Predictable returns and asset allocation:
Should a skeptical investor time the market?*

Jessica A. Wachter

University of Pennsylvania and NBER

Missaka Warusawitharana

Board of Governors of the Federal Reserve System

September 10, 2008

Predictable returns and asset allocation: Should a skeptical investor time the market?

Abstract

We investigate optimal portfolio choice for an investor who is skeptical about the degree to which excess returns are predictable. Skepticism is modeled as an informative prior over the R^2 of the predictive regression. We find that the evidence is sufficient to convince even an investor with a highly skeptical prior to vary his portfolio on the basis of the dividend-price ratio and the yield spread. The resulting weights are less volatile and deliver superior out-of-sample performance as compared to the weights implied by an entirely model-based or data-based view.

Introduction

Are excess returns predictable, and if so, what does this mean for investors? In classic studies of rational valuation (e.g. Samuelson (1965, 1973), Shiller (1981)), risk premia are constant over time and thus excess returns are unpredictable.[1] However, an extensive empirical literature has found evidence for predictability in returns on stocks and bonds by scaled-price ratios and interest rates.[2]

Confronted with this theory and evidence, the literature has focused on two polar viewpoints. On the one hand, if models such as Samuelson (1965) are correct, investors should maintain constant weights rather than form portfolios based on possibly spurious evidence of predictability. On the other hand, if the empirical estimates capture population values, then investors should time their allocations to a large extent, even in the presence of transaction costs and parameter uncertainty.[3] Between these extremes, however, lies an interesting intermediate view: that both data and theory can be helpful in forming portfolio allocations.

This paper models this intermediate view in a Bayesian setting. We consider an investor who has a prior belief about the R^2 of the predictive regression. We implement this prior by specifying a normal distribution for the regression coefficient on the predictor variable. As the variance of this normal distribution approaches zero, the prior belief becomes dogmatic that there is no predictability. As the variance approaches infinity, the prior is diffuse: all levels of predictability are equally likely. In between, the distribution implies that the

---

[1]Examples of general-equilibrium models that imply excess returns that are largely unpredictable include Abel (1990, 1999), Backus, Gregory, and Zin (1989), Campbell (1986), Cecchetti, Lam, and Mark (1993), Kandel and Stambaugh (1991) and Mehra and Prescott (1985).

[2]See, for example, Fama and Schwert (1977), Keim and Stambaugh (1986), Campbell and Shiller (1988), Fama and French (1989), Cochrane (1992), Goetzmann and Jorion (1993), Hodrick (1992), Kothari and Shanken (1997), Lettau and Ludvigson (2001), Lewellen (2004) and Ang and Bekaert (2007).

[3]See, for example, Brennan, Schwartz, and Lagnado (1997) and Campbell and Viceira (1999) for stocks and Sangvinatsos and Wachter (2005) for long-term bonds. Balduzzi and Lynch (1999) show that predictability remains important even in the presence of transaction costs, while Barberis (2000) and Xia (2001) show, respectively, that predictability remains important in the presence of estimation risk and learning. An exception is the case of buy-and-hold portfolios with horizons of many years (Barberis (2000), Cochrane (1999), Stambaugh (1999)). Brennan and Xia (2005) construct a long-run measure of expected returns and derive implications for optimal portfolios. They show that this long-run measure often implies a less extreme response to predictability than regression-based measures.

investor is skeptical about predictability: predictability is possible, but it is more likely that predictability is "small" rather than "large". By conditioning this normal distribution on both the unexplained variance of returns and on the variance of the predictor variable, we create a direct mapping from the investor's prior beliefs on model parameters to a well-defined prior over the R^2.

In our empirical implementation, we consider returns on a stock index and on a long-term bond. The predictor variables are the dividend-price ratio and the yield spread between Treasuries of different maturities. We find that the evidence is sufficient to convince an investor who is quite skeptical about predictability to vary his portfolio on the basis of these variables. The resulting weights, however, are much less volatile than for an investor who allocates his portfolio purely based on data. To see whether the skeptical prior would have been helpful in the observed time series, we implement an out-of-sample analysis. We show that weights based on skeptical priors deliver superior out-of-sample performance when compared to diffuse priors, dogmatic priors, and to a simple regression-based approach.

Our study builds on previous work that has examined predictability from a Bayesian investment perspective. Kandel and Stambaugh (1996) show that predictive relations that are weak in terms of standard statistical measures can nonetheless affect portfolio choice.[4] They conduct a simulation experiment such that predictability is present with modest significance, and examine the portfolio choices of a Bayesian investor who views the simulated data. In contrast, our study bases its inference on the historical time series of returns and predictor variables. We ask whether an investor whose priors imply skepticism about the existence of predictability would find it optimal to vary their investments in risky assets over time. Other studies make use of informative priors in a setting of return predictability. Avramov (2002) and Cremers (2002) show that Bayesian inference and informative priors can lead to superior model selection. Shanken and Tamayo (2005) jointly model time variation in risk

---

[4]Subsequently, a large literature has examined the portfolio consequences of return predictability in a Bayesian framework. Barberis (2000) considers the optimization problem of a long-horizon investor when returns are predictable. Xia (2001) considers the effect of learning about the predictive relation in a dynamic setting with hedging demands. Brandt, Goyal, Santa-Clara, and Stroud (2005) and Skoulakis (2007) extend this work to allow for uncertainty and learning about the other parameters in the predictive system. Johannes, Polson, and Stroud (2002) model the mean and volatility of returns as latent factors. Our methods build directly on those of Stambaugh (1999), who studies the impact of changes in the prior and changes in the likelihood. In contrast to the present study, these papers assume diffuse priors.

and expected return in a Bayesian setting. Shanken and Tamayo incorporate model-based intermediate views on the relation between expected return and risk. In what follows, we compare the prior beliefs we assume to those in each of these related studies.

Our study is also related to that of Poirier (1996), who calculates the prior distribution on the R^2 that is implied by the prior distributions on the regression coefficients and on the standard deviation of errors. Poirier points out that it is often easier to elicit priors on goodness-of-fit measures, like the R^2, as compared to priors on the regression coefficients. Our motivation for using the R^2 is similar. Our study differs from that of Poirier's in that we assume a time-series setting in which the regressors are not pre-determined. We also explicitly calculate the implied posteriors and develop the implications for portfolio choice.

Our use of model-based informative priors has parallels in a literature that examines the portfolio implications of the cross-section of stock returns. Motivated by the extreme weights and poor out-of-sample performance of mean-variance efficient portfolios (Best and Grauer (1991), Green and Hollifield (1992)), Black and Litterman (1992) propose using market weights as a benchmark, in effect using both data and the capital asset pricing model to form portfolios. Recently, Bayesian studies such as Pastor (2000), Avramov (2004) and Wang (2005) construct portfolios incorporating informative beliefs about cross-sectional asset pricing models.[5] Like the present study, these studies show that allowing models to influence portfolio selection can be superior to using the data alone. While these studies focus on the cross-section of returns, we apply these ideas to the time series.

The remainder of this paper is organized as follows. Section 1 describes the assumptions on the likelihood and prior, the calculation of the posterior, and the optimization problem of the investor. Section 2 applies these results to data on stock and bond returns, describes the posterior distributions, the portfolio weights, and the out-of-sample performance across different choices of priors. Section 3 concludes.

---

[5]Related approaches to improving performance of efficient portfolios include Bayesian shrinkage (Jobson and Korkie (1980), Jorion (1985)) and portfolio constraints (Frost and Savarino (1988), Jagannathan and Ma (2003)). Cvitanic, Lazrak, Martellini, and Zapatero (2006) incorporate analyst forecasts in a dynamic setting with parameter uncertainty and learning. Garlappi, Uppal, and Wang (2007) take a multi-prior approach to portfolio allocation that allows for ambiguity aversion. Tu and Zhou (2007) impose priors that ensure that portfolio weights fall into a certain range. Unlike the present study, these papers assume that the true distribution of returns is iid and focus on the cross-section.

1 Portfolio choice for a skeptical investor

Given observations on returns and a predictor variable, how should an investor allocate his wealth? One approach would be to estimate the predictability relation, treat the point estimates as known, and solve for the portfolio that maximizes utility. An alternative approach, adopted in Bayesian studies, is to specify prior beliefs on the parameters. The prior represents the investor's beliefs about the parameters before viewing data. After viewing data, the prior is updated to form a posterior distribution; the parameters are then integrated out to form a predictive distribution for returns, and utility is maximized with respect to this distribution. This approach incorporates the uncertainty inherent in estimation into the decision problem (see Klein and Bawa (1976), Bawa, Brown, and Klein (1979), Brown (1979)).[6] Rather than assuming that the investor knows the parameters, it assumes, realistically, that the investor estimates the parameters from the data. Moreover, this approach allows for prior information, perhaps motivated by economic models, to enter into the decision process. This section describes the specifics of the likelihood function, the prior, and the posterior used in this study.

1.1 Likelihood

This subsection constructs the likelihood function. Let r_{t+1} denote an $N \times 1$ vector of returns on risky assets in excess of a riskless asset from time t to $t+1$, and x_t a scalar predictor variable at time t. The investor observes data on returns $r_1, \ldots, r_T$, and the predictor variable $x_0, \ldots, x_T$. Let

$$D \equiv \{r_1, \ldots, r_T, x_0, x_1, \ldots, x_T\}$$

represent the total data available to the investor. Our initial assumption is that there is a single predictor variable that has the potential to predict returns on (possibly) multiple assets. Allowing multiple predictor variables complicates the problem without contributing to the intuition. For this reason, we discuss the case of multiple predictor variables in Appendix A.

---

[6]Like these papers and like the portfolio choice papers cited in the introduction, this paper studies an investor who should not be viewed as representative. By definition, the representative investor must hold the market portfolio.

The data generating process is assumed to be

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

$$x_{t+1} = \theta_0 + \theta_1 x_t + v_{t+1}, \tag{2}$$

where

$$\begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix} \mid r_t, \ldots, r_1, x_t, \ldots, x_0 \sim N\left(0, \Sigma\right), \tag{3}$$

α and β are $N \times 1$ vectors and Σ is an $(N+1) \times (N+1)$ symmetric and positive definite matrix.[7] It is useful to partition Σ so that

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

where Σ_u is the variance-covariance matrix of u_{t+1}, $\sigma_v^2 = \Sigma_v$ is the variance of v_{t+1}, Σ_{uv} is the $N \times 1$ vector of covariances of v_{t+1} with each element of u_{t+1}, and $\Sigma_{vu} = \Sigma_{uv}^\top$. This likelihood is a multi-asset analogue of that assumed by Kandel and Stambaugh (1996), Campbell and Viceira (1999), and many subsequent studies.

It is helpful to group the regression parameters in (1) and (2) into a matrix:

$$B = \begin{bmatrix} \alpha^\top & \theta_0 \\ \beta^\top & \theta_1 \end{bmatrix},$$

and to define matrices of the observations on the the left hand side and right hand side variables:

$$Y = \begin{bmatrix} r_1^\top & x_1 \\ \vdots & \vdots \\ r_T^\top & x_T \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_0 \\ \vdots & \vdots \\ 1 & x_{T-1} \end{bmatrix}.$$

As shown in Barberis (2000) and Kandel and Stambaugh (1996), the likelihood conditional on the first observation takes the same form as in a regression model with non-stochastic regressors. Let $p(D|B, \Sigma, x_0)$ denote the likelihood function. From results in Zellner (1996), it follows that

$$p(D|B, \Sigma, x_0) = |2\pi\Sigma|^{-\frac{T}{2}} \exp\left\{ -\frac{1}{2} \mathrm{tr}\left[(Y - XB)^\top (Y - XB)\Sigma^{-1} \right] \right\}, \tag{4}$$

---

[7]Results assuming a multivariate t-distribution are similar to those reported below and available from the authors upon request.

where $\text{tr}(\cdot)$ denotes the sum of the diagonal elements of a matrix.[8]

The likelihood function (4) conditions on the first observation of the predictor variable, x_0. In contrast, observations $1, \ldots, T$ are treated as draws from the data-generating process. An alternative, implemented in a return-predictability setting by Stambaugh (1999), is to treat x_0 as a draw from the data-generating process as well. The resulting likelihood function fully incorporates the information contained in x_0, while (4) does not. It may at first seem that this choice should make little difference, since only one observation is involved. However, Poirier (1978) shows that the consequences can be quite substantial because the first observation is transformed in a different way than the remaining observations.

In constructing the likelihood that does not condition on x_0, we assume that the process for x_t is stationary and has run for a substantial period of time. Results in Hamilton (1994, p. 53) imply that x_0 is a draw from a normal distribution with mean

$$\mu_x \equiv E\left[x_t \mid B, \Sigma\right] = \frac{\theta_0}{1 - \theta_1} \tag{5}$$

and variance

$$\sigma_x^2 \equiv E\left[(x_t - \mu_x)^2 \mid B, \Sigma\right] = \frac{\sigma_v^2}{1 - \theta_1^2}. \tag{6}$$

Combining the likelihood of the first observation with the likelihood of the remaining T observations produces

$$
\begin{aligned}
p(D|B,\Sigma) \ &= \ p(D|x_0, B, \Sigma)p(x_0|B,\Sigma) \\
&= \ \left(2\pi\sigma_x^2\right)^{-\frac{1}{2}} |2\pi\Sigma|^{-\frac{T}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2}\sigma_x^{-2}\left(x_0 - \mu_x\right)^2 - \frac{1}{2}\text{tr}\left[(Y - XB)^\top (Y - XB)\Sigma^{-1}\right]\right\}. \tag{7}
\end{aligned}
$$

Equation (7) is the likelihood function used in our analysis. Following Box, Jenkins, and Reinsel (1970), we refer to (7) as the exact likelihood, and to (4) as the conditional likelihood.

---

[8]Maximizing the conditional likelihood function (4) implies estimates of β that are the same as those obtained by ordinary least squares regression. These estimates are biased (see Bekaert, Hodrick, and Marshall (1997), Nelson and Kim (1993) and Stambaugh (1999)), and standard asymptotics provide a poor approximation to the distribution of test statistics in small samples (Cavanagh, Elliott, and Stock (1995), Elliott and Stock (1994), Mankiw and Shapiro (1986), Richardson and Stock (1989)). An active literature based in classical statistics focuses on correcting for these problems (e.g. Amihud and Hurvich (2004), Campbell and Yogo (2006), Eliasz (2004), Ferson, Sarkissian, and Simin (2003), Lewellen (2004), Torous, Valkanov, and Yan (2004)).

1.2 Prior beliefs

This subsection describes the prior. We specify a class of prior distributions that range from being "uninformative" in a sense we will make precise, to "dogmatic". The uninformative priors imply that all amounts of predictability are equally likely, while the dogmatic priors rule out predictability all together; the investor believes that returns are not predictable regardless of what data are observed. Between these extremes lie priors that downweight the return predictability. These informative priors imply that large values of the R^2 from predictive regressions are unlikely, but not impossible.

Before discussing the specifics of our informative priors, we briefly discuss how asset-pricing theory guides us toward these priors. Models with constant risk aversion, rational agents, and homoskedastic endowments (e.g. Abel (1990, 1999), Backus, Gregory, and Zin (1989), Barro (2006) and the benchmark case in Bansal and Yaron (2004)) imply that the R^2 in the predictability regression is exactly zero.[9] Recently, several papers propose general equilibrium models that are capable of generating time-varying expected excess returns (risk premia). Models that successfully account for time-varying risk premia have done so through one of several mechanisms: time-varying risk aversion (e.g. Campbell and Cochrane (1999)), non-rational investors (e.g. Barberis, Huang, and Santos (2001)), or time-varying volatility in the endowment process (e.g. Whitelaw (2000), Bansal and Yaron (2004)). An investor whose prior beliefs are that constant risk aversion is more likely than time-varying risk aversion, that investors are more likely to be rational and that the consumption process is more likely to be homoskedastic would therefore be led to a prior that favors a low R^2 for the predictability regression.[10]

---

[9]Models with regime shifts in the endowment process, such as Cecchetti, Lam, and Mark (1993), Kandel and Stambaugh (1991), Mehra and Prescott (1985), Reitz (1988) imply a small amount of heteroskedasticity, and therefore of excess return predictability. The amount of excess return predictability generated through the regime-shift mechanism is very small, however. Allowing for this small amount of predictability in the prior would greatly complicate the analysis and is unlikely to affect the results.

