Title
Bayesian Analysis for Asset Allocation with Investor's Views Considered

Author
Andrei, Mihnea Stefan

Publication Date
2019

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

Bayesian Analysis for Asset Allocation with Investor's Views Considered

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Statistics and Applied Probability

by

Mihnea S. Andrei

Committee in charge:
John S.J. Hsu, Committee Chair
Jean-Pierre Fouque
Yuedong Wang

June, 2019

The dissertation of Mihnea S. Andrei is approved.

Jean-Pierre Fouque

Yuedong Wang

John S.J. Hsu, Committee Chair

June, 2019

Dedication

This Doctorate Dissertation is dedicated to my adviser (John. S.J. Hsu) and the other 2 members of my committee (Jean-Pierre Fouque and Yuedong Wang). Last, but not least, it is dedicated to my parents (Elena Andrei and Marius Andrei), who made sure I always went to the best schools and to as many Mathematics contests as possible as I was growing up in Romania.

Acknowledgments

For this Doctorate Dissertation, I would like to thank:

- My adviser for 3 years, professor John S.J. Hsu, without who's guidance, this research would have not been possible.

- My other 2 committee members (professors Jean-Pierre Fouque and Yuedong Wang) for the support and for the suggestion to apply the models to the whole market ($S\&P500$) also.

- The Center for Scientific Computing supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC) at UC Santa Barbara through NSF DMR 1720256 and NSF CNS 1725797: CNSI link. Without those clusters, the sensitivity analyses in this dissertation would have not been possible.

Vita of Mihnea Stefan **Andrei**

June 2019

(+1) 774-329-5544 | mandrei@ucsb.edu | mandrei3.wixsite.com/home | www.linkedin.com/in/mihnea-andrei-2a798980

Research Interests

Multivariate Bayesian Statistics (especially applied to Finance), High-Dimensional Tensors in Bayesian Statistics, High-Dimensional Covarince Matrix Estimation, Bayesian Statistics in Astronomy (especially the Drake equation), Computational Statistics, Default of Inter-banking Networks.

Education

University of California, Santa Barbara (UCSB) *Santa Barbara, CA, USA*

Ph.D. in Statistics-GPA $3.72/4.00$ *Aug. 2014 - Exp. June 2019*

- UCSB Scholarship, Fellowship and TAship.
- Ph.D. thesis: *Bayesian Alternatives to Black-Litterman*.
- Ph.D. committee: John S.J. Hsu (Ph.D. adviser), Jean-Pierre Fouque, Yuedong Wang.

Worcester Polytechnic Institute (WPI) *Worcester, MA, USA*

M.Sc. in Financial Mathematics-BS/MS GPA $3.91/4.00$ *Aug. 2012 - May 2014*

- Started the Master's program during junior year, while finishing Bachelor's requirements.
- Master's thesis: *A Discrete Model for the Default Risk of Inter-banking Networks*, advised by Stephan Sturm.

Worcester Polytechnic Institute (WPI) *Worcester, MA, USA*

B.S. with major in Actuarial Mathematics and minor in Business-BS/MS GPA $3.91/4.00$ *Aug. 2010 - May 2014*

- Charles O. Thompson Scholar in the class of 2014 for outstanding performance by first-year students.
- Olson Award in 2012 for outstanding performance in introductory mathematics courses.

Papers and Research

Bayesian Factor Model Alternatives to the Black-Litterman Model *UCSB*

Paper-work in progress: Personal website link *Sep. 2018 - PRESENT*

Bayesian Alternatives to the Black-Litterman Model *UCSB*

Paper: ARXIV link, Personal website link *Jul. 2016 - Aug. 2018*

A Discrete Model for the Default Risk of Inter-banking Networks *WPI*

Master's Thesis: URL link *Jun. 2013 - May 2014*

Former SATMAP.inc, current Afiniti *WPI*

Researcher *Aug. 2013 - May 2014*

- Worked on a model for call-agent matching in a company's calling center that would maximize revenue.

Teaching and Mentoring Experience

University of California, Santa Barbara *Santa Barbara, CA, USA*

Teaching Assistant *Aug. 2015 - PRESENT*

- Leading discussion sections for undergraduate classes: Stochastic Processes (PSTAT 160A and 160B), Probability and Statistical Theory (PSTAT 120A and 120B), SAS Base Programming (PSTAT 130), Introduction to Statistics (PSTAT 5A), Statistics for Economics (PSTAT 109).

University of California, Santa Barbara-Summer Sessions *Santa Barbara, CA, USA*

Teaching Associate *Jun.-Aug. 2018*

- Instructor for an introductory statistics class offered by UCSB's Statistics and Applied Probability Department:PSTAT 5A.
- Lead lectures and created all the course lecture notes, exams, labs and homework assignments.

UCSB Research Mentorship Program *Santa Barbara, CA, USA*

RESEARCH MENTOR *Jun.-Aug. 2016 and 2017*

- Aneesh Nathani, Leonardo Glikbarg-*Predicting Terrorist Attacks in Afghanistan*. Taught them data visualization, fitting and diagnostics techniques for GLM and ARIMA.
- Aayush Patel-Used data mining techniques (neural networks, conditional forests, etc.) for predicting crimes in US states.
- Pranav Ahluwalia, Casey Blout-*Predicting Stock Prices with ARIMA Time Series Models*. Taught them fitting and diagnostics techniques for ARIMA.

Worcester Polytechnic Institute *Worcester, MA, USA*

PEER LEARNING ASSISTANT *Aug. 2011 - May 2014*

- Undergraduate classes: Matrices & Linear Algebra I (almost always-MA 2071), Ordinary Differential Equations (one time-MA 2051), Calculus IV (one time-MA 1024).
- Lead discussion sections, graded homework, quizzes, labs and exams.

Projects

Winter 2017	**Classical Trading Strategy Fails even when using Non-Paramteric Regression**, conducted in a Non-Parametric Regression class (PSTAT 226). Personal website link	*UCSB*
Winter 2016	**Generalized Linear Model for Body Mass Index**, conducted in an Advanced Statistical Methods B class (PSTAT 220B) - Generalized Linear Models class. Personal website link	*UCSB*
Fall 2016	**Survival Models for the Population of Halibut**, conducted in a Survival Analysis class (PSTAT 275).Personal website link	*UCSB*
Winter 2015	**ARIMA and GARCH for the Stock Market**, conducted in a Time Series class (PSTAT 274).Personal website link	*UCSB*
Fall 2013	**Risk Exposure for Portfolio of Stocks**, conducted in a Market and Credit Risk Models class (MA 575).Personal website link	*WPI*
2012-2013	**Automating the Precision Trading System (Fully Automated Trading System for TradeStation)**, conducted during an Inter-Qualifying Project (IQP). URL link	*WPI*

Conference Talk

Joint Mathematics Meetings *Baltimore, MD, USA*

PRESENTER 17^{th} *Jan. 2014*

- Presented *A Discrete Model for the Default Risk of Inter-banking Networks*.

Skills

Programming R, Python, MATLAB, C/C++, LaTeX, EasyLanguage (TradeStation).
Web PHP/MySQL, CSS.
Languages English (proficient), Romanian (native).

Graduate Courses

Fall 2018	**Machine Learning: Special Topics**, which covers PAC-Learnability, Rademacher complexity, Vapnik–Chervonenkis dimension, high-dimensional Probability and Statistics, optimization theory and practice, support vector machines, kernel methods, neural network methods.	*UCSB*
Spring 2018	**Covariance Matrix Estimation Seminar**, which covers techniques in obtaining well-conditioned estimators for high-dimensional covariance matrices and not only: shrinkage estimation, regularizing eigenstructure, graphical models, non-gaussian models such as exponential family high-dimensional PCA and copulas, Bayesian covariance matrix estimation, covariance regression and spatial Statistics.	*UCSB*
2016-2017	**Probability Theory A, B&C**, which covers generating functions, discrete and continuous time Markov chains, random walks, branching processes, birth-death processes, Poisson processes, point processes, different types of convergence for random variables; characteristic functions, continuity theorem, laws of large numbers, central limit theorem, large deviations, infinitely divisible and stable distributions, uniform integrability, martingales, martingale convergence, stopping times, optional sampling, optional stopping theorems and applications, maximal inequalities, Brownian motion, introduction to diffusions.	*UCSB*

Graduate Courses Cont.

Winter 2017	**Non-Parametric Regression**, which covers an introduction to some statistical regression and classification techniques including kernel smoothing, smoothing spline, local regression.	*UCSB*
Fall 2016	**Survival Analysis**, which covers basic concepts: survival functions, hazard functions, cumulative hazard functions, and censoring types; Kaplan-Meier and Nelson-Fleming-Harrington estimates; log-rank tests; exponential and Weibull models; Cox proportional hazards and accelerated failure time regression models; current software and applications.	*UCSB*
Fall 2016	**Matrix Analysis and Computation**, which covers graduate level matrix theory with introduction to matrix computations; SVD's, pseudoinverses, variational characterization of eigenvalues, perturbation theory, direct and iterative methods for matrix computations.	*UCSB*
Spring 2016	**Data Mining**, which covers data exploration, classification and regression trees, random forests, clustering and association rules; building predictive models focusing on model selection, model comparison and performance evaluation.	*UCSB*
Spring 2016	**Numerical Solution of Partial Differential Equations-Finite Difference Methods**, which covers finite difference methods for hyperbolic, parabolic and elliptic PDE's, with application to problems in science and engineering; convergence, consistency, order and stability of finite difference methods; dissipation and dispersion; finite volume methods; software design and adaptivity.	*UCSB*
Winter 2016	**Statistical Decision Theory**, which covers statistical inference including estimation, testing and multiple decision rules in decision theoretic framework, relationship to game theory, admissibility, optimality including Bayes and minimax rules, empirical and hierarchical Bayes, invariant decisions.	*UCSB*
Winter 2016	**Numerical Simulation**, which covers linear multistep methods and Runge-Kutta methods for ordinary differential equations: stability, order and convergence; stiffness; differential algebraic equations; numerical solution of boundary value problems.	*UCSB*
2015-2016	**Advance Statistical Methods A, B&C**, which covers Linear Regression and Generalized Linear Models (graphical methods, estimation and inference, diagnostics and model selection), Multivariate Analysis and hypothesis testing, Factor Analysis, Principal Component Analysis.	*UCSB*
Fall 2015	**Bayesian Inference**, which covers likelihood principle, the discrete version of Bayes theorem, prior and posterior distributions, Bayesian point and interval estimations, and predictions; Bayesian computational methods such as Laplacian approximations and Markov Chain Monte Carlo simulation.	*UCSB*
Winter 2015	**Time Series**, which covers stationary and non-stationary models, seasonal time series, ARMA models: calculation of ACF, PACF, mean and ACF estimation; Barlett's formula, model estimation: Yule-Walker estimates, ML method; identification techniques, diagnostic checking, forecasting, spectral analysis; current software and applications.	*UCSB*
Winter 2015	**Computational Techniques in Statistics**, which explores computationally-intensive methods in Statistics such as combinatorial optimization, EM optimization, Newton-Raphson, Monte Carlo simulation, Markov Chain Monte Carlo methods and bootstrapping.	*UCSB*
Spring 2013	**Market & Credit Risk Models and Management**, which covers metrics of risk such as volatility, value-at-risk and expected shortfall; fundamental quantitative techniques used in financial risk evaluation and management; volatility modeling, time series, non-normal heavy tailed phenomena and multivariate notions of co-dependence such as copulas, correlations and tail-dependence; valuation of default contingent claims including credit default swaps, structured credit portfolios and CDOs.	*WPI*
Fall 2012	**Computational Methods of Financial Mathematics**, which covers explicit, implicit and Crank-Nicholson finite difference schemes for fixed and free boundary value problems, their convergence and stability; Monte Carlo simulation methods, including random number generation, variance reduction techniques and the use of low discrepancy sequences.	*WPI*
Spring 2012	**Financial Mathematics I**, which covers stochastic calculus, securities markets, arbitrage-based pricing of options and their uses for hedging and risk management, forward and futures contracts, European options, American options, exotic options, binomial stock price models, the Black-Scholes-Merton partial differential equation, risk-neutral option pricing, the fundamental theorems of asset pricing, sensitivity measures (the Greeks) and Merton's credit risk model.	*WPI*

Abstract

The Black-Litterman model combines the market equilibrium with the investor's personal views and gives optimal portfolio weights. In this paper we will review the original Black-Litterman model (**Section 1**), we will modify the model such that it fits in a Bayesian framework by considering the investors' personal views to be a direct prior on the means of the returns and by including a typical Inverse Wishart prior on the covariance matrix of the returns (**Section 2**). We will then consider Leonard and Hsu's (1992)[10] idea for a prior on the logarithm of the covariance matrix (**Section 3**). We encountered both running time and memory allocation problems when we applied the latter version to the whole $S\&P500$. To overcome such computational problems, Bayesian factor models are considered for the analysis. This choice was also motivated by the strong connection between Black-Litterman and the Capital Asset Pricing Model, which itself can be seen as a factor model. Sensitivity analyses for the level of confidence that investors have in their own personal views were performed and performance of the models was assessed on a test data set consisting of returns over the month of January 2018.

Contents

List of Figures

List of Tables

1 The Original Black-Litterman Model

The Black-Litterman model was developed in the early 1990's and has been widely used for asset allocation. This model attempts to combine the market equilibrium [1] with the investor's personal views. Please see **Section 1.2** for an example of how personal views are created and **Section 1.3** for more details on the model.

1.1 Estimating the Market Equilibrium

The market is in equilibrium when all investors hold the market portfolio, w_{eq}. It is when the demand for the assets in this portfolio equals the supply. If we denote by π the market equilibrium returns, then the CAPM equation is $\pi = \lambda \Sigma w_{eq}$. Here, λ is the investor's risk aversion coefficient and Σ is the covariance matrix of the returns on the assets in the portfolio[5]. For more details on the connection between traditional Black-Litterman and the CAPM, please see [4].

1.2 Example of Personal Views

Let us see how personal views are inputted in the traditional model. For example, let us consider four assets: Apple Inc. (AAPL), Amazon.com Inc. (AMZN), Google Inc. (GOOG) and Microsoft Corporation (MSFT). Suppose we believe that AAPL will outperform AMZN by 2% and GOOG will have returns that amount to 5%. Let $\mu = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 & \mu_4 \end{bmatrix}^T$ with μ_1, μ_2, μ_3 and μ_4 representing the mean returns of AAPL, AMZN, GOOG and MSFT, respectively, over the period that the investors decide to trade. The

[1] Please refer to **Section 1.1** for more details on how the market equilibrium is computed

personal views can be represented as $P\mu = q_0$, where the columns in the matrix P represent the 4 stocks in the order in which we enumerated them previously and each row in P and q_0 represents a personal view:

$$P = \begin{array}{c|cccc} & \text{AAPL} & \text{AMZN} & \text{GOOG} & \text{MSFT} \\ \text{view1} & 1 & -1 & 0 & 0 \\ \text{view2} & 0 & 0 & 1 & 0 \end{array}, q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$$

One observation would be that the investor clearly can't input contradicting views such as view1 and view3:

$$P = \begin{array}{c|cccc} & \text{AAPL} & \text{AMZN} & \text{GOOG} & \text{MSFT} \\ \text{view1} & 1 & -1 & 0 & 0 \\ \text{view2} & 0 & 0 & 1 & 0 \\ \text{view3} & 1 & -1 & 0 & 0 \end{array}, q_0 = \begin{bmatrix} 0.02 \\ 0.05 \\ 0.1 \end{bmatrix}$$

1.3 The Black-Litterman Approach

Now that we have seen what the individual pieces of the model are, we are also ready to present the mathematical formulation. We will consider that the investor is looking at n assets and has v different views on those assets. The return of the assets is considered to be random, $r \sim N_n(\mu, \Sigma)$.

Black and Litterman (1992) (please see[5] for a more detailed introduction to the model) introduce a prior on the mean of this return: $\mu \sim N_n(\pi, \tau\Sigma)$. The quantity π represents the market equilibrium returns and it is obtained by using an equation equivalent to the CAPM: $\pi = \lambda\Sigma w_{eq}$, with λ representing the investor's risk aversion parameter, w_{eq} the market equilibrium weights and Σ the covariance matrix. The quantity τ is considered to be a parameter that reflects the uncertainty in the CAPM prior, typically considered to be $\tau = 0.05$. Notice that the smaller the τ, the closer our μ will be to the market equilibrium returns π.

In addition, Black and Litterman (1992) also considered the investor's personal views: $P\mu \sim N_v(q_0, \Omega)$, where v denotes the number of personal views. Each such view has associated with it an uncertainty that the investor has with respect to the view. The measures of confidence are entered as diagonal entries in the matrix Ω. In equations (1), (2) and (3) from below, Ω is a covariance matrix. Hence, on the main diagonal we will have the variances of the returns for the personal views. Therefore, a small value reflects a high confidence in the view and vice-versa.

Hence, the model is represented by:

- A normal distribution on the returns over one period:

$$r \sim N_n(\mu, \Sigma) \tag{1}$$

- A CAPM prior:

$$\mu \sim N_n(\pi, \tau\Sigma) \tag{2}$$

$$\pi = \lambda\Sigma w_{eq}$$

  - π is a vector containing the market equilibrium returns for the stocks.
  - λ is the investor's risk aversion parameter.

- $w_{eq} = \begin{bmatrix} w_1 & ... & w_n \end{bmatrix}^T$ is a vector of market equilibrium weights for the stocks selected. Those can be computed simply by using the following formula:

$$w_i = \frac{\text{outstanding shares for stock i} \cdot \text{price for stock i}}{\sum_{i=1}^{n} \text{outstanding shares for stock i} \cdot \text{price for stock i}}$$

- Investor's views prior:

$$P\mu \sim N_v(q_0, \Omega) \tag{3}$$

By combining (2) and (3), Black and Litterman reported that:

$$\mu \sim N(\bar{\mu}, M^{-1}) \text{ , where} \tag{4}$$

$$M^{-1} = \left((\tau\Sigma)^{-1} + P^T\Omega^{-1}P\right)^{-1} \text{ and}$$
$$\bar{\mu} = \left((\tau\Sigma)^{-1} + P^T\Omega^{-1}P\right)^{-1}\left((\tau\Sigma)^{-1}\pi + P^T\Omega^{-1}q_0\right)$$

According to (1) and (4), the marginal distribution of r, unconditional on μ is:

$$r \sim N(\bar{\mu}, \bar{\Sigma}) \text{ , where } \bar{\Sigma} = M^{-1} + \Sigma \tag{5}$$

In the Black-Litterman model, the return is considered to be random and we have just seen that the posterior distribution is also normal (5). This equation appropriately takes into account market volatility and correlations also. Let us further look at the weights, w, that one would obtain when using the posterior of the returns. The typical approach to the problem that an investor with risk aversion parameter λ has when trying to maximize the returns of the portfolio while minimizing the risk is to maximize the function $w^T \mu_{post} - \frac{\lambda}{2} w^T \Sigma_{post} w$ with respect to w. By taking the derivative with respect to w, we obtain for the optimal weights an equation equivalent to CAPM:

$$\mu_{post} = \lambda \Sigma_{post} w^* \tag{6}$$

One can also observe the fact that if an investor has a different risk aversion parameter, $\hat{\lambda}$, he or she can obtain the optimized portfolio weights by using the equation $\hat{w^*} = \frac{\lambda}{\hat{\lambda}} w^*$.

He and Litterman (2002) [5] also observed that the optimal portfolio weights w^* can be expressed as a function of the market equilibrium portfolio:

$$w^* = \frac{1}{1+\tau}\left(w_{eq} + P^T \Lambda\right)$$

where

$$\Lambda = \frac{\tau}{\lambda}\Omega^{-1} q - \frac{1}{1+\tau} A^{-1} P \Sigma w_{eq} - \frac{1}{\lambda(1+\tau)} A^{-1} P \Sigma P^T \Omega^{-1} q$$

and

$$A = \frac{1}{\tau}\Omega + \frac{1}{1+\tau} P \Sigma P^T$$

In the traditional Black-Litterman approach, it was suggested to replace the covariance matrix Σ by a matrix estimated from historical data, after which it treated Σ as a known covariance matrix in their model. This can be problematic since there is extensive literature (please see [7]) that shows the fact that the sample covariance matrix is not a good estimator when the number of variables (or stocks - n - in our case) increases. The optimal portfolio weights w^* can be obtained by plugging in all parameters: the CAPM prior mean π, the uncertainty parameter τ, the personal views parameters P, q_0, Ω and the covariance matrix Σ. The model they proposed was a probability model. The optimal portfolio weights were easily obtained by plugging in all parameters. No data was collected, only the covariance was obtained from historical data. Instead, in here, we will propose a statistical approach, more specifically, a complete Bayesian statistical approach, which also takes into consideration the investor's views. We will consider two cases: (1) when historical data is available and (2) when historical data is not available.

2 Bayesian Models - Inverse-Wishart prior on covariance of returns

2.1 Introduction

The original Black-Litterman Model is a probability model. The allocations can be determined by inputting all parameters in the model. All parameters are determined based on the knowledge of market economic conditions. In here, we would like to develop a statistical model in which parameters are estimated using current data.

We will first look at the traditional model, look at what we could potentially change and introduce a new approach. Since we would like to develop a statistical approach, we will introduce a sample of returns:

(1) – $r_1, r_2, ..., r_m | \mu, \Sigma \overset{iid.}{\sim} N_n(\mu, \Sigma)$.

  * r_i represents the return in the i^{th} trading period (for example daily return in the i^{th} trading day or the hourly return in the i^{th} hour of the trading period or etc.)

  * m = number of returns = length of the trading period or the length of the period over which the investor is intending to hold the portfolio.

– In the pursuit of creating a Bayesian approach, we consider the commonly used priors; a normal prior on the mean vector μ and an Inverse Wishart prior on the covariance matrix Σ:

$$\mu \sim N_n(\pi, \Lambda)$$
$$\Sigma \sim W^{-1}(\nu, \Sigma_0)$$

Notice that in the traditional model, they considered a special case when $\Lambda = \tau\Sigma$. However, the derivations work in the same way, obtaining the same equations as in (4) with a more general positive definite covariance matrix Λ instead of $\tau\Sigma$.

– But how do we specify the prior parameters using historical data:

  * ν = number of historical returns

  * Σ_0 = sample covariance matrix of historical returns

(2) – According to Black and Litterman we would also like to consider the investor's view $P\mu \sim N_v(q_0, \Omega)$.

– Therefore we would have 2 priors on μ:

$$\begin{cases} \mu \sim N_n(\pi, \Lambda) \\ P\mu \sim N_v(q_0, \Omega) \end{cases} \tag{7}$$

– This creates an inconsistency since if μ follows $N_n(\pi, \Lambda)$ then $P\mu$ follows $N_v(P\pi, P\Lambda P^T)$. But q_0 and Ω are parameters inputted by the investor and, therefore, in general we have that $P\pi \neq q_0$ and $P\Lambda P^T \neq \Omega$.

(3) Since v = number of personal views and n = number of stocks, usually, in practice, we will have that $v < n$. Therefore, the prior in the traditional approach $\mu \sim N_n(\pi, \Lambda)$ contains more information (n pieces) than the prior $P\mu \sim N_v(q_0, \Omega)$ does (v pieces). In order for our prior to contain as much information as the traditional approach, we could construct an augmented matrix P^* by adding rows to the original matrix P such that the resulting $n \times n$ matrix P^* is invertible. We will see more details about how to accomplish this in the next **Section 2.2**.

(4) Consider an augmented P^* and let $r_i^* = P^* r_i$.

– Hence, we obtain that the assumptions of our model are:

$$r_1^*, r_2^*, ..., r_m^* \overset{iid.}{\sim} N_n(\mu^*, \Sigma^*)$$

$$\mu^* \sim N_n({q_0}^*, \Omega^*), \text{ where } \mu^* = P^*\mu$$

$$\Sigma^* \sim W^{-1}(\nu, \Sigma_0), \text{ where } \Sigma^* = P^*\Sigma P^{*T}$$

– According to investor's views represented by the second equation in (7), we replace the expected value of $P\mu$ by q_0 and the covariance matrix of $P\mu$ by Ω. More specifically, we have the following:

$$
{q_0}^* = E\left[\mu^*\right] = E\left[P^*\mu\right] = E\left[\begin{bmatrix} P \\ P_2 \end{bmatrix}\mu\right] = \begin{bmatrix} q_0 \\ q_2 \end{bmatrix} \tag{8}
$$

$$
\Omega^* = Var\left(P^*\mu\right) = Var\left(\begin{bmatrix} P \\ P_2 \end{bmatrix}\mu\right) = \tag{9}
$$

$$
= \begin{bmatrix} \Omega & PVar(\mu)P_2^T \\ P_2Var(\mu)P^T & P_2Var(\mu)P_2^T \end{bmatrix} \tag{10}
$$

– This suggests how we should specify ${q_0}^*$ and Ω^* and it can be done using our historical data. For simplicity, let us assume that the horizon over which the investor is intending to hold the portfolio (or the trading period) is $m = 21$ (the average number of trading days in a month). We split the historical dataset of the transformed returns $r_i^* = P^*r_i$ into months and compute a mean for each month. In the end we would take an average of those monthly means and replace the first entries with q_0. We would take the variance of those monthly means and replace the top left part of the matrix with Ω.

