Dealing with Data: An Empirical Analysis of Bayesian Black-Litterman Model Extensions

Daniel Eller Roeder

Professor Andrew Patton, Economics Faculty Advisor
Professor Scott Schmidler, Statistical Science Faculty Advisor

---

Contents

Acknowledgements

I would like to thank my thesis advisors, Professors Scott Schmidler (Statistics) and Andrew Patton (Economics) for all their help throughout this thesis. I would also like to thank Professors Jerry Reiter, Kent Kimbrough, and Allison Hagy for their help in constructing the thesis. A special thanks goes out to my academic curriculum advisors, Professors Emma Rasiel and Dalene Stangl, whose guidance was invaluable throughout my time at Duke. Mike Cerneant at Global Financial Data also provided excellent advice throughout the data collection and cleaning process.

Most importantly I would like to thank my family, Sandra Eller, Greg Roeder and Rebecca Roeder. Your love, guidance and support has made me who I am today.

Abstract

Portfolio Optimization is a common financial econometric application that draws on various types of statistical methods. The goal of portfolio optimization is to determine the ideal allocation of assets to a given set of possible investments. Many optimization models use classical statistical methods, which do not fully account for estimation risk in historical returns or the stochastic nature of future returns. By using a fully Bayesian analysis, however, this analysis is able to account for these aspects and also incorporate a complete information set as a basis for the investment decision. The information set is made up of the market equilibrium, an investor/expert's personal views, and the historical data on the assets in question. All of these inputs are quantified and Bayesian methods are used to combine them into a succinct portfolio optimization model. For the empirical analysis, the model is tested using monthly return data on stock indices from Australia, Canada, France, Germany, Japan, the U.K. and the U.S.

Keywords: Bayesian Analysis, Mean-Variance Portfolio Optimization, Global Markets

JEL Classification: C1, C11, C58, G11

1   Introduction

Portfolio optimization is one of the fastest growing areas of research in financial econometrics. Only recently has computing power reached a level where analysis on numerous assets is even possible and in the post-crisis economy investors are looking for safer and more proven investment methods, which are exactly what financial models provide. Quantitative investment methods have already begun to take over the market and will only continue to rise in popularity as they become a prerequisite for investment profitability.

There are a number of portfolio optimization models used in financial econometrics and many of them build on aspects of previously defined models. The models defined in this paper combine insights from Markowitz (1952), Black and Litterman (1992) and Zhou (2009). Each of these papers use techniques from the previous one to specify and create a novel modeling technique.

The Markowitz model, often referred to as a mean-variance analysis, uses estimates of the next period's mean return vector and covariance matrix to specify the investment portfolio. Markowitz (1952) uses the historical mean and covariance matrix to estimate these inputs. The model is quite sensitive to any changes in the data inputs and often advises extremely long or short positions in assets, which can be problematic for an investor.

The Black-Litterman (BL) model uses information from the market equilibrium and an investor's personal views to estimate the mean and covariance matrix. Many investors make investment decisions based on how they view the market or a certain asset, so this extension is quite practical. Semi-Bayesian methods are employed by Black and Litterman (1992), but no historical data is used which makes the model inherently not Bayesian.

Bayesian statistical methods specify a few types of functions that are necessary to complete an analysis: the prior distribution, the likelihood function, and the posterior distribution. The prior distribution defines how one expects a certain variable to be distributed before viewing any data. The likelihood function describes the observed data in the study. The posterior distribution is the combination of the prior distribution with the likelihood function and defines the new distribution of a given variable under the prior and the likelihood. The prior is combined with the likelihood by using Bayes theorem, which multiplies the prior times the posterior and divides by the normalizing constant.[1] Prior distributions can be of different weights in the posterior distribution depending on how confident one is in the prior. Bayesian analysis is an ideal method to use in a portfolio optimization problem because investors can estimate how the market will perform in the prior under their own beliefs, and then update those beliefs with actual information.

All of the necessary Bayesian components are incorporated in the model presented by Zhou (2009); the BL estimates act as a joint prior and the historical data defines the likelihood function. This strengthens the analysis by making it mostly consistent with Bayesian principles, though some aspects are still not met. The Zhou model uses the historical covariance matrix in each stage of the analysis (prior and likelihood), which is not a sound Bayesian

---

[1]Bayes Theorem: $P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)\,\mathrm{d}\theta}$

application. The true next period covariance matrix is never observable to an investor, meaning there is inherent uncertainty in estimating the covariance matrix, which must be accounted for in the model. The Zhou model underestimates this uncertainty by using the historical covariance matrix in both the prior and likelihood. This method puts too much confidence in the historical estimate of the next period's covariance.

In the models I propose, I will account for this uncertainty by incorporating an inverse-Wishart prior distribution on the covariance matrix, which originally models the covariance as a distribution and not a point estimate. The inverse-Wishart prior uses the original prior covariance matrix as a starting point, but the investor can now model the covariance matrix as a distribution and adjust confidence in the starting point through a tuning parameter. The capital asset pricing model (CAPM) specified covariance matrix is also employed in the first Bayesian updating stage (in two of my extended models) to avoid the double updating problem. These calculations serve as extensions that must be incorporated to make the model statistically sound, as well as a starting point for more extensive analysis of the covariance matrix.

In my extensions the inverse-Wishart prior is applied to either the equilibrium covariance matrix[2] in the first Bayesian updating stage, or to the BL specified prior in the second Bayesian updating stage. There are therefore four extended models under this application since there are two options for the placement of the prior and two options for the equilibrium covariance matrix. The normality assumption of returns is upheld in these models, meaning the inverse-Wishart prior only affects the evaluation of the covariance matrix, not the mean returns. The model that uses the inverse-Wishart prior on the BL estimates and the historical covariance matrix as the equilibrium estimate performs the best, and even outperforms the Zhou model when the parameter inputs are specified correctly. The other models are still useful, however, particularly in theory and as applied to other investment settings.

The final extension presented in this paper uses a full normal-inverse-Wishart prior on the BL prior estimates, derived from the historical covariance matrix as the equilibrium estimate.[3] The normal-inverse-Wishart prior imposes a normal prior on the mean returns and an inverse-Wishart prior on the covariance matrix. The normality assumption of predictive returns is no longer upheld since the new predictive distribution follows a Student-t distribution. Under Standard Bayesian analysis the posterior predictive distribution should be maximized with respect to the investor's utility. However, this thesis is concerned with analyzing the inputs of the models, not the optimization methods. Therefore, the standard mean-variance formula will be used to calculate portfolio weights for the normal-inverse-Wishart prior extension.

The empirical analysis in Zhou (2009) is based on equity index returns from Australia, Canada, France, Germany, Japan, the United Kingdom and the United States. The dataset in this analysis is comprised of total return indices for the same countries, but the data spans through 2013 instead of 2007 as in Zhou (2009). My dataset is also similar to the one chosen

---

[2]Depending on the extended model, the equilibrium covariance matrix is either defined through the historical covariance matrix or the CAPM covariance matrix.

[3]This is the only model used for the full prior extension since it was proven to perform the best under the inverse-Wishart prior extension.

by Black and Litterman (1992), which was picked in order to analyze different international trading strategies in the equity, bond and currency markets. In my empirical analysis all the models will be tested under my dataset.

The goal of this paper is to extend the Zhou model by relaxing the assumptions on the modeling of the covariance matrix. From this a statistically sound and flexible model is created, usable by any type of investor.

In section 2, the literature on the topic is described in detail. Section 3 defines the baseline and extended models. In Section 4 the dataset is described and descriptive statistics are provided. In section 5 the model implementation method is presented. Section 6 presents and interprets the results and in Section 7 conclusions and possible further extensions are offered.

2   Literature Review

2.1   Models

2.1.1   Markowitz

Harry Markowitz established one of the first frameworks for portfolio optimization in 1952. In his paper "Portfolio Selection", Markowitz calculates the portfolio weights that maximize a portfolio's return (while minimizing the volatility), by maximizing a specified utility function for the investor. The utility function is based on the next period return, μ, and covariance matrix, Σ. The historical moments, μ_h and Σ_h, are used to estimate these values. μ_h and Σ_h are the only inputs so the model tends to be extremely sensitive to changes in either variable.

The sensitivity of the model with regard to the historical inputs is problematic for a couple of reasons. A portfolio that must be constantly updated lends itself to large transaction costs, which diminishes the overall profitability of the model. A small deviation in the expected return vector could cause the model to suggest an extremely long and/or short position (in the case of no constraints), and an investor must pay a fee in order to make such an investment. As the model updates itself, it constantly advises new positions and the investor must keep paying transaction costs in order to keep up.[4] Such extreme positions are also at odds with conventional diversification strategies, such as the equal investment ($\frac{1}{N}$) strategy that invests in all assets equally. In fact, the historical mean-variance model is often outperformed by the equal investment portfolio (Jobson and Korkie, 1981).

The Markowitz model is also unable to account for estimation error in the values of the historical means and variances since they are the only inputs. Estimation error can be better accounted for by using Bayesian methods to specify a distribution on the inputs, as

---

[4]There are models that can directly account for transaction costs, see Pogue (1970) for an example.

well as by using multiple model inputs to calculate the next period mean, μ, and covariance matrix, Σ.

Markowitz (1952) assumes that the returns are independent, identical and normally distributed (i.i.d.), with mean μ_h and covariance Σ_h under his mean-variance optimization model.

2.1.2 Black-Litterman

The difficulties with the Markowitz mean-variance model do not render it useless. In fact, when there are better estimates of μ and Σ (rather than just the historical data), it can perform quite well. Black and Litterman (1992) extend the mean-variance framework by creating an estimation strategy that combines an equilibrium model of asset performance, specified under the assumptions of the CAPM, with the investor's views on the assets in the portfolio. Investors frequently make decisions about their portfolio based on how they expect the market to perform, so it is intuitive to incorporate these views into the model.

The equilibrium model is used to specify a neutral starting point that the investor can adjust using specific views. Many assumptions must be made to calculate an equilibrium set of returns. Black and Litterman (1992) assume that the CAPM holds, that investors have the same views on the market and risk aversion, and that demand equals supply in equilibrium. The weakest of these assumptions is that all investors have the same views, which is unlikely on an individual level. However, when the market is considered holistically, as it should be in an equilibrium sense, this assumption is not as flawed. Due to the common usage of analyst equity reports across the market, many investors do indeed have similar (if not identical) views on assets.

Investor views in the BL model can either be absolute or relative. Absolute views specify the expected return for an individual security; for example, an investor may think that the S&P 500 will return 2% in the next period. Relative views specify the relationship between assets; for example, an investor may think that the London Stock Exchange will have a return 2% higher than the Toronto Stock Exchange in the next period. Views are incorporated in the model through Bayesian updating of the equilibrium estimates. This returns a vector of expected returns that is similar to the market equilibrium but adjusted with respect to the investor's views. The BL portfolio weights only differ from the equilibrium weights for assets that the investor has a view on. The estimate of Σ is also calculated using Bayesian updating methods.

The same mean variance utility function is used by Black and Litterman (1992) as by Markowitz (1952) to calculate the optimal portfolio weights. The utility function inputs are the updated BL expected returns, μ_{BL}, and covariance matrix, Σ_{BL}. The same assumptions are also specified by Black and Litterman (1992) as by Markowitz (1952).

2.1.3 Zhou

The BL framework is taken one step further by Zhou (2009) through the incorporation of historical returns in a second Bayesian updating stage. Through this update Zhou (2009) calculates a new mean estimate, μ_z, and the covariance matrix estimate, Σ_z. Zhou (2009) cites two specific benefits of this extension. First, the equilibrium market weights are subject to error that the data can help fix. The market equilibrium values are based on the validity of the CAPM, which is not always supported by historical data.[5] This does not render the equilibrium model useless; it simply must be supplemented by historical data in order to make the model more robust. The combination of the data with the BL prior is assumed to strengthen the model by combining different means of prediction.

