FLORIDA STATE UNIVERSITY
COLLEGE OF ARTS AND SCIENCES

BAYESIAN PORTFOLIO OPTIMIZATION WITH TIME-VARYING FACTOR MODELS

By

FENG ZHAO

A Dissertation submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
?????? Semester, ??????
ABSTRACT

We develop a modeling framework to simultaneously evaluate various types of predictability in stock returns, including stocks’ sensitivity (“betas”) to systematic risk factors, stocks’ abnormal returns unexplained by risk factors (“alphas”), and returns of risk factors in excess of the risk-free rate (“risk premia”). Both firm-level characteristics and macroeconomic variables are used to predict stocks’ time-varying alphas and betas, and macroeconomic variables are used to predict the risk premia. All of the models are specified in a Bayesian framework to account for estimation risk, and informative prior distributions on both stock returns and model parameters are adopted to reduce estimation error. To gauge the economic significance of the predictability, we apply the models to the U.S. stock market and construct optimal portfolios based on model predictions. Out-of-sample performance of the portfolios is evaluated to compare the models.
CHAPTER 1

INTRODUCTION

Predicting stock returns has been a constant pursuit of both finance academics and investment professionals. A large body of literature has documented evidence that stock returns are predictable to some extent, and a growing number of investment firms have been utilizing quantitative models to take advantage of the predictability for stock selection and portfolio construction. Numerous macroeconomic variables and firm-level characteristics have been identified as predictive of stock returns.¹ The macroeconomic variables include those related to business cycle (e.g., industrial production, GDP growth), financial market condition (e.g., dividend yield, credit spread, bond yield), monetary policy (e.g., Fed funds rate, money supply), and others. The list of firm-level characteristics is equally long, including measures of firm size (e.g., market capitalization), stock valuation (e.g., book-to-price, price-to-earnings), operating quality (e.g., profit margin, asset turnover), financial risk (e.g., debt-to-equity), technical factors (e.g., price momentum, trading volume), and many more. Moreover, researchers have found that returns to firm-level characteristics vary under different macroeconomic conditions, giving rise to a growing body of literature related to “factor timing” or “style rotation”.

Finance theories, such as the Capital Asset Pricing Model (CAPM) of Sharpe (1964) and Lintner (1965), the Intertemporal Capital Asset Pricing Model (ICAPM) of Merton (1973), and the Arbitrage Pricing Theory (APT) of Ross (1976), postulate that stock returns are compensation for the stock’s co-movement with systematic, nondiversifiable risk factors. In efficient markets, stock returns should be unpredictable after accounting for the stock’s exposure to systematic risk factors. Competing theories exist to explain the root of

¹Researchers have also found other sources of predictability, such as calendar date (e.g. the “January effect”) and even weather, and provided explanations rooted in behavioral or cognitive psychology. We do not explore these sources of predictability in this paper.
the documented evidence of stock return predictability. Some argue that, consistent with the ICAPM or APT, the documented predictability is due to additional unidentified risk factors, stocks’ time-varying sensitivity (i.e. “betas”) to risk factors, or time-varying risk premium. See, e.g., Lettau and Ludvigson (2001) and Avramov and Chordia (2006). Another explanation for the predictability is market inefficiency caused by investors’ behavioral or cognitive biases, such as over-confidence, biased self-attribution and emotion-driven decision making, which give rise to predictable abnormal returns (i.e., “alphas”) even after controlling for the exposure to the risk factors. (See, e.g., Barberis, Shleifer and Vishny (1998), Daniel, Hirshleifer and Subrahmanyam (1998), and Hong and Stein (1999).) Still, some people believe the documented predictability is nothing but transient patterns uncovered by researchers’ extensive mining of historical data, which will not persist in the future.

For example, there is still on-going debate on whether the firm size and book-to-price effects are attributable to risk factors or evidence of market inefficiency. Fama and French (1993, 1996) argue that firms with small size or high book-to-price ratio are more sensitive to conditions of financial distress, a systematic risk factor. They extend the CAPM to add two mimicking portfolios for firm size and book-to-price as proxy for a risk factor related to financial distress. Daniel and Titman (1997), however, show that the size and book-to-price characteristics themselves, rather than the co-movement of stocks with the size and book-to-price mimicking portfolios, explain cross-sectional variation in stock returns. They conclude that predictability related to size and book-to-price can not be explained by risk factors.\(^2\)

In this paper, we develop a modeling framework to simultaneously evaluate various types of stock return predictability identified in previous research, including predictability in the variation of alphas, betas and risk premium. We examine major categories of firm-level characteristics (the “alpha” factors), such as value, quality, and price momentum.\(^3\) We further examine whether returns to these alpha factors vary over time, and if so, whether


\(^3\)In the jargon of investment professionals, characteristic-based factors can be either alpha factors or risk factors. Alpha factors are characteristics that tend to provide consistently positive returns over time. Risk factors are those that do not provide consistent outperformance over time but are predictive of the variance-covariance of stock returns (e.g., characteristics in the BARRA Risk Model). In models of factor timing, however, the distinction between risk and alpha factors becomes blurred, as a risk factor may provide consistently positive returns under certain macroeconomic conditions, thus becoming an alpha factor under those conditions. In this paper, we refer to all characteristic-based factors as alpha factors, and portfolio-based factors (such as the market portfolio in CAPM) as risk factors.
they can be predicted by macroeconomic variables. Previous research on alpha factors often focuses on studying the predictive power of a particular characteristic of interest after controlling for a risk model. Previous research on factor timing also tends to examine how macroeconomic variables affect the returns to each individual characteristic. In our model, we simultaneously examine all of the characteristics under various macroeconomic conditions.

Within the same model, we predict time-varying betas using both firm-level characteristics and macroeconomics. Moreover, the risk premium is allowed to change over time, and are predicted using macroeconomic variables. In addition to the full model, we estimate several reduced versions of the model, each ignoring certain types of predictability. By comparing the performance of the different models, we shed light on the sources of stock return predictability.

Another objective of this paper is to assess the economic significance of the predictability in stock returns. We demonstrate whether investors can use the predictive models developed in this paper to construct portfolios with performance superior to certain benchmark portfolios. Realistic constraints and transaction costs are considered in portfolio construction.

Portfolio construction is often cast in the mean-variance setting. Investors are assumed to prefer higher expected return but dislike higher return volatility. For example, the investor may be interested in maximizing expected return while maintaining a hurdle on return volatility. Alternatively, the investor may want to minimize the volatility of the portfolio or maximize the Sharpe Ratio\(^4\). In all these cases, the expected returns and covariance matrix of individual stocks in the investment universe are required inputs for constructing the optimal portfolio. In practice, estimates of the expected returns and covariance matrix derived from a model are fed into an optimization algorithm as if they were the true values. Estimation risk plays an important role in portfolio optimization. Ignoring the estimation error in the expected returns and covariance matrix often results in concentrated portfolios with extremely large and unstable weights for a few stocks and poor out-of-sample performance. Prior research has found that accounting for estimation risk using a Bayesian framework can improve the performance of the constructed portfolio. See, e.g., Kandel and Stambaugh (1996), Barberis (2000), Pastor (2000), Pastor and Stambaugh (1999, 2000, 2002), Avramov (2002, 2004), Tu and Zhou (2004), and Avramov and Wermers (2006). In this paper, we

\(^4\)The Sharpe Ratio is a performance measure popularized by William Sharpe, which is defined as the ratio of the mean to the standard deviation of the portfolio return in excess of the risk free rate.
also take a Bayesian approach and account for estimation risk in predicted stock returns by integrating out the model parameters based on their posterior distribution. Markov Chain Monte Carlo (MCMC) techniques are used to estimate the model parameters and the predictive distribution of stock returns.

In the empirical part of the paper, we apply the proposed modeling framework to the U.S. stock market. We build time-varying factor models to predict the returns of stocks in the S&P 500 Index over a 6-month holding period. We update the models and re-construct optimal portfolios every 6 months. Out-of-sample performance of the optimal portfolios are compared to that of the benchmark portfolio. In addition to the full model, we also fit several restricted models in which certain types of predictability is ignored. By comparing the performance of the optimal portfolio based on the full model to that of restricted models, we shed light on the sources of return predictability.

In related work, Avramov and Chordia (2006) also study whether macroeconomic variables can predict stock returns. Using the CAPM as the risk model, they model the alphas and betas of individual stocks, as well as the equity risk premium, as functions of macroeconomic variables. They also estimate their model using a Bayesian framework. However, they allow the estimates of model parameters to vary by stock by fitting a separate model for each stock. In their empirical work, they find that stock returns are predictable due to variation in alpha, but find little evidence of predictability in the equity risk premium.

Our approach differs from Avramov and Chordia (2006) in several important ways. First, we explicitly incorporate characteristic-based factors into the modeling of alpha variation. As mentioned earlier, previous research has identified numerous characteristics as alpha factors. Including these characteristics will likely improve the model’s predictive power. It also allows us to examine some important questions about the sources of alpha variation, e.g., whether alpha is mainly due to consistent outperformance of the alpha factors or due to time-varying returns to the factors that can be predicted by macroeconomic variables.

Second, we model stocks’ betas to risk factors as functions of macroeconomic variables, firm-level characteristics (such as industry association) and their interactions. Instead of treating individual stocks as unrelated assets, we postulate that shared characteristics play an important role in explaining the variation in betas. For example, stocks that are in the same industry and are similar in key characteristics (e.g., size, leverage, profit margin, etc.) tend to have similar sensitivities to risk factors. By incorporating firm-level characteristics
into the modeling of betas, we hope to improve the robustness of the estimates and the overall performance of the model.

Another benefit of including firm-level characteristics in the modeling of alphas and betas is that we end up with a much more parsimonious model. By fitting a separate model for each stock, Arvamov and Chordia (2006) estimates tens of thousands of parameters to predict the expected returns. In contrast, less than a hundred parameters are required in our model to predict the expected returns. We hope our parsimonious specification further reduces estimation error and leads to better portfolio performance.

Third, we take advantage of the Bayesian approach by incorporating two types of informative prior distributions. One type of informative prior distribution is on the mean and covariance matrix of future stock returns. Researchers have found that, for a big stock universe, the sample covariance matrix of stock returns has large estimation error, due to the large number of parameters estimated from the data. Bayesian techniques have been proposed to combine covariance estimates obtained from the data with informative prior distributions in order to reduce estimation error. For example, Ledoit and Wolf (2003) show that better portfolio performance can be achieved by “shrinking” the sample covariance matrix toward a matrix with constant correlation. We adopt the shrinkage idea and set the covariance matrix of the prior distribution of future stock returns according to a simple one-factor risk model. In addition, we derive the the mean of the prior distribution based on the stocks’ relative weights in the benchmark portfolio. This is similar to the Black-Litterman model (1992), in which they combine an investor’s judgmental view with expected returns implied by stock weights in the market portfolio. However, our motivation is different. While the market implied expected returns in the Black-Litterman model are interpreted as the equilibrium returns in a CAPM world, our objective is to “shrink” the model prediction of expected returns to those implied by the benchmark portfolio, so that the stock weights in the optimized portfolio are “shrunk” towards the stock weights in the benchmark portfolio.

