Journal of Econometrics 148 (2009) 162–178

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

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

Jessica A. Wachter[a,b,*], Missaka Warusawitharana[c,1]

[a] Department of Finance, The Wharton School, University of Pennsylvania, 2300 SH-DH, Philadelphia, PA, 19104, United States

[b] NBER, Cambridge, MA 02138, United States

[c] Division of Research and Statistics, Board of Governors of the Federal Reserve System, Mail Stop 97, 20th and Constitution Ave, Washington DC, 20551, United States

ARTICLE INFO

Article history:
Received 29 May 2007
Received in revised form
16 September 2008
Accepted 16 October 2008
Available online 5 November 2008

ABSTRACT

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

0. Introduction

Are excess returns predictable, and if so, what does this mean for investors? In classic studies of rational valuation (e.g. Samuelson (1965, 1973) and Shiller (1981)), risk premia are constant over time and thus excess returns are unpredictable.[2] However, an extensive empirical literature has found evidence for predictability in returns on stocks and bonds by scaled-price ratios and interest rates.[3]

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.[4] Between these extremes, however, lies an interesting intermediate view: that both data and theory can be helpful in forming portfolio allocations.

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

☆ For helpful comments, we thank the associate editor and two anonymous referees, as well as John Campbell, John Cochrane, Joel Dickson, Itamar Drechsler, Bjorn Eraker, John Geweke, Martin Lettau, Stijn Van Nieuwerburgh, Lubos Pastor, Jay Shanken, Robert Stambaugh, Alexander Stremme, Ivo Welch, Amir Yaron, Motohiro Yogo and seminar participants at the 2005 CIRANO-CIREQ Financial Econometrics Conference, the 2006 AFA meetings, the 2006 SED meetings, the 2007 D-CAF Conference on Return Predictability, and at Harvard University, the Vanguard Group, and the Wharton School. We are grateful for financial support from the Aronson + Johnson + Ortiz fellowship through the Rodney L. White Center for Financial Research. This manuscript does not reflect the views of the Board of Governors of the Federal Reserve System.

* Corresponding author at: Department of Finance, The Wharton School, University of Pennsylvania, 2300 SH-DH, Philadelphia, PA, 19104, United States. Tel.: +1 215 898 7634.

E-mail addresses: jwachter@wharton.upenn.edu (J.A. Wachter), missaka.n.warusawitharana@frb.gov (M. Warusawitharana).

1 Tel.: +1 202 452 3461.

2 Examples of general-equilibrium models that imply excess returns that are largely unpredictable include Abel (1999, 1990), Backus et al. (1989), Campbell (1986), Cecchetti et al. (1993), Kandel and Stambaugh (1991) and Mehra and Prescott (1985).

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

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

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.10.009

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

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

Our study builds on previous work that has examined predictability from a Bayesian investment perspective. Kandel and Stambaugh (1996) show that predictive relations that are weak in terms of standard statistical measures can nonetheless affect portfolio choice.[5] They conduct a simulation experiment such that predictability is present with modest significance, and examine the portfolio choices of a Bayesian investor who views the simulated data. In contrast, our study bases its inference on the historical time series of returns and predictor variables. We ask whether an investor whose priors imply skepticism about the existence of predictability would find it optimal to vary their investments in risky assets over time. Other studies make use of informative priors in a setting of return predictability. Avramov (2002) and Cremers (2002) show that Bayesian inference and informative priors can lead to superior model selection. Shanken and Tamayo (2005) jointly model time variation in risk and expected return in a Bayesian setting. Shanken and Tamayo incorporate model-based intermediate views on the relation between expected return and risk. In what follows, we compare the prior beliefs we assume to those in each of these related studies.

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

Our use of model-based informative priors has parallels in a literature that examines the portfolio implications of the cross-section of stock returns. Motivated by the extreme weights and poor out-of-sample performance of mean–variance efficient portfolios (Best and Grauer, 1991; Green and Hollifield, 1992), Black and Litterman (1992) propose using market weights as a benchmark, in effect using both data and the capital asset pricing model to form portfolios. Recently, Bayesian studies such as Pastor (2000), Avramov (2004) and Wang (2005) construct portfolios incorporating informative beliefs about cross-sectional asset pricing models.[6] Like the present study, these studies show that allowing models to influence portfolio selection can be superior to using the data alone. While these studies focus on the cross-section of returns, we apply these ideas to the time series.