The second benefit of incorporating historical data is that the historical mean returns, μ_h, can play a useful role in determining future stock returns. This is essentially an extension of the last benefit, but now sample means are specifically referenced instead of general trends in the data. It is possible that the equilibrium expected returns could be drastically different from μ_h. If this is the case, the equilibrium model is clearly incomplete so it would be naïve of an investor to not incorporate μ_h when calculating future expected returns. In summary, Zhou (2009) states that there are three elements available to the investor in the portfolio optimization decision problem: the equilibrium model, the investor's views, and the data, and that all of them should be used in the portfolio optimization model.

A very complete description of the market is used in Zhou (2009) through the incorporation of three estimates, but there is an aspect of the model that is neglected: the theoretical framework does not account for uncertainty in the estimate of Σ. It is implied that Σ is described only by Σ_h, as it is the only estimate used within each Bayesian updating stage. This repeated use of Σ_h is similar to the limited inputs problem that arises in Markowitz (1952). The repetitive use is also not sound in a Bayesian statistical sense because the same data used in the likelihood is used to generate the prior. In my analysis, an inverse-Wishart prior distribution is put on Σ, to account for uncertainty in estimation.

The historical covariance matrix does an increasingly worse job in estimating Σ for larger values of N (the number of assets in the portfolio). In my analysis N = 7, which is relatively small, but given the theoretical nature of the paper the final model should be generalizable to larger values of N. To improve the generalizability of the model, Σ_{CAPM} is used in two extended models as a unique method of covariance specification across the first Bayesian updating stage. Σ_{CAPM} is a very simple model, however, and is just an example of a model that may improve the specification of Σ when N is large. An interesting topic of further research lies in other predictive models of Σ that can be employed when N is large.

The mean-variance model was one of the first portfolio optimization models created and is the basis for many different portfolio optimization models in use today. Though the model lacks complexity, it follows a basic decision analysis framework that can be used in models that are infinitely more complex. As in any decision problem, the decision maker

---

[5]For more information regarding the choice of the market equilibrium model, see Black and Litterman (1992).

wants to maximize utility based on the information set. In the mean-variance portfolio optimization problem, the decision maker, the investor, seeks to maximize returns while also minimizing volatility. Though utility functions may be different for individual investors, the real difference in portfolio optimization models arise from the data available to the investor. In my analysis the market equilibrium returns, the investor's views, and the historical data make up the information set, but the method of combining the information set is changed as compared to the Zhou model.

2.2 Investment Strategies

Although the BL model is quantitatively based, it is extremely flexible due to the input of subjective views by the investor. These views are directly specified and can come from any source, whether that is a hunch, the Wall Street Journal, or maybe even an entirely different quantitative model. In the models I propose, a momentum strategy is used to specify the views. This is only one of countless different strategies that could be used, whether they are quantitatively based or not. Based on the nature of the model, the results in this paper are heavily dependent on the view specification. However, the goal of this paper is not to have a perfect empirical analysis, but instead to present a flexible, statistically sound and customizable model that is applicable to any type of investor.

2.2.1 Momentum Strategy

A function based on the recent price movement of the indices, a momentum strategy, is used to specify the investor views in the model. This is in contrast to the conventional investing wisdom that individual asset prices and their movements are unrelated to the asset's value. However, when the correct time frame is analyzed, generally the previous 6-12 months, statistically significant returns can be achieved (Berger et al., 2009). This phenomenon holds for all types of assets, from U.S. stocks to foreign currencies, and has been backed up by extensive research. In the last 5 years alone, over 150 papers have been published investigating and proving the momentum effect (Berger et al., 2009). Foreign indices are not an exception to this phenomenon, as it has been shown that indices with positive momentum perform better than those with negative momentum (Asness et al., 1997).

Though a momentum strategy may seem like far-fetched idea to those who have learned standard investing practice, the intuition behind the momentum effect is almost as strong as the statistically significant results. Berger et al. (2009) present a few behavioral explanations that may help to explain the momentum effect.

Assuming the efficient market hypothesis holds,[6] momentum must be explained by some inefficiency in the incorporation of information or the market in general. There are many behavioral explanations that have been put forth in favor of momentum. One is that certain investors are quicker to respond to new information than others. A hedge fund is obviously

---

[6]The efficient market hypothesis states that all public information about an asset is immediately incorporated into the price of the asset, making it essentially impossible to beat the market.

better equipped to respond quickly to information than an investor who reads the Wall Street Journal every week, so it is illogical to believe that all information is fully and immediately incorporated. If instead it is believed that new information is gradually incorporated as more investors learn of it, the momentum effect is an intuitive extension.

The individual *anchoring effect* is analogous to the unequal dissemination of information explained above. Rather than looking at the incorporation of information across the economy, the anchoring effect instead hypothesizes that many investors only partially incorporate new information into their portfolio at first, while continuing to analyze the asset over time. Only after this further analysis will many investors actually make changes to their portfolio. This individually slow incorporation of information is therefore another behavioral argument in favor of momentum investing.

The two phenomena explained above are based on specific aspects of the economy as well as conscious decisions by investors. However, as humans, investors are prone to certain biases that can alter their investment decisions. One of which is the *disposition effect* which states that investors often sell assets too early in order to guarantee returns and keep assets too long in order to avoid losses. This means that good news on a stock may not be incorporated immediately since the ensuing selling by investors will lower the price. On the contrary, when investors keep stocks they should be selling, the price decreases in a more gradual fashion. This disposition effect again slows down the incorporation of information, providing an even stronger basis for momentum.

One final behavioral explanation is referred to as the *bandwagon effect*. When a stock price starts to rise, investors want to jump on the bandwagon with everyone else so they buy the stock, causing the stock price to go even higher. The opposite explanation also holds for the selling of stocks when they perform poorly. The root cause of this phenomenon is the opposite of the above examples, because now investors are essentially incorporating non-existent information, causing the stock to rise or fall an artificially large amount before correcting itself. This often lasts for a few months before the correction occurs, which is in line with the definition of a momentum strategy (Berger et al., 2009).

All of these explanations are quite plausible and there continues to be much discussion about what causes momentum. What cannot be argued, however, is that there is indeed a momentum effect. Those who fail to exploit it are simply missing what could be defined as a rare arbitrage opportunity. The momentum strategy that I employ gives an investor a simple strategy to follow without putting undue weight on what is still an undoubtedly aggressive investment strategy. The other aspects of the Bayesian model, the market equilibrium and historical data, help to shrink the portfolio weights towards what many would consider to be more reliable estimates.

2.2.2 General Investment Strategies

Though momentum investing is gaining in popularity, there are countless other investment strategies in use today. Value investing attempts to target stocks that are undervalued in the market, so that while the investor is holding them, the market will correct the mis-

pricing which gives the investor a positive return. There are many metrics, like the Price-Earnings and Price-Book ratios, which investors look at in trying to determine if a stock is under-valued. Though value investing is not practical in my empirical analysis (given the international index dataset), the general investor would consider having company equity in his portfolio, and writing a function that incorporates these statistics in specifying views would not be difficult. The same holds for other common investment strategies, like growth investing, where investors target companies that they expect will soon experience significant growth in their business operations and therefore likely an increased stock price.

The potential incorporation of other data to predict stock performance is an interesting topic, particularly in the Zhou model and associated extensions, because there are two different ways an investor could incorporate it. The first is through the use of a separate forecasting model that can be incorporated within the investor's views. This is the method I use in my analysis through the momentum strategy. A second option, as referenced by Zhou (2009), is through the use a return forecasting function instead of the historical returns in the data updating stage of the model. However, this would make the model considerably more complex through a loss of conjugacy in the likelihood updating stage. Many investors, and even skilled portfolio managers, are not necessarily quantitatively advanced, so there is no reason to further complicate the model when an almost identical function can be incorporated in a much simpler manner. Using the simple data generated likelihood function also allows the investor to specifically account for historical returns in the data updating stage, and theoretically there are benefits to this type of analysis.

3   Theoretical Framework

This section further explains Bayesian analysis before presenting the models created by Markowitz (1952), Black and Litterman (1992) and Zhou (2009). Finally, the extended models are presented.

3.1   Bayesian Analysis

The models presented by Black and Litterman (1992) and Zhou (2009), along with my extended models, use Bayesian methods and in this sub-section I will present the general steps of a predictive Bayesian analysis.

The first step in any Bayesian analysis is to define the prior, $P(\theta)$.[7] The likelihood function must be specified next and is defined as $L(\theta; \Phi)$, where Φ represents the data used in the likelihood function.[8] The posterior distribution is calculated as

---

[7]Let $\theta = (\mu, \Sigma)$, the two unknown next period moments that must be modeled in a mean-variance optimization.

[8]In standard Bayesian analysis, this would be the historical or collected data. However, in the BL model the likelihood function is defined by the investor views.

$$P(\theta|\Phi) \propto P(\theta)\,L(\theta;\Phi)\,. \tag{1}$$

The normalizing constant is not included in (1) because each model in this paper uses prior distributions that are conjugate to the likelihood function. The use of a conjugate prior dictates that the posterior distribution is of the same family as the likelihood function, but with updated parameters. In Black and Litterman (1992) and Zhou (2009), conjugate multivariate normal distributions are used, and in my extended normal-inverse-Wishart model, conjugate normal-inverse-Wishart distributions are used.

The posterior predictive distribution is calculated to account for the inherent uncertainty of prediction. It is calculated by,

$$P(r_{T+1}|\Phi) = \int_{\theta} P(r_{T+1}|\theta, \Phi)\,P(\theta|\Phi)\,d\theta\,, \tag{2}$$

where r_{T+1} represents the next period's expected return. θ is integrated out of the posterior predictive distribution since it represents the true next period values of μ and Σ which are never known to the investor.

The final step of the general Bayesian model is to maximize the investor's utility under the posterior predictive distribution of the next period returns. The maximization problem is solved by,

$$\max_{w} = \int_{\theta} U(w_{T+1})\,P(r_{T+1}|\Phi)\,d\,r_{T+1}\,, \tag{3}$$

where $U(w_{T+1})$ represents the investor's utility under the next period's optimal portfolio weights. This integral can be very complex depending on the utility function and posterior predictive distribution. However, the mean-variance optimization method allows the investor to bypass the full integration and use only the posterior predictive moments to calculate the portfolio weights. This method reduces the estimation risk accounted for in the model,[9] but it is still a robust method of analysis. The expected return and volatility are the most important aspects of a portfolio and they are fully accounted for in the general Bayesian mean-variance optimization model. In my analysis I use the mean-variance optimization method without any investment constraints.[10]

3.2   Markowitz

Markowitz (1952) specifies a mean-variance utility function with respect to the portfolio asset weight vector, w. The investor's goal is to maximize the expected return while minimizing the volatility and he does so by maximizing the utility function

---

[9]Point estimates, not full distributions, are used in the final step of mean-variance optimization.
[10]Short selling is allowed.

$$U(w) = E[R_{T+1}] - \frac{\gamma}{2}Var[R_{T+1}] = w'\mu - \frac{\gamma}{2}w'\Sigma w, \tag{4}$$

where R_{T+1} is the next period's return, γ is the investor's risk aversion coefficient, $\mu = \mu_h$, and $\Sigma = \Sigma_h$.[11] (4) is referred to as a two moment utility function as it incorporates the predictive distribution's first two moments, the mean and variance. The first order condition of (4) with respect to w solves to

$$w = \frac{1}{\gamma}\Sigma^{-1}\mu, \tag{5}$$

which is used to solve for the optimal portfolio weights given the historical data.

