The Rodney L. White Center for Financial Research

Stock-Return Predictability and Model Uncertainty

Doron Avramov

12-00

The Wharton School
University of Pennsylvania

The Rodney L. White Center for Financial Research
The Wharton School
University of Pennsylvania
3254 Steinberg Hall-Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104-6367

(215) 898-7616
(215) 573-8084 Fax
http://finance.wharton.upenn.edu/~rlwctr

The Rodney L. White Center for Financial Research is one of the oldest financial research centers in the country. It was founded in 1969 through a grant from Oppenheimer & Company in honor of its late partner, Rodney L. White. The Center receives support from its endowment and from annual contributions from its Members.

The Center sponsors a wide range of financial research. It publishes a working paper series and a reprint series. It holds an annual seminar, which for the last several years has focused on household financial decision making.

The Members of the Center gain the opportunity to participate in innovative research to break new ground in the field of finance. Through their membership, they also gain access to the Wharton School's faculty and enjoy other special benefits.

Stock-Return Predictability and Model Uncertainty

By

Doron Avramov*

The Wharton School

University of Pennsylvania

First draft: April 1999

This revision: May 2000

*Ph.D candidate, Finance department, email: doron51@wharton.upenn.edu, phone: (215) 898-7542, fax: (215) 898-6200. I am indebted to Rob Stambaugh for his guidance and encouragement. I also thank Isabelle Bajeux-Besnainou, Michael Brandt, Jeremy Gold, Anthony Lynch, Craig MacKinlay, Andrew Metrick, Lubos Pastor, Krishna Ramaswamy, Frank Schorfheide, Amir Yaron, and seminar participants at Boston College, Cornell, Goldman Sachs Asset Management, London Business School, University of Maryland, University of Pennsylvania, Vanderbilt, and Washington Area Finance Association for their useful comments. Of course, I am fully responsible for all the weaknesses of this paper.

Stock-Return Predictability and Model Uncertainty

Abstract

We investigate the implications of uncertainty about the return-forecasting model for the investment opportunity set. Asset allocations are computed through various approaches that differ in their treatment of model uncertainty. The optimal portfolio choices can differ to economically significant degrees, especially for short-horizon high risk-tolerance investors. We decompose the variance of predicted stock returns into several components, including model uncertainty and parameter uncertainty. The model-uncertainty component can be significantly higher than the parameter-uncertainty component, especially when predictive variables, such as dividend yield and book-to-market, are at their recently observed levels, and there is substantial prior uncertainty about whether returns are predictable.

Introduction

Portfolio decisions in the presence of forecasting variables and parameter uncertainty have been analyzed in several recent studies. For example, Kandel and Stambaugh (1996) show that even when the statistical evidence of predictability seems rather weak, the optimal equity-versus-cash allocation for a single-period investor can depend strongly on the current value of a predictive variable, such as dividend yield. Barberis (1999) analyzes the portfolio choice of a long-horizon investor and finds that as the length of horizon increases, strong predictability leads to a higher investment in equities. Both studies incorporate "estimation risk," or the investor's uncertainty about the parameters of a given forecasting model. However, the studies do not incorporate the investor's uncertainty about the appropriate forecasting model, and thus, their results are conditioned upon a given model.

A common practice in applied work for dealing with a set of alternative models is to select a single model from the set and then proceed to use that model for the economic decision at hand. For example, Bossaerts and Hillion (1999) and Pesaran and Timmermann (1995) use various model-selection criteria to select a set of variables that predict stock returns. Selecting a single model does not account for model uncertainty, even though the use of a model-selection criterion implicitly recognizes the presence of such uncertainty. The selected model is essentially viewed as having a probability of unity being the "correct" one, thereby overstating the precision of the forecast.[1] One might suspect that uncertainty about the correct model for predicting asset returns could play a role in portfolio decisions, since expected future returns are likely to differ across distinct forecasting models. These differences result in additional uncertainty about future stock returns and, therefore, make

[1] Leamer (1978) p. 91 argues: "Ambiguity about model selection should dilute information about quantities of interest as part of the evidence is spent to specify the model."

equities less attractive to risk-averse investors.

