Pomona College

Mathematics Thesis

How Does a Bayesian Investor Time the Market

Author:
Yuanxi Zhang

Supervisor:
Professor Hardin

May 10, 2013

1 Introduction

1.1 Portfolio Theory

Why does an investor want to invest in a diversified portfolio? Did not legendary investor Warren Buffett once was quoted by Suttmeier(2013) saying that "Diversification is protection against ignorance"? It is true that an investor of Mr. Buffett's expertise level only needs to focus on the investment vehicles that generate more than 20% annualized return on book value. However, few investors can beat the performance of Mr. Buffett's company, Berkshire Hathaway. Moreover, even an investment guru like Mr. Buffett suffers from volatility. In the Great Recession, the share price of Berkshire Hathaway dropped by more than 50% before it finally recovered.

Ordinary investors like us hate to bear the risk of losing our hard-earned money in the stock market. They would rather sacrifice some of the returns in exchange for a much lower risk. David Swensen, the portfolio manager of the Yale Endowment fund, is an advocator and pioneer of investing in a diversifited portfolio. Led by David Swensen, the Yale Endowment Fund allocates the assets in foreign stocks and bonds, private equity, hedge funds, natural resources, and real estate rather than the traditional 60-40 allocation in domestic stocks and bonds. According the the Yale Endowment Report (2012), the portfolio returned 13.7% annually over the last 20 years, which boosted the endowment size by a factor of seven. In his book *Unconventional Success*, Mr. Swensen claimed that 90 percent of the returns comes from the portfolio allocation rather than stock picking and market timing. In fact, according to Frazzini (2012) Swensen's Sharpe ratio (a measure of risk-adjusted return) from 1985 to 2008 was 1.12, about 50% higher than 0.76 Buffett achieved over his investment horizon.

Because of its effect on reducing risks while compromising to a slightly lower returns, constructing a diversified portfolio is a crucial task for investors. Markowitz (1959), who was awarded Nobel Economic Prize in 1990, discussed what is known as Modern Portfolio Theory and claimed that the efficient frontier is what every investor wants to achieve. However, in practice, an investor has to make numerous assumptions on the future returns, volatility and correlation. The model is very sensitive to the inputs, too. A tiny change in one of the inputs may lead to a dramatically different target optimal portfolio allocation. Thus, an investor that has little confidence on his subjective judgment of stocks and bonds will be confused about how to proceed with the model. Kelly (1956) discussed a novel way to determine the size of the bets that will maximize the long-term geometric returns. Some mathmaticians such as Edward Thorpe claimed to make money in risk arbitrage investments while using the system to decide portfolio sizes. However, this method also requires the inputs of predicted probability of winning, odds implied by the market price and correlations among investments, which are very hard to estimate precisely in the context of financial markets. Moreover, because the Kelly Criterion is an one-period model, it is difficult to compare investments with different time horizons.

Alternatively, by observing certain financial data, an investor can predict the future returns based on a simple linear regression. The regression method, although very crude, does not rely on any subjective judgment calls. Since a regression model is likely to have a very low R^2, a rational investor should not fully rely on the predictive values from a single-variate regression when making asset allocation decisions. Nonetheless, the investor is able to use the predicted return distributions as guides in adjusting his portfolio such that he maximizes his expected utility because even a skeptical prior distribution will lead to moderate market timing of the investor. The ability to maximize the expected utility for an investor using his prior belief as well as all the available data is the essential reason for using the Bayesian methods in the portfolio theory.

This paper is based on "Predictable returns and asset allocation: Should a skeptical investor time the market?" by Wachter and Warusawitharana (2009).

Consider an investor making an investment decision between stocks, bonds, and T-bills on January 1st, 2013. He has stock and bond return data from January 1900 to December 2012:

$$r_t = \alpha + \beta x_{t-1} + u_t \tag{1}$$

Where r_t is the continuous return vector on the S&P 500 stock index and Treasury Bonds separately in month t in excess of T-bills, α is a 2x1 vector of intercepts, β is a 2x1 vector of coefficients and x_{t-1} is a predictive variable such as the dividend yield at period $t-1$, which is used to predict the stock returns and the bond returns. The OLS estimators and the two R^2s of the equations can be calculated with the past data. Also known is x_T, which is the most recent dividend yield in December 2012. How will he allocate his portfolio according to this information?

1.2 Literature Review

Fama and Bliss (1987) and Cochrane and Piazzesi (2002) developed predictors of treasury bond returns based on forward rates.Cochrane and Piazzesi found that a tent-shaped function of one- to five-year forward rates forecasts bond returns.Fama and French (1989) and Campbell and Shiller (1991) found that a large term spread predicts higher excess bond returns. In addition, Fama and French (1988) claimed that the dividend yield is a good indicator for stock returns, especially in the long run. Kandel and Stambaugh (1995) argued that usual statistical measures underestimate the economic significance of predictive function when used by Bayesian investor. For example, in looking at a predcitive model which has $R^2 = 0.02$, people may think this predictive model has little influence on the asset allocation and dismiss the predictive function as insignificant. However, the usual measure of R^2 underestimates the economic significance of the predictive function. Kandel and Stambaugh suggested that even an uninformative prior with very low expectation on R^2 can have great influence on the asset allocation of an investor. Avramov (2002) used Bayesian Models to analyze the return predictability. He concluded that the Bayesian Model lead to superior models in terms of predictability as well as raw returns. Shanken and Tamayo (2004) expanded on Kandel and Stambaugh (1995) to accommodate the variation in expected risks as well as in expected returns when deploying the Bayesian Method. Specifically, they argued that the risk actually increases as dividend yield increases. A high dividend yield may reflect the fact that the stock prices have fallen substantially since the last dividend payment, which is a sign that the company may be in trouble. Alternatively, there might be a substantial decrease in the dividend in the next period, resulting in an inflated dividend yield if the last dividend payment is used as the input. Therefore, buying stocks when the dividend yield is high could lead to collecting a basket of companies in trouble. As a result, the optimal allocation may allocate less in stocks than the Kandel and Stambaugh (1995) model suggested.

The paper is organized as follows. Chapter 2 describes the likelihood function, the prior distributions, the calculation of the posterior, the Metropolis-Hastings Algorithm, and the calculation of optimal weights. Chapter 3 describes the results in Wachter (2009) including a comparison between the prior and posterior, the resulting portfolio weights, and the out-of-sample performance across different choices of priors. Chapter 4 concludes the whole paper.

2 Model Setup

2.1 Likelihood Function

At time t, investors can observe the data from T=0 to T=t. r_t is an Nx1 vector describing excess returns of N risky assets in period t, and x_t is a scalar describing the predictive variable at time t. We are using one independent variable to predict the returns of two assets. Let $D = [r_1, ..., r_t, x_1, ..., x_t]$ be the set of all the available information. In the paper, the predictive variable is either the dividend-price ratio or the yield spread between the five-year treasury note and three-month treasury bill. Here, because we only consider two risky assets in the model, namely the S&P 500 Index and the treasury bond, N=2. The data generating process is defined as follows:

$$r_{t+1} = \alpha + \beta x_t + u_{t+1} \tag{2}$$

$$x_{t+1} = \theta + \gamma x_t + v_{t+1} \tag{3}$$

α and β are 2x1 vectors: $\alpha^T = [\alpha_1, \alpha_2]$, and $\beta^T = [\beta_1, \beta_2]$. α_1 and β_1 are coefficients for stock returns, while α_2 and β_2 are coefficients for bond returns in the predictive model. The residuals in the above two equations, u_{t+1} and v_{t+1} have the size 2x1 and 1x1, respectively. They are assumed to follow a normal distribution with mean zero. The 3x3 matrix, Σ can be partitioned as:

$\Sigma = \begin{pmatrix} \Sigma_u & \Sigma_{uv} \\ \Sigma_{vu} & \Sigma_v \end{pmatrix}$

Where Σ_u is a 2x2 matrix describing the variance and covariances of u_{t+1}, Σ_{uv} is a 2x1 vector describing the covariance of u_{t+1} and v_{t+1}. Σ_v is a scalar describing the variance of v_{t+1}.