[10]Of these possibilities, it is perhaps most difficult to conceptualize a prior over endowment heteroskedasticity. This issue is complicated by the fact that a model with endowment heteroskedasticity will not necessarily imply that return heteroskedascity takes a simple form, and that return variance and risk premia are related in a straightforward way. In a study that we discuss further below, Shanken and Tamayo (2005) take a reduced-form approach and assume priors that favor a linear relation between return variance and risk premia. They find that there is significant variation in expected excess returns that does not correspond

We now discuss the specific form assumed for the investor's prior beliefs. The most obvious parameter that determines the degree of predictability is β. Set β to zero, and there is no predictability in the model. However, it is difficult to think of prior beliefs about β in isolation from beliefs about other parameters. For example, a high variance of x_t might lower one's prior on β, while a large residual variance of r_t might raise it. Rather than placing a prior on β directly, we instead place a prior on "normalized" β, that is β adjusted for the variance of x and the variance of u. As we show below, this is equivalent to placing a prior on the population equivalent of the centered, unadjusted R^2. Let C_u be the Cholesky decomposition of Σ_u, i.e. $C_u C_u^\top = \Sigma_u$. Then

$$\eta = C_u^{-1} \sigma_x \beta$$

is normalized β. We assume that prior beliefs on η are given by

$$\eta \sim N(0, \sigma_\eta^2 I_N), \tag{8}$$

where I_N is the $N \times N$ identity matrix.[11]

We implement the prior for η by specifying a hierarchical prior for the primitive parameters. That is, the prior for β is conditional on the remaining parameters:

$$p(B, \Sigma) = p(\beta | \alpha, \theta_0, \theta_1, \Sigma) p(\alpha, \theta_0, \theta_1, \Sigma). \tag{9}$$

Then the specification for the distribution for η, (8), is equivalent to the following specification for the distribution of β:

$$\beta | \alpha, \theta_0, \theta_1, \Sigma \sim N(0, \sigma_\eta^2 \sigma_x^{-2} \Sigma_u). \tag{10}$$

Because σ_x is a function of θ_1 and σ_v, the prior on β is also implicitly a function of these parameters.

For the remaining parameters, we choose a prior that is uninformative in the sense of

---

to variation in volatility, a conclusion also reached in many frequentist studies (see Campbell (2003) for a survey). We hope to investigate the important issue of heteroskedasticity in future research; however, given the results of existing studies we expect our conclusions to be unaffected.

[11]Formally, we can consider a prior on the parameters $p(\alpha, \eta, \theta_0, \theta_1, \Sigma) = p(\eta) p(\alpha, \theta_0, \theta_1, \Sigma)$, where $p(\eta)$ is defined by (8).

Jeffreys (1961).[12] We follow the approach of Stambaugh (1999) and Zellner (1996), and derive a limiting Jeffreys prior as explained in Appendix D. This limiting prior takes the form

$$p(\alpha, \theta_0, \theta_1, \Sigma) \propto \sigma_x |\Sigma_u|^{1/2} |\Sigma|^{-\frac{N+4}{2}}, \tag{11}$$

for $\theta_1 \in (-1, 1)$, and zero otherwise. Therefore the joint prior is given by

$$\begin{aligned} p(B, \Sigma) &= p(\beta | \alpha, \theta_0, \theta_1, \Sigma) p(\alpha, \theta_0, \theta_1, \Sigma) \\ &\propto \sigma_x^{N+1} |\Sigma|^{-\frac{N+4}{2}} \exp\left\{ -\frac{1}{2} \beta^\top \left(\sigma_\eta^2 \sigma_x^{-2} \Sigma_u \right)^{-1} \beta \right\}. \end{aligned} \tag{12}$$

Note that in (11) and (12), σ_x is a nonlinear function of the autoregressive coefficient θ_1 and volatility of the shock to the predictor variable σ_v.

The appeal of linking the prior distribution of β to the distribution of Σ_u and σ_x is that it implies a well-defined distribution on the population R^2. Consider for simplicity the case of a single risky asset. In population, the ratio of the variance of the predictable component of return to the total variance is equal to

$$R^2 = \beta^2 \sigma_x^2 \left(\beta^2 \sigma_x^2 + \Sigma_u \right)^{-1} = \frac{\eta^2}{\eta^2 + 1}. \tag{13}$$

(note that Σ_u is a scalar). Equation (13) is the population equivalent of the centered, unadjusted R^2. In what follows, we will refer to (13) simply as the R^2. Because the R^2 is a function of η alone, specifying a prior on η, and therefore specifying the joint prior (12), implies a well-defined prior distribution on the R^2.

When there are N risky assets, there is a natural extension of (13). Besides implying a distribution on the R^2 of each asset, the distribution of η, (8) implies a joint distribution such that no linear combination of asset returns can have an R^2 that is "too large". To be

---

[12]The notion of an uninformative prior in a time-series setting is a matter of debate. One approach is to ignore the time-series aspect of (1) and (2), treating the right hand side variable as exogenous. This implies a flat prior for α, β, θ_0, and θ_1. When applied in a setting with exogenous regressors, this approach leads to Bayesian inference which is quite similar to classical inference (Zellner (1996)). However, Sims and Uhlig (1991) show that applying the resulting priors in a time series setting leads to different inference than classical procedures when x_t is highly persistent. As a full investigation of these issues is outside the scope of this study, we focus on the Jeffreys prior. Replacing the prior in (11) with one that is implied by exogenous regressors gives results that are similar to our current ones; these are available from the authors.

more precise, let w be an $N \times 1$ vector. Then calculations in Appendix B show that

$$\max_w R^2 = \max_w \frac{w^\top \beta \beta^\top \sigma_x^2 w}{w^\top \beta \beta^\top w \sigma_x^2 + w^\top \Sigma_u w}$$
$$= \frac{\eta^\top \eta}{\eta^\top \eta + 1}. \tag{14}$$

The return on the risky part of an investor's portfolio is a linear combination of the returns on the risky assets; therefore the risky part of any portfolio chosen by the investor must have an R^2 that lies below (14). An alternative strategy would be to link the prior to the R^2 of the optimal portfolio of the investor. However, this optimal portfolio depends not only on the history of the data, it also depends on the particular value of the predictor variable. Both are unknown to the investor when forming the prior. In contrast, placing a prior on the R^2, or on the maximum R^2 of a system of equations, has simple, intuitive appeal.[13]

Figure 1 depicts the distribution of the R^2 implied by these prior beliefs. The figure shows the probability that the R^2 exceeds some value k, $P(R^2 > k)$, as a function of k; it is therefore one minus the cumulative distribution function for the R^2. We construct this figure by simulating draws of η from (8) and, for each draw, constructing a draw from the R^2 distribution using (13). As this figure shows, the parameter σ_η indexes the degree to which the prior is informative. For $\sigma_\eta = 0$, the investor assigns zero probability to a positive R^2; for this reason $P(R^2 > k)$ is equal to one at zero and is zero elsewhere. As σ_η increases, the investor assigns non-zero probability to positive values of the R^2. For $\sigma_\eta = .04$, the probability that the R^2 exceeds .02 is .0005. For $\sigma_\eta = .08$, the probability that the R^2 exceeds .02 is .075. Finally when σ_η is large, approximately equal probabilities are assigned to all values of the R^2. This is the diffuse prior that assigns no skepticism to the data. In what follows, we will consider the implications of these four priors for the individual's investment decisions. We will refer to them by using the corresponding probabilities that the R^2 exceeds .02. We note, however, that we focus on .02 for convenience; any number

---

[13]This prior distribution could easily be modified to impose other restrictions on the coefficients β. In the context of predicting equity returns, Campbell and Thompson (2007) suggest disregarding estimates of β if the expected excess return is negative, or if β has an opposite sign to that suggested by theory. In our model, these restrictions could be imposed by assigning zero prior weight to the appropriate regions of the parameter space. One could also consider a non-zero mean for β, corresponding to a prior belief that favors predictability of a particular sign. For simplicity, we focus on priors that apply to any predictor variable on possibly multiple assets, and leave these extensions to future work.

between 0 and 1 could be substituted.

Finally, we note that Jeffreys invariance theory provides an independent justification for modeling priors on β as in (10). Appendix C shows that the limiting Jeffreys prior for B and Σ equals

$$p(B, \Sigma) \propto \sigma_x^{N+1} |\Sigma|^{-\frac{N+4}{2}} . \tag{15}$$

This prior corresponds to (12) as σ_η approaches infinity. Modeling the prior for β as depending on σ_x not only has an interpretation in terms of the R^2, but also implies that an infinite prior variance represents ignorance as defined by Jeffreys (1961). Note that a prior on β that is independent of σ_x would not have this property. Because the priors in (12) combine an informative ("skeptical") prior on β with a Jeffreys prior on the remaining parameters, we refer to these as *skeptical Jeffreys* priors.

Comparison with related studies

In this section we have described one way of modeling prior information. We now compare this approach to that used in some important Bayesian return predictability studies that make use of informative priors.[14]

Kandel and Stambaugh (1996) derive posteriors assuming the investor has seen, in addition to the actual data, a "prior" sample of the data that has unconditional moments equal to those of the actual sample except without predictability.[15] These priors are also assumed by Avramov (2002, 2004). In a recent working paper, Pastor and Stambaugh (2006) focus on informative priors about the correlation between u_{t+1} and v_{t+1}; however, they also assume an informative prior on the R^2 using the sample variance of stock returns. Cremers (2002) specifies informative prior beliefs motivated by placing a distribution on the expected sample R^2. Cremers's priors assume knowledge of sample moments of the predictor variable. Specifically, if we let X denote a $T \times k$ matrix of values for k predictor variables, Cremers assumes that the investor knows $X^\top X$.

The various priors specified in the above studies all contain the same flaw: they require

---

[14]Goyal and Welch (2007) present an "encompassing forecast", which, while not Bayesian, has similar implications in that it downweights the predictability coefficient estimated from the data. Bayesian methods can be seen as formalizing this approach.

[15]Kandel and Stambaugh discuss the appeal of holding the distribution of the R^2 constant. To accomplish this in their set-up, they let the length of the prior sample increase in the number of predictor variables.

that the investor has knowledge of the sample moments when forming a prior. It is not enough for the investor to make a reasonable guess as to the value of the moments. For their analyses to be correct, the investor must know the value of the moment precisely. From a Bayesian perspective, such knowledge is problematic. To know the moments of the data, either the investor must have seen the data (but then the prior and the posterior would be identical), or somehow intuited the correct moments, without seeing the data. Even if we accept this awkward latter interpretation, to be consistent these moments would have to be treated as constants (namely conditioned on) throughout the analysis, which they are not.[16]

In contrast, we construct a prior that does not require conditioning on the sample moments, but is nonetheless based on a measure over which investors have intuition. Rather than improperly conditioning on sample moments, we do this by forming a prior over the population R^2 itself. We note that the stochastic nature of the regressor is at the core of the problem, and thus of our contribution. If the regressors were non-stochastic, as in standard ordinary least squares (but rarely in predictive regressions), conditioning on $X^\top X$ in the prior would be valid.

Another important study that is related to ours is that of Shanken and Tamayo (2005).[17] Shanken and Tamayo model time-variation in risk as well as in expected returns. Like our priors, the priors in Shanken and Tamayo represent a model-based view that is intermediate between complete faith in a model and complete faith in the data. However, their formulation of priors is less parsimonious, requiring ten parameters in the case of a single asset (a broad stock market portfolio) and predictor variable (the dividend-yield). The prior values are specific to these variables and do not transfer easily to other assets or new predictor variables. The advantage of our method is that it expresses the informativeness of the agent's prior beliefs as a single number which can be mapped into beliefs about the maximum R^2. This is the case regardless of the number of risky assets or the number and characteristics of the

---

[16]Data-based procedures for forming priors are often referred to as "empirical Bayes". However, at least in its classic applications, empirical Bayes implies either the the use of data that is known prior to the decision problem at hand or data from the population from which the parameter of interest can be drawn (Robbins (1964), Berger (1985)). For example, if one is forming a prior on a expected return for a particular security, one might use the average expected return of securities for that industry (Pastor and Stambaugh (1999)).

[17]Technically, the priors in Shanken and Tamayo (2005) suffer from the same difficulty as the ones mentioned in the previous paragraph, as they demean their variables and divide by the standard deviation. This requires that the investor knows the sample mean and variance of these variables when forming a prior.

predictor variables.

Our priors are in fact reminiscent of the choice of prior on the intercepts in cross-sectional studies. Pastor and Stambaugh (1999) and Pastor (2000) place an informative prior on the vector of intercepts from regressions of returns on factors in the cross-section. Building on ideas of MacKinlay (1995), these studies argue that failure to condition the intercepts on the residual variance could lead to very high Sharpe ratios, because there would be nothing to prevent a low residual variance draw from occurring simultaneously with a high intercept draw. Bayesian portfolio choice studies (Baks, Metrick, and Wachter (2001), Jones and Shanken (2005), Pastor and Stambaugh (2002)) place an informative prior on estimates of mutual fund skill (intercepts from regressions of returns on factors), and argue based on related ideas that this informative prior should be conditioned on the residual variance of the fund. In the present study, β plays a role that is roughly analogous to the intercept in these previous studies. $\beta = 0$ implies no predictability, and hence no "mispricing". As in these previous studies, conditioning β on volatility measures ensures that a high draw of β could not coincide with a low draw of Σ_u. However, in the time-series setting, it is not sufficient to condition β on Σ_u; β must also be conditioned on σ_x in order to produce a well-defined distribution for quantities of interest.

1.3 Posterior

This section shows how the likelihood of Section 1.1 and the prior of Section 1.2 combine to form the posterior distribution. From Bayes' rule, it follows that the joint posterior for B, Σ is given by

$$p(B, \Sigma | D) \propto p(D | B, \Sigma) p(B, \Sigma),$$

where $p(D | B, \Sigma)$ is the likelihood and $p(B, \Sigma)$ is the prior. Substituting in the prior (12) and the likelihood (7) produces

$$p(B, \Sigma | D) \propto \sigma_x^N |\Sigma|^{-\frac{T+N+4}{2}} \exp\left\{ -\frac{1}{2} \beta^\top \left(\sigma_\eta^2 \sigma_x^{-2} \Sigma_u \right)^{-2} \beta \right\} \exp\left\{ -\frac{1}{2} \sigma_x^{-2} (x_0 - \mu_x)^2 \right\}$$

$$\exp\left\{ -\frac{1}{2} \text{tr} \left[(Y - XB)^\top (Y - XB) \Sigma^{-1} \right] \right\} \quad (16)$$

as a posterior.

