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

Jessica A. Wachter
University of Pennsylvania and NBER

Missaka Warusawitharana
University of Pennsylvania

August 29, 2005

Comments Welcome

*Wachter: Department of Finance, The Wharton School, University of Pennsylvania, 2300 SH-DH, Philadelphia, PA, 19104. jwachter@wharton.upenn.edu, (215)898-7634. Warusawitharana: Department of Finance, The Wharton School, University of Pennsylvania, 2300 SH-DH, Philadelphia, PA, 19104. missaka@wharton.upenn.edu, (215)898-1209. We are grateful to John Campbell, Itamar Drechsler, Bjorn Eraker, Lubos Pastor, Robert Stambaugh, Amir Yaron, Motohiro Yogo, and seminar participants at the 2005 CIRANO-CIREQ Financial Econometrics Conference and at the Wharton School for helpful comments.

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

Abstract

Are excess returns predictable and if so, what does this mean for investors? Previous literature has tended toward two polar viewpoints: that predictability is useful only if the statistical evidence for it is incontrovertible, or that predictability should affect portfolio choice, even if the evidence is weak according to conventional measures. This paper models an intermediate view: that both data and theory are useful for decision-making. We investigate optimal portfolio choice for an investor who is skeptical about the amount of predictability in the data. Skepticism is modeled as an informative prior over the improvement in the Sharpe ratio generated by using the predictor variable. 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, as we show, deliver superior out-of-sample performance compared with weights implied by diffuse priors, dogmatic priors, and ordinary least squares regression.

Introduction

Are excess returns predictable, and if so, what does this mean for investors? Classic studies by Samuelson (1965, 1973) and Shiller (1981) show that in models with rational valuation, returns on risky assets over riskless assets should be constant over time, and thus unpredictable by investors. However, an extensive empirical literature has found evidence for predictability in returns on stocks and bonds by scaled-price ratios and interest rates.[1]

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.[2] 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 that the regression coefficients on the predictor variables are normally distributed around zero. 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 investor is skeptical about predictability: predictability is possible, but it is more likely that predictability is "small" rather than "large".

An important aspect of this study is defining what "small" and "large" mean in the context of predictability. We employ an economic metric: the expected improvement in the squared maximum Sharpe ratio from conditioning portfolio choice on the predictor variable. This statistic is analogous to the Gibbons, Ross, and Shanken (1989) statistic for measuring deviations from the CAPM, and, in the case of a single predictor variable, is closely related to the population R^2 implied by the predictive regression. By carefully specifying the variance of the distribution for the predictor variables, we achieve a prior distribution over this statistic.

In our empirical implementation, we consider returns on a stock index and on a long-term

[1]See, for example, Campbell and Shiller (1988, 1991), Fama and French (1989), Fama and Schwert (1977), Keim and Stambaugh (1986), Kothari and Shanken (1997).

[2]See, for example, Brennan, Schwartz, and Lagnado (1997), 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.

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 an investment perspective. Kandel and Stambaugh (1996) show in a Bayesian framework that predictive relations that are weak in terms of standard statistical measures can nonetheless have large impacts on portfolio choice.[3] Kandel and Stambaugh explore informative priors but adopt an empirical Bayes approach: the prior is formed by viewing data that is equal to the actual data in every way except that there is no predictability. This prior, while analytically convenient, requires knowledge of the entire sample of data. Shanken and Tamayo (2004) model time-varying risk and expected return together in a Bayesian framework. The priors assumed by Shanken and Tamayo are informative, but, like the priors in Kandel and Stambaugh, require knowledge of the entire time series of the predictor variable.[4] Our approach differs from these previous studies in that both stocks and bonds, rather than only a stock index, are considered. More importantly, our approach allows an investor to form a prior on the amount of predictability that is truly "prior", namely that does not require knowledge of the moments of the data.

Besides introducing an informative prior that can be specified without recourse to the data, our study incorporates the findings of Stambaugh (1999) in a setting with informative priors. Stambaugh shows that incorporating the first observation on the predictor variable into the likelihood can make a substantial difference for portfolio choice; previous studies had conditioned

[3]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. Brandt, Goyal, Santa-Clara, and Stroud (2003) develop a simulation-based approach to consider learning about other unknown parameters. Johannes, Polson, and Stroud (2002) model the mean and volatility of returns as latent factors. In contrast to the present study, these papers choose priors to be diffuse.

[4]Bayesian methods have also been shown to be useful for model selection. Avramov (2002) and Cremers (2002) show that Bayesian procedures, combined with informative priors, lead to superior choices of predictor variables when many such combinations of variables are considered. Like the studies mentioned above, these studies make use of informative priors which require conditioning on the entire data set.

on this observation. Moreover, the choice among "uninformative" priors can make a difference as well: a prior that is uninformative in the sense of Jeffreys (1961) has different properties than the priors that have been chosen by previous studies in the portfolio choice literature. Building on the work of Stambaugh, this study also incorporates information contained in the first observation on the predictor variable, and makes use of Jeffreys priors. We show that Jeffreys invariance theory offers an independent justification for defining the prior over the change in the squared maximum Sharpe ratio. As the degree of skepticism goes to zero, the prior satisfies the Jeffreys condition for invariance.

Our use of informative priors has parallels in studies that examine the cross section of stock and mutual fund returns. Pastor and Stambaugh (1999) show how informative beliefs about asset pricing models can be incorporated into calculations of the cost of capital, while Pastor (2000) shows how informative priors can enter into portfolio allocation decisions. Baks, Metrick, and Wachter (2001) incorporate prior beliefs that are skeptical about fund manager skill into the decision problem of investing in actively managed mutual funds. These studies illustrate how informative prior beliefs can help improve decision making when applied to the cross section of returns. Here, we show how related ideas can be applied 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. These sections assume, for simplicity, that there is a single predictor variable. Section 3 extends the methods to allow for multiple predictor variables. Section 4 concludes.

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 Brown (1979)). 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. The likelihood is described in Section 1.1, the prior in Section 1.2, and the posterior in Section 1.3. Section 1.4 describes optimal portfolio choice given the posterior distribution.

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 data on the predictor variable $x_0, \ldots, x_T$. Let

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

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, rather than a single, risky asset introduces little in the way of complication, while allowing multiple predictor variables complicates the problem without contributing to the intuition. For this reason, we postpone the discussion of multiple predictor variables until Section 3.

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. 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 $p(\mathcal{Y}|B, \Sigma, x_0)$ denote the likelihood function. From results in Zellner (1996), it follows that

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

where $\mathrm{tr}(\cdot)$ denotes the sum of the diagonal elements of a matrix.[5]

The likelihood function (4) conditions on the first observation of the predictor variable, x_0. Stambaugh (1999) argues for treating x_0 and $x_1, \ldots, x_T$ symmetrically: as random draws from the data generating process. If the process for x_t is stationary and has run for a substantial period of time, then results in Hamilton (1994, p. 53) imply that x_0 is a draw from a normal distribution with mean

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

and variance

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

Combining the likelihood of the first observation with the likelihood of the remaining T obser-

[5]Maximizing the conditional likelihood function (4) is equivalent to running a vector auto-regression. Stambaugh (1999) points out that the resulting estimates for β are biased; Cavanagh, Elliott, and Stock (1995) show that the t-test for the significance of β has the incorrect size. An active literature in classical statistics focuses on solutions to these problems (e.g. Amihud and Hurvich (2004), Campbell and Yogo (2004), Eliasz (2004), Lewellen (2004), Torous, Valkanov, and Yan (2005)).

vations produces

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

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.

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 possibility of predictability. These informative priors imply that large gains from exploiting predictability are unlikely, but not impossible.

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. To capture this dependence, we specify a hierarchical prior for β. That is, we specify a prior on B, Σ so that the prior for β is conditional on the remaining parameters:

$$p(B,\Sigma) = p(\beta|\alpha,\theta_0,\theta_1,\Sigma)p(\alpha,\theta_0,\theta_1,\Sigma).$$

We further assume that the prior for β conditional on the remaining parameters is multivariate normal:

$$\beta|\alpha,\theta_0,\theta_1,\Sigma \sim N(0,\sigma_\beta^2\sigma_x^{-2}\Sigma_u), \tag{8}$$

where σ_β is a non-negative scalar. Because σ_x is a function of θ_1 and σ_v, the prior on β is also implicitly a function of these parameters. The parameter σ_β indexes the degree to which the prior is informative. We show that as $\sigma_\beta \to \infty$, the prior over β becomes uninformative; all values of β are viewed as equally likely. As $\sigma_\beta \to 0$, the prior converges to a point mass at zero and the prior beliefs assign a probability of 1 to no predictability.

Finite positive values of σ_β involve some skepticism about the amount of predictability in the data. The dependence between β, σ_x, and Σ_u allow this skepticism to be expressed in economic terms, namely as the expected change in the squared maximum Sharpe ratio that results from taking predictability into account. Conditional on B and Σ, the squared maximum Sharpe ratio for an investor who conditions on x_t is equal to

$$(\alpha + \beta x_t)^\top \Sigma_u^{-1} (\alpha + \beta x_t).$$

The squared maximum Sharpe ratio for an investor who does not condition on x_t is equal to

$$(\alpha + \beta \bar{x})^\top \Sigma_u^{-1} (\alpha + \beta \bar{x})$$

(see Campbell, Lo, and MacKinlay (1997, pp. 184–188)). Therefore the expected gain from incorporating information on x_t equals

$$\begin{aligned}
\Delta(\mathrm{SR})^2 &= E\left[(\alpha + \beta x_t)^\top \Sigma_u^{-1}(\alpha + \beta x_t) - (\alpha + \beta \bar{x})^\top \Sigma_u^{-1}(\alpha + \beta \bar{x}) \mid B, \Sigma\right] \\
&= E\left[(\beta x_t)^\top \Sigma_u^{-1}(\beta x_t) - (\beta \bar{x})^\top \Sigma_u^{-1}(\beta \bar{x}) \mid B, \Sigma\right] \\
&= E\left[(\beta x_t)^\top \Sigma_u^{-1}(\beta x_t) - 2(\beta \bar{x})^\top \Sigma_u^{-1}(\beta x_t) + (\beta \bar{x})^\top \Sigma_u^{-1}(\beta \bar{x}) \mid B, \Sigma\right] \\
&= E\left[(x_t - \bar{x})^\top \beta^\top \Sigma_u^{-1} \beta (x_t - \bar{x}) \mid B, \Sigma\right] \\
&= \sigma_x^2 \beta^\top \Sigma_u^{-1} \beta. \qquad (9)
\end{aligned}$$

The first three lines follow from the definition of $\bar{x}$ and the fourth line follows because x_t is scalar.[6] Because (8) implies that $\sigma_x C_u^{-1} \beta \sim N(0, I_N)$, where C_u is such that $C_u C_u^\top = \Sigma_u$, the change in the squared maximum Sharpe ratio has a scaled chi-squared distribution with N degrees of freedom.[7]

Equation (9) is a useful summary statistic for the amount of predictability in the data. As is the case for the Gibbons, Ross, and Shanken (1989) statistic, the gain in the squared maximum Sharpe ratio is quadratic. This is because both positive βs and negative βs are helpful for the investor. For example, if the predictor variable is positive, negative βs indicate that the investor should allocate less to the risky assets than he would otherwise, positive βs indicate that he should

[6]Section 3 extends this calculation to the case of multiple predictor variables.

[7]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 (2004) 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.

allocate more. Either way he experiences a gain in the squared maximum Sharpe ratio relative to assuming the βs are equal to zero. Pre- and post-multiplying Σ_u^{-1} by the βs implies that the investor does not downweight predictability on specific assets per se, but on predictability on the mean-variance efficient portfolio. It is not the amount of predictability in individual assets that matters, but predictability in "the best" portfolio; this is the same intuition for choosing Σ_u^{-1} as a weighting matrix for pricing errors in the Gibbons, Ross, and Shanken statistic. Finally, the gain depends on σ_x: the greater is σ_x, the greater the gain because there is a greater range of outcomes for the predictor variable.[8]

For the single-asset case, (9) is closely related to the R^2, or the percent variance of the return that is explained by the predictor variable. The population R^2 is equal to

$$R^2 = \beta^2\sigma_x^2\left(\beta^2\sigma_x^2 + \Sigma_u\right)^{-1},$$

where Σ_u is now a scalar. It follows from (9), that

$$R^2 = \frac{\Delta(\mathrm{SR})^2}{\Delta(\mathrm{SR})^2 + 1}. \tag{10}$$

Thus when $N = 1$, placing a prior distribution on the change in the squared maximum Sharpe ratio is equivalent to placing a prior distribution on the R^2 of the regression (given a draw from $\Delta(\mathrm{SR})^2$, a draw from the R^2 can be obtained from (10)). In the case of $N > 1$, our prior is equivalent to assuming a joint distribution on the R^2 of each of the return equations. In particular, (8) implies that each component of β is normally distributed with variance determined by the appropriate diagonal element of Σ_u. The previous argument can then be applied to individual components of β.