This study incorporates the additional dimension of model uncertainty into an empirical investigation of stock-return predictability. The Bayesian approach used here does not identify a single "best" model. Instead, for any portfolio whose returns are to be predicted, we consider all possible combinations of a set of M forecasting variables that are believed to govern the evolution of expected stock returns. Each of the 2^M combinations constitutes a unique linear data-generating process having a posterior model probability. Model uncertainty is incorporated via Bayesian model averaging, which uses the posterior probabilities as weights on the individual data-generating models to obtain one overall weighted predictive distribution. The latter is used to explore the potential impact of model uncertainty on the investment opportunity set.

Departing from the traditional focus on a single investable risky asset, we consider buy-and-hold, long-horizon investors who incorporate estimation risk (parameter uncertainty) in allocating funds among size-sorted portfolios and the risk-free Treasury bill.[2] We decompose the variance of predicted cumulative returns on these portfolios over the investment horizon into three components: the cross-model uncertainty; the within-model parameter uncertainty; and the uncertainty about future returns computed as though the forecasting model and the model-specific parameters were known.

We show that if investors display substantial prior uncertainty about whether stock returns are predictable and consider the events of predictability versus no predictability as equally likely *ex ante*, then the model-uncertainty component can be significantly higher

[2]To the best of our knowledge, we are the first to analyze portfolio decisions across multiple risky assets in the presence of both estimation risk and predictive variables.

than the parameter-uncertainty component, and the sum of these two components can account for a significant portion of the overall predictive variance. This finding is especially prominent when predictive variables that are perceived to have been indicators of fundamental values – such as book-to-market, dividend yield, and earnings yield – are at their recently observed level. To elaborate, in recent years, equity markets have been overwhelmingly bullish. As a result, the current values of book-to-market, dividend yield, among other variables, which are inversely related to a stock-price index level, have been substantially distant from their sample means, thereby giving rise to the model-uncertainty component.

Based on the weighted model, which incorporates model uncertainty, the overall investment in equities at the end of 1998 rises with the investment horizon. This effect occurs as a result of both an increase in conditional expected returns towards long-run means and a reduction in the standard deviation due to predictability in stock returns. The increase in expected returns in conjuction with the decrease in the predictive variance makes stocks look more attractive to a long-horizon, risk-averse investor who, therefore, allocates more to equities. We show that cross-sectional differences in predictability across size-sorted portfolios produce different patterns in optimal asset allocations. For example, with a lengthening horizon the investment in large versus either small or medium-size stocks becomes more attractive.

There is not a unique way to ignore model uncertainty, and asset allocations are also computed under specifications that differ with respect to how they ignore that uncertainty. Investors might, among such possibilities, select one model based on a formal selection criterion or simply retain all entertained predictive variables in an "all-inclusive" model. We compare the asset allocation obtained using the weighted model to those computed based on the highest-posterior-probability model, the all-inclusive model, and the model that drops all entertained predictors - the *iid* model of stock returns. We find that optimal portfolio allocations can differ across models to economically significant degrees, depending on investors' risk tolerance and the investment horizon. In particular, the differences in optimal portfolios are especially apparent for short-horizon high risk-tolerance investors.

The Bayesian approach to model uncertainty has not been widely adopted in applied work. The two main reasons are the difficulty in computing the marginal likelihood (and hence the posterior probability) and sensitivity to priors. Not only is computing the marginal likelihood often analytically intractable, but it also requires the elicitation of informative prior belief about all of the model parameters. Otherwise, the posterior model probabilities are not determined uniquely, but instead depend on arbitrary normalizations. Because informative priors often require subjective judgments, they need not reflect a consensus view.[3] In this study, the prior is determined using two sources of information: investors' beliefs about the correct data-generating model, and information conveyed by a "training" sample. The latter is merely an initial sample observed prior to the primary counterpart. The marginal likelihood is then derived using the Gibbs sampler technique advocated by Chib (1995). To account for prior sensitivity, we evaluate the marginal likelihood over a range of possible priors.