This posterior does not take the form of a standard density function because of the presence of σ_x^2 in the prior and in the term in the likelihood involving x_0 (note that σ_x^2 is

13

a nonlinear function of θ_1 and σ_v). However, we can sample from the posterior using the Metropolis-Hastings algorithm (see Chib and Greenberg (1995)). Define column vectors

$$
\begin{aligned}
b &= \operatorname{vec}(B) = [\alpha_1, \ \beta_1, \ \cdots \ \alpha_N, \ \beta_N, \ \theta_0, \ \theta_1]^\top \\
b_1 &= [\alpha_1, \ \beta_1, \ \cdots \ \alpha_N, \ \beta_N]^\top \\
b_2 &= [\theta_0, \ \theta_1]^\top .
\end{aligned}
$$

The Metropolis-Hastings algorithm is implemented "block-at-a-time", by first sampling from $p(\Sigma|b, D)$, then $p(b_1|b_2, \Sigma, D)$, and finally $p(b_2|b_1, \Sigma, D)$. The proposal density for the conditional probability of Σ is the inverted Wishart with $T + 2$ degrees of freedom and scale factor of $(Y - XB)^\top (Y - XB)$. The accept-reject algorithm of Chib and Greenberg (1995, Section 5) is used to sample from the target density, which takes the same form as (16). The proposal densities for b_1 and b_2 are multivariate normal. For b_1, the proposal and the target are equivalent, while for b_2, the accept-reject algorithm is used to sample from the target density. Details are given in Appendix E. As described in Chib and Greenberg, drawing successively from the conditional posteriors for Σ, b_1, and b_2 produces a density that converges to the full posterior.

1.4   Predictive distribution and portfolio choice

This section describes how we determine optimal portfolio choice based on the posterior distribution. Consider an investor who maximizes expected utility at time $T + 1$ conditional on information available at time T. The investor solves

$$
\max E_T[U(W_{T+1})|D] \tag{17}
$$

where $W_{T+1} = W_T[w_T^\top r_{T+1} + r_{f,T}]$, w_T are the weights in the N risky assets, and $r_{f,T}$ is the total return on the riskless asset from time T to $T + 1$ (recall that r_{T+1} is a vector of excess returns). The expectation in (17) is taken with respect to the predictive distribution

$$
p(r_{T+1}|D) = \int p(r_{T+1}|x_T, B, \Sigma) p(B, \Sigma|D)\, dB\, d\Sigma. \tag{18}
$$

Following previous single-period portfolio choice studies (see, e.g. Baks, Metrick, and Wachter (2001) and Pastor (2000)), we assume that the investor has quadratic utility. The advantage of quadratic utility is that it implies a straightforward mapping between the

moments of the predictive distribution of returns and portfolio choice. However, because our method produces an entire distribution function for returns, it can be applied to other utility functions, and to buy-and-hold investors with horizons longer than one quarter.

Let $\tilde{E}$ denote the expectation and $\tilde{V}$ the variance-covariance matrix of the N assets corresponding to the predictive distribution (18). For a quadratic-utility investor, optimal weights w^* in the N assets are given by

$$w^* = \frac{1}{A}\tilde{V}^{-1}\tilde{E}, \tag{19}$$

where A is a parameter determining the investor's risk aversion. The weight in the riskless bond is equal to $1 - \sum_{i=1}^{N} w_i^*$. Note that our method makes no assumptions on the riskfree rate as the level of the riskfree rate is not relevant for portfolio choice under our assumptions.

Given draws from the posterior distribution of the parameters $\alpha^j, \beta^j, \Sigma_u^j$, and a value of x_t, a draw from the predictive distribution of asset returns is given by

$$r^j = \alpha^j + \beta^j x_t + u^j,$$

where $u^j \sim N(0, \Sigma_u^j)$. The optimal portfolio is then the solution to (19), with the mean and variance computed by simulating draws r^j.

2   Results

We consider the problem of a quadratic utility investor who allocates wealth between a riskless asset, a long-term bond, and a stock index. We estimate two versions of the system given in (1)–(3), one with the dividend-price ratio as the predictor variable and one with the yield spread. An appeal of these variables is that they are related to excess returns through present value identities for bonds and stocks (see Campbell and Shiller (1988, 1991)).

2.1   Data

All data are obtained from the Center for Research on Security Prices (CRSP). Excess stock and bond returns are formed by subtracting the quarterly return on the three-month Treasury bond from the quarterly return on the value-weighted NYSE-AMEX-NASDAQ index and the ten-year Treasury bond (from the CRSP indices file) respectively. The dividend-price ratio

is constructed from monthly return data on the stock index as the sum of the previous twelve months of dividends divided by the current price. The natural logarithm of the dividend-price ratio is used as the predictor variable. The yield spread is equal to the continuously compounded yield on the zero-coupon five year bond (from the Fama-Bliss data set) less the continuously compounded yield on the three-month bond. Data on bond yields are available from the second quarter of 1952. We therefore consider quarterly observations from the second quarter of 1952 until the last quarter of 2004.

2.2 Posterior means, expected returns and portfolios conditional on the full sample

This section quantitatively describes the posterior beliefs of an investor who views the entire data set. For both predictor variables, one million draws from the posterior distribution are simulated as described in Section 1.3. An initial 100,000 "burn-in" draws are discarded.

We first examine the posterior distribution over the maximum R^2. We focus on this statistic because it completely summarizes the prior distribution. It is therefore a good place to start when comparing the posterior and the prior. Recall that we are using the term R^2 to describe the population analogue of the usual sample statistic. Just as we can construct a posterior distribution over β, α, and over the volatility parameters, it is possible to construct a posterior distribution over the maximum R^2, which is a function of these primitive parameters.

Because our empirical implementation assumes two assets, we compute the prior distribution of (14) assuming $N = 2$. The left panel of Figure 2 reports the probability that the maximum R^2 exceeds k, as a function of k for both the prior with $\sigma_\eta = .08$ (dotted line) and for the posterior distribution given this prior. The solid line gives the posterior assuming that the dividend-price ratio is the predictor variable, while the dashed line gives the posterior assuming that the yield spread is the predictor variable. The right panel of Figure 2 shows the probability density function of the posterior and of the prior.

We first discuss the result for the dividend-price ratio and then for the yield spread. Below $k = .02$, the posterior probability that the R^2 exceeds k is above the prior probability. Above .02, the posterior probability that the R^2 exceeds k is lower for the posterior than for the prior. While the prior density is decreasing in the R^2 over this range, the posterior

density is hump-shaped with a maximum at about 0.02.

For the yield spread, the agent places more weight on relatively high values of the R^2 as compared with results for the dividend yield. The probability that the R^2 exceeds k is larger for the posterior than the prior across the entire range that we consider. The posterior distribution for the R^2 still peaks at about 0.02, but falls off less quickly than in the case of the dividend yield.

Table 1 reports posterior means for values of σ_η equal to 0, .04, .08, and ∞. To emphasize the economic significance of these priors, we report the corresponding probabilities that the R^2 exceeds .02: 0, .0005, .075, and .999.[18] The predictor variable is the dividend-price ratio. Posterior standard deviations are reported in parentheses. The table also shows results from estimation by ordinary least squares (OLS). For the OLS values, standard errors are reported in parentheses.

As Panel A of Table 1 shows, the dividend-price ratio predicts stock returns but not bond returns. The posterior mean for the β for bond returns is negative and small in magnitude. The posterior mean for the β for stock returns is positive for all of the priors we consider and for the OLS estimate. For the diffuse prior, the posterior mean of β equals to 1.46, below the OLS estimate of 2.72. As the prior becomes more informative, the posterior mean for β becomes smaller: for $P(R^2 > .02) = .075$, the estimate is 1.41, while for $P(R^2 > .02) = .0005$, it is 0.69.

Even though all of the priors are uninformative with respect to the autoregressive coefficient θ_1, the posterior mean of θ_1 nonetheless increases as the priors become more informative over β. The reason is the negative correlation between draws for θ_1 and draws for β. As Stambaugh (1999) shows, the negative correlation between shocks to returns and shocks to the predictor variable implies that when β is below its OLS value, θ_1 tends to be above its OLS value. The reason is that if β is below its OLS value, it must be that the lagged predictor variable and returns have an unusually high covariance in the sample (because the OLS value is "too high"). When this occurs, the predictor variable tends to have an unusually low autocorrelation; thus the OLS estimate for θ_1 is too low and the posterior mean will be above the OLS value. Therefore, placing a prior that weights the posterior mean of β toward zero raises the posterior mean of θ_1.[19]

---

[18]These values are the marginal probability that the R^2 for a single equation exceeds 0.02.

[19]The value for β under the diffuse prior is also substantially below the OLS value, and the value for θ_1 is

Table 1 also reports posterior means and standard deviations for the means of the predictor variable and of stock returns. For example, for returns the table reports $E\left[E[r_{t+1}|B,\Sigma]|D\right] = E\left[\alpha + \beta\frac{\theta_0}{1-\theta_1}|D\right]$ and $\left(\text{Var}\left[\alpha + \beta\frac{\theta_0}{1-\theta_1}|D\right]\right)^{1/2}$. The OLS mean is set equal to $\hat{\alpha} + \hat{\beta}\frac{\hat{\theta}_0}{1-\hat{\theta}_1}$, where $\hat{}$ denotes the OLS estimate of a parameter. The unconditional means are of interest because they help determine the average level of the portfolio allocation. The Bayesian approach implies about the same unconditional mean for the dividend-price ratio, regardless of the prior. This is close to -3.50, the mean in the data. However, the mean implied by OLS is -3.72. The reason that the Bayesian approach is able to identify this mean is the presence of the unconditional distribution term in the likelihood.

These differences in the mean of x translate into differences in the unconditional means for returns. Table 1 shows that for the stock index, the posterior mean equals 1.17%, while the OLS value is 1.09% per quarter. The sample mean for stocks in this time period was 1.67%. The difference between the OLS and the sample mean arises mechanically from the difference between $\frac{\hat{\theta}_0}{1-\hat{\theta}_1}$ (equal to -3.72), and the sample mean of the dividend-price ratio (equal to -3.50). The difference between the sample and the Bayesian posterior mean occurs for a more subtle reason. Because the dividend-price ratio in 1952 is above its conditional maximum likelihood estimate (-3.72), it follows that shocks to the dividend-price ratio were negative on average during the time period. Because of the negative correlation between the stock return and the dividend-price ratio, shocks to stock returns must be positive on average. The exact likelihood function therefore implies a posterior mean that is below the sample mean. Similar reasoning holds for bond returns, though here, the effect is much smaller because of the low correlation between the dividend-price ratio and bond returns. This effect is not connected with the ability of the dividend-price ratio to predict returns, as it operates equally for all values of the prior.

Panel B Table 1 reports analogous results for the yield spread. The yield spread predicts both bond and stock returns with a positive sign. As the prior becomes more diffuse, the posterior mean of the β coefficients go from from 0 to the OLS estimate. As in the case of the dividend-price ratio, both the posterior mean of long-run expected returns and the long-run mean of x_t are nearly the same across the range of prior distributions.

We now examine the consequences of these posterior means for the predictive distribution

---

higher. The reason is that, relative to the flat prior, the Jeffreys prior favors higher values of θ_1.

of returns and for portfolio choice. Figure 3 plots expected excess returns (top two plots) and optimal portfolio holdings (bottom two plots) as functions of the log dividend-price ratio. Graphs are centered at the sample mean. Diamonds denote plus and minus one and two sample standard deviations of the dividend-price ratio. We report results for the four prior beliefs discussed above; to save notation we let $P_{.02} = P(R^2 > .02)$.

The linear form of (1) implies that expected returns are linear in the predictor variables, conditional on the past data. The slope of the relation between the conditional return and x_t equals the posterior mean of β. Figure 3 shows large deviations in the expected return on the stock on the basis of the dividend-price ratio. As the dividend-price ratio varies from -2 standard deviations to +2 standard deviations, the expected return varies from 0% per quarter to 2% per quarter. On the other hand, the dividend-price ratio has virtually no predictive power for returns on the long-term bond.

The bottom panel of Figure 3 shows that the weight on the stock index also increases in the dividend-price ratio. Bond weights decrease in the dividend-price ratio because bond and stock returns are positively correlated, so an increase in the mean of the stock return, without a corresponding increase in the bond return, results in an optimal portfolio that puts less weight on the bond.

For the diffuse prior, weights on the stock index vary substantially, from -30% when the dividend-price ratio is two standard deviations below its mean to 100% when the dividend-price ratio is two standard deviations above its mean. As the prior becomes more informative, expected returns and weights both vary less. However, this change happens quite slowly. Conditional expected returns under a prior that assigns only a .075 chance of an R^2 greater than 2% are nearly identical to conditional expected returns with a diffuse prior. There is sufficient evidence to convince even this skeptical investor to vary her portfolio to nearly the same degree as an investor with no skepticism at all. For a more skeptical prior with $P_{.02} = .0005$, differences emerge: the slope of the relation between expected returns and the dividend-price ratio is about half of what it was with a diffuse prior.

Figure 4 displays analogous plots for the yield spread. Both the conditional expected bond return and the stock return increase substantially in the yield spread. For bonds, these expected returns vary between -2% and 2% per quarter as the yield spread varies between -2 and +2 standard deviations. For stocks, expected returns vary between 0% and 3%, similar to the variation with respect to the dividend-price ratio. These large variations in

expected returns lead to similarly large variation in weights for the diffuse prior: for bonds, the weights vary between -200% and 200% as the yield spread varies between -2 and +2 standard deviations from the mean. For the stock, the weights vary between 0 and 75%. The variation in the weights on the stock appears less than the variation in expected returns on the stock; this is due to the positive correlation in return innovations between stocks and bonds.

Figure 4 also shows that the more informative the prior, the less variable the weights. However, when the predictor variable is the yield spread, inference based on a skeptical prior with $P_{.02} = .075\%$ differs noticeably from inference based on a diffuse prior. Nonetheless, even the investors with skeptical priors choose portfolios that vary with the yield spread.

This section has shown that an investor who is skeptical about predictability, when confronted with historical data, does indeed choose to time the market. The next two sections show the consequences of this for the time series of portfolio weights and for out-of-sample performance.

2.3   Posterior means and asset allocation over the post-war period

We next describe the implications of various prior beliefs for optimal weights over the post-war period. Starting in 1972, we compute the posterior distribution conditional on having observed data up to and including that year. We start in 1972 because this allows for twenty years of data for the first observation; this seems reasonable given the persistence in the data and the fact that there are 12 parameters to be estimated. Starting the analysis in 1982 with thirty years of data leads to very similar results.[20] The posterior is computed by simulating 200,000 draws and dropping the first 50,000. Each quarter, the investor updates the portfolio weights using that quarter's observation of the predictor variable. The assets are the stock, the long-term bond, and a riskless asset.

Panel A of Figure 5 displays the de-meaned dividend-price ratio and the weights in the long-term bond and the stock for the most diffuse prior. On average, the investor allocates a positive fraction of wealth to the stock, a negative fraction to the long-term bond, and

---

[20]We update the posterior distribution yearly rather than quarterly to save on computation time, which is a particular concern when we assess the significance of our results using a Monte Carlo procedure. The results from updating quarterly are very close to those from updating yearly and are available from the authors upon request.

a positive fraction to the riskfree asset. The shorting of the long-term bond arises because of its relatively poor risk-return tradeoff (as compared with the stock), as well as from the positive correlation of bond and stock returns. This latter property makes bonds a good hedge against risks in the more attractive stock portfolio. Campbell, Chan, and Viceira (2003) study allocation to stocks, long-term bonds and a riskfree asset in a non-Bayesian setting that is otherwise similar to ours. They also find persistent shorting of the long-term bond.

Figure 5 also shows that the weights in the stock co-move with the dividend-price ratio. As discussed above, the higher is the dividend-price ratio, the higher the expected return on the stock, and the higher the portfolio weight. Less correlation is apparent for bond returns, due to the fact that investors hold bonds in part to hedge the risk from holding equities. From the mid-90's, on, this correlation is reduced for both assets: despite the continued decline in the dividend-price ratio, the allocation to stocks levels off and the allocation to bonds rises. As Figure 5 shows, under diffuse priors, portfolio weights are highly variable and often extreme.