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

1. Portfolio choice for a skeptical investor

Given observations on returns and a predictor variable, how should an investor allocate his wealth? One approach would be to estimate the predictability relation, treat the point estimates as known, and solve for the portfolio that maximizes utility. An alternative approach, adopted in Bayesian studies, is to specify prior beliefs on the parameters. The prior represents the investor's beliefs about the parameters before viewing data. After viewing data, the prior is updated to form a posterior distribution; the parameters are then integrated out to form a predictive distribution for returns, and utility is maximized with respect to this distribution. This approach incorporates the uncertainty inherent in estimation into the decision problem (see Klein and Bawa (1976), Bawa et al. (1979) and Brown (1979)).[7] Rather than assuming that the investor knows the parameters, it assumes, realistically, that the investor estimates the parameters from the data. Moreover, this approach allows for prior information, perhaps motivated by economic models, to enter into the decision process. This section describes the specifics of the likelihood function, the prior, and the posterior used in this study.

1.1. Likelihood

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

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

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

The data generating process is assumed to be

$$r_{t+1} = \alpha + \beta x_t + u_{t+1} \quad (1)$$

$$x_{t+1} = \theta_0 + \theta_1 x_t + v_{t+1}, \quad (2)$$

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

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

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

where

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

α and β are $N \times 1$ vectors and Σ is an $(N+1) \times (N+1)$ symmetric and positive definite matrix.[8] 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) and 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:

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

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

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

where $\mathrm{tr}(\cdot)$ denotes the sum of the diagonal elements of a matrix.[9]

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

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

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

and variance

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

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

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

Eq. (7) is the likelihood function used in our analysis. Following Box et al. (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 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. Between these extremes lie priors that downweight empirical evidence on return predictability. These informative priors imply that large values of the R^2 from predictive regressions are unlikely, but not impossible.

Before discussing the specifics of our informative priors, we briefly discuss how asset-pricing theory guides us toward these priors. Models with constant relative risk aversion, rational agents, and homoskedastic endowments (e.g. Abel (1990, 1999), Backus et al. (1989) and Barro (2006) and the benchmark case in Bansal and Yaron (2004)) imply that the R^2 in the predictability regression is exactly zero.[10] Recently, several papers propose general equilibrium models that are capable of generating time-varying expected excess returns (risk premia). Models that successfully account for time-varying risk premia have done so through one of several mechanisms: time-varying relative risk aversion (e.g. Campbell and Cochrane (1999)), non-rational investors (e.g. Barberis et al. (2001)), or time-varying volatility in the endowment process (e.g. Whitelaw (2000) and Bansal and Yaron (2004)). An investor whose prior is that constant relative risk aversion is more likely than time-varying relative risk aversion, that investors are more likely to be rational and that aggregate consumption growth is more likely to be homoskedastic would therefore place low probability on a high R^2 from predictive regressions.[11]

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

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

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

11 Of these possibilities, it is perhaps most difficult to conceptualize a prior over endowment heteroskedasticity. This issue is complicated by the fact that a model with endowment heteroskedasticity will not necessarily imply that return heteroskedascity takes a simple form, and that return variance and risk premia are related in a straightforward way. In a study that we discuss further below, Shanken and Tamayo (2005) take a reduced-form approach and assume priors that favor a linear relation between the return variance and the risk premium. They find that there is significant variation in expected excess returns that does not correspond to variation in volatility, a conclusion also reached in many frequentist studies (see Campbell (2003) for a survey). We hope to investigate the important issue of heteroskedasticity in future research; however, given the results of existing studies we expect our conclusions to be unaffected.

