Herfindahl-Hirschman Index (HHI)

Strategy	**Average HHI**	**HHI Std Dev**
Unconstrained Ridge	0.0748	0.0155
Bayesian Diffuse_ridge	0.0950	0.0350
Bayesian Non-informative Conjugate Prior_ridge	0.0661	0.0095

All strategies exhibit significant reductions in portfolio concentration following the introduction of ridge regularization. Turnover metrics also improve across specifications, with the Bayesian conjugate model demonstrating the lowest average turnover and standard deviation. These results indicate that the conjugate formulation derives the greatest benefit from ridge penalisation, achieving superior stability in portfolio allocations relative to both the diffuse and classical counterparts.

Table 11: Turnover Statistics

Strategy	**Average Turnover**	**Turnover Std Dev**
Unconstrained Ridge	0.1393	0.0542
Bayesian Diffuse_ridge	0.1462	0.0837
Bayesian Non-informative Conjugate Prior_ridge	0.1177	0.0430

6 Economically Informed Model: CAPM

6.1 CAPM Framework and Assumptions

The Capital Asset Pricing Model (CAPM) [10] provides a linear formulation for the cross-asset variation in expected returns. It assumes that the excess return of any risky asset is driven entirely by its exposure to a single systematic factor, the excess return of the market portfolio. Formally, for each asset i at time t, the CAPM assumes the following structure:

$$R_{i,t} = \alpha_i + \beta_i S_t + \varepsilon_{i,t}, \tag{16}$$

where $R_{i,t}$ denotes the excess return of asset i, S_t denotes the excess return of the market benchmark, β_i is the market loading or beta coefficient, α_i is the pricing intercept or mispricing component, and $\varepsilon_{i,t}$ is an idiosyncratic zero-mean disturbance, uncorrelated with the market factor.

Conditional on the model, the expected excess return of asset i is given by

$$\mathbb{E}[R_{i,t}] = \alpha_i + \beta_i \mathbb{E}[S_t]. \tag{17}$$

This CAPM formulation implies that systematic exposure fully determines expected returns if and only if $\alpha_i = 0$ for all i.

In empirical applications, however, nonzero values of α_i are frequently observed. These mispricing terms quantify the extent to which asset prices deviate from those implied by the model and arise due to violations of the CAPM's idealised assumptions. Rather than discarding the model entirely, the Bayesian framework adopts the CAPM as an economically motivated approximation and incorporates model uncertainty through a prior distribution over α_i.

This approach allows for the structured inclusion of external beliefs. For example, one may encode the view that assets are approximately, but not exactly, priced by the market by setting the prior mean of α_i to zero and assigning a nonzero prior variance to reflect uncertainty. In this way, the CAPM provides not only a descriptive model of equilibrium pricing, but also a mechanism for defining economically coherent priors in Bayesian portfolio optimization.

As described in Chapter 2, excess returns $R_{i,t}$ are defined as the difference between monthly

total returns and the one-month U.S. Treasury yield. The benchmark factor S_t corresponds to the excess return of the S&P 500 index, which serves as a proxy for the market portfolio.

The CAPM relies on a set of economic assumptions. On the investor side, agents are assumed to be risk-averse and to select portfolios over a single-period horizon based solely on the mean and variance of returns. On the market side, capital markets are assumed to be perfectly competitive and frictionless, and investors may borrow and lend unlimited amounts at a common risk-free rate.

In addition to the economic assumptions, the empirical application of the CAPM often relies on several technical assumptions regarding the statistical properties of the error term and, consequently, the returns themselves. Specifically, the CAPM assumes that the idiosyncratic disturbances $\varepsilon_{i,t}$, uncorrelated with the benchmark return series, are independently and identically normally distributed with a mean of zero and a constant variance $\sigma^2_{\varepsilon_i}$ across time. Formally, we can write:

$$\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_{\varepsilon_i})$$

Given these assumptions on the error term $\varepsilon_{i,t}$, the excess return of asset i, $R_{i,t}$, is also normally distributed. Specifically,

$$R_{i,t} \mid \alpha_i, \beta_i, \sigma^2_{\varepsilon_i}, S_t \sim \mathcal{N}(\alpha_i + \beta_i S_t, \sigma^2_{\varepsilon_i})$$

Clearly, the parameters $\alpha_i, \beta_i, \sigma^2_{\varepsilon_i}$ and the (future) distribution of returns for the benchmark factor, S_t, are the main subjects of our analysis.

6.2 Bayesian CAPM

The classical (frequentist) Capital Asset Pricing Model (CAPM) approach typically relies on frequentist estimation methods, such as Ordinary Least Squares (OLS) regression, to determine the parameters α_i, β_i, and $\sigma^2_{\varepsilon_i}$. Specifically, for each asset i, a linear regression of the excess return $R_{i,t}$ on the market excess return S_t provides point estimates for α_i and β_i. While these methods offer intuitive point estimates, they do not inherently quantify the full probabilistic uncertainty surrounding these parameter estimates.

Moreover, since the ultimate objective is to optimise future portfolio performance, the classical CAPM inherits the same fundamental shortcoming identified in the classical mean-variance

optimization framework: it fails to incorporate estimation risk. In particular, when forecasting future returns, the expected value of the benchmark factor $\mathbb{E}[S_t]$ is treated as known, despite being estimated from historical data. This results in an understatement of uncertainty, analogous to the estimation risk associated with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ in the classical MVO framework discussed in Chapter 3.