3.3 Black-Litterman

The first step of the BL model is in calculating the expected market equilibrium returns, μ_e, from (5). The historical covariance matrix, Σ_h, and the market equilibrium weight vector, w, are plugged into (5) to obtain μ_e. The market equilibrium weights are simply the percentage that each country's market capitalization makes up of the total portfolio market capitalization. Algebraically, this can be presented as

$$w_i = \frac{MktCap_i}{\sum\limits_{i=1}^{n} MktCap_i}, \tag{6}$$

where w_i is the ith asset's market capitalization weight, and n is the number of assets under analysis.

In equilibrium, if it is assumed that the CAPM holds and that all investors have the same risk aversion and views on the market, the demand for any asset will be equal to the available supply. Therefore, the weight of each asset in the optimal portfolio (demand) will be equal to the equilibrium weight from (6) (supply).

Black and Litterman (1992) model the true equilibrium excess return vector, μ, as normally distributed with mean μ_e and covariance matrix $\tau\Sigma_h$. This is written as

$$\mu = \mu_e + \epsilon^e, \quad \epsilon^e \sim N(0, \tau\Sigma_h), \tag{7}$$

where τ is a scalar indicating how μ is modeled by μ_e. A small value of τ is used consistently throughout the literature, as in Lee (2000), where τ is set between 0.01 and 0.05[12]. In practice this is not a rule of the model as there are countless methods used to specify τ.

---

[11] Historical mean and covariance matrix

[12] A small value is used under the assumption that equilibrium returns are less volatile than historical returns

Satchell and Scowcroft (2000) use $\tau = 1$, and since there is no consensus value for the variable I will present the results of my models under multiple values of τ.

The investor views are modeled by

$$P\mu = \mu_v + \epsilon^v, \ \ \epsilon^v \sim N(0, \Omega), \tag{8}$$

where P is a K×N linkage matrix that specifies K views on the N assets and Ω is the covariance matrix explaining the degree of confidence that the investor has in the views. Ω is defined as a diagonal matrix since it is assumed that each view is independent.[13] Ω is one of the more difficult variables to specify in the model, but He and Litterman (1999) provide an elegant method that also helps with the specification of τ. Each diagonal element of Ω can be thought of as the variance of a view, which can be calculated as $P_i \Sigma_h P_i'$, where P_i is an individual row (view) from P (He and Litterman, 1999). By using Σ_h to model Ω, He and Litterman (1999) assume that the variance of each view is proportional to the variance of the historical asset returns.

He and Litterman (1999) calibrate the confidence of each view by shrinking the diagonal error terms by τ. This makes the value of τ irrelevant in calculating the expected return vector specified in (9) since τ is now simplified out of the expected return result.[14] Therefore, τ now acts a tuning parameter for the investor's confidence in the views. When τ is increased, so too are the diagonal error terms of Ω, meaning the investor is less confident in the views.

There are other useful methods used to specify Ω that do not account for τ. One of the most intuitive is presented by Idzorek (2005), in which the investor specifies a confidence interval for each individual view. For example, an investor might assume that the S&P 500 will return 4% more than the Nikkei in the next period, with 95% confidence that the difference in return will be between 2% and 6%. Assuming that the view is normally distributed, the only parameter not specified in the confidence interval is the variance, which can be easily solved for. The equation for a confidence interval is (X1,X2) = X $\pm\sigma$ Z*, where (X1,X2) is the specified range of confidence (2%, 6% in this example), X is the central estimate (4% in this example), and Z* is the z-score for the associated confidence level (95% in this example, with an associated Z* = 1.96), Since σ is the only unknown it can be solved for using simple algebra. This value of σ is squared to calculate the variance, which is the value input in Ω. The investor can also directly specify a variance for each view, but using confidence intervals is a more intuitive way of applying this approach.

The method presented by He and Litterman (1999) will be used throughout this paper to specify Ω. This provides a consistent specification method that can be employed across all rolling window iterations.

(7) and (8) are be combined through Bayesian updating methods, giving the BL mean and variance,[15]

---

[13]This mean all view covariance elements, the non-diagonal elements of Ω, are set to 0.
[14]See Appendix 1 for derivation.
[15]See Appendix 1 for derivation.

$$\mu_{BL} = [(\tau\Sigma_h)^{-1} + P'\Omega^{-1}P]^{-1}[(\tau\Sigma_h)^{-1}\mu_e + P'\Omega^{-1}\mu_v] \tag{9}$$

$$\Sigma_{BL} = \Sigma_h + [(\tau\Sigma_h)^{-1} + P'\Omega^{-1}P]^{-1} . \tag{10}$$

It is assumed that both the market equilibrium and the investor's views follow a multivariate normal distribution, so the posterior is also multivariate normal due to conjugacy, with (9) and (10) as the primary moments. To calculate the optimal portfolio weights, (9) and (10) are plugged into (5).

The BL posterior covariance matrix is simply $[(\tau\Sigma_h)^{-1} + P'\Omega^{-1}P]^{-1}$. The extra addition of Σ_h occurs because the investor must account for the added uncertainty of making a future prediction through the posterior predictive distribution. Empirically, this uncertainty is represented through the extra addition of Σ_h in specifying the next period covariance matrix. For a derivation of both the posterior and posterior predictive distributions, see Appendix 1.

3.4 Zhou

μ_{BL} and Σ_{BL} act as the prior estimates for the Bayesian extension engineered by Zhou (2009). The normal likelihood describing the data is defined through the one constant data generating function,

$$R_T = \mu_h + \epsilon^h, \ \ \epsilon^h \sim N(0, \Sigma_h), \tag{11}$$

where R_T is the current period's return.

The posterior predictive mean and covariance matrix in the Zhou model are defined under Bayesian updating methods as,[16]

$$\mu_z = [\Delta^{-1} + (\Sigma_h/S)^{-1}]^{-1}[\Delta^{-1}\mu_{BL} + (\Sigma_h/S)^{-1}\mu_h] \tag{12}$$

$$\Sigma_z = \Sigma_h + [(\Delta^{-1} + (\Sigma_h/S)^{-1}]^{-1}, \tag{13}$$

where, $\Delta = [(\tau\Sigma_h)^{-1} + P'\Omega^{-1}P]^{-1}$ is the posterior BL covariance matrix and S is the sample size of the data, the weight prescribed to the sample data. The larger the sample size chosen, the larger the weight the data has in the results.

It is known that both the prior and likelihood follow a multivariate normal distribution, so, due to the conjugacy of the distributions, the same is true of the posterior. μ_z is essentially

---

[16]The same assumption that returns are i.i.d. is made by Zhou (2009) as by Black and Litterman (1992) and Markowitz (1952).

a weighted average of μ_{BL} and μ_h dependent on the investor's confidence in the data. As S increases, so does the weight of μ_h in μ_z. In the limit, if S $= \infty$ then the portfolio weights are identical to the mean-variance weights. If S $= 0$ then the portfolio weights are identical to the BL weights.

Analogous to the BL model, the posterior estimate of Σ in Zhou (2009) is $[(\Delta^{-1}+(\Sigma_h/S)^{-1}]^{-1}$. The addition of Σ_h to the posterior in calculating Σ_z is necessary to account for the added uncertainty of the posterior predictive distribution. The same derivation holds here as in Black and Litterman (1992) and can be referenced in Appendix 1.

3.5 Extensions

In this subsection I will first explain how the CAPM covariance matrix is specified before presenting the inverse-Wishart and normal-inverse-Wishart models. There are four extended inverse-Wishart models in the analysis to account for the two options of inverse-Wishart placement and equilibrium matrix specification, and one normal-inverse-Wishart prior on the best performing model from the other four extensions.

3.5.1 CAPM Matrix Specification

The CAPM covariance matrix is investigated as an input for the equilibrium covariance matrix.[17] Estimates of Σ that are independent of the historical data are particularly useful when N is large, since the historical covariance matrix does a poor job estimating Σ in high dimensional settings.

To calculate Σ_{CAPM}, the CAPM regression must first be defined,

$$R_{it} = \beta_i R_{mt} + \epsilon_{it}, \tag{14}$$

where R_{it} is asset i's excess return at period t, R_{mt} is the market return at period t, β_i is the relative riskiness of the asset compared to the market, and ϵ_{it} is the error term of the regression. $|\beta_i| > 1$ means the asset is more volatile than the market, and $|\beta_i| < 1$ means the asset is less volatile than the market. This is intuitive because for a 1% change in the market return, if the individual asset has an associated change of more than 1%, then it is clearly more volatile than the market. The market portfolio in this analysis is the MSCI World Price Index, collected from Global Financial Data. The MSCI World Price Index was chosen because, given the international index dataset, the only fitting market portfolio is a world index.

The variance of each asset's expected CAPM return is calculated by

$$V[R_{it}] = \beta_i^2 V[R_{mt}] + V[\epsilon_{it}], \tag{15}$$

---

[17]The estimate is still scaled by τ.

where V[.] denotes the variance of the given variable.

The n x n CAPM covariance matrix is defined below,

$$\Sigma_{CAPM} = \begin{pmatrix} \beta_1^2\sigma_m^2 + \sigma_1^2 & \beta_1\beta_2\sigma_m^2 & \cdots & \beta_1\beta_n\sigma_m^2 \\ \beta_2\beta_1\sigma_m^2 & \beta_2^2\sigma_m^2 + \sigma_2^2 & \cdots & \beta_2\beta_n\sigma_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_n\beta_1\sigma_m^2 & \beta_n\beta_2\sigma_m^2 & \cdots & \beta_n^2\sigma_m^2 + \sigma_n^2 \end{pmatrix},$$

where β_i is the β of asset i, $\sigma_m^2 = V[R_{mt}]$ is the variance of the market portfolio, and $\sigma_i^2 = V[\epsilon_{it}]$ is the variance of the error term for asset i, where $i \in \{1, 2, ..., n\}$.[18] σ_i^2 is only included in the diagonal elements of the matrix because it is assumed that the error terms of different assets are independent.

3.5.2 Inverse-Wishart Extension

The model by Zhou (2009) uses Σ_h in the prior generating stage, and then updates the prior estimate using Σ_h as the likelihood covariance estimate. Under fully Bayesian methods historical data is not incorporated outside of the likelihood function.

In my analysis an inverse-Wishart prior is put on Σ to account for the uncertainty of estimating Σ with Σ_h in both Bayesian updates. The Zhou model has two Bayesian updating stages, so the inverse-Wishart prior is put on both priors in alternating models. In one analysis the prior is put on the equilibrium estimate and in the other it is put on the BL estimate. The Bayesian updating stage that does not incorporate the prior is left untouched.

The inverse-Wishart prior I employ changes only the specification of Σ, not μ[19], and is specified by $\Sigma \sim \mathcal{IW}(\Psi^{-1}, v_0)$ where Ψ is the prior mean of the covariance matrix, and v_0 is the degrees of freedom of the distribution. The larger the degrees of freedom, the more confidence the investor has in Ψ as an estimate of Σ. The prior is then updated by the likelihood function. When the inverse-Wishart prior is on the equilibrium estimate, the likelihood is defined by the investor's views, whereas when the prior is on the BL estimate, the likelihood is defined by the historical data. The weight of the likelihood function is determined by S, the specified sample size of the data used in the likelihood function.[20]

When the prior is on the equilibrium estimate, v_0 is the number of assumed observations used in the equilibrium specification and S (from now on referred to as SS when the prior is on the equilibrium) is the number of assumed observations used in the view specification. When the prior is on the BL estimate, v_0 is the number of observations used to calculate Σ_{BL}, while S is the number of observations used to calculate Σ_h. The values of v_0, S and SS are determined by the investor. These parameters can be thought of as confidence parameters, where a larger value specifies more confidence in the given estimate.

---

[18] n is equal to the number of assets in the analysis, 7 in this case.

[19] Though the modeling of μ is unchanged, S is still used in (16) so the changing of confidence parameters related to the inverse-Wishart prior affects μ_{ext}.

[20] See Appendix 2 for derivation of inverse-Wishart extension.