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

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

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

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

where I_N is the $N \times N$ identity matrix.[12]

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

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

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

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

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

For the remaining parameters, we choose a prior that is uninformative in the sense of Jeffreys (1961).[13] We follow the approach of Stambaugh (1999) and Zellner (1996), and derive a limiting Jeffreys prior as explained in Appendix D. This limiting prior takes the form

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

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

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

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

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

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

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

When there are N risky assets, there is a natural extension of (13). Besides implying a distribution on the R^2 of each asset, the distribution of η, (8) implies a joint distribution such that no linear combination of asset returns can have an R^2 that is "too large". To be more precise, let w be an $N \times 1$ vector. Then calculations in Appendix B show that

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

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

Fig. 1 depicts the distribution of the R^2 implied by these prior beliefs. The figure shows the probability that the R^2 exceeds some value k, $P(R^2 \geq k)$, as a function of k; it is therefore one minus the cumulative distribution function for the R^2. We construct this figure by simulating draws of η from (8) and, for each draw, constructing a draw from the R^2 distribution using (13). As this figure shows, the parameter σ_η indexes the degree to which the prior is informative. For $\sigma_\eta = 0$, the investor assigns zero probability to a positive R^2; for this reason $P(R^2 \geq k)$ is equal to one at zero and is zero elsewhere. As σ_η increases, the investor assigns non-zero probability to positive values of the R^2. For $\sigma_\eta = 0.04$, the probability that the R^2 exceeds 0.02 is 0.0005. For $\sigma_\eta = 0.08$, the probability that the R^2 exceeds 0.02 is 0.075. Finally when σ_η is large, approximately equal probabilities are assigned to all values of the R^2. This is the diffuse prior that expresses no skepticism with regard to the data. In what follows, we will consider the implications of these four priors for the individual's investment decisions. We will refer to them by using the corresponding probabilities that the R^2 exceeds 0.02. We note, however, that we focus on 0.02 for convenience; any number between 0 and 1 could be substituted.

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

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

[14] 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 (2008) suggest disregarding estimates of β if the expected excess return is negative, or if β has an opposite sign to that suggested by theory. In our model, these restrictions could be imposed by assigning zero prior weight to the appropriate regions of the parameter space. One could also consider a non-zero mean for β, corresponding to a prior belief that favors predictability of a particular sign. For simplicity, we focus on priors that apply to any predictor variable on possibly multiple assets, and leave these extensions to future work.

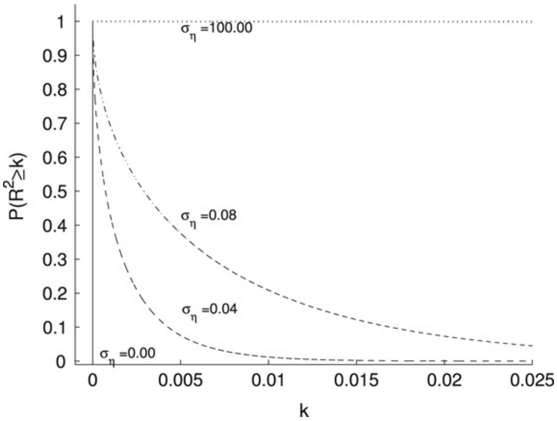

Fig. 1. Prior on the R^2 of the predictive regression. (Notes: The prior probability that the R^2 exceeds a value k, for k ranging from 0 to 0.025 implied by skeptical Jeffreys priors. Prior beliefs are indexed by σ_η, the prior standard deviation of the normalized coefficient on the predictor variable.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

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

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

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

1.2.1. Comparison with related studies

In this section we have described one way of modeling prior information. We now compare this approach to that used in some important Bayesian return predictability studies that make use of informative priors.[15]

Kandel and Stambaugh (1996) derive posteriors assuming the investor has seen, in addition to the actual data, a "prior" sample of the data that has unconditional moments equal to those of the actual sample except without predictability.[16] These priors are also assumed by Avramov (2002, 2004). In a recent working paper, Pastor and Stambaugh (2006) focus on informative priors about the correlation between u_{t+1} and v_{t+1}; they also construct an informative prior on the R^2 using the sample variance of stock returns. Cremers (2002) specifies informative prior beliefs motivated by placing a distribution on the expected sample R^2. Cremers's priors assume knowledge of sample moments of the predictor variable. Specifically, if we let X denote a $T \times k$ matrix of values for k predictor variables, Cremers assumes that the investor knows $X^\top X$.