Remark 1. *Please notice that in the prior of $\mu^* \sim N_n({q_0}^*, \Omega^*)$, with the above specification of parameters, v pieces of information are from the investor's view (the q_0 part of (8)) and $n - v$ pieces* *of information from history (the q_2 part of (8) are estimations obtained from historical data). Similarly, in the traditional approach we have n pieces of information from historical data due to the CAPM prior (2) and v from the investor's views prior (3).*

Also, once we have an invertible P, we can follow two approaches:

- Obtain the distribution of μ, which could be easily done if P is invertible.
- From the very beginning transform the returns into the personal view space: $r_i^* = Pr_i$. This procedure will still require P to be invertible since after obtaining the posterior in the transformed space, we have to be able to transform back.

Hence, either way, we would need to have a matrix P that is invertible and this brings us to the following discussion.

2.2 Creating an Invertible P

The matrix of our personal views is very likely not invertible since most of the times it is not even square. As we will see, the v views that we will have (the number of rows in P) will be smaller than the n assets that we are considering to trade (the number of columns in P). In this section, we will present a method in which we can add rows to P such that the resulting square matrix P^* is invertible. The main idea is based on the way in which one would row reduce a matrix to the echelon form.

Besides the fact that the investor clearly can't input inconsistent views, as we have seen in **Section 1.2**, there is another important remark that can be made:

Remark 2. *Views (which are the rows in the matrix P) can be inputted by the investor such that they are linearly independent. It is simpler to see this in an example. Let us consider that the investor inputs views which are linearly dependent such as:*

$$P = \begin{array}{c c c c c} & AAPL & AMZN & GOOG & MSFT \\ view1 & 1 & -1 & 0 & 0 \\ view2 & 0 & 1 & -1 & 0 \\ view3 & 1 & 0 & -1 & 0 \end{array}, q_0 = \begin{bmatrix} 0.02 \\ 0.05 \\ 0.07 \end{bmatrix}$$

The first 2 views from above imply that the relationships among investors' expected returns over the period of length m are:

$$\begin{cases} E\left[R_{AAPL}\right] - E\left[R_{AMZN}\right] = 0.02 \\ E\left[R_{AMZN}\right] - E\left[R_{GOOG}\right] = 0.05 \end{cases} \Rightarrow E\left[R_{AAPL}\right] - E\left[R_{GOOG}\right] = 0.07$$

Therefore, the third view is redundant and should not be inputted.

With the above remark, we are ready to proceed with the methodology of adding rows to our matrix P (which already has linearly independent rows). It is well known that a matrix is invertible if and only if its row reduced echelon form is the identity matrix. This gives us the idea of making it invertible by adding rows to it in the following way:

- For each column in P which contains only 0's, we have to create a new row that will have only one 1 in the respective column and 0's in all the others.
- If a row has more than 1 nonzero entry, for each one except the entries in the pivot columns, we have to create a row in which we have a 1.

For example, if we consider the matrix P from **Section 1.2**, the above procedure gives us:

$$P = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} = P^* = \begin{bmatrix} P \\ P_2 \end{bmatrix}$$

Please notice that we denoted by P^* the augmented invertible matrix based on P, and by P_2 the part that was added to P.

2.3 Derivation of Posterior Distributions

Now that we have found a method to augment P to a matrix P^* that is invertible and we also managed to create corresponding ${q_0}^*$ and Ω^*, the problem is posed in a more typical Bayesian framework:

$$r_1^*, r_2^*, ..., r_m^* | \mu^*, \Sigma^* \overset{iid.}{\sim} N_n(\mu^*, \Sigma^*) \tag{11}$$

$$\mu^* \sim N_n({q_0}^*, \Omega^*) \tag{12}$$

$$\Sigma^* \sim W^{-1}(\nu, \Sigma_0) \tag{13}$$

From (11), we obtain that the joint density of our returns is:

$$f(r_1^*, ..., r_m^* | \mu^*, \Sigma^*) \propto det(\Sigma^*)^{-\frac{m}{2}} exp\left\{ -\frac{1}{2} \sum_{i=1}^{m} (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) \right\}$$

From (12), we obtain that the density for μ^* is:

$$\pi(\mu^*) \propto det(\Omega^*)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\mu^* - q_0{}^*)^T \Omega^{*-1}(\mu^* - q_0{}^*)\right\}$$

Similarly, using (13), we obtain that the density for Σ^* is:

$$\pi(\Sigma^*) \propto det(\Sigma^*)^{-\frac{\nu+n+1}{2}} exp\left\{-\frac{1}{2}Tr\left(\Sigma_0 \Sigma^{*-1}\right)\right\}$$

Here $Tr(A)$ represents the trace of the matrix A. Hence, by multiplying the above 3 equations, we obtain that the joint density for all of them is:

$$\pi(r_1^*, ..., r_m^*, \mu^*, \Sigma^*) \propto det(\Sigma^*)^{-\frac{\nu+m+n+1}{2}} exp\left\{-\frac{1}{2}Tr\left(\Sigma_0 \Sigma^{*-1}\right)\right\} det(\Omega^*)^{-\frac{1}{2}}$$
$$\times exp\left\{-\frac{1}{2}\left((\mu^* - q_0{}^*)^T \Omega^{*-1}(\mu^* - q_0{}^*) + \sum_{i=1}^{m}(r_i^* - \mu^*)^T \Sigma^{*-1}(r_i^* - \mu^*)\right)\right\} \tag{14}$$

Let us focus on the parenthesis in the second exponential and let us prove the following result.

Lemma 1. *The following equality holds, where* $\bar{r}^* = \frac{\sum_{i=1}^m r_i^*}{m}$*:*

$$\sum_{i=1}^{m}(r_i^* - \mu^*)^T \Sigma^{*-1}(r_i^* - \mu^*) = \sum_{i=1}^{m}(r_i^* - \bar{r}^*)^T \Sigma^{*-1}(r_i^* - \bar{r}^*)$$
$$+ m(\bar{r}^* - \mu^*)^T \Sigma^{*-1}(\bar{r}^* - \mu^*)$$

Proof. We will start by manipulating the right hand side:

$$RHS = \sum_{i=1}^{m}\left(r_i^{*T} \Sigma^{*-1} r_i^* - r_i^{*T} \Sigma^{*-1} \bar{r}^* - \bar{r}^{*T} \Sigma^{*-1} r_i^* + \bar{r}^{*T} \Sigma^{*-1} \bar{r}^*\right)$$
$$+ m\bar{r}^{*T} \Sigma^{*-1} \bar{r}^* - m\bar{r}^{*T} \Sigma^{*-1} \mu^* - m\mu^{*T} \Sigma^{*-1} \bar{r}^* + m\mu^{*T} \Sigma^{*-1} \mu^*$$

But since $m\bar{r}^* = \sum_{i=1}^m r_i^* \Rightarrow m\bar{r}^{*T} = \sum_{i=1}^m r_i^{*T}$, we obtain that:

$$RHS = \sum_{i=1}^m (r_i^{*T}\Sigma^{*-1}r_i^* - r_i^{*T}\Sigma^{*-1}\bar{r}^* - \bar{r}^{*T}\Sigma^{*-1}r_i^*) + 2m\bar{r}^{*T}\Sigma^{*-1}\bar{r}^*$$
$$- \left(\sum_{i=1}^m r_i^{*T}\right)\Sigma^{*-1}\mu^* - \mu^{*T}\Sigma^{*-1}\left(\sum_{i=1}^m r_i^*\right) + \sum_{i=1}^m \mu^{*T}\Sigma^{*-1}\mu^*$$
$$= \sum_{i=1}^m (r_i^{*T}\Sigma^{*-1}r_i^* - r_i^{*T}\Sigma^{*-1}\bar{r}^* - \bar{r}^{*T}\Sigma^{*-1}r_i^*) + 2m\bar{r}^{*T}\Sigma^{*-1}\bar{r}^*$$
$$- \sum_{i=1}^m r_i^{*T}\Sigma^{*-1}\mu^* - \sum_{i=1}^m \mu^{*T}\Sigma^{*-1}r_i^* + \sum_{i=1}^m \mu^{*T}\Sigma^{*-1}\mu^*$$

We observe that Σ^{*-1} and $\bar{r}^*$ do not depend on the sum. Hence, we can factor them out:

$$RHS = \sum_{i=1}^m (r_i^{*T}\Sigma^{*-1}r_i^* - r_i^{*T}\Sigma^{*-1}\mu^* - \mu^{*T}\Sigma^{*-1}r_i^* + \mu^{*T}\Sigma^{*-1}\mu^*)$$
$$+ 2m\bar{r}^{*T}\Sigma^{*-1}\bar{r}^* - \left(\sum_{i=1}^m r_i^{*T}\right)\Sigma^{*-1}\bar{r}^* - \bar{r}^*\Sigma^{*-1}\left(\sum_{i=1}^m r_i^*\right)$$
$$= \sum_{i=1}^m (r_i^* - \mu^*)^T\Sigma^{*-1}(r_i^* - \mu^*) + 2m\bar{r}^{*T}\Sigma^{*-1}\bar{r}^* - m\bar{r}^{*T}\Sigma^{*-1}\bar{r}^*$$
$$- m\bar{r}^{*T}\Sigma^{*-1}\bar{r}^* = \sum_{i=1}^m (r_i^* - \mu^*)^T\Sigma^{*-1}(r_i^* - \mu^*)$$

□

Let us make a notation before we proceed: $s^2 = \frac{1}{m-1}\sum_{i=1}^m (r_i^* - \bar{r}^*)^T\Sigma^{*-1}(r_i^* - \bar{r}^*)$.

Now, by using **Lemma 1**, we are ready to come back to the parenthesis in the second exponential from the joint density of $(r_1^*, ..., r_m^*, \mu^*, \Sigma^*)$ (equation

(14)):

$$
\begin{aligned}
(\mu^* - {q_0}^*)^T \Omega^{*-1} (\mu^* - {q_0}^*) + \sum_{i=1}^{m} (r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*) \\
= (m-1)s^2 + m(\bar{r}^* - \mu^*)^T \Sigma^{*-1} (\bar{r}^* - \mu^*) + (\mu^* - {q_0}^*)^T \Omega^{*-1} (\mu^* - {q_0}^*) \\
= (m-1)s^2 + (\bar{r}^* - \mu^*)^T (m\Sigma^{*-1}) (\bar{r}^* - \mu^*) \\
(\mu^* - {q_0}^*)^T \Omega^{*-1} (\mu^* - {q_0}^*)
\end{aligned}
\tag{15}
$$

Lemma 2. ***(Completing the square)*** *For any $A \in \mathbb{R}^{p \times p}$ positive definite, $B \in \mathbb{R}^{p \times p}$ positive semi-definite and $a, b \in \mathbb{R}^p$ the following identity holds:*

$$
\begin{aligned}
(y-a)^T A (y-a) + (y-b)^T B (y-b) = (y-y^*)^T (A+B)(y-y^*) + \\
+ (a-b)^T H (a-b),
\end{aligned}
$$

where $y^* = (A+B)^{-1}(Aa + Bb)$ and $H = A(A+B)^{-1}B$. If, furthermore, B is positive definite, then $H = (A^{-1} + B^{-1})^{-1}$. [11]

Since both of our normal distributions are not degenerated because we can have inverses for both Σ^* and Ω^*, we conclude that they do not have any eigenvalues equal to 0. Moreover, since they are covariance matrices, we know that they are positive semi-definite. Therefore their eigenvalues are greater than or equal to 0. But since they can't be 0, we observe that they have to be strictly greater than 0. This implies that both matrices are positive definite and therefore we can use the second formula for H in **Lemma 2**.

Now, if we apply this result to equation (15) for $y = \mu^*$, $a = \bar{r}^*$, $b = {q_0}^*$, $A = m\Sigma^{*-1}$ and $B = \Omega^{*-1}$, we obtain:

$$(m-1)s^2 + (\mu^* - \overline{\mu^*})^T (m\Sigma^{*-1} + \Omega^{*-1})(\mu^* - \overline{\mu^*}) +$$
$$+ (\bar{r}^* - {q_0}^*)^T H (\bar{r}^* - {q_0}^*),$$

where

$$\overline{\mu^*} = \left(m\Sigma^{*-1} + \Omega^{*-1}\right)^{-1} \left(m\Sigma^{*-1}\bar{r}^* + \Omega^{*-1}{q_0}^*\right) \text{ and}$$
$$H = \left(\frac{1}{m}\Sigma^* + \Omega^*\right)^{-1}$$

If we go back with this result in the joint density represented by equation (14), we obtain that:

$$f(r_1^*, ..., r_m^*, \mu^*, \Sigma^*) \propto$$
$$\propto det(\Sigma^*)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}(\mu^* - \overline{\mu^*})^T (m\Sigma^{*-1} + \Omega^{*-1})(\mu^* - \overline{\mu^*})\right\}$$
$$\times exp\left\{-\frac{1}{2}(\bar{r}^* - {q_0}^*)^T H (\bar{r}^* - {q_0}^*) + (m-1)s^2\right\}$$
$$\times det(\Omega^*)^{-\frac{1}{2}} det(\Sigma^*)^{\frac{\nu+n+1}{2}} exp\left\{-\frac{1}{2}Tr\left(\Sigma_0\Sigma^{*-1}\right)\right\}$$

Since the only part that depends on μ^* is the first line of the above equation, we conclude that:

$$\pi(\mu^* | r_1^*, ..., r_m^*, \Sigma^*) \propto exp\left\{-\frac{1}{2}(\mu^* - \overline{\mu^*})^T (m\Sigma^{*-1} + \Omega^{*-1})(\mu^* - \overline{\mu^*})\right\}$$

Therefore, we conclude:

$$\mu^*|r_1^*,...,r_m^*,\Sigma^* \sim N_n\left(\overline{\mu^*},\overline{\Sigma^*}\right), \text{ where}$$
$$\overline{\mu^*} = \left(m\Sigma^{*-1}+\Omega^{*-1}\right)^{-1}\left(m\Sigma^{*-1}\bar{r}^* + \Omega^{*-1}{q_0}^*\right)$$
$$\overline{\Sigma^*} = \left(m\Sigma^{*-1}+\Omega^{*-1}\right)^{-1} \tag{16}$$

In order to find the posterior of Σ^*, it is easier to start from the original joint density represented by equation (14). By collecting the terms that depend on Σ^* we obtain:

$$\pi(\Sigma^*|r_1^*,...,r_m^*,\mu^*) \propto det(\Sigma^*)^{-\frac{\nu+m+n+1}{2}}$$
$$\times exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{m}(r_i^*-\mu^*)^T\Sigma^{*-1}(r_i^*-\mu^*) + Tr\left(\Sigma_0\Sigma^{*-1}\right)\right)\right\} \tag{17}$$

We notice that this is quite close to another Inverse Wishart distribution, the only step left that we have to make is to manipulate the exponential. Note that:

$$\sum_{i=1}^{m}(r_i^*-\mu^*)^T\Sigma^{*-1}(r_i^*-\mu^*) = Tr\left(\sum_{i=1}^{m}(r_i^*-\mu^*)^T\Sigma^{*-1}(r_i^*-\mu^*)\right)$$
$$= \sum_{i=1}^{m} Tr\left((r_i^*-\mu^*)^T\Sigma^{*-1}(r_i^*-\mu^*)\right)$$

But inside $Tr(\cdot)$, matrices are cyclically commutative as long as the dimensions agree:

$$\sum_{i=1}^{m} Tr\left((r_i^* - \mu^*)^T \Sigma^{*-1} (r_i^* - \mu^*)\right) = \sum_{i=1}^{m} Tr\left((r_i^* - \mu^*)(r_i^* - \mu^*)^T \Sigma^{*-1}\right)$$
$$= Tr\left(\sum_{i=1}^{m} (r_i^* - \mu^*)(r_i^* - \mu^*)^T \Sigma^{*-1}\right)$$

Finally, by using this result and equation (17), we obtain:

$$\pi(\Sigma^* | r_1^*, ..., r_m^*, \mu^*) \propto det(\Sigma^*)^{-\frac{\nu+m+n+1}{2}}$$
$$\times exp\left\{-\frac{1}{2} Tr\left(\left(\Sigma_0 + \sum_{i=1}^{m} (r_i^* - \mu^*)(r_i^* - \mu^*)^T\right) \Sigma^{*-1}\right)\right\}$$

We notice that this is the kernel of an Inverse Wishart distribution. Therefore, we can conclude that:

$$\Sigma^* | r_1^*, ..., r_m^*, \mu^* \sim W^{-1}\left(\nu + m, \Sigma_0 + \sum_{i=1}^{m} (r_i^* - \mu^*)(r_i^* - \mu^*)^T\right) \quad (18)$$

Now that we have the posterior distributions, we can implement a Gibbs Sampler, which we will see in the following section, where we will also look at how the parameters of the model were estimated.

2.4 Implementation

For implementation purposes, 4 stocks were chosen: Apple(AAPL), Amazon(AMZN), Google(GOOG) and Microsoft(MSFT). Closing prices for the 4 from 1/2/2015 until 5/1/2017 were considered and the returns were computed. Now, this data is split into 2 parts, one representing the *current data* (the last m returns $r_1, ..., r_m$, here $m = 21$) and the rest representing *historical data* used to estimate the parameters in the model. The reason why $m = 21$ was chosen is because we are thinking of modeling the returns that happen within a period of approximately a month and 21 is the average number of trading days in a month. Hence, in this example, the trading period for such an investor would be over a month. Next step is to augment P as discussed in **Section 2.2**. Once P^* is created, we can just create our transformed returns $r_i^* = P^* r_i$. For this example, the following personal views were chosen (the columns represent AAPL, AMZN, GOOG, MSFT, respectively), which also yielded the following augmented P^*:

$$P = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, P^* = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

If we look at the second assumption in the model represented by equation (12), we notice that ${q_0}^*$ and Ω^* are, respectively, the mean and covariance matrix for μ^*, which is in turn a mean of returns from a particular month (again, in this example $m = 21$, approximately a month). Hence, one solution for estimating the parameters would be to take the returns from each month in the historical data and to compute their means. This way, we would have estimates for the monthly mean returns $\hat{\mu_i^*}$, with i an integer between 1 and the number of months in the historical data. Once we obtain those, we can estimate $\hat{q_0}^*$ and $\hat{\Omega}^*$ by taking the mean and the covariance of $\hat{\mu_i^*}$.

But, we have to remember that we need to reflect our personal views in the estimation presented above. In equations (8) and (9), we have showed how one should combine the estimates from the procedure just presented with the investor's personal views:

- Equation (8) shows that we should take the $\hat{q_0}^*$ obtained through the above estimation and replace the first v entries with q_0 (v, as mentioned at the beginning, was the number of personal views). For example, in our implementation we obtain:

$$q_0^* = \begin{bmatrix} 0.02 \\ 0.05 \\ 0.0011579235 \\ 0.0007917503 \end{bmatrix}$$

- Equation (9) shows that we should take the obtained $\hat{\Omega}^*$ and replace the top left $v \times v$ matrix with our personal choice of Ω:

$$\begin{bmatrix} \omega_1 & 0 & -1.072918 \cdot 10^{-5} & -2.665874 \cdot 10^{-7} \\ 0 & \omega_2 & 1.980838 \cdot 10^{-6} & -5.312208 \cdot 10^{-6} \\ -1.072918 \cdot 10^{-5} & 1.980838 \cdot 10^{-6} & 1.487749 \cdot 10^{-5} & 3.732911 \cdot 10^{-6} \\ -2.665874 \cdot 10^{-7} & -5.312208 \cdot 10^{-6} & 3.732911 \cdot 10^{-6} & 9.331705 \cdot 10^{-6} \end{bmatrix}$$

Now that the parameters of our model are estimated, a typical Gibbs Sampler was used based on the posteriors represented by equations (18) and (16).

A burning period of 10^3 was chosen and the number of iterations for the Gibbs Sampler is 10^4. After the Gibbs sampler is completed, one would only have to take the mean of the simulated $\mu^{*(t)}$, call it $\hat{\bar{\mu}}^*$, and the average of the simulated $\Sigma^{*(t)}$, call it $\hat{\bar{\Sigma}}^*$. However, one has to remember that those

Algorithm 1 Gibbs Sampler

1: $\Sigma^{*(t+1)}|r_1^*, ..., r_m^*, \mu^{*(t)} \sim W^{-1}\left(\nu + m, \Sigma_{W^{-1}}\right)$, where

$$\Sigma_{W^{-1}} = \Sigma_0 + \sum_{i=1}^{m}(r_i^* - \mu^{*(t)})(r_i^* - \mu^{*(t)})^T$$

2: $\mu^{*(t+1)}|r_1^*, ..., r_m^*, \Sigma^{*(t+1)} \sim N_n\left(\overline{\mu}^{*(t+1)}, \overline{\Sigma}^{*(t+1)}\right)$, where

$$\overline{\mu}^{*(t+1)} = \left(m\Sigma^{*(t+1)^{-1}} + \Omega^{*-1}\right)^{-1}\left(m\Sigma^{*(t+1)^{-1}}\bar{r}^* + \Omega^{*-1}q_0{}^*\right)$$

$$\overline{\Sigma}^{*(t+1)} = \left(m\Sigma^{*(t+1)^{-1}} + \Omega^{*-1}\right)^{-1}$$

were transformed using P^*, hence now we would have to transform them back into the original space: $\hat{\bar{\mu}} = P^{*-1}\hat{\bar{\mu}}^*, \hat{\bar{\Sigma}} = P^{*-1}\hat{\bar{\Sigma}}^* P^{*-T}$. Just like in the original model, in order to get the weights, one would use an equation similar to the CAPM one presented in **Section 1.3**: $w = \frac{1}{\lambda}\hat{\bar{\Sigma}}^{-1}\hat{\bar{\mu}}$. Here, $\lambda = 2.5$, as chosen in the original model. Also there has been extensive research when it comes to choosing λ. For trading stocks a risk aversion coefficient between 2 and 3 is reasonable.[6] Finally, we are ready to compare the results obtained under the original model with the ones obtained from this one.

2.5 Results Comparison

Before we delve into how we compare the 2 approaches, let us make the observation that in order to make any kind of comparison, one has to make sure that the same data sets were used and the parameters were estimated in the same way. Albeit the same personal views were imputed (same P, Ω, q_0), the two approaches differ in the fact that the extension has a prior on Σ and the original makes use of the market equilibrium returns, which are estimated using $\pi = \lambda\Sigma w_{eq}$. In the following table, we can look at the setup for both side by side:

Extension	Original
$r_1^*, r_2^*, ..., r_m^* \overset{iid.}{\sim} N_n(\mu^*, \Sigma^*)$	$r \sim N(\mu, \Sigma)$
$\mu^* \sim N(q_0{}^*, \Omega^*)$	$\mu \sim N(\pi, \tau\Sigma)$

Instead of the market equilibrium, the extension simply has another parameter, which is estimated as mentioned in **Section 2.4** (also the extension has a prior on Σ and takes into consideration current data). Besides this difference, the two are using the same data sets and the same parameters. Now, the question becomes how should one compare the two. One obvious approach would be to see how the two would perform if one would use them on the real market, which will be presented in the results section for the models that will follow later in this paper. However, it is of more interest to us to check how close to our personal opinion is the posterior mean obtained from the Gibbs Sampler.

Remark 3. *Since for both models we have that $P\mu \sim N(q_0, \Omega)$, the smaller the uncertainty in our views (the diagonal entries of Ω), the smaller the standard deviation and, hence, the more certain the investor is about that particular view.*

Hence, from the above remark, we will look at how $P\hat{\mu}$ behaves as we look at small values for the diagonal entries of Ω. But how should one define "small"? As we have seen in **Section 2.4**, the expected returns for the views were $q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$. Hence, even a value of 10^{-4} is quite large since this would be the variance of our view and, therefore, the standard deviation would become 10^{-2}. Hence, a 95% confidence interval for the first view would be $(0, 0.04)$. If one tries to input even smaller ω, the Inverse Wishart random generator gives a non-singularity error. Hence, we conclude that we compare the models on values of the diagonal of the matrix Ω that are between 0 and 10^{-4}. Albeit we can't input smaller ω, for the purposes of checking the following remark, we changed q_0 to $q_0 = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$. Hence, for both models an exhaustive method was implemented that would compute for each pair of diagonal entries in Ω a posterior mean $\hat{\overline{\mu}}$. Once this is obtained, the distance $|P\hat{\overline{\mu}} - q_0|$ can be calculated for both models.