The posterior mean of the inverse-Wishart distribution[21] is used as the posterior covariance matrix in these extended models. The posterior predictive distribution is derived in the same manner as in the baseline models; the posterior matrix under the second Bayesian updating stage is added to Σ_h. When the inverse-Wishart prior is used on the equilibrium estimate, the predictive update is not immediately necessary because only the posterior covariance is used within the second Bayesian update. Therefore, only after Σ_{BL} (which has the inverse-Wishart prior incorporated) is updated by the data in the second Bayesian update is Σ_h added to the posterior covariance matrix to calculate the posterior predictive estimate.

When the prior is used on Σ_{BL}, however, Σ_h is added to the posterior inverse-Wishart mean since the inverse-Wishart prior is used within the second Bayesian updating stage.

Algebraically, this can be shown as,

$$\mu_{ext} = [\Delta^{-1} + (\Sigma_h/S)^{-1}]^{-1}[\Delta^{-1}\mu_{BL} + (\Sigma_h/S)^{-1}\mu_h] \tag{16}$$

$$\Sigma_{ext} = \Sigma_h + E[\Sigma \,|\, \mu, y_1, ..., y_n] \,, \tag{17}$$

where $E[\Sigma \,|\, \mu, y_1, ..., y_n]$ is the posterior expectation of Σ under the inverse-Wishart prior.

It must be noted that (16) and (17) are not the true Bayesian posterior predictive moments, but are simply an empirical estimate. However, these extensions are still useful as they help account for uncertainty in estimating Σ and determine which modeling procedure will perform best under the full normal-inverse-Wishart prior.

3.5.3   Normal-Inverse-Wishart Extension

In this fully Bayesian extension, a normal-inverse-Wishart prior is imposed on both BL prior estimates, μ_{BL} and Σ_{BL}, which are derived through the use of Σ_h in the equilibrium model. This estimation strategy performs best under the inverse-Wishart prior on Σ, so it is further tested under the full prior. This model is the most statistically robust of the models presented in the paper.

The normal-inverse-Wishart prior is defined algebraically below,[22]

$$\Sigma \sim \mathcal{IW}(\Psi^{-1}, v_0) \tag{18}$$

$$\mu|\Sigma \sim \mathcal{N}(\mu_0, \Sigma/k_0) \tag{19}$$

$$p(\mu, \Sigma) \stackrel{\text{def}}{=} \mathcal{NIW}(\mu_0, k_0, \Psi, v_0) \,, \tag{20}$$

---

[21]Defined in Appendix 2

[22]$\mathcal{IW}$ refers to the inverse-Wishart distribution, and $\mathcal{NIW}$ refers to the normal-inverse-Wishart distribution.

where Ψ represents the prior estimate of Σ, v_0 represents the prior degrees of freedom, or the number of estimates on which Ψ is based, μ_0 represents the prior estimate of μ, and k_0 represents the number of prior observations on which μ_0 is based. In my analysis $\Psi = \Sigma_{BL}$ and $\mu_0 = \mu_{BL}$. The values of k_0 and v_0 are determined by the investor depending on their confidence in the BL prior estimates. It may make sense to let $k_0 = v_0$ since both the mean and covariance estimates are derived from the same models. However, if the investor has more confidence in the prior estimates of μ_{BL} or Σ_{BL}, then the confidence parameters should reflect those views. In fact, it turns out that if $k_0 = v_0$ then the results of the NIW model are poorly specified.

The likelihood function is normal and defined by the sample moments of the data,

$$L(\mu, \Sigma; \Phi) \stackrel{\text{def}}{=} \mathcal{N}(\mu_h, \Sigma_h) , \tag{21}$$

where Φ represents the data collected.

The posterior distribution is calculated through a Bayesian update where the result is

$$P(\mu, \Sigma \,|\, \mu_0, k_0, \Psi, v_0, \mu_h, \Sigma_h) \;=\; \mathcal{NIW}(\mu_n, k_n, \Psi_n, v_n) . \tag{22}$$

The values of the updated parameters, μ_n, k_n, Ψ_n, and v_n, are defined in Appendix 2.

The posterior predictive distribution is calculated through a final Bayesian update where the result is

$$P(r_{T+1}|\Phi) \;=\; t_{v_n-n+1}(\mu_n, \frac{\Psi_n \,(k_n+1)}{k_n \,(v_n-n+1)}) . \tag{23}$$

The posterior predictive distribution is therefore a multivariate student t-distribution with $(v_n - n + 1)$ degrees of freedom and primary moments described in Appendix 2.

4   Data

4.1   Data Source and Description

Monthly stock prices from 1970-2013 for the indices on Australia, Canada, France, Germany, Japan, the U.K. and the U.S. were obtained from Global Financial Data, and were used to calculate the monthly percent return for each index. Respectively, the indices used are the Australia ASX Accumulation Index-All Ordinaries, the Canada S&P/TSX-300 Total Return Index, the France CAC All-Tradable Total Return Index, the Germany CDAX total Return

Index, the Japan Nikko Securities Total Return Index, the U.K. FTSE All-Share Return Index and the S&P 500 Total Return Index. The analysis is based on excess returns so a risk-free rate is also needed. The 3-month U.S. Treasury Bill return is used as the risk-free rate in my analysis.

There are 528 monthly returns in the dataset. While this is not an extremely large sample, it is not prudent to extend the dataset further in the past because including data that is too old will only weaken the analysis. As time goes on trends in the economy change, meaning very old data is not as useful in explaining today's global investment environment.

Data must also be incorporated to describe the market equilibrium, which is determined by the indices' relative market capitalizations. This data was also collected from Global Financial Data and defines the market capitalizations of the entire stock markets in each country from January, 1980 to December, 2013. For a few of the country indices (for the first few years), only yearly data was available so the yearly values were appended to the missing months of each year. Though this may not be completely accurate, total stock market capitalization is not a particularly volatile statistic so it is very unlikely this will significantly affect the results. The market capitalization data also does not describe the specific total return stock indices,[23] but it is still a valid description of the dollar amount of assets in the chosen indices, since the indices are formed to represent the stock market. As done with the index stock prices, all currency values are converted to USD.[24]

4.2   Descriptive Statistics

Table 1 presents descriptive statistics for the seven country indices. The mean annualized monthly excess returns are all close to seven percent and the standard deviations are all close to 20 percent. The volatility for the U.S. is much smaller than for the other countries. Safer investments generally have less volatility in returns, and the S&P 500 is probably the safest of the indices in question. All countries exhibit relatively low skewness, and most countries have a kurtosis that is not much larger than the normal distributions kurtosis of 3.[25] The U.K. deviates the most from the normality assumption given it has the largest absolute value of skewness and a kurtosis that is almost two times as large as the next largest kurtosis. These values are not particularly concerning, however, because the dataset is large and the return distribution does not drastically differ from a normal distribution. The U.K. has a particularly large kurtosis which is less problematic than a large skewness. The skewness is greatly influenced by one particularly large return that occurred in January of 1975 when the U.K. was recovering from a recession. During the recession the U.K.'s GDP decreased by almost 4% and inflation reached as high as 20%(Zarnowitz and Moore, 1977). Inflation was still rampant when the recession ended in January, 1975, which creates a perfect storm for such a high monthly return. Even though the data is total return adjusted, and therefore

---

[23]This data was unavailable on Global Financial Data.

[24]All conversions between currencies were handled automatically through Global Financial Data.

[25]Skewness obviously measures the skewness of the distribution is in a particular direction, where a true normal distribution has a skewness of 0. Kurtosis measures the peakedness of the distribution where a kurtosis >3 means that the distribution is more peaked than a normal distribution.

inflation adjusted, it is difficult to account for such an acute spike in inflation when making such adjustments. The combination of these factors likely leads to the extremely high return value in January 1975, which in turn leads to the high skewness for the U.K. Though the observation is an outlier, it seems to have occurred under legitimate market circumstances and so it is included in the analysis.

Table 1: Analysis of Country Index Returns

Country	Mean (%)	St. Dev. (%)	Skewness (%)	Kurtosis
Australia	7.86	23.68	-0.84	7.54
Canada	5.91	19.44	-0.62	5.57
France	7.43	22.83	-0.18	4.39
Germany	6.83	21.30	-0.37	4.46
Japan	6.11	21.04	0.23	3.79
UK	7.97	22.38	0.98	13.60
US	6.15	15.49	-0.45	4.78

5 Model Implementation

5.1 Rolling Window

A predictive model is best tested under repeated conditions where a subset of the data is used as in-sample data to predict the out-of-sample optimal portfolio weights. For each iteration, only the in-sample data is used to calculate the next period's weights. This simulates how a model would be implemented in a real investment setting since there is obviously no data incorporated in the model for the future prediction period. If the true returns of the predicted periods were included in the analysis, the predictive power of the model would be artificially increased.

In my analysis a 10-year rolling window is used as the in-sample data to predict the following month's out-of-sample optimal portfolio weights. The first set of in-sample data is the first ten years of the data set, January, 1970 - December, 1980, and is used to predict the optimal asset weights for the following month, January 1981.[26] The window then slides over one month and February, 1970 - January, 1981 is used to predict the optimal asset allocations for February, 1981. The dataset extends through 2013, giving 528 individual returns, and given the 10-year rolling window, 408 iterations of the model will be run.

The number of observations used in the in-sample dataset must be considerably larger than the number of parameters estimated by the data. In this analysis, there are 56 different parameters that must be estimated (49 in the 7x7 covariance matrix, and 7 historical mean returns). This ratio of data observations to estimated parameters is sufficient as there are

---

[26]The models in this paper are all one period models, so only the following month's optimal weights are calculated.

many more observations than estimates. If a larger window were chosen there would be fewer calculated iterations of the models, in turn making the model testing procedure less robust. A 10-year rolling window is ideal to maximize iterations while also accounting for enough observations to estimate the parameters.

Another commonly used window is an expanding window, which starts at the beginning of the dataset and expands to include each successive month. One drawback of this method is that for each iteration, the early data points become increasingly far from the period they are predicting. In the final iteration of a model tested under an expanding window (under this dataset), data from 1970 would be used to estimate the optimal weights for December, 2013. This is not ideal because information on the market in 1970 is likely not useful in calculating portfolio weights for a period more than 40 years in the future. The expanding window also decreases the relative weight of each observation as the window expands, because as the number of observations in the in-sample data increases, each individual observation affects the results less and less. There is also less independence across iterations under an expanding window, since each new iteration contains the same in-sample data of the previous iteration, plus one more observation.

Under the rolling window it is quite simple to assess model performance since there is data on each realized return. For each iteration, the realized return of portfolio is calculated by multiplying each individual index's calculated portfolio weight by its corresponding realized return. There are no investment constraints in the model so the amount invested in, or borrowed from, the risk-free rate must also be accounted for. The difference between the sum of all the calculated portfolio weights and 1 is the amount invested or borrowed from the risk-free rate. For example, if all the weights add up to 1.5, this means the risk-free rate was sold short at a weight of .5. Therefore, to calculate the total realized portfolio return the implied risk-free allocation must be multiplied by the corresponding monthly risk-free rate and added to the total return of the assets in the portfolio.

5.2   Momentum-Based Views

In order to use the iterative rolling window an updating view function must be specified to create views that can be imposed on each iteration. The momentum strategy employed in this analysis is based on methodology proposed by Fabozzi et al. (2006), who employs a cross-sectional ranking momentum strategy. The previous 9-month return is calculated for each asset, and assets are ranked by their return. Positive weights are imposed on the top half of the ranked assets and negative weights are imposed on the bottom half. Then, in a one-line relative view vector, all of the indices are weighted by both their volatility and a specified scaling factor that is used to set a specific volatility for the view-specific portfolio.