In contrast, the Bayesian framework provides a more comprehensive methodology for estimating the CAPM parameters, α_i, β_i, and $\sigma^2_{\varepsilon_i}$, by treating them as random variables. In addition, to address the benchmark estimation risk limitation, the Bayesian approach models the uncertainty in the distribution of the market factor. As in the treatment of asset returns, we adopt a simplified assumption that benchmark returns follow a normal distribution:

$$S_t \sim \mathcal{N}(\mu_S, \sigma_S^2),$$

where μ_S and σ_S^2 represent the unknown mean and variance of the benchmark excess return, respectively.

Our beliefs about all unknown parameters (α_i, β_i, $\sigma^2_{\varepsilon_i}, \mu_S, \sigma_S^2$) are initially expressed through prior probability distributions, which are then systematically updated and refined based on the observed historical data through Bayes' theorem.

This estimation procedure naturally follows a two-stage structure. First, conditional on the historical benchmark returns S_t, the asset-specific parameters $\alpha_i, \beta_i, \sigma^2_{\varepsilon_i}$ are updated using standard Bayesian linear regression. Second, the parameters governing the benchmark distribution μ_S and σ_S^2 are themselves updated based on the observed time series of S_t. Combining the posterior distributions from both stages yields the posterior predictive distribution of future asset returns $R_{i,t+1}$, which fully incorporates parameter and factor uncertainty.

6.2.1 Likelihood Function

Given the assumption that the idiosyncratic disturbances $\varepsilon_{i,t}$ are independently and identically normally distributed, and treating the market factor S_t as observed from historical data, the likelihood for the excess return of asset i, $R_{i,t}$, is proportional to:

$$L(\mathbf{R}_i \mid \alpha_i, \beta_i, \sigma^2_{\varepsilon_i}, \mathbf{S}) \propto (\sigma^2_{\varepsilon_i})^{-T/2} \exp\left(-\frac{1}{2\sigma^2_{\varepsilon_i}} \sum_{t=1}^{T} \left(R_{i,t} - \alpha_i - \beta_i S_t\right)^2\right)$$

where $\mathbf{R}_i = (R_{i,1}, \ldots, R_{i,T})^\top$ and $\mathbf{S} = (S_1, \ldots, S_T)^\top$.

For the market factor itself, assuming that benchmark returns follow a normal distribution with unknown mean μ_S and variance σ_S^2, the likelihood for the observed series $\mathbf{S} = (S_1, \ldots, S_T)^\top$ is given by:

$$L(\mathbf{S} \mid \mu_S, \sigma_S^2) \propto (\sigma_S^2)^{-T/2} \exp\left(-\frac{1}{2\sigma_S^2} \sum_{t=1}^{T} (S_t - \mu_S)^2\right)$$

6.2.2 Prior Distributions

The Bayesian framework requires the specification of prior distributions for all unknown parameters.

For each asset i, we place a conjugate prior on the regression coefficients (α_i, β_i) and the idiosyncratic variance $\sigma_{\varepsilon_i}^2$. Specifically, we assume:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \Big| \, \sigma_{\varepsilon_i}^2 \sim \mathcal{MVN}(\mathbf{b}_0, \sigma_{\varepsilon_i}^2 \mathbf{\Omega}_0), \tag{18}$$

$$\sigma_{\varepsilon_i}^2 \sim \text{Inv-}\chi^2\left(\nu_0, c_0^2\right). \tag{19}$$

Where prior mean vector and covariance matrix are defined as

$$\mathbf{b}_0 = \begin{pmatrix} a_0 \\ \beta_0 \end{pmatrix}, \qquad \mathbf{\Omega}_0 = \begin{pmatrix} \sigma_\alpha^2 \frac{1}{\sigma_{\varepsilon_i}^2} & 0 \\ 0 & \sigma_\beta^2 \end{pmatrix}.$$

The prior mean a_0 is interpreted as the prior belief about mispricing, which could either be set to 0 (as done in the backtesting), or reflect additional external information. The prior mean β_0 has been set to one, implying an ex-ante average market exposure. The parameter σ_α^2, set to 1.0, determines the degree of confidence in the pricing restriction, with lower values implying stronger belief in the CAPM. The parameter σ_β^2 (1.0) reflects the prior uncertainty about the slope coefficient and is typically set to a large value to encode vague prior information. The inverse-χ^2 prior on $\sigma_{\varepsilon_i}^2$ reflects uncertainty about residual risk, with ν_0 representing the degrees of freedom, and c_0^2 being the scaling parameter, respectively set to 1 and 0.1 reflect weak prior beliefs.

As for the parameters of the normal distribution of the benchmark factor, we assume a Jeffreys'

prior distribution:

$$p(\mu_S, \sigma_S^2) \propto \frac{1}{\sigma_S^2}. \tag{20}$$

This is the standard non-informative reference prior for the mean and variance of a univariate normal distribution, representing the univariate equivalent of the prior distribution used in Chapter 5.2.

6.2.3 Posterior Distributions

Let $X = (\boldsymbol{x}_1^\top, \ldots, \boldsymbol{x}_T^\top) \in \mathbb{R}^{T\times 2}$, where $\boldsymbol{x}_t = (1, S_t)^\top$, and let $\boldsymbol{R}_i = (R_{i,1}, \ldots, R_{i,T})^\top$ be the vector of observed excess returns for asset i. Moreover, let's denote by $\hat{\boldsymbol{b}}_i$ the ordinary least squares estimator of the regression coefficients. Since the classical linear regression model assumes Gaussian errors, this estimator also coincides with the maximum likelihood estimate:

$$\hat{\boldsymbol{b}}_i = (X^\top X)^{-1} X^\top \boldsymbol{R}_i.$$

Given the conjugate priors specified in equations (18)–(19), and the Gaussian likelihood from the CAPM regression model, the posterior distributions for the parameters $\boldsymbol{b}_i = (\alpha_i, \beta_i)^\top$ and $\sigma_{\varepsilon_i}^2$, after applying Bayes' Theorem, retain the same functional forms, with updated parameters.