In other words, $[u_{t+1}, v_{t+1}] \mid [r_1, ..., r_t, x_0, ...x_t] \sim N(0, \Sigma)$.

In simple form, the predictive function can be written as $E[Y] = XB$, where

Y=

$$\begin{pmatrix} \mathrm{r}_1^T & \mathrm{x}_1 \\ \vdots & \vdots \\ \mathrm{r}_T^T & \mathrm{x}_T \end{pmatrix}$$

X=

$$\begin{pmatrix} 1 & \mathrm{x}_0 \\ \vdots & \vdots \\ 1 & \mathrm{x}_{T-1} \end{pmatrix}$$

B=

$$\begin{pmatrix} \alpha^t & \theta \\ \beta^t & \gamma \end{pmatrix}$$

Y is a Tx3 matrix; X is a Tx2 matrix, and B is a 2x3 matrix. Let p(D| B,Σ, x_0) be the likelihood function. Consider x_1, x_2,...,x_t drawn from a normal distribution, if they form a simple regression $y = xb + e_t$, the likelihood function is

$$p(D \mid b, \sigma_x, x_0) = (2\pi\sigma_x^2)^{-\frac{T}{2}} \exp(-\sum_{t=1}^{T}(\frac{1}{2\sigma_x^2}[(y_t - x_t b)^2]) \tag{4}$$

The likelihood function is calculated by multiplying the probability densitity function of x_1 through the probability density function of x_t. Similarly, when we have a multi-variate regression, we only need to make some small changes. From results in Zeliner (1996), the modified version of the likelihood function is as follows:

$$p(D \mid B, \Sigma, x_0) = \mid 2\pi\Sigma \mid^{-\frac{T}{2}} \exp(-\frac{1}{2}tr[(Y - XB)^T(Y - XB)\Sigma^{-1}]) \tag{5}$$

Here, the x_0 is treated as an observation we have, or a fixed number rather than a stochastic number. According to Stambaugh (1999), this treatment needs some in-depth discussion. If x_0 is a draw from normal distribution, it contains information on θ, γ, and σ, the information that a fixed point lacks. According to Stambaugh (1999), the bias can be as much as one third of the estimates. In addition, Kandel and Stambaugh (1995) suggest that if x_0 is regarded as nonstochastic, an investor loses information on x_0 because it assumes that the prior belief does not depend on x_0, even though his information set includes the pre-sample x_0 and the observed data x_1,...x_T.

According to the calculation by Poirier (1978), the difference between the unbiased method and the biased method is significantly non-zero. Therefore, we need to modify equation (5) to make it unbiased by treating x_0 as a draw from a normal distribution.

As Hamilton(1994) suggested, x_0 is treated as a draw from a normal distribution, and we use the following autoregressive function
$x_{t+1} = \theta + \gamma x_t + v_{t+1}$. Two assumptions are required here: v_is are indepenent and γ is less than 1. According to Hamilton(1994), the following steps can be taken to derive the likelihood of x_0. First, plug in the previous iteration:
$x_t = \theta + \gamma x_{t-1} + v_t$, and
$x_{t+1} = \theta + \theta\gamma + \gamma^2 x_{t-1} + \gamma v_t + v_{t+1}$. Continue the iteration process and use the formula for the sum of geometric series, we get the following:

$$
\begin{aligned}
& x_{t+1} \\
&= \theta + v_{t+1} + \gamma(\theta + v_t) + \gamma^2((\theta + v_{t-1}) + ... \\
&= \theta/((1-\gamma) + v_{t+1} + \gamma v_t + \gamma^2 v_{t-1})...
\end{aligned}
$$

The expectations of v_i are zero as we assumed. By taking the expectation of x_{t+1}, we see that

$$
\begin{aligned}
& \mu_x \\
&= E[x_{t+1} \mid B, \Sigma] = \theta/(1-\gamma) + 0 + \gamma * 0 + \gamma^2 * 0... \\
&= \theta/(1-\gamma)
\end{aligned}
$$

Therefore, the mean of x_{t+1} is $\mu = \frac{\theta}{1-\gamma}$. Given the assumption that the variances of v_i are constant and equal to σ_v^2 and v_i are independent, the variance of x_{t+1} is calculated by:

$$
\begin{aligned}
\sigma_x^2 & \\
&= E[(x_{t+1} - \mu_x)^2 \mid B, \Sigma] \\
&= E(v_{t+1} + \gamma v_t + \gamma^2 v_{t-1} ...)^2 \\
&= (1 + \gamma^2 + \gamma^4 + ...)\sigma_v^2 \\
&= \frac{\sigma_v^2}{1 - \gamma^2}
\end{aligned}
$$

.

Using the formula for geometric series, the expectation of $(x_t - \mu_x)^2$ is given by $\frac{\sigma_v^2}{1-\gamma^2}$. The likelihood function of x_0 is given by a standard normal distribution likelihood

$$p(x_0 \mid B, \Sigma) = (2\pi\sigma_x^2)^{-\frac{1}{2}} \exp(-\frac{1}{2\sigma_x^2}(x_0 - \mu_x)^2)$$

Combining the likelihood conditional on x_0 and the likelihood of x_0, we get the likelihood function of D, or the entire information set.

$$p(D \mid B, \Sigma) = p(D \mid x_0, B, \Sigma)p(x_0 \mid B, \Sigma) \tag{6}$$

$$= (2\pi\sigma_{x^2})^{-\frac{1}{2}} \mid 2\pi\Sigma \mid^{-\frac{T}{2}} \exp(-\frac{1}{2}\sigma_x^{-2}(x_0 - \mu_x)^2 - \frac{1}{2}tr[(Y - XB)^T(Y - XB)\Sigma^{-1}])$$

Equation (6) is the likelihood function we will use.

2.2 Prior Beliefs

Depending on the investor's beliefs, the prior belief β can be diffuse, meaning the predictive function does not contain much information and therefore has little or no predictability; or the prior belief on β can be dogmatic, meaning the predictive function gets a lot of trust. In between these two extremes lies the informative prior, which contains some information but not too much.

In the same line with other academic papers, investor's prior belief on β is assumed by Wachter and Warusawitharana (2009) to follow a normal distribution in this paper. β is the most important parameter that determines the predictability of the predictive function. The prior belief on β, however, is not isolated. It is also related to the other parameters.

Let $C_u C_u^T$ be the Cholesky decomposition of Σ_u. (i.e. $C_u C_u^T = \Sigma_u$). Basically we are taking the square root of Σ_u, a matrix.

Let $\eta = C_u^{-1}\sigma_x\beta$ to be the normalized β. Wachter and Warusawitharana (2009) made the following assumption about η: $\eta \sim N(0, \sigma_\eta^2 I_n)$, where I_n is identity NxN matrix. $\sigma_e ta$ is a constant assumed by investors.