Remark 4. *Since $P\mu \sim N(q_0, \Omega)$, we have that $\lim_{\Omega \to O_2} P\mu = q_0$ a.s.*

Therefore, as the diagonal entries of Ω get smaller and smaller we expect to get closer and closer to q_0.

Remark 5. *If we look at the posterior of μ^* we have that:*

$$\mu^*|r_1^*, ..., r_m^*, \Sigma^* \sim N_n\left(\overline{\mu}^*, \overline{\Sigma}^*\right), \text{ where}$$
$$\overline{\mu}^* = \left(m\Sigma^{*-1} + \Omega^{*-1}\right)^{-1}\left(m\Sigma^{*-1}\overline{r}^* + \Omega^{*-1}{q_0}^*\right)$$
$$\overline{\Sigma}^* = \left(m\Sigma^{*-1} + \Omega^{*-1}\right)^{-1}$$

If we consider a small Ω^*, its inverse $\left(\Omega^{*-1}\right)$ is large. Therefore, the whole term $m\Sigma^{*-1} + \Omega^{*-1} \approx \Omega^{*-1}$, which implies that $(m\Sigma^{*-1} + \Omega^{*-1})^{-1} \approx \Omega^*$. Similarly, $(m\Sigma^{*-1}\overline{r}^* + \Omega^{*-1}{q_0}^*) \approx \Omega^{*-1}{q_0}^*$ for small enough Ω^*. Hence, we would expect that the mean of the simulated $\mu^{*(t)}$ is close to ${q_0}^*$. Or, with the notation already used, $\hat{\overline{\mu}}^* \approx {q_0}^*$. Hence, by using the previous remark also, we obtain that $P(P^{*-1}\overline{\mu}^*) \approx q_0$.

The following graphs have as 2 of the axis the 2 diagonal entries in Ω and the third one represents the distance $|P\overline{\mu} - q_0| = |P(P^{*-1}\overline{\mu}^*) - q_0|$:

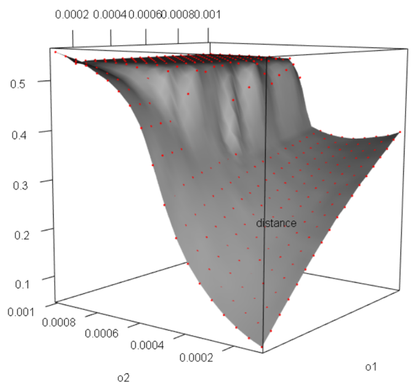

Figure 1: Results of Ω for the extension model

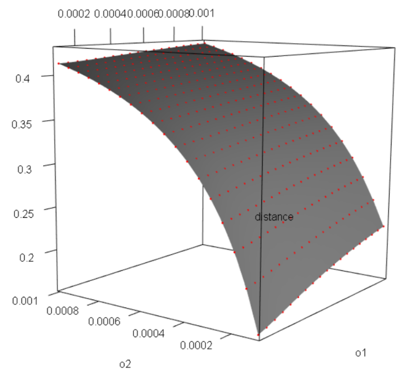

Figure 2: Results of Ω for original model

We notice from the z-axis, which represents the distance mentioned above, that the modified model more closely follows the personal views. This is according to our intuition: in **Remark 1** we have noticed that the modified model contains v pieces of prior information from investor's views and $n - v$ from historical data. Meanwhile, the original model contains v pieces from investor's personal views and n ($n > v$ in practice usually) pieces of information from the historical data through the CAPM prior. Hence, the original model contains v more pieces of information in the prior from historical data and, therefore, we would expect the original model to follow history more closely and to converge to the personal views more slowly, a fact that can be observed from the previous figures.

We can also look at some specific values of the distance for different pairs of ω_1 and ω_2 in Table 1.

ω_1	ω_2	Original	Extension
10^{-4}	0.0001	0.154	0.052
10^{-4}	0.00015	0.2	0.063
10^{-4}	0.0002	0.235	0.077
10^{-4}	0.00025	0.263	0.118
10^{-4}	0.0003	0.286	0.118
10^{-4}	0.00035	0.305	0.145
10^{-4}	0.0004	0.321	0.177
10^{-4}	0.00045	0.335	0.222
10^{-4}	0.0005	0.347	0.282
10^{-4}	0.00055	0.357	0.354

Table 1: Table with specific distance values

However, we would like to see if the structure of $P\hat{\overline{\mu}}$ is similar to q_0. For this we keep the two entries in Ω equal, we exhaustively search over small ω s.t. $\Omega = \omega\mathbb{I}$ and we plot the two entries of $P\hat{\overline{\mu}}$ together with the respective ω. Please note that the blue point (the one at coordinate $(o, q1, q2) = (0, 0.2, 0.5))$ in Figures 3 and 4 represents the exact value of $q_0 = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$, which would be obtained for $\omega = 0$.

By comparing Figures 3 and 4, we notice that not only the point simulations represented by the red points are closer, but the whole curve (which was obtained by interpolation) seems to be closer to the theoretical value represented by the blue point. Also, we notice that in both cases, as ω increases, $P\hat{\overline{\mu}}$ gets further away from q_0, which is what theoretically should happen.

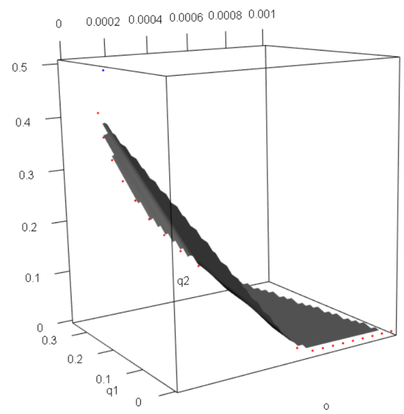

Figure 3: Results of Ω for the extension model

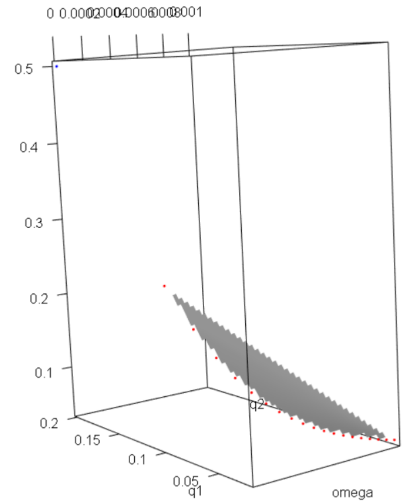

Figure 4: Results of Ω for original model

2.6 But do we need an invertible P?

In **Section 2.2**, we introduced a method of creating an invertible matrix P by adding rows. This has both advantages and disadvantages:

- Disadvantages:
  - The method presented in **Section 2.2** for augmenting P in order to become invertible is not unique.
- Advantages:
  - When augmenting P, v pieces of prior information come from the personal views and $n - v$ pieces of prior information come from history, as we have seen in **Remark 1**.

We now consider the posteriors when P is unaugmented from what the investor is inputting. Hence, in this section, we will consider the same setup as before, with the only difference being the fact that P is not even square:

$$r_1, r_2, ..., r_m|\mu, \Sigma \overset{iid.}{\sim} N_n(\mu, \Sigma)$$
$$P\mu \sim N_v(q_0, \Omega)$$
$$\Sigma \sim W^{-1}(\nu, \Sigma_0)$$

Since P shows up in the second equation of our model assumptions, the only posterior that will change from what we had previously will be that for μ. Hence, in the joint distribution, we will consider only the terms depending on μ:

$$\pi(\mu|r_1, ..., r_m) \propto exp\left\{-\frac{1}{2}\sum_{i=1}^{m}(r_i - \mu)^T\Sigma^{-1}(r_i - \mu)\right\}$$
$$\times\, exp\left\{-\frac{1}{2}(P\mu - q_0)^T\Omega^{-1}(P\mu - q_0)\right\}$$

For the first exponential we can use **Lemma 1**. This yields:

$$\pi(\mu|r_1, ..., r_m) \propto exp\left\{-\frac{1}{2}\left((m-1)s^2 + m(\bar{r} - \mu)^T\Sigma^{-1}(\bar{r} - \mu)\right)\right\}$$
$$\times\, exp\left\{-\frac{1}{2}(q_0 - P\mu)^T\Omega^{-1}(q_0 - P\mu)\right\}$$

We remember that $s^2 = \frac{1}{m-1}\sum_{i=1}^{m}(r_i - \bar{r})^T\Sigma^{-1}(r_i - \bar{r})$ and hence this term does not depend on μ. Now, let us focus on the remaining terms in the exponential:

$$(\overline{r}-\mu)^T(m\Sigma^{-1})(\overline{r}-\mu)+(q_0-P\mu)^T\Omega^{-1}(q_0-P\mu)$$
$$=\overline{r}^T(m\Sigma^{-1})\overline{r}-2\overline{r}^T(m\Sigma^{-1})\mu+\mu^T(m\Sigma^{-1})\mu+{q_0}^T\Omega^{-1}q_0-2{q_0}^T\Omega^{-1}P\mu$$
$$+\mu^TP^T\Omega^{-1}P\mu=\mu^T\left(m\Sigma^{-1}+P^T\Omega^{-1}P\right)\mu-2\left(\overline{r}^T(m\Sigma^{-1})+{q_0}^T\Omega^{-1}P\right)\mu$$
$$+\overline{r}^T(m\Sigma^{-1})\overline{r}+{q_0}^T\Omega^{-1}q_0$$

Since only the first two terms depend on μ, we obtain that:

$$\pi(\mu|r_1,...,r_m,\Sigma)\propto exp\left\{-\frac{1}{2}\mu^T(m\Sigma^{-1}+P^T\Omega^{-1}P)\mu\right\}$$
$$\times\, exp\left\{-\frac{1}{2}2(m\Sigma^{-1}\overline{r}+P^T\Omega^{-1}q_0)^T\mu\right\}$$

Lemma 3. *Let M be a symmetric and invertible matrix, then the following identity holds:*

$$x^TMx-2b^Tx=(x-M^{-1}b)^TM(x-M^{-1}b)-b^TM^{-1}b$$

Proof. We just need to expand the quadratic term:

$$(x-M^{-1}b)^TM(x-M^{-1}b)=x^TMx-2b^TM^{-1}Mx+b^TM^{-1}MM^{-1}b$$
$$=x^TMx-2b^Tx+b^TM^{-1}b$$

□

Hence, if we apply this lemma for $x=\mu$, $M=m\Sigma^{-1}+P^T\Omega^{-1}P$ and $b=m\Sigma^{-1}\overline{r}+P^T\Omega^{-1}q_0$, we obtain that the exponential in the distribution of the posterior of μ is (the $-\frac{1}{2}$ still sits in front of the formula, we just omit it in the following for simplicity of writing):

$$\left(\mu - (m\Sigma^{-1} + P^T\Omega^{-1}P)^{-1}(m\Sigma^{-1}\overline{r} + P^T\Omega^{-1}q_0)\right)^T (m\Sigma^{-1} + P^T\Omega^{-1}P)$$
$$\times \left(\mu - (m\Sigma^{-1} + P^T\Omega^{-1}P)^{-1}(m\Sigma^{-1}\overline{r} + P^T\Omega^{-1}q_0)\right) - b^T M^{-1} b$$

Lastly, we notice that b and M do not depend on μ, and hence, the posterior of μ is dictated by the first big term, which is actually the density of a normal distribution:

$$\mu | r_1, ..., r_m, \Sigma \sim N\left(\mu_{post}, \Sigma_{post}\right), \text{ where}$$
$$\mu_{post} = (m\Sigma^{-1} + P^T\Omega^{-1}P)^{-1}(m\Sigma^{-1}\overline{r} + P^T\Omega^{-1}q_0)$$
$$\Sigma_{post} = \left(m\Sigma^{-1} + P^T\Omega^{-1}P\right)^{-1}$$

This posterior is very close to the one obtained by using the first approach (represented by equation (16)), the only difference being the fact that in this new approach the matrix P shows up. This is because here we did not change the investor inputted matrix P, while in the previous approach we augmented P in order for it to be invertible.

2.7 Implementation

Implementing this model is straightforward since it is very similar to the previous version. The only difference is the fact that in the posterior for μ we have P appearing, while in the previous model there was no P since we were adding rows to it so that it becomes invertible. We remind ourselves that this was the first approach because we can take the inverse and easily find the prior distribution of μ from the prior distribution of $P\mu$. Using the derived posteriors, the Gibbs Sampler is:

Algorithm 2 Gibbs Sampler

1: $\Sigma^{(t+1)}|r_1,...,r_m,\mu^{(t)} \sim W^{-1}\left(\nu+m,\Sigma_0+\sum_{i=1}^{m}(r_i-\mu^{(t)})(r_i-\mu^{(t)})^T\right)$

2: $\mu^{(t+1)}|r_1,...,r_m,\Sigma^{(t+1)} \sim N\left({\mu_{post}}^{(t+1)},{\Sigma_{post}}^{(t+1)}\right)$, where

$$
\begin{aligned}
{\mu_{post}}^{(t+1)} &= (m{\Sigma^{(t+1)}}^{-1}+P^T\Omega^{-1}P)^{-1}(m{\Sigma^{(t+1)}}^{-1}\overline{r}+P^T\Omega^{-1}q_0) \\
{\Sigma_{post}}^{(t+1)} &= \left(m{\Sigma^{(t+1)}}^{-1}+P^T\Omega^{-1}P\right)^{-1}
\end{aligned}
$$

2.8 Results

Just like before, we will try to look at the sensitivity of our model to different confidence levels. **Remarks 3 and 4** made when we presented the results for the previous model still hold. Since in Ω we have on the main diagonal (call them ω_i) the variances in our views $P\mu$, the smaller the ω_i, the more certain we are in view i. This should also be reflected in our posterior: if we provide very large ω_i, it means that we are very uncertain about the views and the model should take into consideration the history a lot more, while if we provide very small values for ω_i, it means that we are very certain about the views and the model should take them into consideration a lot more than the history.

Just like before, in order to quantify and visualize the model's sensitivity to different confidence levels, we will look at the distance $|P\mu_{post}-q_0|$ (which will be on one of the axis in our plots) over different combinations of ω_i. The same 4 stocks from before were chosen (AAPL,AMZN,GOOG,MSFT), but since this work is more recent, the daily returns are from 1/2/2014 to 12/29/2017. The views are (rows are views and the columns represent the 4 stocks in the order AAPL,AMZN,GOOG,MSFT):

$$q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, P = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

When it comes to the confidence levels in the 2 views, one can input values as small as 10^{-7} without encountering any numerical issues, like we did previously when we were augmenting the matrix P. Hence, one doesn't need to make any change to the model when implementing it or when inputting any value. We take a grid of equally spaced points (ω_1, ω_2) between 10^{-7} and $2 \cdot 10^{-5}$. The burn period was set to 10^3 and the number of iterations in the Gibbs Sampler was set to 10^4.

However, one could also use the same views, but considering the daily returns for the whole $S\&P500$ instead of just for 4 stocks. For this, we need the daily returns of companies actively traded in $S\&P500$ over the period mentioned before. We won't have to change q_0 at all, but P has more columns since they would represent the stocks in the famous index and it will still have 2 rows for the same 2 views. One would fill out P by making sure that in the first row and the column corresponding to AAPL we will have a -1, in the first row and the column corresponding to AMZN we will have a 1 and similarly for the second row. Of course, the dimension of some of the matrices and vectors will be much bigger and therefore, all computations will be more expensive. Hence, this version was parallelized and the number of iterations in the Gibbs Sampler decreased to 10^3 (as we will see, even with so few iterations, convergence for the mean is achieved, but convergence for the covariance matrix is not). The interval 10^{-7} to 10^{-5} for the confidence levels was split into 4 parts, in the following way:

(1) $\omega_1 \in \{10^{-7}, 2.5 \cdot 10^{-7}, 4 \cdot 10^{-7}, 5.5 \cdot 10^{-7}, 7 \cdot 10^{-7}, 8.5 \cdot 10^{-7}, 10^{-6}\}$ and $\omega_2 \in \{10^{-7}, 2.5 \cdot 10^{-7}, 4 \cdot 10^{-7}, 5.5 \cdot 10^{-7}, 7 \cdot 10^{-7}, 8.5 \cdot 10^{-7}, 10^{-6}\}$, with each possible pair (ω_1, ω_2) ran on one core.

(2) $\omega_1 \in \{10^{-6}, 2.5 \cdot 10^{-6}, 4 \cdot 10^{-6}, 5.5 \cdot 10^{-6}, 7 \cdot 10^{-6}, 8.5 \cdot 10^{-6}, 10^{-5}\}$ and $\omega_2 \in \{10^{-6}, 2.5 \cdot 10^{-6}, 4 \cdot 10^{-6}, 5.5 \cdot 10^{-6}, 7 \cdot 10^{-6}, 8.5 \cdot 10^{-6}, 10^{-5}\}$, with each possible pair (ω_1, ω_2) ran on one core.

(3) $\omega_1 \in \{10^{-7}, 2.5 \cdot 10^{-7}, 4 \cdot 10^{-7}, 5.5 \cdot 10^{-7}, 7 \cdot 10^{-7}, 8.5 \cdot 10^{-7}, 10^{-6}\}$ and $\omega_2 \in \{10^{-6}, 2.5 \cdot 10^{-6}, 4 \cdot 10^{-6}, 5.5 \cdot 10^{-6}, 7 \cdot 10^{-6}, 8.5 \cdot 10^{-6}, 10^{-5}\}$, with each possible pair (ω_1, ω_2) ran on one core.

(4) $\omega_1 \in \{10^{-6}, 2.5 \cdot 10^{-6}, 4 \cdot 10^{-6}, 5.5 \cdot 10^{-6}, 7 \cdot 10^{-6}, 8.5 \cdot 10^{-6}, 10^{-5}\}$ and $\omega_2 \in \{10^{-7}, 2.5 \cdot 10^{-7}, 4 \cdot 10^{-7}, 5.5 \cdot 10^{-7}, 7 \cdot 10^{-7}, 8.5 \cdot 10^{-7}, 10^{-6}\}$, with each possible pair (ω_1, ω_2) ran on one core.

Each pair (ω_1, ω_2) took a little more than 4 hours to run.

- When $\omega_1 = 10^{-6}$, with 95% confidence, the return on the first view would be in the interval $(0.018, 0.022)$, which would show that the investor is very confident.

- When $\omega_1 = 10^{-4}$, with 95% confidence, the return on the first view would be in the interval $(0, 0.04)$, which would show that the investor is not as confident.

In the figures presented, we notice that both curves have similar shapes, albeit the one on the right converges slower to 0 as o_i become smaller (ω_i in our model). Also, the curve on the right seems to be underneath the one on the left. Intuitively, this is because there is a lot more information

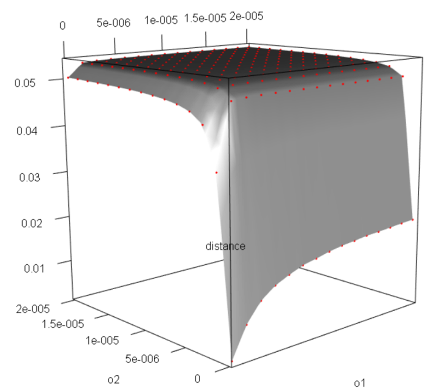

Figure 5: $|P\mu_{post} - q_0|$ when taking only the 4 stocks

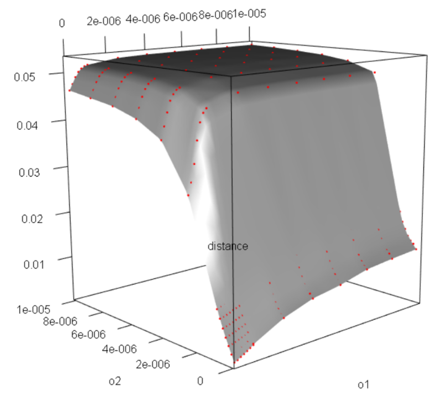

Figure 6: $|P\mu_{post} - q_0|$ when taking $S\&P500$

in our prior for Σ when we take the whole $S\&P500$. Moreover, both have very similar shapes. The distances go to 0 as ω_i go to 0. This is in tune with our intuition of how the model should behave like: as one gets more and more confident in their inputted views, the model should put a lot of importance on them and not on the historical data. Vice-versa, in both figures the distance seems to converge to a certain value as ω_i become bigger and bigger. Again, this is what we would think that the model should do since large ω_i, suggests that one is uncertain about the personal view and therefore, the history should play a more important role. Indeed, if we would only take the historical returns, an unbiased estimate for μ is $\overline{r}$ and the distance becomes $|P\overline{r} - q_0| = 0.05388875$, which is what the plots seem to tend to converge to.

We will move our focus towards looking at the profits (or losses) that one would obtain when using the model to trade over the month of January 2018 (testing data consisting of daily returns between 1/2/2018 and 1/30/2018) using an initial capital of $\$100,000$ (this does not include any capital requirements for short selling). We remember that in order to get portfolio weights we use the same approach as before. From Gibbs Sampling we estimate μ_{post} and Σ_{post} and we use the CAPM equation 6: $w = \frac{1}{2.5}\Sigma_{post}{}^{-1}\mu_{post}$.

Albeit when we took the whole $S\&P500$ the number of iterations in the Gibbs Sampler was small, we notice from the above analysis that we still get very good estimates for μ_{post} since the posterior distance behaves exactly like our intuition suggests it should do. The running averages for the mean also converge fast for small ω_i. However, because of the size of Σ_{post} and because of the fact that one has to take its inverse in order to compute the portfolio weights w, the number of iterations is not enough to give accurate predictions of profits. Nevertheless, for completeness, the average profit when considering the whole $S\&P500$ is $\$13,191.39$ with a standard deviation of $\$2,908.134$.

We will now present the profits obtained when using only 4 stocks. We notice that the first view has a bigger impact on the profits curve than the second view. Moreover, as the confidence in the first view increases (as ω_1 goes to 0), the profits sky rocket. This is because over the month of January 2018 AMZN outperformed AAPL by 23.997% and our view was indeed that AMZN will overrun AAPL (albeit by only 2%, a 10^{th} of what actually happened in reality).

AMZN outperforming AAPL by nearly 24% in one month is uncommon. Therefore, next we will present the same results, the only change made is that we replace AMZN with FB (Facebook). The same data sets were used and all other inputs stay exactly the same as we just presented at the beginning of this section, except q_0. We will also look at how the model behaves when the investor inputs a personal view exactly like what happened during the month of January 2018 (very "informed" investor) and exactly the

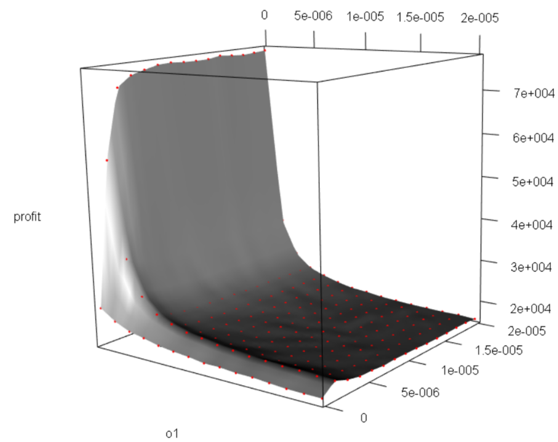

Figure 7: Profit when taking only 4 stocks

opposite of what happened during the month of January (very "uninformed" investor). Therefore, we will also look at what happens when we choose $q_0 = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ and $q_0 = -\begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$, respectively.

