

**Finance and Economics Discussion Series  
Divisions of Research & Statistics and Monetary Affairs  
Federal Reserve Board, Washington, D.C.**

**What is the Chance that the Equity Premium Varies Over Time?  
Evidence from Predictive Regressions**

**Jessica A. Wachter and Missaka Warusawitharana**

**2009-26**

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

What is the Chance that the Equity Premium Varies  
over Time?  
Evidence from Predictive Regressions \*

Jessica A. Wachter

University of Pennsylvania and NBER

Missaka Warusawitharana

Board of Governors of the Federal Reserve System

April 2, 2009

---

\*Wachter: Department of Finance, The Wharton School, University of Pennsylvania, 2300 SH-DH, Philadelphia, PA, 19104. [jwachter@wharton.upenn.edu](mailto:jwachter@wharton.upenn.edu), (215)898-7634. Warusawitharana: Division of Research and Statistics, Board of Governors of the Federal Reserve System, Mail Stop 97, 20th and Constitution Ave, Washington D.C, 20551. [missaka.n.warusawitharana@frb.gov](mailto:missaka.n.warusawitharana@frb.gov), (202)452-3461. We are grateful to Sean Campbell, Michael Johannes, Matthew Pritsker, Robert Stambaugh, Stijn van Nieuwerburgh, Jonathan Wright, Moto Yogo, Hao Zhou and seminar participants at the 2008 meetings of the American Finance Association, the 2007 CIRANO Financial Econometrics Conference, the 2007 Winter Meeting of the Econometric Society, the Federal Reserve Board, the University of California at Berkeley and the Wharton School for helpful comments. 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.

# What is the Chance that the Equity Premium Varies over Time? Evidence from Predictive Regressions

## **Abstract**

We examine the evidence on stock return predictability in a Bayesian setting that includes uncertainty about both the existence and strength of predictability. We consider an investor who believes that excess stock returns exhibit predictability with prior probability  $q < 1$ . In addition, the investor downweights observed predictability by placing a prior distribution on the  $R^2$  of the predictability regression. When we apply our analysis to the dividend-price ratio, we find that even investors who are quite skeptical about the existence and strength of predictability sharply modify their views in favor of predictability when confronted by the evidence. We depart from previous model-selection work by treating the regressor as stochastic rather than known; we find that this has a large impact on inference about time-varying expected returns.

# 1 Introduction

This paper investigates the evidence in favor of stock return predictability from a model-selection perspective. Much recent empirical work has focused on the predictive regression

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

where  $r_{t+1}$  denotes the return on a broad stock index in excess of the riskfree rate,  $x_t$  denotes a predictor variable, and  $u_{t+1}$  is a noise term. Taking expectations implies that  $\alpha + \beta x_t$  is the conditional equity premium. If  $\beta$  is not equal to zero, then the equity premium varies over time.

One approach to investigating whether stock returns are predictable involves running an ordinary least squares regression (OLS) on (1) and asking whether the predictive coefficient  $\beta$  is significantly different from zero. As emphasized in a simulation study by Kandel and Stambaugh (1996), however, this approach has the disadvantage that classical significance may not be indicative of whether the level of predictability is of economic significance. If  $\beta$  is found to be insignificant, or only marginally significant, one cannot conclude that predictability “does not exist” as far as economic agents are concerned.

In this study we adopt a Bayesian approach to inference on (1) that takes model uncertainty as well as parameter uncertainty into account. An investor evaluates the evidence in favor of equation (1) as opposed to a null hypothesis

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

The investor assigns a prior probability  $q$  to a state of the world where (1) describes returns (i.e. the equity premium is time-varying) and thus a prior probability  $1 - q$  to the state of the world where (2) describes returns (i.e. the equity premium is constant). The investor’s beliefs about returns after viewing the data involves assigning a posterior probability to (1), as well as a posterior distribution to the parameters of interest.

Our paper builds on several strands of the recent portfolio allocation literature. Once such strand studies properties of Bayesian estimation of predictive regressions (e.g. Bar-

beris (2000), Johannes, Polson, and Stroud (2002), Brandt, Goyal, Santa-Clara, and Stroud (2005), Pastor and Stambaugh (2008), Skoulakis (2007), Stambaugh (1999), Wachter and Warusawitharana (2009)), but assumes that the predictive model is known. A second strand focuses on model uncertainty, but assumes that the parameters within the model are known (e.g. Chen, Ju, and Miao (2009), Maenhout (2006), Hansen (2007)). A third strand allows for both model and parameter uncertainty, but assumes returns are independent and identically distributed (e.g. Chen and Epstein (2002), Garlappi, Uppal, and Wang (2007)).<sup>1</sup> Our paper builds on this work by assuming that the investor faces both parameter and model uncertainty, and considers the possibility that returns are predictable.

Our paper also builds on the literature on return predictability and model selection (Pesaran and Timmermann (1995), Avramov (2002), Cremers (2002)); these papers make the assumption that the future time path of the regressor is known, an assumption that is frequently satisfied in a standard ordinary least squares regression, but rarely satisfied in a predictive regression. By making use of methods developed in Wachter and Warusawitharana (2009), we are able to formulate and solve the investor’s problem when the regressor is stochastic. Our paper therefore incorporates the insights of the frequentist literature on predictive return regressions (e.g. Cavanagh, Elliott, and Stock (1995), Nelson and Kim (1993), Stambaugh (1999), Lewellen (2004), Torous, Valkanov, and Yan (2004), Campbell and Yogo (2006)) into a Bayesian portfolio selection setting.

When we apply our methods to predicting returns by the dividend-price ratio, we find that an investor who believes that there is a 20% probability of predictability prior to seeing the data updates to a 65% posterior probability after viewing quarterly postwar data. An advantage of modeling the stochastic process for the regressor is that we are able to compute certainty equivalent returns from exploiting predictability that do not depend on a particular value for the regressor. We find certainty equivalent returns of 1.16% per year when the

---

<sup>1</sup>Some of this work considers model uncertainty together with ambiguity aversion. In order to better focus on the affect of parameter and model uncertainty on the investor’s decision-making, we do not consider ambiguity aversion here.

dividend-price ratio is used as a predictor variable for an investor whose prior probability in favor of predictability is just 20%. For an investor who believes that there is a 50/50 chance of return predictability, certainty equivalent returns are 1.83%.

We also empirically evaluate the effect of using a full Bayes, exact likelihood approach as opposed to the conditional likelihood, and as opposed to empirical Bayes. A common approach to Bayesian inference in a time series setting is to treat the first observation of the predictor variable as a known parameter rather than a draw from the data generating process. However, we find that conditioning on the first observation results in Bayes factors (the ratio of the likelihood of model (1) to (2)) that are substantially smaller as compared to when the initial observation is treated as a draw from the data generating process. The posterior for the unconditional risk premium is highly unstable when we condition on the first observation. However, when this is treated as a draw from the data generating process, the expected return is estimated in a reliable way. In addition, using an empirical Bayes approach, which involves using data on the regressor to determine the prior, implies Bayes factors that are larger than those implied by the fully Bayesian approach. Conditioning on the first observation and using empirical Bayes are often regarded as approximation techniques to the full Bayes exact likelihood approach that we emphasize (e.g. Box and Tiao (1973), Chipman, George, and McCulloch (2001)). Our results suggest that, at least for some purposes, this approximation may be less accurate than previously believed.

## 2 Model

### 2.1 Data generating processes

Let  $r_{t+1}$  denote continuously compounded excess returns on a stock index from time  $t$  to  $t+1$  and  $x_t$  the value of a (scalar) predictor variable. We assume that this predictor variable follows the process

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

Stock returns can be predictable, in which case they follow the process (1) or unpredictable, in which case they follow the process (2). In either case, errors are serially uncorrelated, homoskedastic, and jointly normal:

$$\begin{bmatrix} u_{t+1} \\ v_{t+1} \end{bmatrix} | r_t, \dots, r_1, x_t, \dots, x_0 \sim N(0, \Sigma), \quad (4)$$

and

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}. \quad (5)$$

As we show below, the correlation between innovations to returns and innovations to the state variable implies that (3) affects inference about returns, even when there is no predictability.

When the process (3) is stationary, i.e.  $\rho$  is between -1 and 1, the state variable has an unconditional mean of

$$\mu_x = \frac{\theta}{1 - \rho} \quad (6)$$

and a variance of

$$\sigma_x^2 = \frac{\sigma_v^2}{1 - \rho^2}. \quad (7)$$

These follow from taking unconditional means and variances on either side of (3). Note that these are population values conditional on knowing the parameters. Given these, the population  $R^2$  is defined as

$$\text{Population } R^2 = \frac{\beta^2 \sigma_x^2}{\beta^2 \sigma_x^2 + \sigma_u^2}.$$

## 2.2 Prior Beliefs

An investor's prior views on predictability can be elicited by the answer to two straightforward questions.<sup>2</sup> Consider data generating processes of the form (1) and (2). Given these processes, the investor should answer:

---

<sup>2</sup>The basic structure of these prior beliefs is analogous to that used by Baks, Metrick, and Wachter (2001) in the setting of mutual fund performance evaluation.

- [Question 1] What is the probability that predictability exists, i.e. that equation (1) describes returns for some  $\beta \neq 0$ ? (Call this answer  $q$ .)
- [Question 2] Given that predictability exists, what is the probability that the  $R^2$  exceeds 1%? (Call this answer  $P_{.01}$ .)

The answer to Question 2 will be conditional on the frequency; for most of our results, quantities will be measured at an annual frequency. Note that Question 2 is not asking about the probability of achieving an  $R^2$  in a given sample, which depends on sampling variability. It is asking about the  $R^2$  that would result if the time period goes to infinity. The use of 1% is arbitrary; any other value that is greater than 0 could be substituted.

We now demonstrate how to specify priors given the answers to these questions. An appeal of this approach is that it is not necessary to specify aspects of the distribution of the predictor variable and of returns other than those given above. The prior beliefs are invariant to changes to these aspects of the distribution.

### 2.2.1 Full Bayes priors

Let  $H_0$  denote the state of the world in which excess returns are unpredictable (the “null”) and  $H_1$  denote the state of the world in which there is some amount of excess return predictability. Then  $q$  is the prior probability of  $H_1$ , i.e.  $q = p(H_1)$ . In what follows, we construct priors for the parameters conditional on  $H_0$  and on  $H_1$ . It is convenient to group the regression parameters in equations (1), (2) and (3) into vectors

$$b_0 = [\alpha, \theta, \rho]^\top$$

and

$$b_1 = [\alpha, \beta, \theta, \rho]^\top.$$

We then specify the prior  $p(b_0, \Sigma | H_0)$ , which is the prior on  $b_0$  and  $\Sigma$  conditional on no predictability and the prior  $p(b_1, \Sigma | H_1)$ , which is the prior on  $b_1$  and  $\Sigma$  conditional on the existence of predictability.<sup>3</sup>

---

<sup>3</sup>Formally we could write down  $p(b_1, \Sigma | H_0)$  by assuming  $p(\beta | b_0, \Sigma, H_0)$  is a point mass at zero.



Note that  $p(b_1, \Sigma | H_1)$  can also be written as  $p(\beta, b_0, \Sigma | H_1)$ . We set the prior on  $b_0$  and  $\Sigma$  so that

$$p(b_0, \Sigma | H_0) = p(b_0, \Sigma | H_1) = p(b_0, \Sigma).$$

We assume the investor has uninformative beliefs on these parameters. We follow the approach of Stambaugh (1999) and Zellner (1996), and derive a limiting Jeffreys prior as explained in Appendix A. As Appendix A shows, this limiting prior takes the form

$$p(b_0, \Sigma) \propto \sigma_x \sigma_u |\Sigma|^{-\frac{5}{2}}, \quad (8)$$

for  $\rho \in (-1, 1)$ , and zero otherwise.