B contains β, α, γ and θ. Assume that the prior belief for β is conditional on the other parameters:

$$p(B, \Sigma) = p(\beta \mid \alpha, \theta, \gamma, \Sigma)p(\alpha, \theta, \gamma, \Sigma) \tag{7}$$

Since $\eta \sim N(0, \sigma_e ta^2 I_n)$, and $\eta = C_u^{-1}\sigma_x\beta$, the mean of β is also 0 and the variance of β is $\sigma_e ta^2 \sigma_x^{-2} C_u I_n C_u^T = \sigma_n^2 \sigma_x^{-2}\Sigma_u$. It is much easier to find a prior on η than on β because η is closely related with R^2, which will be shown in the later part of the thesis. Therefore, the specification of distribution of β is:

$$\beta \mid \alpha, \theta, \gamma, \Sigma \sim N(0, \sigma_\eta^2 \sigma_x^{-2}\Sigma_u) \tag{8}$$

The priors on the rest of the parameters are independent, and the joint prior is the product of the priors. According to Stambaugh(1999) and Zellner(1996), the uninformative Jeffery's prior has the following form:

$$p(\alpha, \theta, \gamma, \Sigma) \propto \sigma_x \mid \Sigma_u \mid^{\frac{1}{2}} \mid \Sigma \mid^{-\frac{N+4}{2}} \tag{9}$$

Therefore, according to the appendix C and appendix D in Wachter and Warusawitharana (2009), the joint distribution of the prior is given by

$$p(B, \Sigma) = p(\beta \mid \alpha, \theta, \gamma, \Sigma)p(\alpha, \theta, \gamma, \Sigma) \propto \sigma_x^{N+1} \mid \Sigma \mid^{-\frac{N+4}{2}} exp(-\frac{1}{2}\beta^T(\sigma_\eta^2 \sigma_x^{-2}\Sigma_u)^{-1}\beta) \tag{10}$$

Since the prior distribution for β is a normal distribution, and the data are assumed to be normal, the posterior for β should also be in the form of a normal distribution.

The reason for separating out β in the prior is that β is closedly related to the R^2 of the predictive function. In fact, in the case of one asset,
$R^2 = \beta^2\sigma_x^2(\beta^2\sigma_x^2 + \Sigma_u)^{-1} = \frac{\eta^2}{\eta^2+1}$

In the case of multiple assets, max $R^2 = \frac{\eta\eta^T}{\eta\eta^T+1}$ Since R^2 can be expressed by functions containing η alone, a prior on η implies a prior on R^2. Depending on different beliefs of R^2, the specifications of η also varies.

2.3 Posterior Beliefs

According to Bayes' rule, the posterior is equal to a constant times likelihood function times prior. The mathematical formula is denoted by:

$$p(B, \Sigma \mid D) \propto p(D \mid B, \Sigma)p(B, \Sigma) \tag{11}$$

Where $p(D \mid B, \Sigma)$ is the likelihood of the data and $p(\mathrm{B}, \Sigma)$ is the joint prior distribution on all the parameters. After we plug in the results from previous parts, the equation does not take the form of standard density function. Thus, it is intractable to get the constant and come up with the exact distribution function. However, with the Metropolis-Hastings algorithm, we are able to approximate the posterior distribution. In a Metropolis-Hastings algorithm, we start from a point and a starting distribution and in each subsequent step, we sample a candidate point from the jumping distribution q, which is $\mid \Sigma \mid^{-\frac{T+N+4}{2}} \exp(-\frac{1}{2}tr[(Y - XB)^T(Y - XB)\Sigma^{-1}])$ when we sample Σ. Then, the acceptance ratio calculated and the next candidate point is given by either the new candidate point or the previous candidate point. In the end, the distribution of the simulated values will converge to a stationary posterier. The acceptance/rejection rule is that we accept the new candidate point with the probability of $a = \min(1, \frac{\pi(x*)q(x_t|x*)}{\pi(x_t)q(x*|x_t)})$, where $\pi(x)$ is the target posterior distribution of Σ. Otherwise, the old point is kept. The Metropolis-Hastings allows the jumping distribution to be asymmetrical, meaning $q(x* \mid x_t)$ does not have to equal to $q(x_t \mid x*)$. The speed for finding the stationary point in substantially increased, while both methods lead to convergence to the unique stationary distribution.

Here, we want to show that the Metropolis-Hastings method will approximate the posterier disbution that we are looking for. There are two steps involved in the proof: first, to prove that the algorithm generates a unique stationary distribution, and second, to prove that the stationary distribution equals the target posterior distribution.

First, we show that the algorithm generates a unique stationary distribution. According to Tierney(1994), suppose q is the jumping distribution and Π is a stationary distribution for the markov chain process. If a markov chain process is $\Pi - irreducible$ and $\Pi q = \Pi$, then the process is positive recurrent and Π is the unique stationary distribution. If the process is aperiodic, then the unique stationary distribution converages no matter what the initial distribution is. For a Metropolis-Hastings algorithm, irreducible means that any state has a positive probability of reaching to any other state; and aperiodic means the common divisor of the set of times that the chain returns to the initial point is 1. $\Pi q = \Pi$ imply that if the initial distribution equals π, then the distribution at time 1 also equals Π. With these conditions, the Metropolis-Hastings algorithm with jumping distribution q will generate a stationary distribution.

It is assumed that the algorithm used in the Metropolis-Hastings method has the characteristics of irreducible, positive recurrent and aperiodic. Therefore, the Metropolis-Hastings algorithm will generate a unique stationary distribution.

Second, we would like to show that the stationary distribution Π equals the target posterior distribution π. In the case of this paper, the posterior of Σ was derived first. The proof idea is from Christensen, Johnson, Branscum Hanson (2010).

Theorem: Choose any initial point Σ_0. For each step j=1,2,...M, a candidate point Σ_j* is sampled from the jumping distribution $q(\Sigma_{j*} \mid \Sigma_j)$. Let $Z = \Sigma_{j*}$, $Y = \Sigma_j$, and U have a uniform distribution $\mathcal{U}(0,1)$. Let

$$X = \Sigma_{j+1} = \begin{cases} \Sigma_j*, & \text{if U <a(Z} \mid Y) \quad (12) \\ \Sigma_j, & \text{otherwise} \quad (13) \end{cases}$$

where $a(Z,Y) = \min(1, \frac{\pi(Z)q(Y|Z)}{\pi(Y)q(Z|Y)})$. Assume that the following equation is satisfied:

$$a(Z,Y)\pi(Y)q(Z,Y) = a(Y,Z)\pi(Z)q(Y,Z).$$

We will show that if the density of Y is $\pi(Y)$ then the density of X, g(X), equals to the target posterior distribution $\pi(X)$.

Proof of the theorem: Consider the two possible outcomes of the step: $U < a(Z \mid Y)$ and $U > a(Z \mid Y)$. In the first case, Z becomes the new X. In the second case, X remains at Y. Assume that $\delta(X - Y) = 1$ if $X = Y$ and $\delta(X - Y) = 0$ otherwise; $\delta(X - Z) = 1$ if $X = Z$ and $\delta(X - Z) = 0$ otherwise. Then, we have the derivations below.

$$\begin{aligned}
g(X) &= \sum_Z \sum_Y \pi(Y) q(Z \mid Y) a(Z, Y) \delta(X - Z) + \\
&\quad \sum_Z \sum_Y \pi(Y) q(Z \mid Y)(1 - a(Z, Y)) \delta(X - Y) \\
&\quad \text{g(x) is the sum of two cases mentioned above} \\
&= \sum_Y \pi(Y) q(X \mid Y) a(X, Y) + \\
&\quad \sum_Y \delta(X - Y) \sum_Z [1 - a(Z, Y)] q(Z \mid Y) \pi(Y) \\
&\quad \text{Since in the first case Z is accepted as X, we can replace Z with X.} \\
&= H1 + H2
\end{aligned}$$

Let H1 and H2 denote the two parts of the g(X). First consider the case in which $U < a(Z \mid Y)$ happened, denoted by H1. Divide the case into two subcases: $a(X, Y) = 1$ and $a(X, Y) \neq 1$. Then, plug in the value of a(X,Y) and we have the following function.