Another type of informative prior knowledge is on the distribution of model intercept and coefficients. In cross-sectional studies, Pastor and Stambaugh (1999) and Pastor (2000) specify an informative prior on the model intercepts. In predicting the risk premium of the stock and bond markets, Wachter and Warusawitharana (2009) specify an informative prior on the model coefficients, which is linked to an informative prior on the $R^2$ of the regression model. We extend the ideas of previous research and adopt informative prior distributions
on both the intercept and coefficients of the regression models. The informative priors are normal distributions with mean 0 and a finite variance. Effectively, the informative priors reduce estimation error by “shrinking” the model intercept and coefficients towards zero.

The rest of the paper is organized as follows. In Chapter 2, we specify the models, discuss the informative prior distributions, and derive the estimation methods. In Chapter 3, we test our estimation algorithm on a simulated data set. Details of the real data used in the empirical application are provided in Chapter 4. In Chapter 5, we report the empirical results of model estimation and portfolio optimization. Chapter 6 concludes.
CHAPTER 2

PROPOSED MODELS

In this chapter, we specify the models and derive the estimation methods. Section 2.1 specifies the models. Because our proposed models are motivated by the model of Avramov and Chordia (2006), we provide the specifications of both their model and our models and compare their differences. In addition to the full version of our proposed model, we introduce several models of reduced form by restricting certain types of predictability, in order to assess the sources of return predictability. In Sections 2.2 and 2.3, details of two types of informative prior distributions (for stock returns and for model parameters) are developed. We estimate our models using Markov Chain Monte Carlo (MCMC) technique, and derive the conditional posterior distributions in Section 2.4. Section 2.5 describes the portfolio optimization framework based on model predictions.

2.1 Model Specifications


We first provide the model specification of Avramov and Chordia (2006). For the ease of comparison between their model and our proposed models, we have changed the variable notations and matrix representations used in their model; however, the model structure is not altered. Their model is specified as follows. For $t = 0, \ldots, T - 1$, 

\begin{align*}
    r_{t+1} &= \mathbf{a}_{t+1} + \mathbf{B}_{t+1} f_{t+1} + \mathbf{\epsilon}_{t+1} \quad (2.1) \\
    \mathbf{a}_{t+1} &= \Lambda_0 \mathbf{z}_t \quad (2.2) \\
    \mathbf{B}_{t+1} &= (\mathbf{I}_{N_t} \otimes \mathbf{z}_t') \Theta_0 \quad (2.3) \\
    f_{t+1} &= \Delta_0 \mathbf{z}_t + \mathbf{e}_{t+1} \quad (2.4) \\
    \mathbf{z}^*_t &= \Delta_1 \mathbf{z}_t + \mathbf{v}_{t+1} \quad (2.5) \\
    \mathbf{\epsilon}_{t+1} &\sim N(\mathbf{0}, \mathbf{\Sigma}_{t+1}), \quad \mathbf{e}_{t+1} \sim N(\mathbf{0}, \mathbf{\Omega}), \quad \mathbf{v}_{t+1} \sim N(\mathbf{0}, \mathbf{\Phi}). \quad (2.6)
\end{align*}

The following variables are observable from the data:

- \( r_{t+1} \) is an \( N_t \)-vector of Period \((t+1)\) stock returns in excess of the risk-free rate for the \( N_t \) stocks in the investment universe available at the end of Period \( t \). Note that \( N_t \) varies by \( t \) as the investment universe can change over time due to new stocks entering the market and old stocks disappearing because of mergers, bankruptcy and etc.

- \( f_{t+1} \) is a \( K \)-vector of returns in excess of the risk-free rate for the \( K \) portfolio-based factors (i.e., the risk factors) in Period \((t+1)\). For example, the CAPM is a one-factor risk model in which the risk factor is the market return.

- \( \mathbf{z}^*_t \) is an \( M \)-vector of macroeconomic variables observable at the end of Period \( t \). \( \mathbf{z}_t = (1, \mathbf{z}_t^*) \) is the \((M+1)\)-vector with the first entry being the constant scalar 1.

The following notations are either unobservable random noises or parameters to be estimated from the model.

- \( \mathbf{a}_{t+1} \) is an \( N_t \)-vector of abnormal returns ("alphas") unexplained by the portfolio-based factors in Period \((t+1)\).

- \( \mathbf{B}_{t+1} \) is an \( N_t \times K \) matrix of risk loadings ("betas") of the stocks to the portfolio-based factors.
\[ \Lambda_0 \] is an \( N_t \times (M+1) \) matrix of coefficients predicting the alphas \( a_{t+1} \) as linear functions of the macroeconomic variables \( z_t \). The model assumes that the coefficients vary by stock and estimates a separate set of coefficients for each stock (with the \( i \)-th row of \( \Lambda_0 \) being the coefficients for the \( i \)-th stock).

\[ \Theta_0 \] is a \( N_t(M+1) \times K \) matrix of coefficients predicting the risk loadings \( B_t \) as linear functions of the macroeconomic variables \( z_t \). The \( \otimes \) sign in \( I_{N_t} \otimes z_t' \) denotes the Kronecker product, and \( I_{N_t} \) denotes the \( N_t \times N_t \) identity matrix. Again, the model assumes that the coefficients vary by stock and estimates a separate set of coefficients for each stock.

\[ \Delta_0 \] is a \( K \times (M+1) \) matrix of coefficients predicting the excess returns to the portfolio-based factors \( f_{t+1} \) as linear functions of the macroeconomic variables \( z_t \).

\[ \Delta_1 \] is an \( M \times (M+1) \) matrix of coefficients predicting the macroeconomic variables \( z_{t+1}^* \) as linear functions of their lagged values \( z_t^* \) (with an intercept). That is, the macroeconomic variables follow a vector autoregressive model VAR(1).

\[ \epsilon_{t+1} \] is an \( N_t \)-vector of idiosyncratic stock returns, distributed as \( N(0, \Sigma_{t+1}) \). Let \( N \) be the total number of unique stocks available at the end of any period \( t \), \( t = 0, \ldots, T-1 \), and let \( \Sigma = \text{diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2\} \) be the covariance matrix of the idiosyncratic stock returns. Then \( \Sigma_{t+1} \) is an \( N_t \times N_t \) sub-matrix of \( \Sigma \) with entries for the \( N_t \) stocks available at the end of Period \( t \). \( \Sigma \) is assumed to be a diagonal matrix, because the correlation of returns among stocks is explained by the risk factors. \( \epsilon_{t+1} \) is a \( K \)-vector of random noise distributed as \( N(0, \Omega) \). \( v_{t+1} \) is an \( M \)-vector of random noise distributed as \( N(0, \Phi) \). \( \epsilon_{t+1} \) and \( \epsilon_{t+1} \) are assumed to be uncorrelated. The covariance matrix between \( \epsilon_{t+1} \) and \( v_{t+1} \) is \( V_{\epsilon v} \), and the covariance matrix between \( \epsilon_{t+1} \) and \( v_{t+1} \) is \( V_{ev} \).

To summarize, Equation (2.1) decomposes a stock’s excess return into three components: a component attributable to the stock’s correlation with systematic risk factors, an abnormal return component (“alpha”) unexplained by the risk factors, and an idiosyncratic component which is a white noise. Equations (2.2), (2.3) and (2.4) predict stocks’ alphas and betas, as well as the excess returns to the risk factors (“risk premia”), as linear functions of
macroeconomic variables. The alphas and betas are assumed to vary by stock, and a separate set of model coefficients are estimated for each stock. The same set of macroeconomic variables are used as predictors in each of the three regressions, and the macroeconomic variables themselves are assumed to follow a vector autoregressive VAR(1) process in Equation (2.5).

2.1.2 Specification of Our Proposed Models

The specification of our Full Model (Model A) is the following. For \( t = 0, \ldots, T - 1 \),

\[
\begin{align*}
\mathbf{r}_{t+1} &= \mathbf{a}_{t+1} + \mathbf{B}_{t+1}\mathbf{f}_{t+1} + \mathbf{e}_{t+1} \quad (2.8) \\
\mathbf{a}_{t+1} &= \mathbf{X}a_{t}\mathbf{z}_{at} \quad (2.9) \\
\mathbf{B}_{t+1} &= (\mathbf{X}b_{t} \otimes \mathbf{z'}_b)\mathbf{\Theta} \quad (2.10) \\
\mathbf{f}_{t+1} &= \Delta \mathbf{z}_{ct} + \mathbf{e}_{t+1} \quad (2.11) \\
\mathbf{e}_{t+1} &\sim N(\mathbf{0}, \Sigma_{t+1}), \quad \mathbf{e}_{t+1} \sim N(\mathbf{0}, \Omega), \quad \text{cov}(\mathbf{e}_{t+1}, \mathbf{e}_{t+1}) = 0. \quad (2.12)
\end{align*}
\]

Equation (2.8) is the same as Equation (2.1). Also, the notations \( \mathbf{r}_{t+1}, \mathbf{a}_{t+1}, \mathbf{B}_{t+1}, \mathbf{f}_{t+1}, \mathbf{e}_{t+1}, \Sigma_{t+1}, \) and \( \Omega \) are the same as in the model of Avramov and Chordia (2006). Our model removes several variables and parameters used in their model and adds the following new ones.

The following new variables are observable from the data:

- \( \mathbf{X}_{at} \) is an \( N_t \times (J_a + 1) \) matrix of \( J_a \) characteristic-based alpha factors observable at the end of Period \( t \), plus a constant vector of 1 as the first column. Similarly, \( \mathbf{X}_{bt} \) is an \( N_t \times (J_b + 1) \) matrix of \( J_b \) characteristic-based beta factors observable at the end of Period \( t \), plus a constant vector of 1 as the first column. For example, \( \mathbf{X}_{at} \) may include variables related to value, quality and price momentum of the stocks. \( \mathbf{X}_{bt} \) may include dummy variables indicating the stock’s industry classification, as stocks in the same industry tend to have similar betas to the risk factors.

- \( \mathbf{z}_{at} \) is an \( (M_a + 1) \)–vector of \( M_a \) macroeconomic variables observable at the end of Period \( t \) (plus the constant scalar 1 as the first row), which are predictive of the time-varying return to the alpha factors. Similarly, \( \mathbf{z}_{bt} \) is an \( (M_b + 1) \)–vector of \( M_b \)
macroeconomic variables observable at the end of Period $t$ (plus the constant scalar 1 as the first row), which are predictive of the time-varying betas. For example, both $z_{at}$ and $z_{bt}$ may include macroeconomic variables related to financial market condition (bull market vs. bear market), as both alpha factor returns and stocks’ betas may vary by financial market condition. $z_{ct}$ is an $(M_c + 1)$–vector of $M_c$ macroeconomic variables observable at the end of Period $t$ (plus the constant scalar 1 as the first row), which are predictive of the risk premia. For example, in the CAPM where the risk factor is the market risk premium, prior research has identified many macroeconomic variables that are predictive of the market risk premium, such as short-term interest rate, credit spread, consumption growth, and etc.