It is instructive to compare our choice of prior for β to 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. Baks, Metrick, and Wachter (2001) 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

[8] The idea of limiting Sharpe ratios relates to the work of Cochrane and Saa-Requejo (2000), who show how limiting the Sharpe ratio available from trading options puts bounds on option prices. Here, we impose a distribution that makes large increases in the Sharpe ratio unlikely.

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, leading to a very high R^2 and Sharpe ratio improvement. 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 the Sharpe ratio improvement and the R^2.

For the remaining parameters, we choose a prior that is uninformative in the sense of Jeffreys (1961). Jeffreys argues that a reasonable property of a "no-information" prior is that inference be invariant to one-to-one transformations of the parameter space. Given a set of parameters μ, data $\mathcal{Y}$, and a log-likelihood $l(\mu; \mathcal{Y})$, Jeffreys shows that invariance is equivalent to specifying a prior as

$$p(\mu) \propto \left| -E\left(\frac{\partial^2 l}{\partial\mu\partial\mu^\top}\right)\right|. \tag{11}$$

Besides invariance, this formulation of the prior has other advantages such as minimizing asymptotic bias and generating confidence sets that are similar to their classical counterparts (see Phillips (1991)).[9]

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

$$p(\alpha, \theta_0, \theta_1, \Sigma) \propto \sigma_x |\Sigma_u|^{1/2} |\Sigma|^{-\frac{N+4}{2}}, \tag{12}$$

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

$$\begin{aligned} p(B, \Sigma) &= p(\beta|\alpha, \theta_0, \theta_1, \Sigma) p(\alpha, \theta_0, \theta_1, \Sigma) \\ &\propto \sigma_x^{N+1} |\Sigma|^{-\frac{N+4}{2}} \exp\left\{-\frac{1}{2}\beta^\top \left(\sigma_\beta^2 \sigma_x^{-2} \Sigma_u\right)^{-1} \beta\right\}. \end{aligned} \tag{13}$$

(note that σ_β is a constant).

[9]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. Phillips (1991) derives an exact Jeffreys (1961) prior and shows that the inference with this prior leads to different conclusions than inference with a prior that is flat for the regression coefficients. As a full investigation of these issues is outside the scope of this study, we focus on the Jeffreys prior and explore the robustness of our conclusions to other interpretations of "uninformativeness" in Appendix D.

Jeffreys invariance theory provides an independent justification for modeling priors on β as (8). Appendix A shows that the limiting Jeffreys prior for B and Σ equals

$$p(B,\Sigma) \propto |\Sigma_x|^{\frac{N+1}{2}} |\Sigma|^{-\frac{N+4}{2}}. \tag{14}$$

This prior corresponds to (13) as σ_β approaches infinity. Modeling the prior for β as depending on σ_x not only has an interpretation in terms of the squared maximum Sharpe ratio, 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.

Figure 1 plots the cumulative density function of the change in the squared maximum Sharpe ratio for prior beliefs corresponding to $\sigma_\beta = .04$ and $\sigma_\beta = .08$ for one and two assets. While the two-asset case is the focus of this paper, the one-asset case is of interest because the prior distribution for the R^2 of each of the regressions is almost identical to the prior distribution of the change in the squared maximum Sharpe ratio for $N = 1$.[10] As Figure 1 shows, for $\sigma_\beta = .04$, the investor assigns only a 4% chance to a change in the squared maximum Sharpe ratio over 0.01, and less than a 1% chance of an R^2 of over 0.01. Thus this prior is quite skeptical. A slightly less skeptical prior is given by $\sigma_\beta = .08$. This prior assigns a 50% probabilility to a change in the squared maximum Sharpe ratio over .01, and a 20% probability to an R^2 of greater than .01.

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

$$p(B,\Sigma|\mathcal{Y}) \propto p(\mathcal{Y}|B,\Sigma)p(B,\Sigma),$$

where $p(\mathcal{Y}|B,\Sigma)$ is the likelihood and $p(B,\Sigma)$ is the prior. Substituting in the prior (13) and the likelihood (7) produces

$$p(B,\Sigma|\mathcal{Y}) \propto \sigma_x^N |\Sigma|^{-\frac{T+N+4}{2}} \exp\left\{-\frac{1}{2}\beta^\top \left(\sigma_\beta^2 \sigma_x^{-2}\Sigma_u\right)^{-2} \beta\right\} \exp\left\{-\frac{1}{2}\sigma_x^{-2}(x_0-\bar{x})^2\right\}$$
$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left[(Y-XB)^\top (Y-XB)\Sigma^{-1}\right]\right\} \tag{15}$$

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 a nonlinear

[10] This follows from (10), because when x is small, $x/(1+x) = 1/(1+1/x) \approx 1/(1/x) = x$.

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

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

The Metropolis-Hastings algorithm is implemented "block-at-a-time", by first sampling from $p(\Sigma|b, \mathcal{Y})$, then $p(b_1|b_2, \Sigma, \mathcal{Y})$, and finally $p(b_2|b_1, \Sigma, \mathcal{Y})$. The proposal density for the conditional probability of Σ is the inverted Wishart with $T + 2$ degrees of freedom and scale factor of $(Y - XB)(Y - XB)^\top$. 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 (15). 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 C. 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 the posterior determines portfolio choice. Consider an investor who maximizes expected utility at time $T + 1$ conditional on information available at time T. The investor solves

$$\max E_T[U(W_{T+1})|\mathcal{Y}] \tag{16}$$

where

$$W_{T+1} = W_T[w_T^\top r_{T+1} + r_{f,T}],$$

w 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 (16) is taken with respect to the predictive distribution

$$p(r_{T+1}|\mathcal{Y}) = \int p(r_{T+1}|x_T, B, \Sigma)p(B, \Sigma|\mathcal{Y})\, dB\, d\Sigma. \tag{17}$$

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. It therefore allows us to illustrate the implications of our methodology in a particularly clear way. 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 (17). For a quadratic-utility investor, optimal weights w^* in the N assets are given by

$$w^* = \frac{1}{A}\tilde{V}^{-1}\tilde{E}, \tag{18}$$

where A is a parameter determining the investor's risk aversion. The weight in the riskless bond is equal to $1 - \sum_{i=1}^{N} w_i^*$.

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 (18), with the mean and variance computed by simulating draws r^j.

2 Results

To illustrate the methods described in the previous section, we consider the problem of a quadratic utility investor who allocates wealth between a riskfree 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)). This section focuses on the priors described in Section 1.2. To verify the robustness of our results, we also investigate a second specification of "uninformativeness" for the parameters α, θ_0, θ_1, and Σ, described in Appendix D. As Appendix D shows, the implications of this second set of priors are very similar to the implications of the priors in Section 1.2.

The data are described in Section 2.1. Section 2.2 describes aspects of the posterior distribution, and Section 2.3 examines expected returns and portfolio weights. Results in Sections 2.2 and 2.3 condition on the full data set. Section 2.4 describes the time series of posterior means and portfolio weights in post-war data implied by conditioning only on the data observed up until each quarter. Section 2.5 performs out-of-sample analyses.

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

We first quantitatively describe 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. Table 1 reports posterior means implied by the exact likelihood (7) and the prior (13), for values of σ_β ranging from zero to infinity. The predictor variable is the dividend-price ratio. Posterior standard deviations are reported in parentheses. The table also shows results from estimating (1)–(3) by ordinary least squares (OLS). For the OLS values, standard errors are reported in parentheses.

As 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 values of $\sigma_\beta > 0$, and for the OLS estimate. These posterior means are consistent with the classic findings of Campbell and Shiller (1988) and Fama and French (1989) that the dividend-price ratio predicts stock returns with a positive sign. 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 $\sigma_\beta = .08$, the estimate is 1.41, while for $\sigma_\beta = .04$, it is 0.69.

Table 1 also reports posterior means and standard deviations of "unconditional" means of bond and stock returns. That is, the table reports

$$E\left[E[r_{t+1}|B,\Sigma]|\mathcal{Y}\right] = E\left[\alpha + \beta\frac{\theta_0}{1-\theta_1}|\mathcal{Y}\right]. \tag{19}$$

This can also be thought of as the long-run mean of the asset return. 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.

Table 1 shows that while the posterior means are virtually identical across the range of priors, the unconditional means implied by OLS are strikingly different. For the long-term bond, the posterior mean is about 0.18% per quarter for all values of the prior, while the OLS value is 0.23%. For the stock index, the posterior mean equals 1.17%, while the OLS value is 1.09%. In order to understand the discrepancy between the mean implied by OLS and the posterior mean, it is helpful to separate the discrepancy into two parts: the difference between the OLS mean and the sample mean (1.67% for the stock index in this time period), and the difference between the sample mean and the posterior mean. 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 posterior mean and the sample mean is less mechanical. The posterior mean lies below the sample mean because of the x_0 term in the exact likelihood (7). Because the dividend-price ratio in 1952 is above its conditional maximum likelihood estimate, the exact likelihood function adjusts the mean of the dividend-price ratio slightly upward. Because of the negative correlation between the stock return and the dividend-price ratio, the mean return is adjusted downward. 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. Interestingly, this effect is not connected with the ability of the dividend-price ratio to predict stock returns, as it operates equally for all values of the prior.

Table 2 repeats this analysis when the yield spread is the predictor variable. The yield spread predicts both bond and stock returns with a positive sign. Unlike for the dividend-price ratio, OLS and the most diffuse prior lead to very similar results. This is because the yield spread is less correlated with returns than the dividend-price ratio, and because it is less persistent. Both of these characteristics of the dividend yield cause bias in the case of the OLS regressor (Stambaugh (1999)). The posterior mean for β for the most dogmatic prior is, not surprisingly, zero. As the prior becomes more diffuse, the posterior mean of β goes from from 0 to the OLS estimate. As in the case of the dividend-price ratio, the posterior mean of long-run expected returns, the persistence, and the long-run mean of the x_t are nearly the same, regardless of the informativeness of the prior.

This section has shown the effect of different priors on the posterior means of β. The following section examines the predictive distribution of returns and the implications for portfolio choice.

2.3 Conditional returns and portfolios from the full sample

Figure 2 plots expected excess returns (top two plots) and optimal portfolio holdings (bottom two plots) as functions of the log dividend-price ratio. Each plot shows results for $\sigma_\beta = 0, .04, .08$ and for the diffuse prior. The graphs are centered at the sample mean. Diamonds denote plus and minus one and two sample standard deviations of the dividend-price ratio.

The linear form of (1) implies that expected returns are linear in the predictor variables, conditional on the past data. Let $\bar{\alpha}$ and $\bar{\beta}$ denote the posterior means of α and β respectively. Then the posterior mean equals

$$\tilde{E} \;=\; E\left[\alpha + \beta x_t + u_{t+1} | \mathcal{Y}\right] = E\left[\alpha | \mathcal{Y}\right] + E\left[\beta | \mathcal{Y}\right] x_t = \bar{\alpha} + \bar{\beta} x_t$$

The slope of the relation between the conditional return and x_t is therefore the posterior mean of β. Figure 2 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 -1% per quarter to 3% 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 2 shows that the weight on the stock index also increases in the dividend-price ratio. The relation is very nearly linear. The linear relation is not exact because the predictive variance changes slightly with x_t.[11] 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, will result in an optimal portfolio that puts less weight on the bond.