The conditional posterior distribution of the regression coefficients is multivariate normal:

$$\boldsymbol{b}_i \mid \sigma_{\varepsilon_i}^2, \boldsymbol{R}_i, X \sim \mathcal{MVN}(\boldsymbol{b}_i^*, \sigma_{\varepsilon_i}^2 \Omega_i^*), \tag{21}$$

where

$$\Omega_i^* = \left(\Omega_0^{-1} + X^\top X\right)^{-1}, \tag{22}$$

$$\boldsymbol{b}_i^* = \begin{pmatrix} \alpha_i^* \\ \beta_i^* \end{pmatrix} = \Omega_i^* \left(\Omega_0^{-1} \boldsymbol{b}_0 + X^\top X \hat{\boldsymbol{b}}_i\right). \tag{23}$$

As expected, the posterior mean $\boldsymbol{b}_i^*$ combines the the prior mean $\boldsymbol{b}_0$ and the maximum likelihood estimate $\hat{\boldsymbol{b}}_i$, with relative importance determined by the prior precision Ω_0^{-1} and the sample precision $X^\top X$.

The posterior distribution for the idiosyncratic variance is inverse-χ^2, with updated degrees of

freedom and scale:

$$\sigma_{\varepsilon_i}^2 \mid \boldsymbol{R}_i, X \sim \text{Inv-}\chi^2(\nu_i^*, c_i^{*2}), \tag{24}$$

where

$$\nu_i^* = \nu_0 + T, \tag{25}$$

$$c_i^{*2} = \frac{1}{\nu_i^*}\left[\nu_0 c_0^2 + (\boldsymbol{R}_i - X\hat{\boldsymbol{b}}_i)^\top(\boldsymbol{R}_i - X\hat{\boldsymbol{b}}_i) + (\boldsymbol{b}_0 - \hat{\boldsymbol{b}}_i)^\top K_i(\boldsymbol{b}_0 - \hat{\boldsymbol{b}}_i)\right], \tag{26}$$

and

$$K_i = \left(\Omega_0^{-1} + X^\top X\right)^{-1}.$$

The posterior scale c_i^{*2} aggregates three distinct sources of variation. The first term, $\nu_0 c_0^2$, reflects the prior belief about residual variability. The second term, $(\boldsymbol{R}_i - X\hat{\boldsymbol{b}}_i)^\top(\boldsymbol{R}_i - X\hat{\boldsymbol{b}}_i)$, corresponds to the residual sum of squares from the regression model. The third term, $(\boldsymbol{b}_0 - \hat{\boldsymbol{b}}_i)^\top K_i(\boldsymbol{b}_0 - \hat{\boldsymbol{b}}_i)$, penalizes deviations of the prior mean from the maximum likelihood estimate, representing a form of shrinkage toward prior beliefs. Intuitively, the posterior scale c_i^{*2} reflects the overall alignment between prior information and observed data; higher disagreement increases posterior uncertainty, while greater consistency results in tighter posterior variance.

As for the posterior distributions for the parameters of the benchmark factor μ_S and σ_S^2, let $\bar{S} = \frac{1}{T}\sum_{t=1}^T S_t$ be the sample mean of the benchmark returns, and $SS_S = \sum_{t=1}^T (S_t - \bar{S})^2$ represent the sum of squared deviations from the sample mean.

The marginal posterior distribution of the benchmark variance is an inverse-χ^2 distribution:

$$\sigma_S^2 \mid \mathbf{S} \sim \text{Inv-}\chi^2\left(T-1, \frac{SS_S}{T-1}\right). \tag{27}$$

The conditional posterior distribution of the benchmark mean, given the variance, is a Normal distribution:

$$\mu_S \mid \sigma_S^2, \mathbf{S} \sim \mathcal{N}\left(\bar{S}, \frac{\sigma_S^2}{T}\right). \tag{28}$$

As in the diffuse multivariate case 5.2, the non informative prior leads to a dependent posterior, with the posterior mean of the benchmark factor return depending on the posterior variance.

6.2.4 Predictive Moments of Future Asset Returns

The ultimate objective is to obtain the moments of the posterior predictive distribution for future asset returns, $p(R_{i,t+1}|\mathbf{R}_i, \mathbf{S})$, in order to retrieve the economically informed mean vector and covariance matrix of future predictive returns. In order to do that, we first analyse the predictive distribution of the market benchmark, S_{t+1}. Second, the results will be used to find the moments for the individual asset returns, $R_{i,t+1}$.

To begin, we derive the moments of the predictive distribution for the next period's market excess return, S_{t+1}. Given the Normal-Inverse-χ^2 posterior for the benchmark parameters (μ_S, σ_S^2), the posterior predictive distribution for a new observation S_{t+1} is a Student's t-distribution with $T-1$ degrees of freedom. Therefore the predictive mean and variance of the future benchmark returns are:

$$\mathbb{E}[S_{t+1} \mid \mathbf{S}] = \bar{S}. \tag{29}$$

$$\text{Var}(S_{t+1} \mid \mathbf{S}) = \frac{(T+1)}{T(T-3)} SS_S \tag{30}$$

Having obtained the predictive moments of the market factor , we can now derive the moments for an individual asset's excess return, $R_{i,t+1}$.