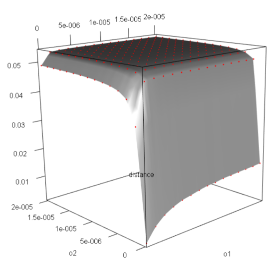

Figure 8: 4 stocks,FB in and $q_0 = [0.02, 0.05]^T$

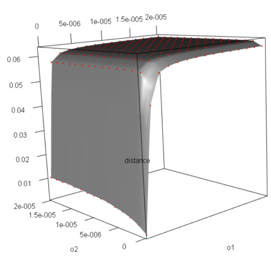

Figure 9: 4 stocks, FB in and view exactly like reality

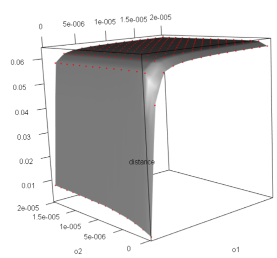

Figure 10: 4 stocks, FB in and view opposite of reality

Again, just like before, we notice that, as ω_i get smaller and smaller, when taking into account the whole $S\&P500$, the curve seems to be under and closer to 0 than the one when taking into account only 4 stocks. This might be because the prior on the covariance matrix containing the whole

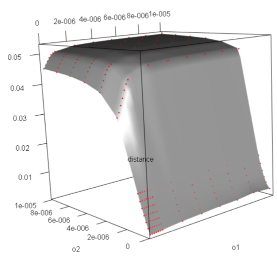

Figure 11: $S\&P500$, FB in and $q_0 = [0.02, 0.05]^T$

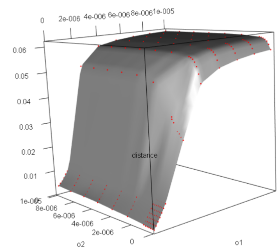

Figure 12: $S\&P500$, FB in and view exactly like reality

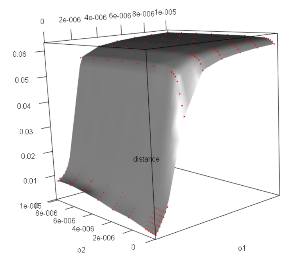

Figure 13: $S\&P500$ FB in and view opposite of reality

$S\&P500$ has more information than the one which only has 4 stocks. Moreover, for the same q_0, the curves have a similar orientation and general shape. Hence, this confirms the belief that albeit a small number of iterations was used for the Gibbs Sampler that takes into account the whole $S\&P500$, the estimated posterior mean is still accurate. However, as mentioned before, the estimate for Σ_{post}^{-1} when it's size is big is not accurate enough to have very reliable profit estimates.

Nevertheless, for completeness of this analysis, we proceed by leaving all the inputs mentioned before unchanged and keeping $q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$. When taking into account the whole $S\&P500$, the average profit over the before mentioned range of simulated pairs (ω_1, ω_2) is \$11,619.97 with a standard deviation of \$2,852.246. In the next plot we can observe the profits obtained when considering just the 4 stocks mentioned before.

From the figure, one can see that the first view has a higher influence on the profits than the second view. This is because if we let ω_2 constant the resulting curve increases a lot faster than the curve obtained by keeping ω_1 constant.

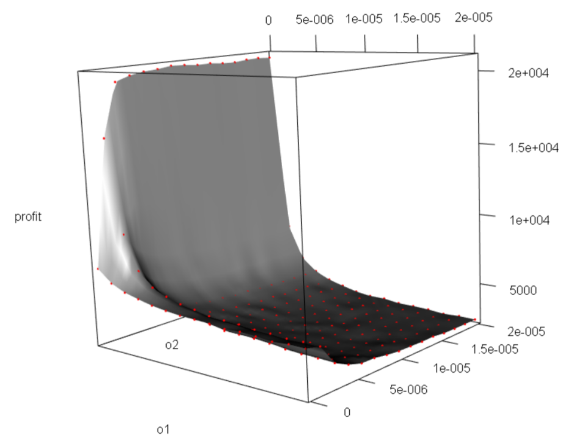

Figure 14: Profit 4 stocks, FB in and $q_0 = [0.02, 0.05]^T$

3 Bayesian Models - Leonard-Hsu prior on covariance of returns

3.1 Introduction

Just like when introducing the approach with an Inverse Wishart prior, let us see what we would like to improve on it:

- It has been shown by Alvarez, Niemi and Simpson in [1] that it creates a strong a priori dependence between the correlation and the variance.
- With an Inverse-Wishart prior on $\Sigma \sim W^{-1}(\nu, \Sigma_0)$, all its entries depend on two parameters: ν and Σ_0.

Therefore, two of the assumptions will be unchanged:

$$r_1, r_2, ..., r_m|\mu, \Sigma \overset{iid.}{\sim} N_n(\mu, \Sigma)$$
$$P\mu \sim N_v(q_0, \Omega)$$

A very interesting idea for a different prior on the covariance matrix is presented by Leonard and Hsu (1992)[10]. As the title of this section is hinting, this prior will actually be on $log(\Sigma)$. In order to better understand Leonard and Hsu's idea, let us look at the distribution:

$$f(r_1, ..., r_m|\mu, \Sigma) = (2\pi)^{-\frac{mn}{2}} det(\Sigma)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}\sum_{i=1}^{m}(r_i - \mu)^T\Sigma^{-1}(r_i - \mu)\right\}$$

Let $A = log(\Sigma)$, λ_{Ai} and $\lambda_{\Sigma i}$ (with $i = \{1, 2, ..., n\}$) be the eigenvalues of A and Σ respectively. Since $A = log(\Sigma)$ we obtain that $\lambda_{Ai} = log(\lambda_{\Sigma i}) \Rightarrow \lambda_{\Sigma i} = e^{\lambda_{Ai}}$. Finally, by remembering that the determinant is the product of the eigenvalues and that the trace of a matrix is the sum of the eigenvalues, we notice that $det(\Sigma) = \prod_{i=1}^{n} \lambda_{\Sigma i} = \prod_{i=1}^{n} e^{\lambda_{Ai}} = e^{Tr(A)}$. By using this in the joint distribution of the returns and by noticing that $(r_i - \mu)^T\Sigma^{-1}(r_i - \mu) \in \mathbb{R}$ we obtain:

$$f(r_1, ..., r_m|\mu, \Sigma) = (2\pi)^{-\frac{mn}{2}} exp\left\{-\frac{1}{2}Tr\left(\sum_{i=1}^{m}(r_i - \mu)^T\Sigma^{-1}(r_i - \mu)\right)\right\}$$
$$\times exp\left\{-\frac{m}{2}Tr(A)\right\} = (2\pi)^{-\frac{mn}{2}} exp\left\{-\frac{1}{2}Tr\left(\sum_{i=1}^{m}(r_i - \mu)(r_i - \mu)^T\Sigma^{-1}\right)\right\}$$
$$\times exp\left\{-\frac{m}{2}Tr(A)\right\} = (2\pi)^{-\frac{mn}{2}} exp\left\{-\frac{m}{2}Tr\left(A + Se^{-A}\right)\right\}$$

Here, $S = \frac{1}{m}\sum_{i=1}^{m}(r_i - \mu)(r_i - \mu)^T$. Before we continue, let us define an operator and make a few notations.

Definition 1. *Let A be a $n \times n$ matrix, $A = (a_{ij})_{i,j=\{1,2,...,n\}}$, then we define* *an operator that stacks in a vector the entries parallel to the main diagonal:*

$$Vec^*(A) = \begin{bmatrix} a_{11} & a_{22} & ... & a_{nn}| & a_{12} & a_{23} & ... & a_{n-1n}|...|a_{1n} \end{bmatrix}^T$$

We notice that if A is $n \times n$, $Vec^*(A)$ is $\frac{1}{2}n(n+1) \times 1$. This definition brings us to the following notations:

Notation 1.

$$\lambda = Vec^*(log(S)), \alpha = Vec^*(log(\Sigma))$$
$$\Lambda = log(S), A = log(\Sigma), d = \frac{1}{2}n(n+1)$$

The idea that Leonard and Hsu had was to approximate $f(r_1, ..., r_m|\mu, \Sigma)$ by approximating e^{-A}. The approximation makes use of the fact that $X(\omega) = e^{-A\omega}$ satisfies a Volterra integral equation[3]:

$$X(t) = S^{-t} - \int_0^t S^{s-t}(A - \Lambda)X(v)dv, 0 < t < \infty,$$

By letting $t = 1$, by iterative substitution of $X(v)$ and by using the spectral decomposition of matrix S we obtain that the approximation is (please see Appendix A for the proof):

$$f^*(r_1, ..., r_m|\alpha) = (2\pi e)^{-\frac{mn}{2}} det(S)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}(\alpha - \lambda)^T Q(\alpha - \lambda)\right\} \quad (19)$$

In order to see how to compute Q, we first have to introduce a couple more notations. If we let e_i, d_i to be the i^{th} normalized eigenvector with its corresponding eigenvalue, respectively, then f_{ij} is obtained by looking at the equation $Vec^*(log(\Sigma))^T f_{ij} = e_i^T log(\Sigma) e_j$ and identifying the coefficients of the entries in the $log(\Sigma)$ matrix. With those f_{ij}, we can finally compute Q:

$$Q = \frac{m}{2}\sum_{i=1}^{n} f_{ii}f_{ii}^T + m\sum_{i<j}^{n}\xi_{ij}f_{ij}f_{ij}^T, \text{ where}$$
$$\xi_{ij} = \frac{(d_i - d_j)^2}{d_i d_j (log(d_i) - log(d_j))^2} \tag{20}$$

Remark 6. *The approximate distribution is:* $\alpha | r_1, ..., r_m \approx\sim N(\lambda, Q^{-1})$

Now we are ready to move on to the next section and resent the assumptions of the model.

3.2 The Model

As mentioned in the previous section, we will have a prior on the $log(\Sigma)$. But how would one construct an intuitive distribution? The simplest distribution that one could work with is the multivariate normal, in which the variance terms on the main diagonal have a mean θ_1 and a variance σ_1^2 and the covariance terms, which are on the off diagonal, have another mean θ_2 and another variance σ_2^2. Hence, we arrive at the following model:

$$r_1, ..., r_m | \mu, \Sigma \overset{iid.}{\sim} N(\mu, \Sigma) \tag{21}$$

$$P\mu \sim N(q_0, \Omega) \tag{22}$$

$$\alpha | \theta, \Delta \sim N(J\theta, \Delta) \tag{23}$$

Where we have the following uninformative priors:

$$\pi(\theta) \propto 1$$
$$\pi(\sigma_1^2, \sigma_2^2) \propto 1$$

We introduced the following notations:

Notation 2.

$$J = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix}, \Delta = \begin{bmatrix} \sigma_1^2 I_n & \mathbb{O} \\ \mathbb{O} & \sigma_2^2 I_{d-n} \end{bmatrix}, \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$$

Please note that this approach has a few advantages over the classical Inverse-Wishart one:

- There are 2 parameters that determine the entries in the covariance matrix: σ_1^2 and σ_2^2 (θ is integrated out as shown in Appendix B).
- We do not need good estimates for the hyper-parameters σ_1^2 and σ_2^2.
- From a modeling perspective, it has been studied before (please see [1]) that a model which allows flexibility by allowing both covariances and variances to be modeled by the data is more appealing.

3.3 Derivation of Posterior Distributions

If we let θ to have a uniform prior ($\theta \propto 1$) by integrating it out from the density in equation (23), we obtain:

Proposition 1.

$$f(\alpha|\sigma_1^2, \sigma_2^2) = \int_\theta det(\Delta)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\alpha - J\theta)^T \Delta^{-1}(\alpha - J\theta)\right\} d\theta =$$
$$= 2\pi det(\Delta)^{-\frac{1}{2}} det(J^T \Delta^{-1} J)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\alpha^T G\alpha\right\}, \; where$$
$$G = \left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right)^T \Delta^{-1} \left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right)$$

For the proof, please see the Appendix B.

Now, by using this distribution together with the approximation obtained from the Volterra integral of the distribution of returns denoted by equation (19) and with the prior on $P\mu$ represented by equations (22), we can finally obtain the approximate joint distribution:

$$\begin{aligned} f(\alpha, \mu, \sigma_1^2, \sigma_2^2, r_1, ..., r_m) \approx\propto\; & det(\Delta)^{-\frac{1}{2}} det(J^T\Delta^{-1}J)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\alpha^T G\alpha\right\} \\ & \times det(S)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}(\alpha - \lambda)^T Q(\alpha - \lambda)\right\} \\ & \times det(\Omega)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(P\mu - q_0)^T \Omega^{-1}(P\mu - q_0)\right\} \end{aligned} \tag{24}$$

We will first proceed with finding the posterior of α. Hence, we have to collect all the terms depending on α. Since one of those is the approximation obtained from the Volterra integral, the posterior is going to be an approximate distribution:

$$\pi^*(\alpha|r_1,...,r_m,\sigma_1^2,\sigma_2^2,\mu) \approx\propto exp\left\{-\frac{1}{2}\left(\alpha^T G\alpha + (\alpha-\lambda)^T Q(\alpha-\lambda)\right)\right\}$$

We can apply **Lemma 2 (Completing the square)** with $y=\alpha, a = 0, A = G, b = \lambda, B = Q$ and we obtain that:

$$\alpha|r_1,...,r_m,\sigma_1^2,\sigma_2^2,\mu \approx\sim N(\alpha^*,(Q+G)^{-1}), \text{ where } \alpha^* = (Q+G)^{-1}Q\lambda \tag{25}$$

Moving to the posterior of σ_1^2, σ_2^2, we have to collect the terms depending on Δ, which also includes G. We note that the term obtained from the Volterra integral approximation of the matrix exponential does not show up in this posterior. Hence, this will be an exact distribution:

$$\pi(\sigma_1^2,\sigma_2^2|\alpha,\mu,r_1,...,r_m) \propto det(\Delta)^{-\frac{1}{2}} det(J^T\Delta^{-1}J)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\alpha^T G\alpha\right\}$$

However, one can write the above distribution in scalar form. By applying **Lemma 4** which can be found in **Appendix B**, one finds that the joint posterior distribution of σ_1^2, σ_2^2 is equal to:

$$\pi(\sigma_1^2,\sigma_2^2|\alpha,\mu,r_1,...,r_m) \propto \left(\sigma_1^2\right)^{-\frac{n-1}{2}} \left(\sigma_2^2\right)^{-\frac{d-n-1}{2}} exp\left\{-\frac{1}{2}\alpha^T G\alpha\right\}$$

Furthermore, by applying **Lemma 5** which can also be found in **Appendix B**, we obtain that the scalar version for the equation is:

$$\pi(\sigma_1^2, \sigma_2^2|\alpha, \mu, r_1, ..., r_m) \propto \left(\sigma_1^2\right)^{-\frac{n-1}{2}} exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i - \overline{\alpha}_v)^2\right\}$$
$$\times \left(\sigma_2^2\right)^{-\frac{d-n-1}{2}} exp\left\{-\frac{1}{2\sigma_2^2}\sum_{i=n+1}^{d}(\alpha_i - \overline{\alpha}_c)^2\right\}$$

Here, $\overline{\alpha}_v$ are the averages of the log of the variance terms and $\overline{\alpha}_c$ are the averages of the log of the covariance terms:

$$\overline{\alpha}_v = \frac{\sum_{i=1}^{n}\alpha_i}{n} \text{ and } \overline{\alpha}_c = \frac{\sum_{i=n+1}^{d}\alpha_i}{d-n}$$

Hence, both posteriors of σ_1^2 and σ_2^2 are following Inverse Gamma distributions and they are independent:

$$\sigma_1^2|\alpha, \mu, r_1, ..., r_m \sim IG\left(\frac{n-3}{2}, \frac{1}{2}\sum_{i=1}^{n}(\alpha_i - \overline{\alpha}_v)^2\right)$$
$$\sigma_2^2|\alpha, \mu, r_1, ..., r_m \sim IG\left(\frac{d-n-3}{2}, \frac{1}{2}\sum_{i=n+1}^{d}(\alpha_i - \overline{\alpha}_c)^2\right) \quad (26)$$

We are finally ready to compute the posterior for μ also by collecting the terms that depend on it. We notice that the term obtained from the Volterra integral approximation of the matrix exponential does not show up in the posterior. Therefore, like the posteriors of σ_1^2 and σ_2^2, this will be an exact distribution. Moreover, we notice that the first two equations in the assumptions of our model (equations (21) and (22)) are the same as when we used an Inverse Wishart prior. Therefore, the derivation for the posterior for μ will be the same, yielding:

$$\mu|\alpha, \sigma_1^2, \sigma_2^2, r_1, ..., r_m \sim N\left(\mu_{post}, \Sigma_{post}\right), \text{ where}$$
$$\mu_{post} = (m\Sigma^{-1} + P^T\Omega^{-1}P)^{-1}(m\Sigma^{-1}\overline{r} + P^T\Omega^{-1}q_0)$$
$$\Sigma_{post} = \left(m\Sigma^{-1} + P^T\Omega^{-1}P\right)^{-1}$$

3.4 Implementation

Now that we have derived our posteriors, we are ready to implement it, using a Gibbs Sampler. The only difference from before is that we will use a Metropolis-Hastings algorithm for sampling α, for which we need the exact posterior distribution. This will be proportional to the distribution obtained from collecting all terms with an α from the joint distribution represented by equation (24):

$$exp\left\{-\frac{1}{2}\alpha^T G\alpha\right\} det(S)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}(\alpha-\lambda)^T Q(\alpha-\lambda)\right\}$$

We have seen that it results in the posterior:

$$\alpha|r_1, ..., r_m, \sigma_1^2, \sigma_2^2, \mu \approx\sim N(\alpha^*, (Q+G)^{-1}), \text{ where } \alpha^* = (Q+G)^{-1}Q\lambda$$
$$\pi^*(\alpha|r_1, ..., r_m, \sigma_1^2, \sigma_2^2, \mu) \approx\propto exp\left\{-\frac{1}{2}(\alpha-\alpha^*)^T(Q+G)(\alpha-\alpha^*)\right\}$$

This is an approximation since $det(S)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}(\alpha-\lambda)^T Q(\alpha-\lambda)\right\}$ is an approximation of the pdf of a multivariate normal using the Volterra integral equation. If we would replace it with the exact distribution, we would obtain:

$$\pi(\alpha|r_1, ..., r_m, \sigma_1^2, \sigma_2^2) \propto exp\left\{-\frac{m}{2}Tr\left(A + Se^{-A}\right) - \frac{1}{2}\alpha^T G\alpha\right\}$$

The Metropolis-Hastings step at t^{th} iteration would be that we would simulate a candidate value from the approximate posterior distribution: $\widetilde{\alpha} \approx\sim N(\alpha^*, (Q+G)^{-1})$ and we would accept it with probability $min(\rho, 1)$, where

$$\rho = \frac{\pi\left(\widetilde{\alpha}|r_1, ..., r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)}\right)}{\pi\left(\alpha^{(t)}|r_1, ..., r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)}\right)} \cdot \frac{\pi^*\left(\alpha^{(t)}|r_1, ..., r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)}\right)}{\pi^*\left(\widetilde{\alpha}|r_1, ..., r_m, \sigma_1^{2(t)}, \sigma_2^{2(t)}, \mu^{(t)}\right)}$$

It is useful at this point to remember that because of the notation introduced in **Notation 1**, we have a connection between π^* and π since there is one between A and α, namely:

$$\alpha = Vec^*(A)$$

Using the Metropolis Hastings step that was just discussed, we arrive at the following Gibbs Sampler:

Algorithm 3 Gibbs Sampler $log(\Sigma)$

1: $\alpha^{(t+1)} = \begin{cases} \widetilde{\alpha} \sim N\left(\left(Q^{(t)} + G^{(t)}\right)^{-1} Q^{(t)}\lambda^{(t)}, \left(Q^{(t)} + G^{(t)}\right)^{-1}\right) \text{w.p. } min(\rho, 1) \\ \alpha^{(t)}\text{otherwise} \end{cases}$

2: Since $\alpha = Vec^*(log(\Sigma)) \Rightarrow \begin{cases} \text{compute } \Sigma^{(t+1)} = exp\left\{Vec^{*-1}\left(\alpha^{(t+1)}\right)\right\} \\ \text{keep } \Sigma^{(t)} \end{cases}$

3: $\begin{cases} {\sigma_1^2}^{(t+1)} \sim IG\left(\frac{n-3}{2}, \frac{1}{2}\sum_{i=1}^{n}\left({\alpha_i}^{(t+1)} - {\overline{\alpha}_v}^{(t+1)}\right)^2\right) \\ {\sigma_2^2}^{(t+1)} \sim IG\left(\frac{d-n-3}{2}, \frac{1}{2}\sum_{i=n+1}^{d}\left({\alpha_i}^{(t+1)} - {\overline{\alpha}_c}^{(t+1)}\right)^2\right) \end{cases} \Rightarrow$

$\Rightarrow \Delta^{(t+1)} = \begin{bmatrix} {\sigma_1^2}^{(t+1)} I_n & \mathbb{O} \\ \mathbb{O} & {\sigma_2^2}^{(t+1)} I_{d-n} \end{bmatrix}$

4: Let $\Sigma_\mu = \left(m{\Sigma^{(t+1)}}^{-1} + P^T\Omega^{-1}P\right)^{-1}$, $\mu^{(t+1)} \sim N\left(\Sigma_\mu\left(m{\Sigma^{(t+1)}}^{-1}\overline{r} + P^T\Omega^{-1}q_0\right), \Sigma_\mu\right)$

5: Compute $S^{(t+1)} = \frac{1}{m}\sum_{i=1}^{m}\left(r_i - \mu^{(t+1)}\right)\left(r_i - \mu^{(t+1)}\right)^T$, $\lambda^{(t+1)} = Vec^*\left(log\left(S^{(t+1)}\right)\right)$, ${d_j}^{(t+1)}$ and ${e_j}^{(t+1)}$ the eigenvalue and normalized eigenvector of $S^{(t+1)}$ respectively.

6: Compute $f_{ij}^{(t+1)}$ by identifying the coefficients of the entries of the $log(\Sigma)$ matrix from the equation $Vec^*\left(log\left(\Sigma^{(t)}\right)\right)^T {f_{ij}}^{(t+1)} = {{e_i}^{(t+1)}}^T log\left(\Sigma^{(t)}\right) {e_j}^{(t+1)}$

7: Compute $\xi_{ij}^{(t+1)} = \frac{({d_i}^{(t+1)} - {d_j}^{(t+1)})^2}{{d_i}^{(t+1)}{d_j}^{(t+1)}\left(log\left({d_i}^{(t+1)}\right) - log\left({d_j}^{(t+1)}\right)\right)^2}$

8: Compute $Q^{(t+1)} = \frac{m}{2}\sum_{i=1}^{n} {f_{ii}}^{(t+1)}{{f_{ii}}^{(t+1)}}^T + m\sum_{i<j}^{n} {\xi_{ij}}^{(t+1)}{f_{ij}}^{(t+1)}{{f_{ij}}^{(t+1)}}^T$

9: Compute

$$G^{(t+1)} = \left(I_d - J(J^T{\Delta^{(t+1)}}^{-1}J)^{-1}J^T{\Delta^{(t+1)}}^{-1}\right)^T {\Delta^{(t+1)}}^{-1} \times$$
$$\times \left(I_d - J(J^T{\Delta^{(t+1)}}^{-1}J)^{-1}J^T{\Delta^{(t+1)}}^{-1}\right)$$

3.5 Results

Just like we did before, in this section we will depict the sensitivity of the model to changes in confidence levels (ω_i) in terms of both the distance of the posterior to investor's view and the profits obtained if one would use this model to trade.

Before we delve into the actual results for this version of the model, we notice that **Remarks (3) and (4)** both hold. Basically, this means that as the diagonal entries in Ω get smaller, the more confident we are in the views because we have the assumption that $P\mu \sim N(q_0, \Omega)$. Same assumption points out the fact that the smaller Ω is, the closer $P\mu$ should be to q_0. Hence, a very small Ω shows the fact that the investor is very confident in this view and, therefore, the posterior should also be close to q_0. Therefore, the smaller our Ω is, the closer $P\mu_{\text{post}}$ should be to q_0. In the first part of this section we will present some plots similar to the ones presented before. We will take 2 views and do an exhaustive search over possible combinations of pairs of values for the 2 diagonal entries of Ω (which are depicted as 2 axis) and compute the same distance as before:$|P\mu_{\text{post}} - q_0|$ (which is depicted as 1 axis).

We chose the same 4 stocks (AAPL, AMZN, GOOG, MSFT), and we will use the same data set as when we presented the results in **Section 2.8**: daily returns from 1/2/2014 to 12/29/2017. We will use the following inputs (again the columns are in order AAPL, AMZN, GOOG, MSFT and the rows represent the views):

$$q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, P = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Just like when we had a P non-square and an Inverse Wishart prior, in this version of the model, one can use smaller confidence levels than when we were just using an Inverse Wishart prior and the augmented matrix P. This time one can choose ω_i (which were defined as the entries in the main diagonal of Ω) of the order 10^{-7} without getting any numerical issues. For the results presented here, we let (ω_1, ω_2) range between 10^{-6} to 10^{-4}.

However, one can imagine that this approach is more computationally expensive than just having an Inverse Wishart prior on Σ. Therefore, the sensitivity analysis was ran in parallel on multiple cores (each core running the Gibbs Sampler for 1 pair (ω_1, ω_2)) and the range itself was split into 4 ranges:

(1) $\omega_1 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$ and $\omega_2 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$

(2) $\omega_1 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$ and $\omega_2 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$

(3) $\omega_1 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$ and $\omega_2 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$

(4) $\omega_1 \in \{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}\}$ and $\omega_2 \in \{10^{-6}, 4 \cdot 10^{-6}, 7 \cdot 10^{-6}, 10^{-5}\}$

The Gibbs Sampler was ran on one core for each possible pair (ω_1, ω_2) within the same range.

The burn period was set to 10^3 and the iterations to 10^4. Albeit those seem relatively small, convergence is actually achieved very fast when ω_i are small.