The remainder of the paper proceeds as follows. In Section 1, we establish the framework

[3]Kass and Raftery (1995) discuss several cases in which model probabilities can be computed with improper priors. Sensitivity of model probabilities to priors is not confined to small size samples. As the sample grows, the prior exerts less influence on the posterior of parameters within a model (and hence on model estimation). In contrast, the influence of the prior need not vanish in computing the marginal likelihood and, therefore, not in hypothesis testing and model selection.

for a long-horizon optimal portfolio choice in the presence of model uncertainty, multiple investable assets, and multiple predictive variables. We also provide the preliminaries of the Bayesian model averaging method in the context of portfolio decisions. Section 2 describes the sample data that contains histories of returns on size-sorted portfolios and realizations of several variables that are suspected to have been relevant in forecasting stock returns. Section 3 contains results. Conclusions and ideas for future research are provided in Section 4. Technical issues are discussed in the appendix.

Literature related to this work follows several distinct veins, including the investigation of parameter uncertainty and its implications for optimal portfolio choice in the context of both *iid* returns[4] and time-varying expected returns.[5] Studies that explore asset-return predictability include Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988a, 1988b), Fama and French (1988, 1989), Chen (1991), Kothari and Shanken (1997), and Lo and MacKinlay (1997), among others. Lastly, a comprehensive treatment of model uncertainty and variable selection can be found, for example, in Leamer (1978), Kass and Raftery (1995), and Raftery, Madigan, and Hoeting (1997).

I Long-Horizon Portfolio Choice and Model Uncertainty

This section develops a framework for analyzing long-horizon portfolio choices in the presence of model uncertainty, parameter uncertainty within any forecasting model, multiple investable assets, and multiple predictive variables. It constitutes the methodological part

[4]See, for example, Bawa, Brown, and Klein (1979), Jobson and Korkie (1980), Jorion (1985), and Frost and Savarino (1986).

[5]See, for example, Kandel and Stambaugh (1996) and Barberis (1999). Studies on long-horizon portfolio choice in the presence of time-varying expected returns that do not account for parameter uncertainty include Brennan, Schwartz, and Lagnado (1997), Brandt (1999), and Campbell and Viceira (1999).

of the paper and is divided into four subsections. Subsection A describes the implications of uncertainty about the return forecasting model for the probability distribution function of future stock returns. In particular, it introduces the concept of model averaging as the solution to portfolio decision in the presence of such uncertainty. In the Bayesian approach used here, the *a posteriori* dynamics of stock returns and predictive variables are perceived to have been governed by a combination of a prior belief about the unknown parameters characterizing the dynamics, and the likelihood function of the data conditioning upon these parameters. Subsection B specifies the assumptions on these dynamics and the resulting likelihood function. In Subsection C, the variance of future cumulative stock returns is decomposed into three components, including the cross-model uncertainty and the within-model parameter uncertainty. The contribution of each component to the overall perceived variance of future returns is analyzed in Section 3. Subsection D discusses the elicitation of prior belief about the unknown parameters, the resulting posterior density, and the computation of posterior probabilities for all models under consideration.

It should be emphasized at the outset that previous studies that employ the Bayesian model averaging technique have treated the explanatory variables in the regression model as fixed non-stochastic.[6] A built-in advantage in doing so is the ability to combine various prior specifications, such as the natural conjugate priors, with the conditional likelihood function to produce an analytical posterior density. The posterior probability of a model can thereby be obtained analytically. However, in that approach, the posterior probability is computed only for a subset of the overall data points - the dependent variables in the regression model. In time-series regressions, although the explanatory variables are pre-assigned, they evolve stochastically, and as such, they should be treated as part of the data. Indeed, to the best of

[6]See, for example, Raftery, Madigan, and Hoeting (1997) and the reference therein.

our knowledge, we are the first to compute model posterior probabilities for all the sample data containing stock returns and realizations of several variables that are believed to be relevant in forecasting stock returns.