To find the future mean of returns, we use Adam's rule to get rid of the conditioning on S_{t+1}:

$$\begin{aligned}
\mathbb{E}[R_{i,t+1} \mid \mathbf{R}_i, \mathbf{S}] &= \mathbb{E}_{S_{t+1}}[\mathbb{E}[R_{i,t+1} \mid S_{t+1}, \mathbf{R}_i, X]] && (31)\\
&= \mathbb{E}_{S_{t+1}}[\alpha_i^* + \beta_i^* S_{t+1}] && (32)\\
&= \alpha_i^* + \beta_i^* \mathbb{E}[S_{t+1} \mid \mathbf{S}] && (33)\\
&= \alpha_i^* + \beta_i^* \bar{S}. && (34)
\end{aligned}$$

As for the predictive variance, it is derived using the law of total variance:

$$\begin{aligned}
\text{Var}(R_{i,t+1} \mid \mathbf{R}_i, \mathbf{S}) &= \mathbb{E}[\text{Var}(R_{i,t+1} \mid S_{t+1}, \alpha_i, \beta_i, \sigma_{\varepsilon_i}^2, \mathbf{R}_i, \mathbf{S})] + \text{Var}[\mathbb{E}(R_{i,t+1} \mid S_{t+1}, \alpha_i, \beta_i, \sigma_{\varepsilon_i}^2, \mathbf{R}_i, \mathbf{S})]\\
&= \mathbb{E}[\sigma_{\varepsilon_i}^2 \mid \mathbf{R}_i, \mathbf{S}] + \text{Var}[\alpha_i + \beta_i S_{t+1} \mid \mathbf{R}_i, \mathbf{S}]
\end{aligned}$$

The first term is the expected value of $\sigma^2_{\varepsilon_i}$ given its Inverse-$\chi^2(\nu_i^*, c_i^{*2})$ posterior distribution:

$$\mathbb{E}[\sigma^2_{\varepsilon_i} \mid \mathbf{R}_i, \mathbf{S}] = \frac{\nu_i^* c_i^{*2}}{\nu_i^* - 2}$$

with $\nu_i^* = \nu_0 + T$.

The second term, $\text{Var}[\alpha_i + \beta_i S_{t+1} \mid \mathbf{R}_i, \mathbf{S}]$, is derived using the law of total variance one more time:

$$\begin{aligned}
\text{Var}[\alpha_i + \beta_i S_{t+1} \mid \mathbf{R}_i, \mathbf{S}] &= \mathbb{E}_{\alpha_i,\beta_i}[\text{Var}(\alpha_i + \beta_i S_{t+1} \mid \alpha_i, \beta_i, \mathbf{S})] + \text{Var}_{\alpha_i,\beta_i}[\mathbb{E}(\alpha_i + \beta_i S_{t+1} \mid \alpha_i, \beta_i, \mathbf{S})] \\
&= \mathbb{E}_{\alpha_i,\beta_i}[\beta_i^2 \text{Var}(S_{t+1} \mid \mathbf{S}) \mid \mathbf{R}_i, \mathbf{S}] + \text{Var}_{\alpha_i,\beta_i}[\alpha_i + \beta_i \bar{S} \mid \mathbf{R}_i, \mathbf{S}] \\
&= \mathbb{E}[\beta_i^2 \mid \mathbf{R}_i, \mathbf{S}] \text{Var}(S_{t+1} \mid \mathbf{S}) + \text{Var}[\alpha_i + \beta_i \bar{S} \mid \mathbf{R}_i, \mathbf{S}]
\end{aligned}$$

We use the following intermediate results:

$$\text{Var}(S_{t+1} \mid \mathbf{S}) = \frac{(T+1)}{T(T-3)} SS_S \qquad \text{Cov}((\alpha_i, \beta_i)^\top \mid \mathbf{R}_i, \mathbf{S}) = \frac{\nu_i^* c_i^{*2}}{\nu_i^* - 2} \Omega_i^*$$

$$\mathbb{E}[\beta_i^2 \mid \mathbf{R}_i, \mathbf{S}] = (\beta_i^*)^2 + \frac{\nu_i^* c_i^{*2}}{\nu_i^* - 2} \Omega^*_{i;\,2,2} \qquad \text{Var}[\alpha_i + \beta_i \bar{S} \mid \mathbf{R}_i, \mathbf{S}] = \tilde{\boldsymbol{x}}^\top \left(\frac{\nu_i^* c_i^{*2}}{\nu_i^* - 2} \Omega_i^* \right) \tilde{\boldsymbol{x}}$$

where $\tilde{\boldsymbol{x}} = (1, \bar{S})^\top$.

Substituting these, we obtain:

$$\text{Var}[\alpha_i + \beta_i S_{t+1} \mid \mathbf{R}_i, \mathbf{S}] = \left((\beta_i^*)^2 + \frac{\nu_i^* c_i^{*2}}{\nu_i^* - 2} \Omega^*_{2,2} \right) \frac{(T+1)}{T(T-3)} SS_S + \tilde{\boldsymbol{x}}^\top \left(\frac{\nu_i^* c_i^{*2}}{\nu_i^* - 2} \Omega_i^* \right) \tilde{\boldsymbol{x}}$$

Combining all terms, the predictive variance of $R_{i,t+1}$ is:

$$\text{Var}(R_{i,t+1} \mid \mathbf{R}_i, \mathbf{S}) = \frac{(T+\nu_0) c_i^{*2}}{(T+\nu_0) - 2} \left[1 + \tilde{\boldsymbol{x}}^\top \Omega_i^* \tilde{\boldsymbol{x}} + \Omega^*_{2,2} \frac{(T+1)}{T(T-3)} SS_S \right] + (\beta_i^*)^2 \frac{(T+1)}{T(T-3)} SS_S \tag{35}$$

This expression is valid for $\nu_0 + T > 2$ and $T > 3$. It decomposes the total predictive variance into three components: expected idiosyncratic variance, variance from uncertainty in the regression coefficients, and variance from uncertainty in the future market factor.