The various priors specified in the above studies all contain the same flaw: they require that the investor has knowledge of the sample moments when forming a prior. It is not enough for the investor to make a reasonable guess as to the value of the moments. For the analyses to be correct, the investor must know the value of the moment precisely. From a Bayesian perspective, such knowledge is problematic. To know the moments of the data, either the investor must have seen the data (but then the prior and the posterior would be identical), or somehow intuited the correct moments, without seeing the data. Even if we accept this awkward latter interpretation, to be consistent these moments would have to be treated as constants (namely conditioned on) throughout the analysis, which they are not.[17]

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

Another important study that is related to ours is that of Shanken and Tamayo (2005).[18] Shanken and Tamayo model time-variation in risk as well as in expected returns. Like our priors, the priors in Shanken and Tamayo represent a model-based view that is intermediate between complete faith in a model and complete faith in the data. However, their formulation of priors is less parsimonious, requiring ten parameters in the case of a single asset (a broad stock market portfolio) and predictor variable (the dividend-yield). The prior values are specific to these variables and do not transfer easily to other assets or new predictor variables. The advantage of our method is that it expresses the informativeness of the agent's prior beliefs as a single number which can be mapped into beliefs about the maximum R^2. This is the case regardless of the number of risky assets or the number and characteristics of the predictor variables.

Our choice of priors is in fact reminiscent of the choice of priors over intercepts in cross-sectional studies. Pastor and Stambaugh (1999) and Pastor (2000) specify 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 imply very high Sharpe ratios, because there would be nothing to prevent a low residual variance draw from occurring simultaneously with a high intercept draw. Bayesian portfolio choice studies (Baks et al., 2001; Jones and Shanken, 2005; Pastor and Stambaugh, 2002) specify an informative prior on estimates of mutual fund skill. Because skill is measured as the intercept on a regression of fund returns on factors, this is analogous to the prior used in the cross-sectional studies. Likewise, these mutual fund studies condition prior beliefs about skill on the residual variance of the fund. In the present study, β plays a role that is roughly analogous to the intercept in these previous studies: $\beta = 0$ implies no predictability, and hence no "mispricing". However, in the time-series setting, it is not sufficient to condition β on residual variance Σ_u; β must also be conditioned on σ_x in order to produce a well-defined distribution for quantities of interest.

---

15 Goyal and Welch (2008) present an "encompassing forecast", which, while not Bayesian, has similar implications in that it downweights the predictability coefficient estimated from the data. Bayesian methods can be seen as formalizing this approach.

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

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

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

1.3. Posterior

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

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

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

$$\begin{aligned} p(B, \Sigma|D) \propto\ & \sigma_x^N |\Sigma|^{-\frac{T+N+4}{2}} \exp\left\{-\frac{1}{2}\beta^\top \left(\sigma_\eta^2 \sigma_x^{-2} \Sigma_u\right)^{-2} \beta\right\} \\ & \times \exp\left\{-\frac{1}{2}\sigma_x^{-2}(x_0 - \mu_x)^2\right\} \\ & \times \exp\left\{-\frac{1}{2}\mathrm{tr}\left[(Y - XB)^\top (Y - XB)\Sigma^{-1}\right]\right\} \end{aligned} \quad (16)$$

as a posterior.

This posterior does not take the form of a standard density function because of the presence of σ_x^2 in the prior and in the term in the likelihood involving x_0 (note that σ_x^2 is a nonlinear function of θ_1 and σ_v). However, we can sample from the posterior using the Metropolis–Hastings algorithm (see Chib and Greenberg (1995)). Define column vectors

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

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

1.4. Predictive distribution and portfolio choice

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

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

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

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

Following previous single-period portfolio choice studies (see, e.g. Baks et al. (2001) and Pastor (2000)), we assume that the investor has quadratic utility. The advantage of quadratic utility is that it implies a straightforward mapping between the moments of the predictive distribution of returns and portfolio choice. However, because our method produces an entire distribution function for returns, it can be applied to other utility functions, and to buy-and-hold investors with horizons longer than one quarter.