The parameter that distinguishes  $H_0$  from  $H_1$  is  $\beta$ . One approach would be to write down a prior distribution for  $\beta$  unconditional on the remaining parameters. However, it is difficult to think about priors on  $\beta$  in isolation from beliefs about other parameters. For example, a high variance of  $x_t$  might lower one's prior on  $\beta$ , while a large residual variance of  $r_t$  might raise it. Rather than placing a prior on  $\beta$  directly, we follow Wachter and Warusawitharana (2009) and place a prior on the population  $R^2$ . To implement this prior on the  $R^2$ , we place a prior on "normalized"  $\beta$ , that is  $\beta$  adjusted for the variance of  $x$  and the variance of  $u$ . Let

$$\eta = \sigma_u^{-1} \sigma_x \beta.$$

denote normalized  $\beta$ . We assume that prior beliefs on  $\eta$  are given by

$$\eta | H_1 \sim N(0, \sigma_\eta^2) \quad (9)$$

The population  $R^2$  is closely related to  $\eta$ :

$$\text{Population } R^2 = \frac{\beta^2 \sigma_x^2}{\beta^2 \sigma_x^2 + \sigma_u^2} = \frac{\eta^2}{\eta^2 + 1}. \quad (10)$$

Equation (10) provides a mapping between a prior distribution on  $\eta$  and a prior distribution on the population  $R^2$ . Given an  $\eta$  draw, an  $R^2$  draw can be computed using (10).

A prior on  $\eta$  implies a hierarchical prior on  $\beta$ . Because

$$p(\beta, b_0, \Sigma | H_1) = p(\beta | b_0, \Sigma, H_1) p(b_0, \Sigma | H_1),$$

it suffices to choose a prior for  $\beta$  conditional on the other parameters. The prior for  $\eta$ , (9), implies

$$\beta|\alpha, \theta, \rho, \Sigma \sim N(0, \sigma_\beta^2), \quad (11)$$

where

$$\sigma_\beta = \sigma_\eta \sigma_x^{-1} \sigma_u.$$

Because  $\sigma_x$  is a function of  $\rho$  and  $\sigma_v$ , the prior on  $\beta$  is also implicitly a function of these parameters. The parameter  $\sigma_\eta$  indexes the degree to which the prior is informative. As  $\sigma_\eta \rightarrow \infty$ , the prior over  $\beta$  becomes uninformative; all values of  $\beta$  are viewed as equally likely. As  $\sigma_\eta \rightarrow 0$ , the prior converges to  $p(b_0, \Sigma)$  multiplied by a point mass at 0, implying a dogmatic view in no predictability. Combining (11) with (8) implies the joint prior under  $H_1$ :

$$\begin{aligned} p(b_1, \Sigma|H_1) &= p(\beta|b_0, \Sigma, H_1)p(b_0|H_1) \\ &\propto \frac{1}{\sqrt{2\pi\sigma_\eta^2}} \sigma_x^2 |\Sigma|^{-\frac{5}{2}} \exp \left\{ -\frac{1}{2} \beta^2 (\sigma_\eta^2 \sigma_x^{-2} \sigma_u^2)^{-1} \right\}. \end{aligned} \quad (12)$$

Jeffreys invariance theory provides an independent justification for modeling priors on  $\beta$  as (11). Stambaugh (1999) shows that the limiting Jeffreys prior for  $b_1$  and  $\Sigma$  equals

$$p(b_1, \Sigma|H_1) \propto \sigma_x^2 |\Sigma|^{-\frac{5}{2}}. \quad (13)$$

This prior corresponds to the limit of (12) as  $\sigma_\eta$  approaches infinity. Modeling the prior for  $\beta$  as depending on  $\sigma_x$  not only has a convenient interpretation in terms of the distribution of the  $R^2$ , but also implies that an infinite prior variance represents ignorance as defined by Jeffreys (1961). Note that a prior on  $\beta$  that is independent of  $\sigma_x$  would not have this property.

Figure 1 shows the resulting distribution for the population  $R^2$  for various values of  $\sigma_\eta$ . Panel A shows the distribution conditional on  $H_1$  while Panel B shows the unconditional distribution. More precisely, for any value  $k$ , Panel A shows the prior probability that the  $R^2$  exceeds  $k$ , conditional on the existence of predictability. For large values of  $\sigma_\eta$ , e.g. 100, the prior probability that the  $R^2$  exceeds  $k$  across the relevant range of values for the  $R^2$  is

close to one. The lower the value of  $\sigma_\eta$ , the less variability in  $\beta$  around its mean of zero, and the lower the probability that the  $R^2$  exceeds  $k$  for any value of  $k$ . Panel B shows the unconditional probability that the  $R^2$  exceeds  $k$  for any value of  $k$ , assuming that the prior probability of predictability,  $q$ , is equal to 0.5. By the definition of conditional probability:

$$p(R^2 > k) = p(R^2 > k | H_1)q.$$

Therefore Panel B takes the values in Panel A and scales them down by 0.5. To distinguish (8) and (12) from an alternative set of priors that we describe in the following section, we refer to these as *full Bayes priors*.

### 2.2.2 Empirical Bayes priors

A second approach to formulating priors involves conditioning on moments of the data. Let  $T$  denote the length of the sample and  $\hat{\sigma}_x$  the sample variance of  $x$ :

$$\hat{\sigma}_x = \frac{1}{T} \sum_{t=1}^T \left( x_t - \frac{1}{T} \sum_{s=1}^T x_s \right)^2.$$

One specification for the prior, introduced by Fernandez, Ley, and Steel (2001), is as follows:

$$p(\beta | \sigma_u^2, H_1) = N(0, \kappa \sigma_u^2 \hat{\sigma}_x^{-1}), \quad (14)$$

where  $\kappa$  is a constant that determines the informativeness of the prior, and

$$p(\sigma_u) \propto \sigma_u^{-1}. \quad (15)$$

The specification is completed by setting

$$p(\alpha) \propto 1. \quad (16)$$

These assumptions on the prior are combined with the likelihood

$$p(D | \alpha, \beta, \sigma_u, H_1) = (2\pi\sigma_u^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \sum_{t=0}^{T-1} (r_{t+1} - \alpha - \beta x_t)^2 \sigma_u^{-2} \right\} \quad (17)$$

and

$$p(D | \alpha, \beta, \sigma_u, H_0) = (2\pi\sigma_u^2)^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \sum_{t=0}^{T-1} (r_{t+1} - \alpha)^2 \sigma_u^{-2} \right\}. \quad (18)$$

Very similar specifications are employed by Chipman, George, and McCulloch (2001), Cremers (2002), Wright (2003) and Stock and Watson (2005). Note that these equations display the marginal likelihood over the return equations (1) and (2) rather than the full likelihood that includes the data generating process for  $x_t$ . An appeal of this formulation for the prior is that it leads to analytical expressions for the posterior distribution and for the Bayes factor (in fact, it is closely related to the “ $g$ -prior” of Zellner (1996)).

The above assumptions are most reasonable in the case where  $x_1, \dots, x_T$  are observed at time 0. While this holds in many applications of OLS regression, it holds rarely, if ever, in the case of predictive regressions in financial time series. Moreover, were  $x_1, \dots, x_T$  observed, the contemporaneous correlation between  $x_t$  and  $r_t$  would invalidate the likelihoods (17) and (18) because the value of  $x_t$  would convey information about  $r_t$  not reflected in these likelihoods. One way to interpret the above in the setting where  $x$  is stochastic is to assume that, while the data on  $x_t$  themselves are unobserved, certain functions of the data, namely sample moments of  $x_t$  such as  $\hat{\sigma}_x$ , are observed. Allowing data to influence the prior is generally referred to as the “empirical Bayes” method.<sup>4</sup> For this reason, the formulation of priors that use moments from the sample could be thought of as an example of empirical Bayes, at least if one accepts a broad definition of the term.<sup>5</sup>

Regardless of its theoretical attractiveness, it is of interest to ask whether the use of empirical Bayes in this setting make a difference in practice. There are a number of differences

---

<sup>4</sup>However, in traditional applications of empirical Bayes, the term has generally implied 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 firms in that industry (Pastor and Stambaugh (1999)).

<sup>5</sup>Avramov (2002) uses marginal likelihoods analogous to (17) and (18), but formulates the prior by assuming that the agent observes a prior sample with moments similar to the existing sample, but without predictability. This is also an example of the empirical Bayes approach.

between the specification described in (14)–(18) and ours. Most importantly, by assuming the investor knows the sample moments of  $x$ , the above approach avoids the need to make explicit assumptions on the prior for the parameters of the  $x$  process and for the likelihood of the  $x$  process. However, as we show, these assumptions, whether hidden or explicit, have important consequences for the posterior distribution.

Leaving these issues aside for the moment, our immediate goal is to write down a version of the above specification that is close enough to our model so that differences in results stemming from the link (or lack thereof) between the distribution of  $\sigma_x$  and that of  $\beta$  can be interpreted. To this end, we consider the specification

$$p(\beta|b_0, \Sigma, H_1) \sim N(0, \hat{\sigma}_\beta^2),$$

where

$$\hat{\sigma}_\beta = \sigma_\eta \hat{\sigma}_x^{-1} \hat{\sigma}_u.$$

We compute  $\hat{\sigma}_u$  as the standard deviation of the residual from OLS regression of the predictive regression.<sup>6</sup> Note that these priors do not imply a proper prior distribution for the  $R^2$ . Therefore they cannot be used to answer Question 2 posed above. In order to compare the empirical Bayes and the full Bayes priors, we use the same values of  $\sigma_\eta$  to form  $\hat{\sigma}_\beta$  as we use to form  $\sigma_\beta$ .

We assume a standard uninformative prior for the remaining parameters (see Zellner (1996) and Gelman, Carlin, Stern, and Rubin (2004)): with a normal distribution for  $\beta$ , where the prior covariance reflects the agent’s beliefs about predictability. We also ensure that  $x_t$  is stationary. That is:

$$p(b_0, \Sigma|H_1) = p(b_0, \Sigma|H_0) \propto |\Sigma|^{-\frac{3}{2}}, \tag{19}$$

for  $\rho \in (-1, 1)$ , and zero otherwise. It follows that

$$p(b_1, \Sigma|H_1) \propto \frac{1}{\sqrt{2\pi\hat{\sigma}_\beta^2}} |\Sigma|^{-\frac{3}{2}} \exp \left\{ -\frac{1}{2} \beta^2 \hat{\sigma}_\beta^{-2} \right\} \tag{20}$$

---

<sup>6</sup>For simplicity, we do not incorporate a link between  $\hat{\sigma}_u$  and  $\beta$  as in (14). Because  $\sigma_u$  is estimated very precisely (unlike  $\sigma_x$ ), this is unlikely to make a large difference in the results.

These priors may be thought of as the simplest set of priors which contain information about the distribution of  $\beta$ , the coefficient on return predictability. In what follows, we refer to these as *empirical Bayes priors*. We combine these priors with the same likelihood as used for the full Bayes prior, described below.

## 2.3 Likelihood

### 2.3.1 Likelihood under $H_1$

Under  $H_1$ , returns and the state variable follow the joint process given in (1) and (3). It is convenient to group observations on returns and contemporaneous observations on the state variable into a matrix  $Y$  and lagged observations on the state variable and the constant into a matrix  $X$ . Let

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

and let

$$z = \text{vec}(Y)$$

$$Z_1 = I_2 \otimes X.$$

In the above, the  $\text{vec}$  operator stacks the elements of the matrix columnwise. It follows that the likelihood conditional on  $H_1$  and on the first observation  $x_0$  takes the form of

$$p(D|b_1, \Sigma, x_0, H_1) = |2\pi\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2}(z - Z_1 b_1)^\top (\Sigma^{-1} \otimes I_T) (z - Z_1 b_1) \right\} \quad (21)$$

(see Zellner (1996)).