Figure 6 offers another perspective on the relation between the predictor variable and the allocation. Panel A shows the posterior means of β for the stock index and the long-term bond. The posterior means are shown for priors ranging from dogmatic to diffuse ($P_{.02}$ ranging from 0 to 1) and for the point estimates of β from OLS. The left side of Panel A shows that, for the stock, the OLS betas lie above the posterior mean for the entire sample. The posterior means for $P_{.02} = .999$, and for $P_{.02} = 0.075$ decline around 1995, and then rise again around 2000.[21]

The posterior mean of the stock beta for $P_{.02} = 0.075$ lies above the posterior mean for the diffuse prior after 2000. This may seem surprising, as the role of the prior is to shrink the βs toward zero. However, the prior shrinks the total amount of predictability as measured by the R^2. This can be accomplished not only by shrinking β, but also by shrinking the persistence relative to the diffuse prior. In contrast to the posterior means for the less informative priors, the posterior mean for $P_{.02} = .0005$ remains steady throughout the sample and actually increases after 2000.

Panel B of Figure 6 shows holdings in the stock and bond for a range of beliefs about

---

[21]This plot is suggestive of parameter instability in the postwar sample; indeed evidence of such instability is found by Lettau and Van Nieuwerburgh (2007), Paye and Timmermann (2006), and Viceira (1996).

predictability. Also displayed are holdings resulting from OLS estimation. Volatility in holdings for the stock decline substantially as the prior becomes more dogmatic. For the fully dogmatic prior, the weight on the stock index displays some initial volatility, and then stays at about 40% after about 1976. The prior that is close to dogmatic, $P_{.02} = .0005$, implies some market timing for the stock portfolio based on the dividend-price ratio. In the early part of the sample the weight in the stock index implied by this prior is about 50%, declining to zero at the end of the sample. Of course, $P_{.02} = .075$ and the diffuse prior imply greater amounts of market timing. These priors imply time-varying weights that fluctuate both at a very slow frequency, and at a higher frequency. However, the weights that display the most volatility are those arising from ordinary least squares regression.

Figure 6 also shows corresponding results for the bond. The right side of Panel A shows that for most of the sample, the dividend yield predicts bond returns with a positive sign, just as it does with stock returns. Starting in the late-90s, this predictability begins to decline. Volatility in bond holdings also decline substantially as the prior becomes more dogmatic. Holdings corresponding to yield spread predictability (shown in Panel B of Figure 5 and in Figure 7) share these features.

Figure 6 demonstrates that using predictive variables in portfolio allocations need not lead to extreme weights. Combining the sample evidence with priors that are skeptical about return predictability leads to a moderate amount of market timing. We now turn to the out-of-sample performance of these strategies.

2.4 Out-of-sample performance

The previous sections show how skeptical priors can inform portfolio selection. The results indicate that investors who are highly skeptical about return predictability nonetheless choose time-varying weights, but that these weights are less variable and extreme than the weights for investors with diffuse priors. In this section, we assess out-of-sample performance of these priors.[22] The goal is not to find out which prior is "best"; indeed this analysis should

---

[22] Avramov (2004) and Handa and Tiwari (2006) also highlight how informative non-dogmatic priors can improve ex post performance of portfolios. Unlike the present study, these studies use informative priors for the cross-section (in the style of Pastor and Stambaugh (1999)), and show that shrinking a broad cross-section of returns toward what would be implied by asset pricing models can lead to better ex post performance in settings that include return predictability. As discussed in Section 1.2, Avramov also examines the effects of

not be interpreted as suggesting that one prior is superior to another. Rather, the purpose of this section is to relate our findings to those in papers that critique the evidence for predictability based on out-of-sample performance (Bossaerts and Hillion (1999), Goyal and Welch (2007)).

To assess out-of-sample performance in a way that controls for risk, we adopt a certainty equivalent approach. The certainty equivalent return (CER) answers the question: "what riskless rate would the investor be willing to accept in exchange for not following this strategy?" That is

$$\text{CER} = E[r_p] - A\frac{1}{2}\text{Var}[r_p], \tag{20}$$

where A is the appropriate risk-aversion parameter (see, e.g., Brennan and Xia (2001)). In this analysis, the mean and variance in (20) are computed using the sample mean and variance that result from following strategies associated with a given prior belief. That is, for each quarter, we apply the weights described in the previous section to the actual returns realized over the next quarter. This gives us a time series of 120 quarterly returns to use in computing the means and variances. In reporting the certainty equivalent returns, we multiply by 400 to express the return as an annual percentage. Results are reported for values of A equal to 2 and 5. As an additional metric, we also report out-of-sample Sharpe ratios. These are equal to the sample mean of excess returns, divided by the sample standard deviation. Excess returns are quarterly and constructed as described above. In reporting Sharpe ratios, we multiply by 2 to annualize.

Panel A of Table 2 reports CERs and Sharpe ratios when the dividend-price ratio is the predictor variable. For both metrics, the weights implied by ordinary least squares deliver worse performance than the weights implied by the dogmatic prior, which implies no market timing. A similar result is found by Goyal and Welch (2007), who argue against the use of predictability in portfolio allocation. We find, however, that market timing can increase out-of-sample performance, if the investor treats the evidence with some skepticism. The intermediate prior with $P_{.02} = .0005$ has the best out-of-sample performance regardless of the level of risk aversion and whether the CER or Sharpe ratio metric is used. While ignoring the predictability evidence results in better performance than applying a diffuse prior, simply

introducing a "prior" sample with no predictability and no mispricing relative to the 3-factor model. When there is such a sample, performance is often better than when there is no such prior sample. However, it is not clear whether these results are due to the prior over predictability or mispricing.

looking at these two extreme positions hides the better performance that can be achieved by taking the intermediate view.

Panel B of Table 2 reports analogous results for the yield spread as the predictor variable. For the certainty equivalent return metric, OLS again performs the worst, while the diffuse and dogmatic priors perform somewhat better. However, skeptical priors perform better still.[23] The differences can be large. When the risk aversion parameter equals 2, for example, a diffuse prior results in a CER of 5.47% per annum, while a dogmatic prior results in a CER of 5.61%. However, for the intermediate prior of $P_{.02} = .0005$, the CER is 7.58% per annum; for the prior with $P_{.02} = .075$, the CER is 9.19% per annum. These results show that priors indicating skepticism about the degree of predictability lead to superior out-of-sample performance over the postwar period, as well as to less extreme portfolio allocations, and to more stable parameter estimates.

We now ask whether the out-of-sample results reported in Table 2 are statistically significant. Unlike the rest of the paper, this exercise is purely frequentist in nature. More precisely, we take as our null hypothesis that returns are unpredictable. We then ask, what is the probability, calculated across repeated samples, that our strategies would have led to superior performance? Because these repeated samples are generated under the null of no predictability, any outperformance would presumably be spurious.

To correctly assess the statistical significance of our results, we use a Monte Carlo procedure. The goal of this procedure is to capture the dependencies inherent in our out-of-sample return observations (even though the shocks to returns and the predictor variable may be independent across time, portfolio returns depend on the investor's posterior, and therefore in principle on all previous observations). We simulate 300 samples of data designed so that means, variances, and covariances of returns and of the state variable, and the autocorrelation of the state variable are the same as in the data. However, in simulated data, returns are unpredictable. We then calculate out-of-sample performance, exactly repeating the procedure we used to calculate performance in actual data.[24]

---

[23]For the Sharpe ratio metric, OLS and the diffuse priors also outperform the most dogmatic prior.

[24]We consider only 300 draws because of the heavy computational requirements. While sampling from the posterior distribution for a given dataset takes under an hour, obtaining out-of-sample statistics requires that this procedure be repeated 32 times (once for every sample year since 1972). The whole procedure then needs to be repeated 300 times for the Monte Carlo. The entire set of Monte Carlo results (for two predictor variables and four priors) requires about three months of computation time on a 10-node Beowulf cluster.

Panel A of Table 3 shows the results of this exercise for the dividend-price ratio. The first row of Panel A reports the difference between the CER of each skeptical prior (and the regression estimate) and that of the dogmatic prior seen in the data for a risk aversion of 2. These differences are a measure of the ex-post gains from market timing. The next two rows report the 5th and the 95th percentile of the Monte-Carlo-generated statistic. The results show that the outperformance of the prior $P_{.02} = .0005$ could have occurred by chance with at most a 10% probability using a two-tailed test (because computation costs force us to limit the Monte Carlo to 300 draws, critical values corresponding to higher levels of significance would be unreliable). In the data, outperformance is 69 basis points per year while the 95th percentile is 60 basis points. The subsequent rows report the results of a similar exercise for the Sharpe ratio. The difference in the data between the skeptical and the dogmatic prior is 0.025; the 95th percentile is only 0.006.

Panel B of Table 3 reports analogous results for the yield spread. Both skeptical priors ($P_{.02} = .0005$ and $P_{.02} = .075$) outperform the dogmatic prior. This outperformance is far above the 95th percentile for the Monte Carlo draws for both the CER and the Sharpe ratio measure. For example, when risk aversion is 2, $P_{.02} = .075$ implies a performance differential of 3.58%. The 95th percentile is 1.55%. When risk aversion is 5, this prior implies a performance differential of 1.46%; the 95th percentile is 0.62%. Taken together, our Monte Carlo results suggest that it is quite unlikely that the superior performance of the skeptical priors could have occurred by chance.

Finally, we examine the performance of our strategies over subperiods of the data. The results above use data from 1952 to 2004. Here, we repeat the analysis with data ending in 1984 and 1994. The results for the dividend-price ratio are shown in Panel A of Table 4. For each subsample, OLS continues to give the worst performance as measured by certainty equivalent returns. However, skeptical priors deliver superior performance relative to the dogmatic prior. Similar results hold when the yield spread is the predictor variable, as shown in Panel B. In both subsamples, and across all measures of performance, the skeptical priors outperform the dogmatic prior.

To summarize, we find that intermediate priors lead to strategies that exhibit superior out of sample performance. They outperform strategies implied by the diffuse prior, which

---

This prevents us from embarking on more detailed statistical analyses, e.g. correcting for the impact of searching over the two skeptical priors.

takes an entirely data-driven view. They also outperform strategies implied by the dogmatic prior. While this latter result may seem unsurprising in light of the in-sample evidence that the yield spread and the dividend-price ratio predict returns, it is in contrast to the findings of Goyal and Welch (2007) and Bossaerts and Hillion (1999) that show that strategies making use of predictability have poor out-of-sample performance.

3   Conclusions

How much evidence on predictability is enough to influence portfolio choice? One view is that predictability should be taken into account only if the statistical evidence for it is incontrovertible. An opposite view is that investors should time their allocations to a large extent, even if the evidence for predictability is weak according to conventional measures. The first view states that investors should be extremely skeptical when viewing data showing evidence of predictability, while the second view states that no skepticism is necessary at all.

In this paper, we modeled the portfolio choice problem of an investor who has prior beliefs on the amount of predictability in the data. These prior beliefs put "skepticism" about predictability on a sound decision-theoretic basis. The skeptical investor believes that, while predictability is possible, large values of the R^2 are unlikely.

We applied our method to post-war data on bond and stock returns, with the yield spread and the dividend-price ratio as predictor variables. We found that even investors with a high degree of prior skepticism still vary their allocations to long-term bonds and stocks based on both of these variables. Thus the amount of predictability in the data is sufficient to influence investment, even if the investor is skeptical about the strength of this predictability.

To see the implications of various prior beliefs for portfolio allocations over the postwar period, we implemented an out-of-sample analysis. For each quarter starting in 1972, the posterior and optimal portfolio weights were determined based on previous data. Parameter estimates implied by skeptical priors were more stable than those implied by the diffuse prior and by ordinary least squares regression. Moreover, the resulting weights are less volatile and deliver superior out-of-sample performance as compared to the weights implied by an entirely model-based or data-based view.

This study provides a method for rigorously implementing an intermediate view on pre-

dictability. The resulting portfolio weights are more reasonable, and in fact perform better out of sample than either extreme view. The question remains as to why the skeptical prior outperforms. Skepticism can be motivated by a theory of rational markets (see Samuelson (1965)). Our results suggest that there may be value to this theory even if it does not hold exactly. More broadly, this study supports the idea that using models to downweight unreasonable regions of the parameter space may improve decision making.

Appendix

A   Extensions to multiple predictor variables

This section extends the results in Section 1 to the case of multiple predictor variables. We continue to assume (1) and (2), except that x_t is now allowed to be $K \times 1$, β is $N \times K$, θ_0 is $K \times 1$, and θ_1 is $K \times K$. Define

$$B = \begin{bmatrix} \alpha^\top & \theta_0^\top \\ \beta^\top & \theta_1^\top \end{bmatrix}$$

The multivariate analogues of (5) and (6) are

$$\mu_x = E\left[x_t \mid B, \Sigma\right] = (I_K - \theta_1)^{-1} \theta_0, \tag{21}$$

and

$$\text{vec}(\Sigma_x) = \left(I_{K^2} - (\theta_1 \otimes \theta_1)\right)^{-1} \text{vec}(\Sigma_v), \tag{22}$$

where $\Sigma_x = E\left[(x_t - x_0)(x_t - x_0)^\top\right]$ (see Hamilton (1994, p. 265)).

As in Section 1, we specify a prior over normalized β. Let C_x be a lower triangular matrix such that $C_x C_x^\top = \Sigma_x$, and

$$\eta = C_u^{-1} \beta C_x \tag{23}$$

(recall that C_u is such that $C_u C_u^\top = \Sigma_u$). Assume prior beliefs on η are given by

$$\text{vec}(\eta) \sim N(0, \sigma_\eta^2 I_{NK}).$$

Then (23) implies

$$\text{vec}(\eta) = \left(C_x^\top \otimes C_u^{-1}\right) \text{vec}(\beta)$$

and therefore,

$$\text{vec}(\beta) \mid \alpha, \theta_0, \theta_1, \Sigma \sim N\left(0, \sigma_\eta^2 \left(\Sigma_x^{-1} \otimes \Sigma_u\right)\right). \tag{24}$$

Appendix D shows that the Jeffreys prior on $(\alpha, \theta_0, \theta_1, \Sigma)$ takes the form

$$p(\alpha, \theta_0, \theta_1 \Sigma) \propto |\Sigma|^{-\frac{N+2K+2}{2}} |\Sigma_u|^{K/2} |\Sigma_x|^{K/2}. \tag{25}$$

Because

$$\begin{aligned}
\text{vec}(\beta)^\top \left(\Sigma_u \otimes \Sigma_x^{-1}\right)^{-1} \text{vec}(\beta) &= \text{vec}(\beta)^\top \left(\Sigma_u^{-1} \otimes \Sigma_x\right) \text{vec}(\beta) \\
&= \text{tr}\left(\beta^\top \Sigma_x \beta \Sigma_u^{-1}\right),
\end{aligned}$$

the joint prior is given by

$$p(B, \Sigma) \propto |\Sigma_x|^{\frac{N+K}{2}} |\Sigma|^{-\frac{N+2K+2}{2}} \exp\left\{-\frac{1}{2}\sigma_\eta^{-2}\mathrm{tr}\left(\beta^\top \Sigma_x \beta \Sigma_u^{-1}\right)\right\}. \tag{26}$$

As σ_η approaches zero, $p(B, \Sigma)$ approaches a dogmatic, no-predictability prior. Appendix C shows that $p(B, \Sigma)$ approaches a Jeffreys prior as σ_η approaches infinity. Intermediate levels of σ_η allow for skepticism about the level of predictability by placing an informative prior on the maximum R^2 implied by the equations (1) and (2). The generalization of the equation for the maximum R^2 is

$$\max_w R^2 = \max_w \frac{w^\top \beta \Sigma_x \beta^\top w}{w^\top \beta \Sigma_x \beta^\top w + w^\top \Sigma_u w}$$

$$= \frac{\eta^\top \eta}{\eta^\top \eta + 1} \tag{27}$$

as shown in Appendix B.