$H1$

$$
\begin{aligned}
&= \sum_{Y} \pi(Y) q(X \mid Y) a(X, Y) \\
&= \sum_{(Y: a(X,Y)=1)} \pi(Y) q(X \mid Y) a(X, Y) \\
&\quad + \sum_{(Y: a(X,Y)\neq 1)} \pi(Y) q(X \mid Y) a(X, Y)
\end{aligned}
$$

Divide the case into two subcases: a(X,Y)=1 and $a(X,Y) \neq 1$

$$
\begin{aligned}
&= \sum_{(Y: a(X,Y)=1)} \pi(Y) q(X \mid Y) \\
&\quad + \sum_{(Y: a(X,Y)\neq 1)} \pi(Y) q(X) \\
&\quad * \frac{\pi(X) q(Y \mid X)}{\pi(Y) q(X \mid Y)}
\end{aligned}
$$

Plug in the values of a(X,Y) in the two subcases, based on the definition of a(X,Y) specified earlier.

$$
= \sum_{(Y: a(X,Y)=1)} \pi(Y) q(X \mid Y) + \sum_{(Y: a(X,Y)\neq 1)} \pi(X) q(Y \mid X)
$$

Then, consider the case in which $U > a(Z \mid Y)$ happened, denoted by H2. Again, divide this case into two subcases and plug in the values of $a(Z, Y)$ in each subcase.

$H2$

$$= \sum_{Y} \delta(X-Y) \sum_{Z} [1-a(Z,Y)] q(Z \mid Y) \pi(Y)$$

$$= \sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)=1)} [1-a(Z,Y)] q(Z \mid Y) \pi(Y) +$$

$$\sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} [1-a(Z,Y)] q(Z \mid Y) \pi(Y)$$

Divide the case into two subcases: a(X,Y)=1 and $a(X,Y) \neq 1$

$$= 0 + \sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} [1-a(Z,Y)] q(Z \mid Y) \pi(Y)$$

After we plug a=1 for the first term, we see that

the first term become 0. Only the second term is left.

$$= \sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} q(Z \mid Y) \pi(Y) -$$

$$\sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} q(Z \mid Y) \pi(Y) a(Z,Y)$$

$$= \sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} q(Z \mid Y) \pi(Y) -$$

$$\sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} q(Z \mid Y) \pi(Y)$$

$$* \frac{\pi(Z) q(Y \mid Z)}{\pi(Y) q(Z \mid Y)}$$

Here we plug in the value of $a(Z,Y)$when $a(Z,Y) \neq 1$

$$= \sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} q(Z \mid Y) \pi(Y) -$$

$$\sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} \pi(Z) q(Y \mid Z)$$

$$= \sum_{Y} \delta(X-Y) \sum_{(Z:a(Z,Y)\neq 1)} [q(Z \mid Y) \pi(Y) -$$

$$\pi(Z) q(Y \mid Z)]$$

$$= \sum_{(Z:a(Z,X)\neq 1)} [q(Z \mid X) \pi(X) - \pi(Z) q(X \mid Z)]$$

In this case, the value of Y is adopted as the value for X

$$= \sum_{(Y:a(Y,X)\neq 1)} [q(Y \mid X) \pi(X) - \pi(Y) q(X \mid Y)]$$

Change the variable name from Z to Y

By definition of the function a(X,Y), $(Y : a(X,Y) = 1) = (Y : a(Y,X) \neq 1) \cup (Y : q(Y \mid X)\pi(X) = \pi(Y)q(X \mid Y)$ because the left side of the equation implies either $a(Y,X) = 1$ or $a(Y,X) < 1$. We will use this condition in the following derivation. In addition, we know that $\sum_Y q(Y \mid X) = 1$ as q(Y $\mid X)$ is a well-defined conditional density. Then, the above equation can be modified as follows.

$g(X)$

$$= \sum_{(Y:a(Y,X)\neq 1)} [q(Y \mid X)\pi(X) - \pi(Y)q(X \mid Y)] + \sum_{(Y:a(X,Y)=1)} \pi(Y)q(X \mid Y) + \sum_{(Y:a(X,Y)\neq 1)} \pi(X)q(Y \mid X)$$

Take the sum of H1 and H2

$$= \sum_{(Y:a(Y,X)\neq 1)} [q(Y \mid X)\pi(X) - \pi(Y)q(X \mid Y)] + \sum_{(Y:a(X,Y)\neq 1)} \pi(X)q(Y \mid X) + \sum_{(Y:a(Y,X)\neq 1)} \pi(Y)q(X \mid Y) + \sum_{(Y:q(Y|X)\pi(X)=\pi(Y)q(X|Y)} \pi(Y)q(X \mid Y)$$

Use the condition that$(Y : a(X,Y) = 1) = (Y : a(Y,X) \neq 1) \cup (Y : q(Y \mid X)\pi(X) = \pi(Y)q(X \mid Y))$

$$= \sum_{(Y:a(X,Y)\neq 1)} \pi(X)q(Y \mid X) + \sum_{(Y:q(Y|X)\pi(X)=\pi(Y)q(X|Y)} \pi(X)q(Y \mid X) + \sum_{(Y:a(Y,X)\neq 1)} \pi(X)q(Y \mid X)$$

Under the condition of $(q(Y \mid X)\pi(X) = \pi(Y)q(X \mid Y)$, we can exchange $q(Y \mid X)\pi(X)$with $\pi(Y)q(X \mid Y)$ Also, the plus and minus $\sum_{(Y:a(Y,X)\neq 1)} \pi(Y)q(X \mid Y)$ cancel out each other

$$= \Sigma_Y q(Y \mid X)\pi(X)$$

The previous step included all three possible scenarios of Y

$$= \pi(X)$$

because $\sum_Y q(Y \mid X) = 1$

Therefore, $g(X) = \pi(X)$ and X has the distribution π, which is the target posterior distribution. We have shown that starting from any distribution, using Metropolis-Hastings algorithm we will get the stationary distribution Π that is equal to the target distribution π.

According to Chib and Greenberg (1995), it is faster to use the block-at-a-time algorithm. Here, suppose the vector B is devided into two blocks, with b_1 consists of all the α_i and β_i, and b_2 consists of the θ and γ in the predictive functions. Then, by the block-at-a-time algorithm, Wachter and Warusawitharana (2009) first sample from $p(\Sigma \mid B, D)$, then sample from $p(b_1 \mid b_2, \Sigma, D)$, and finally sample from $p(b_2 \mid b_1, \Sigma, D)$. The jumping distribution for Σ is $\mid \Sigma \mid^{-\frac{T+N+4}{2}} \exp(-\frac{1}{2}tr[(Y - XB)^T(Y - XB)\Sigma^{-1}])$. The proposal distribution for b_1 and b_2 are both normal. After sampling from the conditional posteriors with the order of Σ, b_1, and b_2 , it is possible to arrive at the converging full posterior on all of the variables.