For the diffuse prior, weights on the dividend-price ratio 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,

[11] The predictive variance is equal to

$$\tilde{V} = E\left[\left((\alpha - \bar{\alpha}) + (\beta - \bar{\beta})x_t + u_{t+1}\right)^2\right],$$

which in principle depends on x_t. Empirically this effect turns out to be very small.

expected returns and weights both vary less. However, this change happens quite slowly. Conditional expected returns under a prior with $\sigma_\beta = .08$ are nearly identical to conditional expected returns with a diffuse prior. Given that this prior assigned only a 20% probability of an increase in the maximum squared Sharpe ratio over .02, this prior could be considered skeptical. Yet there is sufficient evidence in the data to convince this investor to vary her portfolio to nearly the same degree as an investor with no skepticism at all. For a more skeptical prior with $\sigma_\beta = .04$, differences begin to 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. Not surprisingly, with a dogmatic no-predictability prior, the relation between expected returns and the log dividend-price ratio is completely flat, and weights are a constant function of x_t. For this prior, the level of the conditional expected return is also equal to the level of the expected return unconditional on x_t, which, as described above, is lower than the sample mean of the return.

Figure 3 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 goes between -2 and +2 long-run 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 3 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 prior with $\sigma_\beta = .08$ is substantially different than inference based on a diffuse prior. Skepticism reduces portfolio timing more for the yield spread than for the dividend-price ratio, even though the classical evidence in favor of the yield spread is, if anything, stronger than that for the dividend-price ratio (see, e.g., Ang and Bekaert (2003)). Nonetheless, even the investors with skeptical priors ($\sigma_\beta = .08, .04$) choose portfolios that vary with the yield spread.

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

We next describe the implications of various prior beliefs for optimal weights over the postwar period. Starting with the first quarter of 1972, we compute the posterior (15) conditional on having observed data up to and including that quarter. The posterior is computed by simulating 500,000 draws and dropping the first 50,000. The assets are the stock, the long-term bond, and a riskfree asset. We consider prior beliefs with $\sigma_\beta = 0, .04, .08$, and ∞.

Figure 4 plots the weights in the long-term bond and the stock, along with the de-meaned dividend-price ratio for the most diffuse prior. For most of the sample, the weights in the stock are highly positively correlated with the dividend-price ratio. Less correlation is apparent for bond returns. 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 4 shows, under diffuse priors, portfolio weights are highly variable and often extreme.

Figure 5 offers another perspective on the relation between the predictor variable and the allocation. The top panel of the figure shows the posterior mean of β for the stock index, and the bottom panel shows the posterior mean of β for the bond. The posterior means are shown for priors ranging from dogmatic to diffuse (σ_β ranging from 0 to ∞) and for the point estimates of β from OLS. The top panel shows that for stock returns, the OLS beta lies above the posterior mean for the entire sample. The OLS estimates, the posterior mean when $\sigma_\beta = \infty$, and the posterior mean when $\sigma_\beta = 0.08$ decline around 1995, and then rise again around 2000, but do not reach their former levels. The posterior mean for $\sigma_\beta = .08$ 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 expected change in the maximum squared Sharpe ratio. 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 $\sigma_\beta = .04$ remains steady throughout the sample and actually increases after 2000.

Figure 6 plots holdings in the bond and the stock for a range of beliefs about predictability. Also plotted are holdings resulting from estimating (1) – (3) using ordinary least squares regression. Volatility in holdings for both the bond and 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 weight on the bond for the fully dogmatic prior is more volatile. In both cases, any volatility in the weight is due to changes in the parameter estimates, as the predictor variable plays no role in portfolio allocation for the dogmatic prior. The prior that is close to dogmatic, $\sigma_\beta = .04$, 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, $\sigma_\beta = .08$ 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 by far the greatest fluctuations are those arising from ordinary least squares regression.

Figures 7, 8 and 9 show the results of repeating this analysis for the yield spread. Figure 7 plots the weights in the long-term bond and the stock, along with the de-meaned yield spread. Figure 7 shows that weights in both the bond and the stock are clearly positively correlated with the yield spread, as also reflected in Figure 3, and are highly variable over time. This variation takes place at a higher frequency than the variation for the dividend-price ratio, as a result of the lower auto-correlation of the yield spread.

Figure 8 plots the posterior means of βs for priors with $\sigma_\beta = 0, .04, .08, \infty$, and the point estimates of β from OLS. The βs on the yield spread display considerably more stability than the βs on the dividend-price ratio. For all non-dogmatic priors, the posterior means of β are positive for the entire sample and remain largely unchanged following the mid-1980s.

Figure 9 shows portfolio weights for a range of priors, and includes the weights implied by ordinary least squares regression. The weight on the bond for the most dogmatic prior rises over the sample, as the investor updates his beliefs about the parameters for the bond. In contrast, the parameters for the stock stay approximately constant. The non-dogmatic priors all display higher frequency movements associated with changes in the yield spread. These changes while noticeable, are much less dramatic for the skeptical prior beliefs than for the diffuse prior beliefs. In particular, the most diffuse prior beliefs indicate an allocation to the long-term bond that is greater than 100% several times over the sample. However, the allocations for the skeptical priors never rise above 100%. For the yield spread, the weights implied by an ordinary least squares regression are nearly identical to the weights implied by the diffuse prior.

Figures 6 and 9 demonstrate 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.

2.5 Out-of-sample performance

The previous analysis shows how skeptical priors can inform portfolio selection. The results in the previous section show that even highly skeptical investors choose time-varying weights, but 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 the priors that are previously considered. The goal is not to find which prior is "correct", or to prove definitively whether predictability is present or not. Rather, the purpose of this section is to relate the findings of this paper to recent critiques of the predictability evidence (see, e.g., Goyal and Welch (2004)).

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 riskfree 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. In this analysis, the mean and variance in (20) are computed using the sample mean and variance that result from following strategies associated with given priors beliefs. 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 (20). In reporting the certainty equivalent returns, we multiply by 400 to express the return as an annual percentage. Because of the high variability of expected returns and weights on the long-term bond, we also report certainty equivalent returns on the portfolio that only allocates wealth between the stock and the riskfree rate. 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 in the paragraph above. In reporting Sharpe ratios, we multiply by 2 to annualize.

Table 3 reports CERs and Sharpe ratios when the dividend-price ratio is the predictor variable. For both metrics, and for both types of portfolios (bonds and stocks, or stocks only), the weights implied by ordinary least squares deliver worse performance than the weights implied by the dogmatic prior. A similar result is found in Goyal and Welch (2004), who argue against the use of predictability in portfolio allocation. Moreover, the OLS weights perform worse than all of the priors, with the most diffuse prior featuring the next-worst performance. We find, however, that predictability can increase out-of-sample performance, if the investor treats the evidence with some skepticism: the intermediate prior of $\sigma_\beta = .04$ has the best out-of-sample performance across both the CER and Sharpe ratio metrics, regardless of the level of risk aversion, or whether stocks, or stocks and bonds are in the portfolio. While it is the case that ignoring the predictability evidence results in better performance than applying a diffuse prior, simply looking at these two extreme positions hides the better performance that can be achieved by taking the intermediate view.

Table 4 reports analogous results for the yield spread as the predictor variable. When the certainty equivalent return metric is considered, the dogmatic prior outperforms the OLS estimates and the most diffuse prior. For the Sharpe ratio, the OLS estimates outperform the dogmatic prior, but the diffuse prior outperforms both. However, as in the case of the dividend-price ratio, the best performance occurs with the intermediate priors.[12] Both $\sigma_\beta = .04$ and $\sigma_\beta = .08$ deliver superior out of sample performance when compared to the diffuse and dogmatic priors. The differences can be large. When the risk aversion parameter equals 2, for example, a diffuse prior loses money (the CER is equal to -7% per annum). The dogmatic prior does better, delivering a CER of 5.27% per annum. However, for the intermediate prior of $\sigma_\beta = .08$, the CER is 7.58% per annum. Thus priors indicating skepticism about the degree of predictability lead to less extreme portfolio allocations, and to more stable parameter estimates and superior out-of-sample performance over the postwar period.

3 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 $K \times K$. Let Θ be the set of $K \times K$ matrices θ_1 such that the eigenvalues of θ_1 are less than one in absolute value. Define

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

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

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

and

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

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 hierarchical prior for β. In the case of multiple predictor variables, β is a matrix. It is convenient to stack the elements of β into a vector $\phi = \mathrm{vec}(\beta^\top)$.

[12] Across the table, the one exception is for the Sharpe ratio with the stock only portfolio. Here, the diffuse prior performs the best.

The prior for β conditional on the remaining parameters is given by

$$\phi \,|\, \alpha, \theta_0, \theta_1, \Sigma \sim N\left(0, \sigma_\beta \Sigma_u \otimes \Sigma_x^{-1}\right). \tag{23}$$

This prior for β also implies a chi-squared distribution for the improvement in the squared maximum Sharpe ratio. Let C_x be a lower triangular matrix such that $C_x C_x^\top = \Sigma_x$, and

$$\eta = \left(C_u^{-1} \otimes C_x^\top\right)\phi.$$

(recall that C_u is such that $C_u C_u^\top = \Sigma_u$). Then $\eta \sim N(0, \sigma_\beta I_{NK})$, and

$$\begin{aligned}
\Delta(\mathrm{SR})^2 &= E\left[(x_t - \bar{x})^\top \beta^\top \Sigma_u^{-1} \beta (x_t - \bar{x}) \,|\, B, \Sigma\right] \\
&= E\left[\mathrm{tr}\left((x_t - \bar{x})^\top \beta^\top \Sigma_u^{-1} \beta (x_t - \bar{x})\right) \,|\, B, \Sigma\right] \\
&= E\left[\mathrm{tr}\left(\beta^\top \Sigma_u^{-1} \beta (x_t - \bar{x})(x_t - \bar{x})^\top\right) \,|\, B, \Sigma\right] \\
&= \mathrm{tr}\left(\beta^\top \Sigma_u^{-1} \beta E\left[(x_t - \bar{x})(x_t - \bar{x})^\top \,|\, B, \Sigma\right]\right) \\
&= \mathrm{tr}(\beta^\top \Sigma_u^{-1} \beta \Sigma_x) \\
&= \phi^\top (\Sigma_u^{-1} \otimes \Sigma_x)\phi = \eta^\top \eta.
\end{aligned}$$

The second line follows because $(x_t - \bar{x})^\top \beta^\top \Sigma_u^{-1} \beta (x_t - \bar{x})$ is a scalar, the third line follows from the fact that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, and the fourth line follows from the linearity of the expectations operator.

The framework above assumes that the prior standard deviation of the orthogonalized βs, σ_β, is the same for all predictor variables. It is possible to relax this assumption, at the cost of a slightly more complicated specification. The definition of C_x implies

$$C_x^{-1} x_t | B, \Sigma \sim N(0, I_K).$$

Fix a choice of C_x that is lower triangular. The first element of $C_x^{-1} x_t$ is proportional to the first predictor variable, the second element is proportional to the component of the second predictor variable uncorrelated with the first, and so on. Define a diagonal matrix

$$\Sigma_\beta = \begin{pmatrix} \sigma_{\beta,1} & & \\ & \ddots & \\ & & \sigma_{\beta,K} \end{pmatrix},$$

and let

$$\phi \,|\, \alpha, \theta_0, \theta_1, \Sigma \sim N\left(0, \Sigma_u \otimes \left[\left(C_x^{-1}\right)^\top \Sigma_\beta C_x^{-1}\right]\right). \tag{24}$$

The elements of Σ_β can then be interpreted with reference to the ordering of the predictor variable. For each k, setting $\sigma_{\beta,k} = 0$ implies a dogmatic prior on the effect of adding a kth predictor variable, assuming that predictor variables $1, \ldots k-1$ are already in the regression, and that variables $k+1, \ldots K$ are not.

Specification (23) is a special case of (24), where all the diagonal elements of Σ_β are equal. For this reason, we focus on (24) for the remainder of this section. The probability density function corresponding to (24) is given by

$$p(\phi \,|\, \alpha, \theta_0, \theta_1, \Sigma) \propto |\Sigma_x|^{N/2} |\Sigma_u|^{-K/2} \exp\left\{-\frac{1}{2}\phi^\top \left(\Sigma_u \otimes \left[\left(C_x^{-1}\right)^\top \Sigma_\beta C_x^{-1}\right]\right)^{-1} \phi\right\}.$$

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

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

Therefore, the joint prior is given by

$$p(B, \Sigma) \propto |\Sigma_x|^{\frac{N+K}{2}} |\Sigma|^{-\frac{N+2K+2}{2}} \exp\left\{-\frac{1}{2}\phi^\top \left(\Sigma_u \otimes \left[\left(C_x^{-1}\right)^\top \Sigma_\beta C_x^{-1}\right]\right)^{-1} \phi\right\}. \tag{26}$$

This prior generalizes (13) to multiple predictor variables, while retaining its appealing properties. As elements of Σ_β approach zero, $p(B, \Sigma)$ approaches a dogmatic, no-predictability prior. Appendix A shows that $p(B, \Sigma)$ approaches a Jeffreys prior as the elements of Σ_β approach infinity. Intermediate levels of the elements in Σ_β allow for skepticism about the level of predictability in the same sense as σ_β in Section 1.2. Finally, as the above discussion shows, this prior implies well-defined prior beliefs on the improvement in the squared maximum Sharpe ratio.