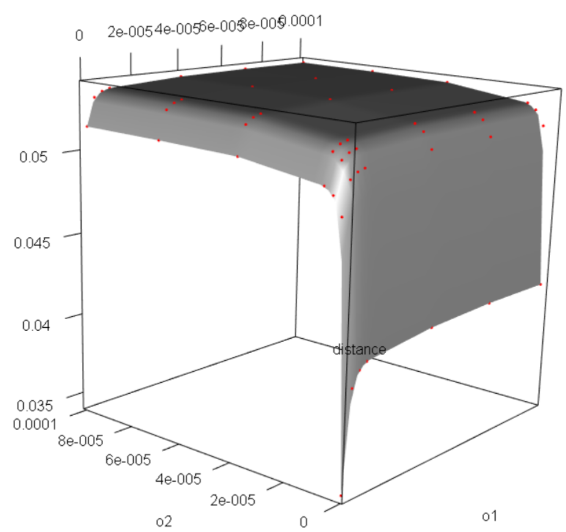

Figure 15: Distances for $log(\Sigma)$ prior

We notice that in this version of the model, the distance converges to 0 very fast as o1 (ω_1 in the model) and o2 (ω_2 in the model) go to 0. Also, we notice that as o1 and o2 get bigger, it converges very fast to a stabilizing distance. This is consistent with our intuition since if we are very confident in our views, the model should put a lot more importance on them, while if we are not confident at all in our views, the model should just take into consideration the history. Indeed, if we use only the history, the unbiased estimator for μ is the sample mean of the returns ($\overline{r}$) and therefore the distance becomes $|P\overline{r} - q_0| = 0.05388875$.

We also notice that the second view (corresponding to $o2$) has more influence on the posterior than the first view. This is because the $3D$ curve would leave a $2D$ line on a section parallel to the "o2 vs distance" plane that converges to 0 as o2 gets very small much faster than a section parallel to the "o1 vs distance" plane would when o1 gets very small.

We will proceed by looking at profits (losses) that we would obtain by using this model trained on the same daily returns between 1/2/2014 and 12/29/2017. We would estimate using Gibbs Sampling the posterior mean (μ_{post}) and the posterior covariance (Σ_{post}) and we use the CAPM equation (6) to obtain the weights to be $w = \frac{1}{2.5}{\Sigma_{post}}^{-1}\mu_{post}$. With those weights we compute the profits that we would obtain over the month January 2018 (just like before, daily returns between 1/2/2018 and 1/30/2018) with an initial investment of $\$100,000$. Here, one could use a different investment horizon also.

The same P, q_0, grid for ω_i, burn period, iteration period were used as before. The following is a $3D$ plot of the sensitivity of the profits to changes in confidence:

We observe a profit that is approximately between $\$10,000$ and $\$58,000$. In order to interpret this curve, we would have to know what actually happened in the month of January 2018 using the views inputted. More specifically, over the month of January 2018, $Pr_{\text{Jan2018}} = \begin{bmatrix} 0.23996743 \\ 0.01366718 \end{bmatrix}$. Albeit

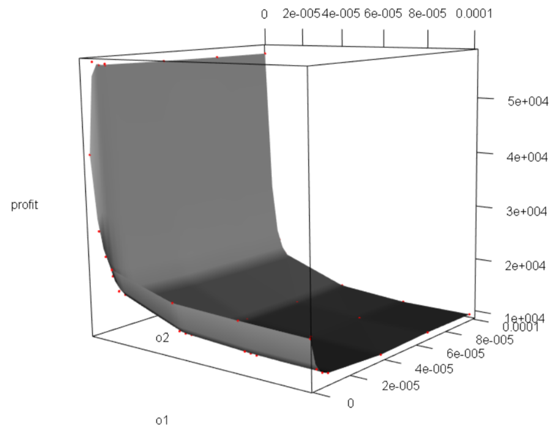

Figure 16: Profits with AMZN in and $q_0 = [0.02, 0.05]^T$

the inputted 1^{st} view is a 10^{th} of what happened in reality (AMZN outperformed AAPL by almost 24% in January 2018), the model puts a higher importance on it than on the 2^{nd} view. Indeed, the profits increase drastically as we decrease ω_1 and keep ω_2 constant. Profits do not increase much as we decrease ω_2 and keep ω_1 constant.

Just like we did before, since a 24% gain on AAPL in a month is an extreme scenario, let us consider a different stock instead of AMZN. We will replace AMZN with FB (Facebook) and we will keep all the inputs the same as before, except that we will vary q_0. In the following 3 figures we will present the results for profits when the investor considers $q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$, $q_0 = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ which is exactly what happened during the month of January 2018 (the "well informed" investor) and $q_0 = \begin{bmatrix} -0.06212815 \\ -0.01366718 \end{bmatrix}$ which is exactly the opposite of what happened during the month of January 2018 (the "poorly informed" investor):

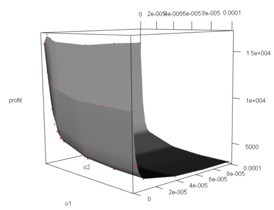

Figure 17: Profits FB instead of AMZN and $q_0 = [0.02, 0.05]^T$

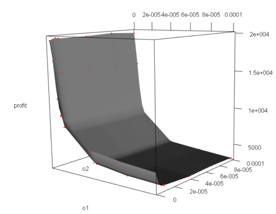

Figure 18: Profits FB instead of AMZN and view exactly like reality

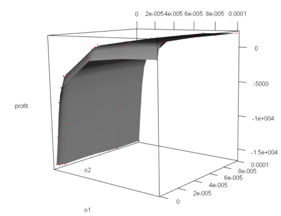

Figure 19: Profits FB instead of AMZN and view opposite of reality

- Since $Pr_{\text{Jan2018}} = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$, the view in which $q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}$ has returns that are much closer to what happened in reality than when we had AMZN instead of FB (especially the first view is closer). We notice that the second view has a greater influence on the profits than what we have seen in Figure 16 and this can be clearly noticed in Figure 17 from above.

- If the investor has a view exactly like the reality (Figure 18), the first view has more influence on the profits as ω_1 gets smaller and smaller.

- Moreover, if we compare Figures 18 and 19, we notice that they seem to be a reflection of each other with respect to a plane parallel to the "o1 vs o2" plane. This would make sense since the only difference between the two is that in Figure 18 we have a $q_0 = \begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$ and in Figure 19 we have a $q_0 = -\begin{bmatrix} 0.06212815 \\ 0.01366718 \end{bmatrix}$.

3.6 Limitations

In the previous section, we haven't presented any results for the whole $S\&P500$. This is because we have encountered both memory allocation and running time problems. Both arise from the size of the matrices which makes all matrix computations and sampling from multivariate distributions time consuming. The biggest issue is with the construction of the matrix Q. We remind ourselves that we have to compute f_{ij} by looking at the equation $Vec^*(log(\Sigma))^T f_{ij} = e_i^T log(\Sigma) e_j$ and identifying the coefficients of the entries in the $log(\Sigma)$ matrix. With those f_{ij}, we can finally compute Q:

$$Q = \frac{m}{2}\sum_{i=1}^{n} f_{ii} f_{ii}^T + m \sum_{i<j}^{n} \xi_{ij} f_{ij} f_{ij}^T, \text{ where}$$
$$\xi_{ij} = \frac{(d_i - d_j)^2}{d_i d_j (log(d_i) - log(d_j))^2}$$

It is easy to compute ξ_{ij} and the elegant way to compute the $f's$ is by coding a 4 way tensor and applying the function $Vec^*(\cdot)$ to 2 of its entries (one can see the pattern more easily by taking a small dimensional example). However, this is not the fastest way since one can actually fill out each entry in Q directly. In both situations, the dimensionality problem still exists. When we take into consideration the whole $S\&P500$, the number of rows and columns are of size $d = \frac{500 \cdot 501}{2}$, but since Q is symmetric we would have to store a little more than half of the entries in Q (albeit this approach makes all the formulas in the posterior a lot messier). Even so, the size of such an object is approximately 53 GB. Even with the biggest server at $UCSB$, for which a node has 1 TB of RAM memory, we could only run this in parallel on at most 20 cores.

The memory allocation problem combined with a running time that is a lot bigger than just the 4 hours that took to run the simulations presented in **Section 2.8** makes this approach computationally not feasible for a large data set.

We have looked at a couple of ideas to remedy the problem:

- Writing the matrix Q to the disk. Unfortunately, one would need a high speed connection (for example SSD) to be able to write it fast enough that it doesn't make the running time even longer. This is of paramount importance since we have to compute Q at each iteration of the Gibbs Sampler.

- We have looked at parallelizing the Gibbs Sampler itself (which is a Markov Chain). More precisely, in the general setting of Markov Chains, we have looked at independently starting at m initial points and, from each initial point, starting independent Markov Chains. It has been shown[2] that for one single Markov Chain that satisfies Doob's conditions, the ergodic average converges geometrically:

$$P\left(\frac{1}{n}\sum_{k=1}^{n} f(X_k) > \epsilon \middle| X_0 = x_0\right) \leq A(\epsilon)\rho(\epsilon)^n \text{ ,where}$$
$$(\exists) d_0, t_0 \text{ s.t. } \rho(\epsilon) = \Phi(d_0, t_0)^{\frac{1}{d_0}} + \eta \text{ with } \eta \text{ s.t. } \rho(\epsilon) < 1,$$
$$\Phi(d_0, t_0) = \sup_{x_0} E\left[e^{t_0 \sum_{k=1}^{d_0} f(X_k)} \middle| x_0\right]$$

  By using this result, one can easily show that for running m Markov Chains in parallel we obtain the following bound:

$$P\left(\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{k=1}^{n} f(X_{ik}) > \epsilon \middle| x_0\right) \leq e^{-{t_0}^* mn\epsilon} A^*(\epsilon)^m \rho^*(\epsilon)^{mn}$$

  Here, the existence of ${d_0}^*, {t_0}^*$ and the definitions of $A^*(\cdot), \rho^*(\cdot)$ are in the same way as before. The problem is that we cannot compare the right hand sides of the 2 inequalities from above because the $A(\cdot), A^*(\cdot)$ and $\rho(\cdot), \rho^*(\cdot)$ are different since this is a proof of existence.

4 Bayesian Factor Black-Litterman Models

The running time and memory allocation problems encountered when using the whole market would suggest that one has to reduce the dimensionality. Moreover, there is a strong connection between the original Black-Litterman model and CAPM (which can be seen as a factor analysis model in statistics). This gave us the idea of adding a fully Bayesian specified factor model to the Bayesian extensions presented in this paper. All the posteriors have already been derived for those. In this section we will provide a brief introduction to the work presented here so far and to the classical factor analysis model.

4.1 Introduction

In previous chapters, we discussed two Bayesian versions for the Black-Litterman model:

- One with an Inverse-Wishart prior on the covariance matrix of the returns:

$$\begin{aligned} r_1, r_2, ..., r_T | \mu, \Sigma &\overset{iid.}{\sim} N_n(\mu, \Sigma) \\ P\mu &\sim N_v(q_0, \Omega) \\ \Sigma &\sim W^{-1}(\nu, \Sigma_0) \end{aligned} \tag{27}$$

- The other one has a prior on the logarithm of the covariance matrix, inspired from the work of Leonard and Hsu [10]:

$$
\begin{aligned}
r_1, ..., r_T|\mu, \Sigma &\overset{iid.}{\sim} N_n(\mu, \Sigma) \\
P\mu &\sim N_v(q_0, \Omega) \\
\alpha|\theta, \Delta = Vec^*(log(\Sigma))|\theta, \Delta &\sim N_{\frac{1}{2}n(n+1)}(J\theta, \Delta)
\end{aligned}
\tag{28}
$$

Where the variables were introduced in **Notations 1 and 2** and the operator $Vec^*(\cdot)$ was defined in **Definition 1**.

Just like in the original Black-Litterman, P is the matrix of personal views, q is a vector that contains return on those views, and Ω is a diagonal matrix containing the confidence in each view. For example, if the investor believes that Amazon will outperform Apple by 2% and that Google will outperform Microsoft by 5%, they will have the following setup:

$$
q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, P = \begin{array}{c|cccc} & \text{AAPL} & \text{FB} & \text{GOOG} & \text{MSFT} \\ \text{view1} & -1 & 1 & 0 & 0 \\ \text{view2} & 0 & 0 & 1 & -1 \end{array}
$$

As we have seen in (25), when using the version with prior on the logarithm of the covariance matrix of the returns (28), the approximated posteriors for α using the Volterra integral equation are:

$$
\alpha|r_1, ..., r_m, \sigma_1^2, \sigma_2^2, \mu \approx\sim N(\alpha^*, (Q + G)^{-1}), \text{ where } \alpha^* = (Q + G)^{-1}Q\lambda
$$

The matrix Q, defined as in equation (20), is of size $d \times d = \frac{1}{2}n(n+1) \times \frac{1}{2}n(n+1)$ and is randomly generated at each iteration in a Gibbs Sampler. Therefore, if one considers the whole $S\&P500$, the size of this matrix in terms of memory would be of around 106GB. Because of this issue, we decided to introduce factors in order to reduce the dimension. Hence, as we will see in the following sections, after applying factor models, we will introduce priors on the covariance matrix of the common factors instead of introducing priors directly on the covariance matrix of the returns. The dimension of the covariance matrix of the common factors is $q \times q$ (q=number of factors), which is much smaller than $n \times n$ (n=number of stocks), the dimension of the covariance matrix of the returns.

4.2 Factor Analysis

The observable vector of returns at time t satisfies the following equation:

$$r_t - \mu = \Lambda f_t + \epsilon_t$$

Here we introduced the following notation:

Notation 3.

- *r_t is a vector of size $n \times 1$ (n =number of stocks) which represents the observed returns for each individual stock at time t.*
- *μ is a vector of size $n \times 1$ representing the means of the returns for each individual stock.*
- *Λ is a $n \times q$ matrix of factor weights.*
- *f_t is a vector of size $q \times 1$ representing the common factors at time t.*
- *ϵ_t is a vector of size $n \times 1$.*

We also have the following assumptions:

(1) $E[\epsilon_t] = 0$ and $Cov(\epsilon_t) = \Psi$. Hence, we obtain that $E\left[\epsilon_t \epsilon_t^T\right] = \Psi$.

(2) $E[f_t] = 0$ and $Cov(f_t) = \Phi$. Hence, we obtain that $E\left[f_t f_t^T\right] = \Phi$

(3) ϵ_t and f_t are independent. Hence, we obtain that $Cov(\epsilon_t, f_t) = 0$ or, equivalently, $E\left[f_t \epsilon_t^T\right] = 0$

Remark 7. *The covariance matrix of the returns is:*

$$Cov(r_t) = \Sigma = \Lambda \Phi \Lambda^T + \Psi$$

Proof.

$$\Sigma = Cov(r_t) = E\left[(r_t - \mu)(r_t - \mu)^T\right] = E\left[(\Lambda f_t + \epsilon_t)(\Lambda f_t + \epsilon_t)^T\right] =$$

$$= E\left[\Lambda f_t f_t^T \Lambda^T + 2\Lambda f_t \epsilon_t^T + \epsilon_t \epsilon_t^T\right] = \Lambda E\left[f_t f_t^T\right] \Lambda^T + 2\Lambda E\left[f_t \epsilon_t^T\right] + E\left[\epsilon_t \epsilon_t^T\right] =$$

$$= \Lambda \Phi \Lambda^T + \Psi$$

□

If we allow in the above remark $\Phi^{\frac{1}{2}}$ to be the Cholesky decomposition matrix of Φ, and we denote by $L = \Lambda \Phi^{\frac{1}{2}}$, we obtain that $\Sigma = LL^T + \Psi$.

The Principal Factor Method is taking advantage of the spectral decomposition of Σ and the above remark. Let λ_{0i} and e_i (where $i = \{1, 2, ..., n\}$) be the eigenvalues and eigenvectors of Σ, respectively. Also, let us assume that the eigenvalues are ordered in descending order: $\lambda_{01} \geq ... \geq \lambda_{0n}$ Then, the spectral decomposition of Σ can be represented as:

$$\Sigma = \sum_{i=1}^{n} \lambda_{0i} e_i {e_i}^T$$

By keeping the largest q eigenvalues and discarding the smaller $n - q$, we obtain an approximation to Σ:

$$\Sigma \approx \sum_{i=1}^{q} \lambda_{0i} e_i {e_i}^T$$

Hence, we would obtain an exact equality if we would add to the above approximation the error term:

$$\Sigma = \sum_{i=1}^{q} \lambda_{0i} e_i {e_i}^T + \Psi = \begin{bmatrix} \sqrt{\lambda_{01}} e_1 & \dots & \sqrt{\lambda_{0q}} e_q \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_{01}} e_1 \\ \vdots \\ \sqrt{\lambda_{0q}} e_q \end{bmatrix} + \Psi =$$

$$= LL^T + \Psi$$

4.3 The Models

The reduction in dimension is not the only reason of using factor models. The other motivation is, as Cheng showed in [4], that the original Black-Litterman [5] is closely related to the Capital Asset Pricing Model (CAPM), which actually is itself a factor model. Our work consists of combining the two Bayesian versions for the Black-Litterman model ((27) and (28)) with the work of Lee, Poon and Song (2007) in [8] and the work of Lee and Shi (2000) in [9].

4.4 Assumptions for Inverse-Wishart prior on covariance of common factors

We introduce a factor model on the returns (n =number of stocks, T =number of returns considered, v =number of views):

$$r_t = \mu + \Lambda f_t + e_t$$
$$e_t|\Psi \overset{iid.}{\sim} N_n(0, \Psi) \text{ for all } t = \{1, 2, ..., T\}, \text{ where } \Psi = diag(\Psi_1, ..., \Psi_n)$$

(1) Hence, by letting the parameters μ, f_t, Λ, Ψ be random so that we can put priors on them, we obtain that the conditional distribution of the returns r_t is:

$$r_t|\mu, f_t, \Lambda, \Psi \sim N_n(\mu + \Lambda f_t, \Psi) \text{ for all } t = \{1, 2, ..., T\}$$

(2) Next, let us introduce priors on all parameters:

$$f_t|\Phi \overset{iid.}{\sim} N_q(0, \Phi) \text{ for all } t = \{1, 2, ..., T\}$$
$$\Lambda_k|\Psi_k \overset{indep.}{\sim} N_q(\Lambda_{0k}, \Psi_k H_k)$$
$$\Psi_k \overset{indep.}{\sim} IG(\alpha_k, \beta_k) \text{ for all } k = \{1, 2, ..., n\}$$

Here, Λ_k^T is the k^{th} row in Λ.

(3) Following the Black-Litterman approach, we introduce a prior on the mean of the returns, which is projected through the investor's views:

$$P\mu \sim N_v(q_0, \Omega)$$

(4) Moreover, similar to (27) and (28), we introduce two different priors on

the covariance matrix of common factors, which has dimension $q \times q$. This is smaller than $n \times n$, which is the size of the covariance matrix of the returns. The first one that we will focus on is the typical Inverse-Wishart prior and the second one will be a logarithmic prior, following the work of Leonard and Hsu in [10]:

$$\Phi \sim W^{-1}(\nu_0, R_0)$$

Therefore, all the model assumptions are:

$$\begin{aligned} r_t|\mu, f_t, \Lambda, \Psi \sim N_n(\mu + \Lambda f_t, \Psi), \text{for all } t = \{1, 2, ..., T\} \\ P\mu \sim N_v(q_0, \Omega) \\ f_t|\Phi \overset{iid.}{\sim} N_q(0, \Phi),\ \text{for all } t = \{1, 2, ..., T\} \\ \Phi \sim W^{-1}(\nu_0, R_0) \\ \Lambda_k|\Psi_k \overset{iid.}{\sim} N_q(\Lambda_{0k}, \Psi_k H_k) \\ \Psi_k \overset{iid.}{\sim} IG(\alpha_k, \beta_k),\ \text{for all } k = \{1, 2, ..., n\} \end{aligned} \tag{29}$$

We will proceed by computing the posteriors for this simpler version, which has an Inverse-Wishart prior on the covariance matrix of the common factors f_t.

4.5 Posteriors for Inverse-Wishart Prior on covariance of common factors

From the model assumptions in (29), we find that the joint distribution is:

$$f(\cdot) \propto \det(\Psi)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(r_t-\mu-\Lambda f_t)^T\Psi^{-1}(r_t-\mu-\Lambda f_t)\right\}$$
$$\times \det(\Omega)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(P\mu-q_0)^T\Omega^{-1}(P\mu-q_0)\right\}$$
$$\times \det(\Phi)^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T} f_t^T\Phi^{-1}f_t\right\}\det(\Phi)^{-\frac{\nu_0+q+1}{2}}$$
$$\times \exp\left\{-\frac{1}{2}Tr(R_0\Phi^{-1})\right\}\prod_{k=1}^{n}\Psi_k^{-\alpha_k-1}\exp\left\{-\sum_{k=1}^{n}\frac{\beta_k}{\Psi_k}\right\}$$
$$\times \prod_{k=1}^{n}\det(\Psi_k H_k)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{n}(\Lambda_k-\Lambda_{0k})^T\frac{1}{\Psi_k}H_k^{-1}(\Lambda_k-\Lambda_{0k})\right\} \quad (30)$$

We start by finding the updated density of f_t:

$$\pi(f_t|\cdot) \propto \exp\left\{-\frac{1}{2}\left((\Lambda f_t-(r_t-\mu))^T\Psi^{-1}(\Lambda f_t-(r_t-\mu))+f_t^T\Phi^{-1}f_t\right)\right\}$$

Let us focus on the term in the exponential:

$$f_t^T\Lambda^T\Psi^{-1}\Lambda f_t - 2f_t^T\Lambda^T\Psi^{-1}(r_t-\mu)+(r_t-\mu)^T\Psi^{-1}(r_t-\mu)+f_t^T\Phi^{-1}f_t$$
$$= f_t^T(\Lambda^T\Psi^{-1}\Lambda+\Phi^{-1})f_t - 2(r_t-\mu)^T\Psi^{-1}\Lambda f_t + (r_t-\mu)^T\Psi^{-1}(r_t-\mu)$$

Just like before, we will repeatedly make use of **Lemma 3**. We first apply it for $x = f_t, M = \Lambda^T\Psi^{-1}\Lambda+\Phi^{-1}, b^T = (r_t-\mu)^T\Psi^{-1}\Lambda$ and we obtain that the term in the exponential for the posterior of f_t is:

$$\left(f_t-(\Lambda^T\Psi^{-1}\Lambda+\Phi^{-1})^{-1}\Lambda^T\Psi^{-1}(r_t-\mu)\right)^T(\Lambda^T\Psi^{-1}\Lambda+\Phi^{-1})$$
$$\times\left(f_t-(\Lambda^T\Psi^{-1}\Lambda+\Phi^{-1})^{-1}\Lambda^T\Psi^{-1}(r_t-\mu)\right)$$
$$-(r_t-\mu)^T\Psi^{-1}\Lambda(\Lambda^T\Psi^{-1}\Lambda+\Phi^{-1})^{-1}\Lambda^T\Psi^{-1}(r_t-\mu)$$

Here, only the first term depends on f_t and we actually observe that it is the kernel of a normal distribution. Therefore, we obtain that:

$$f_t|\cdot \overset{indep.}{\sim} N_q\left((\Lambda^T\Psi^{-1}\Lambda+\Phi^{-1})^{-1}\Lambda^T\Psi^{-1}(r_t-\mu),(\Lambda^T\Psi^{-1}\Lambda+\Phi^{-1})^{-1}\right)$$

Now, we are ready to find the posterior for μ:

$$\pi(\mu|\cdot)\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(r_t-\Lambda f_t-\mu)^T\Psi^{-1}(r_t-\Lambda f_t-\mu)\right\}$$
$$\times\exp\left\{-\frac{1}{2}(P\mu-q_0)^T\Omega^{-1}(P\mu-q_0)\right\}$$

Let $r_t^*=r_t-\Lambda f_t$, for all $t=\{1,2,...,T\}$ and $\overline{r}^*=\frac{\sum_{t=1}^T r_t^*}{T}=\frac{\sum_{t=1}^T(r_t-\Lambda f_t)}{T}$. If we focus only on the first exponential, we can apply the typical trick of subtracting and adding $\overline{r}^*$ and we obtain that the term in the first exponential is equal to:

$$\sum_{t=1}^{T}(r_t^*-\mu)^T\Psi^{-1}(r_t^*-\mu)=$$
$$\sum_{t=1}^{T}(r_t^*-\overline{r}^*)^T\Psi^{-1}(r_t^*-\overline{r}^*)+T(\overline{r}^*-\mu)^T\Omega^{-1}(\overline{r}^*-\mu)$$