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

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

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

Given draws from the posterior distribution of the parameters α^j, β^j, Σ_u^j, and a value of x_t, a draw from the predictive distribution of asset returns is given by

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

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

2. Results

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

2.1. Data

All data are obtained from the Center for Research on Security Prices (CRSP). Excess stock and bond returns are formed by subtracting the quarterly return on the three-month Treasury bond from the quarterly return on the value-weighted NYSE-AMEX-NASDAQ index and the ten-year Treasury bond (from the CRSP indices file) respectively. The dividend-price ratio is constructed from monthly return data on the stock index as the sum of the previous twelve months of dividends divided by the current price. The natural logarithm of the dividend-price ratio is used as the predictor variable. The yield spread is equal to the continuously compounded yield on the zero-coupon five year bond (from the Fama–Bliss data set) less the continuously compounded yield on the three-month bond. Data on bond yields are available from the second quarter of 1952. We therefore consider quarterly observations from the second quarter of 1952 until the last quarter of 2004.

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

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

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

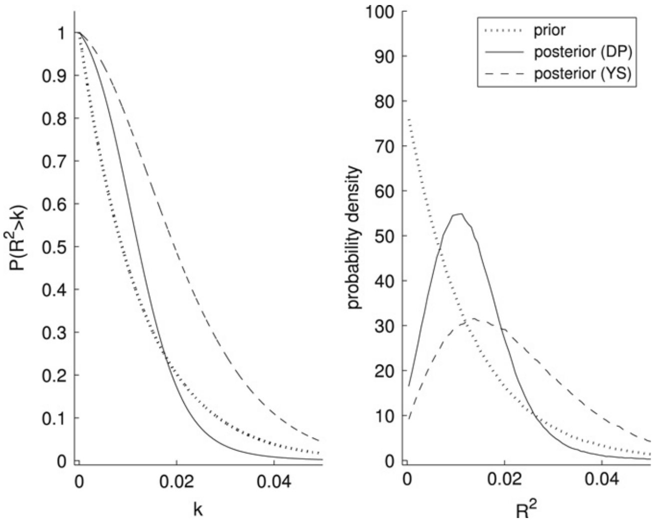

Fig. 2. Posterior distribution on the maximum R^2 (Notes: The left panel shows the prior and the posterior probability that the maximum R^2 over two assets exceeds a value k. The right panel shows the probability density function for the prior and the posterior of the maximum R^2. The prior volatility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

Because our empirical implementation assumes two assets, we compute the prior distribution of (14) assuming $N = 2$. The left panel of Fig. 2 reports the probability that the maximum R^2 exceeds k, as a function of k for both the prior with $\sigma_\eta = 0.08$ (dotted line) and for the posterior distribution given this prior. The solid line gives the posterior assuming that the dividend-price ratio is the predictor variable, while the dashed line gives the posterior assuming that the yield spread is the predictor variable. The right panel of Fig. 2 shows the probability density function of the posterior and of the prior.

We first discuss the result for the dividend-price ratio and then for the yield spread. Below $k = 0.02$, the posterior probability that the R^2 exceeds k is above the prior probability. Above 0.02, the posterior probability that the R^2 exceeds k is lower for the posterior than for the prior. While the prior density is decreasing in the R^2 over this range, the posterior density is hump-shaped with a maximum at about 0.02.

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

Table 1 reports posterior means for values of σ_η equal to 0, 0.04, 0.08, and ∞. To emphasize the economic significance of these priors, we report the corresponding probabilities that the R^2 exceeds 0.02: 0, 0.0005, 0.075, and 0.999.[19] The predictor variable is the dividend-price ratio. Posterior standard deviations are reported in parentheses. The table also shows results from estimation by ordinary least squares (OLS). For the OLS values, standard errors are reported in parentheses.

As Panel A of Table 1 shows, the dividend-price ratio predicts stock returns but not bond returns. The posterior mean for the β for bond returns is negative and small in magnitude. The posterior mean for the β for stock returns is positive for all of the priors we consider and for the OLS estimate. For the diffuse prior, the posterior mean of β equals to 1.46, below the OLS estimate of 2.72. As the prior becomes more informative, the posterior mean for β becomes smaller: for $P(R^2 > 0.02) = 0.075$, the estimate is 1.41, while for $P(R^2 > 0.02) = 0.0005$, it is 0.69.