This method is somewhat limiting because it is quite possible that more than half of the stocks could have positive or negative momentum at a given time period. Therefore, in my analysis, positive weightings are imposed on the assets with positive 9-month returns, and negative weightings are imposed on the assets with negative 9-month returns. The relative

weights are determined by the market capitalization weighting method presented by Idzorek (2005), which is similar to the method used to calculate the equilibrium returns. The positive return assets are weighted by the ratio of their individual market capitalizations to the sum of the positive return assets' market capitalizations, and the same goes for the negative return assets. This puts more weight on large indices, which is intuitive because there is likely more potential for realized returns.

To calculate the expected return of the view, a positive, market capitalization weighted mean is calculated for the positive return assets and a negative market capitalization weighted mean is calculated for the negative return assets. The difference between these values is therefore the estimated amount that the positive return capitalization weighted portfolio is expected to return over the negative market capitalization weighted portfolio.

The Omega entry is calculated using the method specified by He and Litterman (1999).

6 Results

6.1 Baseline Models

The results of the three baseline models are presented below in Table 2. The results are dependent on the input parameters γ, τ, and S (the sample size of the data specified in the historical updating stage). The parameters are set as $\gamma = 2.5$, $\tau = .025$, S = 60 for the results in Table 2.[27] A sensitivity analysis of the baseline models can be found in Appendix 3. The sensitivity analysis is quite important to interpreting the results because by varying the parameter inputs it is possible to see which estimates of μ and Σ are most important to the empirical success of the models.

It also must be considered that the results are heavily dependent on both the dataset and the view specifying function, two aspects of the model that are not necessarily generalizable to any investor. Further empirical analysis of the models is therefore necessary to determine which is best under the varying conditions of the current investment market and under different investor views.

Table 2: Summary Statistics for Baseline Models

Model	Mean Return (%)	Volatility (%)	Skewness (%)	Kurtosis
Markowitz	2.82	39.94	-0.39	4.46
BL	7.24	15.02	0.51	4.19
Zhou	4.62	27.77	-0.47	4.70

Note: The values in this table are specified under $\gamma = 2.5$, $\tau = .025$ and S = 60

---

[27]The Sample Size specification of 60 is based off the idea that the investor is 50% confident in the data, since the true sample size of each iterative result is 120. This tuning parameter is the least specified in the literature, and therefore most dependent on the investor's discretion.

The Markowitz model performs the worst of the baseline models, both in terms of the mean return and volatility. The results imply that given the dataset, μ_h and Σ_h do not do a great job on their own as data inputs in the mean-variance portfolio optimization problem. This is consistent with the original hypothesis that further data inputs are necessary in conjunction with a more robust modeling procedure to improve the overall model.

Comparisons between the Markowitz model and the others are difficult to make outside of the overall conclusion that the BL and Zhou models outperform the Markowitz model, both in the mean return and volatility. The BL model uses completely different data inputs, and though Zhou model uses historical data as an input like Markowitz, the BL prior estimates make it difficult to directly explain why the Zhou model improves upon the Markowitz model. The main conclusion is that using only historical data in the mean-variance analysis is not optimal, especially when more robust data inputs and modeling procedures are available.

Comparing the BL and Zhou models is much easier since the only difference between the two models lies is in the use of historical data. The BL model outperforms the Zhou model in the mean return and volatility, meaning that in this analysis the incorporation of historical data is not optimal. However, this does not render the Zhou model useless since repeated empirical analysis is necessary to determine the actual effects of the historical data. Zhou (2009) only calculates one iteration of the model as a brief example, so there is currently no sufficient literature on whether the historical data is truly an optimal addition. A robust model testing procedure could be employed by running a rolling-window model testing procedure on many datasets, and then running t-tests on the set of mean returns and volatilities specified under each dataset to determine if one model consistently outperforms the other.

A more in depth analysis of the Markowitz, BL, and Zhou models is possible by examining how the varying of tuning parameters affects the results.[28]

The Markowitz model performs increasing well under larger values of γ. The model often specifies particularly risky positions, so it is intuitive that increasing the risk-aversion of the investor will lower the volatility. However, it is surprising that a larger γ also increases the mean return, since lower volatility is often associated with a lower mean return.

The BL model is largely resistant to changes in γ. The sensitivity table shows identical results for each value of γ,[29] which is simply a result of the model set-up. γ is used in both the market equilibrium calculation in the prior generating stage, as well as in the calculation of the final weights in the mean-variance optimization. This essentially wipes out the effect γ because the market equilibrium values are determined by γ, and the mean-variance optimization, which also uses γ, is calculated mostly through the equilibrium values.[30]

In the BL model τ specifies the investor's confidence in the views, where a larger value is associated with less confidence. Increasing τ improves the model with respect to volatility, but not the mean return. When τ is increased from .01 to 1, the mean return is only slightly

---

[28]Sensitivity tables are presented in Appendix 3.

[29]The results do change very slightly in smaller decimal places.

[30]The views seem to be playing a minimal role with respect to a changing γ.

diminished while the volatility decreases by almost 50%. This is a great trade-off for almost any investor. These results imply that while the momentum strategy does work, it does not have enough predictive power to be used with complete confidence. This finding is in line with the literature on momentum, the momentum strategy is useful because momentum is a significant phenomenon in the market, but it is not robust enough to merit extreme confidence.

The results of the Zhou model are greatly affected by all three of the tuning parameters. Increasing values of γ are associated with an increasing mean return and a decreasing volatility. γ significantly affects the results of the Zhou model, unlike the BL model, because the parameter no longer has a similar affect across multiple stages of the model. γ is used in calculating the market equilibrium returns, μ_e, in the first stage, but is not used again in the model until after the historical data is incorporated. At this point the estimates are considerably changed from the equilibrium values so γ is not identically accounted for when used again in the mean-variance optimization. Larger values of γ improve the model both in the mean return and volatility which is consistent with the results of the Markowitz model. This consistency must occur because γ has the same affect across the use of the same historical data.

In the Zhou model τ is still used as the parameter that determines the relative confidence in the investor's views, as compared to the equilibrium estimate, but its effect on the results is different due to the incorporation of historical data. When τ is increased, the model inherently puts more weight in both the equilibrium *and* the data. It has already been shown that the incorporation of historical data hurts the results in this empirical analysis, so by increasing the relative weight of the data by increasing τ it follows that the results are hurt both in the mean return and volatility.

Increasing values of S are associated with a decreasing mean return and an increasing volatility, which is expected in this empirical analysis because using the data does not improve the portfolio. This is now a direct effect since S specifically determines the weight of the data.

6.2 Extended Models

The results of the five proposed extended models are presented below in Table 3. There are four inverse-Wishart extensions that differ in their use of Σ_{CAPM} or Σ_h in the equilibrium stage and in the location of the inverse-Wishart prior. The results are calculated using parameter inputs of $\gamma = 2.5$, $\tau = .025$ and S = 60. For the models with the inverse-Wishart prior on the equilibrium estimate, the prior degrees of freedom (v_0) is equal to N+2[31] and the posterior sample size on the views (SS) is equal to 1 (the investor's views have a weight of one observed data point). v_0 is also equal to N+2 for the models with the inverse-Wishart prior on the BL covariance estimate. It must be noted that when Σ_{CAPM} is used as the

---

[31]N+2 is the least informative possibility, any smaller value will cause the prior matrix to be non-semi definite and the analysis will not work.

equilibrium estimate, it is also used throughout the entire BL prior generating stage which means Ω is derived from Σ_{CAPM}.

In the fifth extended model that incorporates a full normal-inverse-Wishart prior on the BL prior estimates (μ_{BL}, Σ_{BL}), derived from Σ_h, it is prudent for the investor to specify more confidence in the BL estimates (without confident views) than the data. This parameterization ensures that the posterior predictive covariance matrix is not overspecified. The results for the NIW model are therefore presented under $v_0 = k_0 = 120$ and S = 15 and $\tau = 1$.

The extended models are referred to throughout the section as follows:

1. **Equil-Historical**: Inverse-Wishart prior on the equilibrium estimate $\tau \Sigma_h$

2. **Equil-CAPM**: Inverse-Wishart prior on the equilibrium estimate $\tau \Sigma_{CAPM}$

3. **BL-Historical**: Inverse-Wishart prior on Σ_{BL}, derived from Σ_h

4. **BL-CAPM**: Inverse-Wishart prior on Σ_{BL}, derived from Σ_{CAPM}

5. **NIW**: Full normal-inverse-Wishart prior on μ_{BL} and Σ_{BL}, derived from Σ_h

Table 3: Summary Statistics for Baseline Models

Model	Mean Return (%)	Volatility (%)	Skewness (%)	Kurtosis
Equil-Historical	7.53	28.29	-0.22	4.57
Equil-CAPM	6.51	40.93	-0.22	4.19
BL-Historical	4.75	9.52	-0.47	4.67
BL-CAPM	36.52	79.47	1.03	11.99
NIW	6.70	15.83	-0.57	0.12

Note: For the first for models, the values in this table are specified under $\gamma = 2.5$, $\tau = .025$, $v_0 = N+2$, SS = 1 and S = 60. For the final model, $v_0 = k_0 = 120$, S = 15 and $\tau = 1$

It can be seen in Table 3 that of the four inverse-Wishart prior models, the BL-Historical model performs the worst in the mean return but the best in volatility. The other models all have extremely high volatilities, even higher than the original Zhou model, so in spite of the mean return being higher there is not much of an overall improvement. The BL-Historical model performs quite well, however, and even beats the BL model in volatility and Zhou model in both the mean return and volatility. The goal of the inverse-Wishart prior is to reduce volatility by accounting for the uncertainty in the estimation of Σ, and in this empirical analysis it appears to do so.

I am most confident in the results of the NIW model due to the robust Bayesian methods used in the model. When confidence parameters are specified well, the NIW model improves upon the Zhou model both in the mean return and volatility.

The results of each extended model are further explained in the subsections below. As in the baseline analysis, the results are heavily dependent on the values of the parameter inputs. Sensitivity tables displaying the parameters' effects on the results can be found in Appendix 4.

6.2.1 Equil-Historical

The confidence parameters in the Equil-Historical model can be difficult to specify since v_0 and SS are both related to τ. v_0 directly specifies confidence in the prior (equilibrium), SS directly specifies confidence in the likelihood (views), and τ determines the relative confidence in the views and equilibrium. Increasing values of τ are almost always associated with a decreasing mean return and an increasing volatility.[32] When SS is large and v_0 is small, however, increasing τ has an opposite effect because it corrects the over-weighting of the views in the SS and v_0 specification. In general, the results are less influenced by τ in this model because v_0 and SS account for similar specifications.

The model performs best under low prior degrees of freedom (v_0), high likelihood sample sizes (SS) and a low data sample size (S). This finding is consistent with the results thus far because it was found in the Zhou results that more weight in the data hurts the portfolio. The specification of v_0 and SS is not entirely consistent, however, because in the primary results the Zhou model does better when less weight is put in the views, which is the opposite of what is done with a high SS and low v_0. The specification is somewhat different here, however, since the posterior covariance matrix in the first Bayesian updating stage is now a weighted average of the prior estimate ($\tau\Sigma_h$) and the likelihood estimate (Ω). The results therefore imply that the model does better when the diagonal Ω carries more weight. This means Σ_{ext} is over-specified when too much weight is put in the prior $\tau\Sigma_h$ matrix. By giving more weight to Ω in the prior, it allows the covariance elements of the estimated Σ_{ext} to be specified mostly by Σ_h in the second Bayesian updating stage.

The Equil-Historical model performs best under low values of v_0, high values of SS and low values of S, holding the other parameters constant. This allows the data inputs to be combined such that the views are influential in specifying returns, and in not over-specifying the covariance elements of Σ. This model is ideal for an investor that is willing to take risks, since both the mean return and volatility are large.

6.2.2 Equil-CAPM

Σ_{CAPM} appears to do a poor job in estimating Σ as the Equil-CAPM model performs considerably worse than the Equil-Historical model, for all values of v_0 and SS.