6.2.5 Derivation of Predictive Covariance

To complete the inputs for portfolio optimization, we now derive the future predictive covariance between the excess returns of two distinct assets, $R_{i,t+1}$ and $R_{j,t+1}$. Under the single-factor Bayesian framework, we consider the standard factor model specification for future returns:

$$R_{i,t+1} = \alpha_i + \beta_i S_{t+1} + \varepsilon_{i,t+1} \tag{36}$$

$$R_{j,t+1} = \alpha_j + \beta_j S_{t+1} + \varepsilon_{j,t+1} \tag{37}$$

Here, S_{t+1} represents the future excess return of the market benchmark. A key assumption is that the idiosyncratic disturbances $\varepsilon_{i,t+1}$ and $\varepsilon_{j,t+1}$ are independent of each other (i.e., $\mathrm{Cov}(\varepsilon_{i,t+1}, \varepsilon_{j,t+1}) = 0$ for $i \neq j$) and are also uncorrelated with the market factor S_{t+1}. Furthermore, all model parameters $(\alpha_i, \beta_i, \alpha_j, \beta_j, \sigma^2_{\varepsilon_i}, \sigma^2_{\varepsilon_j}, \mu_S, \sigma^2_S)$ are treated as random variables, reflecting our posterior uncertainty. Importantly, the parameters for asset i (i.e., $\alpha_i, \beta_i, \sigma^2_{\varepsilon_i}$) are independent of the parameters for asset j (i.e., $\alpha_j, \beta_j, \sigma^2_{\varepsilon_j}$) conditional on the observed market factor $\mathbf{S}$.

We apply the Law of Total Covariance to decompose the predictive covariance:

$$\begin{aligned}\mathrm{Cov}(R_{i,t+1}, R_{j,t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}) = \mathbb{E}[\mathrm{Cov}(R_{i,t+1}, R_{j,t+1} \mid S_{t+1}, \boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \mathbf{R}_i, \mathbf{R}_j, \mathbf{S})] \\ + \mathrm{Cov}(\mathbb{E}[R_{i,t+1} \mid S_{t+1}, \boldsymbol{\theta}_i, \mathbf{R}_i, \mathbf{S}], \mathbb{E}[R_{j,t+1} \mid S_{t+1}, \boldsymbol{\theta}_j, \mathbf{R}_j, \mathbf{S}])\end{aligned}$$

where $\boldsymbol{\theta}_k = (\alpha_k, \beta_k, \sigma^2_{\varepsilon_k})$.

The first term, the expected conditional covariance, simplifies due to the assumed independence of the idiosyncratic disturbances. Conditional on the future market factor S_{t+1} and the parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_j$, the only remaining uncertainty in $R_{i,t+1}$ and $R_{j,t+1}$ comes from their respective idiosyncratic errors. Since $\varepsilon_{i,t+1}$ and $\varepsilon_{j,t+1}$ are uncorrelated:

$$\begin{aligned}&\mathrm{Cov}(R_{i,t+1}, R_{j,t+1} \mid S_{t+1}, \boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}) \\ &\quad = \mathrm{Cov}(\alpha_i + \beta_i S_{t+1} + \varepsilon_{i,t+1}, \alpha_j + \beta_j S_{t+1} + \varepsilon_{j,t+1} \mid S_{t+1}, \boldsymbol{\theta}_i, \boldsymbol{\theta}_j, \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}) = 0\end{aligned}$$

For the second term, the covariance of conditional expectations, we first evaluate the conditional

expected returns. These expectations are conditional on S_{t+1} and the parameters of each asset:

$$\mathbb{E}[R_{k,t+1} \mid S_{t+1}, \alpha_k, \beta_k, \mathbf{R}_k, \mathbf{S}] = \alpha_k + \beta_k S_{t+1}$$

Now, we need to compute the covariance of these conditional expectations:

$$\begin{aligned}
&\text{Cov}(\alpha_i + \beta_i S_{t+1}, \alpha_j + \beta_j S_{t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}) \\
&= \mathbb{E}[(\alpha_i + \beta_i S_{t+1})(\alpha_j + \beta_j S_{t+1}) \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}] \\
&\quad - \mathbb{E}[\alpha_i + \beta_i S_{t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}]\mathbb{E}[\alpha_j + \beta_j S_{t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}]
\end{aligned}$$

Now, for the first term, we recall that the parameters (α_i, β_i) are independent of (α_j, β_j) conditional on $\mathbf{S}$. Therefore, we can decompose expectations of products of parameters from different assets:

$$\begin{aligned}
&\mathbb{E}[(\alpha_i + \beta_i S_{t+1})(\alpha_j + \beta_j S_{t+1}) \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}] \\
&= \mathbb{E}[\alpha_i\alpha_j + \alpha_i\beta_j S_{t+1} + \beta_i\alpha_j S_{t+1} + \beta_i\beta_j S_{t+1}^2 \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}] \\
&= \alpha_i^*\alpha_j^* + \alpha_i^*\beta_j^*\bar{S} + \beta_i^*\alpha_j^*\bar{S} + \beta_i^*\beta_j^*\text{Var}(S_{t+1} \mid \mathbf{S}) + \beta_i^*\beta_j^*\bar{S}^2
\end{aligned}$$

Moreover, having already derived the posterior and predictive distributions, the second term is:

$$\mathbb{E}[\alpha_i + \beta_i S_{t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}]\mathbb{E}[\alpha_j + \beta_j S_{t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}] = (\alpha_i^* + \beta_i^*\bar{S})(\alpha_j^* + \beta_j^*\bar{S})$$