The following notations are new parameters to be estimated from the model.

- $\Lambda$ is a $(J_a + 1) \times (M_a + 1)$ matrix of coefficients predicting $a_{t+1}$ as linear functions of the macroeconomic variables $z_{at}$, the alpha factors $X_{at}$, and their interactions. Thus, the quantity $\Lambda z_{at}$ measures the time-varying return to the alpha factors $X_{at}$ in Period $(t + 1)$.

- $\Theta$ is a $(J_b + 1)(M_b + 1) \times K$ matrix of coefficients predicting the risk loadings (“betas”) $B_{t+1}$ as linear functions of the beta factors $X_{bt}$, the macroeconomic variables $z_{bt}$, and their interactions.

- $\Delta$ is a $K \times (M_c + 1)$ matrix of coefficients predicting the returns to the portfolio-based factors $f_{t+1}$ as linear functions of the macroeconomic variables $z_{ct}$.

Comparison with the Avramov and Chordia (2006) Model

Our model differs from the model of Avramov and Chordia (2006) in several ways, some of which have been outlined in Chapter 1. Here we provide a summary of the differences.

- When predicting time-varying alphas and betas as linear functions of macroeconomic variables, their model estimates a separate set of model coefficients for each stock. Prior research has documented many firm-level characteristics that are also predictive of stocks’ alphas or betas. We explicitly incorporate these characteristics as additional predictors of alphas and betas to improve the predictive power of the model. Another
benefit of incorporating firm-level characteristics is that our model is much more parsimonious: while their model estimates tens of thousands of model coefficients, our model only has a few dozen coefficients. We expect that the parsimony of our model will result in more robust parameter estimation and return forecast.

- Their model uses the same set of macroeconomic variables to predict the time-varying alphas, betas and risk premia. In our model, we use a different set of macroeconomic variables to predict the time-varying alphas, betas and risk premia. A large number of macroeconomic variables have been identified as predictive of the market risk premium; however, there has been limited research on which macroeconomic variables are predictive of the alphas and betas and on how they interact with the characteristic-based alpha and beta factors. In our empirical work, we only include macroeconomic variables related to the financial market condition (bull market vs. bear market) to predict the time-varying effect of alphas and betas. Again, the objective is to avoid overfitting and improve the robustness of the model.

- Their model assumes that the macroeconomic variables follow a VAR(1) process, and estimates the covariance matrices $V_{\epsilon v}$ and $V_{\epsilon v}$. Thus, Equations (2.1), (2.4) and (2.5) together form a Seemingly Unrelated Regression (SUR) model. If the VAR(1) model correctly describes the data generating process of the macroeconomic variables, the SUR model may improve the asymptotic efficiency of the parameter estimates. Although it is common in the literature to model macroeconomic variables using a VAR(1) model, the adequacy of such a model is questionable. Including the VAR(1) model may actually reduce the estimation efficiency of parameters in Equations (2.1) and (2.4) because more parameters need to be estimated from the same amount of data. In our model, we opt for a more parsimonious specification by omitting the VAR(1) model for the macroeconomic variables.

- Another important difference is in the estimation method and specification of prior distributions. Although Avramov and Chordia (2006) also account for estimation error of the parameters in portfolio optimization, they adopt only non-informative priors for the model parameters and estimate the model using the maximum likelihood method. In our model, we evaluate both non-informative and informative prior distributions.
for both stock returns and model parameters, in order to take full advantage of the Bayesian framework. We use Markov Chain Monte Carlo (MCMC) techniques to estimate the models.

Specification of Models of Reduced Form

Our full model (Model A) postulates that risk premia are predictable using macroeconomic variables and that stocks’ alphas and betas are predictable using both macroeconomic variables and firm-level characteristics. We further specify the following models of reduced form by restricting certain types of predictability. In our empirical work, we compare the performance of optimal portfolios constructed based on the different models in order to assess the sources of return predictability.

- **Model B: Static Alpha Model.** Replace $z_{at}$ in Equation (2.9) by the constant scalar of 1. This is a reduced form of Model A that assumes that returns to the alpha factors are not predictable using macroeconomic variables. Note that the alpha of individual stocks can still change over time due to changes in the stocks’ value of the alpha factors.

  - **Model B2: Constant Alpha Model.** Replace $z_{at}$ in Equation (2.9) by the constant scalar of 1, and replace $X_{at}$ in Equation (2.9) by the constant vector of $1$. This is a reduced form of Model B, which assumes that all stocks have the same alpha that does not change over time.

- **Model C: Static Beta Model.** Replace $z_{bt}$ in Equation (2.10) by the constant scalar of 1. This is a reduced form of Model A that assumes that stocks’ betas are not predictable by macroeconomic variables. Note that the betas of individual stocks can still change over time due to changes in the stocks’ value of the beta factors.

  - **Model C2: Constant Beta Model.** Replace $z_{bt}$ in Equation (2.10) by the constant scalar of 1, and replace $X_{bt}$ in Equation (2.10) by the constant vector of $1$. This is a reduced form of Model C, which assumes that all stocks have the same betas that do not change over time.

- **Model D: Constant Risk Premia Model.** Replace $z_{ct}$ in Equation (2.11) by the constant scalar of 1. This is a reduced form of Model A that assumes that the risk premia are constant over time.
2.2 Informative Prior Distributions for Stock Returns

When the number of model parameters is large relative to the size of the sample used to estimate the models, the estimation error can be substantial. Portfolios constructed using these model estimates as inputs often have poor out-of-sample performance. Estimates from a more parsimonious model, on the other hand, may be biased, but tend to have smaller estimation error. To reduce estimation error, we construct informative prior distributions for $r_{T+1}$ using simple models. We assume that the prior distribution of $r_{T+1}$ follows a normal distribution:

$$r_{T+1} \sim N(\mu_0, \Sigma_0). \quad (2.13)$$

To construct $\Sigma_0$, we adopt a single-factor risk model based on the market portfolio. Let $r_{i,t}$ denote the excess return of Stock $i$ in Period $t$, and $r_{M,t}$ denote the excess market return in Period $t$, we estimate each stock’s beta to the market by running the following regression:

$$r_{i,t} = \alpha_i + \beta_i r_{M,t} + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma_i^2)$$

Note that we do not necessarily believe that CAPM holds, as we do not restrict $\alpha_i$ to zero. Let $\hat{\beta}_i$ and $\hat{\sigma}_i$ be the estimates of $\beta_i$ and $\sigma_i$, respectively, and let $\hat{\sigma}_M$ be the sample standard deviation of the market return. All estimates are based on 10 years of monthly return data up to Period $T$, and the S&P 500 index is used as the proxy for the market. Let $\sigma_{i,j}$ denote the $(i,j)$–entry of $\Sigma_0$. We define

$$\sigma_{i,j} = \begin{cases} \hat{\beta}_i^2 \hat{\sigma}_M^2 + \hat{\sigma}_i^2, & \text{if } i = j \\ \hat{\beta}_i \hat{\beta}_j \hat{\sigma}_M^2, & \text{if } i \neq j. \end{cases}$$

We choose $\mu_0$ such that a mean-variance efficient investor who believes in the prior distribution would end up holding the benchmark portfolio. Thus $\mu_0$ depends on the choice of benchmark portfolio. Let $w_0$ be the vector of stock weights of the benchmark portfolio in terms of market value observed at the end of period $T$. Following Black and Litterman (1992), we assume that the investor is mean-variance efficient who chooses a portfolio with a weight vector $w$ that maximizes the following utility function

$$U(w) = w' \mu_0 - \frac{\gamma}{2} w' \Sigma_0 w. \quad (2.14)$$

Taking the derivative of Equation (2.14) with respect to $w$ and setting it to 0, we get

$$\mu_0 - \gamma \Sigma_0 w = 0.$$
If the investor holds the benchmark portfolio, i.e., $w = w_0$, then the implied expected return is

$$\mu_0 = \gamma \Sigma_0 w_0$$

The above derivation of $\mu_0$ is the same as in the Black-Litterman model (1992). However, the motivations are different. The Black-Litterman model builds a framework to tilt an investor’s judgmental views about expected stock returns towards the expected return implied by an equilibrium asset-pricing model. As such, $\mu_0$ represents the equilibrium expected return according to the CAPM, and $w_0$ is the vector of stock weights of the market portfolio. In our setting, we do not necessarily believe that the CAPM holds, and $\mu_0$ does not represent the equilibrium expected return. Rather, our objective is to tilt the optimal portfolio derived from our model toward the benchmark portfolio by “shrinking” individual stock weights in the optimal portfolio towards the stock weights in the benchmark portfolio. Hence, we advocate setting $w_0$ to the vector of stock weights in the chosen benchmark portfolio, which may not be the market portfolio. For example, when the objective is to construct a market-neutral long-short portfolio, the risk-free rate (e.g., 3-month T-bill yield) is often chosen as the benchmark portfolio. Thus $w_0$ is a zero vector. Consequently, $\mu_0$ is also a zero vector. That is, the prior distribution assumes that the expected excess return is zero for all stocks.

When the benchmark is the market portfolio, $\gamma$ is the risk aversion parameter of an investor who invests all of her wealth in the market portfolio. Avramov and Chordia (2006) compute the time-varying $\gamma$ for such an investor, and report that $\gamma$ has a mean of 1.97 and a standard deviation of 0.59 over the period between August 1972 and December 2003. For simplicity, we set $\gamma$ to the constant 2 for the purpose of computing $\mu_0$ in our empirical analysis.

### 2.3 Informative Prior Distribution for Model Coefficients and Intercepts

When no informative prior knowledge is available about the model coefficients and intercepts $(\Lambda, \Theta, \Delta)$, we specify a non-informative prior distribution using a diffuse normal distribution with mean 0 and infinite variance. The diffuse prior distribution assumes no knowledge on the magnitude of the intercepts and coefficients. However, it is reasonable to believe that extremely large alphas, and thus intercepts, are unlikely. Also, past empirical research
suggests that stock returns are only weakly predictable, and even the weak evidence of predictability is still been disputed. Since the magnitude of the model coefficients is related to the predictiveness of the regression models, it is reasonable to believe that extremely large coefficients are also unlikely. Therefore, incorporating informative prior distributions with finite variances may help reduce estimation error.

Several authors have incorporated this type of informative priors in earlier studies. Pastor and Stambaugh (1999) estimate stocks’ expected returns using factor-based pricing models. They specify several informative prior distributions for the model intercept allowing the variance of the distribution to take on a wide range of values. They also specify an informative prior on the model coefficients with the variance of the distribution estimated from the sample. Pastor (2000) studies the implications of uncertainty in asset pricing models on portfolio selection decisions, and compares several informative prior distributions on the model intercept, with both the mean and the variance of the distributions taking on a range of values. Wachter and Warusawitharana (2009) evaluates informative priors on the coefficients of asset pricing models in predicting the equity and bond risk premia, and show that the informative prior distributions are linked to prior knowledge on the $R^2$ of the asset pricing models.