The computations for the likelihood and the posterior closely follow those in Section 1. Define

$$Y = \begin{bmatrix} r_1^\top & x_1^\top \\ \vdots & \vdots \\ r_T^\top & x_T^\top \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_0^\top \\ \vdots & \vdots \\ 1 & x_{T-1}^\top \end{bmatrix}.$$

The exact likelihood is given by

$$p(D|B, \Sigma) = |2\pi\Sigma_x|^{-\frac{1}{2}}|2\pi\Sigma|^{-\frac{T}{2}}$$

$$\exp\left\{-\frac{1}{2}(x_0 - \mu_x)^\top \Sigma_x^{-1}(x_0 - \mu_x) - \frac{1}{2}\mathrm{tr}\left[(Y - XB)^\top(Y - XB)\Sigma^{-1}\right]\right\}. \tag{28}$$

Applying Bayes rule to (26) and (28) leads to the posterior

$$p(B, \Sigma|D) \propto |\Sigma_x|^{\frac{N+K-1}{2}}|\Sigma|^{-\frac{T+N+2K+2}{2}} \exp\left\{-\frac{1}{2}(x_0 - \mu_x)^\top \Sigma_x^{-1}(x_0 - \mu_x)\right\}$$

$$\exp\left\{-\frac{1}{2}\sigma_\eta^{-2}\mathrm{tr}\left(\beta^\top \Sigma_x \beta \Sigma_u^{-1}\right) - \frac{1}{2}\mathrm{tr}\left[(Y - XB)^\top(Y - XB)\Sigma^{-1}\right]\right\}. \tag{29}$$

The procedure for sampling from this posterior is very similar to that described in Section 1.3. The Metropolis-Hastings algorithm is used to draw blocks of parameters at a time. First, Σ is drawn from $p(\Sigma|B, D)$ using the inverted Wishart distribution with $T + K + 1$ degrees of freedom. Second, α and β are drawn from $p(\alpha, \beta|\theta_0, \theta_1, D)$, and finally θ_0 and θ_1 are drawn from $p(\theta_0, \theta_1|\alpha, \beta, \Sigma, D)$. The computation of the conditional posteriors for (α, β), and (θ_0, θ_1) is exactly as in Section 1.3.

B   Derivation of the maximum R^2

This section derives equation (14) and (27). The maximum R^2 satisfies

$$\max R^2 = \max_w \frac{w^\top \beta \Sigma_x \beta^\top w}{w^\top \beta \Sigma_x \beta^\top w + w^\top \Sigma_u w}, \tag{30}$$

where w is an $N \times 1$ nonzero vector (it is not necessary that the elements of w sum to 1). Define $\bar{w} = C_u^\top w$, where $C_u C_u^\top = \Sigma_u$. Rewrite (30) as

$$\max R^2 = \max_{\bar{w}} \frac{\bar{w}^\top \eta \eta^\top \bar{w}}{\bar{w}^\top \eta \eta^\top \bar{w} + \bar{w}^\top \bar{w}}. \tag{31}$$

The Cauchy-Schwartz inequality implies that $\frac{\bar{w}^\top \eta \eta^\top \bar{w}}{\bar{w}^\top \bar{w}} \leq \eta^\top \eta$ for all nonzero vectors $\bar{w}$, and that equality is achieved by choosing $\bar{w}$ to be proportional to η. Therefore $\bar{w} = \eta$ is a solution to (31). Substituting η in for $\bar{w}$ in the argument of (31) implies (27). Equation (14) is a special case.

C   Jeffreys prior on B, Σ

Our derivation for the limiting Jeffreys prior on B, Σ follows Stambaugh (1999). We derive the prior density for $p(B, \Sigma^{-1})$ and then transform this into the density for $p(B, \Sigma)$ using the Jacobian. Let $b = \text{vec}(B)$, and

$$\zeta = (\sigma^{11}, \sigma^{12}, \ldots, \sigma^{1,N+K}, \sigma^{22}, \sigma^{23}, \ldots, \sigma^{2,N+K}, \ldots, \sigma^{N+K,N+K}),$$

where σ^{ij} denotes element (i,j) of Σ^{-1}. Let $l(B, \Sigma; D) = \log p(D|B, \Sigma, x_0)$ denote the natural log of the conditional likelihood. The definition of the Jeffreys prior implies

$$p(B, \Sigma^{-1}) \propto \left| -E \left(\begin{bmatrix} \frac{\partial^2 l}{\partial b \partial b^\top} & \frac{\partial^2 l}{\partial b \partial \zeta^\top} \\ \frac{\partial^2 l}{\partial \zeta \partial b^\top} & \frac{\partial^2 l}{\partial \zeta \partial \zeta^\top} \end{bmatrix} \right) \right|^{1/2}. \tag{32}$$

Computing the expectation on the right hand side of (32) yields

$$p(B, \Sigma^{-1}) \propto \left| \begin{matrix} \Sigma^{-1} \otimes \Psi & 0 \\ 0 & \frac{\partial^2 \log |\Sigma|}{\partial \zeta \partial \zeta^\top} \end{matrix} \right|^{1/2}, \tag{33}$$

where

$$\Psi = \begin{bmatrix} 1 & \mu_x^\top \\ \mu_x & \mu_x \mu_x^\top + \Sigma_x \end{bmatrix}.$$

Note that Ψ is $(K+1) \times (K+1)$. From the formula for the determinant of a partitioned matrix (Green (1997, p. 33)), it follows that $|\Psi| = \Sigma_x$.

Box and Tiao (1973, pp. 474–475) show that

$$\left| \frac{\partial^2 \log |\Sigma|}{\partial \zeta \partial \zeta^\top} \right|^{1/2} = |\Sigma|^{\frac{N+K+1}{2}} \tag{34}$$

(recall that Σ is $(N+K) \times (N+K)$). It then follows from (33) that

$$\begin{aligned} p(B, \Sigma^{-1}) &\propto |\Sigma^{-1} \otimes \Psi|^{1/2} |\Sigma|^{\frac{N+K+1}{2}} \\ &= \left(|\Sigma|^{-(K+1)} |\Psi|^{N+K} \right)^{1/2} |\Sigma|^{\frac{N+K+1}{2}} \\ &= |\Psi|^{\frac{N+K}{2}} |\Sigma|^{N/2} \\ &= |\Sigma_x|^{\frac{N+K}{2}} |\Sigma|^{N/2}. \end{aligned}$$

From results in Zellner (1996, p. 226), it follows that Jacobian of the transformation from Σ^{-1} to Σ is $|\Sigma|^{-(N+K+1)}$. Therefore

$$p(B, \Sigma) \propto |\Sigma_x|^{\frac{N+1}{2}} |\Sigma|^{-\frac{N+2K+2}{2}}. \tag{35}$$

D   Jeffreys prior on $\alpha, \theta_0, \theta_1, \Sigma$

We calculate the prior for $(\alpha, \theta_0, \theta_1, \Sigma^{-1})$, and use the determinant of the Jacobian to transform this into a prior for $(\alpha, \theta_0, \theta_1, \Sigma)$. Define blocks of Σ^{-1} as

$$\Sigma^{-1} = \begin{bmatrix} (\Sigma^{-1})_{11} & (\Sigma^{-1})_{12} \\ (\Sigma^{-1})_{21} & (\Sigma^{-1})_{22} \end{bmatrix}.$$

Here, $(\Sigma^{-1})_{11}$ is $N \times N$, $(\Sigma^{-1})_{12}$ is $N \times K$, $(\Sigma^{-1})_{21} = (\Sigma^{-1})_{12}^\top$, and $(\Sigma^{-1})_{22}$ is $K \times K$.

The starting point for the calculation is the information matrix for B, Σ^{-1} given in Appendix C. The information matrix for $\alpha, \theta_0, \theta_1, \Sigma^{-1}$ can be obtained by removing the rows and columns corresponding to derivatives with respect to β_{ij} from (33). Without loss of generality, the rows and columns of (33) can be re-ordered so that

$$p(B, \Sigma^{-1}) \propto \left| \begin{matrix} \Psi \otimes \Sigma^{-1} & 0 \\ 0 & \frac{\partial^2 \log |\Sigma|}{\partial \zeta \partial \zeta^\top} \end{matrix} \right|^{1/2}.$$

31

This corresponds to taking second derivatives of l with respect to $\text{vec}(B^\top)$ rather than $\text{vec}(B)$. Because $\text{vec}(B^\top) = (\alpha^\top, \theta_0^\top, \text{vec}(\beta)^\top, \text{vec}(\theta_1)^\top)^\top$, removing the rows and columns corresponding to $\text{vec}(\beta)$ leads to

$$p(\alpha, \theta_0, \theta_1, \Sigma^{-1}) \propto \begin{vmatrix} \Phi & 0 \\ 0 & \frac{\partial^2 \log |\Sigma|}{\partial \zeta \partial \zeta^\top} \end{vmatrix}^{1/2}, \tag{36}$$

where

$$\Phi = \begin{bmatrix} \Sigma^{-1} & \mu_x^\top \otimes \begin{bmatrix} (\Sigma^{-1})_{12} \\ (\Sigma^{-1})_{22} \end{bmatrix} \\ \mu_x \otimes [(\Sigma^{-1})_{21}, (\Sigma^{-1})_{22}] & (\Sigma_x + \mu_x \mu_x^\top) \otimes (\Sigma^{-1})_{22} \end{bmatrix}.$$

From the formula for the determinant of a partitioned matrix, it follows that

$$|\Phi| = |\Sigma^{-1}| \left| (\Sigma_x + \mu_x \mu_x^\top) \otimes (\Sigma^{-1})_{22} - (\mu_x \otimes [(\Sigma^{-1})_{21}, (\Sigma^{-1})_{22}]) \Sigma \left(\mu_x^\top \otimes \begin{bmatrix} (\Sigma^{-1})_{12} \\ (\Sigma^{-1})_{22} \end{bmatrix} \right) \right|.$$

Because

$$\Sigma \begin{bmatrix} (\Sigma^{-1})_{12} \\ (\Sigma^{-1})_{22} \end{bmatrix} = \begin{bmatrix} 0_{N \times K} \\ I_K \end{bmatrix},$$

it follows that

$$\begin{aligned} |\Phi| &= |\Sigma^{-1}| \left| (\Sigma_x + \mu_x \mu_x^\top) \otimes (\Sigma^{-1})_{22} - \mu_x \mu_x^\top \otimes (\Sigma^{-1})_{22} \right| \\ &= |\Sigma^{-1}| \left| \Sigma_x \otimes (\Sigma^{-1})_{22} \right| \\ &= |\Sigma|^{-1} |\Sigma_x|^K \left| (\Sigma^{-1})_{22} \right|^K. \end{aligned}$$

Applying the formula for the determinant of a partitioned matrix to Σ produces

$$|\Sigma| = |\Sigma_u| \left| \Sigma_v - \Sigma_{vu} \Sigma_u^{-1} \Sigma_{uv} \right|.$$

By the formula for the inverse of a partitioned matrix (Green (1997, p. 33)),

$$\left| (\Sigma^{-1})_{22} \right| = \left| \Sigma_v - \Sigma_{vu} \Sigma_u^{-1} \Sigma_{uv} \right|^{-1}.$$

Therefore,

$$\begin{aligned} |\Phi| &= |\Sigma|^{-(K+1)} |\Sigma|^K \left| (\Sigma^{-1})_{22} \right|^K |\Sigma_x|^K \\ &= |\Sigma|^{-(K+1)} |\Sigma_u|^K |\Sigma_x|^K. \end{aligned}$$

32

Finally, from (34) and (36),

$$p(\alpha, \theta_0, \theta_1, \Sigma^{-1}) \quad \propto \quad |\Phi|^{1/2} |\Sigma|^{\frac{N+K+1}{2}}$$

$$= \quad |\Sigma|^{N/2} |\Sigma_u|^{K/2} |\Sigma_x|^{K/2}.$$

which, by the Jacobian of the transformation from Σ^{-1} to Σ (see Appendix C) completes the proof of (25).

E   Sampling from the posterior

The conditional density $p(\Sigma|B, D)$ has an expression identical to (16). The proposal density is the inverted Wishart with $T + 2$ degrees of freedom (Zellner (1996, p. 395)):

$$f(\Sigma|B, D) \propto |\Sigma|^{-\frac{T+N+4}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[(Y - XB)^\top (Y - XB)\Sigma^{-1}\right]\right\}.$$

Because the target density takes the form

$$p(\Sigma|B, D) \propto \psi(\Sigma) \times \text{proposal},$$

we can use the results in Chib and Greenberg (1995, Section 5) to sample from the posterior.

The density for $p(B|\Sigma, D)$ is sampled from in two steps: first we sample from $p(b_1|b_2, \Sigma, D)$, and next we sample from $p(b_2|b_1, \Sigma, D)$. For the first of these steps, we can sample directly from the true density, without using the accept-reject algorithm. Note that

$$\text{tr}\left[(Y - XB)^\top (Y - XB)\Sigma^{-1}\right] \quad = \quad (b - \hat{b})^\top (\Sigma^{-1} \otimes X^\top X)(b - \hat{b}) + \text{terms independent of } B$$

where $\hat{b} = \text{vec}(\hat{B})$, and $\hat{B} = (X^\top X)^{-1} X^\top Y$. Let $V = \left(\Sigma^{-1} \otimes X^\top X\right)^{-1}$, and partition V so that

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix},$$

where V_{11} is $2N \times 2N$, V_{12} is $2N \times 2$, $V_{21} = V_{12}^\top$ and V_{22} is 2×2. Then

$$(b - \hat{b})^\top (\Sigma^{-1} \otimes X^\top X)(b - \hat{b}) =$$

$$\left(b_1 - \hat{b}_1 - V_{12}V_{22}^{-1}(b_2 - \hat{b}_2)\right)^\top \left(V_{11} - V_{12}V_{22}^{-1}V_{21}\right)^{-1} \left(b_1 - \hat{b}_1 - V_{12}V_{22}^{-1}(b_2 - \hat{b}_2)\right)$$

$$+ \text{ terms independent of } b_1.$$

(see Green (1997, Chapter 3.1)). Then under the prior (12),

$$p(b_1 \mid b_2, \Sigma, D) \propto \exp\left\{ -\frac{1}{2}\beta^\top \left(\sigma_\eta^2 \sigma_x^{-2}\Sigma_u\right)^{-1}\beta \right\}$$

$$\exp\left\{ -\frac{1}{2}\left(b_1 - \hat{b}_1 - V_{12}V_{22}^{-1}(b_2 - \hat{b}_2)\right)^\top \left(V_{11} - V_{12}V_{22}^{-1}V_{21}\right)^{-1}\left(b_1 - \hat{b}_1 - V_{12}V_{22}^{-1}(b_2 - \hat{b}_2)\right) \right\}.$$

We now complete the square to derive the distribution for b_1. Let

$$\Omega = \Sigma_u^{-1} \otimes \begin{bmatrix} 0 & 0 \\ 0 & \left(\sigma_\eta^2\sigma_x^{-2}\right)^{-1} \end{bmatrix}. \tag{37}$$

The posterior distribution for b_1 conditional on b_2 and Σ is normal with variance

$$V_{11}^* = \left(\Omega + (V_{11} - V_{12}V_{22}^{-1}V_{21})^{-1}\right)^{-1}$$

and mean

$$b_1^* = V_{11}^*(V_{11} - V_{12}V_{22}^{-1}V_{21})^{-1}\left(\hat{b}_1 + V_{12}V_{22}^{-1}(b_2 - \hat{b}_2)\right).$$

Similar reasoning can be used to draw from the posterior for $b_2 = [\theta_0, \theta_1]$ conditional on the other parameters, Applying the results of Green (1997, Chapter 3.1),

$$(b - \hat{b})^\top(\Sigma^{-1} \otimes X^\top X)(b - \hat{b}) =$$

$$\left(b_2 - \hat{b}_2 - V_{21}V_{11}^{-1}(b_1 - \hat{b}_1)\right)^\top \left(V_{22} - V_{21}V_{11}^{-1}V_{12}\right)^{-1}\left(b_2 - \hat{b}_2 - V_{21}V_{11}^{-1}(b_1 - \hat{b}_1)\right)$$

$$+ \text{ terms independent of } b_2.$$

The density is sampled from using an accept-reject algorithm. The proposal density is normal with mean $\hat{b}_2 + V_{12}^\top V_{11}^{-1}(b_1 - \hat{b}_1)$ and variance $V_{22} - V_{12}^\top V_{11}^{-1}V_{12}$. The target density is

$$p(b_2|b_1, \Sigma, D) \propto \sigma_x^N \exp\left\{ -\frac{1}{2}\beta^\top \left(\sigma_\eta^2\sigma_x^{-2}\Sigma_u\right)^{-1}\beta - \frac{1}{2}\sigma_x^{-2}(x_0 - \mu_x)^2 \right\} \times \text{proposal}.$$

for $\theta_1 \in (0, 1)$ and zero otherwise.[25]

---

[25]For the case of multiple predictor variables, the procedure for simulating from the posterior is very similar. In this case, V_{11} is $NK \times NK$ and

$$\Omega = \Sigma_u^{-1} \otimes \begin{bmatrix} 0 & 0 \\ 0 & \sigma_\eta^{-2}\Sigma_x \end{bmatrix}.$$

The proposal density for Σ is an inverted Wishart with $T + K + 1$ degrees of freedom.