A The Decision-Maker's Optimization Problem

Let us consider a long-horizon risk-averse investor who allocates funds across a risk-free asset and N risky assets. The portfolio decision is made based upon the sample data Φ containing a finite history of returns on the risky assets and realizations of several predictive variables that are believed to have been governing the evolution of asset returns. It is assumed that the investor's portfolio decision has no effect on the probability distribution of asset returns. Lastly, capital markets are assumed to be frictionless.

For the above-described investor, the quantity of interest is the cumulative excess return on the chosen portfolio over the investment horizon. The vector of cumulative excess log returns on N investable assets is given by $R_{T+K} = \sum_{k=1}^{K} r_{T+k}$, where r_{T+k} is an $N \times 1$ vector of excess log returns on the N assets at time $T+k$, T denotes the beginning of the investment horizon or the end of the sample period, and K is the length of the horizon. The investor considers an initial set of J distinct models to construct the predictive distribution of stock returns. The model space is denoted by $\mathcal{M}$, where $\mathcal{M} = \{\mathcal{M}_1, \ldots, \mathcal{M}_J\}$.

The investor's tradeoff between risk and return is reflected through an iso-elastic utility function for wealth at the end of the investment horizon:

$$U(W_{T+K}) = \begin{cases} \frac{1}{1-\gamma} W_{T+K}^{1-\gamma} & \text{for } \gamma > 0 \text{ and } \gamma \neq 1, \\ \ln(W_{T+K}) & \text{for } \gamma = 1, \end{cases} \tag{1}$$

where γ is the decision-maker's relative risk-aversion coefficient, and W_{T+K} denotes the terminal wealth. The latter is computed as:

$$W_{T+K} = W_T \left[(1 - \omega' \iota_N) \exp(r_f K) + \omega' \exp(r_f K \iota_N + R_{T+K}) \right], \tag{2}$$

where W_T is the wealth at the end of the sample period, ω is an $N \times 1$ vector denoting portfolio weights chosen for N risky assets at time T, ι_N is an $N \times 1$ vector of ones, and r_f is the continuously compounded risk-free rate of return.[7] From the properties of the iso-elastic utility function, the fraction of wealth invested in equities is independent of W_T, which is, therefore, normalized to unity.

The predictive distribution of cumulative stock returns that depends solely upon the sample data averages over all the conditional predictive distributions of individual models belonging to $\mathcal{M}$ and is computed as:[8]

$$P(R_{T+K}|\Phi) = \sum_{j=1}^{J} P(R_{T+K}|\mathcal{M}_j, \Phi) \, P(\mathcal{M}_j|\Phi). \tag{3}$$

The first factor on the right-hand side of equation (3) is the predictive distribution for long-horizon returns pertaining to model $\mathcal{M}_j$ and is given by:

$$P(R_{T+K}|\mathcal{M}_j, \Phi) = \int_{\Theta_j} P(R_{T+K}|\Theta_j, \mathcal{M}_j, \Phi) \, P(\Theta_j|\mathcal{M}_j, \Phi) d\Theta_j, \tag{4}$$

where Θ_j denotes the vector of parameters in $\mathcal{M}_j$ and $P(\Theta_j|\mathcal{M}_j, \Phi)$ is the decision-maker's posterior belief about these parameters. The second factor on the right-hand side is the posterior probability of $\mathcal{M}_j$, which is given by:

$$P(\mathcal{M}_j|\Phi) = \frac{P(\Phi|\mathcal{M}_j) \, P(\mathcal{M}_j)}{\sum_{i=1}^{J} P(\Phi|\mathcal{M}_i) \, P(\mathcal{M}_i)}, \tag{5}$$

[7] The risk-free rate of return is assumed to be constant over the investment horizon. Its monthly rate is set equal to the rate on a one-month Treasury bill prevailing at the end of the sample period which is 4.55% in annualized terms.