2.4 Predictive Distribution

Given the histogram of posterier distribution $p(B, \Sigma \mid D)$, we want to use the information to determine portfolio decisions. To do so, we need to derive the predictative probability function $p(r_{T+1} \mid D)$. First, we multiply the posterier by the likelihood function of future observation.

$$p(r_{T+1}, B, \Sigma \mid D) = p(r_{T+1} \mid B, \Sigma, D)p(B, \Sigma \mid D) \tag{14}$$

Integrating the joint density function gives the predictive density distribution function.

$$p(r_{T+1} \mid D) = \int_p (r_{T+1}, B, \Sigma \mid D)dBd\Sigma = \int p(r_{T+1} \mid B, \Sigma, D)p(B, \Sigma \mid D)dBd\Sigma \tag{15}$$

Utility is a measurement of happiness of an investor associated with the return in a specific state. An investor has to maximize the expected utility $E(U)$ next period given all the available information at time T. He has to achieve $\max E_T[U(W_{T+1}) \mid D]$ by adjusting his portfolio weights between the stock index, risky bond, and risk-free bonds. W_{T+1} is the wealth of the investor in the next period and is calculated by taking the weighted average of expected returns of the asset returns in his portfolio according to his allocation weights. The wealth next period is thus given by:

$$W_{T+1} = W_T[w_1(r_{t+1,1} + i_{T+1}) + w_2(r_{t+1,2} + i_{T+1}) + (1 - w_1 - w_2)i_{T+1}] \quad (16)$$

Where $r_{t+1,1}$ is the predicted return for the stock index at time $T+1$, $r_{t+1,2}$ is the predicted return for the Treasury bonds at time $T+1$, and i_{T+1} is the return for risk-free cash savings observed at time T. w_1, w_2 are weights for the stock index and bonds at time T, respectively. 1-$w_1 - w_2$ is the weight on the risk-free asset. The function is set up so that the total return in the next period is determined by the weighted average of returns in different assets in the current period.

Depending on different assumptions on the utility, the optimal weight is also different. Wachter and Warusawitharana (2009) used a quadratic utility function. The function is easy to analyze mathematically. However, the function is unusual because economics theory usually suggests a diminishing yet positive marginal utility. A diminishing marginal utility implies that for someone who has USD 100 of wealth another USD 100 gives him substantial utility; while for someone who has USD 10000000 of wealth, another USD 100 means almost nothing to him. In contrast, in a quadratic utility function, there is a satiation level beyond which the investor prefers less returns to more. This is an implausible assumption as people typically do prefer a high return. A typical risk-averse person in economics is sometimes assumed to have iso-elastic utility instead. An iso-elastic utility function describes a risk-averse person that has diminishing marginal utility. Iso-elastic functions have the same elasticity, which is the ratio of percentage change in the dependent variable to the percentage change in the independent variable. Suppose that we do use the quadratic utility. Then, the investor demands the greatest risk-adjusted return $R = E[r_p] - \frac{A}{2}var[r_p]$, where A is the risk-averse coefficient with a larger value indicating a less risk-averse person. The risk-adjusted return is also called certainty-equivalent return (CER). The risk-adjusted return under the quadratic utility assumes that people are risk averse. Thus, a certain 3% return is better than a fair coin flip with 0% and 6% payoff. Moreover, given the same expected return, the larger the variance of the payoffs, the less desirable the strategy is. Furthermore, if the risk-averse coefficient for someone is large, he dislikes volatility and will get a lower utility than someone with a small risk-tolerant coefficient when judging a same portfolio.

To maximize the risk-adjusted return, we just need to solve the problem: $\max[w_1(r_{t+1,1}+i_{T+1})+w_2(r_{t+1,2}+i_{T+1})+(1-w_1-w_2)i_{T+1}]-\frac{A}{2}var[w_1(r_{t+1,1}+i_{T+1})+w_2(r_{t+1,2}+i_{T+1})+(1-w_1-w_2)i_{T+1}]$, such that the sum of three weights is 1

Simplifiying the above equation, it is equivalent to $\max[i_{t+1}+w_1r_{t+1,1}+w_2r_{t+1,2}-\frac{A}{2}(\sum_{i=1}^{2}\sum_{j=1}^{2}\sigma_{ij}w_iw_j]$, where σ_{ij} is the variance-covariance matrix of stock returns and bond returns:

$\sigma_{ij}=\begin{pmatrix}\sigma_{1,1} & \sigma_{1,2}\\ \sigma_{2,1} & \sigma_{2,2}\end{pmatrix}$

Take the first derivative and let the derivative equal to zero, we have $E-A\sum_{j=1}^{2}\sigma_{ij}w_j=\vec{0}$ where E^T is a 2x1 vector $[r_{t+1,1},r_{t+1,2}]$. Then, it is possible to solve for the weight in stocks and bonds. Since the sum of weights allocated in stocks, bonds, and riskless assets equals to 1, we can solve for the weight of wealth allocated to riskless asset once we know w_1 and w_2 By solving the equation above, we get $w=\frac{\sigma^{-1}E}{A}$. This weight vector is the one that optimizes the overall utility of an investor, given all the data, the investor's prior beliefs and the investor's risk tolerance level. After the weight vector is calculated each period, the investor's actual return in the period can also be calculated by the end of the period. It is then possible to compare realized annualized returns from different prior beliefs and different risk tolerance levels.

3 Results and Discussion

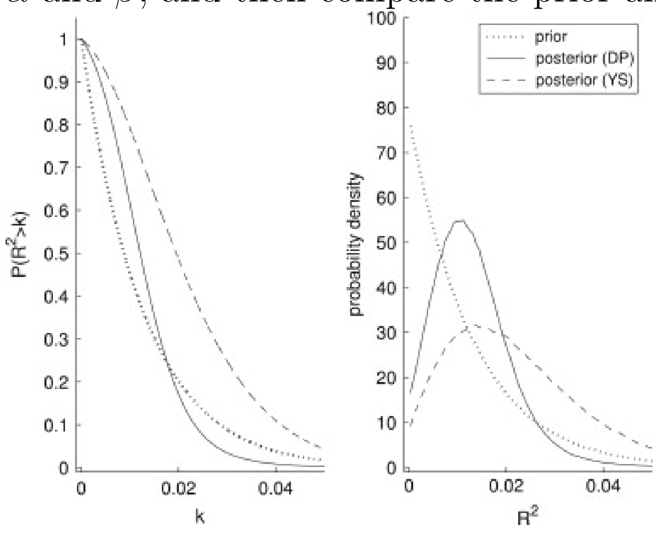
Wachter and Warusawitharana (2009) tried to build an asset allocation with risk-free bonds, long term treasury bonds, and the S&P 500 index. Two sets of regressions are run, first using the dividend-to-price ratio, and second using the yield spread. Furthermore, Wachter and Warusawitharana (2009) seeked to optimize the utility which is assumed to be quadratic.

3.1 Data

Quarterly stock index, 3-month treasury bond, and 10-year treasury bond data are collected. The excess returns for stocks and bonds are calculated by subtracting the quarterly returns of 3-month treasury bond from the that of the stock index and the 10-year treasury bond. The dividend-price ratio is the amount of total dividends paid in the index in the previous 12 months divided by the price level. The yield spread is calculated by subtracting the yield of 3-month treasury bond from the yield of 5-year treasury bond. The data were collected from 1952 to 2004.