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

$$Y = \begin{bmatrix} r_1^\top & x_1^\top \\ \vdots & \vdots \\ r_T^\top & x_T^\top \end{bmatrix}$$

and

$$X = \begin{bmatrix} 1 & x_0^\top \\ \vdots & \vdots \\ 1 & x_{T-1}^\top \end{bmatrix}.$$

The exact likelihood is given by

$$p(\mathcal{Y}|B, \Sigma) = |2\pi\Sigma_x|^{-\frac{1}{2}} |2\pi\Sigma|^{-\frac{T}{2}}$$
$$\exp\left\{-\frac{1}{2}\left(x_0 - \bar{x}\right)^\top \Sigma_x^{-1} \left(x_0 - \bar{x}\right) - \frac{1}{2}\mathrm{tr}\left[(Y - XB)^\top (Y - XB)\Sigma^{-1}\right]\right\}. \tag{27}$$

Applying Bayes rule to (26) and (27) leads to the posterior

$$p(B,\Sigma|\mathcal{Y}) \propto |\Sigma_x|^{\frac{N+K-1}{2}}|\Sigma|^{-\frac{T+N+2K+2}{2}} \exp\left\{-\frac{1}{2}(x_0-\bar{x})^\top \Sigma_x^{-1}(x_0-\bar{x})\right\}$$
$$\exp\left\{-\frac{1}{2}\phi^\top\left(\Sigma_u \otimes \left[\left(C_x^{-1}\right)^\top \Sigma_\beta C_x^{-1}\right]\right)^{-1}\phi - \frac{1}{2}\mathrm{tr}\left[(Y-XB)^\top(Y-XB)\Sigma^{-1}\right]\right\}. \quad (28)$$

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 $p(\Sigma|B,\mathcal{Y})$ using the inverted Wishart distribution with $T+K+1$ degrees of freedom. Second, α and β are drawn from $p(\alpha,\beta|\theta_0,\theta_1,\mathcal{Y})$, and finally θ_0 and θ_1 are drawn from $p(\theta_0,\theta_1|\alpha,\beta,\Sigma,\mathcal{Y})$. The computation of the conditional posteriors for (α,β), and (θ_0,θ_1) is exactly as in Section 1.3.

4 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 investment gains from exploiting predictability are relatively unlikely.

To quantify the size of the investment gains, we introduced a measure that is analogous to the Gibbons, Ross, and Shanken (1989) measure of mispricing. Just as the Gibbons, Ross, and Shanken measure shows the gain in the squared maximum Sharpe ratio from investing in additional assets, our measure shows the gain in the squared maximum Sharpe ratio from exploiting predictability. By placing a prior distribution on this measure, the investor also places a prior on the R^2 from the predictability equations. The skepticism of the investor's beliefs is controlled by a single parameter, σ_β. As σ_β approaches zero, the beliefs become dogmatic that there is no predictability, as σ_β approaches infinity, the prior becomes uninformative in the sense of Jeffreys (1961).

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 existence 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, these skeptical priors offered superior out-of-sample performance when compared to both diffuse priors, dogmatic priors, and an ordinary least squares regression.

This study provides a method for rigorously implementing an intermediate view on predictability. 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 Jeffreys prior on B, Σ

Our derivation for the limiting Jeffreys prior on B, Σ follows Stambaugh (1999). Zellner (1996, pp. 216-220) derives a limiting Jeffreys prior by applying (11) to the likelihood (7) and retaining terms of the highest order in T. Stambaugh shows that Zellner's approach is equivalent to applying (11) to the conditional likelihood (4), and taking the expectation in (11) assuming that x_0 is multivariate normal with mean (21) and variance (22) (or mean (5) and variance (6) in the $K = 1$ case). We adopt this approach.

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 $b = \mathrm{vec}(B)$, 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

$$l(B, \Sigma; \mathcal{Y}) = \log p(\mathcal{Y}|B, \Sigma, x_0). \tag{29}$$

denote the natural log of the likelihood. The definition of the Jeffreys prior requires

$$p(B, \Sigma^{-1}) \propto \left| -E\left(\begin{bmatrix} \frac{\partial^2 l}{\partial b \partial b^\top} & \frac{\partial^2 l}{\partial b \partial \zeta^\top} \\ \frac{\partial^2 l}{\partial \zeta \partial b^\top} & \frac{\partial^2 l}{\partial \zeta \partial \zeta^\top} \end{bmatrix} \right) \right|^{1/2} \tag{30}$$

Computing the expectation on the right hand side of (30) yields

$$p(B, \Sigma^{-1}) \propto \begin{vmatrix} \Sigma^{-1} \otimes \Psi & 0 \\ 0 & \frac{\partial^2 \log |\Sigma|}{\partial \zeta \partial \zeta^\top} \end{vmatrix}^{1/2}, \tag{31}$$

where

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

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

$$\left| \frac{\partial^2 \log |\Sigma|}{\partial \zeta \partial \zeta^\top} \right|^{1/2} = |\Sigma|^{\frac{N+K+1}{2}} \tag{32}$$

(recall that Σ is $(N+K)\times(N+K)$). It then follows from (31) 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

$$p(B,\Sigma)\propto|\Sigma_x|^{\frac{N+1}{2}}\,|\Sigma|^{-\frac{N+2K+2}{2}}. \tag{33}$$

B 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 (11) for B,Σ^{-1} given in Appendix A. 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 (31). Without loss of generality, the rows and columns of (31) can be re-ordered so that

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

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

$$p(\alpha,\theta_0,\theta_1,\Sigma^{-1})\propto\begin{vmatrix}\Phi & 0\\ 0 & \frac{\partial^2\log|\Sigma|}{\partial\zeta\partial\zeta^\top}\end{vmatrix}^{1/2}, \tag{34}$$

where

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

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

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

Because

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

it follows that

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

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)),

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

Therefore,

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

Finally, from (32) and (34),

$$\begin{aligned} p(\alpha, \theta_0, \theta_1, \Sigma^{-1}) &\propto |\Phi|^{1/2}|\Sigma|^{\frac{N+K+1}{2}} \\ &= |\Sigma|^N|\Sigma_u|^{K/2}|\Sigma_x|^{K/2}. \end{aligned}$$

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

C Sampling from the posterior (14)

The conditional density $p(\Sigma|B, \mathcal{Y})$ has an expression identical to (15). The proposal density is the inverted Wishart with $T+2$ degrees of freedom (Zellner (1996, p. 395)):

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

Because the target density takes the form

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

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

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

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

where $\hat{b} = \mathrm{vec}(\hat{B})$, 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 $2N \times 2$, and V_{22} is 2×2. Then

$$\begin{aligned}(b-\hat{b})^\top&(\Sigma^{-1}\otimes X^\top X)(b-\hat{b}) = \\ &\left(b_1-\hat{b}_1-V_{21}^\top V_{22}^{-1}(b_2-\hat{b}_2)\right)^\top \left(V_{11}-V_{21}^\top V_{22}^{-1}V_{21}\right)^{-1}\left(b_1-\hat{b}_1-V_{21}^\top V_{22}^{-1}(b_2-\hat{b}_2)\right) \\ &\qquad + \text{terms independent of } b_1.\end{aligned}$$

(see Green (1997, Chapter 3.1)). Then under the prior (13),

$$\begin{aligned}p(b_1\,|\,b_2,\Sigma,\mathcal{Y}) &\propto \exp\left\{-\frac{1}{2}\beta^\top\left(\sigma_\beta^2\sigma_x^{-2}\Sigma_u\right)^{-1}\beta\right\} \\ \exp&\left\{-\frac{1}{2}\left(b_1-\hat{b}_1-V_{21}^\top V_{22}^{-1}(b_2-\hat{b}_2)\right)^\top\left(V_{11}-V_{21}^\top V_{22}^{-1}V_{21}\right)^{-1}\left(b_1-\hat{b}_1-V_{21}^\top V_{22}^{-1}(b_2-\hat{b}_2)\right)\right\}.\end{aligned}$$

This equation also holds under the prior (36). We now complete the square to derive the distribution for b_1. Let

$$\Omega = \Sigma_u^{-1}\otimes\begin{bmatrix} 0 & 0 \\ 0 & \left(\sigma_\beta^2\sigma_x^{-2}\right)^{-1}\end{bmatrix}. \tag{35}$$

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

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

and mean

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

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),

$$(b - \hat{b})^\top (\Sigma^{-1} \otimes X^\top X)(b - \hat{b}) = \\ \left(b_2 - \hat{b}_2 - V_{12}^\top V_{11}^{-1}(b_1 - \hat{b}_1)\right)^\top \left(V_{22} - V_{12}^\top V_{11}^{-1} V_{12}\right)^{-1} \left(b_2 - \hat{b}_2 - V_{12}^\top V_{11}^{-1}(b_1 - \hat{b}_1)\right) \\ + \text{terms independent of } b_2.$$

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

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

for $\theta_1 \in (0, 1)$ and zero otherwise.[13]

D Results for the Flat Prior on θ_1

An alternative specification, motivated by the multivariate approach with exogenous regressors, is to assume a flat prior for $(\alpha, \theta_0, \theta_1)$ and consider a prior for Σ that is proportional to $|\Sigma|^{-\frac{N+2}{2}}$ (see, e.g., Pastor and Stambaugh (1999)). In this case,

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

and

$$\begin{aligned} p(B, \Sigma) &\propto |\sigma_\beta^2 \sigma_x^{-2} \Sigma_u|^{-1/2} |\Sigma|^{-\frac{N+2}{2}} \exp\left\{-\frac{1}{2}\beta^\top \left(\sigma_\beta^2 \sigma_x^{-2} \Sigma_u\right)^{-1} \beta\right\} \\ &\propto \sigma_x^N |\Sigma_u|^{-1/2} |\Sigma|^{-\frac{N+2}{2}} \exp\left\{-\frac{1}{2}\beta^\top \left(\sigma_\beta^2 \sigma_x^{-2} \Sigma_u\right)^{-1} \beta\right\}. \end{aligned} \tag{36}$$

As $\sigma_\beta \to \infty$, we obtain a prior considered by Stambaugh (1999).

Combining the prior (36) with the exact likelihood (7) produces

$$p(B, \Sigma|\mathcal{Y}) \propto \sigma_x^{N-1} |\Sigma_u|^{-1/2} |\Sigma|^{-\frac{T+N+2}{2}} \exp\left\{-\frac{1}{2}\beta^\top \left(\sigma_\beta^2 \sigma_x^{-2} \Sigma_u\right)^{-2} \beta\right\} \exp\left\{-\frac{1}{2}\sigma_x^{-2}(x_0 - \bar{x})^2\right\} \\ \exp\left\{-\frac{1}{2}\mathrm{tr}\left[(Y - XB)^\top (Y - XB)\Sigma^{-1}\right]\right\}. \tag{37}$$

[13] 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

$$\Omega = \Sigma_u^{-1} \otimes \begin{bmatrix} 0 & 0 \\ 0 & \left(C_x^{-1} \Sigma_\beta C_x\right)^{-1} \end{bmatrix}.$$

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

as a posterior. The strategy for sampling from (37) is the same as for the posterior under the limiting Jeffreys prior. The proposal densities are the same, except that Σ has an inverted-Wishart density with T, rather than $T + 2$ degrees of freedom. The target densities are different, but can be derived in the same way as the target densities for (15).

Tables 5–8 report results for the prior (36). As Tables 5 and 6 show, the posterior means implied by the flat prior are qualitatively similar to those implied by the limiting Jeffreys prior. Indeed, for the yield spread the results are nearly indistinguishable. This is not surprising: the flat prior and the Jeffreys prior differ in the weight they give to values of θ_1 near to -1 and 1, with the Jeffreys prior placing more weight at these extremes. However, the OLS estimate for the yield spread persistence is 0.74, far away from these extremes. Thus the yield spread lacks sufficient persistence to create a divergence in these two priors. In contrast, these two priors do imply somewhat different results for the dividend-price ratio. The Jeffreys prior leads to posterior means of θ_1 that are closer to 1, and to posterior means of beta that are somewhat higher (for the stock, the beta is 1.99 under the flat prior and 1.46 under the Jeffreys prior).