Therefore, we obtain that the posterior of μ is:

$$\pi(\mu|\cdot)\propto \exp\left\{-\frac{1}{2}T(\overline{r}^*-\mu)^T\Psi^{-1}(\overline{r}^*-\mu)\right\}$$
$$\times\exp\left\{-\frac{1}{2}(P\mu-q_0)^T\Omega^{-1}(P\mu-q_0)\right\}$$

Again, let us turn our attention to the term in the exponentials:

$$\overline{r}^{*T}\Psi^{-1}\overline{r}^{*} - 2\overline{r}^{*T}(T\Psi^{-1})\mu + \mu^T T\Psi^{-1}\mu + \mu^T P^T\Omega^{-1}P\mu$$
$$-2q_0^T\Omega^{-1}P\mu + q_0^T\Omega^{-1}q_0$$

Hence, the posterior of μ is:

$$\pi(\mu|\cdot) \propto \left\{-\frac{1}{2}\left(\mu^T(T\Psi^{-1} + P^T\Omega^{-1}P)\mu - 2(\overline{r}^{*T}T\Psi^{-1} + q_0^T\Omega^{-1}P)\mu\right)\right\}$$

Finally, we managed to arrive at an equation to which we can again apply **Lemma 3**. With $x = \mu$, $M = T\Psi^{-1} + P^T\Omega^{-1}P$, $b^T = \overline{r}^{*T}T\Psi^{-1} + q_0^T\Omega^{-1}P$, we obtain that the term in the exponential is:

$$\left(\mu - (T\Psi^{-1} + P^T\Omega^{-1}P)^{-1}(T\Psi^{-1}\overline{r}^{*} + P^T\Omega^{-1}q_0)\right)^T (T\Psi^{-1} + P^T\Omega^{-1}P)$$
$$\times \left(\mu - (T\Psi^{-1} + P^T\Omega^{-1}P)^{-1}(T\Psi^{-1}\overline{r}^{*} + P^T\Omega^{-1}q_0)\right)$$
$$-(T\Psi^{-1}\overline{r}^{*} + P^T\Omega^{-1}q_0)^T(T\Psi^{-1} + P^T\Omega^{-1}P)^{-1}(T\Psi^{-1}\overline{r}^{*} + P^T\Omega^{-1}q_0)$$

Since the first term is the only one that depends on μ and since we recognize this to be the kernel of a normal distribution, we eventually obtain that:

$$\mu|\cdot \sim N\left((T\Psi^{-1} + P^T\Omega^{-1}P)^{-1}(T\Psi^{-1}\overline{r}^{*} + P^T\Omega^{-1}q_0), (T\Psi^{-1} + P^T\Omega^{-1}P)^{-1}\right),$$
$$\text{where } \overline{r}^{*} = \frac{\sum_{t=1}^{T}(r_t - \Lambda f_t)}{T}$$

We move next to finding the posterior of Φ:

$$\pi(\Phi|\cdot) \propto \det(\Phi)^{\frac{\nu_0+q+1+T}{2}} \exp\left\{\sum_{t=1}^{T} f_t^T\Phi^{-1}f_t + Tr(R_0\Phi^{-1})\right\}$$

Again, let us focus on the terms in the exponential. By using the fact that $Tr(\cdot)$ is cyclically commutative (as long as dimensions agree), we obtain that:

$$\sum_{t=1}^{T} f_t^T \Phi^{-1} f_t + Tr(R_0 \Phi^{-1}) = \sum_{t=1}^{T} Tr(f_t^T \Phi^{-1} f_t) + Tr(R_0 \Phi^{-1})$$
$$= \sum_{t=1}^{T} Tr(f_t f_t^T \Phi^{-1}) + Tr(R_0 \Phi^{-1}) = Tr\left(\left(R_0 + \sum_{t=1}^{T} f_t f_t^T\right)\Phi^{-1}\right)$$

Therefore, the posterior of Φ is:

$$\Phi|\cdot \sim W^{-1}\left(\nu_0 + T, R_0 + \sum_{t=1}^{T} f_t f_t^T\right)$$

Finally, we are left to compute the posteriors of Λ and Ψ, which we will do in one step. This is because if we let $\tilde{\theta}$ be the vector of all parameters except Λ and Ψ, we obtain that:

$$\pi(\Lambda, \Psi|\tilde{\theta}) = \pi(\Lambda|\Psi, \tilde{\theta})\pi(\Psi|\tilde{\theta}) \tag{31}$$

By looking at the likelihood in equation (30) and by collecting the terms depending on Λ and Ψ, we obtain:

$$\pi(\Lambda, \Psi|\tilde{\theta}) \propto \prod_{k=1}^{n}\left(\Psi_k^{-\frac{T}{2}-\alpha_k-1}\det(\Psi_k H_k)^{-\frac{1}{2}}\right)\exp\left\{-\sum_{k=1}^{n}\frac{\beta_k}{\Psi_k}\right\}$$
$$\times \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(r_t - \mu - \Lambda f_t)^T \Psi^{-1}(r_t - \mu - \Lambda f_t)\right\}$$
$$\times \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}(\Lambda_k - \Lambda_{0k})^T \frac{1}{\Psi_k} H_k^{-1}(\Lambda_k - \Lambda_{0k})\right\}$$

Let us first focus our attention on the last two exponentials. We notice that one sum is over columns (the one over t), while the other sum is over the rows (the one over k). However, we can write the sum over t as a sum over k in the following way:

$$\sum_{t=1}^{T}(r_t - \mu - \Lambda f_t)^T \Psi^{-1}(r_t - \mu - \Lambda f_t) = \sum_{t=1}^{T}\sum_{k=1}^{n}(r_{kt} - \mu_k - f_t^T \Lambda_k)^2 \frac{1}{\Psi_k}$$
$$= \sum_{k=1}^{n}(r_{k\cdot}^T - \mu_k \vec{1} - F^T \Lambda_k)^T \frac{1}{\Psi_k}(r_{k\cdot}^T - \mu_k \vec{1} - F^T \Lambda_k)$$

Notation 4. *Here, we introduced the following notation:*

- $\vec{1} = \begin{bmatrix} 1 & 1 & ... & 1 \end{bmatrix}^T$, *of size* $T \times 1$.
- $r_{k\cdot}$ = *the* k^{th} *row in the matrix of returns* $R = \begin{bmatrix} r_1 & ... & r_T \end{bmatrix}$
- $F = \begin{bmatrix} f_1 & ... & f_T \end{bmatrix}$ *is the matrix in which we have as columns the common factors.*
- μ_k *is the* k^{th} *entry in the vector of means* μ.
- Λ_k *is the* k^{th} *row in the matrix* Λ.

Since we managed to change the summation so that it is with respect to the rows, we can now combine the last two exponentials from the joint posterior density presented above:

$$\pi(\Lambda_k, \Psi_k|\tilde{\theta}) \propto \left(\Psi_k^{-\frac{T}{2}-\alpha_k-1} \det(\Psi_k H_k)^{-\frac{1}{2}}\right) \exp\left\{-\frac{\beta_k}{\Psi_k}\right\}$$
$$\times \exp\left\{-\frac{1}{2}(r_{k\cdot}^T - \mu_k \vec{1} - F^T \Lambda_k)^T \frac{1}{\Psi_k}(r_{k\cdot}^T - \mu_k \vec{1} - F^T \Lambda_k)\right\}$$
$$\times \exp\left\{-\frac{1}{2}(\Lambda_k - \Lambda_{0k})^T \frac{1}{\Psi_k} H_k^{-1}(\Lambda_k - \Lambda_{0k})\right\}$$

We will focus only on the terms in the exponentials:

$$\frac{\beta_k}{\Psi_k} + (r_{k\cdot}^T - \mu_k \vec{1} - F^T \Lambda_k)^T \frac{1}{\Psi_k} (r_{k\cdot}^T - \mu_k \vec{1} - F^T \Lambda_k)$$
$$+(\Lambda_k - \Lambda_{0k})^T \frac{1}{\Psi_k} H_k^{-1} (\Lambda_k - \Lambda_{0k})$$
$$= \frac{\beta_k}{\Psi_k} + \Lambda_k^T F \frac{1}{\Psi_k} F^T \Lambda_k + \Lambda_k \frac{1}{\Psi_k} H_k^{-1} \Lambda_k - 2\Lambda_k^T F \frac{1}{\Psi_k} (r_{k\cdot}^T - \mu_k \vec{1})$$
$$-2\Lambda_k^T \frac{1}{\Psi_k} H_k^{-1} \Lambda_{0k} + (r_{k\cdot}^T - \mu_k \vec{1})^T \frac{1}{\Psi_k} (r_{k\cdot}^T - \mu_k \vec{1}) + \Lambda_{0k}^T \frac{1}{\Psi_k} H_k^{-1} \Lambda_{0k}$$
$$= \Lambda_k^T \frac{1}{\Psi_k} (FF^T + H_k^{-1}) \Lambda_k - 2\Lambda_k \frac{1}{\Psi_k} \left(F(r_{k\cdot}^T - \mu_k \vec{1}) + H_k^{-1} \Lambda_{0k} \right)$$
$$+(r_{k\cdot}^T - \mu_k \vec{1})^T \frac{1}{\Psi_k} (r_{k\cdot}^T - \mu_k \vec{1}) + \Lambda_{0k}^T \frac{1}{\Psi_k} H_k^{-1} \Lambda_{0k} + \frac{\beta_k}{\Psi_k}$$

Since the only terms that depend on Λ_k are the first two, we can focus for now only on them and it will give us the posterior. However, we keep in mind that we still have three other terms remaining in the exponential, which will give us the posterior of Ψ_k (please see (31)). Now, for the first two terms, we can apply again **Lemma 3** for $x = \Lambda_k$, $M = \frac{1}{\Psi_k}(FF^T + H_k^{-1})$, $b = \frac{1}{\Psi_k} \left(F(r_{k\cdot}^T - \mu_k \vec{1}) + H_k^{-1} \Lambda_{0k} \right)$ and we obtain:

$$\left(\Lambda_k - \Psi_k (FF^T + H_k^{-1})^{-1} \frac{1}{\Psi_k} \left(F(r_{k\cdot}^T - \mu_k \vec{1}) + H_k^{-1} \Lambda_{0k} \right) \right)^T$$
$$\times \frac{1}{\Psi_k} (FF^T + H_k^{-1})$$
$$\times \left(\Lambda_k - \Psi_k (FF^T + H_k^{-1})^{-1} \frac{1}{\Psi_k} \left(F(r_{k\cdot}^T - \mu_k \vec{1}) + H_k^{-1} \Lambda_{0k} \right) \right)$$
$$-\frac{1}{\Psi_k} \left(F(r_{k\cdot}^T - \mu_k \vec{1}) + H_k^{-1} \Lambda_{0k} \right)^T \Psi_k (FF^T + H_k^{-1})^{-1} \frac{1}{\Psi_k}$$
$$\times \left(F(r_{k\cdot}^T - \mu_k \vec{1}) + H_k^{-1} \Lambda_{0k} \right) + (r_{k\cdot}^T - \mu_k \vec{1})^T \frac{1}{\Psi_k} (r_{k\cdot}^T - \mu_k \vec{1})$$
$$+\Lambda_{0k}^T \frac{1}{\Psi_k} H_k^{-1} \Lambda_{0k} + \frac{\beta_k}{\Psi_k}$$

Notation 5. *Let us make another notation:*

$$\overline{\Omega}_k = (FF^T + H_k^{-1})^{-1}$$
$$\overline{\mu}_k = \overline{\Omega_k}\left(F(r_{k\cdot}^T - \mu_k\vec{1}) + H_k^{-1}\Lambda_{0k}\right)$$

With the above notation, we finally found the posterior of Λ_k to be:

$$\Lambda_k|\cdot \overset{indep.}{\sim} N(\overline{\mu}_k, \Psi_k\overline{\Omega}_k)$$

All we have left is to put together the last four terms in the above equation and, after noticing that the first term is simply $\frac{1}{\Psi_k}(\overline{\Omega}_k^{\ -1}\overline{\mu}_k)^T\overline{\Omega}_k\overline{\Omega}_k^{\ -1}\overline{\mu}_k = \frac{1}{\Psi_k}\overline{\mu}_k^T\overline{\Omega}_k^{\ -1}\overline{\mu}_k$, we obtain that the posterior of Ψ_k is:

$$\Psi_k|\cdot \overset{indep.}{\sim} IG\left(\alpha_{\Psi_k}, \beta_{\Psi_k}\right), \text{ where}$$
$$\alpha_{\Psi_k} = \frac{T}{2} + \alpha_k$$
$$\beta_{\Psi_k} = \beta_k + \frac{1}{2}\left((r_{k\cdot}^T - \mu_k\vec{1})^T(r_{k\cdot}^T - \mu_k\vec{1}) + \Lambda_{0k}H_k^{-1}\Lambda_{0k} - \overline{\mu}_k{}^T\overline{\Omega}_k^{\ -1}\overline{\mu}_k\right)$$

4.6 Assumptions for Leonard-Hsu prior on covariance of common factors

The only change from the assumptions presented in **Section 4.4** is the prior on Φ. As mentioned previously, we reduce the dimension of the covariance matrix of the returns by introducing a prior on the covariance matrix of the common factors. This has dimension $q \times q$ ($q =$number of factors), which is much smaller than $n \times n$ ($n =$number of stocks). Hence, the only equation that changes in the following set of equations is the last one:

$$
\begin{aligned}
r_t|\mu, f_t, \Lambda, \Psi &\sim N_n(\mu + \Lambda f_t, \Psi), \text{for all } t = \{1, 2, ..., T\} \\
P\mu &\sim N_v(q_0, \Omega) \\
f_t|\Phi &\overset{iid.}{\sim} N_q(0, \Phi) \text{for all } t = \{1, 2, ..., T\} \\
\Lambda_k|\Psi_k &\overset{iid.}{\sim} N_q(\Lambda_{0k}, \Psi_k H_k), \text{for all } k = \{1, 2, ..., n\} \\
\Psi_k &\overset{iid.}{\sim} IG(\alpha_k, \beta_k), \text{for all } k = \{1, 2, ..., n\} \\
Vec^*(log(\Phi))|\theta, \Delta &\sim N(J\theta, \Delta)
\end{aligned}
\tag{32}
$$

Where the variables were introduced in **Notations 1 and 2** and the operator $Vec^*(\cdot)$ was defined in **Definition 1**.

4.7 Posteriors for Leonard-Hsu prior on covariance of common factors

The only change from the version just introduced is that we replace the Inverse-Wishart prior with:

$$\alpha|\theta, \Delta = Vec^*(log(\Phi))|\theta, \Delta \sim N(J\theta, \Delta)$$

The framework is very similar to what was introduced in **Section 3** (more specifically equations (21), (22) and (23), with the common factors playing the role of the returns. We obtain that the posteriors in this case are:

$$
\begin{gathered}
\alpha|\cdot \approx\sim N(\alpha^*, (Q+G)^{-1}), \text{ where } \alpha^* = (Q+G)^{-1}Q\lambda, \\
G = \left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right)^T \Delta^{-1} \left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right)
\end{gathered}
$$

Moreover, the matrix Q is computed in a similar way as before (please see equation 20). If we let e_i, d_i to be the i^{th} normalized eigenvector with its corresponding eigenvalue, respectively, then f_{ij} is obtained by looking at the equation $Vec^*(\log(\Phi))^T f_{ij} = e_i^T \log(\Phi) e_j$ and identifying the coefficients of the entries in the $\log(\Phi)$ matrix. With those f_{ij}, we can finally compute Q:

$$Q = \frac{m}{2}\sum_{i=1}^{q} f_{ii}f_{ii}^T + m\sum_{i<j}^{q} \xi_{ij} f_{ij} f_{ij}^T, \text{ where}$$

$$\xi_{ij} = \frac{(d_i - d_j)^2}{d_i d_j (log(d_i) - log(d_j))^2}$$

Furthermore, the posteriors for σ_1^2 and σ_2^2 are very similar to what we obtained in equations (26). The only difference is that the number of stocks n is replaced by the number of factors q and $d = \frac{1}{2}q(q+1)$:

$$\sigma_1^2|\cdot \sim IG\left(\frac{q-3}{2}, \frac{1}{2}\sum_{i=1}^{q}(\alpha_i - \overline{\alpha}_v)^2\right)$$

$$\sigma_2^2|\cdot \sim IG\left(\frac{d-q-3}{2}, \frac{1}{2}\sum_{i=q+1}^{d}(\alpha_i - \overline{\alpha}_c)^2\right)$$

4.8 Sensitivity Analysis

As mentioned in the introductory **Section 1**, the motivation behind incorporating a factor model to our previous models was spurred by the memory allocation problem that we ran into when using the alternative with a logarithmic prior on the covariance matrix of the returns of the whole $S\&P500$. Therefore, in this section, we will present simulation results only for the model that has a logarithmic prior on Φ.

Furthermore, the Inverse-Wishart version is much simpler to implement. We would just make the observation that the way in which we would specify the parameters of the $\mathcal{W}^{-1}(\nu_0, R_0)$ is similar to our previous alternatives, which did not contain the factor models. The only difference is that in the previous versions this was a distribution on the covariance matrix of the returns, while here it is on the covariance matrix of the common factors f_t. Hence, we will use the historical common factors to estimate ν_0 and R_0. More specifically, we would determine the optimal number of factors for example from a scree plot of eigenvalues vs number of factors. We consider the whole $S\&P500$ consisting of daily returns between 1/2/2014 and 12/29/2017 (the same dataset considered in the results section for our version in which P is not augmented - **Section 2.8** and for our version which has a prior on $log(\Sigma)$ - **Section 3.5)**. We determined that the optimal number of factors for this dataset is $q = 18$. Next, we would fit to this dataset a factor model with $q = 18$ factors, we would take the common factors f_t from the output of the function and we would consider $\hat{R_0} = Cov(f_t)$ and ν_0 =number of f_t's$= T$.

4.9 Implementation for Leonard-Hsu prior on covariance of common factors

Now that we have derived our posteriors, we are ready to implement it, using a Gibbs Sampler. We will use a Metropolis-Hastings algorithm for sampling α, for which we need both the exact posterior distribution and the approximation obtained with the Volterra integral equation. It is an approach introduced by Leonard and Hsu in [10] and also very similar to the one used in **Section 3.4**. The exact distribution is:

$$exp\left\{-\frac{1}{2}\alpha^T G\alpha\right\} det(S)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}(\alpha-\lambda)^T Q(\alpha-\lambda)\right\}$$

This results in the following posterior:

$$\alpha|\cdot \approx\sim N(\alpha^*, (Q+G)^{-1}), \text{ where } \alpha^* = (Q+G)^{-1}Q\lambda$$
$$\pi^*(\alpha|\cdot) \approx\propto exp\left\{-\frac{1}{2}(\alpha-\alpha^*)^T(Q+G)(\alpha-\alpha^*)\right\}$$

This is an approximation since $det(S)^{-\frac{m}{2}} exp\left\{-\frac{1}{2}(\alpha-\lambda)^T Q(\alpha-\lambda)\right\}$ is an approximation of the pdf of a multivariate normal using the Volterra integral equation. If we replace it with the exact distribution, we would obtain:

$$\pi(\alpha|\cdot) \propto exp\left\{-\frac{m}{2}Tr\left(A+Se^{-A}\right) - \frac{1}{2}\alpha^T G\alpha\right\}$$

The Metropolis-Hastings step at t^{th} iteration would be that we would simulate a candidate value from the approximate posterior distribution: $\widetilde{\alpha} \approx\sim N(\alpha^*, (Q+G)^{-1})$ and we would accept it with probability $min(\rho, 1)$, where

$$\rho = \frac{\pi\left(\widetilde{\alpha}|\cdot\right)}{\pi\left(\alpha^{(t)}|\cdot\right)} \cdot \frac{\pi^*\left(\alpha^{(t)}|\cdot\right)}{\pi^*\left(\widetilde{\alpha}|\cdot\right)}$$

It is useful at this point to remember that because of the notation introduced in **Notation 1**, we have a connection between π^* and π since there is one between A and α, namely:

$$\alpha = Vec^*(log(\Phi)), A = log(\Phi)$$

One of the big advantages of using a Bayesian framework is that we do not need good estimates for the initial staring points for the Gibbs Sampler. This is because the Gibbs Sampler is a Markov chain that satisfies Doob's conditions and, therefore, it forgets the initial starting points and, eventually, it converges to the stationary distribution. There is extensive literature that shows, for example, that the sample covariance matrix is a bad estimator (ill-conditioned) when the number of parameters is large in comparison to the amount of data used to estimate it. One of the most famous papers, which also introduces a correction, is Ledoit and Wolf [7].

Therefore, albeit the Gibbs Sampler converges to the same distribution no matter the starting points, we should try to initialize it with good estimates. Also, we have to make sure that we specify the hyper-parameters with values that would make sense in the real world:

- T = number of returns in the historical dataset= number of returns from 1/2/2014 to 12/29/2017.
- $\hat{F_{init}} = \left[\hat{f}_1 \hat{f}_2 ... \hat{f}_T\right]$, where $\hat{f}_t$ for $t = \{1, 2, ..., T\}$ are the common factors obtained by fitting a factor model on the historical dataset with an optimal number of factors of $q = 18$ determined from a scree plot of eigenvalues.
- We also have that $f_t|\Phi \overset{indep.}{\sim} N_q(0, \Phi)$. In order to specify $\hat{\Phi_{init}}$, we take the covariance of the above found common factors: $\hat{\Phi_{init}} = Cov(\hat{f}_t)$.
- We also have the following assumption.

$$\alpha|\theta, \Delta = Vec^*(log(\Phi))|\theta, \Delta \sim N_d \left(\begin{bmatrix} \theta_1 \vec{1}_q \\ \theta_2 \vec{1}_{d-q} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 I_q & \mathbb{O} \\ \mathbb{O} & \sigma_2^2 I_{d-q} \end{bmatrix} \right)$$

- In order to initialize σ_1^2, we have to take the variance of the first q entries in $Vec^*(log(\hat{\Phi_{init}}))$.
- In order to initialize σ_2^2, we have to take the variance of the last $d-q$ entries in $Vec^*(log(\hat{\Phi_{init}}))$.

- We remember that $\Lambda_k|\Psi_k \overset{indep.}{\sim} N_q(\Lambda_{0k}, \Psi_k H_k)$. Since, in general, we do not have any prior information on the factor weights, we specify the hyper-parameters to be:
  - $\hat{\Lambda_{0kinit}} = \vec{0}$
  - We initialize the variance $\Psi_k H_k$ with a big value: $\Psi_{kinit} = 1$ and $H_{kinit} = 10^{10}\mathbb{I}_q$.
- Also, we remember that $\Psi_k \overset{indep.}{\sim} IG(\alpha_k, \beta_k)$, for all $k \in \{1, 2, ..., n\}$. Similarly to the previous point made, in the real world, we do not have any prior information on Ψ_k and this should be reflected in our choice of α_k and β_k. If we let $\alpha_k \to 0$ and $\beta_k \to 0$ in the pdf of the $IG(\alpha_k, \beta_k)$, we notice that we obtain an uninformative prior. Therefore, we initialize $\alpha_{kinit} = \beta_{kinit} = 10^{-10}$.

Using the Metropolis Hastings step that was just discussed, we arrive at the following Gibbs Sampler:

Algorithm 4 Gibbs Sampler $log(\Phi)$

1: $\alpha^{(t+1)} = \begin{cases} \widetilde{\alpha} \sim N\left(\left(Q^{(t)} + G^{(t)}\right)^{-1} Q^{(t)}\lambda^{(t)}, \left(Q^{(t)} + G^{(t)}\right)^{-1}\right) \text{ w.p. } min(\rho, 1) \\ \alpha^{(t)} \text{otherwise} \end{cases}$

2: Since $\alpha = Vec^*(log(\Phi)) \Rightarrow \begin{cases} \text{compute } \Phi^{(t+1)} = exp\left\{Vec^{*-1}\left(\alpha^{(t+1)}\right)\right\} \\ \text{keep } \Phi^{(t)} \end{cases}$

3: $\begin{cases} \sigma_1^{2\,(t+1)} \sim IG\left(\frac{q-3}{2}, \frac{1}{2}\sum_{i=1}^{q}\left(\alpha_i{}^{(t+1)} - \overline{\alpha}_v{}^{(t+1)}\right)^2\right) \\ \sigma_2^{2\,(t+1)} \sim IG\left(\frac{d-q-3}{2}, \frac{1}{2}\sum_{i=q+1}^{d}\left(\alpha_i{}^{(t+1)} - \overline{\alpha}_c{}^{(t+1)}\right)^2\right) \end{cases} \Rightarrow$

$$\Rightarrow \Delta^{(t+1)} = \begin{bmatrix} \sigma_1^{2\,(t+1)} I_q & \mathbb{O} \\ \mathbb{O} & \sigma_2^{2\,(t+1)} I_{d-q} \end{bmatrix}$$

4: Let $\Sigma_\mu = \left(T\Psi^{(t)^{-1}} + P^T\Omega^{-1}P\right)^{-1} \Rightarrow \mu^{(t+1)} \sim N\left(\Sigma_\mu\left(T\Psi^{(t)^{-1}}\overline{r}^{*(t)} + P^T\Omega^{-1}q_0\right), \Sigma_\mu\right)$, where $\overline{r}^{*(t)} = \frac{1}{T}\sum_{i=1}^{T}\left(r_i - \Lambda^{(t)}f_i^{(t)}\right)$.