The likelihood function (21) conditions on the first observation of the predictor variable,  $x_0$ . Stambaugh (1999) argues for treating  $x_0$  and  $x_1, \dots, x_T$  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. 265) imply that  $x_0$  is a draw from a multivariate normal distribution with mean  $\mu_x$  and standard deviation  $\sigma_x$ . Combining

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

$$p(D|b_1, \Sigma, H_1) = |2\pi\sigma_x^2|^{-\frac{1}{2}}|2\pi\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2}(x_0 - \mu_x)^2 \sigma_x^{-2} - \frac{1}{2}(z - Z_1 b_1)^\top (\Sigma^{-1} \otimes I_T) (z - Z_1 b_1) \right\}. \quad (22)$$

Following Box and Tiao (1973), we refer to (21) as the *conditional likelihood* and (22) as the *exact likelihood*.

### 2.3.2 Likelihood under $H_0$

Under  $H_0$ , returns and the state variable follow the processes given in (2) and (3). Let

$$Z_0 = \begin{bmatrix} \iota_T & 0_{T \times 2} \\ 0_{T \times 1} & X \end{bmatrix},$$

where  $\iota_T$  is the  $T \times 1$  vector of ones. Then the conditional likelihood can be written as

$$p(D|b_0, \Sigma, x_0, H_0) = |2\pi\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2}(z - Z_0 b_0)^\top (\Sigma^{-1} \otimes I_T) (z - Z_0 b_0) \right\}. \quad (23)$$

Using similar reasoning as in the  $H_1$  case, the exact likelihood is given by

$$p(D|b_0, \Sigma, H_0) = |2\pi\sigma_x^2|^{-\frac{1}{2}}|2\pi\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2}(x_0 - \mu_x)^2 \sigma_x^{-2} - \frac{1}{2}(z - Z_0 b_0)^\top (\Sigma^{-1} \otimes I_T) (z - Z_0 b_0) \right\}. \quad (24)$$

As above, we refer to (23) as the *conditional likelihood* and (24) as the *exact likelihood*.

## 2.4 Posterior distribution

The investor updates his prior beliefs to form the posterior distribution upon seeing the data. As we discuss below, this posterior requires the computation of two quantities: the posterior of the parameters conditional on the absence or existence of return predictability, and the posterior probability that returns are predictable. Given these two quantities, we can simulate from the posterior distribution.

To compute the posteriors conditional on the absence or existence of return predictability, we apply Bayes' rule conditioning on  $H_0$  and conditioning on  $H_1$ . It follows from Bayes' rule that

$$p(b_0, \Sigma | H_0, D) \propto p(D | b_0, \Sigma, H_0) p(b_0, \Sigma | H_0) \quad (25)$$

is the posterior conditional on  $H_0$  and that

$$p(b_1, \Sigma | H_1, D) \propto p(D | b_1, \Sigma, H_1) p(b_1, \Sigma | H_1) \quad (26)$$

is the posterior conditional on  $H_1$ . Because  $\sigma_x$  is a nonlinear function of the underlying parameters, the posterior distributions conditional on  $H_0$  and  $H_1$  are nonstandard and must be computed numerically. We can sample from these distributions quickly and accurately using the Metropolis-Hastings algorithm (see Chib and Greenberg (1995), Johannes and Polson (2006)). See Appendix B for details.

Let  $\bar{q}$  denote the posterior probability that excess returns are predictable. By definition,

$$\bar{q} = p(H_1 | D).$$

It follows from Bayes' rule, that

$$\begin{aligned} \bar{q} &= \frac{p(D | H_1) q}{p(D | H_1) q + p(D | H_0) (1 - q)} \\ &= \frac{\mathcal{B}_{10} q}{\mathcal{B}_{10} q + (1 - q)}, \end{aligned} \quad (27)$$

where

$$\mathcal{B}_{10} = \frac{p(D | H_1)}{p(D | H_0)} \quad (28)$$

is the Bayes factor for the alternative hypothesis of predictability against the null of no predictability. The Bayes factor is a likelihood ratio in that it is the likelihood of return predictability divided by the likelihood of no predictability. However, it differs from the standard likelihood ratio in that the likelihoods  $p(D | H_i)$  are not conditional on the values of the parameters. In fact, these likelihoods can be formally written as

$$p(D | H_0) = \int p(D | b_0, \Sigma, H_0) p(b_0, \Sigma | H_0) db_0 d\Sigma \quad (29)$$



and

$$p(D|H_1) = \int p(D|b_1, \Sigma, H_1)p(b_1, \Sigma|H_1) db_1 d\Sigma. \quad (30)$$

To form  $p(D|H_0)$  and  $p(D|H_1)$ , the likelihood conditional on parameters (the likelihood function generally used in classical statistics) is integrated over the prior distribution of the parameters. Under our distributions, these integrals cannot be computed analytically. However, the Bayes factor (28) can be computed directly using the generalized Savage-Dickey ratio (Dickey (1971), Verdinelli and Wasserman (1995)). Details can be found in Appendix C.

Putting these two pieces together, we draw from the posterior parameter distribution by drawing from  $p(b_1, \Sigma|D, H_1)$  with probability  $\bar{q}$  and from  $p(b_0, \Sigma|D, H_0)$  with probability  $1 - \bar{q}$ .

### 3 Results

We now apply the above framework to understanding the predictive power of the dividend-price ratio and payout yield for the excess return on a broad equity index.

#### 3.1 Data

We use data from the Center for Research on Security Prices (CRSP). We compute excess stock returns by subtracting the continuously compounded 3-month treasury bill return from the return on the value-weighted CRSP index at annual and quarterly frequencies. Following a large portfolio selection literature (see, e.g., Brennan, Schwartz, and Lagnado (1997), Campbell and Viceira (1999)), we focus on the dividend-price ratio as the predictive factor. The dividend-price ratio is computed by dividing the dividend payout over the previous 12 months with the current price of the stock index. The use of 12 months of data accounts for seasonalities in dividend payments. We use the logarithm of the dividend-price ratio as the predictive factor. We also use the repurchases-adjusted payout yield of Boudoukh, Michaely, Richardson, and Roberts (2007) as a predictive factor. Data are annual data from 1927 to

the beginning of 2005; we also report results with the dividend-price ratio at a quarterly frequency from 1952 onwards.

### 3.2 Bayes factors and posterior means

Table 1 reports Bayes factors and posterior means when the payout yield is used as a predictor variable. Table 2 and 3 report analogous results for the dividend-price ratio in annual data and in quarterly postwar data respectively. Each table reports results for full Bayes priors combined with the exact likelihood, for full Bayes priors combined with the conditional likelihood and for empirical Bayes priors combined with the exact likelihood. For each prior and likelihood combination, four values of  $\sigma_\eta$  are considered: 0.05, 0.09, 0.15 and 100. For the full Bayes priors, these translate into values of  $P_{.01}$  (the prior probability that the  $R^2$  exceeds 0.01) equal to 0.05, 0.25, 0.50 and 0.99 respectively. For the empirical Bayes priors, the prior distribution over the  $R^2$  is not well defined. We construct these priors using the same values of  $\sigma_\eta$  as the full Bayes counterparts. Because the results are qualitatively similar across the three data sets, we focus on results for the payout yield in Table 1.

Table 1 shows that the Bayes factor is hump-shaped in  $P_{.01}$  for each prior-likelihood combination. For small values of  $P_{.01}$ , the Bayes factor is close to one. For large values, the Bayes factor is close to zero. Both results can be understood using the formula for the Bayes factor in (28) and for the likelihoods  $p(D | H_1)$  and  $p(D | H_0)$  in (29) and (30). For low values of  $P_{.01}$ , the investor imposes a very tight prior on the  $R^2$ . Therefore the hypotheses that returns are predictable and that returns are unpredictable are nearly the same. It follows from (29) and (30) that the likelihoods of the data under these two scenarios are nearly the same and that the Bayes factor is nearly one. This is intuitive: when two hypotheses are close, a great deal of data are required to distinguish one from the other.

The fact that the Bayes factor approaches zero as  $P_{.01}$  increases is less intuitive. The reduction in Bayes factors implies that, as the investor allows a greater range of values for the  $R^2$ , the posterior probability that returns are predictable approaches zero. This effect

is known as Bartlett’s paradox, and was first noted by Bartlett (1957) in the context of distinguishing between uniform distributions. As Kass and Raftery (1995) discuss, Bartlett’s paradox makes it crucial to formulate an informative prior on the parameters that differ between  $H_0$  and  $H_1$ . The mathematics leading to Bartlett’s paradox are most easily seen in a case where Bayes factors can be computed in closed form. However, we can obtain an understanding of the paradox based on the form of the likelihoods  $p(D | H_1)$  and  $P(D | H_0)$ . These likelihoods involve integrating out the parameters using the prior distribution. If the prior distribution on  $\beta$  is highly uninformative, the prior places a large amount of mass in extreme regions of the parameter space. In these regions, the likelihood of the data conditional on the parameters will be quite small. At the same time, the prior places a relatively small amount of mass in the regions of the parameter space where the likelihood of the data is large. Therefore  $P(D | H_1)$  (the integral of the likelihood under  $H_1$ ) is small relative to  $P(D | H_0)$  (the integral of the likelihood under  $H_0$ ).

Table 1 also shows that there are substantial differences between the Bayes factors resulting from the exact versus the conditional likelihood and from empirical versus full Bayes. The Bayes factors resulting from the exact likelihood are larger than those resulting from the conditional likelihood, thus implying a greater posterior probability of return predictability. The Bayes factors resulting from full Bayes are smaller than those resulting from empirical Bayes, implying a lower posterior probability of return predictability.

In what follows, we seek to explain these patterns in the Bayes factors. Let  $\bar{\beta}$  be the posterior mean of  $\beta$  conditional on predictability and  $\bar{\rho}$  the posterior mean of  $\rho$  conditional on predictability. As Table 1 shows, differences in Bayes factors between specifications reflect differences in  $\bar{\beta}$ . That is, for any given value of  $P_{.01}$ ,  $\bar{\beta}$  is higher for the exact likelihood than for the conditional likelihood, and lower for full Bayes than for empirical Bayes. Moreover, the opposite pattern is evident for  $\bar{\rho}$ . The negative correlation between  $\rho$  and  $\beta$  is also noted by Stambaugh (1999)). The source of this negative relation is the negative correlation between shocks to returns and shocks to the predictor variable. Suppose that a draw of  $\beta$  is below its value predicted by ordinary least squares (OLS). This implies that the OLS value

for  $\beta$  is “too high”, i.e. in the sample shocks to the predictor variable are followed by shocks to returns of the same sign. Therefore shocks to the predictor variable tend to be followed by shocks to the predictor variable that are of different signs. Thus the OLS value for  $\rho$  is “too low”. This explains why values of  $\bar{\rho}$  are higher for low values of  $P_{.01}$  (and hence low values of  $\bar{\beta}$ ) than for high values, and higher than the ordinary least squares estimate.

We can use the connection between  $\bar{\rho}$ ,  $\bar{\beta}$  and the Bayes factor to account for differences between the Bayes factors between the prior and likelihood specifications. As Table 1 shows, using the exact likelihood leads to lower posterior values of  $\rho$ . This is because the exact likelihood leads to more precise estimates of  $\mu_x$ . By the argument in the previous paragraph, this implies greater posterior values for  $\beta$  and higher Bayes factors.