[8] Throughout the study, P will serve as a generic notation for probability density function.

where $P(\mathcal{M}_j)$ is the decision-maker's prior belief that $\mathcal{M}_j$ is the correct model, and $P(\Phi|\mathcal{M}_j)$ is the marginal likelihood of the data. The latter is computed as:

$$P(\Phi|\mathcal{M}_j) = \int_{\Theta_j} P(\Phi|\Theta_j, \mathcal{M}_j)\, P(\Theta_j|\mathcal{M}_j) d\Theta_j, \tag{6}$$

where $P(\Phi|\Theta_j, \mathcal{M}_j)$ and $P(\Theta_j|\mathcal{M}_j)$ denote the likelihood of the data conditioning upon the set of parameters in $\mathcal{M}_j$ and the investor's prior belief about these parameters, respectively. Equation (6) indicates that the marginal likelihood is merely the normalizing constant of the posterior density.

With a dogmatic belief about the return forecasting model, using the predictive density in (4) is the well-known Bayesian solution to deriving optimal portfolio choices in the presence of estimation risk. The predictive density is obtained after integrating out the parameter space Θ, which explicitly takes account of the uncertainty about the model parameters. The Bayesian solution in control problems was introduced by Zellner and Chetty (1965) and has been extensively used in portfolio decisions in the context of both *iid* returns and time-varying expected returns.

What if investors encounter a set of plausible forecasting models *a priori*? For example, investors might be uncertain about whether or not stock returns are predictable or, conditional upon having predictability, investors might be concerned with selecting the relevant forecasting variables. In that case, the notion of model uncertainty arises. Forecasting stock returns conditional upon a single model selected from the feasible set, as is common practice in econometrics, will tend to overstate the precision of the forecast by not taking account of the uncertainty about which model is the correct one.[9]

[9] Leamer (1983) pp. 36-37 describes the nature of the common practice of selecting an econometric model as follows: *"The concepts of unbiasedness, consistency, efficiency, maximum-likelihood estimation, in fact,*

The desire to account for model uncertainty in deriving a sensible portfolio choice dictates using a predictive distribution that averages out the model space - the unconditional predictive distribution displayed in equation (3). By doing so, the investor's problem can be described as the maximization of the unconditional expected utility in the presence of both uncertainty about the return forecasting model and parameter uncertainty within any forecasting model belonging to $\mathcal{M}$.

The decision-maker's optimization problem in the presence of such uncertainty is formulated as:

$$\omega^* = \arg\max_{\omega} \int_{R_{T+K}} U\left(W_{T+K}\right) \left[\sum_{\mathcal{M}_j \in \mathcal{M}} P\left(R_{T+K} | \mathcal{M}_j, \Phi\right) P\left(\mathcal{M}_j | \Phi\right) \right] dR_{T+K}, \qquad (7)$$

subject to the feasibility constraint displayed in equation (2). We do not truncate the tails of the stock-return distribution. Therefore, short selling and buying on margin are precluded; otherwise, the expected utility would be equal to $-\infty$. $C(\omega)$, the feasible set of allocations, thus satisfies:

$$C(\omega) = \left\{ \omega \ : \ 0 \le \omega_i < 1 \ \forall \, i, \text{ and } \omega' \iota_N < 1 \right\}. \qquad (8)$$

The optimal portfolio choice, ω^*, in equation (7) cannot be obtained analytically. Therefore, we approximate the integral in that equation by taking draws $\left\{ R_{T+K}^{(g)} \right\}_{g=1}^{G}$ from the unconditional (weighted) predictive distribution and using numerical optimization code to maximize the quantity:

$$\mathrm{E}\left[U(W_{T+K}(\omega))\right] = \frac{1}{G} \sum_{g=1}^{G} \frac{\left\{ (1 - \omega' \iota_N) \exp(r_f K) + \omega' \exp(r_f K \iota_N + R_{T+K}^{(g)}) \right\}^{1-\gamma}}{1-\gamma}, \qquad (9)$$

---

all the concepts of traditional theory, utterly lose their meaning by the time an applied researcher pulls from the bramble of computer output the one thorn of a model he likes best, the one he chooses to portray as a rose." Moulton (1991) employs this argument to motivate his Bayesian model averaging application.

where G denotes the number of draws.