The same difficulties in parameter specification arise in this model, as in the Equil-Historical model. However, τ is much less influential in this model due to use of Σ_{CAPM} as the equilibrium covariance. Under the BL prior, Σ_{CAPM} is used to calculate both μ_e and Σ_{BL}. The results imply that Σ_{CAPM} does a poor job in both functions. μ_e is very important to all the models since it serves as the primary estimates of μ. In faultily estimating μ_e the model is essentially tainted beyond retrieve. Therefore, there is no value of τ that can account for the non-optimal use of Σ_{CAPM} in the prior stage.

Increasing values of v_0 are associated with a decreasing mean return and volatility. In this

---

[32]This is consistent with the results of the standard Zhou model.

model the (diagonal) elements of Ω are specified as proportional to the diagonal elements of Σ_{CAPM}, which is now known to be a faulty estimator. It is therefore better to put confidence in the minimally informative and uniquely specified Σ_{CAPM}, rather than the much less informative Ω.

Increasing values of SS are associated with a drastically increasing mean return and volatility. By putting confidence in the views the investor is benefiting in returns by using the momentum strategy over the faultily specified equilibrium values, but is hurt in volatility because the momentum strategy in itself is more volatile, and Ω is poorly modeled through the CAPM matrix.

Increasing values of S are associated with a decreasing mean return and volatility. The historical data does a poor job estimating μ, but in this case putting more weight in the data helps to decrease the volatility. The historical matrix does relatively well modeling Σ when the number of assets under analysis is low, and it does even better in this model since the other estimate of Σ in this model is Σ_{CAPM}. It is therefore beneficial in volatility to put more weight in Σ_h rather than Σ_{CAPM} through a larger S.

The Equil-CAPM model performs best under moderate values of v_0, SS and S, holding the other parameters constant. This parameterization allows the data inputs to be combined without any undue weight on any specific one. This model should not be used in a similar empirical setting as it is consistently outperformed by Equil-historical. However, in a setting where many more assets are under consideration, Σ_{CAPM} (or a different non-historical data generated covariance matrix) will have more predictive power making the Equil-CAPM model applicable.

6.2.3   BL-Historical

The confidence parameters in the BL-Historical model are more intuitive than those in the equilibrium models since there is no longer an overlap in specification between τ, v_0 and SS.[33] The inverse-Wishart prior is no longer used in the first updating stage, meaning τ is now the only parameter that defines the relative confidence in the equilibrium and views. v_0 now specifies the confidence in the BL estimates (μ_{BL}, Σ_{BL}), while S specifies the confidence in the historical estimates (μ_h, Σ_h).

τ has an identical effect on the results of this model as it does on the results of the baseline Zhou model since it is used in the exact same manner. The effect is more pronounced for larger values of v_0, since this puts more weight in the BL estimate where τ is employed.

Unlike in the equilibrium models, there is no strict rule for how v_0 and S affect the mean return. Under large values of S, increasing values of v_0 are associated with a decreasing mean return, while the opposite is true for small values of S. This implies that both the BL estimates and the historical data play useful roles, since over-weighting either estimator diminishes the mean return. The mean return is not particularly volatile with respect to the parameter inputs, while the volatility is.

---

[33]SS is no longer a confidence parameter in this model as it is only needed in the equilibrium.

The pattern of changing portfolio volatility is also more consistent with respect to S and v_0 as increasing values of the parameters are associated with an increasing volatility. In this model, larger parameter values are associated with a posterior covariance matrix with smaller elements because the mean of the posterior inverse-Wishart distribution, as defined in Appendix 2, is used as the posterior input. To calculate the posterior predictive Σ_{ext}, the posterior inverse-Wishart mean is added to Σ_h. Therefore, smaller parameter values are associated with more weight in the posterior estimate, as opposed to Σ_h, in the posterior predictive estimate. The posterior inverse-Wishart estimate uses more information (through the Bayesian updates) than Σ_h to estimate Σ, so it is fitting that the model performs better when the posterior mean is weighted more heavily in Σ_{ext}.

The BL-Historical model performs best under low values of v_0 and S, holding the other parameters constant. When parameters are correctly specified, it outperforms the other models presented in this section. It is the most applicable of the four inverse-Wishart extensions and will be further analyzed in the NIW model.

6.2.4   BL-CAPM

The confidence parameters in the BL-CAPM model are interpreted in an identical manner to the parameters in the BL-Historical model.

τ is a particularly important confidence parameter in this model since it is solely used to define the relative confidence in the equilibrium and views. Increasing values of τ are associated with a drastically decreasing mean return and volatility. Under large values of τ, the investor implies minimal confidence in the momentum strategy. Therefore, the mean return falls because the investor does not fully exploit the momentum strategy. However, the volatility also falls because the overall investment strategy is less aggressive.

Increasing values of v_0 are associated with an increasing mean return and volatility. Under large values of v_0 more weight is put in μ_{BL}, which has been shown to do a better job as an estimator than μ_h in maximizing the portfolio return. However, Σ_{CAPM} does a much worse job than Σ_h in estimating Σ, so it is fitting that the volatility increases along with the returns.

The effect of S is not as consistent; when v_0 is large compared to S, increasing values of S are associated with decreasing returns and volatility but when v_0 is small compared to S, increasing values of S are associated with increasing returns and volatility until an inflection point is reached and the returns and volatility begin to decrease. This implies that neither the BL estimates nor the historical data should have an overwhelming weight in the posterior. Even though the use of historical data does not consistently improve the portfolio, it does indeed help if the investor is too confident in the BL estimates. In this model, the BL estimates are not ideally specified under Σ_{CAPM}, so it is fitting that the data needs to play a bigger role in determining the portfolio.

The BL-CAPM model performs best under moderate values of v_0 and S, holding the other parameters constant This allows both the BL estimates and the historical data to play a role

in the portfolio optimization problem. Though this model does not perform as well as the BL-Historical model, it still may be useful when N is large, making a non-data estimate of Σ more applicable.

6.2.5 NIW

The NIW model uses the most robust Bayesian statistical methods to specify the mean-variance inputs by putting a normal prior (conditional on Σ) on μ_{BL} and an inverse-Wishart prior on Σ_{BL}. The posterior predictive t-distribution is fully calculated and the estimated moments are used in the standard mean-variance optimization method presented in (5).

There is an additional confidence parameter used in the normal-inverse-Wishart prior, k_0, which is the number of observations that the investor assumes went into the calculation of μ_{BL}. The confidence parameters of v_0 and S are defined the same as in the other extended models.

If the investor assumes $k_0 = v_0$,[34] as S increases the mean return and volatility converge to the original Markowitz solution. This is no different than an increasing S in the original Zhou model because the data is playing a larger role in the optimization model. As $k_0 = v_0$ increases, both the mean return and volatility increase significantly. This occurs because the momentum strategy is given too much weight, both in the mean and variance. The problem is known to be in the view specification because if τ is increased, the changes in the mean return and volatility under an increasing $k_0 = v_0$ are much less pronounced. The results therefore imply that by heavily weighting the BL estimates under confident views, Σ_{ext} is over-specified

If $k_0 \neq v_0$, it is up to the investor to decide whether μ_{BL} (associated with k_0) or Σ_{BL} (associated with v_0) should carry more weight. The basis of this paper is in trying to account for uncertainty in estimating Σ, so it is intuitive that $k_0 > v_0$ should be used. Through this specification, the investor is able to benefit from the momentum strategy in returns without over-specifying Σ_{ext} through multiple updates of heavily weighted covariance matrix. If τ is increased, the results still improve both in the mean return and volatility but not as extensively as under a large value of v_0.

The NIW model is able to robustly account for the uncertainty in estimating both μ and Σ through the incorporation of a normal-inverse-Wishart prior. An investor can now directly specify their confidence in μ_{BL}, Σ_{BL} and the data while also accounting for the uncertainty of modeling Σ through Σ_h in multiple Bayesian updates.

---

[34]This means the investor assumes the same number of observations went into the calculation of μ_{BL} and Σ_{BL}.

7 Conclusion

In exploring the results of my extensions to the Zhou model, it is clear that fully Bayesian mean-variance specification methods outperform loosely Bayesian methods when parameters are specified correctly. Through the four extensions under the inverse-Wishart prior, it was found that BL-Historical extension outperforms the Zhou model in volatility. With this information in hand, a full normal-inverse-Wishart prior was used on the same prior estimates to create a robust and fully Bayesian mean-variance specification model.

The BL model, which is used as a joint prior in the Zhou and extended models, allows the investor to incorporate specific views on the market. The views can be determined in a one-off nature or by a complex function specifying an investment strategy. The former would likely be employed by an amateur, independent investor while the latter by a professional or investment team. All the models presented in this paper use a market capitalization weighted momentum strategy to specify the views in each iteration.

The data updating stage of the Zhou model has similar flexibility in that the historical means, or a more complex data modeling mechanism, can be employed depending on the quantitative skills of the investor. The incorporation of a predictive model is a topic of further research that could significantly increase the profitability of the Bayesian model. However, this application would also greatly increase the complexity of the model. Asset return predictions models can also be incorporated in a much simpler manner through the use of absolute views on a specific asset.

The inverse-Wishart prior is used to model the uncertainty of predicting the next period's covariance matrix, which is not fully accounted for in the original Zhou model. This method works well empirically in the BL-Historical model. The models that use the CAPM covariance matrix may be useful under large values of N, as the historical covariance matrix does a poor job estimating the next period's covariance matrix in this setting. However, the CAPM covariance calculation is a simple method that is used as an example of other potential covariance matrix specification methods. The study of additional covariance matrix inputs serves as another topic of further research within the general model.

The normal-inverse-Wishart prior is used in the NIW model to fully account for the uncertainty of estimating the mean and covariance matrix. This model is the most statistically robust given the fully Bayesian techniques used to estimate the mean-variance inputs. It also performs well empirically as it outperforms the Zhou model in returns and volatility under correct parameterization. To determine the parameters, the investor can run an iterative model on previous investment periods to see which parameter values are associated with investment profitability. This will give the investor a sense of which model inputs are most important in estimating the next period's return and covariance matrix. This method is particularly important in determining S because the inclusion of historical data may not be optimal, as in my empirical analysis. If this is the case, the investor should set a small S or simply use the Black-Litterman model which does not incorporate historical data. Though these historical parameters will not perfectly specify the next period's parameters, they give the investor a method of determining how the future returns are predicted by the various

inputs of model.

The NIW model uses fully Bayesian methods to specify the mean and covariance matrix inputs, but the mean-variance method is still used to calculate the optimal portfolio weights instead of a fully Bayesian method. Bayesian optimization methods are still applicable to this analysis given that the posterior predictive t-distribution is fully calculated in the NIW model. In a fully Bayesian optimization model the investor would maximize investment utility with respect to the posterior predictive distribution. This application serves as a particularly important topic of further research within the realm of Bayesian portfolio optimization as there are many different investor utility functions that can be employed.

Through the use of a rolling window, the results presented in this paper give an idea of how the models perform under repeated conditions. However, each iteration of the rolling window is very similar to the previous one given that all but one data point is identical. In order to confidently determine if one model outperforms another, it is necessary to do an empirical analysis on multiple datasets.

There are countless strategies within the Bayesian mean-variance model for both input specification (in an economic sense) and input combination (in a statistical sense). The importance of input specification is exemplified by the sub-optimal performance of the CAPM covariance matrix in the equilibrium model, while the importance of input combination is seen by the optimal performance of the NIW model. Through this Bayesian mean-variance specification model, the investor has a straightforward quantitative algorithm that can help improve investment success. Investors base their decisions off how they view the assets in the market, and by using this model they can greatly improve their chance of profitability by using robust methods of prediction outside of their views.

References

Asness, C., Liew, J., and Stevens, R. (1997). Parallels between the cross-sectional predictability of stock and country returns. *Journal of Portfolio Management*, 23(3):79–87.