We adopt the same ideas and explore informative priors on both the intercepts and coefficients of our regression models (i.e., all entries of $\Lambda$, $\Theta$, and $\Delta$). Our informative priors are assumed to be independent normal distributions with mean zero and a finite variance. Because the variance of coefficients depends on the variance of the corresponding predictor variables, Wachter and Warusawitharana (2009) specify the informative priors on “normalized” coefficients that are adjusted for the variance of predictors. To simply the specification of the prior distributions, we take a different approach. We normalize each continuous predictor in our model by subtracting its sample mean and dividing by its standard deviation, and fit the models using the normalized predictors. We then set the variance (or standard deviation) of our informative priors to a single constant. In our empirical work, we evaluate various values of the standard deviation, such as 1%, 2%, 5% and 10%.
2.4 Estimation Method and Conditional Posterior Distributions

In this section, we provide the method for estimating the model parameters and the distribution future stock returns. We focus on the full model (Model A) only. The reduced-form models B, B2, C, C2 and D have the same structure as the full model, except that some of the predictors are reduced to constants. The same estimation method applies to these models. We do not treat them separately in the discussion of estimation methods.

We explicitly take estimation risk into account when estimating the distribution of future stock returns. At the end of Period T, we estimate the predictive distribution of $r_{T+1}$ using all data available up to Period T. The mean and covariance matrix of this predictive distribution are used as inputs in portfolio optimization. The predictive distribution is obtained by integrating out the model parameters based on their posterior distribution:

$$p(r_{T+1} \mid D_T) = \int_{\Psi} p(r_{T+1} \mid \Psi, D_T) \ p(\Psi \mid D_T) \ d\Psi \quad (2.15)$$

where $\Psi$ denotes the collection of parameters in the model, and $D_T$ denotes data observable up to the end of Period T. $p(\Psi \mid D_T)$ is the joint posterior distribution of all parameters conditional on available data, and $p(r_{T+1} \mid \Psi, D_T)$ is the density function of $r_{T+1}$ given data and model parameters.

We use Markov Chain Monte Carlo (MCMC) techniques to estimate the model parameters and the predictive distribution of $r_{T+1}$. We first derive the conditional posterior distribution of each parameter, conditional on the data and all the other parameters. A Gibbs sampler is then used to iteratively draw samples of the model parameters from the conditional posterior distributions. At the end of each iteration, we also draw $f_{T+1}$ from its distribution conditional on the model parameters, and draw $r_{T+1}$ from its distribution conditional on the model parameters and $f_{T+1}$. The mean and covariance of the predictive distribution of $r_{T+1}$ are then computed from the sample.

In the remainder of this section, we derive the conditional posterior distributions of the model parameters. We also derive the conditional distribution of $r_{T+1}$, given both its non-informative and informative prior distributions.
Substituting Equations (2.9) and (2.10) into Equation (2.8), we get:

\[
\begin{align*}
    r_{t+1} &= X_{at} \Lambda z_{at} + (X_{bt} \otimes z'_{bt}) \Theta f_{t+1} + \epsilon_{t+1} \\
    &= (X_{at} \otimes z'_{at}) \text{vec}(\Lambda') + (X_{bt} \otimes z'_{bt} \otimes f'_{t+1}) \text{vec}(\Theta') + \epsilon_{t+1} \\
    &= (X_{at} \otimes z'_{at} X_{bt} \otimes z'_{bt} \otimes f'_{t+1}) \begin{pmatrix}
        \text{vec}(\Lambda') \\
        \text{vec}(\Theta')
    \end{pmatrix} + \epsilon_{t+1}
\end{align*}
\]

where \( \text{vec}(Y) \) denotes the vector formed by stacking the columns of any matrix \( Y \). Equation (2.17) is a linear regression model, in which \( \Lambda \) is modeled as linear functions of the characteristic-based factors \( X_{at} \), the macroeconomic variables \( z_{at} \) and their interactions, and \( \Theta \) is modeled as linear functions of the characteristic-based factors \( X_{bt} \), portfolio-based factors \( f_{t+1} \), macroeconomic variables \( z_{bt} \) and their interactions.

Equation (2.11) can be rewritten as

\[
f_{t+1} = (I_K \otimes z'_{at}) \text{vec}(\Delta') + e_{t+1}
\]

where \( I_K \) is a \( K \times K \) identity matrix. Equation (2.18) is also a linear regression model. Moreover, \( \epsilon_{t+1} \) and \( e_{t+1} \) are assumed to be uncorrelated.

The model parameters are \( \Lambda, \Delta, \Theta, \Sigma, \) and \( \Omega \). Further, \( r_{T+1} \) depends on both the model parameters and \( f_{T+1} \).

2.4.1 Conditional Posterior Distributions for \( \Sigma = \text{diag}\{\sigma^2_1, \sigma^2_2, \ldots, \sigma^2_N\} \)

We specify the prior distribution of \( \sigma^2_i, (i = 1, 2, \ldots, N) \) as a Scaled Inverse Chi-square distribution with variance \( S_0^2 \) and degrees of freedom \( d_0 \):

\[
\sigma^2_i \sim \frac{d_0 S_0^2}{\lambda d_0}
\]

with density function

\[
p(\sigma^2_i) \propto (\sigma^2_i)^{-\left(\frac{d_0}{2} + 1\right)} \exp\left\{-\frac{d_0 S_0^2}{2\sigma^2_i}\right\}
\]

The prior distributions are assumed to be independent among the \( N \) stocks. Further, we assume the prior distributions are non-informative. When \( d_0 <= 4 \), the distribution is diffuse with infinite variance. Thus, we set \( d_0 = 4 \). The value of \( S_0^2 \) is less critical for a non-informative prior. Rossi, Allenby and McCulloch (2005) suggest setting \( S_0^2 \) to a value close to the sample variance. We set \( S_0^2 = 100 \), which corresponds to a standard deviation of 10%.
Before deriving the conditional posterior distribution of \( \sigma_i^2 \), we define some additional terms. Let \( \mathcal{T}_i = \{t_1, t_2, \ldots, t_{n_i}\} \) be the collection of periods in which Stock \( i \) is available, where \( n_i \) is the dimension of \( \mathcal{T}_i \) (i.e., the number of periods in which Stock \( i \) is available). For \( t \in \mathcal{T}_i \), let \( r_{i,t+1} \) be the excess return of Stock \( i \) in Period \( t + 1 \), \( \epsilon_{i,t+1} \) be the idiosyncratic return of Stock \( i \) in Period \( (t+1) \), \( x_{ai,t} \) be the \((J_a+1)\) vector of alpha factors for Stock \( i \) observed at the end of Period \( t \), and \( x_{bi,t} \) be the \((J_b+1)\) vector of beta factors for Stock \( i \) observed at the end of Period \( t \). According to Equation (2.17),

\[
  r_{i,t+1} = (x'_{ai,t} \otimes z_{at}') x'_{bi,t} \otimes f'_{t+1} + \epsilon_{i,t+1}
\]

\( \epsilon_{i,t+1} \sim N(0, \sigma_i^2) \).

The density function of the conditional posterior distribution of \( \sigma_i^2 \) is:

\[
p(\sigma_i^2 | \{r_{i,t+1}\}, \{f_{i+1}\}, \Lambda, \Theta) \propto \prod_{t \in \mathcal{T}_i} p(r_{i,t+1} | \sigma_i^2, \Lambda, \Theta, f_{t+1}) p(\sigma_i^2)
\]

\[
\propto \prod_{t \in \mathcal{T}_i} \frac{1}{\sqrt{2\pi \sigma_i}} \exp \left\{ -\frac{\epsilon_{i,t+1}^2}{2\sigma_i^2} \right\} (\sigma_i^2)^{-\frac{d_0}{2}+\frac{1}{2}} \exp \left\{ -\frac{d_0 S_0^2 + n_i S^2}{2\sigma_i^2} \right\}
\]

\[
\propto (\sigma_i^2)^{-\frac{d_0+n_i+1}{2}} \exp \left\{ -\frac{d_0 S_0^2 + n_i S^2}{2\sigma_i^2} \right\}
\]

where

\[
S^2 = \frac{1}{n_i} \sum_{t \in \mathcal{T}_i} \epsilon_{i,t+1}^2
\]

Thus, the conditional posterior distribution of \( \sigma_i^2 \) is a Scaled Inverse Chi-square distribution

\[
[\sigma_i^2 | \{r_{i,t+1}\}, \{f_{i+1}\}, \Lambda, \Theta] \sim \frac{d_1 S_1^2}{\chi^2_{d_1}}
\]

with degrees of freedom \( d_1 = d_0 + n_i \) and variance \( S_1^2 = \frac{d_0 S_0^2 + n_i S^2}{d_0 + n_i} \).

### 2.4.2 Conditional Posterior Distributions for \( \Omega \)

We specify the prior distribution of \( \Omega \) as an Inverted Wishart\((g_0, G_0)\) with density function

\[
p(\Omega) \propto |\Omega|^{-(g_0+K+1)/2} \exp \left\{ \text{trace}(-G_0 \Omega^{-1} / 2) \right\}.
\]

19
We assume the prior distribution is non-informative by setting $g_0 = K + 3$, and $G_0 = I_K$, a $K \times K$ identity matrix.

The density function of the conditional posterior distribution of $\Omega$ is

$$p(\Omega \mid \{f_{t+1}\}, \Delta) \propto \left[ \prod_{t=0}^{T-1} p(f_{t+1} \mid \Delta, \Omega) \right] p(\Omega)$$

$$\propto |\Omega|^{-T/2} \exp \left\{ -\frac{1}{2} \sum_{t=0}^{T-1} (e'_{t+1} \Omega^{-1} e_{t+1}) \right\} |\Omega|^{-(g_0 + K + 1)/2} \exp \left\{ \text{trace}(-G_0 \Omega^{-1}/2) \right\}$$

$$\propto |\Omega|^{-(g_0 + K + T + 1)/2} \exp \left\{ \text{trace}(-G_1 \Omega^{-1}/2) \right\}$$

where $G_1 = G_0 + \sum_{t=0}^{T-1} (e_{t+1} e'_{t+1})$. Thus, the conditional posterior distribution of $\Omega$ is Inverted Wishart($g_0 + T$, $G_1$).

### 2.4.3 Conditional Posterior Distributions for $\Delta$

To derive the conditional posterior distribution of $\Delta$, we will apply the following theorem.