Now, combining the two terms, we obtain:

$$\begin{aligned}
&\text{Cov}(R_{i,t+1}, R_{j,t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S}) \\
&= \left(\alpha_i^*\alpha_j^* + \alpha_i^*\beta_j^*\bar{S} + \beta_i^*\alpha_j^*\bar{S} + \beta_i^*\beta_j^*\text{Var}(S_{t+1} \mid \mathbf{S}) + \beta_i^*\beta_j^*\bar{S}^2\right) \\
&\quad - (\alpha_i^* + \beta_i^*\bar{S})(\alpha_j^* + \beta_j^*\bar{S}) \\
&= \left(\alpha_i^*\alpha_j^* + \alpha_i^*\beta_j^*\bar{S} + \beta_i^*\alpha_j^*\bar{S} + \beta_i^*\beta_j^*\text{Var}(S_{t+1} \mid \mathbf{S}) + \beta_i^*\beta_j^*\bar{S}^2\right) \\
&\quad - \left(\alpha_i^*\alpha_j^* + \alpha_i^*\beta_j^*\bar{S} + \beta_i^*\alpha_j^*\bar{S} + \beta_i^*\beta_j^*\bar{S}^2\right) \\
&= \beta_i^*\beta_j^*\text{Var}(S_{t+1} \mid \mathbf{S}) = \beta_i^*\beta_j^*\frac{(T+1)}{T(T-3)}SS_S
\end{aligned}$$

Hence, the predictive covariance between any two assets is driven by their posterior mean

sensitivities to the common market factor (β_i^* and β_j^*) and the predictive variance of that benchmark factor $\text{Var}(S_{t+1} \mid \mathbf{S})$.

6.3 Using Posterior CAPM for Portfolio Optimization

Given the derived predictive mean vector and covariance for all assets, we can express the generic entries for the mean and covariance as follows:

$$\mu_i = \mathbb{E}[R_{i,t+1} \mid \mathbf{R}_i, \mathbf{S}] \qquad \Sigma_{i,j} = \text{Cov}(R_{i,t+1}, R_{j,t+1} \mid \mathbf{R}_i, \mathbf{R}_j, \mathbf{S})$$

These inputs can then be used in portfolio optimization, either by directly plugging them into the optimization problem as classical estimates 3 or by using them as prior mean and variance in the Bayesian models 5.

6.3.1 Naive MVO with Bayesian CAPM

In this implementation, the predictive mean vector and covariance matrix derived from the Bayesian CAPM are used directly as inputs to the classical unconstrained mean-variance optimization problem.

Coherently with the backtesting approach, at each rebalancing date, the CAPM moments are estimated using a rolling window of sixty months of historical excess returns. These estimates are then used to compute the optimal portfolio weights, which are held constant over a two-month investment horizon. After each holding period, the model is re-estimated using the updated data.

The expected improvement lies in the more accurate estimation of predictive moments, which are now informed by the CAPM structure and explicitly account for estimation risk in the benchmark factor. These adjustments are expected to enhance out-of-sample performance, improve weight stability, and reduce realized portfolio volatility.

6.3.2 Bayesian Conjugate CAPM Model

In the Bayesian Conjugate CAPM model, the predictive moments derived from the Bayesian CAPM framework are used to construct an informative prior for the Normal-Inverse Wishart model 5.3. At each rebalancing point, the CAPM is estimated using a 60-month rolling window of excess returns. The resulting predictive mean vector and covariance matrix are then embedded

into the conjugate prior structure.

Specifically, the prior mean vector $\boldsymbol{\mu}_0$ is set equal to the CAPM-implied predictive mean, while the prior scale matrix $\mathbf{S}_0$ is taken as the CAPM-implied predictive covariance scaled by $\nu_0 - N - 1$:

$$\mathbf{S}_0 = (\nu_0 - N - 1)\,\hat{\boldsymbol{\Sigma}}_{\text{CAPM}}.$$

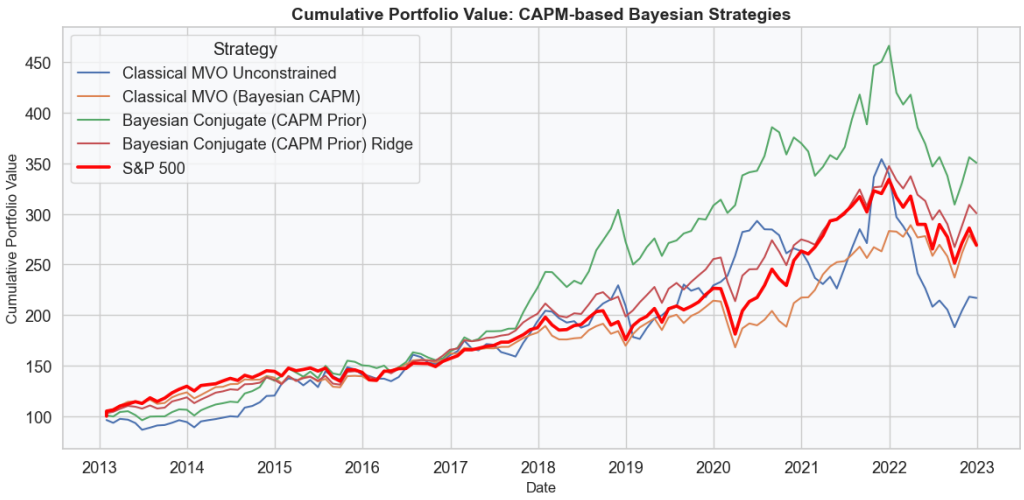
The prior degrees of freedom ν_0 and the mean confidence parameter κ_0 regulate the informativeness of the prior. In the implementation, we set $\nu_0 = N + 10$, assigning moderate confidence to the CAPM-implied covariance, and $\kappa_0 = 30$, implying that one-third ($\frac{30}{30+60}$) of the weight in the posterior mean is attributed to the CAPM-implied mean, and two-thirds to the sample data:

$$\mathbb{E}[\boldsymbol{\mu}_{\text{post}}] = \frac{\kappa_0}{\kappa_0 + T}\,\boldsymbol{\mu}_0 + \frac{T}{\kappa_0 + T}\,\hat{\mu}.$$