5: Let

$$\Sigma_f = \left(\Lambda^{(t)^T}\Psi^{(t)^{-1}}\Lambda^{(t)} + \Phi^{(t+1)^{-1}}\right)^{-1} \Rightarrow f_i^{(t+1)} \sim N\left(\Sigma_f\Lambda^{(t)^T}\Psi^{(t)^{-1}}\left(r_i - \mu^{(t+1)}\right), \Sigma_f\right)$$

6:

$$\Psi_k^{(t+1)} \overset{indep.}{\sim} IG\left(\alpha_{\Psi_k}, \beta_{\Psi_k}\right), \text{where}$$

$$\alpha_{\Psi_k} = \frac{T}{2} + \alpha_k$$

$$\beta_{\Psi_k} = \beta_k + \frac{1}{2}\left(\left(r_{k\cdot}^T - \mu_k^{(t+1)}\vec{1}\right)^T\left(r_{k\cdot}^T - \mu_k^{(t+1)}\vec{1}\right) + \Lambda_{0k}H_k^{-1}\Lambda_{0k} - \overline{\mu}_k{}^{(t+1)^T}\overline{\Omega}_k{}^{(t+1)^{-1}}\overline{\mu}_k{}^{(t+1)}\right)$$

$$\overline{\Omega}_k{}^{(t+1)} = \left(F^{(t+1)}F^{(t+1)^T} + H_k^{-1}\right)^{-1}$$

$$\overline{\mu}_k{}^{(t+1)} = \overline{\Omega}_k{}^{(t+1)}\left(F^{(t+1)}\left(r_{k\cdot}^T - \mu_k^{(t+1)}\vec{1}\right) + H_k^{-1}\Lambda_{0k}\overline{\Omega}_k{}^{(t+1)}\right)$$

7: $\Lambda_k^{(t+1)} \overset{indep.}{\sim} N\left(\overline{\mu}_k{}^{(t+1)}, \Psi_k^{(t+1)}\overline{\Omega}_k{}^{(t+1)}\right)$

8: Compute $S_f^{(t+1)} = \frac{\sum_{i=1}^{T} f_i f_i^T}{T}$, $\lambda^{(t+1)} = Vec^*\left(log\left(S_f^{(t+1)}\right)\right)$, $d_j{}^{(t+1)}$ and $e_j{}^{(t+1)}$ the eigenvalue and normalized eigenvector of $S_f^{(t+1)}$ respectively.

9: Compute $f_{ij}^{(t+1)}$ by identifying the coefficients of the entries of the $log(\Phi)$ matrix from the equation $Vec^*\left(log\left(\Phi^{(t+1)}\right)\right)^T f_{ij}{}^{(t+1)} = e_i{}^{(t+1)^T} log\left(\Phi^{(t+1)}\right) e_j{}^{(t+1)}$

10: Compute $\xi_{ij}^{(t+1)} = \frac{\left(d_i{}^{(t+1)} - d_j{}^{(t+1)}\right)^2}{d_i{}^{(t+1)}d_j{}^{(t+1)}\left(log\left(d_i{}^{(t+1)}\right) - log\left(d_j{}^{(t+1)}\right)\right)^2}$

11: Compute $Q^{(t+1)} = \frac{T}{2}\sum_{i=1}^{q} f_{ii}{}^{(t+1)}f_{ii}{}^{(t+1)^T} + T\sum_{i<j}^{q}\xi_{ij}{}^{(t+1)}f_{ij}{}^{(t+1)}f_{ij}{}^{(t+1)^T}$

12: Compute

$$G^{(t+1)} = \left(I_d - J(J^T\Delta^{(t+1)^{-1}}J)^{-1}J^T\Delta^{(t+1)^{-1}}\right)^T\Delta^{(t+1)^{-1}} \times$$
$$\times\left(I_d - J(J^T\Delta^{(t+1)^{-1}}J)^{-1}J^T\Delta^{(t+1)^{-1}}\right)$$

4.10 Results-Personal Views on 4 Stocks

In this section we will present the results of sensitivity analysis identical to some presented in our previous results subsections. This is because we would like to check if the extensions introduced behave similarly. We will depict the sensitivity of the model to changes in confidence levels (ω_i) in terms of both the distance of the posterior to investor's view and the profits obtained if one would use this model to trade. An analysis will be presented in the next section, when many industry sectors from $S\&P500$ will be involved in the investor's views (for a brief introduction to industry sectors, please see Appendix C).

Before we delve into the actual results for this version of the model, we notice that **Remarks 3 and 4** both hold. Basically, this means that the smaller the diagonal entries in Ω, the more confident we are in the views because we have the assumption that $P\mu \sim N(q_0, \Omega)$. Same assumption points out the fact that the smaller Ω is, the closer $P\mu$ should be to q_0. Hence, a very small Ω shows the fact that the investor is very confident in this view and, therefore, the posterior should also be close to q_0. Therefore, the smaller our Ω is, the closer $P\mu_{\text{post}}$ should be to q_0. We will consider 2 views and do an exhaustive search over possible combinations of pairs of values for the 2 diagonal entries of Ω (which are depicted as 2 axis) and compute the distance:$||P\mu_{\text{post}} - q_0||$ (which is depicted as 1 axis).

Just like in the results **Section 3.5**, we chose to have views for the same 4 stocks (AAPL, FB, GOOG, MSFT), and we will use the same data set: daily returns from 1/2/2014 to 12/29/2017. We will use the following inputs (again the columns are in order AAPL, FB, GOOG, MSFT and the rows represent the views). Please notice that the matrix P in our implementation has a lot more columns (one for each stock actively traded in $S\&P500$), but the vast majority of the entries are 0:

$$q_0 = \begin{bmatrix} 0.02 \\ 0.05 \end{bmatrix}, P = \begin{array}{c|cccc} & \text{AAPL} & \text{FB} & \text{GOOG} & \text{MSFT} \\ \text{view1} & -1 & 1 & 0 & 0 \\ \text{view2} & 0 & 0 & 1 & -1 \end{array}$$

Albeit the memory allocation problem encountered in our previous extensions was solved, the one presented in this paper is far more computationally expensive since we have to sample from more distributions. Therefore, again the exhaustive search was ran in parallel on multiple cores (each core running the Gibbs Sampler for 1 pair (ω_1, ω_2), which was split into 16 different ranges, each one running 6 simulations on an evenly split grid). The burn period is 10^3 and the iteration period is 10^4, just as they were in **Section 3.5**.

In the following plot, 2 of the axis are represented by the two confidence levels (ω_1 and ω_2) and the third one is represented by the distance $|P\mu_{\text{post}} - q_0|$. As mentioned previously (**Remarks 3 and 4**), this distance should go to 0 as ω_1 and ω_2 go to 0, which can easily be observed in the following figure:

Furthermore, similarly to the versions introduced previously, as ω_1 and ω_2 increase, the distance converges to the same number. Since ω_1 and ω_2 are standard deviations, a high standard deviation represents a lack of confidence in the personal views inputted. Therefore, intuitively, the model should only take into consideration the history. This is precisely how the model behaves. If we just consider the historical returns, the unbiased estimator for μ is the sample mean of the returns, $\bar{r}$. The distance

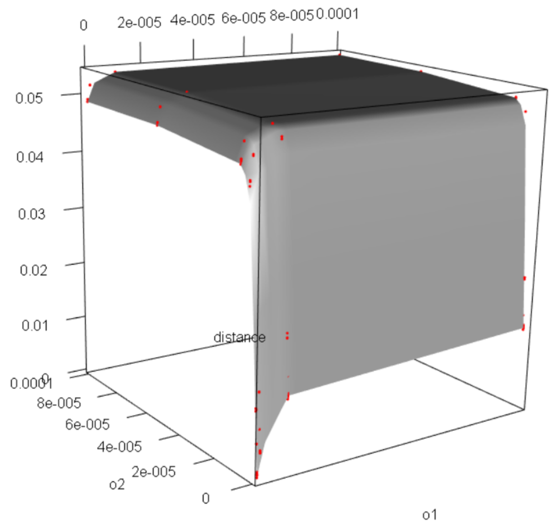

Figure 20: Distance when considering only 4 stocks

$|P\overline{r} - q_0| = 0.05386381$, which is the level at which the curve in the presented picture flattens.

One could use this model to hold a portfolio over a testing dataset consisting of the daily returns during the month of January 2018 with an initial starting capital of $\$100,000$. We remember that in order to obtain portfolio weights, we estimate from the Gibbs Sampler $\Sigma_{post} = \Lambda_{post}\Phi_{post}{\Lambda_{post}}^T + \Psi_{post}$ and we try to maximize the portfolio returns, while minimizing the portfolio risk. Hence, we would like to find $\max_w w^T\mu_{post} - \frac{\lambda}{2}w^T\Sigma_{post}w$, where λ is the investor's risk aversion coefficient. In his paper [6], Janecek suggests that $\lambda = 2.5$ is a reasonable choice for equities. By making the derivative with respect to w equal to 0, and by solving the resulting equation for w, we obtain: $w^* = \frac{1}{2.5}\Sigma_{post}^{-1}\mu_{post}$. The profits without considering any fees on a testing dataset consisting of the returns over the month of January 2018 for all the previously mentioned combinations of confidence levels

(ω_1 and ω_2) averaged \$40,075.87 with a standard deviation of \$14,373.88

4.11 Results-Personal Views on Industry Sectors

In this section, we will present similar results to the ones presented in the previous section. We will have the exact same training and testing datasets as before. The only change is in the personal views inputted in our model. However, this time we would like to enter personal views about different industry sectors.

In order to have good personal views and not just random guesses, as we have done so far, we will use the weighting recommendations provided by CFRA[2], an independent fundamental and forensic investment research firm. Each stock within the same sector receives equal weight that sum up to 1, with a positive weight for the ones outperforming and a negative weight for the ones under-performing. We will have the following 4 personal views (for details on which companies are in each industry sector, please consult Appendix C):

(1) *Information technology* outperforms *utilities* by 0.2% with confidence level ω_1.

(2) *Energy* outperforms *industrials* by 0.1% with confidence level ω_2.

(3) *Real Estate* outperforms *consumer staples* by 0.2% with confidence level ω_2.

(4) *Consumer discretionary* outperforms *financials* by 0.3% with confidence level ω_2.

[2]CFRA: Fidelity Investments link

Hence, we have that $q_0 = (0.002, 0.001, 0.002, 0.003)^T$.

We picked 2 distinct confidence levels for the 4 views simply because we wanted to have another 3D plot with 2 of the axis represented by ω_1 and ω_2 and the third axis represented by the distance $||P\mu_{post} - q_0||$. Just as before, the exhaustive search was ran in parallel on multiple cores. The same burning and iteration periods were used also.

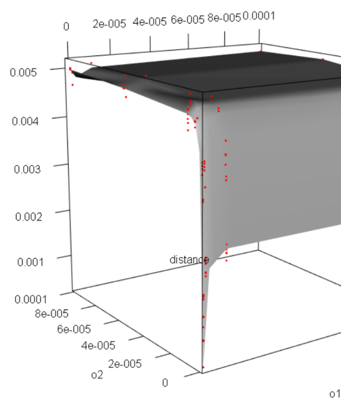

Figure 21: Distance when considering industry sectors

Again, the model behaves exactly as our intuition and as **Remarks 3 and 4** would suggest. As ω_1 and ω_2 go to 0, the distance $||P\mu_{post} - q_0||$ converges to 0. Moreover, for bigger values of ω_1 and ω_2 (small confidence in views) the distance converges to $0.004999748 = ||P\overline{r} - q_0||$. This confirms our intuition that the less confident the investor is in his or her views, the more the model takes into consideration the history.

Moving on to presenting the profits, we used the same starting capital of $\$100,000$, the same testing dataset over the month of January 2018 and the same methodology for computing the portfolio weights. The mean of the profits over all the simulated pairs (ω_1, ω_2) was $\$37,576.68$ with a standard deviation of $5,857.198.

A

Proof of Approximation using Volterra Integrals

As mentioned before, Bellman in his book *Introduction to Matrix Analysis* shows an even more general result than what we need. The matrix exponential $X(t) = e^{(A_0+cB_0)t}$ satisfies the Volterra integral equation:

$$X(t) = e^{A_0 t} + c\int_0^t e^{A_0(t-s)} B_0 X(s) ds, 0 < t < \infty$$

Now if we let in the above equation $A_0 = -\Lambda$, $B_0 = \Lambda - A$, $c = 1$ and remembering that $\Lambda = log(S)$ we obtain:

$$X(t) = S^{-t} - \int_0^t S^{s-t}(A - \Lambda)X(v)dv, 0 < t < \infty,$$

Since we want to approximate e^{-A}, we let in the above equation $t = 1$ and we repeatedly replace X:

$$\begin{aligned}
e^{-A} = X(1) = S^{-1} - \int_0^1 S^{s-1}(A-\Lambda)S^{-s}ds \\
= S^{-1} - \int_0^1 S^{s-1}(A-\Lambda)\left(S^{-s} - \int_0^s S^{u-s}(A-\Lambda)X(u)du\right)ds \\
= S^{-1} - \int_0^1 S^{s-1}(A-\Lambda)S^{-s}ds + \int_0^1\int_0^s S^{s-1}(A-\Lambda)S^{u-s}(A-\Lambda) \\
\times\left(S^{-u} - \int_0^u S^{v-u}(A-\lambda)X(v)dv\right)duds \\
\approx S^{-1} - \int_0^1 S^{s-1}(A-\Lambda)S^{-s}ds \\
+\int_0^1\int_0^s S^{s-1}(A-\Lambda)S^{u-s}(A-\Lambda)S^{-u}duds
\end{aligned}$$

Where this is an approximation because the triple and higher order integrals were ignored. The conditional pdf of the returns is:

$$f(r_1, ..., r_m|\mu, \Sigma) \propto \exp\left\{-\frac{m}{2}Tr(A + Se^{-A})\right\}$$

Hence, from the Volterra approximation, by multiplying by S and taking the trace, we obtain:

$$Tr(Se^{-A}) \approx n - \int_0^1 Tr\left(S^s(A-\Lambda)S^{-s}\right)ds$$
$$+\int_0^1\int_0^s Tr\left(S^s(A-\Lambda)S^{u-s}(A-\Lambda)S^{-u}duds\right)$$

The first integral is easier to compute:

$$\int_0^1 Tr\left(S^s(A-\Lambda)S^{-s}\right)ds = \int_0^1 Tr\left(A-\Lambda\right)ds = Tr\left(A-\Lambda\right)$$

The second integral requires more calculations. Before we delve into them, let us write the spectral decomposition of S as $S = E_0D_0E_0^T$. If we define the matrix *log* through the Taylor series expansion, and by suing the fact that E_0 is orthonormal, we obtain that the spectral decomposition of $log(S)$ is $\Lambda = log(S) = E_0 log(D_0)E_0^T$. Also, let us make another notation: $B = E_0^T(A-\Lambda)E_0 \Rightarrow E_0BE_0^T = A - \Lambda$:

$$Tr\left(S^s(A-\Lambda)S^{u-s}(A-\Lambda)S^{-u}\right) = Tr\left((A-\Lambda)S^{u-s}(A-\Lambda)S^{-(u-s)}\right)$$
$$= Tr\left(E_0BD_0^{u-s}BD_0^{-(u-s)}E_0^T\right) = Tr\left(BD_0^{u-s}BD_0^{-(u-s)}\right)$$

In order to compute the integral of this *Trace* term, we will try to put it in scalar form:

$$BD_0^{u-s} = \begin{bmatrix} b_{11}d_1^{u-s} & b_{12}d_2^{u-s} & ... & b_{1n}d_n^{u-s} \\ : & : & ... & : \\ b_{n1}d_1^{u-s} & b_{12}d_2^{u-s} & ... & b_{1n}d_n^{u-s} \end{bmatrix}$$

For the matrix $BD_0^{-(u-s)}$ we obtain a similar result, the only difference is that d_i^{u-s} is replaced by $\frac{1}{d_i^{u-s}}$. Also, from the spectral decomposition, please note that d_i are the eigenvalues of S.

Since we need the $Tr\left(BD_0^{u-s}BD_0^{-(u-s)}\right)$, we will only compute the diagonal entries of this matrix:

$$diag\left(BD_0^{u-s}BD_0^{-(u-s)}\right) =$$
$$\begin{bmatrix} b_{11}^2 + b_{12}b_{21}\left(\frac{d_2}{d_1}\right)^{u-s} + b_{13}b_{31}\left(\frac{d_3}{d_1}\right)^{u-s} + \cdots + b_{1n}b_{n1}\left(\frac{d_n}{d_1}\right)^{u-s} \\ b_{21}b_{12}\left(\frac{d_1}{d_2}\right)^{u-s} + b_{22}^2 + \cdots + b_{2n}b_{n2}\left(\frac{d_n}{d_2}\right)^{u-s} \\ : \\ b_{n1}b_{1n}\left(\frac{d_n}{d_1}\right)^{u-s} + b_{n2}b_{2n}\left(\frac{d_n}{d_2}\right)^{u-s} + \cdots + b_{nn}^2 \end{bmatrix}$$

But we know that B is symmetric. Therefore, we obtain that:

$$\int_0^1 \int_0^s Tr\left(BD_0^{u-s}BD_0^{-(u-s)}\right) duds = \sum_{i=1}^n \int_0^1 \int_0^s b_{ii}^2 duds$$

$$+\sum_{i\neq j}^n \int_0^1 \int_0^s b_{ij}^2 \left(\frac{d_i}{d_j}\right)^{u-s} duds, \text{ where we have that}$$

$$\sum_{i=1}^n \int_0^1 \int_0^s b_{ii}^2 duds = \sum_{i=1}^n \frac{b_{ii}^2}{2} \text{ and also}$$

$$\sum_{i\neq j}^n \int_0^1 \int_0^s b_{ij}^2 \left(\frac{d_i}{d_j}\right)^{u-s} duds = \sum_{i\neq j}^n \int_0^1 b_{ij}^2 \left(\frac{d_i}{d_j}\right)^{u-s}$$

$$\times \frac{1}{\log(d_i)-\log(d_j)}\Big|_0^s ds = \sum_{i\neq j} \frac{b_{ij}^2}{\log(d_i)-\log(d_j)} \int_0^1 1-\left(\frac{d_i}{d_j}\right)^{-s} ds$$

$$= \sum_{i\neq j} \frac{b_{ij}^2}{\log(d_i)-\log(d_j)} \left(1-\left(\frac{d_j}{d_i}\right)^s \frac{1}{\log(d_j)-\log(d_i)}\right)\Big|_0^1$$

$$= \sum_{i\neq j} \frac{b_{ij}^2}{\log(d_i)-\log(d_j)} + \sum_{i\neq j} b_{ij}^2 \frac{\frac{d_j}{d_i}-1}{(\log(d_i)-\log(d_j))^2}$$

$$= \sum_{i<j} \left(\frac{b_{ij}^2}{\log(d_i)-\log(d_j)} + \frac{b_{ji}^2}{\log(d_j)-\log(d_i)}\right)$$

$$+\sum_{i<j} b_{ij}^2 \frac{\frac{d_j}{d_i}+\frac{d_i}{d_j}-2}{(\log(d_i)-\log(d_j))^2} = 0 + \sum_{i<j} b_{ij}^2 \frac{\frac{d_j}{d_i}+\frac{d_i}{d_j}-2}{(\log(d_i)-\log(d_j))^2}$$

Finally, by adding the two double integrals, we obtain that

$$\int_0^1 \int_0^s Tr\left(BD_0^{u-s}BD_0^{-(u-s)}\right) duds = n - Tr(A) + Tr(\Lambda) +$$

$$+\frac{1}{2}\sum_{i=0}^n b_{ii}^2 + \sum_{i<j} b_{ij}^2 \frac{\frac{d_j}{d_i}+\frac{d_i}{d_j}-2}{(\log(d_i)-\log(d_j))^2}$$

With the notation of the ξ_{ij} introduced in the paper, we obtain the Volterra approximation represented by equation (19).

B

Proof of Proposition 1

The following equality holds:

$$f(\alpha|\sigma_1^2,\sigma_2^2) = \int_\theta det(\Delta)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(\alpha - J\theta)^T\Delta^{-1}(\alpha - J\theta)\right\} d\theta$$
$$= 2\pi det(\Delta)^{-\frac{1}{2}} det(J^T\Delta^{-1}J)^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\alpha^T G\alpha\right\}, \text{ where}$$
$$G = \left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right)^T \Delta^{-1}\left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right)$$

Proof. Before we actually attempt to compute the integral, we would like to put all the quantities in scalar form since this would make our life easier. This brings us to the following two lemmas:

Lemma 4. $det(\Delta)^{-\frac{1}{2}} det(J^T\Delta^{-1}J)^{-\frac{1}{2}} = \frac{1}{\sqrt{n(d-n)}}\left(\sigma_1^2\right)^{-\frac{n-1}{2}}\left(\sigma_2^2\right)^{-\frac{d-n-1}{2}}$

Proof.

$$J^T\Delta^{-1}J = \begin{bmatrix} \frac{1}{\sigma_1^2} & \dots & \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \dots & 0 & \frac{1}{\sigma_2^2} & \dots & \frac{1}{\sigma_2^2} \end{bmatrix} J = \begin{bmatrix} \frac{n}{\sigma_1^2} & 0 \\ 0 & \frac{d-n}{\sigma_2^2} \end{bmatrix}$$

Hence, we obtain that $det(J^T\Delta^{-1}J)^{-\frac{1}{2}} = \frac{1}{\sqrt{n(d-n)}}\left(\sigma_1^2\right)^{\frac{1}{2}}\left(\sigma_2^2\right)^{\frac{1}{2}}$ Also, clearly since Δ is diagonal, we obtain that:

$$det(\Delta)^{-\frac{1}{2}} = \left(\sigma_1^2\right)^{-\frac{n}{2}}\left(\sigma_2^2\right)^{-\frac{d-n}{2}}$$

Multiplying the two determinants, we obtain the desired result. □

Now let us turn our attention to writing in scalar form the term in the exponential:

Lemma 5. $\alpha^T G\alpha = \frac{1}{\sigma_1^2}\sum_{i=1}^{n}(\alpha_i - \overline{\alpha_v})^2 + \frac{1}{\sigma_2^2}\sum_{i=n+1}^{d}(\alpha_i - \overline{\alpha_c})^2$, *where* $\overline{\alpha_v}$ *is the average of the* $\alpha' s$ *on the main diagonal (i.e. those that originate from the log of the variance terms of the returns) and* $\overline{\alpha_c}$ *is the average of all the* α*'s that are on the off diagonal (i.e. those that originate from the log of the covariance terms of the returns).*

Proof. First of all, one can notice that the formula for G can be simplified for calculation purposes:

$$
\begin{aligned}
G &= \left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right)^T \Delta^{-1}\left(I_d - J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}\right) \\
&= \Delta^{-1} - \Delta^{-1}J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1} - \Delta^{-1}J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1} \\
&+\Delta^{-1}J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1} = \Delta^{-1} - \Delta^{-1}J(J^T\Delta^{-1}J)^{-1}J^T\Delta^{-1}
\end{aligned}
$$

We remember that we have computed $J^T\Delta^{-1}J$ in **Lemma 4**:

$$J^T \Delta^{-1} J = \begin{bmatrix} \frac{n}{\sigma_1^2} & 0 \\ 0 & \frac{d-n}{\sigma_2^2} \end{bmatrix} \text{ and } \Delta^{-1} J = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ \vdots & \vdots \\ \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \\ \vdots & \vdots \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix}$$

$$\Rightarrow \Delta^{-1} J \left(J^T \Delta^{-1} J \right) = \begin{bmatrix} \frac{1}{n} & 0 \\ \vdots & \vdots \\ \frac{1}{n} & 0 \\ 0 & \frac{1}{d-n} \\ \vdots & \vdots \\ 0 & \frac{1}{d-n} \end{bmatrix}$$

$$\Rightarrow \Delta^{-1} J \left(J^T \Delta^{-1} J \right) J^T \Delta^{-1} = \begin{bmatrix} \frac{1}{n\sigma_1^2} & ... & \frac{1}{n\sigma_1^2} & 0 & ... & 0 \\ \vdots & ... & \vdots & \vdots & ... & \vdots \\ \frac{1}{n\sigma_1^2} & ... & \frac{1}{n\sigma_1^2} & 0 & ... & 0 \\ 0 & ... & 0 & \frac{1}{(d-n)\sigma_2^2} & ... & \frac{1}{(d-n)\sigma_2^2} \\ \vdots & ... & \vdots & \vdots & ... & \vdots \\ 0 & ... & 0 & \frac{1}{(d-n)\sigma_2^2} & ... & \frac{1}{(d-n)\sigma_2^2} \end{bmatrix}$$