**THEOREM 1:** Suppose $y_t \sim N(\alpha_t + X_t \beta, \Sigma_t)$, and $\beta \sim N(\beta_0, \Psi_0)$. Further assume $y_{t_1}$ and $y_{t_2}$ are independent when $t_1 \neq t_2$. Then $[\beta \mid \{y_t\}] \sim N(\beta_1, \Psi_1)$, where

$$\Psi_1 = \left\{ \Psi_0^{-1} + \sum_t (X' \Sigma_t^{-1} X) \right\}^{-1}$$

$$\beta_1 = \Psi_1 \{ \Psi_0^{-1} \beta_0 + \sum_t (X' \Sigma_t^{-1} (y_t - \alpha_t)) \}.$$  

The proof of this theorem follows closely that of the Seemingly Unrelated Regression (SUR) model. (See Zellner (1971).)

We specify the prior distribution of vec($\Delta'$) as $N(q_0, Q_0)$, and set $q_0$ to a vector of zero, and $Q_0$ to a diagonal matrix. In the case of non-informative prior on model coefficients and intercepts, we set the diagonal elements of $Q_0$ to infinity (approximated by a very large number); in the case of informative prior, we set the values a finite number (such as the square of 1%, 2%, 5% and 10%).

From (2.18),

$$[f_{t+1} \mid \Delta, \Omega] \sim N((I_K \otimes z'_{ad}) \text{vec}(\Delta'), \Omega).$$
The density function of the conditional posterior distribution of $\text{vec}(\Delta')$ is

$$p(\text{vec}(\Delta') \mid \{f_{t+1}\}, \Omega) \propto \prod_{t=0}^{T-1} p(f_{t+1} \mid \Delta, \Omega) p(\text{vec}(\Delta')).$$

Applying Theorem 1, the conditional posterior distribution of $\text{vec}(\Delta')$ is:

$$[\text{vec}(\Delta') \mid \{f_{t+1}\}, \Omega] \sim N(q_1, Q_1),$$

with

$$Q_1 = \left\{ Q_0^{-1} + \sum_{t=0}^{T-1} [(I_K \otimes z'_{ct})' \Omega^{-1} (I_K \otimes z_{ct})] \right\}^{-1},$$

$$q_1 = Q_1 \left\{ Q_0^{-1} q_0 + \sum_{t=0}^{T-1} [(I_K \otimes z'_{ct})' \Omega^{-1} f_{t+1}] \right\}.$$

### 2.4.4 Conditional Posterior Distributions for $\Lambda$ and $\Theta$

We specify the prior distribution of $[\text{vec}(\Lambda') \mid \text{vec}(\Theta')]$ as $N(w_0, W_0)$, and set $w_0$ to a vector of zero, and $W_0$ to a diagonal matrix. In the case of non-informative prior on model coefficients and intercepts, we set the diagonal elements of $W_0$ to infinity (approximated by a very large number); in the case of informative prior, we set the values to a finite number (such as the square of 1%, 2%, 5% and 10%).

The density function of the conditional posterior distribution of $[\text{vec}(\Lambda') \mid \text{vec}(\Theta')]$ is:

$$p \left( \begin{pmatrix} \text{vec}(\Lambda') \\ \text{vec}(\Theta') \end{pmatrix} \mid \{r_{t+1}\}, \{f_{t+1}\}, \Sigma \right) \propto \prod_{t=0}^{T-1} p(r_{t+1} \mid f_{t+1}, \Sigma, \Lambda, \Theta) p \left( \begin{pmatrix} \text{vec}(\Lambda') \\ \text{vec}(\Theta') \end{pmatrix} \right).$$

From Equation (2.17) and Theorem 1, the conditional posterior distribution of $[\text{vec}(\Lambda') \mid \text{vec}(\Theta')]$ is:

$$[\begin{pmatrix} \text{vec}(\Lambda') \\ \text{vec}(\Theta') \end{pmatrix} \mid \{r_{t+1}\}, \{f_{t+1}\}, \Sigma] \sim N(w_1, W_1),$$

with

$$W_1 = \left\{ W_0^{-1} + \sum_{t=0}^{T-1} \begin{pmatrix} X_{at} \otimes z'_{at} & X_{bt} \otimes z'_{bt} \otimes f'_{t+1} \end{pmatrix} \Sigma_{t+1}^{-1} \begin{pmatrix} X_{at} \otimes z'_{at} & X_{bt} \otimes z'_{bt} \otimes f'_{t+1} \end{pmatrix} \right\}^{-1},$$

$$w_1 = W_1 \left\{ W_0^{-1} w_0 + \sum_{t=0}^{T-1} \begin{pmatrix} X_{at} \otimes z'_{at} & X_{bt} \otimes z'_{bt} \otimes f'_{t+1} \end{pmatrix} \Sigma_{t+1}^{-1} r_{t+1} \right\}.$$
2.4.5 Conditional Distributions for \( f_{T+1} \) and \( r_{T+1} \)

At the end of each iteration of the Gibbs sampler, we draw \( f_{T+1} \) according to its conditional distribution

\[
[f_{T+1} \mid \Delta, \Omega] \sim N(\Delta z_T, \Omega).
\]

When no prior distribution is specified for \( r_{T+1} \), we draw \( r_{T+1} \) from the following conditional distribution (see Equation (2.16)):

\[
[r_{T+1} \mid f_{T+1}, \Sigma_{T+1}, \Lambda, \Theta] \sim N(X_a T \Lambda z_{aT} + (X_b T \otimes z'_{bT}) \Theta f_{T+1}, \Sigma_{T+1}) \tag{2.19}
\]

We now consider the case in which an informative prior distribution of \( r_{T+1} \) is specified. We assume the prior distribution is

\[
r_{T+1} \sim N(\mu_0, \Sigma_0) \tag{2.20}
\]

where \( \mu_0 \) and \( \Sigma_0 \) are derived in Section 2.2. To incorporate this informative prior distribution, we rewrite Equation (2.20) as:

\[
[\mu_0 \mid r_{T+1}, \Sigma_0] \sim N(r_{T+1}, \Sigma_0) \tag{2.21}
\]

That is, we view the expected return \( \mu_0 \) as part of the data derived from the stock weights in the benchmark portfolio, and Equation (2.21) is its conditional distribution. The density function of the conditional distribution of \( r_{T+1} \) is

\[
p(r_{T+1} \mid f_{T+1}, \Sigma, \Lambda, \Theta, \mu_0, \Sigma_0)
\]

\[
\propto p(r_{T+1} \mid f_{T+1}, \Sigma, \Lambda, \Theta) p(\mu_0 \mid r_{T+1}, \Sigma_0).
\]

From Equation (2.19), Equation (2.21) and Theorem 1, the conditional distribution of \( r_{T+1} \) is:

\[
[r_{T+1} \mid f_{T+1}, \Sigma, \Lambda, \Theta, \mu_0, \Sigma_0] \sim N(\tilde{\mu}, \tilde{\Sigma})
\]

where

\[
\tilde{\Sigma} = (\Sigma_{T+1}^{-1} + \Sigma_0^{-1})^{-1}
\]

\[
\tilde{\mu} = \bar{\Sigma} \{ \Sigma_{T+1}^{-1} [X_a T \Lambda z_{aT} + (X_b T \otimes z'_{bT}) \Theta f_{T+1}] + \Sigma_0^{-1} \mu_0 \}.
\]
2.5 Optimization Framework

For each proposed model, we construct optimal portfolios at the end of each Period $T$ using the forecasted mean and covariance matrix of stock returns in Period $(T + 1)$. We assume that investors are mean-variance efficient (see Markowitz (1952)), and construct optimal portfolios by maximizing the following utility function:

$$U(w) = \mu_p - \frac{\gamma}{2} \sigma_p^2$$  \hspace{1cm} (2.22)

$$= w' \mu_{T+1} - \frac{\gamma}{2} w' \Sigma_{T+1} w$$  \hspace{1cm} (2.23)

where $\mu_p$ and $\sigma_p^2$ are the predicted mean and variance of the excess return of the optimal portfolio, $\gamma$ is the investor’s risk aversion parameter, $w$ is the vector of stock weights in the optimal portfolio, and $\mu_{T+1}$ and $\Sigma_{T+1}$ are the mean and covariance matrix of the excess returns $r_{T+1}$ estimated from its predictive distribution. The objective function balances the trade-off between the expected return and the variance of the optimal portfolio with respect to the investor’s risk aversion level. Note that the objective function is the same as Equation (2.14) used for deriving the informative prior distribution of stock returns.

We focus on constructing long-only portfolios, that is, no short-selling of stocks is allowed. In addition, we assume there is no margin borrowing. However, investing in the risk-free rate is allowed. This is achieved by imposing the following linear constraints on the stock weights:

$$w \geq 0$$  \hspace{1cm} (2.24)

$$w'1 \leq 1$$  \hspace{1cm} (2.25)

This portfolio optimization is a quadratic programming problem, and can be solved by many software specializing in operations research.

Before running the optimization, we need to set the value of the risk aversion parameter $\gamma$. Prior empirical research suggests that the likely range of $\gamma$ is between 2 and 10. In our empirical work, we run optimization for each of the following four values of $\gamma$: 2, 5, 7, and 10.

Because the turnover rate of stock portfolios based on quantitative models tends to be high, portfolio performance is unrealistic if transaction costs are not considered. There are mainly three sources of transaction cost: the bid-ask spread, the brokerage fee, and the
market impact. Although some investment firms utilize sophisticated models to estimate transaction cost, in this paper we simplify it by assuming transaction cost is a fixed percentage of the transaction amount and account for the transaction cost when computing the performance of trading strategies.
CHAPTER 3

Simulation

Before applying the models to real data of the U.S. stock market, we conduct a simulation study in this chapter to test the MCMC estimation method derived in Chapter 2. We first set the true values of the model parameters \((\Lambda, \Theta, \Delta, \Sigma, \Omega)\) to pre-determined values. The values of the predictive variables \((X_{at}, X_{bt}, z_{at}, z_{bt}, z_{ct})\), \(t = 0, \ldots, T\), are also set to pre-determined values. To generate the true values of \((f_{t+1}, r_{t+1})\), we make a random draw of the residuals \((e_{t+1}, e_{t+1})\) from the normal distributions in Equation (2.1.2), and compute \(f_{t+1}\) according to Equation (2.11) and \(r_{t+1}\) according to Equation (2.8). Next, we run the MCMC algorithm on the above data to generate draws of the model parameters \((\Lambda, \Theta, \Delta, \Sigma, \Omega)\) as well as the predicted values of \((f_{T+1}, r_{T+1})\). Summary statistics of the draws are computed and compared to the true values.

In our simulation, we set \(J_a = J_b = 3\), \(M_a = M_b = M_c = 2\), \(K = 1\), \(N = 500\) and \(T = 100\). Therefore, the dimensions of \(\Delta, \Lambda, \Theta, \Omega\), and \(\Sigma\) are \(1 \times 3\), \(4 \times 3\), \(12 \times 1\), \(1 \times 1\), and \(500 \times 500\), respectively. The true value of each entry of \((\Delta, \Lambda, \Theta)\) is independently selected from a standard normal distribution. The true value of each entry of \((X_{at}, X_{bt}, z_{at}, z_{bt}, z_{ct})\) is also independently selected from a standard normal distribution.