Under this approach, the resulting posterior predictive moments used in the portfolio optimization problem integrate three sources of information: the structure imposed by the CAPM, the cross-sectional and time-series information contained in the sample data, and the incorporation of estimation risk through the Bayesian framework.

6.3.3 Empirical Results

The cumulative performance graph below 6.3.3 shows that the Bayesian Conjugate (CAPM Prior) strategy achieves the highest final portfolio value over the investment period. The regularized Bayesian Conjugate (CAPM Prior) Ridge strategy is the second best performer among the tested models, also significantly outperforming the S&P 500 benchmark and other strategies. The naive CAPM implementation offers a substantial improvement over the classical mean-variance approach, aligning with the performance of the SP500 benchmark.



The introduction of CAPM informed priors substantially improves return performance. The Bayesian Conjugate CAPM strategy yields the highest average return (15.01% annualized), significantly surpassing all other models. The ridge regularized version also outperforms the market index, delivering a strong 11.36% annualized return.

Table 12: Mean Monthly and Annualized Returns

Strategy	Mean Monthly Return	Annualized Return
Classical MVO (Unconstrained)	0.0074	9.24%
Naive MVO (Bayesian CAPM)	0.0085	10.65%
Bayesian Conjugate CAPM	0.0110	13.99%
Bayesian Conjugate CAPM Ridge	0.0094	11.86%
S&P 500	0.0060	7.42%

Other than achieving high returns, all CAPM informed strategies significantly reduce volatility with respect to both the SP500 benchmark and the classical MVO. The Bayesian Conjugate CAPM Ridge specification achieves the lowest volatility (0.1344 annualized).

Table 13: Monthly and Annualized Volatility

Strategy	Volatility (Monthly Std)	Annualized Volatility
Classical MVO (Unconstrained)	0.0554	0.1919
Naive MVO (Bayesian CAPM)	0.0401	0.1389
Bayesian Conjugate CAPM	0.0460	0.1593
Bayesian Conjugate CAPM Ridge	0.0388	0.1344
S&P 500	0.0472	0.1635

Consequently, the highest risk adjusted performances are achieved by the Bayesian Conjugate CAPM Ridge strategy, with an Annualized Sharpe Ratio of 0.8374, and the unregularized Bayesian Conjugate CAPM, 0.8266. This is a direct result of their ability to deliver strong returns with the lowest volatility. On the other hand, the Classical MVO and the S&P 500 lag considerably, as evident in the table below.

Table 14: Annualized Sharpe Ratio

Strategy	Annualized Sharpe Ratio
Classical MVO (Unconstrained)	0.4627
Naive MVO (Bayesian CAPM)	0.7318
Bayesian Conjugate CAPM	0.8266
Bayesian Conjugate CAPM Ridge	0.8374
S&P 500	0.4391

All CAPM-informed strategies achieve lower turnover compared to the classical approach. The Naive MVO (Bayesian CAPM) strategy stands out with the lowest turnover. While the conjugate models benefit from strong prior information, their turnover is still influenced by the noisy contribution of sample estimates. Nevertheless, the introduction of ridge regularization significantly improves portfolio stability, reducing the turnover of the Bayesian Conjugate CAPM strategy from 0.5858 to 0.1478. In contrast, the unregularized Classical MVO exhibits very high turnover, signaling unstable portfolio weights.

Table 15: Average Turnover and Herfindahl-Hirschman Index (HHI)

Strategy	**Average Turnover**	**Average HHI**
Classical MVO (Unconstrained)	1.2776	1.1266
Naive MVO (Bayesian CAPM)	0.0585	0.0690
Bayesian Conjugate CAPM	0.5858	0.3658
Bayesian Conjugate CAPM Ridge	0.1478	0.0730

Therefore, combining Bayesian inference with CAPM informed priors leads to superior performance. While the standard Bayesian Conjugate model yields the highest absolute returns, it does so with higher volatility and turnover than its regularized and naive counterparts. The addition of shrinkage regularization via the ridge model creates a very robust strategy, delivering the best risk adjusted returns (Sharpe Ratio) and the lowest volatility, and representing the best strategy implemented so far.

7 Conclusion

7.1 Summary of Results and Main Findings

This thesis has implemented a series variations of portfolio optimization models of increasing structural complexity, beginning with the classical mean-variance optimization (MVO) framework up to Bayesian formulations informed by economic theory. Each model was designed to address shortcomings of the classical MVO, including its sensitivity to estimation error and lack of flexibility in incorporating uncertainty or external information.

The empirical results, summarized in Figure 3, support the thesis that Bayesian methods provide substantial improvements over classical optimization. Even under weakly informative priors, Bayesian models achieve more stable allocations and improved out-of-sample risk characteristics. However, the most significant gains are observed when external structure is introduced via the Capital Asset Pricing Model (CAPM).