34

References

Abel, Andrew, 1990, Asset prices under habit formation and catching up with the Joneses, *American Economic Review Papers and Proceedings* 80, 38–42.

Abel, Andrew, 1999, Risk premia and term premia in general equilibrium, *Journal of Monetary Economics* 43, 3–33.

Amihud, Yakov, and Clifford Hurvich, 2004, Predictive regressions: A reduced-bias estimation method, *Journal of Financial and Quantitative Analysis* 39, 813–841.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *Review of Financial Studies* 20, 651–707.

Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Avramov, Doron, 2004, Stock return predictability and asset pricing models, *Review of Financial Studies* 17, 699–738.

Backus, David, A. Gregory, and Stanley Zin, 1989, Risk premiums in the term structure: Evidence from artificial economies, *Journal of Monetary Economics* 24, 371–399.

Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *Journal of Finance* 56, 45–86.

Balduzzi, Pierluigi, and Anthony W. Lynch, 1999, Transaction costs and predictability: some utility cost calculations, *Journal of Financial Economics* 52, 47–78.

Bansal, Ravi, and Amir Yaron, 2004, Risks for the long-run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.

Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Barberis, Nicholas, Ming Huang, and Tano Santos, 2001, Prospect theory and asset prices, *Quarterly Journal of Economics* 116, 1–53.

Barro, Robert J., 2006, Rare disasters and asset markets in the twentieth century, *The Quarterly Journal of Economics* pp. 823–866.

Bawa, Vijay S., Stephen J. Brown, and Roger W. Klein, 1979, *Estimation risk and optimal portfolio choice.* (North-Holland Amsterdam).

Bekaert, Geert, Robert J. Hodrick, and David A. Marshall, 1997, On biases in tests of the expectations hypothesis of the term structure of interest rates, *Journal of Financial Economics* 44, 309–348.

Berger, James O., 1985, *Statistical decision theory and Bayesian analysis.* (Springer New York).

Best, Michael J., and Robert R. Grauer, 1991, On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results, *Review of Financial Studies* 4, 315–342.

Black, Fischer, and Robert Litterman, 1992, Global portfolio optimization, *Financial Analysts Journal* pp. 28–43.

Bossaerts, Peter, and Pierre Hillion, 1999, Implementing statistical criteria to select return forecasting models: What do we learn?, *Review of Financial Studies* 12, 405–428.

Box, George, Gwilym M. Jenkins, and Gregory Reinsel, 1970, *Time Series Analysis: Forecasting and Control.* (Holden-Day Inc. S. Francisco, CA).

Box, George E.P., and George C. Tiao, 1973, *Bayesian Inference in Statistical Analysis.* (Addison-Wesley Pub. Co. Reading, MA).

Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud, 2005, A simulation approach to dynamics portfolio choice with an application to learning about return predictability, *Review of Financial Studies* 18, 831–873.

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1997, Strategic asset allocation, *Journal of Economic Dynamics and Control* 21, 1377–1403.

Brennan, Michael J., and Yihong Xia, 2001, Assessing asset pricing anomalies, *Review of Financial Studies* 14, 905–942.

Brennan, Michael J., and Yihong Xia, 2005, Persistence, predictability, and portfolio planning, Working paper, University of California at Los Angeles.

Brown, Stephen J., 1979, Optimal portfolio choice under uncertainty: A Bayesian approach, in Vijay S. Bawa, Stephen J. Brown, and Roger W. Klein, eds.: *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).

Campbell, John Y., 1986, Bond and stock returns in a simple exchange model, *Quarterly Journal of Economics* 101, 785–803.

Campbell, John Y., 2003, Consumption-based asset pricing, in *Handbook of the Economics of Finance Volume IB* (North-Holland, Amsterdam).

Campbell, John Y., Y. Lewis Chan, and Luis M. Viceira, 2003, A multivariate model of strategic asset allocation, *Journal of Financial Economics* 67, 41–80.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Robert J. Shiller, 1991, Yield spreads and interest rate movements: A bird's eye view, *Review of Economic Studies* 58, 495–514.

Campbell, John Y., and Samuel Thompson, 2007, Predicting excess stock returns out of sample: Can anything beat the historical average?, forthcoming, *Review of Financial Studies*.

Campbell, John Y., and Luis M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time-varying, *Quarterly Journal of Economics* 114, 433–495.

Campbell, John Y., and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.

Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.

Cecchetti, S.G., P.S. Lam, and N.C. Mark, 1993, The equity premium and the risk-free rate: Matching the moments, *Journal of Monetary Economics* 31, 21–46.

Chib, Siddhartha, and Edward Greenberg, 1995, Understanding the Metropolis-Hastings algorithm, *American Statistician* 49, 327–335.

Cochrane, John H., 1992, Explaining the variance of price-dividend ratios, *Review of Financial Studies* 5, 243–280.

Cochrane, John H., 1999, Portfolio advice for a multifactor world, *Economic Perspectives* Federal Reserve Bank of Chicago 23, 59–78.

Cremers, K.J. Martjin, 2002, Stock return predictability: A Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.

Cvitanic, Jaksa, Ali Lazrak, Lionel Martellini, and Fernando Zapatero, 2006, Dynamic portfolio choice with parameter uncertainty and the economic value of analysts'recommendations, *Review of Financial Studies* 19, 1113–1156.

Eliasz, Piotr, 2004, Optimal median unbiased estimation of coefficients on highly persistent regressors, Working paper, Princeton University.

Elliott, Graham, and James H Stock, 1994, Inference in Time Series Regression When the Order of Integration of a Regressor Is Unknown, *Econometric Theory* 10, 672–700.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F., and William G. Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 115–146.

Ferson, Wayne E., Sergei Sarkissian, and Timothy T. Simin, 2003, Spurious regressions in financial economics?, *Journal of Finance* 58, 1393–1413.

Frost, Peter A., and James Savarino, 1988, For better performance constrain portfolio weights, *Journal of Portfolio Management* 15, 29–34.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang, 2007, Portfolio selection with parameter and model uncertainty: A multi-prior approach, *Review of Financial Studies* 20, 41–81.

Goetzmann, William N., and Philippe Jorion, 1993, Testing the predictive power of dividend yields, *Journal of Finance* 48, 663–679.

Goyal, Amit, and Ivo Welch, 2007, A comprehensive look at the empirical performance of equity premium prediction, forthcoming, *Review of Financial Studies.*

Green, Richard C., and Burton Hollifield, 1992, When will mean-variance efficient portfolios be well diversified?, *Journal of Finance* 47, 1785–1809.

Green, William H., 1997, *Econometric Analysis.* (Prentice-Hall, Inc. Upper Saddle River, NJ).

Hamilton, J. D., 1994, *Time Series Analysis.* (Oxford University Press Princeton, NJ).

Handa, Puneet, and Ashish Tiwari, 2006, Does stock market predictability imply improved asset allocation and performance? Evidence from the U.S. stock market (1954-2002), *Journal of Business* 79, 2423–2468.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraint helps, *Journal of Finance* 58, 1651–1683.

Jeffreys, Harold, 1961, *Theory of Probability.* (Oxford University Press Clarenden).

Jobson, J. D., and Bob Korkie, 1980, Estimation of Markowitz efficient portfolios, *Journal of the American Statistical Association* 75, 544–554.

Johannes, Michael, Nicholas Polson, and Jonathan R. Stroud, 2002, Sequential optimal portfolio performance: Market and volatility timing, Working paper, Columbia University, University of Chicago, and University of Pennsylvania.

Jones, Christopher S., and Jay Shanken, 2005, Mutual fund performance with learning across funds, *Journal of Financial Economics* 78, 507–552.

Jorion, Philippe, 1985, International portfolio diversification with estimation risk, *Journal of Business* 58, 259–278.

Kandel, Shmuel, and Robert F. Stambaugh, 1991, Asset returns and intertemporal preferences, *Journal of Monetary Economics* 27, 39–71.

Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385–424.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Klein, Roger W., and Vijay S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215–231.

Kothari, S. P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time-series analysis, *Journal of Financial Economics* 44, 169–203.

Lettau, Martin, and Sydney C. Ludvigson, 2001, Consumption, aggregate wealth and expected stock returns, *Journal of Finance* 56, 815–849.

Lettau, Martin, and Stijn VanNieuwerburgh, 2007, Reconciling the return predictability evidence, forthcoming, *Review of Financial Studies*.

Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.

MacKinlay, Craig A., 1995, Multifactor models do not explain the CAPM, *Journal of Financial Economics* 38, 3–28.

Mankiw, N.Gregory, and Matthew D. Shapiro, 1986, Do we reject too often? : Small sample properties of tests of rational expectations models, *Economics Letters* 20, 139–145.

Mehra, Rajnish, and Edward Prescott, 1985, The equity premium puzzle, *Journal of Monetary Economics* 15, 145–161.

Nelson, C. R., and M. J. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.

Pastor, Lubos, 2000, Portfolio selection and asset pricing models, *Journal of Finance* 55, 179–223.

Pastor, Lubos, and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance* 54, 67–121.

Pastor, Lubos, and Robert F. Stambaugh, 2002, Investing in equity mutual funds, *Journal of Financial Economics* 63, 351–380.

Pastor, Lubos, and Robert F. Stambaugh, 2006, Predictive systems: Living with imperfect predictors, CRSP Working Paper # 617.

Paye, Bradley S., and Allan Timmermann, 2006, Instability of return prediction models, *Journal of Empirical Finance* 13, 274–315.

Poirier, Dale 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, 67–68.

Poirier, Dale J., 1996, Prior beliefs about fit, in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds.: *Bayesian Statistics 5* (Oxford University Press, Oxford).

Reitz, Thomas A., 1988, The equity risk premium: A solution, *Journal of Monetary Economics* 22, 117–131.

Richardson, Matthew, and James H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323–348.

Robbins, Herbert, 1964, The empirical Bayes approach to statistical decision problems, *The Annals of Mathematical Statistics* 35, 1–20.

Samuelson, Paul, 1965, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review* 6, 41–49.

Samuelson, Paul, 1973, Proof that properly discounted present values of assets vibrate randomly, *Bell Journal of Economics and Management Science* 4, 369–374.

Sangvinatsos, Antonios, and Jessica A. Wachter, 2005, Does the failure of the expectations hypothesis matter for long-term investors?, *Journal of Finance* 60, 179–230.

Shanken, Jay A., and Ane Tamayo, 2005, Dividend yield, risk, and mispricing: A Bayesian analysis, Working paper, Emory University and London Business School.

Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–436.

Sims, Christopher A., and Harald Uhlig, 1991, Understanding unit rooters: A helicopter tour, *Econometrica* 59, 1591–1599.

Skoulakis, Georgios, 2007, Dynamic portfolio choice with Bayesian learning, Working paper, University of Maryland.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.

Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.

Tu, Jun, and Guofu Zhou, 2007, Incorporating economic objectives into Bayesian priors: Portfolio choice under parameter uncertainty, Working paper, Singapore Management University and Washington University in St. Louis.

Viceira, Luis, 1996, Testing for structural change in the predictability of asset returns, Working paper, Harvard University.

Wang, Zhenyu, 2005, A shrinkage approach to model uncertainty and asset allocation, *Review of Financial Studies* 18, 673–705.

Whitelaw, Robert F., 2000, Stock market risk and return: an equilibrium approach, *Review of Financial Studies* 13, 521–547.

Xia, Yihong, 2001, Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation, *Journal of Finance* 56, 205–246.

Zellner, Arnold, 1996, *An introduction to Bayesian inference in econometrics.* (John Wiley and Sons, Inc. New York, NY).

Table 1: Posterior Means

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)
$E\left[r_{\text{bond}}|B,\Sigma\right]$	0.18	0.18	0.18	0.17	0.23
	(0.27)	(0.30)	(0.34)	(1.07)	
$E\left[r_{\text{stock}}|B,\Sigma\right]$	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)
$E\left[r_{\text{bond}}|B,\Sigma\right]$	0.21	0.21	0.21	0.21	0.23
	(0.28)	(0.28)	(0.29)	(0.33)	
$E\left[r_{\text{stock}}|B,\Sigma\right]$	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)	

Notes: Posterior means for the predictive coefficients β, the autoregressive coefficient on the predictor variable θ_1, and unconditional posterior means for returns and the predictor variable. Posterior standard deviations are in parentheses. The assets are the ten-year bond and the stock index; the predictor variables are log dividend-price ratio (Panel A) and the yield spread (Panel B). Prior beliefs are indexed by $P(R^2 > 2\%)$, the probability that the R^2 from the predictive regression exceeds 2%. $P(R^2 > 2\%) = 0$ corresponds to the dogmatic prior; $P(R^2 > 2\%) = 0.999$ corresponds to the diffuse prior. The last column gives results implied by parameters estimated by ordinary least squares regression. For regression estimates, standard errors are in parentheses. $E\left[r|B,\Sigma\right] = \alpha + \beta\frac{\theta_0}{1-\theta_1}$ and denotes the frequentist expectation of excess returns. $E\left[x|B,\Sigma\right] = \frac{\theta_0}{1-\theta_1}$ and denotes the frequentist expectation of the predictor variable. Data are quarterly from 1952 to 2004.

Table 2: Out-of-Sample Results

	$P(R^2 > 2\%)$				
A	0	0.0005	0.075	0.999	Reg.
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

Notes: Certainty equivalent returns and Sharpe ratios when the predictor variables are dividend-price ratio (Panel A) and the yield spread (Panel B) and the assets are the stock index and the ten-year bond. For each year beginning in 1972, the predictive distribution for returns is computed using all data up to that year. Optimal portfolios are then computed quarterly to maximize a mean-variance utility function with risk aversion parameter $A = 2, 5$; these are combined with actual returns over the following quarter to create out-of-sample returns on the investment strategy. Certainty equivalent returns (CERs) are defined as $E[r_p] - \frac{A}{2}\text{Var}[r_p]$, where r_p is the return on the investment strategy and the mean and variance are computed using the sample. Sharpe ratios are the average excess returns on the investment strategy divided by the standard deviation and do not depend on A. Prior beliefs are indexed by $P(R^2 > 2\%)$, the probability that the R^2 from the predictive regression exceeds 2%. $P(R^2 > 2\%) = 0$ corresponds to the dogmatic prior; $P(R^2 > 2\%) = 0.999$ corresponds to the diffuse prior. Reg. denotes results obtained from ordinary least squares regression. Data are from 1952 to 2004. CERs are in annualized percentages ($\times 400$), and Sharpe ratios are annualized ($\times 2$).