Table 3 shows that in spite of these differences, the out-of-sample performance of the dividend-price ratio is virtually the same, regardless of which prior is considered. The intermediate prior with $\sigma_\beta = .04$ still delivers the best performance. Table 4 shows, not surprisingly, that the results for the yield spread are also quite similar to the results for the Jeffreys prior.

References

Amihud, Yakov, and Clifford Hurvich, 2004, Predictive regressions: a reduced-bias estimation method, forthcoming, *Journal of Financial and Quantitative Analysis*.

Ang, Andrew, and Geert Bekaert, 2003, Stock return predictability: Is it there?, Working paper, Columbia University.

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

Baks, Klaas P., Andrew Metrick, and Jessica Wachter, 2001, Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation, *The 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.

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

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, 2003, A simulation approach to dynamics portfolio choice with an application to learning about return predictability, Forthcoming, *Review of Financial Studies*.

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

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., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*. (Princeton University Press Princeton, NJ).

Campbell, John Y., and Robert J. Shiller, 1988, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 58, 495–514.

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, 2004, Predicting the equity premium out of sample: Can anything beat the historical average?, Working paper, Harvard University.

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, 2004, Efficient tests of stock return predictability, Working paper, Harvard University and University of Pennsylvania.

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

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

Cochrane, John, and Jesus Saa-Requejo, 2000, Beyond arbitrage: Good deal asset price bounds in incomplete markets, *Journal of Political Economy* 108, 79–119.

Cremers, K.J. Martjin, 2002, Stock Return Predictability: A Bayesian Model Selection Perspective, *Review of Financial Studies* 15, 1223–1249.

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

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 29, 23–49.

Fama, Eugene F., and William G. Schwert, 1977, Asset returns and inflation, *Journal of Financial Economics* 5, 23–29.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Goyal, Amit, and Ivo Welch, 2004, A Comprehensive Look at the Empirical Performance of Equity Premium Prediction, Working paper, Department of Economics, Yale University.

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).

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

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.

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 Econometrics* 17, 357–390.

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.

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.

Pastor, Lubos, 2000, Portfolio selection and asset pricing models, *The 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.

Phillips, Peter C.B., 1991, To criticize the critics: An objective Bayesian analysis of stochastic trends, *Journal of Applied Econometrics* 6, 333–364.

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 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, 2004, 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.

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

Torous, Walter, Rossen Valkanov, and Shu Yan, 2005, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 78.

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: Dividend-Price ratio

	σ_β				
Parameter	0	0.04	0.08	∞	Reg.
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)
$E\left[r_{\text{bond}}\right]$	0.18	0.18	0.18	0.17	0.23
	(0.27)	(0.30)	(0.34)	(1.07)	
$E\left[r_{\text{stock}}\right]$	1.16	1.17	1.17	1.17	1.09
	(0.29)	(0.24)	(0.28)	(0.72)	
$E[x]$	-3.49	-3.50	-3.50	-3.50	-3.72
	(1.48)	(0.99)	(0.76)	(1.35)	

Notes: Posterior means for parameters in (1)–(3) implied by the likelihood (7) and the prior (13). Posterior standard deviations are in parentheses. The assets are the ten-year bond and the stock index; the log dividend-price ratio is the predictor variable x. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficients on the predictor variable (see (8)). $E[\cdot]$ denotes the posterior mean of the long-run expectation. The last column gives results from ordinary least squares regression, where $E[x] = \hat{\theta}_0/(1-\hat{\theta}_1)$ and $E[r] = \hat{\alpha} + \hat{\beta}E[x]$, and $\hat{}$ denotes the regression estimate of a coefficient. For regression estimates, standard errors are in parentheses. Data are quarterly from 1952 to 2004.

Table 2: Posterior Means: Yield spread

Parameter	σ_β 0	0.04	0.08	∞	Reg.
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)
$E\left[r_{\text{bond}}\right]$	0.21	0.21	0.21	0.21	0.23
	(0.28)	(0.28)	(0.29)	(0.33)	
$E\left[r_{\text{stock}}\right]$	1.67	1.67	1.67	1.67	1.69
	(0.58)	(0.59)	(0.60)	(0.63)	
$E[x]$	0.97	0.97	0.97	0.97	0.99
	(0.19)	(0.19)	(0.19)	(0.21)	

Notes: Posterior means for parameters in (1)–(3) implied by the likelihood (7) and the prior (13). Posterior standard deviations are in parentheses. The assets are the ten-year bond and the stock index; the yield spread is the predictor variable x. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficients on the predictor variable (see (8)). $E[\cdot]$ denotes the posterior mean of the long-run expectation. The last column gives results from ordinary least squares regression, where $E[x] = \hat{\theta}_0/(1-\hat{\theta}_1)$ and $E[r] = \hat{\alpha} + \hat{\beta}E[x]$, and $\hat{}$ denotes the regression estimate of a coefficient. For regression estimates, standard errors are in parentheses. Data are quarterly from 1952 to 2004.

Table 3: Out-of-Sample Results for the Dividend-Price Ratio

Panel A: Certainty Equivalent Returns

		σ_β				
R. A.	Assets	0	0.04	0.08	∞	Reg.
2	Bond and Stock	6.83	7.50	6.53	4.57	-1.81
	Stock Only	8.64	9.22	8.09	7.47	0.34
5	Bond and Stock	6.73	6.98	6.57	5.77	3.20
	Stock Only	7.42	7.64	7.16	6.91	4.04

Panel B: Sharpe Ratios

	σ_β				
Assets	0	0.04	0.08	∞	Reg.
Bond and Stock	0.18	0.21	0.17	0.09	0.10
Stock Only	0.29	0.32	0.26	0.24	0.18

Notes: Certainty equivalent returns (Panel A) and Sharpe ratios (Panel B) when the dividend-price ratio is the predictor variable. For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using all data up to that quarter assuming the likelihood (7) and the prior (13). Optimal portfolios are then determined by (18) 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 obtained as

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

where r_p are the quarterly returns 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 σ_β, the prior standard deviation of the normalized coefficient on the predictor variable (see (8)). Reg. denotes out-of-sample results obtained from estimating (1)–(3) by ordinary least squares. Each panel reports results from investing in both a stock index and the ten-year bond, and the stock index alone. Data are quarterly from 1952 to 2004. CERs are in annualized percentages ($\times$400), and Sharpe ratios are annualized ($\times$2).

Table 4: Out-of-Sample Results for the Yield Spread

Panel A: Certainty Equivalent Returns

		σ_β				
R. A.	Assets	0	0.04	0.08	∞	Reg.
2	Bond and Stock	5.27	6.93	7.58	-7.00	-15.59
	Stock Only	6.85	7.45	7.98	6.02	5.07
5	Bond and Stock	6.12	6.80	7.08	1.33	-2.07
	Stock Only	6.72	6.96	7.17	6.39	6.02

Panel B: Sharpe Ratios

	σ_β				
Assets	0	0.04	0.08	∞	Reg.
Bond and Stock	0.17	0.25	0.32	0.29	0.25
Stock Only	0.24	0.27	0.31	0.34	0.33

Notes: Certainty equivalent returns (Panel A) and Sharpe ratios (Panel B) when the yield spread is the predictor variable. For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using all data up to that quarter assuming the likelihood (7) and the prior (13). Optimal portfolios are then determined by (18) 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 obtained as

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

where r_p are the quarterly returns 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 σ_β, the prior standard deviation of the normalized coefficient on the predictor variable (see (8)). Reg. denotes out-of-sample results obtained from estimating (1)–(3) by ordinary least squares. Each panel reports results from investing in both a stock index and the ten-year bond, and the stock index alone. Data are quarterly from 1952 to 2004. CERs are in annualized percentages ($\times 400$), and Sharpe ratios are annualized ($\times 2$).

Table 5: Posterior Means for the Dividend-Price Ratio with the Prior (36)

	σ_β				
Parameter	0	0.04	0.08	∞	Reg.
β_{bond}	-0.00	0.03	0.00	-0.15	-0.10
	(0.00)	(0.25)	(0.47)	(0.72)	(0.73)
β_{stock}	0.00	0.92	1.77	1.99	2.72
	(0.00)	(0.69)	(1.03)	(1.18)	(1.52)
θ_1	0.997	0.990	0.984	0.983	0.976
	(0.002)	(0.006)	(0.010)	(0.011)	(0.015)
$E\left[r_{\text{bond}}\right]$	0.18	0.18	0.18	0.18	0.23
	(0.28)	(0.30)	(0.32)	(0.61)	
$E\left[r_{\text{stock}}\right]$	1.17	1.17	1.17	1.17	1.09
	(0.31)	(0.21)	(0.23)	(0.40)	
$E[x]$	-3.51	-3.49	-3.50	-3.50	-3.72
	(1.18)	(0.71)	(0.51)	(0.71)	

Notes: Posterior means for parameters in (1)–(3) implied by the likelihood (7) and the prior (36). Posterior standard deviations are in parentheses. The assets are the ten-year bond and the stock index; the log dividend-price ratio is the predictor variable x. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficients on the predictor variable (see (8)). $E[\cdot]$ denotes the posterior mean of the long-run expectation. The last column gives results from ordinary least squares regression, where $E[x] = \hat{\theta}_0/(1-\hat{\theta}_1)$ and $E[r] = \hat{\alpha} + \hat{\beta}E[x]$, and $\hat{}$ denotes the regression estimate of a coefficient. For regression estimates, standard errors are in parentheses. Data are quarterly from 1952 to 2004.

Table 6: Posterior Means for the Yield Spread with the Prior (36)

	σ_β				
Parameter	0	0.04	0.08	∞	Reg.
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)	(0.56)	(0.56)
θ_1	0.73	0.73	0.73	0.75	0.74
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
$E\left[r_{\text{bond}}\right]$	0.21	0.21	0.21	0.21	0.23
	(0.28)	(0.28)	(0.29)	(0.33)	
$E\left[r_{\text{stock}}\right]$	1.67	1.67	1.67	1.67	1.69
	(0.58)	(0.59)	(0.60)	(0.63)	
$E[x]$	0.97	0.97	0.97	0.97	0.99
	(0.19)	(0.19)	(0.19)	(0.20)	

Notes: Posterior means for parameters in (1)–(3) implied by the likelihood (7) and the prior (36). Posterior standard deviations are in parentheses. The assets are the ten-year bond and the stock index; the yield spread is the predictor variable x. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficients on the predictor variable (see (8)). $E[\cdot]$ denotes the posterior mean of the long-run expectation. The last column gives results from ordinary least squares regression, where $E[x] = \hat{\theta}_0/(1-\hat{\theta}_1)$ and $E[r] = \hat{\alpha} + \hat{\beta}E[x]$, and $\hat{}$ denotes the regression estimate of a coefficient. For regression estimates, standard errors are in parentheses. Data are quarterly from 1952 to 2004.

Table 7: Out-of-Sample Results for the Dividend-Price Ratio with the Prior (36)

Panel A: Certainty Equivalent Returns

		σ_β				
R. A.	Assets	0	0.04	0.08	∞	Reg.
2	Bond and Stock	6.87	7.51	6.03	3.56	-1.81
	Stock Only	8.68	9.17	7.51	6.29	0.34
5	Bond and Stock	6.74	6.98	6.36	5.35	3.20
	Stock Only	7.44	7.62	6.93	6.43	4.04

Panel B: Sharpe Ratios

	σ_β				
Assets	0	0.04	0.08	∞	Reg.
Bond and Stock	0.18	0.21	0.15	0.09	0.10
Stock Only	0.29	0.32	0.24	0.21	0.18

Notes: Certainty equivalent returns (Panel A) and Sharpe ratios (Panel B) when the dividend-price ratio is the predictor variable. For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using all data up to that quarter assuming the likelihood (7) and the prior (36). Optimal portfolios are then determined by (18) 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 obtained as

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

where r_p are the quarterly returns 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 σ_β, the prior standard deviation of the normalized coefficient on the predictor variable (see (8)). Reg. denotes out-of-sample results obtained from estimating (1)–(3) by ordinary least squares. Each panel reports results from investing in both a stock index and the ten-year bond, and the stock index alone. Data are quarterly from 1952 to 2004. CERs are in annualized percentages ($\times 400$), and Sharpe ratios are annualized ($\times 2$).