On the other hand, the use of full rather than empirical Bayes implies higher posterior values of  $\rho$ . This occurs because the full Bayes prior, on account of the  $\sigma_x^2$  term, puts more weight on high values of  $\sigma_x$  and therefore high values of  $\rho$ . When  $\beta$  is not far from zero, the posterior distribution is higher for lower values of  $\sigma_\beta$ , and hence higher values of  $\sigma_x$ . This leads to lower posterior means of  $\beta$  and lower Bayes factors.

Tables 1–3 also report the posterior means of excess returns (the equity premium) and of the predictor variable conditional on predictability. In each case, the OLS row reports the sample mean of excess returns and the sample mean of the predictor variable.<sup>7</sup> Posterior means conditional on no predictability are very close to their counterparts for  $P_{.01} = .05$ . Surprisingly, the various choices for the predictor variable and for the prior and likelihood imply different values for the equity premium. For example, the sample average for excess returns over the 1927 to 2004 period is 5.85% per annum. In contrast, the full Bayes exact likelihood approach generates average returns that range from 5.05% to 5.24% per annum

---

<sup>7</sup>Posterior means for  $r$  and  $x$  integrate out over uncertainty in the predictor variables. In the case of returns, for example, we compute

$$E[r|D, H_1] = E \left[ \alpha + \beta \frac{\theta}{1 - \rho} | H_1 \right],$$

where the expectation on the right hand side is taken over the posterior distribution for the parameters.

depending on the informativeness of the prior (the more informative the prior, the higher the excess return).

The differences in the estimates of the equity premium arise from differences in estimates of the mean of the predictor variable. The conditional maximum likelihood estimate of the mean of  $x$  (not reported) is -3.54. The posterior mean implied by the exact likelihood is between -3.16 and -3.17 (depending on the prior). Thus according to the model, shocks to the predictor variable over the sample period must be negative for -3.54 to be the estimated value when the conditional likelihood is used. It follows that the shocks to excess returns must be positive (because of the negative correlation). Therefore the posterior mean is below the sample mean. This effect also operates in the case of the dividend-price ratio and is in fact more dramatic. In annual data from 1927 to 2004, the implied means for excess returns range from 4.02 to 4.71% per annum versus the sample mean of 5.85%.

While the use of empirical Bayes implies values for the posterior mean of  $r$  that are similar to those for full Bayes, the use of the conditional likelihood implies estimates that are highly variable and can even be negative. This is because of the lack of precision in estimating  $\mu_x$ .

Tables 1–3 demonstrate differences in the posterior distribution depending on whether one uses full Bayes or empirical Bayes, and whether one uses the exact likelihood or the conditional likelihood. In what follows, we will examine the full Bayes, exact likelihood case more closely, and show its implications for inference on return predictability. The following two sections examine statistical measures: the posterior likelihood of predictability and the posterior distribution of the  $R^2$ . The final section examines economic significance of the predictability evidence through certainty equivalent returns.

### 3.3 Posterior likelihood of predictability

We now examine the posterior probability that excess returns are predictable. Given a Bayes factor and a prior belief on the existence of predictability  $q$ , the posterior probability

of predictability  $\bar{q}$  can be computed using equation (27). The greater the investor's prior belief about predictability, the greater is his posterior belief. The greater is the Bayes factor, the greater is the posterior belief. As described in the previous section, the Bayes factor itself depends on the other aspect of the investor's prior: the prior probability that the  $R^2$  exceeds 1% should predictability exist.

Table 4 presents the posterior probabilities of predictability as a function of the investor's prior about the existence of predictability,  $q$ , and the prior belief on the strength of predictability,  $P_{.01}$ . We consider the posterior resulting from full Bayes priors and the exact likelihood. The posterior probability is increasing in  $q$  and hump-shaped in  $P_{.01}$ , reflecting the fact that the Bayes factors are hump-shaped in  $P_{.01}$ . The results demonstrate that investors with moderate beliefs on both the existence and strength of predictability revise their beliefs on the existence on predictability sharply upward. For example, an investor with  $q = 0.5$  and  $P_{.01} = 0.50$  conclude that the posterior likelihood of predictability equals 0.88 using the payout yield to predict annual returns. This result is robust to a wide range of choices for  $P_{.01}$ . As the table shows,  $P_{.01} = 0.25$  implies a posterior probability of 0.74. The posterior probability falls off dramatically as  $P_{.01}$  approaches one; for these very diffuse priors (which imply what might be considered an economically unreasonable amount of predictability), the Bayes factors are close to zero.

While the evidence is slightly weaker when the dividend-price ratio is used in annual data, the dividend-price ratio combined with quarterly post-war data implies stronger evidence in favor of predictability. In particular,  $q = .50$  implies posterior probabilities of predictability above 0.80 for all but the most diffuse prior.

This section has examined an important aspect of the posterior distribution: the probability that returns are predictable. In what follows, we examine the full posterior for the  $R^2$  of the predictability relation.

### 3.4 Posterior $R^2$ values

We measure the investor's prior beliefs about the strength of predictability using the metric  $P(R^2 > 1\%|H_1) = P_{.01}$ . It is therefore of interest to examine the posterior beliefs over the  $R^2$ . We consider posteriors derived from the full Bayes prior and the exact likelihood.

Figure 2 shows two plots on the prior and posterior distribution of the  $R^2$  with priors  $P(R^2 > 1\%|H_1) = 0.50$  and  $q = 0.5$  using the payout yield to predict annual returns. Panel A plots  $P(R^2 > k)$  as a function of  $k$  for both the prior and the posterior; this corresponds to 1 minus the cumulative density function of the  $R^2$ .<sup>8</sup> The plot for the  $P(R^2 > k)$  demonstrates a clear rightward shift for the posterior for values of  $k$  up to 0.15 (both the prior and the posterior place similarly low probabilities that the  $R^2$  exceeds 0.15). The strength of the predictability can be seen in that while the prior implies  $P(R^2 > 1\%) = 0.25$ , the posterior implies  $P(R^2 > 1\%)$  close to 0.85. Thus, after observing the data, an investor revises his beliefs on the strength of predictability substantially upward. Panel B plots the probability density function of the  $R^2$ . The full Bayes prior places the highest density on low values of the  $R^2$ . The posterior however places high density in the region around 5% and has lower density than the prior for  $R^2$  values less than 2%. The evidence in favor of predictability, with a moderate  $R^2$ , is sufficient to overcome the investor's initial skepticism.

Figure 3 shows the comparable plots using the dividend-price ratio to predict annual returns. Results are similar to those discussed for the payout yield. The posterior probability of  $P(R^2 > k)$  is again higher than the prior probability for  $k$  ranging from 0 to 15%. The probability that the  $R^2$  exceeds 1% goes from 15% to about 75%. The probability density function also shows lower density than the prior for very low values of the  $R^2$  and again places high density in the region of 5%.

Figure 4 repeats this analysis using the dividend-price ratio to predict quarterly returns. The results show that the posterior clearly favors the existence of a moderate amount of predictability (note that we would expect the  $R^2$  measured at a quarterly horizon to be below

---

<sup>8</sup>This figure shows the unconditional posterior probability that the  $R^2$  exceeds  $k$ ; that is, it does not condition on the existence of predictability.

that for an annual horizon). Panel A shows that the probability that the  $R^2$  exceeds 1% is 25% for the prior but above 80% for the posterior. More generally, the posterior probability that the  $R^2$  exceeds  $k$  is greater for the posterior than for the prior for all  $k < 3\%$ . Panel B shows that the posterior density exhibits a clear spike around 2%.

The above analysis evaluates the statistical evidence on predictability. The Bayesian approach also enables us to study the economic gains from market timing. In particular, we can evaluate the certainty equivalent loss from failing to time the market under different priors on the existence and strength of predictability.

### 3.5 Certainty equivalent returns

We now measure the economic significance of the predictability evidence using certainty equivalent returns. We assume an investor who maximizes

$$E \left[ \frac{W_{T+1}^{1-\gamma}}{1-\gamma} \middle| D \right]$$

for  $\gamma = 5$ , where  $W_{T+1} = W_T(w \exp\{r_{T+1} + r_{f,T}\} + (1-w) \exp\{r_{f,T}\})$ , and  $w$  is the weight on the risky asset. The expectation is taken with respect to the predictive distribution

$$p(r_{T+1} | D) = \bar{q}p(r_{T+1} | D, H_1) + (1 - \bar{q})p(r_{T+1} | D, H_0),$$

where

$$p(r_{T+1} | D, H_i) = \int p(r_{T+1} | x_T, b_i, \Sigma, H_i) p(b_i, \Sigma | D, H_i) db_i d\Sigma$$

for  $i = 0, 1$ .

A draw  $r_{T+1}$  from the distribution  $p(r_{T+1} | x_T, b_1, \Sigma)$  is given by (1) with probability  $\bar{q}$  and (2) with probability  $1 - \bar{q}$ . The posterior distribution of the parameters is described in Section 2.4.

For any portfolio weight  $w$ , we can compute the certainty equivalent return as solving

$$\frac{\exp\{(1-\gamma)\text{CER}\}}{1-\gamma} = E \left[ \frac{(w \exp\{r_{T+1} + r_{f,T}\} + (1-w) \exp\{r_{f,T}\})^{1-\gamma}}{1-\gamma} \middle| D \right]. \quad (31)$$



Following Kandel and Stambaugh (1996), we measure utility loss as the difference between certainty equivalent returns from following the optimal strategy and from following a sub-optimal strategy. We define the sub-optimal strategy as the strategy that the investor would follow if he believes that there is no predictability. Note, however, that the expectation in (31) is computed with respect to the same distribution for both the optimal and sub-optimal strategy.

Table 5 presents the average certainty equivalent loss: we compute the difference in certainty equivalent returns as described above, and then average over the posterior distribution for  $x$ . The data indicate economically meaningful economic losses from failing to time the market. Panel A shows that, for example, an investor with a prior on  $\beta$  such that  $P_{.01} = 0.50$  and a 50% prior belief in the existence of return predictability would suffer a certainty equivalent loss of 0.84% from failing to time the market using the payout yield.<sup>9</sup> Higher values of  $q$  imply greater certainty equivalent losses. Panel B shows somewhat lower certainty equivalent losses for the dividend-price ratio using annual data. However, the certainty equivalent loss is much greater for distributions computed using quarterly postwar data: 1.83% per annum for the investor with  $P_{.01} = 0.50$ , and  $q = 0.50$ , and higher for higher levels of  $q$ .

## 4 Conclusion

This study has taken a Bayesian model selection approach to the question of whether the equity premium varies over time. We considered investors who face uncertainty both over whether predictability exists, and over the strength of predictability if it does exist. We found substantial evidence in favor of predictability when the dividend-price ratio and payout yield were used to predict returns. Moreover, we found large certainty equivalent losses from failing to time the market, even for investors who have strong prior beliefs in a constant equity premium.

---

<sup>9</sup>The low values of the certainty equivalent losses for  $P_{.01} = 0.99$  are a reflection of Bartlett's paradox, as described above.

Finally, we found that taking a fully Bayesian approach that incorporates the exact likelihood function leads to substantially different inference as compared with empirical Bayes or the conditional likelihood function. Empirical Bayes tends to overstate the evidence in favor of predictability while using the conditional likelihood understates the evidence. These results point to the importance of taking into account the stochastic nature of the regressor when studying return predictability from a Bayesian perspective.

# Appendix

## A Jeffreys prior under $H_0$

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  $\mu$ , data  $D$ , and a log-likelihood  $l(\mu; D)$ , Jeffreys shows that invariance is equivalent to specifying a prior as

$$p(\mu) \propto \left| -E \left( \frac{\partial^2 l}{\partial \mu \partial \mu^\top} \right) \right|^{1/2}. \quad (\text{A.1})$$

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)).