Berger, A., Ronen, I., and Moskowitz, T. (2009). The case for momentum investing.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analysts Journal*, 48(5):28.

Fabozzi, F., Focardi, S., and Kolm, P. (2006). Incorporating trading strategies in the black-litterman framework. *CFA Digest*, 36(4).

He, G. and Litterman, R. (1999). The intuition behind black-litterman model portfolios. *Investment Management Research, Goldman Sachs and Co.*

Idzorek, T. (2005). A step-by-step guide to the black-litterman model.

Jobson, J. D. and Korkie, B. (1981). Putting markowitz theory to work. *Journal of Portfolio Management*, 7(4):70–74.

Lee, W. (2000). Advanced theory and methodology of tactical asset allocation.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77.

Pogue, G. (1970). An extension of the markowitz portfolio selection model to include variable transactions' costs, short sales, leverage polices and taxes. *The Journal of Finance*, 25(5):1005–1027.

Satchell, S. and Scowcroft, A. (2000). A demystification of the black-litterman model: Managing quantitative and traditional portfolio construction. *Journal of Asset Management*, 1(2):138–150.

Zarnowitz, V. and Moore, G. (1977). The recession and recovery of 1973-1976. *Explorations in Economic Research*, 4(4):1–87.

Zhou, G. (2009). Beyond black-litterman: Letting the data speak. *The Journal of Portfolio Management*, 36(1):36.

Appendix 1 - Baseline Derivations

Black-Litterman Derivation

Equilibrium Prior Specification

$$p(\mu) = \mathcal{N}(\mu_e, \tau\Sigma_h)$$

$$p(\mu) = 2\pi^{\frac{m}{2}} |\tau\Sigma_h|^{-\frac{1}{2}} exp\{ -\frac{1}{2}(\mu - \mu_e)'(\tau\Sigma_h)^{-1}(\mu - \mu_e) \}$$

$$\propto exp\{ -\frac{1}{2} (\mu - \mu_e)'(\tau\Sigma_h)^{-1}(\mu - \mu_e) \}$$

$$\propto exp\{ -\frac{1}{2} [\mu'(\tau\Sigma_h)^{-1}\mu - \mu'(\tau\Sigma_h)^{-1}\mu_e - \mu_e'(\tau\Sigma_h)\mu + \mu_e'(\tau\Sigma_h)^{-1}\mu_e] \}$$

$$\propto exp\{ -\frac{1}{2} [\mu'(\tau\Sigma_h)^{-1}\mu] + \mu'(\tau\Sigma_h)^{-1}\mu_e \}$$

View Likelihood Specification

$$p(\mu_v|\mu) = \mathcal{N}(P\mu, \Omega)$$

$$p(\mu_v|\mu) = 2\pi^{\frac{m}{2}} |\Omega|^{-\frac{1}{2}} exp\{ -\frac{1}{2}(\mu_v - P\mu)'(\Omega)^{-1}(\mu_v - P\mu) \}$$

$$\propto exp\{ -\frac{1}{2}(\mu_v - P\mu)'(\Omega)^{-1}(\mu_v - P\mu) \}$$

$$\propto exp\{ -\frac{1}{2} [\mu_v'(\Omega)^{-1}\mu_v - (P\mu)'(\Omega)^{-1}\mu_v - \mu_v'(\Omega)^{-1}(P\mu) + P\mu'(\Omega)^{-1}P\mu] \}$$

$$\propto exp\{ -\frac{1}{2} [\mu'(P'\Omega^{-1}P)\mu] + \mu'(P'\Omega^{-1})\mu_v \}$$

Bayesian Updating

$$p(\mu|\mu_v, \Sigma_h) = \frac{p(\mu_v|\mu)\, p(\mu)}{\int_\Theta p(\mu_v|\mu)p(\mu)d\mu}$$

$$\propto p(\mu_v|\mu) \times p(\mu)$$

$$\propto exp\{ -\frac{1}{2} [\mu'(P'\Omega^{-1}P)\mu] + \mu'(P'\Omega^{-1})\mu_v \} \times exp\{ -\frac{1}{2} [\mu'(\tau\Sigma_h)^{-1}\mu] + \mu'(\tau\Sigma_h)^{-1}\mu_e \}$$

$$\propto exp\{ -\frac{1}{2} [\mu'(P'\Omega^{-1}P)\mu] + \mu'(P'\Omega^{-1})\mu_v - \frac{1}{2} [\mu'(\tau\Sigma_h)^{-1}\mu] + \mu'(\tau\Sigma_h)^{-1}\mu_e \}$$

$$\propto exp\{ -\frac{1}{2} \mu'[(\tau\Sigma_h)^{-1} + (P'\Omega^{-1}P)]\mu + \mu'[(\tau\Sigma_h)^{-1}\mu_e + (P'\Omega^{-1})\mu_v]\}$$

This is the kernal of a multivariate normal distribution with mean and covariance

$$\mu_{BL} = [(\tau\Sigma_h)^{-1} + P'\Omega P]^{-1} [(\tau\Sigma_h^{-1}\mu_e + P'\Omega^{-1}\mu_v]$$
$$\Sigma_{bl} = [(\tau\Sigma_h)^{-1} + P'\Omega^{-1}P]^{-1}$$

Incorporating Posterior Predictive

The Bayesian predictive density is derived from:

$$p(r_{T+1}|\mu_v, \Sigma_h) = \int_\Theta p(r_{T+1}|\mu, \Sigma) \, p(\mu|\mu_v, \Sigma_h) \, d\mu$$
$$= \int_\Theta \mathcal{N}(\mu, \Sigma) \, \mathcal{N}(\mu_{BL}, \Sigma_{bl}) d\mu,$$

where μ and Σ are the next period portfolio return and covariance matrix. The standard result of this integral is

$$p(r_{T+1}|\mu_v, \Sigma) = \mathcal{N}(\mu_{BL}, \Sigma_{BL}), \text{where} \Sigma_{BL} = \Sigma_h + \Sigma_{bl}$$

Simplification of μ_{BL} and Σ_{bl} when $\Omega = diag(P\tau\Sigma_h P')$

$$\mu_{BL} = [(\tau\Sigma_h)^{-1} + P'\Omega^{-1}P]^{-1} \, [(\tau\Sigma_h^{-1}\mu_e + P'\Omega^{-1}\mu_v]$$
$$= [(\tau\Sigma_h)^{-1} + P' \, diag(P\tau\Sigma_h P')^{-1}P]^{-1} \, [(\tau\Sigma_h^{-1}\mu_e + P' \, diag(P\tau\Sigma_h P')^{-1}\mu_v]$$
$$= \{\tau^{-1} \, [\Sigma_h^{-1} + P' \, diag(P\Sigma_h P')^{-1}P] \, \}^{-1} \, \{\tau^{-1} \, [\Sigma_h^{-1}\mu_e + P' \, diag(P\Sigma_h P')^{-1}\mu_v] \, \}$$
$$= [\tau\tau^{-1}] \, [\Sigma_h^{-1} + P' \, diag(P\Sigma_h P')^{-1}P]^{-1} \, [\Sigma_h^{-1}\mu_e + P' \, diag(P\Sigma_h P')^{-1}\mu_v]$$
$$\mu_{BL} = [\Sigma_h^{-1} + P' \, diag(P\Sigma_h P')^{-1}P]^{-1} \, [\Sigma_h^{-1}\mu_e + P' \, diag(P\Sigma_h P')^{-1}\mu_v]$$

$$\Sigma_{bl} = [(\tau\Sigma_h)^{-1} + P'\Omega^{-1}P]^{-1}$$
$$= [(\tau\Sigma_h)^{-1} + P'diag(P\tau\Sigma_h P')^{-1}P]^{-1}$$
$$= \tau[\Sigma_h^{-1} + P'diag(P\Sigma_h P')^{-1}P]^{-1}$$

Zhou Derivation

The Zhou model derivation is identical to the BL model derivation, except now the BL estimates act as the prior (rather than the equilibrium estimates), and the data acts as the likelihood (rather than the views).

$$p(\mu) = \mathcal{N}(\mu_{BL}, \Sigma_{BL})$$
$$p(\mu_h|\mu) = \mathcal{N}(\mu_h, \Sigma_h)$$
$$p(\mu|\mu_h) \propto p(\mu_h|\mu) \times p(\mu)$$
$$p(\mu_h|\mu) = \mathcal{N}(\mu_z, \Sigma_z)$$
$$\mu_z = [\Delta^{-1} + (\Sigma_h/T)^{-1}]^{-1}[\Delta^{-1}\mu_{BL} + (\Sigma_h/T)^{-1}\mu_h]$$
$$\Sigma_z = \Sigma_h + [(\Delta^{-1} + (\Sigma_h/T)^{-1}]^{-1}$$
$$\Delta = \Sigma_{bl}$$

The same derivation holds as above for the posterior posterior predictive distribution, which is why the extra Σ_h is added to the posterior to calculate Σ_z.

Appendix 2 - Extension Derivations

Standard Inverse-Wishart Derivation

The estimation of μ is consistent with the methods above, the difference between the models is only in Σ. The derivation for the posterior value of Σ with an inverse-Wishart prior is shown below. In this derivation, v_0 is the degrees of freedom, p is the number of assets, and n is the number of observations in the dataset.

Prior Specification

$$p(\Sigma) = \mathcal{IW}(\Psi^{-1}, v_0)$$

$$p(\Sigma) = \frac{|\Psi|^{\frac{v_0}{2}}}{|\Sigma|^{\frac{v_0+p+1}{2}} \, 2^{\frac{v_0 p}{2}} \, \Gamma_p(\frac{v_0}{2})} \, exp[-\frac{1}{2} tr(\Psi T^{-1})]$$

$$p(\Sigma) \propto |\Sigma|^{-\frac{v_0+p+1}{2}} \, exp[-\frac{1}{2} tr(\Psi \Sigma^{-1})]$$

Likelihood Specification

$$p(y_1, ..., y_n \,|\, \mu, \Sigma) = 2\Pi^{-\frac{np}{2}} \, |\Sigma|^{-\frac{n}{2}} exp[-\frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)' \Sigma^{-1} (y_i - \mu)]$$

$$p(y_1, ..., y_n \,|\, \mu, \Sigma) \propto |\Sigma|^{-\frac{n}{2}} [-\frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)' \Sigma^{-1} (y_i - \mu)]$$

$$\text{Let } S_\mu = \sum_{i=1}^{n} (y_i - \mu)(y_i - \mu)'$$

$$p(y_1, ..., y_n \,|\, \mu, \Sigma) \propto |\Sigma|^{-\frac{n}{2}} \, exp[-\frac{1}{2} tr(S_\mu \Sigma^{-1})]$$

Bayesian Updating

$$p(\Sigma \,|\, \mu, y_1, ..., y_n) \propto p(\Sigma) \times p(y_1, ..., y_n \,|\, \mu, \Sigma)$$

$$\propto |\Sigma|^{-\frac{n}{2}} \, exp[-\frac{1}{2} tr(S_\mu \Sigma^{-1})] \times |\Sigma|^{-\frac{m+p+1}{2}} \, exp[-\frac{1}{2} tr(\Psi \Sigma^{-1})]$$

$$\propto |\Sigma|^{-\frac{v_0+n+p+1}{2}} \, exp\{-\frac{1}{2} - tr[(\Psi + S_\mu)\Sigma^{-1}]\}$$

This is the kernal of the inverse-Wishart density, $\mathcal{IW}((\Psi + S_\mu)^{-1}, v_0 + n)$, which has a mean of

$$E[\Sigma \,|\, \mu, y_1, ..., y_n] = \frac{1}{v_0 + n - p - 1} (\Psi + S_\mu)$$

Normal Inverse-Wishart Parameters

In this section I will present the posterior and posterior predictive values under the normal-inverse-Wishart prior. Note: n (not S like throughout the paper) refers to the number of historical data observations.