Table 8: Out-of-Sample Results for the Yield Spread with the Prior (36)

Panel A: Certainty Equivalent Returns

		σ_β				
R. A.	Assets	0	0.04	0.08	∞	Reg.
2	Bond and Stock	5.35	6.93	7.44	-6.28	-15.59
	Stock Only	6.91	7.50	8.02	6.15	5.07
5	Bond and Stock	6.15	6.80	7.03	1.62	-2.07
	Stock Only	6.74	6.98	7.18	6.44	6.02

Panel B: Sharpe Ratios

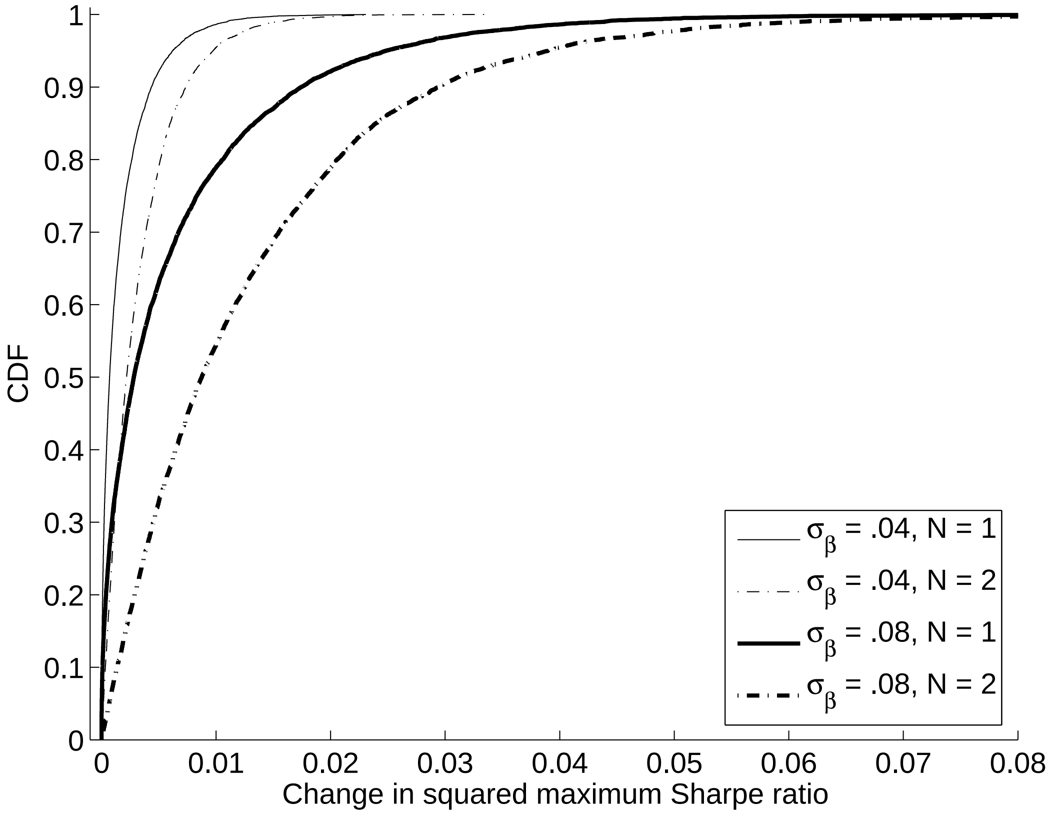
	σ_β				
Assets	0	0.04	0.08	∞	Reg.
Bond and Stock	0.17	0.25	0.31	0.29	0.25
Stock Only	0.24	0.27	0.31	0.34	0.33

Notes: Certainty equivalent returns (Panel A) and Sharpe ratios (Panel B) when the yield spread is the predictor variable. For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using all data up to that quarter assuming the likelihood (7) and the prior (36). Optimal portfolios are then determined by (18) 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 obtained as

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

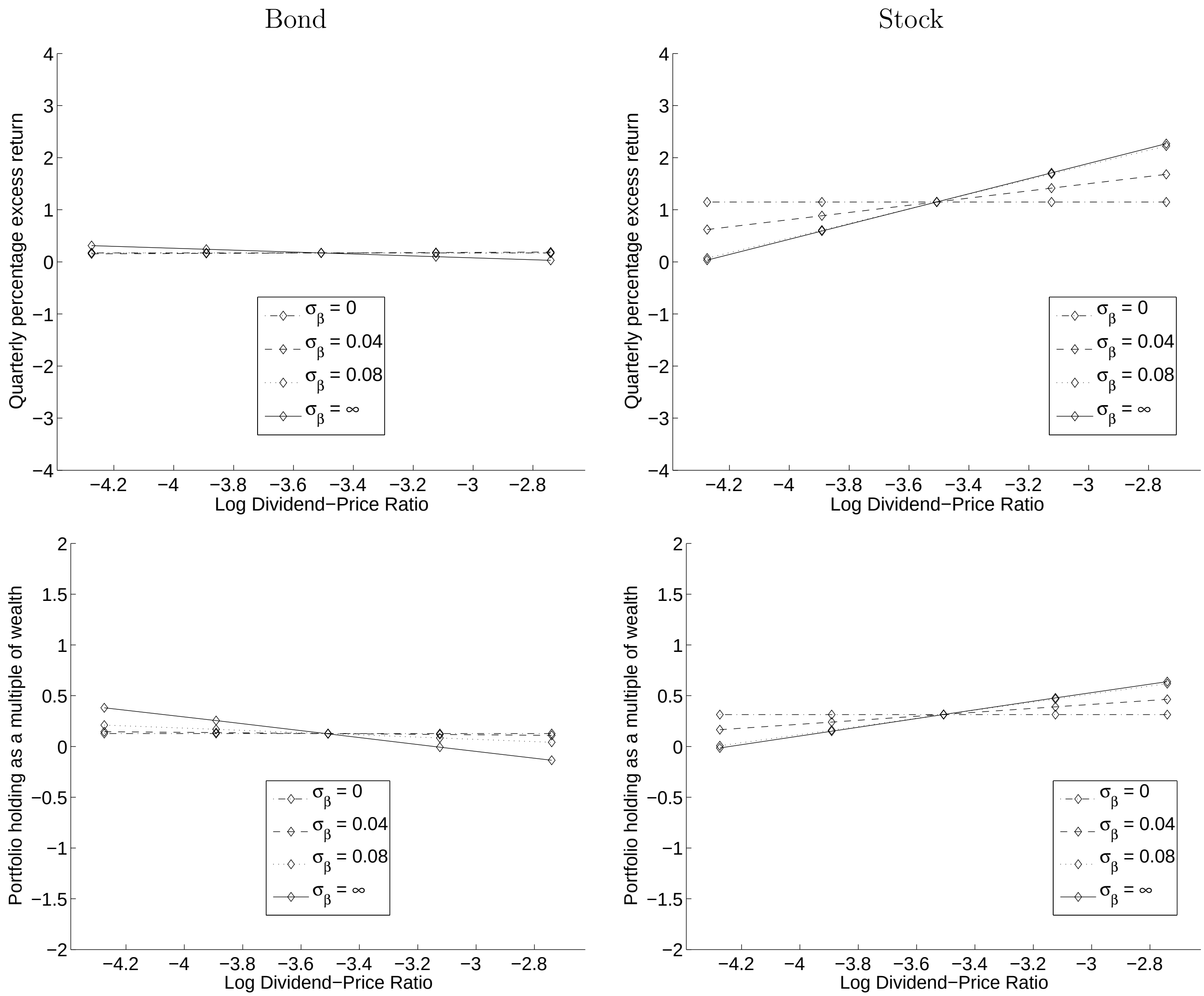
where r_p are the quarterly returns on the investment strategy and the mean and variance are computed using the sample. Sharpe ratios are the average excess returns on the investment strategy divided by the standard deviation and do not depend on A. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficient on the predictor variable (see (8)). Reg. denotes out-of-sample results obtained from estimating (1)–(3) by ordinary least squares. Each panel reports results from investing in both a stock index and the ten-year bond, and the stock index alone. Data are quarterly from 1952 to 2004. CERs are in annualized percentages ($\times 400$), and Sharpe ratios are annualized ($\times 2$).

Figure 1: Prior on the Change in the Squared Maximum Sharpe Ratio



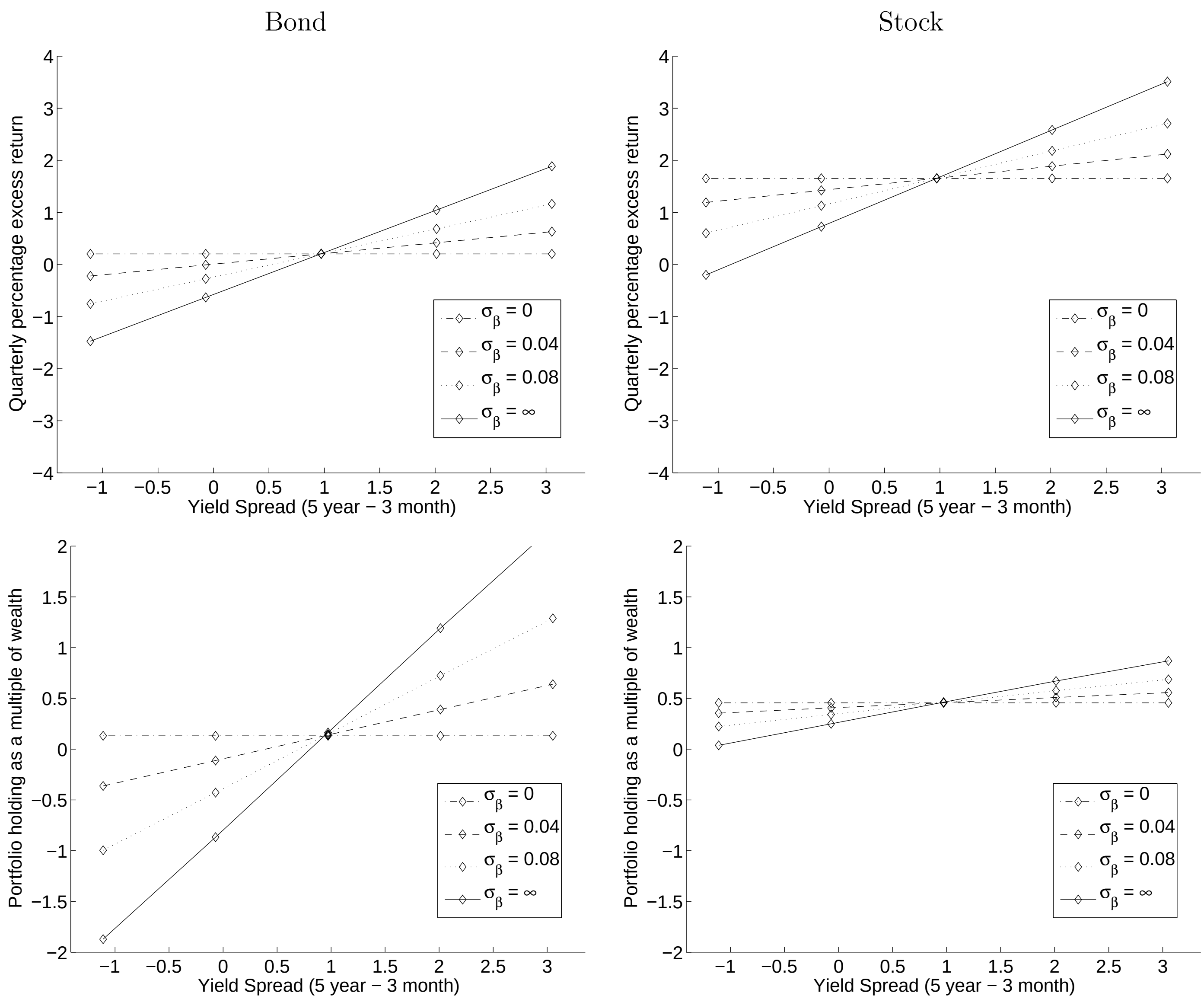
Notes: Prior distribution on the change in the expected squared maximum Sharpe ratio arising from conditioning on the predictor variable. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficients on the predictor variable (see (8)). N denotes the number of assets available to the investor. When $N = 1$, the change in squared maximum Sharpe ratio is a close approximation to the R^2 of the predictive regression.