Our derivation for the limiting Jeffreys prior on  $b_0, \Sigma$  follows Stambaugh (1999). Zellner (1996, pp. 216-220) derives a limiting Jeffreys prior by applying (A.1) to the likelihood (24) and retaining terms of the highest order in  $T$ . Stambaugh shows that Zellner’s approach is equivalent to applying (A.1) to the conditional likelihood (23), and taking the expectation in (A.1) assuming that  $x_0$  is multivariate normal with mean (6) and variance (7). We adopt this approach.

We derive the prior density for  $p(b_0, \Sigma^{-1})$  and then transform this into the density for  $p(b_0, \Sigma)$  using the Jacobian. Let

$$l_0(b_0, \Sigma; D) = \log p(D|b_0, \Sigma, H_0, x_0). \quad (\text{A.2})$$

denote the natural log of the conditional likelihood. Let  $\zeta = [\sigma^{(11)} \ \sigma^{(12)} \ \sigma^{(22)}]^\top$ , where  $\sigma^{(ij)}$  denotes element  $(i, j)$  of  $\Sigma^{-1}$ . Applying (A.1) implies

$$p(b_0, \Sigma^{-1}|H_0) \propto \left| -E \left[ \begin{array}{cc} \frac{\partial^2 l_0}{\partial b_0 \partial b_0^\top} & \frac{\partial^2 l_0}{\partial b_0 \partial \zeta^\top} \\ \frac{\partial^2 l_0}{\partial \zeta \partial b_0^\top} & \frac{\partial^2 l_0}{\partial \zeta \partial \zeta^\top} \end{array} \right] \right|^{1/2}. \quad (\text{A.3})$$

The the form of the conditional likelihood implies that

$$l_0(b_0, \Sigma; D) = -\frac{T}{2} \log |2\pi\Sigma| - \frac{1}{2} (z - Z_0 b_0)^\top (\Sigma^{-1} \otimes I_T) (z - Z_0 b_0). \quad (\text{A.4})$$

It follows from (A.4) that

$$\frac{\partial l_0}{\partial b_0} = \frac{1}{2} Z_0^\top (\Sigma^{-1} \otimes I_T) (z - Z_0 b_0),$$

and

$$\begin{aligned} \frac{\partial^2 l_0}{\partial b_0 \partial b_0^\top} &= -\frac{1}{2} Z_0^\top (\Sigma^{-1} \otimes I_T) Z_0 \\ &= -\frac{1}{2} \begin{bmatrix} \iota_T^\top & 0 \\ 0 & X^\top \end{bmatrix} (\Sigma^{-1} \otimes I_T) \begin{bmatrix} \iota_T & 0 \\ 0 & X \end{bmatrix} \\ &= -\frac{1}{2} \begin{bmatrix} \sigma^{(11)} T & \sigma^{(12)} \iota^\top X \\ \sigma^{(12)} X^\top \iota & \sigma^{(22)} X^\top X \end{bmatrix}. \end{aligned} \quad (\text{A.5})$$

Taking the expectation conditional on  $b_0$  and  $\Sigma$  implies

$$E \left[ \frac{\partial^2 l_0}{\partial b_0 \partial b_0^\top} \right] = -\frac{T}{2} \begin{bmatrix} \sigma^{(11)} & \sigma^{(12)} [1 \ \mu_x] \\ \sigma^{(12)} \begin{bmatrix} 1 \\ \mu_x \end{bmatrix} & \sigma^{(22)} \begin{bmatrix} 1 & \mu_x \\ \mu_x & \sigma_x^2 + \mu_x^2 \end{bmatrix} \end{bmatrix} \quad (\text{A.6})$$

Using arguments in Stambaugh (1999), it can be shown that

$$E \left[ \frac{\partial^2 l_0}{\partial b_0 \partial \zeta^\top} \right] = 0.$$

Moreover,

$$- \left| E \left( \frac{\partial^2 l_0}{\partial \zeta \partial \zeta^\top} \right) \right| = \left| \frac{\partial^2 \log |\Sigma|}{\partial \zeta \partial \zeta^\top} \right| = |\Sigma|^3$$

(see Box and Tiao (1973, pp. 474-475)). Therefore

$$p(b_0, \Sigma^{-1} | H_0) \propto |\Phi|^{\frac{1}{2}} |\Sigma|^{\frac{3}{2}} \quad (\text{A.7})$$

where

$$\Phi = \begin{bmatrix} \Sigma^{-1} & \mu_x \begin{bmatrix} \sigma^{(12)} \\ \sigma^{(22)} \end{bmatrix} \\ \mu_x [\sigma^{(12)} \ \sigma^{(22)}] & (\sigma_x^2 + \mu_x^2) \sigma^{(22)} \end{bmatrix}.$$

This matrix  $\Phi$  has the same determinant as  $-E \left[ \frac{\partial^2 l_0}{\partial b_0 \partial b_0^\top} \right]$  because 2 columns and 2 rows have been reversed.

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

$$|\Phi| = |\Sigma^{-1}| \left| (\sigma_x^2 + \mu_x^2) \sigma^{(22)} - \mu_x^2 [\sigma^{(12)} \ \sigma^{(22)}] \Sigma \begin{bmatrix} \sigma^{(12)} \\ \sigma^{(22)} \end{bmatrix} \right|.$$

Because

$$\Sigma \begin{bmatrix} \sigma^{(12)} \\ \sigma^{(22)} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

it follows that

$$\begin{aligned} |\Phi| &= |\Sigma^{-1}| \left| (\sigma_x^2 + \mu_x^2) \sigma^{(22)} - \mu_x^2 \sigma^{(22)} \right| \\ &= |\Sigma|^{-1} \sigma_x^2 \sigma^{(22)}. \end{aligned}$$

The determinant of  $\Sigma$  equals

$$|\Sigma| = \sigma_u^2 (\sigma_v^2 - \sigma_{uv}^2 \sigma_u^{-2}),$$

while  $\sigma^{(22)} = (\sigma_v^2 - \sigma_{uv}^2 \sigma_u^{-2})^{-1}$ . Therefore,

$$|\Phi| = |\Sigma|^{-2} \sigma_u^2 \sigma_x^2.$$

Substituting into (A.7),

$$p(b_0, \Sigma^{-1} | H_0) \propto |\Sigma|^{\frac{1}{2}} \sigma_u \sigma_x.$$

The Jacobian of the transformation from  $\Sigma^{-1}$  to  $\Sigma$  is  $|\Sigma|^{-3}$ . Therefore,

$$p(b_0, \Sigma | H_0) = |\Sigma|^{-\frac{5}{2}} \sigma_u \sigma_x.$$

## B Sampling from Posterior Distributions

This section describes how to sample from the posterior distributions. In all cases, the sampling procedure for the posteriors under  $H_1$  and  $H_0$  involve the Metropolis-Hastings algorithm. Below we describe the case of the full Bayes exact likelihood in detail. The procedures for the other cases are similar.

## B.1 Posterior distribution under $H_0$

Substituting (8) and (24) into (25) implies that

$$p(b_0, \Sigma | H_0, D) \propto \sigma_u |\Sigma|^{-\frac{T+5}{2}} \exp \left\{ -\frac{1}{2} \sigma_x^{-2} (x_0 - \mu_x)^2 - \frac{1}{2} (z - Z_0 b_0)^\top (\Sigma^{-1} \otimes I_T) (z - Z_0 b_0) \right\}.$$

This posterior does not take the form of a standard density function because of the term in the likelihood involving  $x_0$  (note that  $\sigma_x^2$  is a nonlinear function of  $\rho$  and  $\sigma_v$ ). However, we can sample from the posterior using the Metropolis-Hastings algorithm.

The Metropolis-Hastings algorithm is implemented “block-at-a-time”, by repeatedly sampling from  $p(\Sigma | b_0, H_0, D)$  and from  $p(b_0 | \Sigma, H_0 D)$  and repeating. To calculate a proposal density for  $\Sigma$ , note that

$$(z - Z_0 b_0)^\top (\Sigma^{-1} \otimes I_T) (z - Z_0 b_0) = \text{tr} [(Y - X B_0)^\top (Y - X B_0) \Sigma^{-1}],$$

where

$$B_0 = \begin{bmatrix} \alpha & \theta \\ 0 & \rho \end{bmatrix}.$$

The proposal density for the conditional probability of  $\Sigma$  is the inverted Wishart with  $T + 2$  degrees of freedom and scale factor of  $(Y - X B_0)^\top (Y - X B_0)$ . The target is therefore

$$p(\Sigma | b_0, H_0, D) \propto \sigma_u \exp \left\{ -\frac{1}{2} \beta^2 (\sigma_\eta^2 \sigma_x^{-2} \sigma_u^2)^{-2} - \frac{1}{2} \sigma_x^{-2} (x_0 - \mu_x)^2 \right\} \times \text{proposal}.$$

Let

$$V_0 = (Z_0^\top (\Sigma^{-1} \otimes I_T) Z_0)^{-1}$$

Let

$$\hat{b}_0 = V_0 Z_0^\top (\Sigma^{-1} \otimes I_T) z$$

It follows from completing the square that

$$(z - Z_0 b_0)^\top (\Sigma^{-1} \otimes I_T) (z - Z_0 b_0) = (b_0 - \hat{b}_0)^\top V_0^{-1} (b_0 - \hat{b}_0) + \text{terms independent of } b_0.$$

The proposal density for  $b_0$  is therefore multivariate normal with mean  $\hat{b}_0$  and variance-covariance matrix  $V_0$ . The accept-reject algorithm of Chib and Greenberg (1995, Section 5) is used to sample from the target density, which is equal to

$$p(b_0|\Sigma, H_0, D) \propto \exp \left\{ -\frac{1}{2} (x_0 - \mu_x)^2 \sigma_x^{-2} \right\} \times \text{proposal}.$$

Note that  $\sigma_u$  and  $\Sigma$  are in the constant of proportionality. Drawing successively from the conditional posteriors for  $\Sigma$  and  $b_0$  produces a density that converges to the full posterior conditional on  $H_0$ .

## B.2 Posterior distribution under $H_1$

Substituting (12) and (22) into (26) implies that

$$p(b_1, \Sigma|H_1, D) \propto \sigma_x |\Sigma|^{-\frac{T+5}{2}} \exp \left\{ -\frac{1}{2} \beta^2 (\sigma_\eta^2 \sigma_x^{-2} \sigma_u^2)^{-2} - \frac{1}{2} \sigma_x^{-2} (x_0 - \mu_x)^2 \right\} \\ \exp \left\{ -\frac{1}{2} (z - Z_1 b_1)^\top (\Sigma^{-1} \otimes I_T) (z - Z_1 b_1) \right\}.$$

The sampling procedure is similar to that described in Appendix B.1. Details can be found in Wachter and Warusawitharana (2009). To summarize, we first draw from the posterior  $p(\Sigma | b_1, H_1, D)$ . The proposal density is an inverted Wishart with  $T + 2$  degrees of freedom and scale factor  $(Y - X B_1)^\top (Y - X B_1)$ , where

$$B_1 = \begin{bmatrix} \alpha & \theta \\ \beta & \rho \end{bmatrix}.$$

We then draw from  $p(\theta, \rho | \alpha, \beta, \Sigma, H_1, D)$ . The proposal density is multivariate normal with mean and variance determined by the conditional normal distribution, as described in Wachter and Warusawitharana. Finally, we draw from  $p(\alpha, \beta | \theta, \rho, \Sigma, H_1, D)$ . In this case, the target and the proposal are the same, and are also multivariate normal.