Since \(\Omega\) is a positive scalar when \(K = 1\) and \(\Sigma\) is a diagonal covariance matrix, we randomly select \(\Omega\) and the diagonal elements of \(\Sigma\) from a Uniform(1, \(max\)) distribution. We consider three scenarios, in which \(max = 4, 25\) and \(100\), respectively. We compare the simulation outcome of the three scenarios to assess how the volatility of the error terms \((e_{t+1}, e_{t+1})\) affects the estimation accuracy of the other parameters. The true values of the parameters \((\Delta, \Lambda, \Theta)\) and the variables \((X_{at}, X_{bt}, z_{at}, z_{bt}, z_{ct})\) are kept the same in the three scenarios.

In each scenario, we run the Gibbs sampler with non-informative prior distributions.
for both model parameters and stock returns. The Gibbs sampler is run for a total of 500 iterations, and the first 100 draws are discarded as the “burn-in” period. Using the remaining 400 draws, we compute the sample mean, standard deviation, and the 1st and 99th percentiles for each parameter. If the true value of a parameter falls between the 1st and 99th percentiles of the sample (i.e., inside the 98% empirical confidence interval), we consider the estimation of the parameter to have converged to its true value.

Tables 3.1, 3.2, and 3.3 report the simulation results. In each of the three scenarios, the true values of $\Delta$, $\Lambda$, $\Theta$, $\Omega$ and $f_{T+1}$ all fall inside the 98% confidence interval. We also notice that, as the volatility of $\epsilon_{t+1}$ and $e_{t+1}$ increases, the estimates of the model parameters become less accurate, with larger standard deviations and wider confidence intervals. For example, the standard deviation of $\Delta_{1,1}$ increases from 0.113 in Scenario 1 to 0.171 in Scenario 2, and 0.270 in Scenario 3; its 98% confidence interval also widens from $(-1.639, -1.125)$ in Scenario 1 to $(-1.825, -1.042)$ in Scenario 2, and $(-2.131, -0.905)$ in Scenario 3.

In each scenario, the true values of $\Sigma$ and $r_{T+1}$ also fall inside the 98% confidence interval, except for a small number of cases.\footnote{Specifically, out of 500 stocks, the true values of $\Sigma$ for 13 stocks, and the true values of $r_{T+1}$ for 11 stocks, do not fall inside the 98% confidence interval in Scenario 1. The corresponding numbers are 20 & 11 in Scenario 2, and 22 & 15 in Scenario 3.} We re-ran the simulations for as many as 3000 iterations, and the true values of $\Sigma$ and $r_{T+1}$ for all stocks fall inside the 99.8% confidence interval. For the sake of space, we do not report the detailed results of $\Sigma$ and $r_{T+1}$ in the tables.

The simulation study shows that the MCMC method is capable of uncovering the true values of the model parameters when the specified model correctly describes the data-generating process.
Table 3.1: Simulation Results (Scenario 1)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Mean</th>
<th>Std Dev</th>
<th>1st Pct</th>
<th>99th Pct</th>
<th>Converge(Y/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta_{1,1}$</td>
<td>-1.284</td>
<td>-1.408</td>
<td>0.113</td>
<td>-1.639</td>
<td>-1.125</td>
<td>Y</td>
</tr>
<tr>
<td>$\Delta_{1,2}$</td>
<td>-0.555</td>
<td>-0.757</td>
<td>0.122</td>
<td>-1.031</td>
<td>-0.432</td>
<td>Y</td>
</tr>
<tr>
<td>$\Delta_{1,3}$</td>
<td>-0.088</td>
<td>-0.165</td>
<td>0.114</td>
<td>-0.428</td>
<td>0.104</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{1,1}$</td>
<td>-0.168</td>
<td>-0.170</td>
<td>0.009</td>
<td>-0.191</td>
<td>-0.150</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{2,1}$</td>
<td>0.125</td>
<td>0.132</td>
<td>0.007</td>
<td>0.115</td>
<td>0.149</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{3,1}$</td>
<td>-2.120</td>
<td>-2.122</td>
<td>0.008</td>
<td>-2.140</td>
<td>-2.104</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{4,1}$</td>
<td>-0.709</td>
<td>-0.708</td>
<td>0.007</td>
<td>-0.724</td>
<td>-0.692</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{1,2}$</td>
<td>0.895</td>
<td>0.889</td>
<td>0.006</td>
<td>0.876</td>
<td>0.903</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{2,2}$</td>
<td>0.016</td>
<td>0.019</td>
<td>0.006</td>
<td>0.003</td>
<td>0.032</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{3,2}$</td>
<td>-0.815</td>
<td>-0.813</td>
<td>0.006</td>
<td>-0.828</td>
<td>-0.797</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{4,2}$</td>
<td>-0.609</td>
<td>-0.602</td>
<td>0.006</td>
<td>-0.616</td>
<td>-0.590</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{1,3}$</td>
<td>-0.436</td>
<td>-0.444</td>
<td>0.006</td>
<td>-0.459</td>
<td>-0.428</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{2,3}$</td>
<td>-3.060</td>
<td>-3.068</td>
<td>0.006</td>
<td>-3.081</td>
<td>-3.052</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{3,3}$</td>
<td>0.158</td>
<td>0.172</td>
<td>0.006</td>
<td>0.156</td>
<td>0.185</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{4,3}$</td>
<td>0.298</td>
<td>0.302</td>
<td>0.007</td>
<td>0.283</td>
<td>0.318</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{1,1}$</td>
<td>-0.867</td>
<td>-0.866</td>
<td>0.005</td>
<td>-0.879</td>
<td>-0.853</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{2,1}$</td>
<td>-1.500</td>
<td>-1.501</td>
<td>0.004</td>
<td>-1.510</td>
<td>-1.493</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{3,1}$</td>
<td>-1.505</td>
<td>-1.511</td>
<td>0.004</td>
<td>-1.520</td>
<td>-1.500</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{4,1}$</td>
<td>-1.402</td>
<td>-1.399</td>
<td>0.003</td>
<td>-1.407</td>
<td>-1.391</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{5,1}$</td>
<td>0.185</td>
<td>0.184</td>
<td>0.005</td>
<td>0.173</td>
<td>0.197</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{6,1}$</td>
<td>-0.630</td>
<td>-0.631</td>
<td>0.003</td>
<td>-0.639</td>
<td>-0.623</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{7,1}$</td>
<td>0.631</td>
<td>0.628</td>
<td>0.004</td>
<td>0.621</td>
<td>0.638</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{8,1}$</td>
<td>-1.869</td>
<td>-1.867</td>
<td>0.005</td>
<td>-1.877</td>
<td>-1.856</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{9,1}$</td>
<td>-0.104</td>
<td>-0.106</td>
<td>0.003</td>
<td>-0.115</td>
<td>-0.099</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{10,1}$</td>
<td>0.570</td>
<td>0.567</td>
<td>0.003</td>
<td>0.560</td>
<td>0.576</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{11,1}$</td>
<td>-0.198</td>
<td>-0.204</td>
<td>0.004</td>
<td>-0.215</td>
<td>-0.194</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{12,1}$</td>
<td>-0.180</td>
<td>-0.179</td>
<td>0.003</td>
<td>-0.186</td>
<td>-0.171</td>
<td>Y</td>
</tr>
<tr>
<td>$\Omega_{1,1}$</td>
<td>1.466</td>
<td>1.309</td>
<td>0.177</td>
<td>0.969</td>
<td>1.775</td>
<td>Y</td>
</tr>
<tr>
<td>$f_{T+1}$</td>
<td>-2.723</td>
<td>-1.242</td>
<td>1.093</td>
<td>-3.694</td>
<td>1.176</td>
<td>Y</td>
</tr>
</tbody>
</table>
Table 3.2: Simulation Results (Scenario 2)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Mean</th>
<th>Std Dev</th>
<th>1st Pct</th>
<th>99th Pct</th>
<th>Converge(Y/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta_{1,1}$</td>
<td>-1.284</td>
<td>-1.473</td>
<td>0.171</td>
<td>-1.825</td>
<td>-1.042</td>
<td>Y</td>
</tr>
<tr>
<td>$\Delta_{1,2}$</td>
<td>-0.555</td>
<td>-0.862</td>
<td>0.185</td>
<td>-1.285</td>
<td>-0.359</td>
<td>Y</td>
</tr>
<tr>
<td>$\Delta_{1,3}$</td>
<td>-0.088</td>
<td>-0.205</td>
<td>0.173</td>
<td>-0.605</td>
<td>0.215</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{1,1}$</td>
<td>-0.168</td>
<td>-0.174</td>
<td>0.012</td>
<td>-0.202</td>
<td>-0.146</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{2,1}$</td>
<td>0.125</td>
<td>0.130</td>
<td>0.011</td>
<td>0.103</td>
<td>0.157</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{3,1}$</td>
<td>-2.120</td>
<td>-2.124</td>
<td>0.012</td>
<td>-2.153</td>
<td>-2.096</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{4,1}$</td>
<td>-0.709</td>
<td>-0.702</td>
<td>0.011</td>
<td>-0.728</td>
<td>-0.677</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{1,2}$</td>
<td>0.895</td>
<td>0.890</td>
<td>0.009</td>
<td>0.869</td>
<td>0.911</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{2,2}$</td>
<td>0.016</td>
<td>0.021</td>
<td>0.010</td>
<td>-0.004</td>
<td>0.041</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{3,2}$</td>
<td>-0.815</td>
<td>-0.809</td>
<td>0.010</td>
<td>-0.833</td>
<td>-0.784</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{4,2}$</td>
<td>-0.609</td>
<td>-0.598</td>
<td>0.009</td>
<td>-0.619</td>
<td>-0.579</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{1,3}$</td>
<td>-0.436</td>
<td>-0.448</td>
<td>0.009</td>
<td>-0.472</td>
<td>-0.424</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{2,3}$</td>
<td>-3.060</td>
<td>-3.069</td>
<td>0.010</td>
<td>-3.090</td>
<td>-3.045</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{3,3}$</td>
<td>0.158</td>
<td>0.179</td>
<td>0.009</td>
<td>0.156</td>
<td>0.199</td>
<td>Y</td>
</tr>
<tr>
<td>$\Lambda_{4,3}$</td>
<td>0.298</td>
<td>0.303</td>
<td>0.010</td>
<td>0.274</td>
<td>0.328</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{1,1}$</td>
<td>-0.867</td>
<td>-0.866</td>
<td>0.006</td>
<td>-0.880</td>
<td>-0.852</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{2,1}$</td>
<td>-1.500</td>
<td>-1.500</td>
<td>0.005</td>
<td>-1.512</td>
<td>-1.490</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{3,1}$</td>
<td>-1.505</td>
<td>-1.513</td>
<td>0.005</td>
<td>-1.524</td>
<td>-1.499</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{4,1}$</td>
<td>-1.402</td>
<td>-1.400</td>
<td>0.004</td>
<td>-1.410</td>
<td>-1.389</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{5,1}$</td>
<td>0.185</td>
<td>0.187</td>
<td>0.006</td>
<td>0.174</td>
<td>0.204</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{6,1}$</td>
<td>-0.630</td>
<td>-0.630</td>
<td>0.004</td>
<td>-0.640</td>
<td>-0.621</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{7,1}$</td>
<td>0.631</td>
<td>0.628</td>
<td>0.005</td>
<td>0.618</td>
<td>0.640</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{8,1}$</td>
<td>-1.869</td>
<td>-1.864</td>
<td>0.006</td>
<td>-1.877</td>
<td>-1.850</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{9,1}$</td>
<td>-0.104</td>
<td>-0.104</td>
<td>0.004</td>
<td>-0.115</td>
<td>-0.095</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{10,1}$</td>
<td>0.570</td>
<td>0.569</td>
<td>0.004</td>
<td>0.559</td>
<td>0.579</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{11,1}$</td>
<td>-0.198</td>
<td>-0.203</td>
<td>0.005</td>
<td>-0.216</td>
<td>-0.191</td>
<td>Y</td>
</tr>
<tr>
<td>$\Theta_{12,1}$</td>
<td>-0.180</td>
<td>-0.180</td>
<td>0.004</td>
<td>-0.190</td>
<td>-0.170</td>
<td>Y</td>
</tr>
<tr>
<td>$\Omega_{1,1}$</td>
<td>3.395</td>
<td>3.020</td>
<td>0.408</td>
<td>2.201</td>
<td>4.120</td>
<td>Y</td>
</tr>
<tr>
<td>$f_{T+1}$</td>
<td>-3.576</td>
<td>-1.323</td>
<td>1.661</td>
<td>-5.104</td>
<td>2.387</td>
<td>Y</td>
</tr>
</tbody>
</table>
Table 3.3: Simulation Results (Scenario 3)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Mean</th>
<th>Std Dev</th>
<th>1st Pct</th>
<th>99th Pct</th>
<th>Converge(Y/N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Δ₁,₁</td>
<td>-1.284</td>
<td>-1.580</td>
<td>0.270</td>
<td>-2.131</td>
<td>-0.905</td>
<td>Y</td>
</tr>
<tr>
<td>Δ₁,₂</td>
<td>-0.555</td>
<td>-1.038</td>
<td>0.290</td>
<td>-1.693</td>
<td>-0.264</td>
<td>Y</td>
</tr>
<tr>
<td>Δ₁,₃</td>
<td>-0.088</td>
<td>-0.272</td>
<td>0.271</td>
<td>-0.899</td>
<td>0.369</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₁,₁</td>
<td>-0.168</td>
<td>-0.175</td>
<td>0.015</td>
<td>-0.211</td>
<td>-0.141</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₂,₁</td>
<td>0.125</td>
<td>0.124</td>
<td>0.015</td>
<td>0.087</td>
<td>0.162</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₁,₂</td>
<td>-2.120</td>
<td>-2.126</td>
<td>0.017</td>
<td>-2.165</td>
<td>-2.087</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₂,₂</td>
<td>-0.709</td>
<td>-0.693</td>
<td>0.016</td>
<td>-0.729</td>
<td>-0.660</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₁,₂</td>
<td>0.895</td>
<td>0.891</td>
<td>0.013</td>
<td>0.863</td>
<td>0.920</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₂,₂</td>
<td>0.016</td>
<td>0.023</td>
<td>0.014</td>
<td>-0.012</td>
<td>0.052</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₁,₂</td>
<td>-0.815</td>
<td>-0.805</td>
<td>0.014</td>
<td>-0.838</td>
<td>-0.771</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₂,₂</td>
<td>-0.609</td>
<td>-0.592</td>
<td>0.013</td>
<td>-0.622</td>
<td>-0.566</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₁,₂</td>
<td>-0.436</td>
<td>-0.451</td>
<td>0.013</td>
<td>-0.483</td>
<td>-0.418</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₂,₂</td>
<td>-3.060</td>
<td>-3.067</td>
<td>0.014</td>
<td>-3.096</td>
<td>-3.033</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₁,₂</td>
<td>0.158</td>
<td>0.185</td>
<td>0.013</td>
<td>0.157</td>
<td>0.213</td>
<td>Y</td>
</tr>
<tr>
<td>Λ₂,₂</td>
<td>0.298</td>
<td>0.302</td>
<td>0.014</td>
<td>0.263</td>
<td>0.337</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₁,₁</td>
<td>-0.867</td>
<td>-0.867</td>
<td>0.005</td>
<td>-0.880</td>
<td>-0.854</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₂,₁</td>
<td>-1.500</td>
<td>-1.498</td>
<td>0.005</td>
<td>-1.510</td>
<td>-1.488</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₃,₁</td>
<td>-1.505</td>
<td>-1.513</td>
<td>0.005</td>
<td>-1.524</td>
<td>-1.500</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₄,₁</td>
<td>-1.402</td>
<td>-1.402</td>
<td>0.004</td>
<td>-1.411</td>
<td>-1.391</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₅,₁</td>
<td>0.185</td>
<td>0.192</td>
<td>0.006</td>
<td>0.179</td>
<td>0.207</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₆,₁</td>
<td>-0.630</td>
<td>-0.629</td>
<td>0.004</td>
<td>-0.639</td>
<td>-0.620</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₇,₁</td>
<td>0.631</td>
<td>0.628</td>
<td>0.005</td>
<td>0.618</td>
<td>0.640</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₈,₁</td>
<td>-1.869</td>
<td>-1.864</td>
<td>0.006</td>
<td>-1.876</td>
<td>-1.850</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₉,₁</td>
<td>-0.104</td>
<td>-0.102</td>
<td>0.004</td>
<td>-0.113</td>
<td>-0.093</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₁₀,₁</td>
<td>0.570</td>
<td>0.570</td>
<td>0.004</td>
<td>0.560</td>
<td>0.580</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₁₁,₁</td>
<td>-0.198</td>
<td>-0.200</td>
<td>0.005</td>
<td>-0.213</td>
<td>-0.188</td>
<td>Y</td>
</tr>
<tr>
<td>Θ₁₂,₁</td>
<td>-0.180</td>
<td>-0.180</td>
<td>0.004</td>
<td>-0.190</td>
<td>-0.170</td>
<td>Y</td>
</tr>
<tr>
<td>Ω₁,₁</td>
<td>8.385</td>
<td>7.441</td>
<td>1.007</td>
<td>5.511</td>
<td>10.090</td>
<td>Y</td>
</tr>
<tr>
<td>fᵣ₊₁</td>
<td>-4.999</td>
<td>-1.455</td>
<td>2.605</td>
<td>-7.302</td>
<td>4.308</td>
<td>Y</td>
</tr>
</tbody>
</table>
CHAPTER 4
DATA