Figure 2: Conditional Expected Returns and Holdings when x_t is the Dividend-Price Ratio



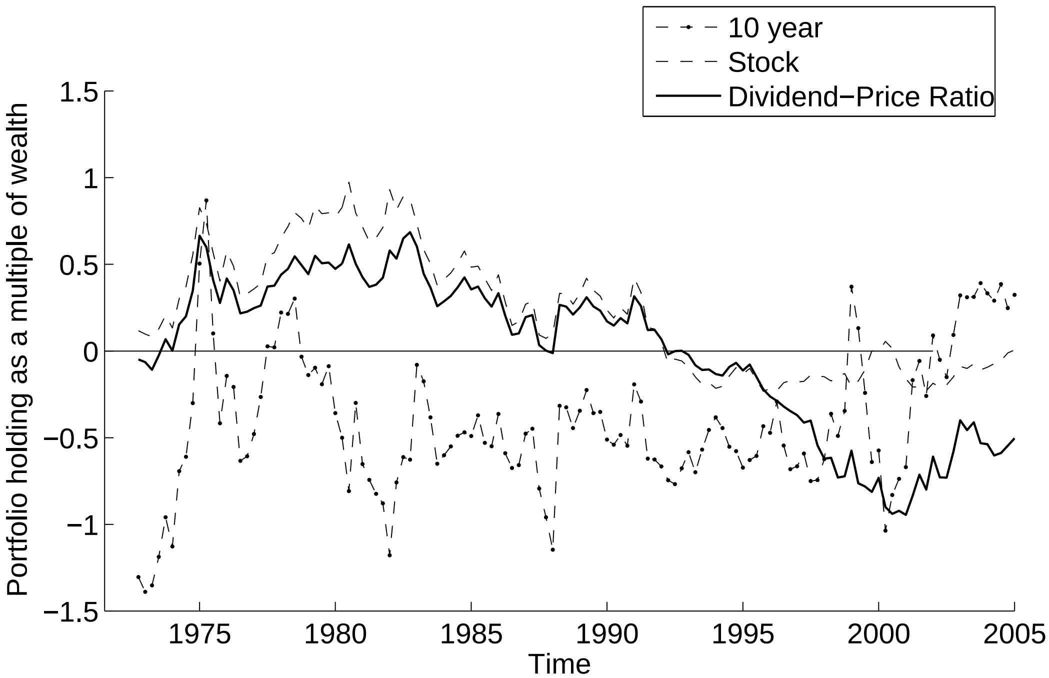
Notes: Conditional expected returns (top two plots) and portfolio holdings (bottom two plots) as functions of the log dividend-price ratio. Conditional expected returns are calculated using the predictive distribution implied by the likelihood (7) and the prior (13) with the ten-year bond and the stock index as risky assets and the dividend-price ratio as the predictor variable x_t. Given the predictive distribution, portfolio holdings are calculated using (18) for risk aversion parameter $A = 5$. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficients on the predictor variable (see (8)). Diamonds correspond to the sample mean and plus and minus one and two sample standard deviations of the predictor variable. Data are quarterly from 1952 to 2004.

Figure 3: Conditional Expected Returns and Holdings when x_t is the Yield Spread



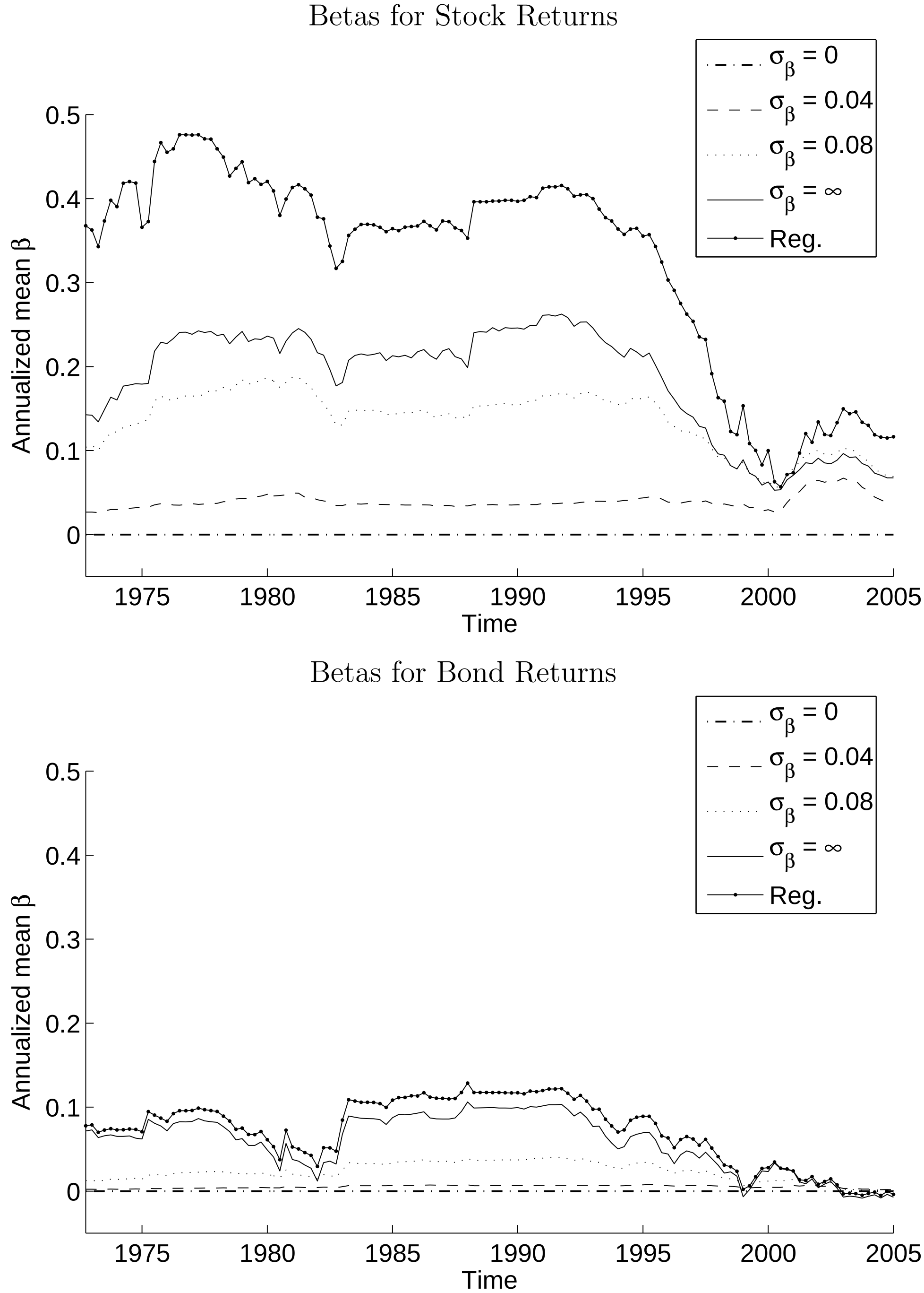
Notes: Conditional expected returns (top two plots) and portfolio holdings (bottom two plots) as functions of the yield spread. Conditional expected returns are calculated using the predictive distribution implied by the likelihood (7) and the prior (13) with the ten-year bond and the stock index as risky assets and the yield spread as the predictor variable x_t. Given the predictive distribution, portfolio holdings are calculated using (18) for risk aversion parameter $A = 5$. Prior beliefs are indexed by σ_β, the prior standard deviation of the normalized coefficients on the predictor variable (see (8)). Diamonds correspond to the sample mean and plus and minus one and two sample standard deviations of the predictor variable. Data are quarterly from 1952 to 2004.

Figure 4: Time Series of the Dividend-Price Ratio and Portfolio Holdings for the Diffuse Prior



Notes: Time series of optimal portfolio holdings in the bond and in the stock index and the de-meaned log dividend-price ratio. For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using the data up to that quarter assuming the dividend-price ratio is the predictor variable, the likelihood is (7), and the prior is (13) with $\sigma_\beta = \infty$. This predictive distribution is used to determine optimal portfolios from (18) with risk aversion parameter $A = 5$. Data are quarterly from 1952 to 2004.

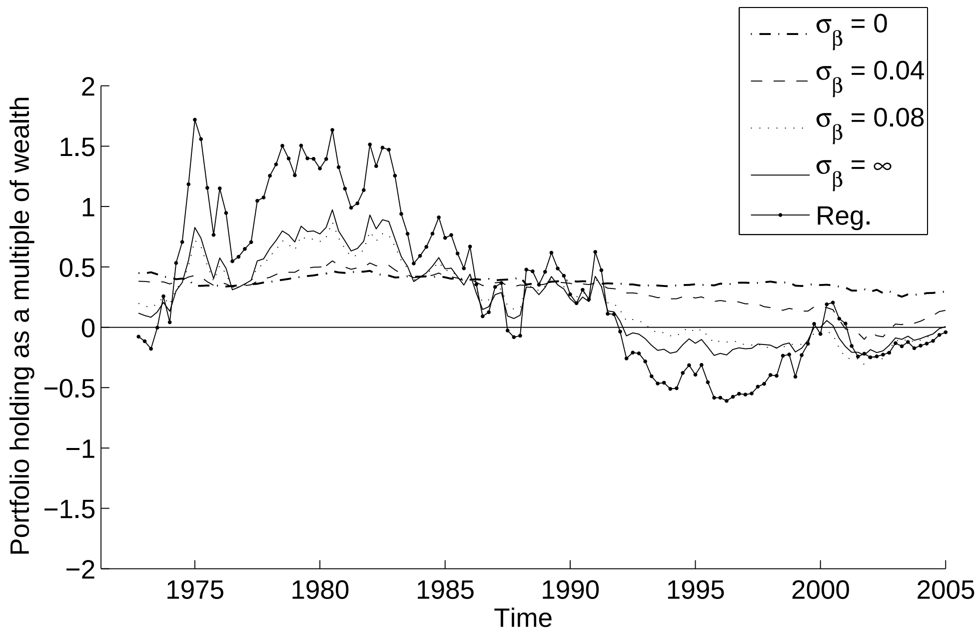
Figure 5: Time Series of Posterior Means of β when x_t is the Dividend-Price Ratio



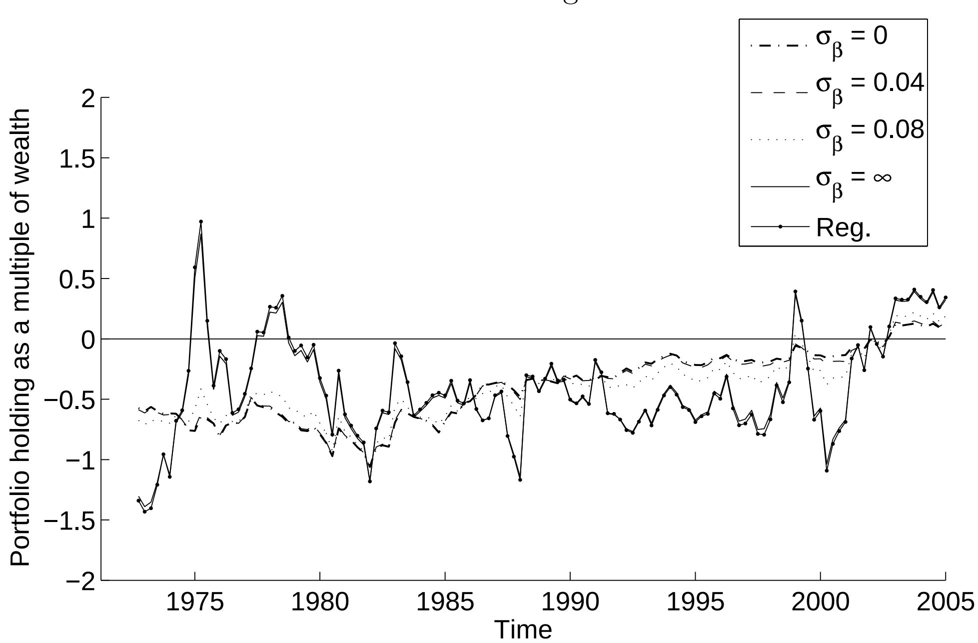
Notes: Time series of the posterior means of β, the coefficient on x_t. The top figure plots β corresponding to the stock return, the bottom figure plots β corresponding to the bond return. For each quarter beginning in the first quarter of 1972, the posterior distribution is computed using the data up to that quarter assuming the dividend-price ratio is the predictor variable, the likelihood is (7), and the prior is (13) for various values of σ_β. Also plotted are the βs implied by estimating (1)–(3) by ordinary least squares regression (Reg). Data are quarterly from 1952 to 2004.