Table 3: Monte Carlo Results

| A	Statistic	$P(R^2 > 2\%)$ 0.0005	0.075	0.999	Reg.
Panel A: Dividend-Price Ratio					
	Certainty Equivalent Returns				
2	Data	0.69	-0.21	-1.72	-9.61
	5th percentile	-0.67	-1.93	-5.99	-9.61
	95th percentile	0.60	1.36	1.46	1.20
5	Data	0.26	-0.13	-0.75	-3.92
	5th percentile	-0.27	-0.77	-2.40	-3.84
	95th percentile	0.24	0.54	0.58	0.48
	Sharpe Ratios				
	Data	0.025	-0.009	-0.046	-0.029
	5th percentile	-0.006	-0.034	-0.111	-0.186
	95th percentile	0.006	0.015	-0.005	0.004
Panel B: Yield Spread					
	Certainty Equivalent Returns				
2	Data	1.97	3.58	-0.15	-4.97
	5th percentile	-2.11	-1.89	-7.91	-6.92
	95th percentile	1.75	1.55	2.47	1.76
5	Data	0.80	1.46	-0.01	-1.90
	5th percentile	-0.84	-0.75	-3.16	-2.77
	95th percentile	0.70	0.62	0.99	0.71
	Sharpe Ratios				
	Data	0.089	0.182	0.242	0.200
	5th percentile	-0.020	-0.024	-0.084	-0.082
	95th percentile	-0.000	-0.011	-0.054	-0.052

Notes: Differences between the out-of-sample performance of skeptical priors (and of the regression estimates) and that of the dogmatic prior. See Table 2 for the description of the performance measures. "Data" denotes the performance difference in the data. The analysis is repeated in samples of simulated data with no predictability. The table displays the 5th and 95th percentiles of the performance difference. Prior beliefs are indexed by $P(R^2 > 2\%)$, the probability that the R^2 from the predictive regression exceeds 2%. $P(R^2 > 2\%) = 0$ corresponds to the dogmatic prior; $P(R^2 > 2\%) = 0.999$ corresponds to the diffuse prior. Reg. denotes results from ordinary least squares regression. A is the risk-aversion parameter.

Table 4: Subperiod Out-of-Sample Results: Dividend-Price Ratio

| 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

Notes: Certainty equivalent returns (CERs) and Sharpe ratios are computed as in Table 2 with 1984 and 1994 as end dates rather than 2004. A is the risk aversion parameter. Prior beliefs are indexed by $P(R^2 > 2\%)$, the probability that the R^2 from the predictive regression exceeds 2%. $P(R^2 > 2\%) = 0$ corresponds to the dogmatic prior; $P(R^2 > 2\%) = 0.999$ corresponds to the diffuse prior. Reg. denotes out-of-sample results obtained from ordinary least squares regression. CERs are in annualized percentages ($\times 400$), and Sharpe ratios are annualized ($\times 2$).

Figure 1: Prior on the R^2 of the Predictive Regression
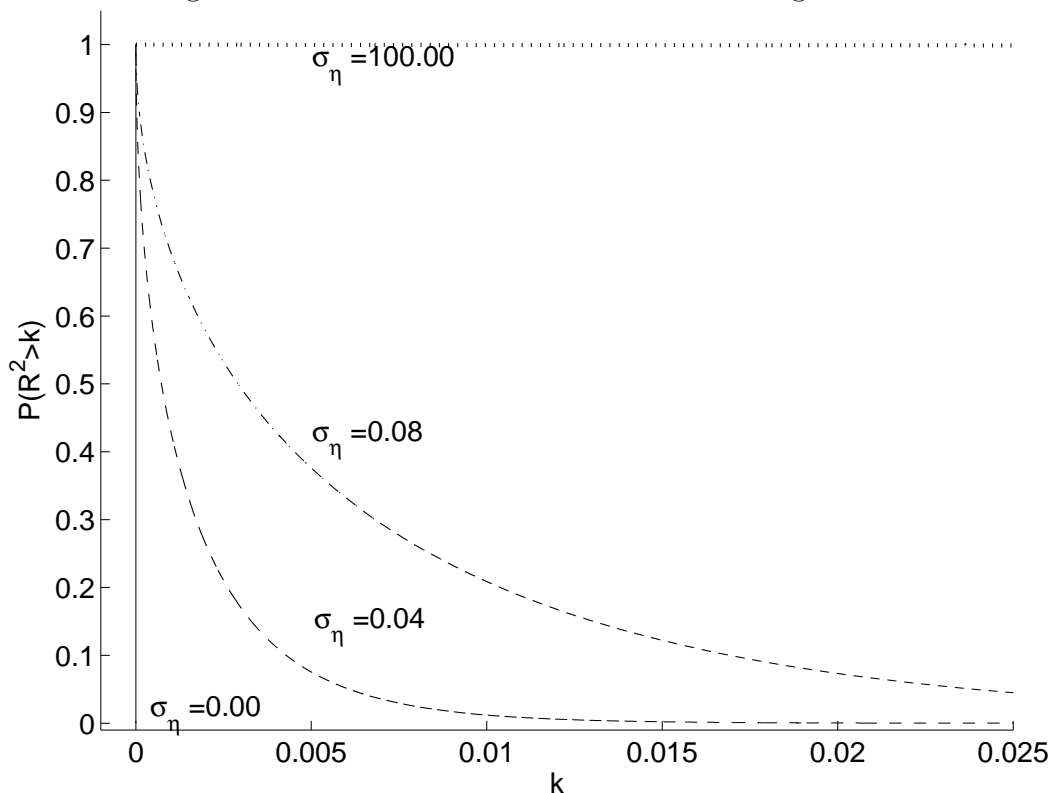
Notes: The prior probability that the R^2 exceeds a value k, for k ranging from 0 to 0.025 implied by skeptical Jeffreys priors. Prior beliefs are indexed by σ_η, the prior standard deviation of the normalized coefficient on the predictor variable. The dogmatic prior is given by $\sigma_\eta = 0$; the diffuse prior by $\sigma_\eta = \infty$. Intermediate priors express some skepticism over return predictability.

Figure 2: Posterior Distribution on the Maximum R^2
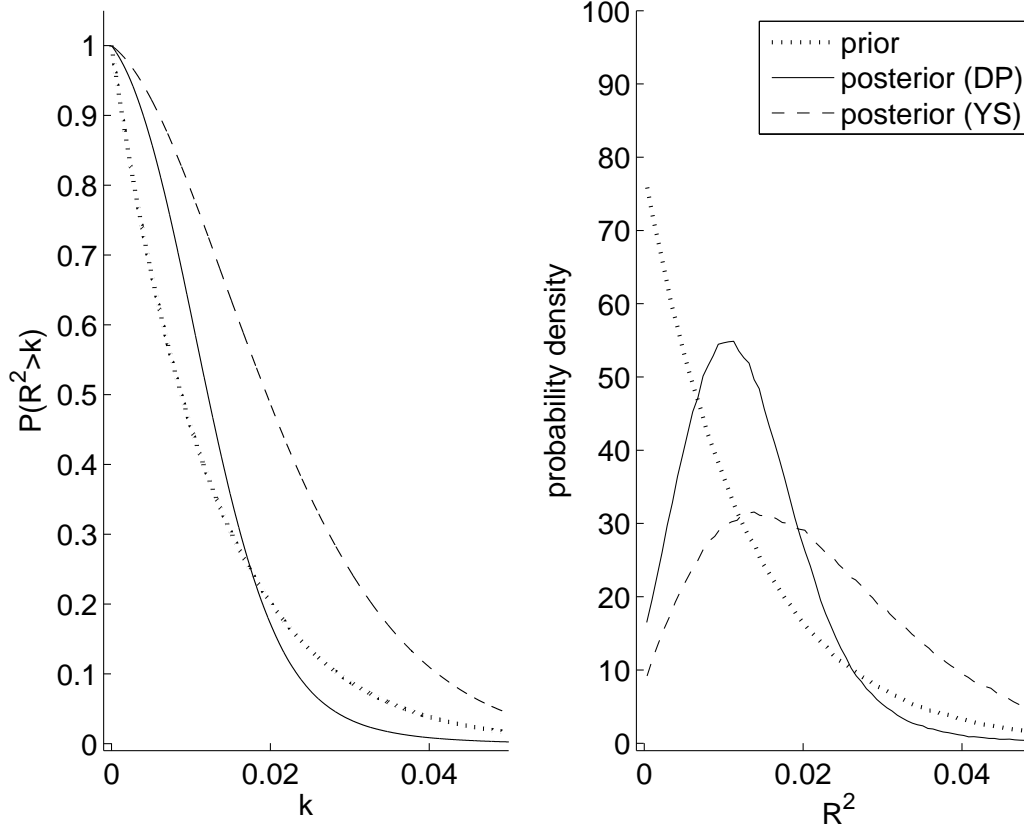
Notes: The left panel shows the prior and the posterior probability that the maximum R^2 over two assets exceeds a value k. The right panel shows the probability density function for the prior and the posterior of the maximum R^2. The prior volatility of normalized β, σ_η, is set equal to 0.08. The assets are a stock index and a long-term bond. The predictor variable is either the dividend-price ratio (solid line) or the yield spread (dashed line). The predictive relation is measured over a quarterly horizon.

Figure 3: Conditional Expected Returns and Holdings when the Dividend-Price Ratio Predicts Returns
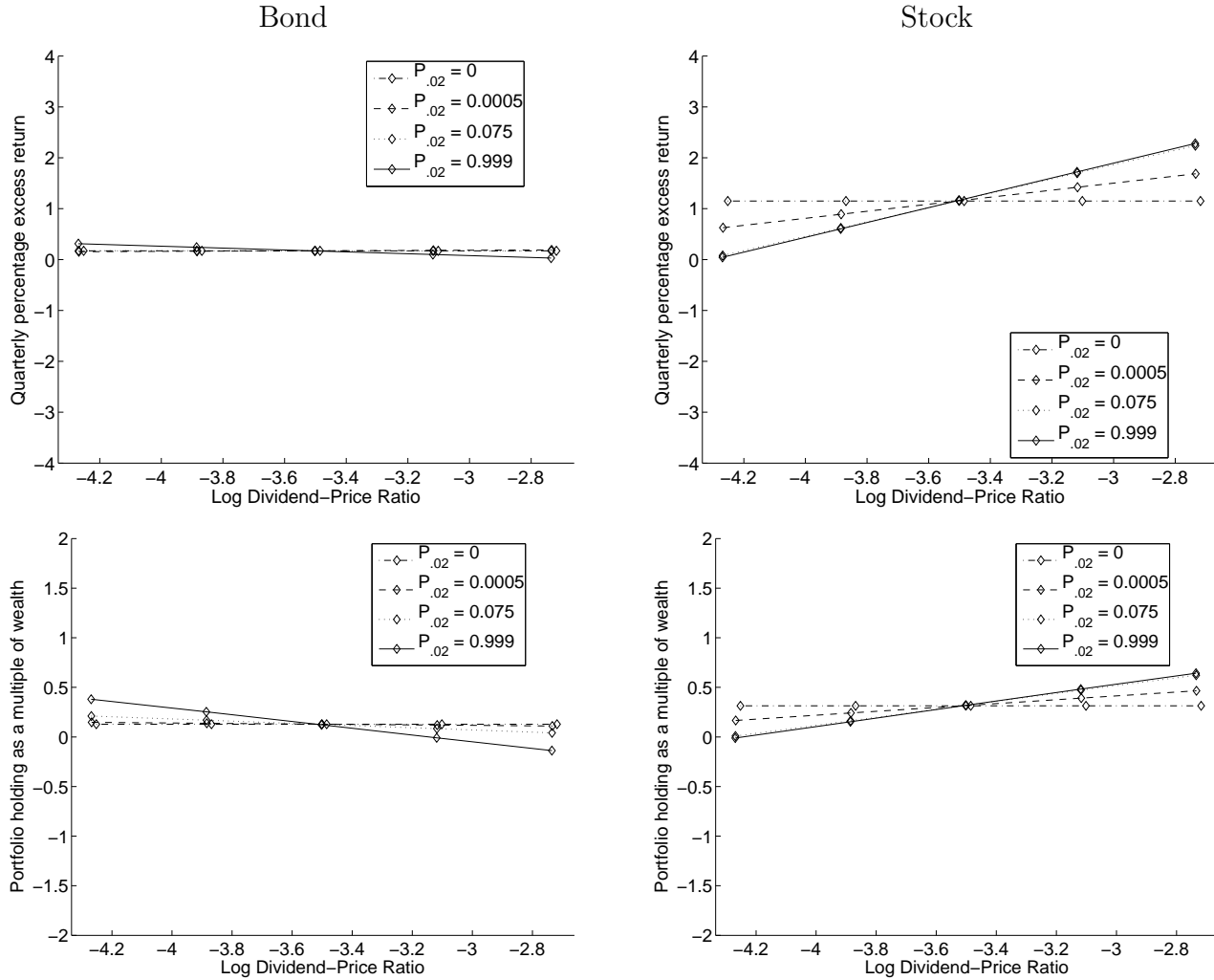

Notes: Conditional expected returns (top two plots) and portfolio holdings (bottom two plots) as functions of the log dividend-price ratio. Conditional expected returns are calculated using the predictive distribution. Given the predictive distribution, portfolios maximize mean-variance utility for risk aversion parameter $A = 5$. Assets are a stock index, a long-term bond, and the riskfree asset (not shown). Prior beliefs are indexed by $P_{.02}$, the probability that the R^2 from the predictive regression exceeds 2%. $P_{.02} = 0$ corresponds to the dogmatic prior; $P_{.02} = 0.999$ corresponds to the diffuse prior. Diamonds correspond to the sample mean and plus and minus one and two sample standard deviations of the predictor variable.