3.2 Comparing the Posterior and Prior

Wachter and Warusawitharana (2009) ran one simulation on each of the predictive variables to get the histogram of posterior on R^2. The simulation is performed using the Metropolis-Hastings method specified above. Some 100,000 initial burn-in data are discarded because they are noisy. Since the prior on η is also the same as the prior on R^2 as we demonstrated before, we can construct the posterior on R^2 based on other primative variables such as α and β, and then compare the prior and the posterior distribution on R^2.



The left side of the graph shows that the investor's belief on the predictability changed from prior to posterior. In other words, the data changed the way the investor think about the prediction model.

From the graph, it's obvious that the posterior probability that R^2 exceeds k is greater than the prior probability when k <0.02 when the posterior is based on the dividend yield. After k=0.02, the posterior probability is lower than the prior probability. When the posterior is based on the yield spread, on the other hand, the posterior probability that R^2 exceeds k is always above the prior probability.

The right side of the graph shows the different probability density function over the whole range of R^2. Again, the data changed the investor's perception of the predictability of the model. While the probabily density function for the prior is strictly decreasing over the whole R^2, both of the probability density functions for posterior first rise, then peak at around R^2=0.02 before they fall back. Since the posterior with the yield spread data generally was always above the posterior with the dividend yield data, it reflects a more diffuse belief, and the probability density function is flatter and has a fatter tail.

Parameter	$P(R^2>2\%)$				Reg.
	0	0.0005	0.075	0.999	
Panel A: Dividend yield					
β_{bond}	0.00	0.02	−0.00	−0.18	−0.10
	(0.00)	(0.22)	(0.43)	(0.72)	(0.73)
β_{stock}	0.00	0.69	1.41	1.46	2.72
	(0.00)	(0.62)	(0.97)	(1.09)	(1.52)
θ_1	0.997	0.993	0.988	0.989	0.976
	(0.002)	(0.006)	(0.009)	(0.010)	(0.015)
$E[r_{\text{bond}}\|B,\Sigma]$	0.18	0.18	0.18	0.17	0.23
	(0.27)	(0.30)	(0.34)	(1.07)	
$E[r_{\text{stock}}\|B,\Sigma]$	1.16	1.17	1.17	1.17	1.09
	(0.29)	(0.24)	(0.28)	(0.72)	
$E[x\|B,\Sigma]$	−3.49	−3.50	−3.50	−3.50	−3.72
	(1.48)	(0.99)	(0.76)	(1.35)	
Panel B: Yield spread					
β_{bond}	0.00	0.20	0.46	0.81	0.80
	(0.00)	(0.14)	(0.20)	(0.26)	(0.26)
β_{stock}	0.00	0.22	0.51	0.89	0.89
	(0.00)	(0.28)	(0.42)	(0.55)	(0.56)
θ_1	0.74	0.73	0.74	0.75	0.74
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
$E[r_{\text{bond}}\|B,\Sigma]$	0.21	0.21	0.21	0.21	0.23
	(0.28)	(0.28)	(0.29)	(0.33)	
$E[r_{\text{stock}}\|B,\Sigma]$	1.67	1.67	1.67	1.67	1.69
	(0.58)	(0.59)	(0.60)	(0.63)	
$E[x\|B,\Sigma]$	0.97	0.97	0.97	0.97	0.99
	(0.19)	(0.19)	(0.19)	(0.21)	

Also shown are the posterior means for values of σ_η equal to 0, 0.04, 0.08, and ∞. This corresponds to a $P(R^2 > 2\%)$ of 0, 0.0005, 0.075, and 0.999, respectively. The different values of σ_η reflects the different prior belief characteristics. A low $P(R^2 > 2\%)$ reflects a dogmatic view that the regression function has no predictive power, while a higher value reflects a diffuse view and uninformative prior. The results of posterier means are also compared with the results from a simple OLS.

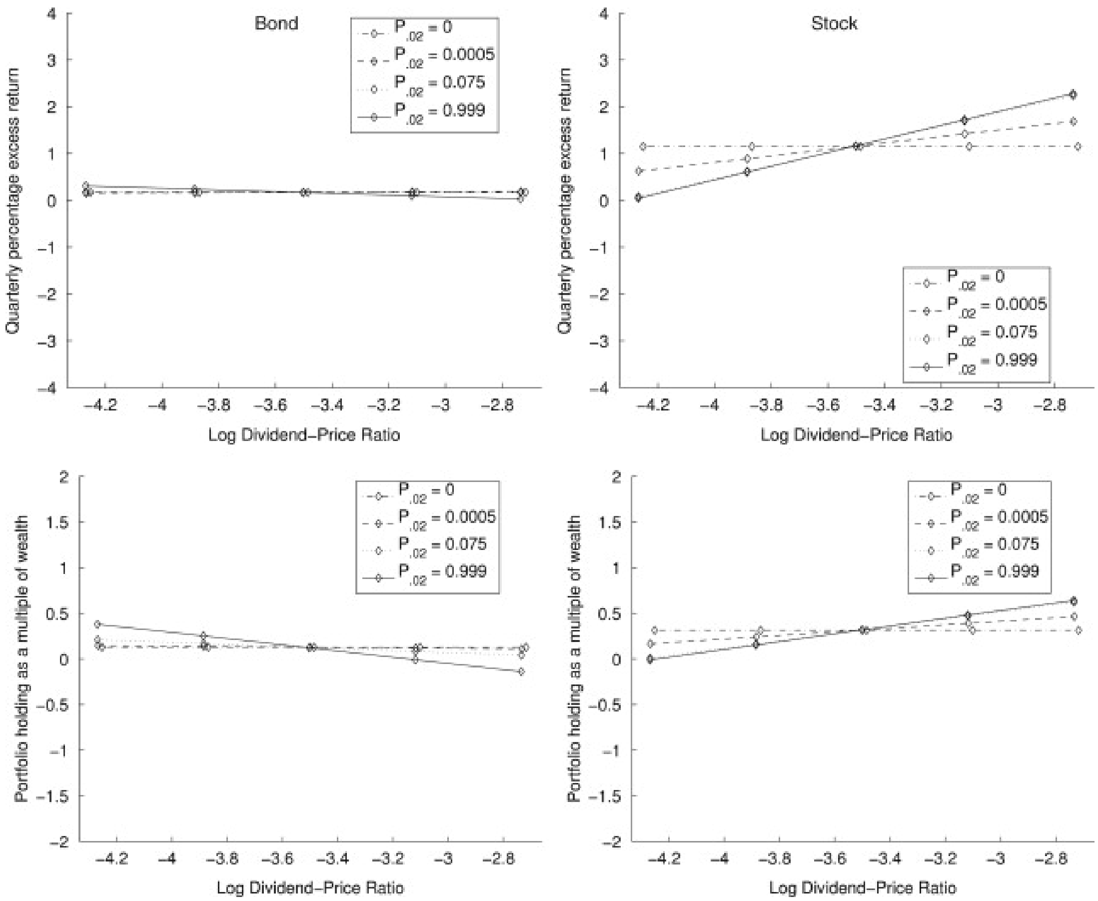
As Panel A shows, the βs for stocks are positive and not close to zero when $P(R^2 > 2\%) = 0.0005$, 0.075, and 0.999, although not statistically significant, implying that dividend yield does predict the stock returns. The dividend yield does not predict bond returns as the βs for bonds are close to zero. As the prior becomes more informative, the β on stocks become closer to zero, since the investor believes that the regression function has less predictive power now. For example, when $P(R^2 > 2\%)$=0.999, the β for stock is 1.46. When $P(R^2 > 2\%)$=0.0005, the β for stock is only 0.69.