Figure 6: Time Series of Portfolio Holdings when x_t is the Dividend-Price Ratio

Stock Holdings

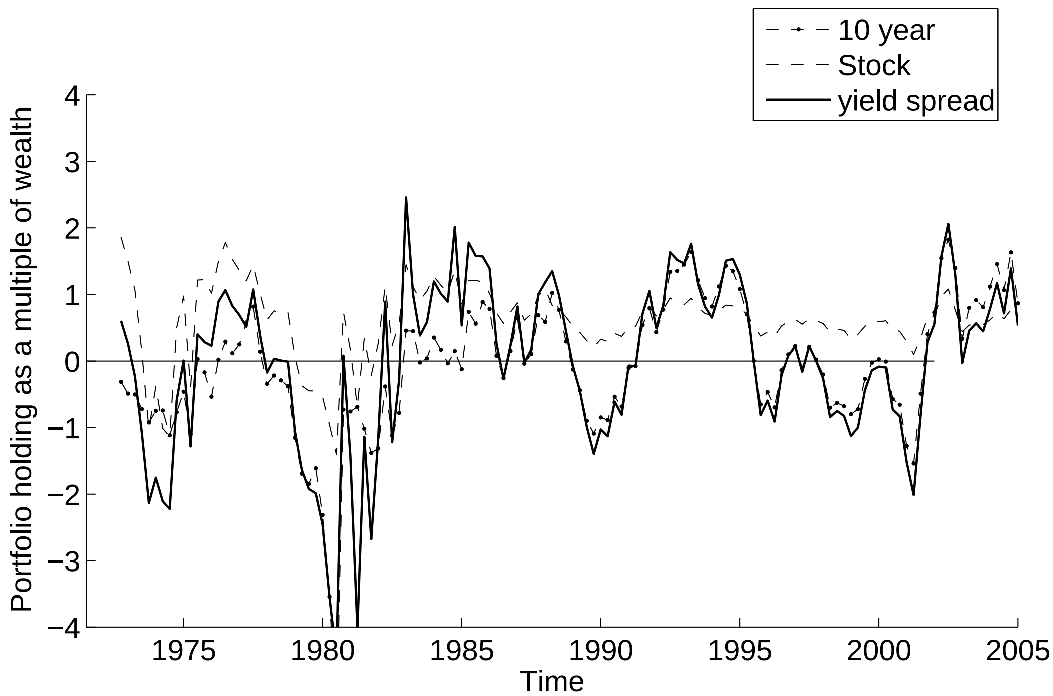


Bond Holdings



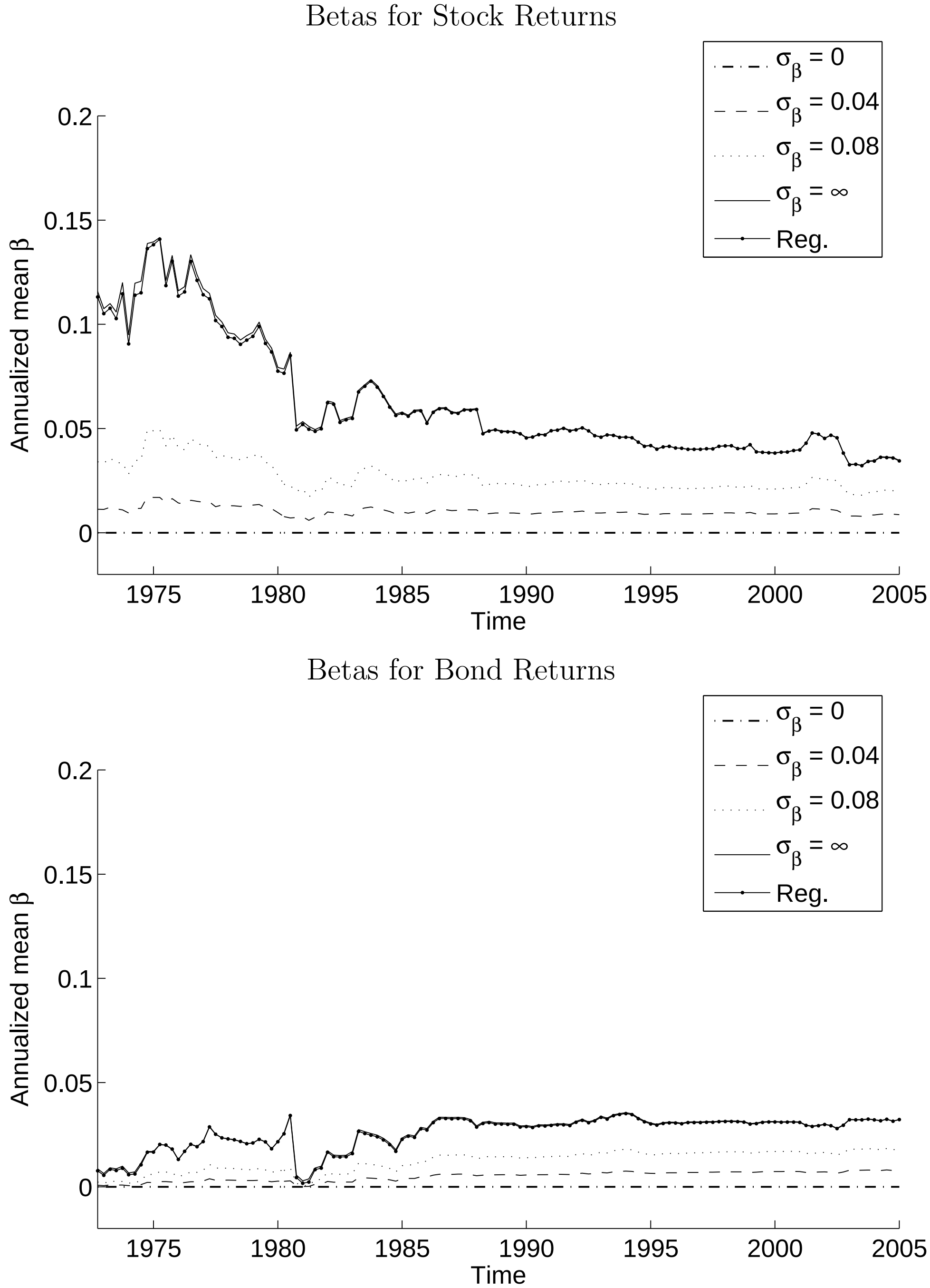
Notes: Time series of optimal portfolio holdings in the stock index (top panel) and in the ten-year bond (bottom panel). For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using the data up to that quarter assuming the dividend-price ratio is the predictor variable, the likelihood is (7), and the prior is (13) for various values of σ_β. This predictive distribution is used to determine optimal portfolios from (18) with risk aversion parameter $A = 5$. Reg. denotes results obtained from estimating (1)–(3) by ordinary least squares. Data are quarterly from 1952 to 2004.

Figure 7: Time Series of the Yield Spread and Portfolio Holdings for the Diffuse Prior



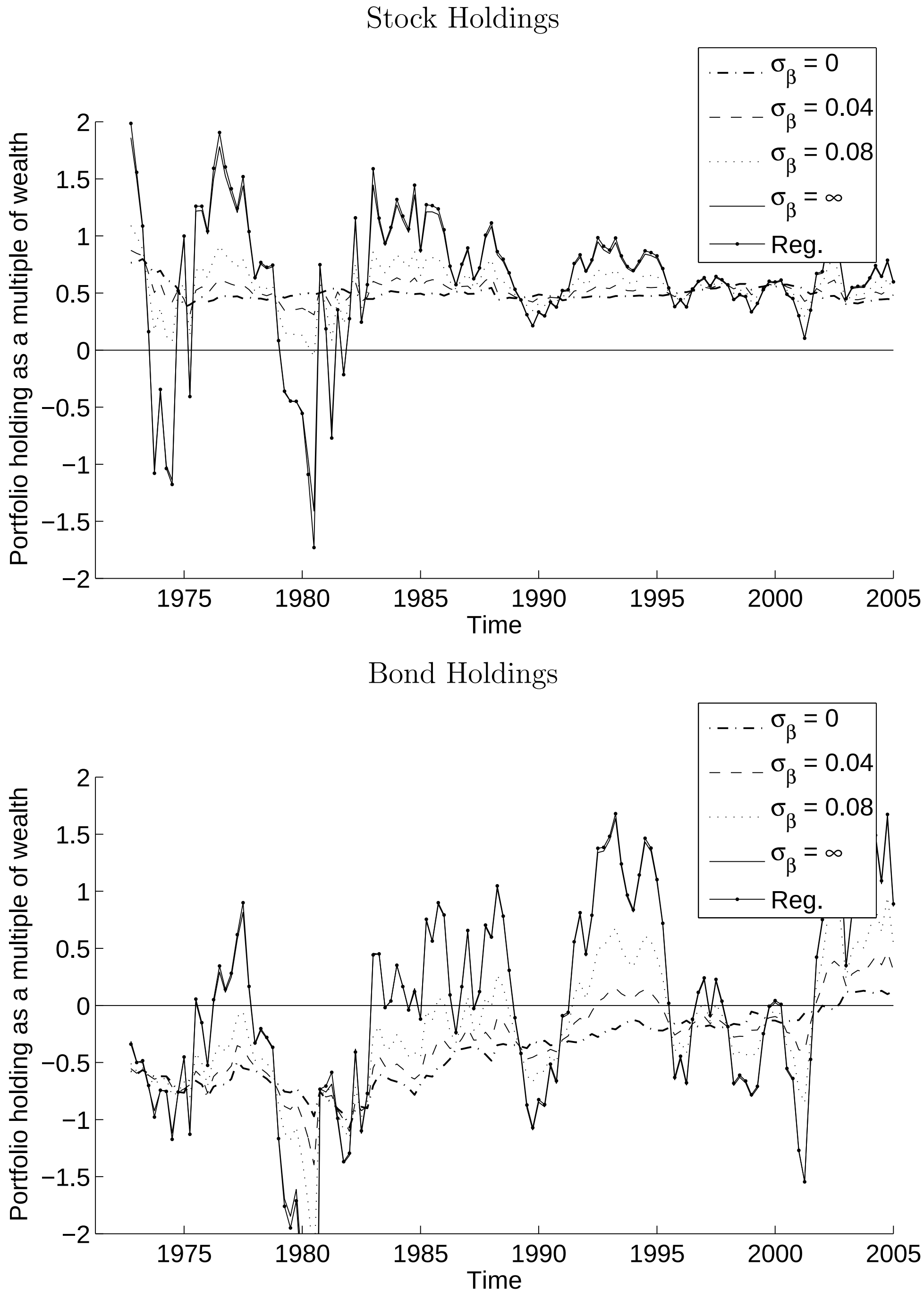
Notes: Time series of optimal portfolio holdings in the bond and in the stock index and the de-meaned yield spread. For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using the data up to that quarter assuming the yield spread is the predictor variable, the likelihood is (7), and the prior is (13) with $\sigma_\beta = \infty$. This predictive distribution is used to determine optimal portfolios from (18) with risk aversion parameter $A = 5$. Data are quarterly from 1952 to 2004.

Figure 8: Time Series of Posterior Means of β when x_t is the Yield Spread



Notes: Time series of the posterior means of β, the coefficient on x_t. The top figure plots β corresponding to the stock return, the bottom figure plots β corresponding to the bond return. For each quarter beginning in the first quarter of 1972, the posterior distribution is computed using the data up to that quarter assuming the yield spread is the predictor variable, the likelihood is (7), and the prior is (13) for various values of σ_β. Also plotted are the βs implied by estimating (1)–(3) by ordinary least squares regression (Reg). Data are quarterly from 1952 to 2004.

Figure 9: Time Series of Portfolio Holdings when x_t is the Yield Spread



Notes: Time series of optimal portfolio holdings in the stock index (top panel) and in the ten-year bond (bottom panel). For each quarter beginning in the first quarter of 1972, the predictive distribution for returns is computed using the data up to that quarter assuming the yield spread is the predictor variable, the likelihood is (7), and the prior is (13) for various values of σ_β. This predictive distribution is used to determine optimal portfolios from (18) with risk aversion parameter $A = 5$. Reg. denotes results obtained from estimating (1)–(3) by ordinary least squares. Data are quarterly from 1952 to 2004.