We apply the modeling and optimization framework developed in this paper to the U.S. stock market. We focus on large-cap stocks and consider stocks in the S&P 500 Index as the investment universe. We consider the CAPM as the risk model and use the S&P500 index as proxy for the market portfolio (i.e., the risk factor). Monthly return data on individual stocks and the S&P500 Index are obtained from the Center for Research of Security Prices (CRSP) database.

The characteristic-based factors are sourced from both CRSP and COMPSTAT, and are grouped into two categories: the alpha factors ($X_{\text{at}}$) and the beta factors ($X_{\text{bt}}$). We collect 9 alpha factors, which are further grouped into the following 4 sub-categories:

- **firm size**: the market capitalization of its common stock. We transform the variable by taking the natural logarithm in order to reduce the skewness.

- **value**: we select 4 value related characteristics: book to price (book value of equity divided by market value of stock), earnings yield (trailing 12-month earnings per share divided by stock price), net payout yield (trailing 12-month dividends yield plus net stock repurchase rate) and sales to price (trailing 12-month total sales divided by market value of stock).

- **quality**: we select 2 quality related characteristics: ROE (12-month trailing return on equity), and income growth (12-month trailing net income growth rate).

- **market sentiment**: we define two variables related to market sentiment: price momentum (stock return over the past 12 months) and trading volume (number of shares traded over the past 12 months divided by total number of shares outstanding).
Past research has documented that the value factors, quality factors and price momentum are positively correlated with future stock returns, while size and trading volume are negatively correlated with future stock returns. Instead of fitting the models using the raw values of the characteristics, we compute a z-score for each characteristic by normalizing it by its cross-sectional sample mean and standard deviation in each month. The z-scores are then capped at ±3 to mitigate undue influence of outliers.

We predict the beta of stocks to the risk premium (i.e., excess return of the S&P 500 Index) using the Global Industry Classification Standard (GICS). The GICS is an industry classification system and classifies each company into one of 10 sectors, 23 industry groups, 59 industries, and 122 subindustries. The 10 sectors are: energy, materials, industrials, consumer discretionary, consumer staples, healthcare, financials, information technology, telecommunications, and utilities. We model beta at the sector level by defining a dummy variable for each sector. Stocks in the same sector are assumed to have the same beta to the risk factor. Thus, $X_{bt}$ is a matrix of dummy variables.

We predict the risk premium (i.e., S&P500 return in excess of the risk-free rate) using 5 macroeconomic and aggregate financial variables ($z_{ct}$):

- **short-term interest rate**: defined as the 3-month treasury bill yield.
- **term spread of interest rate**: defined as the yield spread between 30-year and 3-month treasurys.
- **credit spread**: defined as the yield spread between AAA and BAA rated corporate bonds.
- **consumption shock**: defined as the portion of aggregate consumption growth unexplained by the cointegration relationship among consumption, labor income and asset wealth (see Lettau and Ludvigson (2001)).
- **net payout ratio**: defined as the natural logarithm of the sum of aggregate dividend yield and net repurchase rate of stocks in the S&P500 Index (see Boudoukh, Michaely, Richardson, and Roberts (2007)).

The data for short-term interest rate, term spread and credit spread are downloaded from the website of Federal Reserve Bank of St. Louis. The data for consumption shock
and net payout ratio are obtained from the websites of Prof. Ludvigson and Prof. Richardson, respectively. Past research has shown that the term spread, credit spread, consumption shock and net payout ratio are positively correlated with future risk premium, while short-term rate is negatively correlated with future risk premium. To simplify the specification of informative priors on model coefficients, we standardize each variable by substracting its sample mean and dividing by its sample standard deviation. We then cap each variable at ±3 to mitigate undue influence of outliers.

We model the time-varying component of alpha and beta ($z_{at}$ and $z_{bt}$) using one aggregate financial variable: return of S&P 500 Index in the past 12 months. That is, we assume the alpha factor returns and stocks’ beta to the risk factor differ by market condition (bull market vs. bear market). The variable is standardized by taking the logistic transformation, so that the transformed variable has a continuous distribution between 0 and 1, with values close to 0 indicating a bear market, and values close to 1 indicating a bull market.

Table 4.1 reports the summary statistics of the alpha factors and the macroeconomic and aggregate financial variables. The historical COMPUSTAT data start in 1964, and the macroeconomic data start in 1952.