Now we just have to subtract this matrix from Δ^{-1}, which is just diagonal, and we can finally compute the desired quantity:

$$\alpha^T G\alpha = \sum_{1\le i\ne j\le n} \frac{1}{n\sigma_1^2}\alpha_i\alpha_j + \sum_{i=1}^{n} \frac{n-1}{n\sigma_1^2}\alpha_i^2 + \sum_{n+1\le i\ne j\le d} \frac{1}{(d-n)\sigma_2^2}\alpha_i\alpha_j + \sum_{i=n+1}^{d} \frac{d-n-1}{(d-n)\sigma_2^2}\alpha_i^2$$

By looking at this equation and the one that we have to prove, we realize that if we would manage to show the following identity, we would also prove the lemma:

$$\sum_{1\le i\ne j\le n} \frac{1}{n\sigma_1^2}\alpha_i\alpha_j + \sum_{i=1}^{n} \frac{n-1}{n\sigma_1^2}\alpha_i^2 = \frac{1}{\sigma_1^2}\sum_{i=1}^{n}(\alpha_i - \overline{\alpha_v})^2$$

Let us start from the right hand side:

$$\sum_{i=1}^{n}(\alpha_i - \overline{\alpha_v})^2 = \sum_{i=1}^{n}\alpha_i^2 - n\overline{\alpha}^2 = \sum_{i=1}^{n}\alpha_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}\alpha_i\right)^2$$
$$= \sum_{i=1}^{n}\alpha_i^2 - \frac{1}{n}\sum_{i=1}^{n}\alpha_i^2 - \frac{1}{n}\sum_{i\ne j}\alpha_i\alpha_j = \sum_{i=1}^{n}\frac{n-1}{n}\alpha_i^2 - \frac{1}{n}\sum_{i\ne j}\alpha_i\alpha_j$$

□

Now we finally have all the necessary identities to write the integral in our proposition in scalar form. We would have to prove that:

$$\int_{\theta_1} exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\theta_1)^2\right\}d\theta_1 \int_{\theta_2} exp\left\{-\frac{1}{2\sigma_2^2}\sum_{i=n+1}^{q_0}(\alpha_i-\theta_2)^2\right\}d\theta_2$$
$$= 2\pi\frac{\sigma_1}{\sqrt{n}}exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v})^2\right\}\frac{\sigma_2}{\sqrt{d-n}}exp\left\{-\frac{1}{2\sigma_2^2}\sum_{i=n+1}^{q_0}(\alpha_i-\overline{\alpha_c})^2\right\}$$

Hence, if we manage to show the following identity, we would manage to prove the proposition also:

$$\int_{\theta_1} exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\theta_1)^2\right\}d\theta_1 = \sqrt{2\pi}\frac{\sigma_1}{\sqrt{n}}exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v})^2\right\}$$

Let us start from the left hand side and subtract and add the average $\overline{\alpha_v}$ in each term of the sum from the exponential:

$$\begin{aligned}
LHS &= \int_{\theta_1} exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\theta_1)^2\right\}d\theta_1 \\
&= \int_{\theta_1} exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v}+\overline{\alpha_v}-\theta_1)^2\right\}d\theta_1 \\
&= exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v})^2\right\} \\
&\times \int_{\theta_1} exp\left\{-\frac{1}{2\sigma_1^2}\left(\sum_{i=1}^{n}(\overline{\alpha_v}-\theta_1)^2+2\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v})(\overline{\alpha_v}-\theta_1)\right)\right\}d\theta_1 \\
&= exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v})^2\right\} \\
&\times \int_{\theta_1} exp\left\{-\frac{1}{2\sigma_1^2}\left(n(\overline{\alpha_v}-\theta_1)^2+2(\overline{\alpha_v}-\theta_1)\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v})\right)\right\}d\theta_1 \\
&= exp\left\{-\frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(\alpha_i-\overline{\alpha_v})^2\right\}\int_{\theta_1} exp\left\{-\frac{1}{2\left(\frac{\sigma_1}{\sqrt{n}}\right)^2}(\overline{\alpha_v}-\theta_1)^2\right\}d\theta_1
\end{aligned}$$

Now we recognize that the term inside the integral is close to the density of a normal distribution. Hence, this gives us the idea of doing the change of variables:

$$y_1 = \frac{\theta_1 - \overline{\alpha_v}}{\frac{\sigma_1}{\sqrt{n}}} \Rightarrow dy_1 = \frac{\sqrt{n}}{\sigma_1} d\theta_1 \Rightarrow d\theta_1 = \frac{\sigma_1}{\sqrt{n}} dy_1$$

$$\Rightarrow LHS = \sqrt{2\pi} exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{n} (\alpha_i - \overline{\alpha_v})^2 \right\} \int_{\theta_1} \frac{1}{\sqrt{2\pi}} exp \left\{ -\frac{1}{2} y_1^2 \right\} \frac{\sigma_1}{\sqrt{n}} dy_1$$

$$= \sqrt{2\pi} \frac{\sigma_1}{\sqrt{n}} exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{i=1}^{n} (\alpha_i - \overline{\alpha_v})^2 \right\}$$

As mentioned before, a similar identity can be showed for the second integral that depends solely on θ_2 and this completes the proof of the proposition. $\square$

C

$S\&P500$ Industry Sectors

The stocks in the $S\&P500$ are divided into broad groupings based on economic characteristics. Currently there are 11 industry sectors[3]

- Communication Services: from telephone access to high-speed internet, this sector of the economy keeps us all connected.
- Consumer Discretionary: businesses that have demand that rises and falls based on general economic conditions such as washers and dryers, sporting goods, new cars, and diamond engagement rings
- Consumer Staples: businesses that sell the necessities of life, ranging from bleach and laundry detergent to toothpaste and packaged food.
- Energy: businesses that source, drill, extract, and refine the raw commodities we need to keep the country going, such as oil and gas.
- Financials: banks, insurance companies, real estate investment trusts, credit card issuers, and a host of other money-centric enterprises that keep the debits and credits of the economy flowing.
- Health Care: drug companies, medical supply companies, and other scientific-based operations that are concerned with improving and healing human life.
- Industrials: from railroads and airlines to military weapons and industrial conglomerates.

[3] According to thebalance.com

- Information Technology: hardware, software, computer equipment, and IT services operations.

- Materials sector manufacturers, logs, and mines everything from precious metals, paper, and chemicals to shipping containers, wood pulp, and industrial ore.

- Real Estate: all Real Estate Investment Trusts (REITs) with the exception of Mortgage REITs, which is housed under the financial sector. The sector also includes companies that manage and develop properties.

- Utilities sector is home to the firms that make our lights work when we flip the switch, let our stoves erupt in flame when we want to cook food, make water come out of the tap when we are thirsty, and more.

In this section, we will show specifically which companies were considered in each one of the industry sectors from our 4 personal views introduced in **Section 4.11**.

Information Technology					
Count	Symbol	Name	Count	Symbol	Name
1	ACN	Accenture plc	35	INTC	Intel Corp.
2	ATVI	Activision Blizzard	36	IBM	International Business Machines
3	ADBE	Adobe Systems Inc	37	INTU	Intuit Inc.
4	AMD	Advanced Micro Devices Inc	38	IPGP	IPG Photonics Corp.
5	AKAM	Akamai Technologies Inc	39	JNPR	Juniper Networks
6	ADS	Alliance Data Systems	40	KLAC	KLA-Tencor Corp.
7	GOOGL	Alphabet Inc Class A	41	LRCX	Lam Research
8	GOOG	Alphabet Inc Class C	42	MA	Mastercard Inc.
9	APH	Amphenol Corp	43	MCHP	Microchip Technology
10	ADI	Analog Devices, Inc.	44	MU	Micron Technology
11	ANSS	ANSYS	45	MSFT	Microsoft Corp.
12	AAPL	Apple Inc.	46	MSI	Motorola Solutions Inc.
13	AMAT	Applied Materials Inc.	47	NTAP	NetApp
14	ADSK	Autodesk Inc.	48	NFLX	Netflix Inc.
15	ADP	Automatic Data Processing	49	NVDA	Nvidia Corporation
16	AVGO	Broadcom	50	ORCL	Oracle Corp.
17	CA	CA, Inc.	51	PAYX	Paychex Inc.
18	CDNS	Cadence Design Systems	52	QCOM	QUALCOMM Inc.
19	CSCO	Cisco Systems	53	RHT	Red Hat Inc.
20	CTXS	Citrix Systems	54	CRM	Salesforce.com
21	CTSH	Cognizant Technology Solutions	55	STX	Seagate Technology
22	GLW	Corning Inc.	56	SWKS	Skyworks Solutions
23	DXC	DXC Technology	57	SYMC	Symantec Corp.
24	EBAY	eBay Inc.	58	SNPS	Synopsys Inc.
25	EA	Electronic Arts	59	TTWO	Take-Two Interactive
26	FFIV	F5 Networks	60	TEL	TE Connectivity Ltd.
27	FB	Facebook, Inc.	61	TXN	Texas Instruments
28	FIS	Fidelity National Information Services	62	TSS	Total System Services
29	FISV	Fiserv Inc	63	VRSN	Verisign Inc.
30	FLIR	FLIR Systems	64	V	Visa Inc.
31	IT	Gartner Inc	65	WDC	Western Digital
32	GPN	Global Payments Inc.	66	WU	Western Union Co
33	HRS	Harris Corporation	67	XRX	Xerox Corp.
34	HPQ	HP Inc.	68	XLNX	Xilinx Inc

Table 2: Information Technology stocks

Energy					
Count	Symbol	Name	Count	Symbol	Name
1	APC	Anadarko Petroleum Corp	17	KMI	Kinder Morgan
2	ANDV	Andeavor	18	MRO	Marathon Oil Corp.
3	APA	Apache Corporation	19	MPC	Marathon Petroleum
4	BHGE	Baker Hughes, a GE Company	20	NOV	National Oilwell Varco Inc.
5	COG	Cabot Oil & Gas	21	NFX	Newfield Exploration Co
6	CVX	Chevron Corp.	22	NBL	Noble Energy Inc
7	XEC	Cimarex Energy	23	OXY	Occidental Petroleum
8	CXO	Concho Resources	24	OKE	ONEOK
9	COP	ConocoPhillips	25	PSX	Phillips 66
10	DVN	Devon Energy Corp.	26	PXD	Pioneer Natural Resources
11	EOG	EOG Resources	27	RRC	Range Resources Corp.
12	EQT	EQT Corporation	28	SLB	Schlumberger Ltd.
13	XOM	Exxon Mobil Corp.	29	FTI	TechnipFMC
14	HAL	Halliburton Co.	30	VLO	Valero Energy
15	HP	Helmerich & Payne	31	WMB	Williams Cos.
16	HES	Hess Corporation			

Table 3: Energy stocks

Consumer Staples					
Count	Symbol	Name	Count	Symbol	Name
1	MO	Altria Group Inc	17	HRL	Hormel Foods Corp.
2	ADM	Archer-Daniels-Midland Co	18	SJM	JM Smucker
3	CPB	Campbell Soup	19	K	Kellogg Co.
4	CHD	Church & Dwight	20	KMB	Kimberly-Clark
5	CLX	The Clorox Company	21	KR	Kroger Co.
6	KO	Coca-Cola Company (The)	22	MKC	McCormick & Co.
7	CL	Colgate-Palmolive	23	TAP	Molson Coors Brewing Company
8	CAG	Conagra Brands	24	MDLZ	Mondelez International
9	STZ	Constellation Brands	25	MNST	Monster Beverage
10	COST	Costco Wholesale Corp.	26	PEP	PepsiCo Inc.
11	COTY	Coty, Inc	27	PM	Philip Morris International
12	CVS	CVS Health	28	PG	Procter & Gamble
13	DPS	Dr Pepper Snapple Group	29	SYY	Sysco Corp.
14	EL	Estee Lauder Cos.	30	TSN	Tyson Foods
15	GIS	General Mills	31	WMT	Wal-Mart Stores
16	HSY	The Hershey Company	32	WBA	Walgreens Boots Alliance

Table 4: Consumer Staples stocks

Financials					
Count	Symbol	Name	Count	Symbol	Name
1	AMG	Affiliated Managers Group Inc	33	JPM	JPMorgan Chase & Co.
2	AFL	AFLAC Inc	34	KEY	KeyCorp
3	ALL	Allstate Corp	35	LNC	Lincoln National
4	AXP	American Express Co	36	L	Loews Corp.
5	AIG	American International Group, Inc.	37	MTB	M&T Bank Corp.
6	AON	Aon plc	38	MMC	Marsh & McLennan
7	AJG	Arthur J. Gallagher & Co.	39	MET	MetLife Inc.
8	AIZ	Assurant Inc.	40	MCO	Moody's Corp
9	BAC	Bank of America Corp	41	MS	Morgan Stanley
10	BK	The Bank of New York Mellon Corp.	42	NDAQ	Nasdaq, Inc.
11	BBT	BB&T Corporation	43	NTRS	Northern Trust Corp.
12	BLK	BlackRock	44	PBCT	People's United Financial
13	HRB	Block H&R	45	PNC	PNC Financial Services
14	BHF	Brighthouse Financial Inc	46	PFG	Principal Financial Group
15	COF	Capital One Financial	47	PGR	Progressive Corp.
16	CBOE	Cboe Global Markets	48	PRU	Prudential Financial
17	SCHW	Charles Schwab Corporation	49	RJF	Raymond James Financial Inc.
18	CB	Chubb Limited	50	RF	Regions Financial Corp.
19	CINF	Cincinnati Financial	51	SPGI	S&P Global, Inc.
20	C	Citigroup Inc.	52	STT	State Street Corp.
21	CME	CME Group Inc.	53	STI	SunTrust Banks
22	CMA	Comerica Inc.	54	SIVB	SVB Financial
23	DFS	Discover Financial Services	55	TROW	T. Rowe Price Group
24	ETFC	E*Trade	56	TMK	Torchmark Corp.
25	RE	Everest Re Group Ltd.	57	TRV	The Travelers Companies Inc.
26	FITB	Fifth Third Bancorp	58	USB	U.S. Bancorp
27	BEN	Franklin Resources	59	UNM	Unum Group
28	GS	Goldman Sachs Group	60	WFC	Wells Fargo
29	HIG	Hartford Financial Svc.Gp.	61	WLTW	Willis Towers Watson
30	HBAN	Huntington Bancshares	62	XL	XL Capital
31	ICE	Intercontinental Exchange	63	ZION	Zions Bancorp
32	IVZ	Invesco Ltd.			

Table 5: Financials stocks

Utilities					
Count	Symbol	Name	Count	Symbol	Name
1	AES	AES Corp	14	EXC	Exelon Corp.
2	LNT	Alliant Energy Corp	15	FE	FirstEnergy Corp
3	AEP	American Electric Power	16	NEE	NextEra Energy
4	AWK	American Water Works Company Inc	17	NI	NiSource Inc.
5	CNP	CenterPoint Energy	18	NRG	NRG Energy
6	CMS	CMS Energy	19	PCG	PG&E Corp.
7	ED	Consolidated Edison	20	PNW	Pinnacle West Capital
8	D	Dominion Energy	21	PEG	Public Serv. Enterprise Inc.
9	DTE	DTE Energy Co.	22	SCG	SCANA Corp
10	DUK	Duke Energy	23	SRE	Sempra Energy
11	EIX	Edison Int'l	24	SO	Southern Co.
12	ETR	Entergy Corp.	25	WEC	Wec Energy Group Inc
13	ES	Eversource Energy	26	XEL	Xcel Energy Inc

Table 6: Utilities stocks

Industrials					
Count	Symbol	Name	Count	Symbol	Name
1	MMM	3M Company	33	IR	Ingersoll-Rand PLC
2	AYI	Acuity Brands Inc	34	JEC	Jacobs Engineering Group
3	ALK	Alaska Air Group Inc	35	JBHT	J. B. Hunt Transport Services
4	ALLE	Allegion	36	JCI	Johnson Controls International
5	AAL	American Airlines Group	37	KSU	Kansas City Southern
6	AME	AMETEK Inc.	38	LLL	L-3 Communications Holdings
7	AOS	A.O. Smith Corp	39	LMT	Lockheed Martin Corp.
8	ARNC	Arconic Inc.	40	MAS	Masco Corp.
9	BA	Boeing Company	41	NLSN	Nielsen Holdings
10	CHRW	C. H. Robinson Worldwide	42	NSC	Norfolk Southern Corp.
11	CAT	Caterpillar Inc.	43	NOC	Northrop Grumman Corp.
12	CTAS	Cintas Corporation	44	PCAR	PACCAR Inc.
13	CSX	CSX Corp.	45	PH	Parker-Hannifin
14	CMI	Cummins Inc.	46	PNR	Pentair Ltd.
15	DE	Deere & Co.	47	PWR	Quanta Services Inc.
16	DAL	Delta Air Lines Inc.	48	RTN	Raytheon Co.
17	DOV	Dover Corp.	49	RHI	Robert Half International
18	ETN	Eaton Corporation	50	ROK	Rockwell Automation Inc.
19	EMR	Emerson Electric Company	51	COL	Rockwell Collins
20	EFX	Equifax Inc.	52	ROP	Roper Technologies
21	EXPD	Expeditors International	53	LUV	Southwest Airline
22	FAST	Fastenal Co	54	SRCL	Stericycle Inc
23	FDX	FedEx Corporation	55	TXT	Textron Inc.
24	FLS	Flowserve Corporation	56	TDG	TransDigm Group
25	FLR	Fluor Corp.	57	UNP	Union Pacific
26	FBHS	Fortune Brands Home & Security	58	UAL	United Continental Holdings
27	GD	General Dynamics	59	UPS	United Parcel Service
28	GE	General Electric	60	URI	United Rentals, Inc.
29	GWW	Grainger (W.W.) Inc.	61	UTX	United Technologies
30	HON	Honeywell Int'l Inc.	62	VRSK	Verisk Analytics
31	HII	Huntington Ingalls Industries	63	WM	Waste Management Inc.
32	ITW	Illinois Tool Works	64	XYL	Xylem Inc.

Table 7: Industrials stocks

Real Estate					
Count	Symbol	Name	Count	Symbol	Name
1	ARE	Alexandria Real Estate Equities Inc	17	IRM	Iron Mountain Incorporated
2	AMT	American Tower Corp A	18	KIM	Kimco Realty
3	AIV	Apartment Investment & Management	19	MAC	Macerich
4	AVB	AvalonBay Communities, Inc.	20	MAA	Mid-America Apartments
5	BXP	Boston Properties	21	PLD	Prologis
6	CCI	Crown Castle International Corp.	22	PSA	Public Storage
7	DLR	Digital Realty Trust Inc	23	O	Realty Income Corporation
8	DRE	Duke Realty Corp	24	REG	Regency Centers Corporation
9	EQIX	Equinix	25	SBAC	SBA Communications
10	EQR	Equity Residential	26	SPG	Simon Property Group Inc
11	ESS	Essex Property Trust, Inc.	27	SLG	SL Green Realty
12	EXR	Extra Space Storage	28	VTR	Ventas Inc
13	FRT	Federal Realty Investment Trust	29	VNO	Vornado Realty Trust
14	GGP	General Growth Properties Inc.	30	WELL	Welltower Inc.
15	HCP	HCP Inc.	31	WY	Weyerhaeuser Corp.
16	HST	Host Hotels & Resorts			

Table 8: Real Estate stocks

Consumer Discretionary					
Count	Symbol	Name	Count	Symbol	Name
1	AAP	Advance Auto Parts	39	M	Macy's Inc.
2	AMZN	Amazon.com Inc.	40	MAR	Marriott Int'l.
3	APTV	Aptiv Plc	41	MAT	Mattel Inc.
4	AZO	AutoZone Inc	42	MCD	McDonald's Corp.
5	BBY	Best Buy Co. Inc.	43	MGM	MGM Resorts International
6	BWA	BorgWarner	44	KORS	Michael Kors Holdings
7	KMX	Carmax Inc	45	MHK	Mohawk Industries
8	CCL	Carnival Corp.	46	NWL	Newell Brands
9	CBS	CBS Corp.	47	NWSA	News Corp. Class A
10	CHTR	Charter Communications	48	NWS	News Corp. Class B
11	CMG	Chipotle Mexican Grill	49	NKE	Nike
12	CMCSA	Comcast Corp.	50	JWN	Nordstrom
13	DHI	D. R. Horton	51	NCLH	Norwegian Cruise Line
14	DRI	Darden Restaurants	52	ORLY	O'Reilly Automotive
15	DISCA	Discovery Inc. Class A	53	OMC	Omnicom Group
16	DISCK	Discovery Inc. Class C	54	RL	Polo Ralph Lauren Corp.
17	DISH	Dish Network	55	PHM	Pulte Homes Inc.
18	DG	Dollar General	56	PVH	PVH Corp.
19	DLTR	Dollar Tree	57	ROST	Ross Stores
20	EXPE	Expedia Inc.	58	RCL	Royal Caribbean Cruises Ltd
21	FL	Foot Locker Inc	59	SNA	Snap-On Inc.
22	F	Ford Motor	60	SWK	Stanley Black & Decker
23	GRMN	Garmin Ltd.	61	SBUX	Starbucks Corp.
24	GM	General Motors	62	TPR	Tapestry, Inc.
25	GPC	Genuine Parts	63	TGT	Target Corp.
26	GT	Goodyear Tire & Rubber	64	TIF	Tiffany & Co.
27	HBI	Hanesbrands Inc	65	TJX	TJX Companies Inc.
28	HOG	Harley-Davidson	66	TSCO	Tractor Supply Company
29	HAS	Hasbro Inc.	67	TRIP	TripAdvisor
30	HLT	Hilton Worldwide Holdings Inc	68	FOXA	Twenty-First Century Fox Class A
31	HD	Home Depot	69	FOX	Twenty-First Century Fox Class B
32	IPG	Interpublic Group	70	ULTA	Ulta Beauty
33	KSS	Kohl's Corp.	71	UAA	Under Armour Class A
34	LB	L Brands Inc.	72	VFC	V.F. Corp.
35	LEG	Leggett & Platt	73	VIAB	Viacom Inc.
36	LEN	Lennar Corp.	74	DIS	The Walt Disney Company
37	LKQ	LKQ Corporation	75	WHR	Whirlpool Corp.
38	LOW	Lowe's Cos.	76	YUM	Yum! Brands Inc

Table 9: Consumer Discretionary stocks

References

[1] Alvarez, I., Niemi, J. and Simpson, M. (Jul. 2016) arXiv:1408.4050v2, `https://arxiv.org/pdf/1408.4050.pdf`

[2] Baum, L.E., Katz, M. and Read, R.R. (Feb 1962) *Convergence rates for the Law of Large Numbers*. American Mathematical Society, Vol. 102, No. 2, pp. 187-199

[3] Bellman, R. (1970) *Introduction to Matrix Analysis*. McGraw-Hill, New York.

[4] Cheng, Ing-Haw (2008) *CAPM and Black-Litterman* Princeton University, URL: `http://www.dartmouth.edu/~icheng/capm_BL.pdf`

[5] Guangliang, H. and Litterman, R. (Oct. 2002) *The Intuition Behind Black-Litterman Model Portfolios*. SSRN, 22 Oct. 2002, Web.

[6] Janecek, K. (2004) *What is a Realistic Aversion to Risk for Real-world Individual Investors*. Semanticsscholar, Web.

[7] Ledoit, O. and Wolf, M. (2004) *A well-conditioned estimator for large-dimensional covariance matrices* Journal of Multivariate Analysis, vol. 88, pp. 365-411

[8] Lee, S.Y., Poon, W.Y. and Song X.Y. (2007) *Bayesian Analysis of the Factor Model with Finance Applications*. Quantitative Finance, 7:3, 343-356, DOI:10.1080/14697680601009838.

[9] Lee, S.Y., Shi J.Q. (2000) *Joint Bayesian Analysis of Factor Scores and Structural Parameters in the Factor Analysis Model*. Ann.Inst.Statist.Math., Vol. 52, No. 4, 722-736

[10] Leonard, T. and Hsu, John S.J. (Dec. 1992) *Bayesian Inference for a Covariance Matrix*. The Annals of Statistics, vol. 20, no. 4, pp. 1669-1696

[11] Leonard, T. and Hsu, John S.J. (1999) *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge UK, Cambridge UP.