## C Computing the Bayes factor

Verdinelli and Wasserman (1995) provide an implementable formula for the inverse of the Bayes factor. In our notation, this formula can be written as

$$\mathcal{B}_{10}^{-1} = p(\beta = 0 | H_1, D) E \left[ \frac{p(b_0, \Sigma | H_0)}{p(\beta = 0, b_0, \Sigma | H_1)} \middle| \beta = 0, H_1, D \right]. \quad (\text{C.1})$$

To compute  $p(\beta = 0 | H_1, D)$ , note that

$$p(\beta = 0 | H_1, D) = \int p(\beta = 0 | b_0, \Sigma, H_1, D) p(b_0, \Sigma | H_1, D) db_0 d\Sigma. \quad (\text{C.2})$$

As discussed in Appendix B.2, the posterior distribution of  $\alpha$  and  $\beta$  conditional on the remaining parameters is normal. We can therefore compute  $p(\beta = 0 | b_0, \Sigma, H_1, D)$  (including integration constants) in closed form, by using the properties of the conditional normal distribution. Consider  $N$  draws from the full posterior:  $((b_1^{(1)}, \Sigma^{(1)}), \dots, (b_1^{(N)}, \Sigma^{(N)}))$ , where we can write  $(b_1^{(i)}, \Sigma^{(i)})$  as  $(\beta^{(i)}, b_0^{(i)}, \Sigma^{(i)})$ . We use these draws to integrate out over  $b_0$  and  $\Sigma$ . It follows from (C.2) that

$$p(\beta = 0 | H_1, D) \approx \frac{1}{N} \sum_{i=1}^N p(\beta = 0 | b_0^{(i)}, \Sigma^{(i)}, H_1, D).$$

where the approximation is accurate for large  $N$ .

To compute the second term in (C.1), we observe that

$$\frac{p(b_0, \Sigma | H_0)}{p(\beta = 0, b_0, \Sigma | H_1)} = \frac{p(b_0, \Sigma | H_0)}{p(\beta = 0 | b_0, \Sigma, H_1) p(b_0, \Sigma | H_1)} = \sqrt{2\pi} \sigma_\beta,$$

because  $p(b_0, \Sigma | H_0) = p(b_0, \Sigma | H_1)$ . For the empirical Bayes approach,  $\sigma_\beta$  is a constant and no further simulation is needed. For the full Bayes approach,  $\sigma_\beta = \sigma_\eta \sigma_x^{-1} \sigma_u$ . We require the expectation taken with respect to the posterior distribution conditional on the existence of predictability and the realization  $\beta = 0$ . To calculate this expectation, we draw  $((b_0^{(1)}, \Sigma^{(1)}), \dots, (b_0^{(N)}, \Sigma^{(N)}))$  from  $p(b_0, \Sigma | \beta = 0, H_1, D)$ . This involves modifying the procedure for drawing from the posterior for  $b_1, \Sigma$  given  $H_1$  (see Appendix B.2). We sample from  $p(\Sigma | \alpha, \beta = 0, \theta, \rho, H_1, D)$ , then from  $p(\rho, \theta | \alpha, \beta = 0, \Sigma, H_1, D)$  and finally from  $p(\alpha | \beta = 0, \Sigma, \theta, \rho, H_1, D)$ , and repeat until the desired number of draws are obtained. All



steps except the last are identical to those described in Appendix B.2 (the value of  $\beta$  is identically zero rather than the value from the previous draw). For the last step we derive  $p(\alpha | \beta = 0, \Sigma, \theta, \rho, H_1, D)$  from the joint distribution  $p(\alpha, \beta | \Sigma, \theta, \rho, H_1, D)$ , making use of the properties of the conditional normal distribution.

Given these draws from the posterior distribution, the second term equals

$$E \left[ \frac{p(b_0, \Sigma | H_0)}{p(\beta = 0, b_0, \Sigma | H_1)} \middle| \beta = 0, H_1, D \right] \approx \frac{1}{N} \sum_{i=1}^N \sqrt{2\pi} \sigma_\eta (\sigma_x^{(i)})^{-1} \sigma_u^{(i)}, \quad (\text{C.3})$$

where this approximation is accurate for  $N$  large.

## References

- 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, *Journal of Finance* 56, 45–86.
- Barberis, Nicholas, 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.
- Bartlett, M.S., 1957, Comment on 'A Statistical Paradox' by D. V. Lindley, *Biometrika* 44, 533–534.
- Berger, James O., 1985, *Statistical decision theory and Bayesian analysis*. (Springer New York).
- Boudoukh, Jacob, Roni Michaely, Matthew Richardson, and Michael R. Roberts, 2007, On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* 62, 877–915.
- Box, George E.P., and George C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*. (Addison-Wesley Pub. Co. Reading, MA).
- Brandt, Michael W., Amit Goyal, Pedro Santa-Clara, and Jonathan R. Stroud, 2005, A simulation approach to dynamics portfolio choice with an application to learning about return predictability, *Review of Financial Studies* 18, 831–873.
- Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1997, Strategic asset allocation, *Journal of Economic Dynamics and Control* 21, 1377–1403.
- Campbell, John Y., and Luis M. Viceira, 1999, Consumption and portfolio decisions when expected returns are time-varying, *Quarterly Journal of Economics* 114, 433–495.

- Campbell, John Y., and Motohiro Yogo, 2006, Efficient tests of stock return predictability, *Journal of Financial Economics* 81, 27–60.
- Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.
- Chen, Hui, Nengjiu Ju, and Jianjun Miao, 2009, Dynamic asset allocation with ambiguous return predictability, Working paper, MIT.
- Chen, Zengjing, and Larry Epstein, 2002, Ambiguity, risk and asset returns in continuous time, *Econometrica* 70, 1403–1443.
- Chib, Siddhartha, and Edward Greenberg, 1995, Understanding the Metropolis-Hastings algorithm, *American Statistician* 49, 327–335.
- Chipman, Hugh, Edward I. George, and Robert E. McCulloch, 2001, The practical implementation of Bayesian model selection, in P. Lahiri, eds.: *Model Selection* (IMS Lecture Notes, Bethesda, MA ).
- Cremers, K.J. Martjin, 2002, Stock return predictability: A Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.
- Dickey, James M., 1971, The weighted likelihood ratio, linear hypotheses on normal location parameters, *The Annals of Mathematical Statistics* 42, 204–223.
- Fernandez, Carmen, Eduardo Ley, and Mark F. J. Steel, 2001, Benchmark priors for Bayesian model averaging, *Journal of Econometrics* 100, 381–427.
- Garlappi, Lorenzo, Raman Uppal, and Tan Wang, 2007, Portfolio selection with parameter and model uncertainty: A multi-prior approach, *Review of Financial Studies* 20, 41–81.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin, 2004, *Bayesian Data Analysis*. (Chapman & Hall/CRC Boca Raton, FL).

- Hamilton, J. D., 1994, *Time Series Analysis*. (Oxford University Press Princeton, NJ).
- Hansen, Lars Peter, 2007, Beliefs, doubts and learning: Valuing economic risk, NBER working paper #12948.
- Jeffreys, Harold, 1961, *Theory of Probability*. (Oxford University Press Clarendon).
- Johannes, Michael, and Nicholas Polson, 2006, MCMC methods for financial econometrics, in Yacine Ait-Sahalia, and Lars Hansen, eds.: *Handbook of Financial Econometrics* (Elsevier, North-Holland ).
- 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.
- Kass, R., and A. E. Raftery, 1995, Bayes factors, *Journal of the American Statistical Association* 90, 773–795.
- Lewellen, Jonathan, 2004, Predicting returns with financial ratios, *Journal of Financial Economics* 74, 209–235.
- Maenhout, Pascal, 2006, Robust portfolio rules and detection-error probabilities for a mean-reverting risk premium, *Journal of Economic Theory* 128, 136–163.
- Nelson, C. R., and M. J. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.
- Pastor, Lubos, and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance* 54, 67–121.
- Pastor, Lubos, and Robert F. Stambaugh, 2008, Predictive systems: Living with imperfect predictors, forthcoming, *Journal of Finance*.

- Pesaran, M Hashem, and Allan Timmermann, 1995, Predictability of stock returns: Robustness and economic significance, *Journal of Finance* 50, 1201–1228.
- Phillips, Peter C.B., 1991, To criticize the critics: An objective Bayesian analysis of stochastic trends, *Journal of Applied Econometrics* 6, 333–364.
- Robbins, Herbert, 1964, The empirical Bayes approach to statistical decision problems, *The Annals of Mathematical Statistics* 35, 1–20.
- Skoulakis, Georgios, 2007, Dynamic portfolio choice with Bayesian learning, Working paper, University of Maryland.
- Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.
- Stock, James H., and Mark W. Watson, 2005, An empirical comparison of methods for forecasting using many predictors, Working paper, Harvard University and Princeton University.
- Torous, Walter, Rossen Valkanov, and Shu Yan, 2004, On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* 77, 937–966.
- Verdinelli, Isabella, and Larry Wasserman, 1995, Computing Bayes factors using a generalization of the Savage-Dickey density ratio, *Journal of the American Statistical Association* 90, 614–618.
- Wachter, Jessica A., and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, forthcoming, *Journal of Econometrics*.
- Wright, Jonathan H., 2003, Bayesian model averaging of exchange rate forecasts, International Finance Discussion Papers 779, Board of Governors of the Federal Reserve.

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

Table 1: Bayes factors and posterior means: Payout yield and annual returns

Model	$P_{.01}$	$\mathcal{B}_{10}$	$\bar{\beta}$	$\bar{\rho}$	$\bar{r}$	$\bar{x}$
Full Bayes	0.05	1.68	2.23	0.936	5.24	-3.17
Exact Lkl.	0.50	11.99	12.94	0.889	5.14	-3.16
	0.99	18.20	19.54	0.878	5.05	-3.16
Full Bayes	0.05	1.36	1.39	0.959	5.64	-5.32
Conditional Lkl.	0.50	5.51	10.71	0.910	4.87	-3.76
	0.99	6.54	16.42	0.914	-22.66	-6.24
Empirical Bayes	0.05	2.58	3.99	0.926	5.22	-3.17
Exact Lkl.	0.50	19.43	14.17	0.887	5.13	-3.16
	0.99	27.13	21.90	0.851	5.09	-3.16
OLS			20.89	0.863	5.85	-3.15

Notes:  $P_{.01}$  denotes the prior probability that the  $R^2$  from the predictive regression exceeds .01 conditional on the existence of predictability (this is applicable for full Bayes priors; empirical Bayes priors are constructed to be comparable to full Bayes counterparts).  $\mathcal{B}_{10} = p(D|H_1)/p(D/H_0)$  denotes the Bayes factor in favor of predictability ( $H_1$ ) versus no predictability ( $H_0$ ). The table also reports posterior means of the predictive coefficient  $\beta$ , the autoregressive coefficient  $\rho$ , the excess return  $r$  and the predictor variable  $x$  conditional on  $H_1$ . The predictor variable is the payout yield (the dividend-price ratio adjusted for repurchases) constructed from the value-weighted CRSP index. Continuously compounded stock returns on the value weighted CRSP index are in excess of the continuously-compounded return on the three-month Treasury Bill. Data are annual from 1/1/1927 to 1/1/2004. OLS denotes results obtained from ordinary least squares regression.