Among all strategies, the best-performing in terms of cumulative return is the Bayesian Conjugate CAPM model, which integrates prior beliefs derived from CAPM-implied expected returns and covariances with sample data. Moreover, the strategy that achieves the highest risk-adjusted performance, as measured by the Sharpe ratio, is the Bayesian Conjugate CAPM with ridge regularization, which combines economic information, empirical observations, and shrinkage regularization to control overfitting and improve robustness.

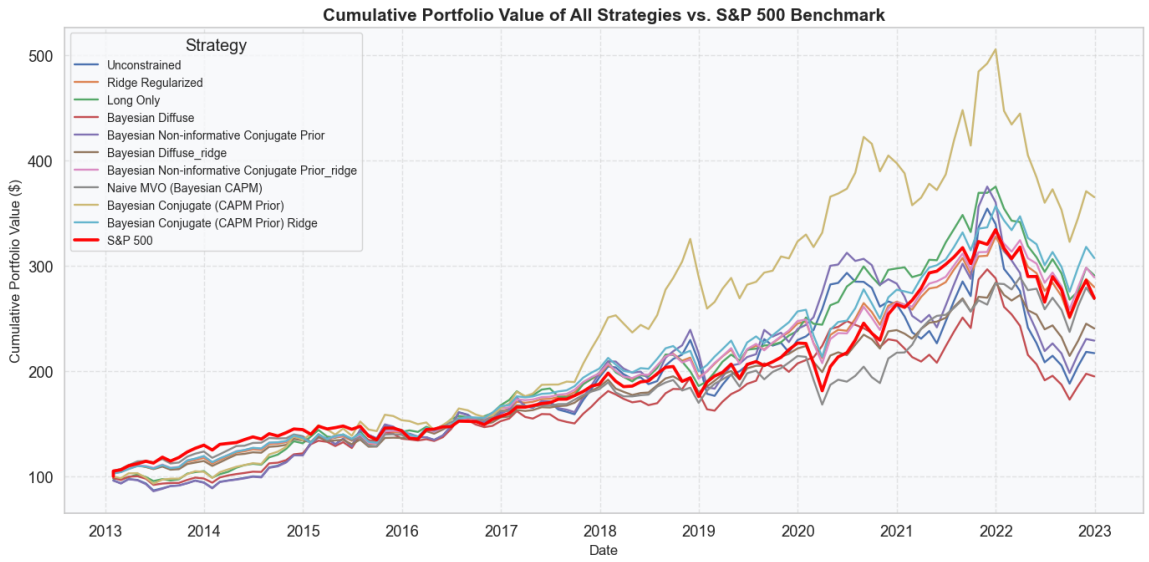


Figure 3: Cumulative portfolio value of all strategies compared to the S&P 500 benchmark.

7.2 Limitations

Despite the results, the derivation and implementation of the models required several simplifications compared to a more realistic setting.

First, the asset universe is fixed ex-ante as of January 2008. While this design avoids survivorship and look-ahead bias, it limits the generalizability of the findings to broader or dynamic investment spaces. Second, the Bayesian CAPM relies on a single-factor model, potentially neglecting relevant sources of systematic risk captured by multifactor models such as the Fama–French three-factor or five-factor specifications, which could lead to mispricing and, consequently, estimation error. Furthermore, transaction costs and other implementation frictions are not considered, potentially overstating cumulative returns, in particular in high turnover strategies. Lastly, performance is evaluated solely through backtesting on historical data, which remains a retrospective assessment and cannot guarantee generalization to future market conditions.

7.3 Directions for Further Research

Given the simplifications introduced in the work, further research could be aimed at relaxing some of the assumptions and extending the approach along multiple dimensions.

As already mentioned, while the turnover metric serves as a proxy for measuring rebalancing frictions, future research could incorporate explicit transaction costs in the backtesting procedure, allowing for a more realistic assessment of net returns. Moreover, the assumption of normally distributed returns, despite being broadly justifiable at the monthly frequency, may underestimate the likelihood of extreme events. Extending the models to allow for heavier-tailed distributions, as well as skewness, would allow for more accurate risk estimation [2]. Finally, the optimization throughout the thesis relies solely on the first two moments of the posterior predictive distribution. As a consequence, all deviations from the mean, regardless of the direction, are penalized equally, despite the fact that investors are more concerned with downside risk than with upside variability. A promising direction for future research would involve formulating portfolio allocation problems that exploit the full predictive distribution of future asset returns. This would allow the introduction of downside-sensitive performance measures such as Conditional Value-at-Risk (CVaR), which focuses explicitly on the tail of the loss distribution, making it more aligned with the risk preferences of investors facing asymmetric or heavy-tailed return scenarios [9].

References

[1] Vijay K Chopra and William T Ziemba. The effect of errors in means, variances, and covariances on optimal portfolio choice. *Journal of Portfolio Management*, 19(2):6–11, 1993.

[2] Campbell R. Harvey and Akhtar Siddique. Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis*, 34(4):465–487, 1999.

[3] Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

[4] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[5] Harry M. Markowitz. *Foundations of Portfolio Theory*. McGraw-Hill, 1991.

[6] Robert C Merton. On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361, 1980.

[7] Richard O. Michaud. The markowitz optimization enigma: Is 'optimized' optimal? *Financial Analysts Journal*, 45(1):31–42, 1989.

[8] Svetlozar T Rachev, John S J Hsu, Biliana S Bagasheva, and Frank J Fabozzi. *Bayesian Methods in Finance*. John Wiley & Sons, 2008.

[9] R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–42, 2000.

[10] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

[11] Wharton Research Data Services. Wharton Research Data Services (WRDS). `https://wrds-www.wharton.upenn.edu/`, 2023. University of Pennsylvania. Available at: `https://wrds-www.wharton.upenn.edu/`.