Even though all of the priors are uninformative with respect to the autoregressive coefficient θ_1, the posterior mean of θ_1 nonetheless increases as the priors become more informative over β. The reason is the negative correlation between draws for θ_1 and draws for β. As Stambaugh (1999) shows, the negative correlation between shocks to returns and shocks to the predictor variable implies that when β is below its OLS value, θ_1 tends to be above its OLS value. The reason is that if β is below its OLS value, it must be that the lagged predictor variable and returns have an unusually high covariance in the sample (because the OLS value is "too high"). When this occurs, the predictor variable tends to have an unusually low autocorrelation; thus the OLS estimate for θ_1 is too low and the posterior mean will be above the OLS value. Therefore, placing a prior that weights the posterior mean of β toward zero raises the posterior mean of θ_1.[20]

Table 1 also reports posterior means and standard deviations for the means of the predictor variable and of excess returns.

Table 1
Posterior means.

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

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

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

20 The value for β under the diffuse prior is also substantially below the OLS value, and the value for θ_1 is higher. The reason is that, relative to the flat prior, the Jeffreys prior favors higher values of θ_1.

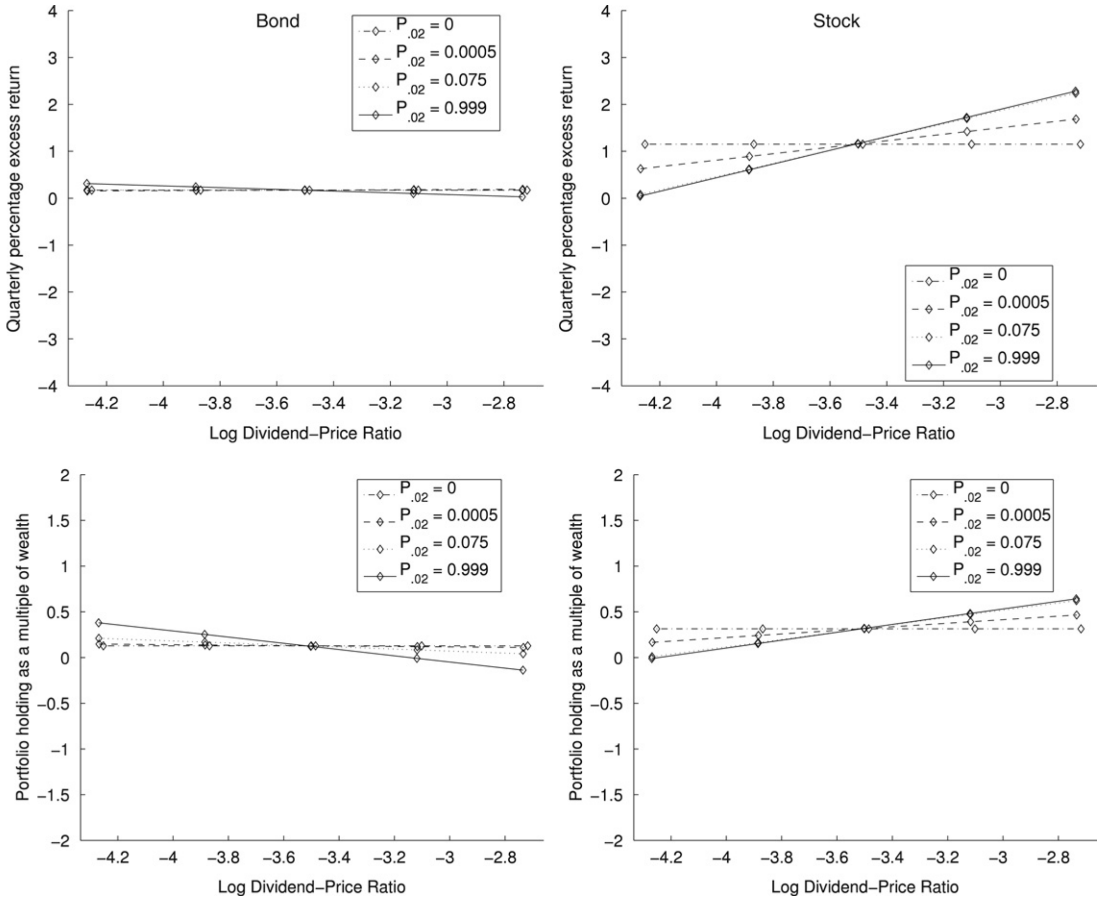

Fig. 3. Conditional expected returns and holdings when the dividend-price ratio predicts returns. (Notes: Conditional expected returns (top two plots) and portfolio holdings (bottom two plots) as functions of the log dividend-price ratio. Conditional expected returns are calculated using the predictive distribution. Given the predictive distribution, portfolios maximize mean–variance utility for risk aversion parameter $A = 5$. Assets are a stock index, a long-term bond, and the riskfree asset (not shown). Prior beliefs are indexed by $P_{.02}$, the probability that the R^2 from the predictive regression exceeds 2%. $P_{.02} = 0$ corresponds to the dogmatic prior; $P_{.02} = 0.999$ corresponds to the diffuse prior. Diamonds correspond to the sample mean and plus and minus one and two sample standard deviations of the predictor variable.)

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

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

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

We now examine the consequences of these posterior means for the predictive distribution of returns and for portfolio choice. Fig. 3 plots expected excess returns (top two plots) and optimal portfolio holdings (bottom two plots) as functions of the log dividend-price ratio. Graphs are centered at the sample mean. Diamonds denote plus and minus one and two sample standard deviations of the dividend-price ratio. We report results for the four prior beliefs discussed above; to save notation we let $P_{.02} = P(R^2 > 0.02)$.