Table 2: Bayes factors and posterior means: Dividend-price ratio and annual returns

Model	$P_{.01}$	$\mathcal{B}_{10}$	$\bar{\beta}$	$\bar{\rho}$	$\bar{r}$	$\bar{x}$
Full Bayes	0.05	1.51	1.48	0.966	4.71	-3.37
Exact Lkl.	0.50	5.73	7.64	0.946	4.37	-3.35
	0.99	6.90	11.30	0.948	4.02	-3.35
Full Bayes	0.05	1.21	0.83	0.980	5.31	-10.24
Conditional Lkl.	0.50	2.78	5.56	0.963	3.15	-6.75
	0.99	3.53	8.90	0.976	-83.53	-16.17
Empirical Bayes	0.05	2.23	2.65	0.960	4.64	-3.36
Exact Lkl.	0.50	9.17	8.85	0.942	4.31	-3.34
	0.99	9.00	13.28	0.925	4.17	-3.33
OLS			11.64	0.944	5.85	-3.27

Notes:  $P_{.01}$  denotes the prior probability that the  $R^2$  from the predictive regression exceeds .01 conditional on the existence of predictability (this is applicable for full Bayes priors; empirical Bayes priors are constructed to be comparable to full Bayes counterparts).  $\mathcal{B}_{10} = p(D|H_1)/p(D/H_0)$  denotes the Bayes factor in favor of predictability ( $H_1$ ) versus no predictability ( $H_0$ ). The table also reports posterior means of the predictive coefficient  $\beta$ , the autoregressive coefficient  $\rho$ , the excess return  $r$  and the predictor variable  $x$  conditional on  $H_1$ . The predictor variable is the dividend-price ratio constructed from the value-weighted CRSP index. Continuously compounded stock returns on the value weighted CRSP index are in excess of the continuously-compounded return on the three-month Treasury Bill. Data are annual from 1/1/1927 to 1/1/2004. OLS denotes results obtained from ordinary least squares regression.



Table 3: Bayes factors and posterior means: Dividend-price ratio and quarterly post-war returns

Model	$P_{.01}$	$\mathcal{B}_{10}$	$\bar{\beta}$	$\bar{\rho}$	$\bar{r}$	$\bar{x}$
Full Bayes	0.05	4.68	1.05	0.990	3.20	-3.49
Exact Lkl.	0.50	7.06	1.87	0.984	3.21	-3.50
	0.99	6.48	2.01	0.983	3.21	-3.50
Full Bayes	0.05	2.14	0.69	0.994	2.68	-8.13
Conditional Lkl.	0.50	2.90	1.51	0.988	0.53	-6.87
	0.99	2.59	1.59	0.988	-4.74	-8.66
Empirical Bayes	0.05	10.57	1.44	0.988	3.20	-3.50
Exact Lkl.	0.50	11.72	2.43	0.979	3.20	-3.50
	0.99	9.34	2.77	0.976	3.20	-3.50
OLS			2.74	0.976	5.22	-3.51

Notes:  $P_{.01}$  denotes the prior probability that the  $R^2$  from the predictive regression exceeds .01 conditional on the existence of predictability (this is applicable for full Bayes priors; empirical Bayes priors are constructed to be comparable to full Bayes counterparts).  $\mathcal{B}_{10} = p(D|H_1)/p(D/H_0)$  denotes the Bayes factor in favor of predictability ( $H_1$ ) versus no predictability ( $H_0$ ). The table also reports posterior means of the predictive coefficient  $\beta$ , the autoregressive coefficient  $\rho$ , the excess return  $r$  and the predictor variable  $x$  conditional on  $H_1$ . The posterior mean of  $r$  is annualized by multiplying by 4. The predictor variable is the dividend-price ratio constructed from the value-weighted CRSP index. Continuously compounded stock returns on the value weighted CRSP index are in excess of the continuously-compounded return on the three-month Treasury Bill. Data are quarterly from 4/1/1952 to 1/1/2005. OLS denotes results obtained from ordinary least squares regression.

Table 4: Posterior probability of predictable excess stock returns for the full Bayes exact likelihood.

Predictor	$P(R^2 > 0.01   H_1)$	Prior prob. of return predictability $q$			
		0.01	0.20	0.50	0.80
Payout Yield	0.05	0.02	0.30	0.63	0.87
Annual	0.50	0.11	0.75	0.92	0.98
	0.99	0.16	0.82	0.95	0.99
Dividend-Price Ratio	0.05	0.02	0.27	0.60	0.86
Annual	0.50	0.05	0.59	0.85	0.96
	0.99	0.07	0.63	0.87	0.97
Dividend-Price Ratio	0.05	0.05	0.54	0.82	0.95
Quarterly	0.50	0.07	0.64	0.88	0.97
	0.99	0.06	0.62	0.87	0.96

Notes: The table reports  $\bar{q}$ , the probability the investor assigns to predictable excess stock returns after seeing the data. Rows vary  $P(R^2 > .01|H_1)$ , the prior probability that the  $R^2$  from the predictability regression exceeds 0.01, conditional on the existence of predictability. Columns vary  $q$ , the prior probability of predictable excess stock returns. The predictor variables include the payout yield and the dividend-price ratio, both constructed from the value-weighted CRSP index. Continuously compounded stock returns on the value-weighted CRSP index are in excess of the continuously-compounded return on the three-month Treasury Bill. The first two panels report results using annual data from 1/1/1927 to 1/1/2004. The last panel reports results using quarterly data from 4/1/1952 to 1/1/2005.

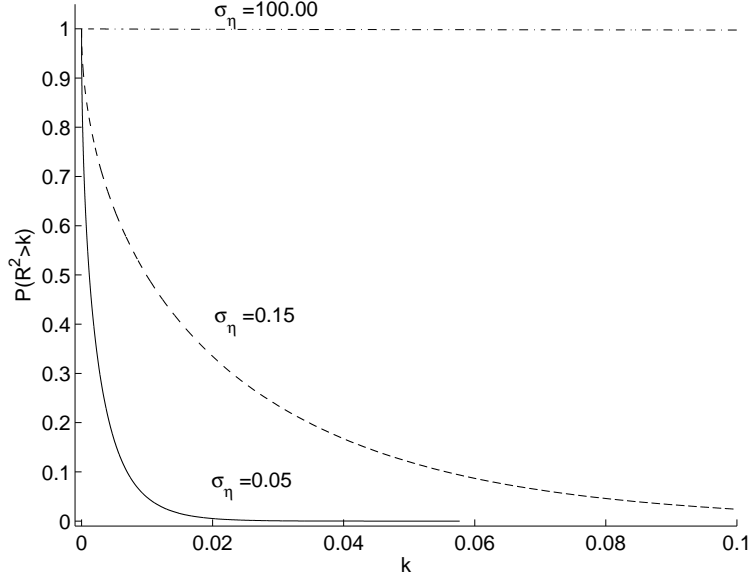
Table 5: Average certainty equivalent returns from timing the market.

Predictor	$P(R^2 > 0.01   H_1)$	Prior prob. of return predictability $q$			
		0.20	0.50	0.80	0.99
Payout Yield	0.05	0.01	0.03	0.05	0.07
Annual	0.50	0.57	0.82	0.92	0.95
	0.99	1.15	1.50	1.61	1.65
Dividend-Price Ratio	0.05	0.01	0.03	0.06	0.08
Annual	0.50	0.37	0.69	0.84	0.90
	0.99	0.97	1.60	1.87	1.98
Dividend-Price Ratio	0.05	0.42	0.86	1.07	1.16
Quarterly	0.50	1.14	1.83	2.11	2.21
	0.99	1.19	1.97	2.30	2.42

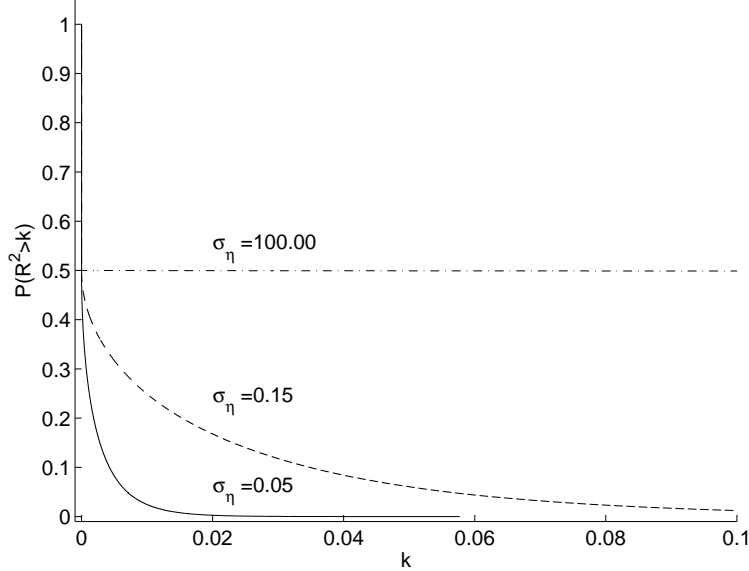
Notes: The table reports the certainty equivalent return to timing the market. Rows vary  $P(R^2 > .01|H_1)$ , the prior probability that the  $R^2$  from the predictability regression exceeds 0.01, conditional on the existence of predictability. Columns vary  $q$ , the prior probability of predictable excess stock returns. The predictor variables include the payout yield and the dividend-price ratio, both constructed from the value-weighted CRSP index. The posterior is constructed using full Bayes priors with the exact likelihood. Continuously compounded stock returns on the value-weighted CRSP index are in excess of the continuously-compounded return on the three-month Treasury Bill. The first two panels report results using annual data from 1/1/1927 to 1/1/2004. The last panel reports results using quarterly data from 4/1/1952 to 1/1/2005. In this panel, returns are annualized by multiplying by 4. The certainty equivalent returns are constructed by averaging over the CER values for 1000 draws of the predictor variable from its unconditional posterior distribution.

Figure 1: Prior Distribution of the  $R^2$

Panel A: Probability of predictability  $q = 1$ .

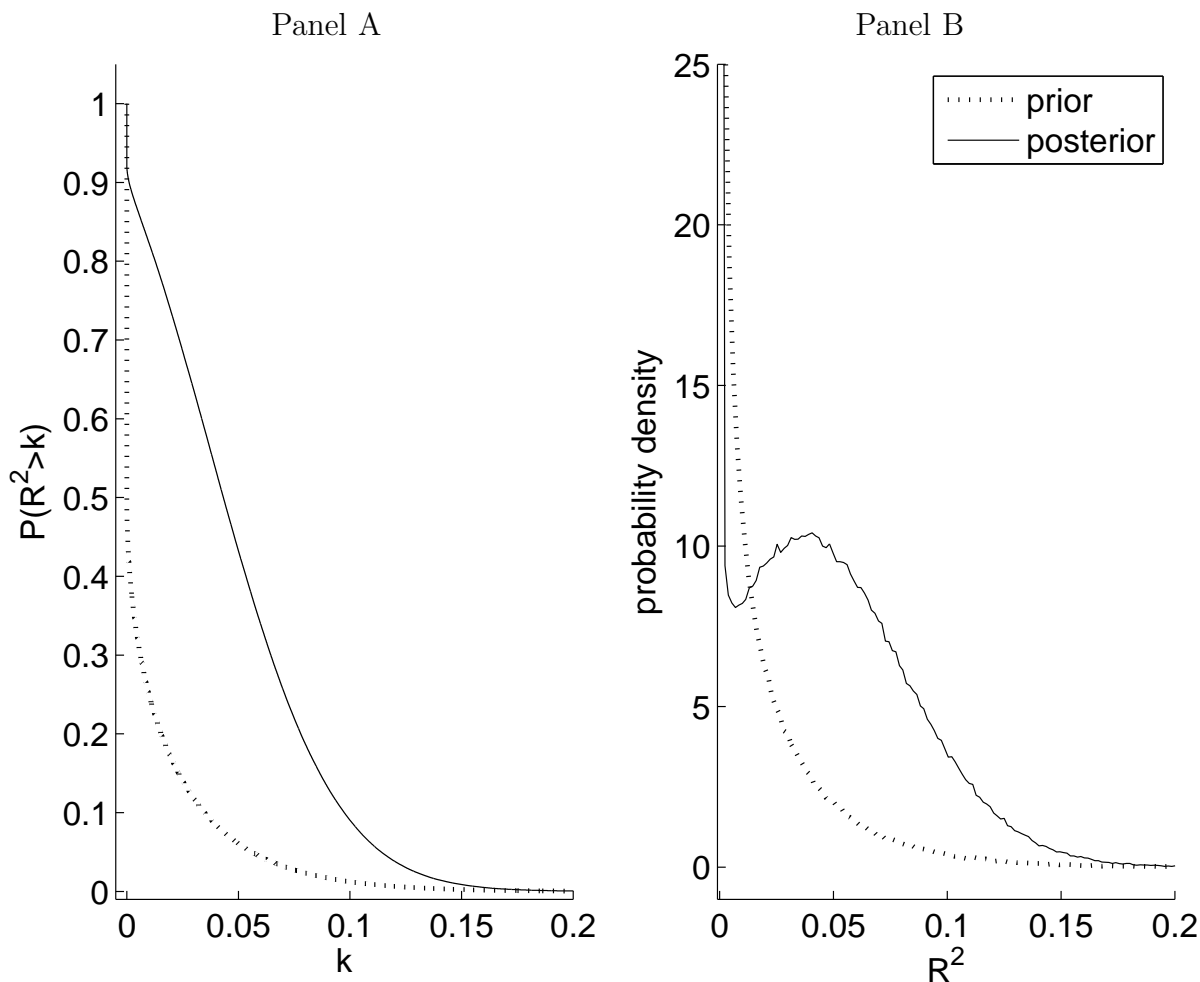


Panel B: Probability of predictability  $q = 0.5$ .