The expectation in equation (9) is taken with respect to the weighted predictive distribution $P(R_{T+K}|\Phi) = \sum_{\mathcal{M}_j \in \mathcal{M}} P(R_{T+K}|\mathcal{M}_j, \Phi) P(\mathcal{M}_j|\Phi)$. Sampling from that distribution is obtained by first drawing a model $\mathcal{M}_j \in \mathcal{M}$ with probability $P(\mathcal{M}_j|\Phi)$. Future long-horizon cumulative excess log returns are then drawn from the conditional predictive distribution $P(R_{T+K}|\mathcal{M}_j, \Phi)$. Computing the posterior probability and drawing stock returns from the conditional predictive distribution will be discussed below.

B The Conditional Distribution of Stock Returns

The continuously compounded excess returns and predictive variables are modeled using a multivariate regression framework. In particular, the $N \times 1$ vector of excess log returns is the dependent variable in the multivariate predictive regression:

$$r_t' = x_{t-1}' B_R + \epsilon_t', \tag{10}$$

where $x_{t-1}' = (1, z_{t-1}')$, and z_{t-1} contains M *ex ante* variables observed at the end of $t-1$. These may include the dividend yield, term structure slope, default spread, and book-to-market ratio, among other variables that are believed to be relevant in forecasting asset returns. B_R is an $(M+1) \times N$ matrix of the multivariate regression coefficients. The disturbance terms, $\epsilon_t,\ t = 1, \ldots, T$, are assumed to obey the standard assumption of a zero expectation conditional on the set of instruments z_{t-1}. It is also assumed that the disturbances are serially uncorrelated and have variance equal to the $N \times N$ matrix Σ_{RR}.[10]

[10]Stambaugh (1986 and 1999) shows that while ϵ_t has zero expectation conditional on z_s for $s < t$, the expectation of ϵ_t conditional on z_s for $s \geq t$ is nonzero. In fact, contrary to the standard regression model, the predictive variables in a time series setting evolve stochastically. Therefore, regression equation (10) departs from the standard Bayesian regression framework discussed, for example, by Zellner (1971) and Box

By letting $y_t' = (r_t', z_t')$, it follows that the data-generating process can be characterized by the multivariate regression (also known as a restricted VAR):

$$y_t' = x_{t-1}' B + u_t', \tag{11}$$

where B is an $(M+1) \times (N+M)$ matrix of regression coefficients whose partition is given by:

$$B = \begin{bmatrix} B_R & B_z \end{bmatrix} = \begin{bmatrix} \alpha_R' & \alpha_z' \\ a_R & a_z \end{bmatrix}, \tag{12}$$

and $u_t \sim iid \;\; N(0, \Sigma)$ with:

$$\Sigma = \begin{bmatrix} \Sigma_{RR} & \Sigma_{Rz} \\ \Sigma_{zR} & \Sigma_{zz} \end{bmatrix}. \tag{13}$$

Note that α_R is the $N \times 1$ vector of intercepts in the multivariate predictive regression displayed in equation (10), and a_R is the $M \times N$ matrix of slopes in that regression. In the same vein, α_z is the $M \times 1$ vector of intercepts, a_z is an $M \times M$ matrix of slopes, and Σ_{zz} is the covariance matrix of the residuals in the multivariate regression of contemporaneous on lagged predictive variables. Lastly, Σ_{Rz} is the matrix of covariances of zero-mean contemporaneous innovations to excess returns and predictive variables.