As Panel B shows, yield spread predict the bond returns well since the coefficients for β are significantly non-zero when . Similarly, as the prior becomes more dogmatic, the β for bonds decreases.

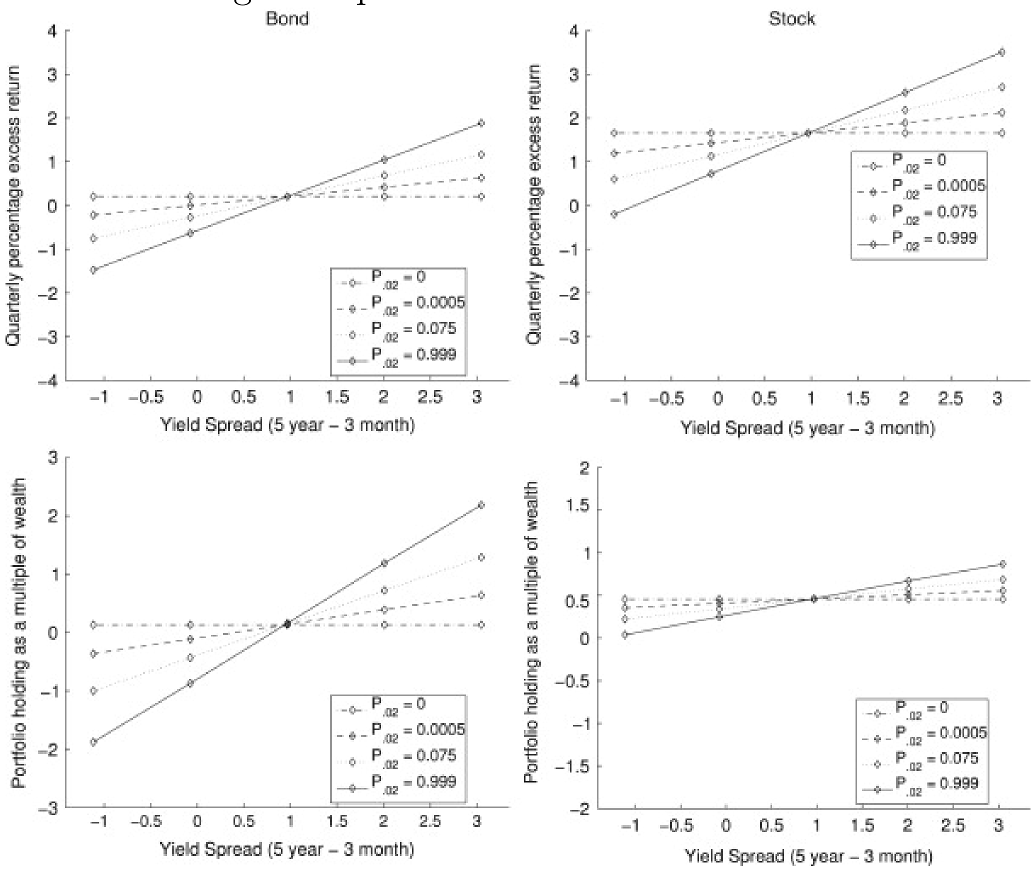
The coefficient for the autoregressive model in Panel A and Panel B, which are denoted as θ_1 in the table, are very close to 1 and statistically significant, showing that the predictive variables themselves have strong correlations over time.

Interestingly, the long-run excess returns for stocks and bonds, which are $E[r_bond \mid B, \Sigma]$ and $E[r_stock \mid B, \Sigma]$, are relatively stable in each Panel, no matter what prior we use. For example, in Panel A, $E[r_bond \mid B, \Sigma]$ equals 0.18 or 0.17 within the whole range of $P(R^2 > 2\%)$. The 0.18 number is also close to the excess return result from the OLS regressions, which gives 0.23 excess returns.

The expected value for the prediction variable, $E[x \mid B, \Sigma]$, also shows stability under different priors. In Panel A, $E[x \mid B, \Sigma]$ equals -3.49 or -3.50; in Panel B, the expected value equals 0.97.



The upper part of the graph shows the quarterly excess returns on stocks and bonds under different priors with dividend yield as predictive variable. On the x-axis is the Log Dividend Yield(predictive variable x_t); on the y-axis is the percentage excess returns. The solid lines denote diffuse priors, and the dotted lines denote dogmatic priors. The slope of the lines correspond to β from the table 1. While the excess returns for bonds don't change much because of poor predictability of the dividend yield on bond returns, the excess return for stocks increases substantially as dividend yield increases for diffuse priors. The excess returns keep at a similar level for dogmatic priors. The bottom part of the graph shows the suggested weight of stocks and bonds. For a diffuse prior, the weights on stocks can vary from 0% to more than 50% as log dividend yield rises from -4.2 to -2.8. The stock weights for dogmatic prior remains at about 30% as the log dividend yield changes. Similarly, the weights on bonds for a diffuse prior falls from less than 50% to less than 0% as the log dividend yield rises; while the weights do not change much for a dogmatic prior.

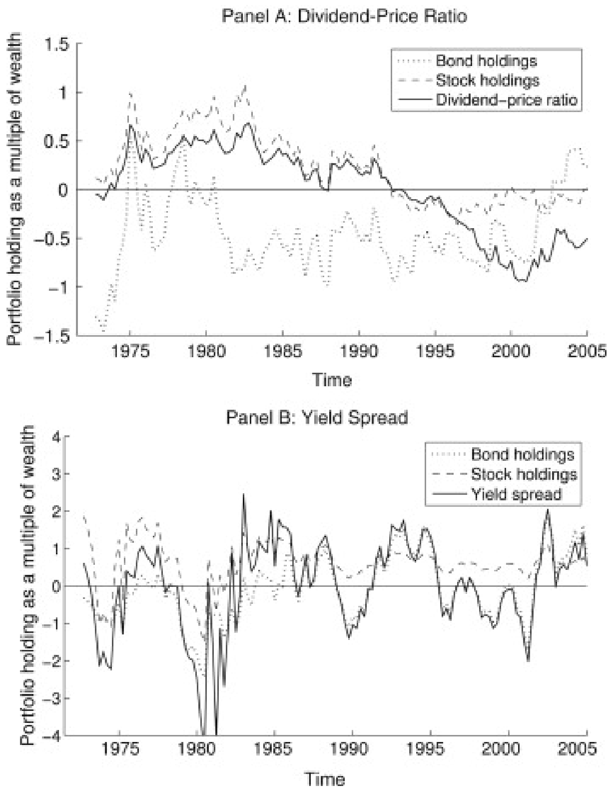


The upper part of the graph shows the quarterly excess returns on stocks and bonds under different priors with yield spread as predictive variable. Same as the previous graph, on the x-axis is the Yield Spread (predictive variable x_t); on the y-axis is the percentage excess returns. The solid lines denote diffuse priors, and the dotted lines denote dogmatic priors. The slope of the lines again correspond to β from the table 1. For a diffuse prior, the excess returns for stocks and bonds increases as yield spreads increases. For example, the excess return for bonds increases from -2% to 2% as yield spread increases from -1% to 3%. The dogmatic prior sees the excess returns remain almost unchanged as yield spread increases. The bottom part of the graph shows the suggested weight of stocks and bonds. For a diffuse prior, the weights on stocks and bonds vary from 0% to more than 75% and from -200% to 200% respectively as yield rises from -1 to 3. The stock and bond weights for dogmatic prior remains about the same as the log dividend yield changes.