The linear form of (1) implies that expected returns are linear in the predictor variables, conditional on the past data. The slope of the relation between the conditional return and x_t equals the posterior mean of β. Fig. 3 shows large deviations in the expected return on the stock on the basis of the dividend-price ratio. As the dividend-price ratio varies from −2 standard deviations to

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

$+2$ standard deviations, the expected return varies from 0% per quarter to 2% per quarter. On the other hand, the dividend-price ratio has virtually no predictive power for returns on the long-term bond.

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

For the diffuse prior, weights on the stock index vary substantially, from -30% when the dividend-price ratio is two standard deviations below its mean to 100% when the dividend-price ratio is two standard deviations above its mean. As the prior becomes more informative, expected returns and weights both vary less. However, this change happens quite slowly. Conditional expected returns under a prior that assigns only a 0.075 chance of an R^2 greater than 2% are nearly identical to conditional expected returns with a diffuse prior. There is sufficient evidence to convince even this skeptical investor to vary her portfolio to nearly the same degree as an investor with no skepticism at all. For a more skeptical prior with $P_{.02} = 0.0005$, differences emerge: the slope of the relation between expected returns and the dividend-price ratio is about half of what it was with a diffuse prior.

Fig. 4 displays analogous plots for the yield spread. Both the conditional expected bond return and the stock return increase substantially in the yield spread. For bonds, these expected returns vary between -2% and 2% per quarter as the yield spread varies between -2 and $+2$ standard deviations. For stocks, expected returns vary between 0% and 3%, similar to the variation with respect to the dividend-price ratio. These large variations in expected returns lead to similarly large variation in weights for the diffuse prior: for bonds, the weights vary between -200% and 200% as the yield spread varies between -2 and $+2$ standard deviations from the mean. For the stock, the weights vary between 0% and 75%. The variation in the weights on the stock appears less than the variation in expected returns on the stock; this is due to the positive correlation in return innovations between stocks and bonds.

Fig. 4 also shows that the more informative the prior, the less variable the weights. However, when the predictor variable is the yield spread, inference based on a skeptical prior with $P_{.02} = 0.075\%$ differs noticeably from inference based on a diffuse prior. Nonetheless, even the investors with skeptical priors choose portfolios that vary with the yield spread.

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

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

We next describe the implications of various prior beliefs for optimal weights over the postwar period. Starting in 1972, we compute the posterior distribution conditional on having observed data up to and including that year. We start in 1972 because this allows for twenty years of data for the first observation; this seems