We build our models to predict stock’s excess return and the equity risk premium over a 6-month holding period, and refit the models every 6 months. To ensure that stock returns are not serially correlated, we fit the models using data with non-overlapping 6-month performance windows, starting in January and July of each year. Further, a stock needs to have a least 2 years of historical data to be included in the sample.

We start our model fitting using all available data up to December 1972, because COMPUSTAT data for stocks in the S&P500 index are sparse prior to that. We then construct optimal portfolios based on the model predictions. Because our model predicts stock returns over a 6-month holding period, we assume that the optimal portfolios are held for 6 months without rebalancing. After every 6 months, we refit the models using all available data up to that date and reconstruct the optimal portfolios. The last model refitting is based on data up to December 2008, because monthly stock return data in the CRSP database end in June 2009 as of our last data collecton. For each investment strategy we consider, this process produces 72 optimal portfolios, each held for 6 months, with performance windows spanning January 1973 and June 2009.
Table 4.1: Summary Statistics of Predictors

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Mean</th>
<th>Median</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>alpha factors</td>
<td>firm size ($ million)</td>
<td>8685.25</td>
<td>2290.23</td>
<td>24179.18</td>
</tr>
<tr>
<td></td>
<td>book to price (%)</td>
<td>90.03</td>
<td>53.96</td>
<td>3396.61</td>
</tr>
<tr>
<td></td>
<td>earnings yield (%)</td>
<td>5.54</td>
<td>6.41</td>
<td>36.58</td>
</tr>
<tr>
<td></td>
<td>dividend yield (%)</td>
<td>3.43</td>
<td>2.56</td>
<td>35.62</td>
</tr>
<tr>
<td></td>
<td>sales to price (%)</td>
<td>227.97</td>
<td>122.94</td>
<td>3143.71</td>
</tr>
<tr>
<td></td>
<td>ROE (%)</td>
<td>13.39</td>
<td>13.57</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>income growth (%)</td>
<td>25.58</td>
<td>10.08</td>
<td>1887.15</td>
</tr>
<tr>
<td></td>
<td>price momentum (%)</td>
<td>14.30</td>
<td>10.57</td>
<td>38.93</td>
</tr>
<tr>
<td></td>
<td>trading volume (%)</td>
<td>81.90</td>
<td>51.95</td>
<td>108.90</td>
</tr>
<tr>
<td>macroeconomic variables</td>
<td>3-month treasury (%)</td>
<td>5.87</td>
<td>5.40</td>
<td>2.79</td>
</tr>
<tr>
<td>predicting risk primum</td>
<td>term spread (%)</td>
<td>1.51</td>
<td>1.56</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>credit spread (%)</td>
<td>1.05</td>
<td>0.92</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>consumption shock</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>net payout ratio (log)</td>
<td>-2.24</td>
<td>-2.21</td>
<td>0.24</td>
</tr>
<tr>
<td>macroeconomic variable</td>
<td>S&amp;P500 past 12-month return (%)</td>
<td>8.46</td>
<td>9.73</td>
<td>15.75</td>
</tr>
<tr>
<td>predicting time-varying alpha</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and beta</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5

Empirical Results

We construct optimal portfolios for various investment strategies according to the objective function and constraints outlined in Section 2.5. We maximize the quadratic utility function in Equation (2.22), which is also commonly known as the Certainty Equivalent Return (CER). The two constraints in Equations (2.24) and (2.25) are imposed on all portfolios. That is, we do not allow short-selling of stocks, nor do we allow leveraging through margin borrowing. However, we assume that the investor may invest in the risk-free rate as well as the individual stocks in the S&P500 index.

5.1 Performance of the Full Model

We first report the performance of optimal portfolios based on the full model (Model A). Four versions of the objective function are considered, with the risk aversion factor $\gamma$ set to 2, 5, 7 and 10, respectively. To assess the predictiveness of our model and the effect of different prior distributions, we compare the performance of the following four investment strategies for each risk aversion level:

- **INDEX**: buy and hold the S&P 500 Index. This is the passive investment strategy that serves as the benchmark portfolio.

- **OPT-NP**: optimal strategy with non-informative prior distributions. In this strategy, we estimate the model using non-informative priors on both stock returns and model coefficients and intercepts.

- **OPT-IP-S**: optimal strategy with informative prior on stock returns. In this strategy, we estimate the model using the informative prior on stock returns detailed in Section 2.2. However, the non-informative prior on model coefficients and intercepts is used.
• **OPT-IP-M**: optimal strategy with informative prior on model coefficients and intercepts. In this strategy, we estimate the model using the informative prior on model coefficients and intercepts detailed in Section 2.3. However, the non-informative prior on stock returns is used. We tested 4 versions of the informative prior distribution on model coefficients and intercepts, with the standard deviation of the distribution set to 1%, 2%, 5%, and 10%, respectively. We find that all of them outperform OPT-NP and INDEX. Particularly, the version with 1% standard deviation performs the best. For the sake of space, we only report the results for this version.

• **OPT-IP-SM**: optimal strategy with informative priors on both stock returns and model coefficients and intercepts. The standard deviation of the informative prior distribution for model coefficients and intercepts is set to 1%, same as in OPT-IP-M.

Table 5.1 reports the portfolio performance of the full model over the entire performance period (from January 1973 through June 2009). Although the portfolios are rebalanced every 6 months, the performance summary statistics are computed using monthly returns. The 3rd and 4th columns of the table report the annualized mean and standard deviation\(^1\) of monthly excess returns. Our primary performance measure for comparing the strategies is the Certainty Equivalent Return (CER) computed according to Equation (2.22)\(^2\), because it is the objective function we are trying to maximize. In addition, we report another commonly used performance measure - the Sharpe Ratio (SR), which is the ratio of the mean to the standard deviation of annualized excess returns. None of the performance measures accounts for transaction costs. We will evaluate the impact of transaction costs in a later section.

We first note that, over the performance period, the stock market has not been very generous to investors who stick to the buy-and-hold strategy. The performance of the INDEX strategy has been very modest, with an annualized mean excess return of only 1.1%, a standard deviation of 15.8%, and a barely positive Sharpe Ratio (0.07). For the INDEX strategy, the CER is the only performance measure that varies by risk aversion level, ranging from -1.4% (for \(\gamma = 2\)) to -11.4% (for \(\gamma = 10\)). Relative to the INDEX strategy, all of the optimization strategies generated higher mean excess returns. However, they are also more risky with higher standard deviations. When the risk aversion level increases, both the mean excess returns and standard deviations of excess returns increase.

\(1\) annualized mean = 12 * monthly mean; annualized std dev = \(\sqrt{12}\) * monthly std dev.

\(2\) Note that when computing CER according to Equation (2.22), \(\mu_p\) and \(\sigma_p\) should be measured in fraction, not in percentage. We scale up CER by 100 when reporting in Table 5.1.
and the standard deviation of the optimization strategies generally decrease, as a higher risk aversion level puts more penalty on risk and favors portfolios that are less risky. At each risk aversion level, all of the optimization strategies delivered much higher Sharpe Ratio than the INDEX strategy.

The OPT-NP strategy, however, does not always outperform the INDEX strategy in terms of CER, the primary performance measure. In fact, the OPT-NP strategy delivers higher CER than the INDEX strategy in only 1 out of 4 risk aversion scenarios (i.e., when $\gamma = 2$). This suggests that the model-based optimal strategy using non-informative priors does not always produce portfolios with better out-of-sample performance than the buy-and-hold strategy. In contrast, all 3 optimization strategies based on informative priors (OPT-IP-S, OPT-IP-M, and OPT-IP-SM) outperform both the INDEX strategy and the OPT-NP strategy, at each risk aversion level and in terms of both CER and Sharpe Ratio. The results indeed support our hypothesis that informative prior distributions reduce estimation error and result in optimal portfolios with better out-of-sample performance.

We further compare the 3 strategies using informative priors. Relative to the OPT-IP-S strategy, the OPT-IP-M strategy has higher or equal CER in all 4 risk aversion scenarios. Moreover, OPT-IP-SM has the highest CER in 3 out of 4 risk aversion scenarios (when $\gamma = 2, 5, 10$). The Sharpe Ratios of the 3 strategies are similar and do not form a consistent ranking. Overall, the results suggest that informative priors on model coefficients and intercepts are somewhat more important than informative priors on stock returns, and that incorporating both types of informative priors is often better than incorporating only one type of informative priors.
Table 5.1: Comparison of Strategy Performance (Model A)

<table>
<thead>
<tr>
<th>Risk Aversion(γ)</th>
<th>Strategy</th>
<th>Mean(%)</th>
<th>Std Dev(%)</th>
<th>CER</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>INDEX</td>
<td>1.1</td>
<td>15.8</td>
<td>-1.4</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>OPT-NP</td>
<td>8.9</td>
<td>29.9</td>
<td>0.0</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-S</td>
<td>11.9</td>
<td>29.9</td>
<td>2.9</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-M</td>
<td>10.4</td>
<td>27.4</td>
<td>2.9</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-SM</td>
<td>13.7</td>
<td>28.2</td>
<td>5.8</td>
<td>0.49</td>
</tr>
<tr>
<td>5</td>
<td>INDEX</td>
<td>1.1</td>
<td>15.8</td>
<td>-5.1</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>OPT-NP</td>
<td>9.9</td>
<td>25.2</td>
<td>-6.0</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-S</td>
<td>11.9</td>
<td>23.6</td>
<td>-2.0</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-M</td>
<td>10.4</td>
<td>21.7</td>
<td>-1.4</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-SM</td>
<td>10.9</td>
<td>22.1</td>
<td>-1.4</td>
<td>0.49</td>
</tr>
<tr>
<td>7</td>
<td>INDEX</td>
<td>1.1</td>
<td>15.8</td>
<td>-7.6</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>OPT-NP</td>
<td>10.4</td>
<td>23.6</td>
<td>-9.1</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-S</td>
<td>11.0</td>
<td>21.7</td>
<td>-5.6</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-M</td>
<td>9.9</td>
<td>19.9</td>
<td>-3.9</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-SM</td>
<td>9.9</td>
<td>20.1</td>
<td>-4.2</td>
<td>0.49</td>
</tr>
<tr>
<td>10</td>
<td>INDEX</td>
<td>1.1</td>
<td>15.8</td>
<td>-11.4</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>OPT-NP</td>
<td>10.5</td>
<td>21.9</td>
<td>-13.5</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-S</td>
<td>9.7</td>
<td>19.7</td>
<td>-9.8</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-M</td>
<td>9.5</td>
<td>18.0</td>
<td>-6.7</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>OPT-IP-SM</td>
<td>9.0</td>
<td>17.2</td>
<td>-5.9</td>
<td>0.52</td>
</tr>
</tbody>
</table>
As the next step, we will complete the following tasks of the empirical work:

- Fit the reduced-form models (B, B2, C, C2, and D). Repeat the analysis of Section 5.1 for each of the models. Compare the performance of portfolios based on different models and assess the source of return predictability.

- Evaluate the impact of transaction cost on portfolio performance.
REFERENCES