Prior Specification

$$\Sigma \sim \mathcal{IW}(\Psi^{-1}, v_0)$$
$$\mu|\Sigma \sim \mathcal{N}(\mu_0, \Sigma/k_0)$$
$$p(\mu, \Sigma) \overset{\text{def}}{=} \mathcal{NIW}(\mu_0, k_0, \Psi, v_0)$$

Likelihood Specification

$$L(\mu, \Sigma; \Phi) \overset{\text{def}}{=} \mathcal{N}(\mu_h, \Sigma_h)$$

Posterior Calculation

$$P(\mu, \Sigma \,|\, \mu_0, k_0, \Psi, v_0, \mu_h, \Sigma_h) = \mathcal{NIW}\,(\mu_n, k_n, \Psi_n, v_n)$$
$$\mu_n = \frac{k_0}{k_0 + n}\mu_0 + \frac{n}{k_0 + n}\mu_h$$
$$k_n = k_0 + n$$
$$v_n = v_0 + n$$
$$\Psi_n = \Psi + \Sigma_h + \frac{k_0 n}{k_0 + n}(\mu_h - \mu_{BL})(\mu_h - \mu_{BL})'$$

The posterior predictive distribution, with parameters defined above, is

$$P(r_{T+1}|\Phi) = t_{v_n - n + 1}\left(\mu_n, \frac{\Psi_n\,(k_n + 1)}{k_n\,(v_n - n + 1)}\right)$$

38

Appendix 3 - Baseline Sensitivity Tables

All Values in the following sensitivity table are percentages (%)

Markowitz Gamma Sensitivity

Table 4: Markowitz Returns Gamma

γ	1	2.5	5	10	20
Mean	-0.18	2.82	3.82	4.32	4.57
SD	99.84	39.94	19.98	10.01	5.08
Skewness	-0.38	-0.39	-0.39	-0.40	-0.40
Kurtosis	4.47	4.46	4.44	4.39	4.26

Black-Litterman

$\tau = .025$ is used in the Gamma sensitivity table, and $\gamma = 2.5$ is used in the Tau sensitivity table.

Table 5: BL Gamma Sensitivity

γ	1	2.5	5	10	20
Mean	7.24	7.24	7.24	7.24	7.24
SD	15.02	15.02	15.02	15.02	15.02
Skewness	-0.51	-0.51	-0.51	-0.51	-0.51
Kurtosis	4.19	4.19	4.19	4.19	4.19

Table 6: BL Tau Sensitivity

τ	0.01	0.025	0.1	0.5	1
Mean	7.27	7.24	7.07	6.47	6.06
SD	15.24	15.02	13.99	10.24	7.68
Skewness	-0.51	-0.51	-0.51	-0.52	-0.53
Kurtosis	4.19	4.19	4.20	4.24	4.28

Zhou

$\tau = .025$, $\gamma = 2.5$ and S = 60 are used to calculate the results below when the parameters are not varied within table

Table 7: Zhou Gamma Sensitivity

γ	1	2.5	5	10	20
Mean	2.83	4.62	5.21	5.51	5.66
SD	63.12	27.77	16.23	10.72	8.20
Skewness	-0.42	-0.47	-0.52	-0.57	-0.58
Kurtosis	4.57	4.70	4.82	4.85	4.72

Table 8: Zhou Tau Sensitivity

τ	0.01	0.025	0.1	0.5	1
Mean	5.62	4.62	3.48	3.00	2.93
SD	21.94	27.77	35.08	38.32	38.79
Skewness	-0.54	-0.47	-0.41	-0.39	-0.39
Kurtosis	4.84	4.70	4.53	4.47	4.46

Table 9: Zhou Sample Size (S) Sensitivity

S	15	30	60	120	240
Mean	6.06	5.37	4.62	3.95	3.47
SD	19.35	23.09	27.77	32.18	35.45
Skewness	-0.57	-0.52	-0.47	-0.43	-0.41
Kurtosis	4.83	4.82	4.70	4.59	4.53

Appendix 4 - Extension Sensitivity Tables

Equil-Historical

Tables 14 and 15 use values of $\tau = .025$, $\gamma = 2.5$ and S $= 60$
Tables 16 and 17 use values of $\tau = .025$, $\gamma = 2.5$, $v_0 = $ n+2, SS $= 30$

Table 10: Equil-Historical v_0, SS Return Sensitivity

		v_0				
		n+2	15	30	60	120
	1	7.53	6.29	5.75	5.39	5.12
	5	9.08	7.26	6.53	6.10	5.74
SS	10	9.61	7.86	6.94	6.44	6.06
	30	10.10	8.94	7.82	7.10	6.62
	60	10.26	9.52	8.51	7.67	7.06
	120	10.35	9.92	9.18	8.35	7.61

Table 11: Equil-Historical v_0, SS Volatility Sensitivity

		v_0				
		n+2	15	30	60	120
	1	28.29	29.18	28.87	28.54	28.28
	5	26.52	28.63	29.19	29.11	28.86
SS	10	26.14	27.88	28.96	29.20	29.09
	30	25.92	26.65	27.94	28.81	29.16
	60	25.89	26.20	27.10	28.13	28.86
	120	25.87	26.00	26.46	27.29	28.22

Table 12: Equil-Historical S Return Sensitivity

15	30	60	120	240
10.85	10.61	10.10	9.45	8.69

Table 13: Equil-Historical S Volatility Sensitivity

15	30	60	120	240
22.82	24.51	25.92	27.04	28.02

Equil-CAPM

Tables 14 and 15 use values of $\tau = .025$, $\gamma = 2.5$.
Tables 16 and 17 use values of $\tau = .025$, $\gamma = 2.5$, $v_0 = $ n+2, SS $= 5$

Table 14: Equil-CAPM v_0, SS Return Sensitivity

		v_0				
		n+2	**15**	**30**	**60**	**120**
	1	6.51	4.98	4.81	4.76	4.74
	5	19.42	8.03	6.60	6.21	6.05
SS	**10**	32.98	10.71	7.66	6.82	6.48
	30	69.89	19.98	10.99	8.38	7.33
	60	100.33	31.89	15.52	10.42	8.32
	120	129.92	50.93	23.79	14.29	10.21

Table 15: Equil-CAPMl v_0, SS Volatility Sensitivity

		v_0				
		n+2	**15**	**30**	**60**	**120**
	1	40.93	40.05	39.97	39.95	39.94
	5	57.90	42.43	41.26	40.97	40.87
SS	**10**	84.40	45.25	42.15	41.46	41.20
	30	171.35	59.07	45.63	42.83	41.89
	60	251.73	82.21	51.80	44.98	42.79
	120	336.27	124.86	66.02	49.99	44.74

Table 16: Equil-CAPM S Return Sensitivity

15	**30**	**60**	**120**	**240**
42.71	26.08	15.52	9.48	6.24

Table 17: Equil-CAPM S Volatility Sensitivity

15	**30**	**60**	**120**	**240**
104.92	69.98	51.80	44.26	41.54

BL-Historical

Tables 18 and 19 use values of $\tau = .025$, $\gamma = 2.5$.

Table 18: BL-Historical v_0, S Return Sensitivity

		v_0				
		n+2	**15**	**30**	**60**	**120**
	15	4.97	5.02	5.12	5.27	5.46
	30	4.94	4.96	4.99	5.05	5.12
S	**60**	4.75	4.75	4.74	4.72	4.70
	120	4.38	4.37	4.34	4.30	4.25
	240	3.91	3.91	3.89	3.86	3.81

Table 19: BL-Historical v_0, S Volatility Sensitivity

		v_0				
		n+2	**15**	**30**	**60**	**120**
	15	2.47	3.16	4.70	7.06	10.05
	30	4.89	5.59	7.12	9.50	12.61
S	**60**	9.52	10.11	11.42	13.53	16.42
	120	16.33	16.71	17.58	19.06	21.23
	240	23.82	24.00	24.43	25.20	26.43

BL-CAPM

Tables 20 and 21 use values of $\tau = .025$, $\gamma = 2.5$.

Table 20: BL-CAPM v_0, S Return Sensitivity

		v_0				
		n+2	**15**	**30**	**60**	**120**
	15	20.84	25.90	36.83	53.36	74.25
	30	29.14	32.72	40.56	52.70	68.56
S	**60**	36.52	38.50	42.94	50.07	59.88
	120	38.27	39.07	40.89	43.97	48.53
	240	32.79	33.01	33.53	34.47	35.97

Table 21: BL-CAPM v_0, S Volatility Sensitivity

		v_0				
		n+2	**15**	**30**	**60**	**120**
	15	43.25	56.89	86.31	130.72	186.69
	30	62.85	72.08	92.27	123.48	164.16
S	**60**	79.47	84.42	95.49	113.26	137.70
	120	84.28	86.26	90.83	98.52	109.93
	240	75.40	75.99	77.38	79.86	83.85

NIW

Tables 22 and 23 use values of $\tau = .025$, $\gamma = 2.5$.
Tables 24 and 25 use values of $\gamma = 2.5$ and S = 15.
Tables 25 and 26 use values of $\tau = .025$, $\gamma = 2.5$ and S = 15.

Table 22: NIW $v_0 = k_0$, S Return Sensitivity

		$v_0 = k_0$				
	n+2	**15**	**30**	**60**	**120**
	15	4.53	5.28	7.33	11.48	19.01
	30	3.82	4.22	5.29	7.49	11.75
S	**60**	3.37	3.59	4.15	5.29	7.56
	120	3.11	3.23	3.52	4.11	5.29
	240	2.97	3.03	3.18	3.49	4.09

Table 23: NIW $v_0 = k_0$, S Volatility Sensitivity

		$v_0 = k_0$				
	n+2	**15**	**30**	**60**	**120**
	15	32.37	37.87	50.66	74.87	118.73
	30	34.97	37.48	43.51	55.41	78.92
S	**60**	37.04	38.14	40.88	46.41	57.84
	120	38.36	38.85	40.09	42.62	47.89
	240	39.11	39.34	39.91	41.07	43.49

Table 24: NIW $v_0 = k_0$, τ Return Sensitivity

		$v_0 = k_0$				
	n+2	**15**	**30**	**60**	**120**
	0.01	4.53	5.29	7.41	11.88	20.60
	0.025	4.53	5.28	7.33	11.48	19.01
τ	**0.1**	4.54	5.24	7.01	10.03	14.26
	0.5	4.58	5.11	6.11	7.23	8.21
	1	4.62	5.03	5.67	6.27	6.70

Table 25: NIW $v_0 = k_0$, τ Volatility Sensitivity

		n+2	15	30	60	120
		$v_0 = k_0$				
τ	0.01	32.66	38.43	52.14	79.17	131.87
	0.025	32.37	37.87	50.66	74.87	118.73
	0.1	31.01	35.31	44.37	58.89	79.24
	0.5	25.32	25.95	26.69	27.52	28.52
	1	20.60	19.49	17.81	16.51	15.83

Table 26: NIW v_0, k_0, Return Sensitivity

		n+2	15	30	60	120
		k_0				
v_0	n+2	4.53	5.17	6.02	6.71	7.17
	15	4.43	5.28	6.41	7.31	7.92
	30	4.20	5.54	7.33	8.77	9.73
	60	3.76	6.02	9.05	11.48	13.11
	120	2.96	6.85	12.04	16.22	19.01

Table 27: NIW v_0, k_0, Volatility Sensitivity

		n+2	15	30	60	120
		k_0				
v_0	n+2	32.37	28.67	24.15	21.10	19.50
	15	42.75	37.87	31.91	27.89	25.77
	30	67.85	60.10	50.66	44.29	40.93
	60	114.64	101.58	85.63	74.87	69.21
	120	196.62	174.24	146.91	128.45	118.73