Figure 4: Conditional Expected Returns and Holdings when the Yield Spread Predicts Returns
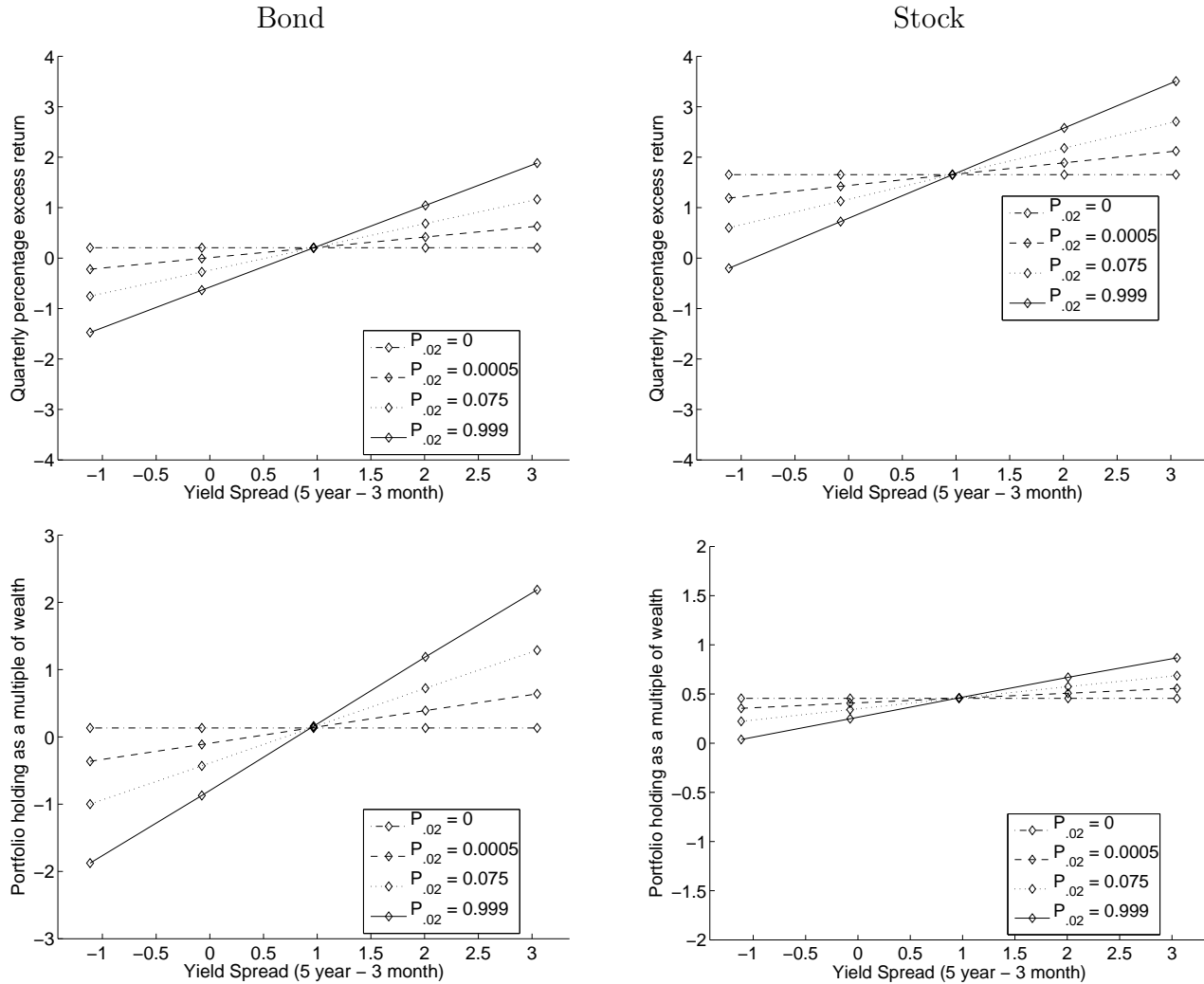
Notes: Conditional expected returns (top two plots) and portfolio holdings (bottom two plots) as functions of the yield spread. Conditional expected returns are calculated using the predictive distribution. Given the predictive distribution, portfolios maximize mean-variance utility for risk aversion parameter $A = 5$. Assets are a stock index, a long-term bond, and the riskfree asset (not shown). Prior beliefs are indexed by $P_{.02}$, the probability that the R^2 from the predictive regression exceeds 2%. $P_{.02} = 0$ corresponds to the dogmatic prior; $P_{.02} = 0.999$ corresponds to the diffuse prior. Diamonds correspond to the sample mean and plus and minus one and two sample standard deviations of the predictor variable.

Figure 5: Time Series of State Variables and Portfolio Holdings for the Diffuse Prior
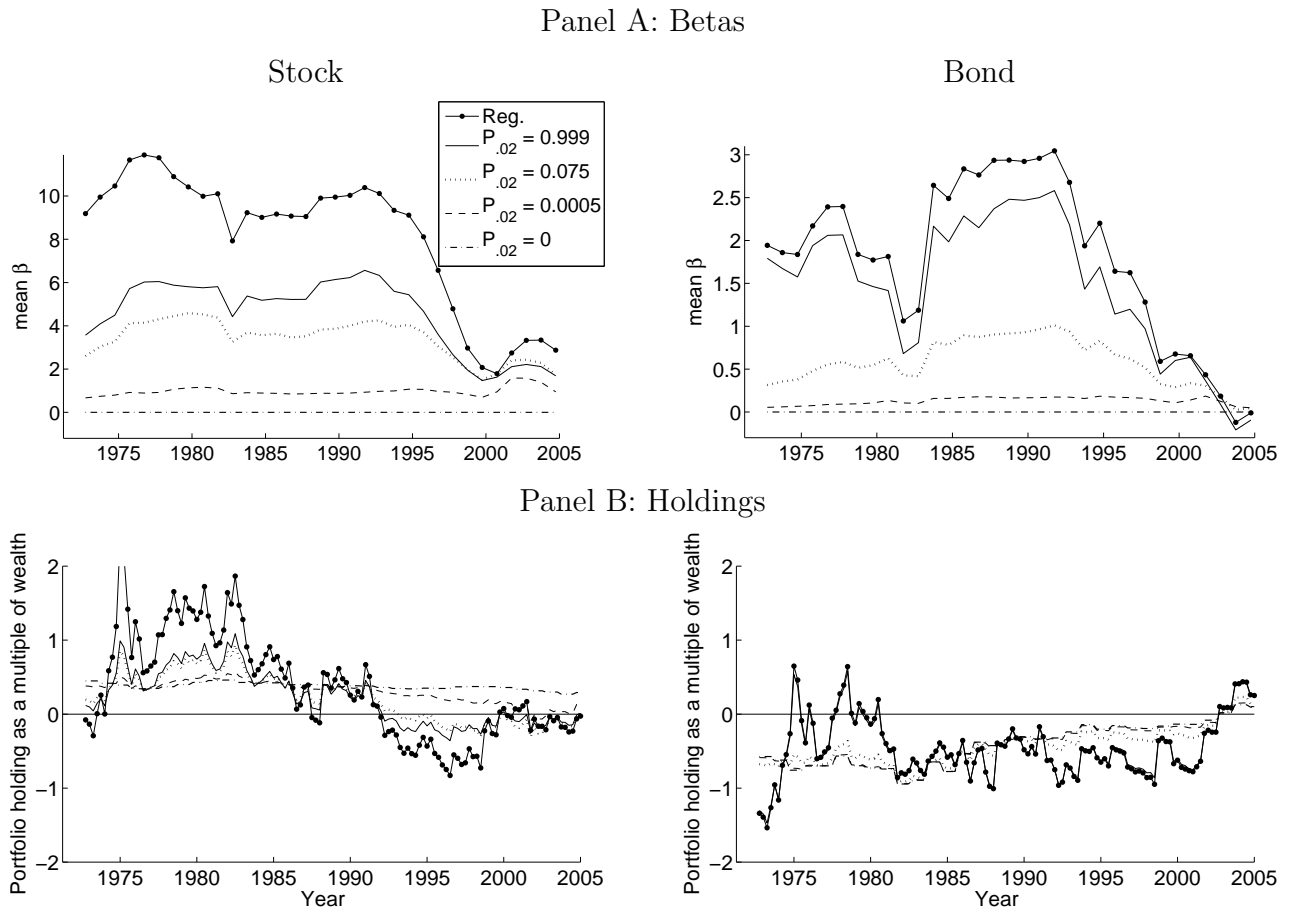
Panel A: Dividend-Price Ratio
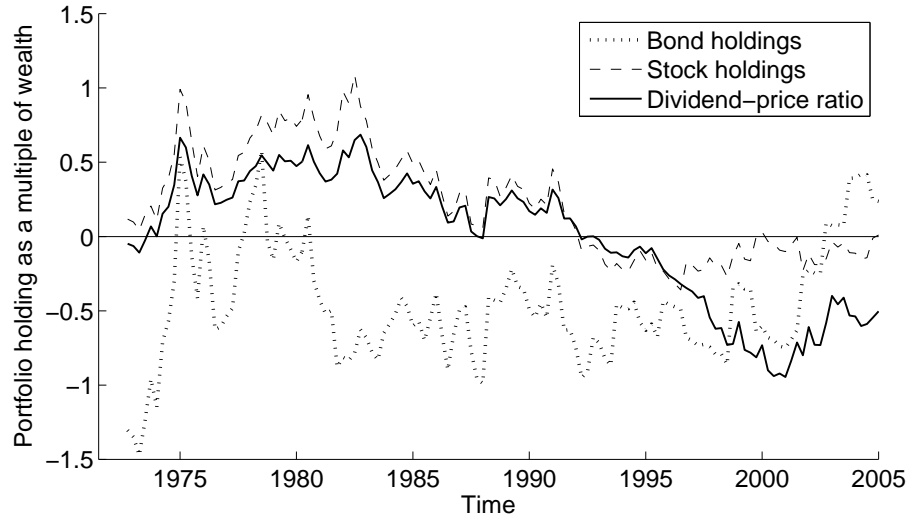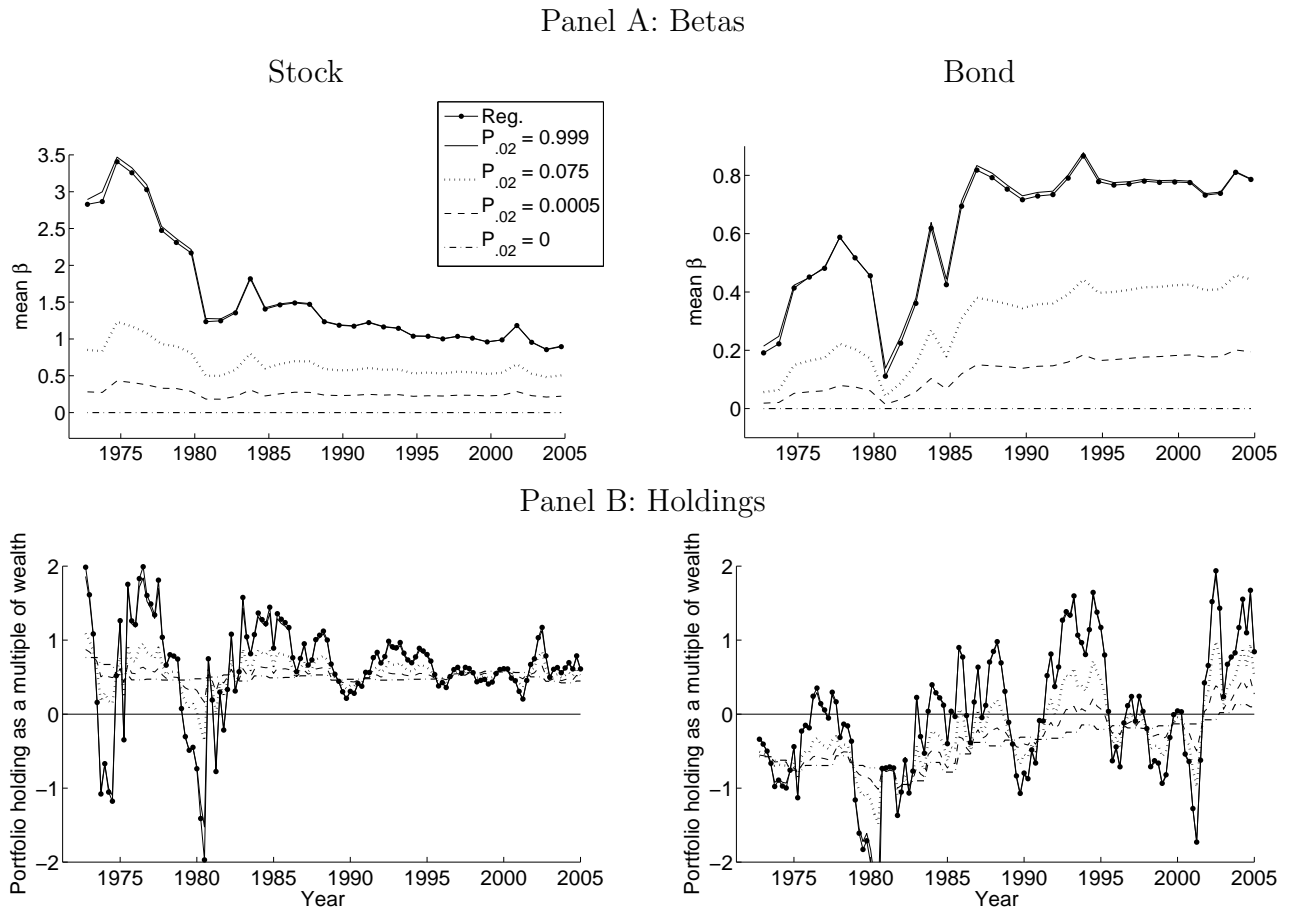
Panel B: Yield Spread
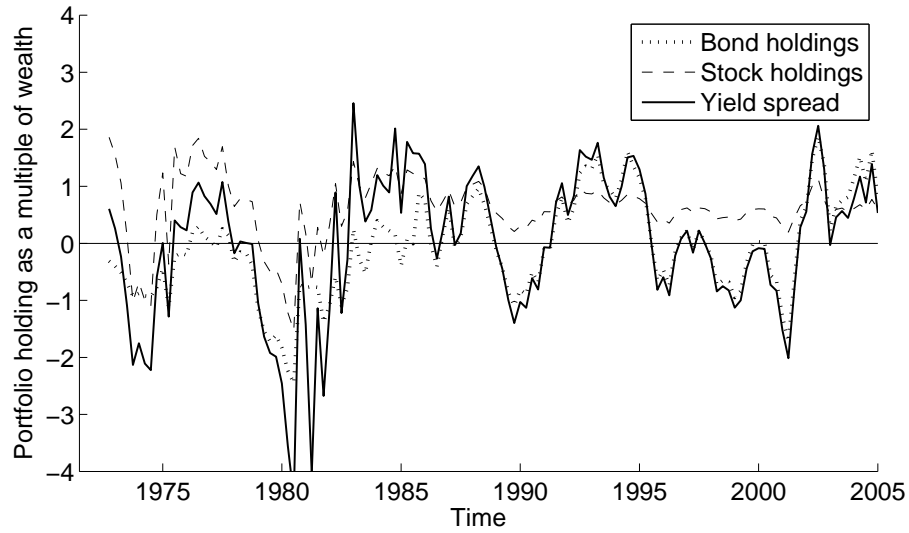
Notes: Time series of optimal portfolio holdings in the long-term (10-year) bond and in the stock index are plotted together with the predictor variable minus its mean. Not shown are holdings in the riskfree asset. For each year beginning in 1972, the predictive distribution for returns is computed using the data up to that year. Given the predictive distribution, the portfolio maximizes mean-variance utility for risk aversion parameter $A = 5$. Panel A shows results assuming that the dividend-price ratios is the predictor variable while Panel B shows results assuming that the yield spread is the predictor variable.

Figure 6: Time Series of Posterior Means and Stock Holdings when the Dividend-Price Ratio Predicts Returns

Notes: Panel A shows the posterior mean of β (the coefficient on the predictor variable for stock returns) for the stock index and the long-term bond, respectively; Panel B shows the corresponding portfolio holdings in the stock index and the long-term bond. For each year beginning in 1972, the posterior distribution is computed using the data up to that year. Optimal portfolios for the mean-variance investor are computed quarterly for risk aversion $A = 5$. Prior beliefs are indexed by $P_{.02}$, the probability that the R^2 from the predictive regression exceeds 2%. $P_{.02} = 0$ corresponds to the dogmatic prior; $P_{.02} = 0.999$ corresponds to the diffuse prior. Also shown are results from ordinary least squares regression (Reg). Note that the left and right graphs in Panel A are on different scales.

Figure 7: Time Series of Posterior Means and Stock Holdings when the Yield Spread Predicts Returns

Panel A: Betas

Panel B: Holdings

Notes: Panel A shows the posterior mean of β (the coefficient on the predictor variable for stock returns) for the stock index and the long-term bond, respectively; Panel B shows the corresponding portfolio holdings in the stock index and the long-term bond. For each year beginning in 1972, the posterior distribution is computed using the data up to that year. Optimal portfolios for the mean-variance investor are computed quarterly for risk aversion $A = 5$. Prior beliefs are indexed by $P_{.02}$, the probability that the R^2 from the predictive regression exceeds 2%. $P_{.02} = 0$ corresponds to the dogmatic prior; $P_{.02} = 0.999$ corresponds to the diffuse prior. Also shown are results from ordinary least squares regression (Reg). Note that the left and right graphs in Panel A are on different scales.