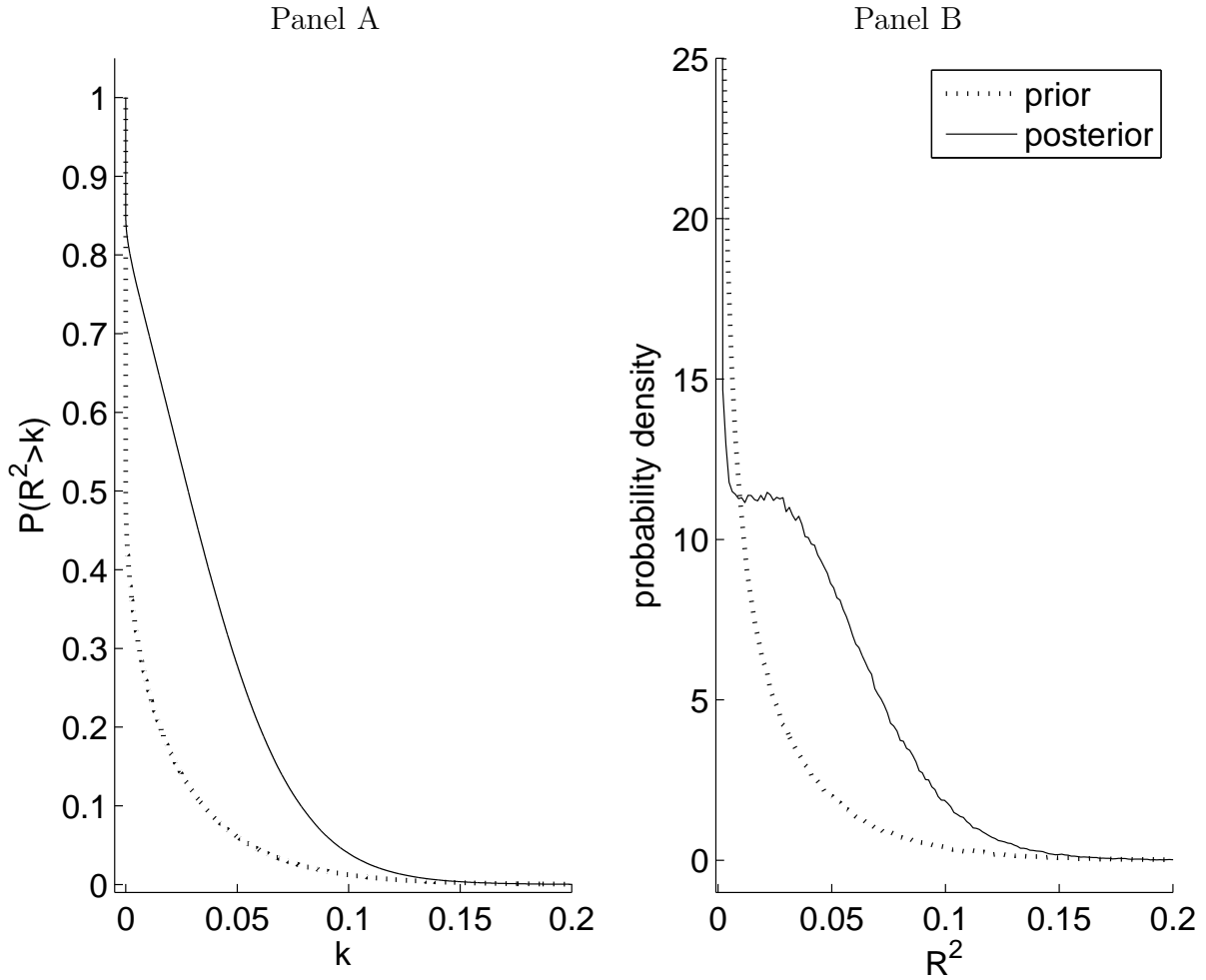
Notes: The figures plot the prior probability that the  $R^2$  will be greater than some value  $k$  for different values of  $k$ . This equals 1 minus the cumulative density function for the distribution on the  $R^2$ . Panel A reports the values conditional on predictability ( $q = 1$ ) and panel B plots the values for a prior value of  $q = 0.5$ .  $\sigma_\eta$  parameterizes the prior variance of  $\beta$  with  $\sigma_\beta = \sigma_\eta \sigma_x^{-1} \sigma_u$ .

Figure 2: Posterior Distribution of the  $R^2$ : Payout Yield and Annual Returns



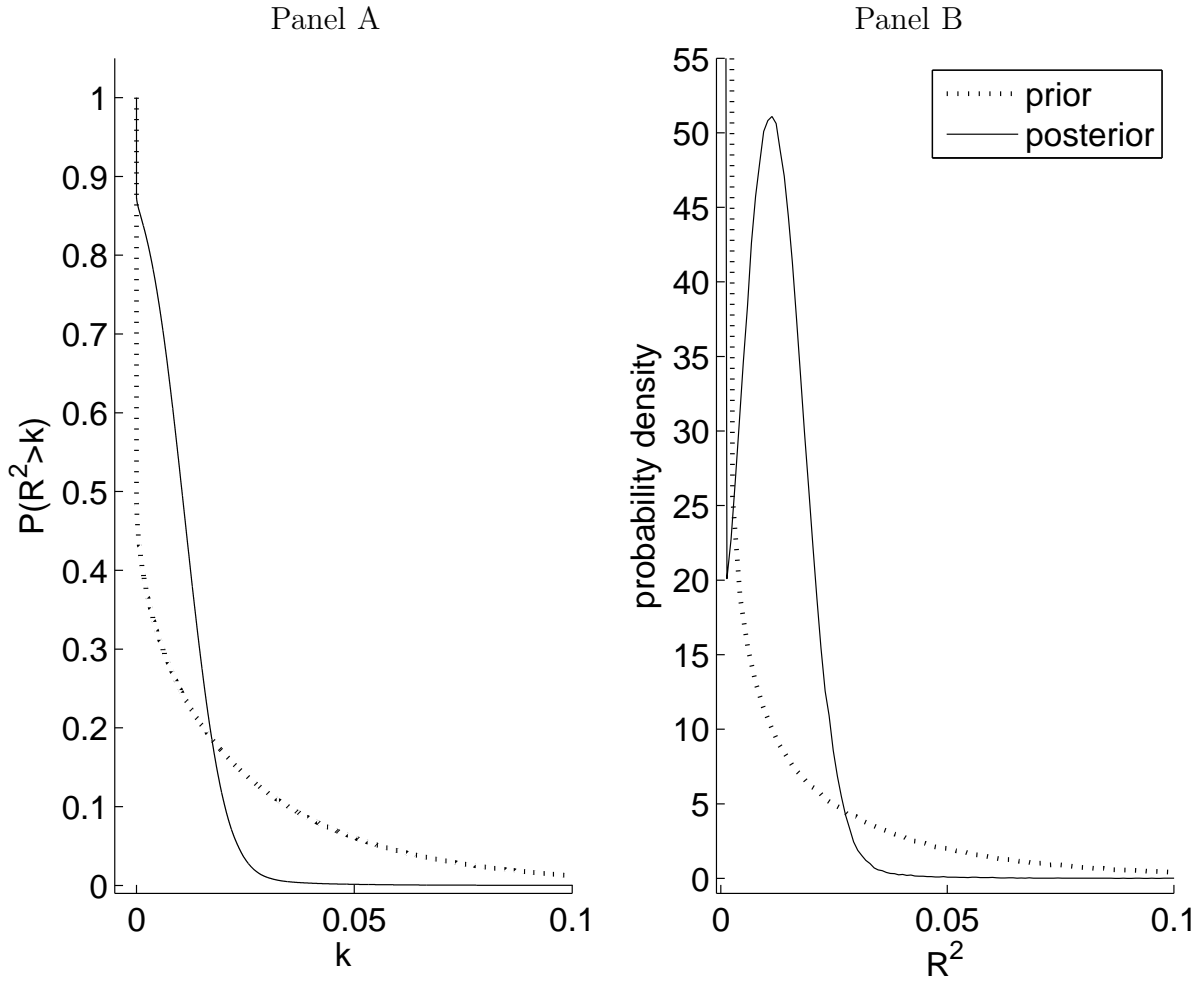
Notes: Panel A plots the probability that the  $R^2$  from a predictive regression of excess stock returns on the payout yield will be greater than some value  $k$  for different values of  $k$ . This equals 1 minus the cumulative density function for the distribution on the  $R^2$ . Panel B plots the probability density function of the  $R^2$  for the same regression. The dashed line signifies the prior and the solid line signifies the posterior distribution for the  $R^2$ . The likelihood function for these plots is the full Bayes exact likelihood with  $P(R^2 > 0.01|H_1) = 0.50$  and  $q = 0.5$ . Data are annual from 1/1/1927 to 1/1/2004.

Figure 3: Posterior Distribution of the  $R^2$ : Dividend-Price Ratio and Annual Returns



Notes: Panel A plots the probability that the  $R^2$  from a predictive regression of excess stock returns on the dividend-price ratio will be greater than some value  $k$  for different values of  $k$ . This equals 1 minus the cumulative density function for the distribution on the  $R^2$ . Panel B plots the probability density function of the  $R^2$  for the same regression. The dashed line signifies the prior and the solid line signifies the posterior distribution for the  $R^2$ . The likelihood function for these plots is the full Bayes exact likelihood with  $P(R^2 > 0.01|H_1) = 0.50$  and  $q = 0.5$ . Data are annual from 1/1/1927 to 1/1/2004.

Figure 4: Posterior Distribution of the  $R^2$ : Dividend-Price Ratio and Quarterly Returns



Notes: Panel A plots the probability that the  $R^2$  from a predictive regression of excess stock returns on the dividend-price ratio will be greater than some value  $k$  for different values of  $k$ . This equals 1 minus the cumulative density function for the distribution on the  $R^2$ . Panel B plots the probability density function of the  $R^2$  for the same regression. The dashed line signifies the prior and the solid line signifies the posterior distribution for the  $R^2$ . The likelihood function for these plots is the full Bayes exact likelihood with  $P(R^2 > 0.01|H_1) = 0.50$  and  $q = 0.5$ . Data are quarterly from 4/1/1952 to 1/1/2005.