3.3 The Implication on Asset Allocation

Wachter and Warusawitharana (2009) simulated the posterior starting in 1972 to allow for enough data. The posterior is approached by simulating 2,000,000 data and discarding the first 50,000. The investor was assumed to have a diffuse prior and risk aversion coefficient of A=5. The investor rebalances the portfolio at the beginning of each quarter. From the simulation, Wachter and Warusawitharana (2009) were able to calculate the out-of-sample performance in the next section. In the graph, the x-axis is time and the y-axis is portfolio weights. Since the graph shows the asset allocation using the most diffuse priors ($\sigma_\eta = 0$), the asset allocation is inevitably more volatile than the allocation with more dogmatic priors . Panel A shows the weights with the dividend yield as the predictive variable. Over the whole period, the weight for long-term bond holding, denoted by the dotted line, was negative (shorting), showing that the bond exhibit unfavorable risk-adjusted returns over the period. The weights for stocks, denoted by dashed line, has been positive until 1993, when the internet bubble and the "new economy" started to develop. From this time, the correlation between stocks and the dividend yield started to decrease. Notice that the dividend yield and the stock holding weights are positively correlated, showing that the higher the dividend yield, the higher the expected stock returnss, and therefore the higher weights are assigned to stocks. The bonds exhibit much smaller correlation with dividend yield.

Panel B shows the change in weights over time with yield spread as the predictive variable. Again, the weights hown here were used to calculate the returns in the next section. Here, the bond weights, denoted by dotted lines, follow the yield spread closely. For example, when the yield spread dropped substantially in 1980, the bond weights also dropped from 0% to −400%. When the yield spread spiked from 2002 to 2003, the bond weights followed the trend and increased from −200% to 200%. The stock weights, denoted by dashed lines, levelled off after mid-1990s.



3.4 Performance

To assess the performance of the model, the concept of risk-adjusted return is again used. This measure tell us how much the strategy worth by calculating how much the investor is willing to accept as certainty returns in exchange for giving up the current strategy. The following table shows the realized risk-adjusted returns for different priors and risk-tolerance coefficients based on two datasets over the twenty-two year period.

A	$P(R^2>2\%)$				Reg.
	0	0.0005	0.075	0.999	
Panel A: Dividend-price ratio					
Certainty equivalent returns					
2	6.98	7.66	6.76	5.25	−2.63
5	6.78	7.05	6.66	6.03	2.86
Sharpe ratios					
	0.19	0.22	0.19	0.15	0.16
Panel B: Yield spread					
Certainty equivalent returns					
2	5.61	7.58	9.19	5.47	0.65
5	6.26	7.06	7.72	6.25	4.35
Sharpe ratios					
	0.19	0.28	0.37	0.43	0.39

The panel A of the above table hows the Sharpe ratios and the risk-adjusted returns when dividend yield is the predictor. The higher the Sharpe ratio and the risk adjusted return, the better. It is interesting that market timing does increase the out-of-sample performance relative to OLS and the dogmatic prior. In fact, $P(R^2 > 2\%) = 0.0005$ has the best risk-adjusted return performance and the highest Sharpe ratio.

Panel B of Table 2 shows analogous results for the yield spread as the predictor variable. OLS again performs the worst in the risk-adjusted return metric, and the diffuse and dogmatic priors perform better. However, skeptical priors perform best. When A (the risk tolerance parameter) equals 2, a diffuse prior results in 5.47% annualized return, while a dogmatic prior returns 5.61% annually. However, For the prior with $P(R^2 > 2\%) = 0.075$, the annualized return is 9.19%. These results show that the skeptical prior lead to superior out-of-sample performance over the postwar period, as well as to less extreme portfolio allocations.

End date	A	$P(R^2>2\%)$				Reg.
		0	0.0005	0.075	0.999	
Panel A: Dividend-price ratio						
Certainty equivalent returns						
1984	2	6.64	7.99	11.05	11.42	−3.14
	5	8.08	8.62	9.86	9.97	4.23
1994	2	6.35	7.27	9.34	9.03	2.60
	5	7.21	7.58	8.40	8.24	5.69
Sharpe ratios						
1984		0.08	0.16	0.33	0.37	0.38
1994		0.10	0.15	0.28	0.29	0.34
Panel B: Yield spread						
Certainty equivalent returns						
1984	2	3.11	4.74	4.83	−9.24	−21.91
	5	6.67	7.36	7.44	1.92	−3.07
1994	2	4.55	6.44	7.63	0.79	−6.26
	5	6.50	7.28	7.78	5.09	2.31
Sharpe ratios						
1984		0.01	0.09	0.20	0.32	0.27
1994		0.08	0.18	0.29	0.38	0.34

As shown from the above graph, the strategy performs quite consistently over subperiods as well. In Panel A, when A=2 and the time period is from 1974 to 1984, we see that an investor with skeptical priors generated risk-adjusted return of 7.99%, which was higher than the 6.64% return generated with dogmatic prior. When A=5, the skeptical investor generated 8.62% in risk-adjusted returns, which is 54 basis points higher than an invesor with dogmatic return. In each subperiod, the skeptical priors outperform the dogmatic priors with the values of risk-tolerance coefficient being 2 or 5.

4 Conclusion

The paper models the portfolio choice problem based on the different beliefs on the prior, ranging from dogmatic to skeptical and diffuse. When the method is used in the post-war data, it is shown that even a very skeptical investor would moderately time the market. In addition, the weights are less volatile and deliever superior out-of-sample performances when the investor is skeptical of the predictability of the regression model.

5 References

Avramov, D.(2002). return predictability and model uncertainty. *Journal of Financial Economics*, 64, 423-458.

Campbell, J. Y. Shiller, R. J. (1991).Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies*, 58, 495-514.

Chib, S. Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49, 327-335.

Christensen, R. Johnson, W. Branscum, A. Hanson,T. (2010). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians.* Boca Raton, FL:CRC Press

Cochrane, J. H. Piazzesi, M. (2002). Bond risk premia. *National Bureau of Economic Research.*

Fama, E. F. Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 680-692.

Fama, E. F. French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22, 3-25.

Fama, E. F. French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25, 23-49.

Frazzini, A. Kabiller, D. Pedersen, L. H. (2012). Buffetts Alpha. Unpublished manuscript, Yale University.

Hamilton, J. D. 1994. *Time series analysis.* Cambridge: Cambridge University Press.

Kandel, S. Stambaugh, R. F. (1995). On the Predictability of Stock Returns: An Asset-Allocation Perspective. *National Bureau of Economic Research*

Kelly, J. B. (1956). A New Interpretation of Information Rate. *Bell System Technical Journal*, 35, 917-926.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investment.* New York: John Wiley and Sons.

Poirier, D.J. (1978). The effect of the first observation in regression models with first-order autoregressive disturbances. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, 27, 6768.

Shanken, J. Tamayo, A. (2004). Dividend yield, risk, and mispricing: A Bayesian analysis. London Business School Accounting Subject Area Working Paper.
Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54, 375-421.
Suttmeier, R.H. (2013, Feb 11). Why It's Time To Move More Into Cash. *Forbes*, retrieved from http://www.forbes.com/pictures/gg45fmimi/2-diversification-wont-protect-you-2/
Swensen, D. (2005). *Unconventional Success: A Fundamental Approach to Personal Investment*. Free Press.
Tierney,L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics*, Vol. 22, No. 4, 1701-1728.
Wachter, J. A. Warusawitharana, M. (2009). Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148, 162 - 178.
Yale Endowment Report, (2012), retrieved from http://investments .yale.edu/images/documents/YaleEndowment12.pdf
Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* New York: J. Wiley