In what follows, it will be convenient to work with the matrix form of the model:

$$Y = XB + U, \tag{14}$$

and Tiao (1997). In the standard Bayesian regression framework, the set of variables z_{t-1} are assumed to be either fixed non-stochastic or stochastic, but distributed independently of the disturbances of ϵ_t for $t = 1, \ldots, T$, and with a probability density function not involving the parameters B_R and Σ_{RR}.

where

$$Y = \begin{bmatrix} r_1', z_1' \\ \vdots \\ r_T', z_T' \end{bmatrix}, \; X = \begin{bmatrix} 1, z_0' \\ \vdots \\ 1, z_{T-1}' \end{bmatrix}, \; U = \begin{bmatrix} u_1' \\ \vdots \\ u_T' \end{bmatrix}.$$

Under the assumptions set forth above, it follows that

$$\text{vec}(U) \sim N\left(0, \Sigma \otimes I_T\right), \tag{15}$$

where $\text{vec}(U)$ is the column-wise vectorization of the matrix U, and I_T is the $T \times T$ identity matrix.

It is assumed throughout that z_0, an $M \times 1$ vector of initial values of the predictive variables, is non-stochastic. Therefore, the likelihood function of the data, $\mathcal{L}(B, \Sigma; Y, x_0)$, is proportional to the probability density function $P\left(Y|B, \Sigma, x_0\right)$ which is given by:

$$\begin{aligned} P\left(Y|B, \Sigma, x_0\right) &= (2\pi)^{-\frac{T(N+M)}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[(Y - XB)'(Y - XB)\Sigma^{-1}\right]\right), \qquad (16) \\ &= (2\pi)^{-\frac{T(N+M)}{2}} |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[S + (B - \hat{B})'X'X(B - \hat{B})\right]\Sigma^{-1}\right), \end{aligned}$$

where

$$\begin{aligned} \hat{B} &= (X'X)^{-1}X'Y, \\ S &= (Y - X\hat{B})'(Y - X\hat{B}), \end{aligned}$$

and $\text{tr}(\cdot)$ stands for the trace operator.

The framework set forth above and further developed below facilitates analyzing portfolio decisions in the presence of model uncertainty, parameter uncertainty within any forecasting model, multiple investable assets, and multiple predictive variables. In particular, the conditional distribution of future cumulative excess log returns, R_{T+K}, is multivariate normally distributed with mean and variance depending on B and Σ, which are uniquely determined for any given model, as discussed below, and the most recent realizations of the predictive variables, z_T. Formally, it follows that

$$R_{T+K}|B, \Sigma, \Phi, \mathcal{M}_j \sim N\left(\lambda_{\mathcal{M}_j}, \Upsilon_{\mathcal{M}_j}\right), \tag{17}$$

where Φ contains Y and z_0.

We show in part A of the appendix that if the stochastic processes governing the evolution of stock returns and predictive variables are both covariance-stationary, then $\lambda_{\mathcal{M}_j}$ and $\Upsilon_{\mathcal{M}_j}$ are given by (the notational dependence of the right-hand side parameters on $\mathcal{M}_j$ is suppressed):

$$\begin{aligned} \lambda_{\mathcal{M}_j} &= K\alpha_R \\ &+ a_R'\left[a_z'\left((a_z')^{K-1} - I_M\right)(a_z' - I_M)^{-1} - (K-1)I_M\right](a_z' - I_M)^{-1}\alpha_Z \\ &+ a_R'\left[((a_z')^K - I_M)(a_z' - I_M)^{-1}\right] z_T, \\ \Upsilon_{\mathcal{M}_j} &= K\Sigma_{RR} + \sum_{k=1}^{K}\phi(k)\Sigma_{zz}\phi(k)' + \sum_{k=1}^{K}\Sigma_{Rz}\phi(k)' + \sum_{k=1}^{K}\phi(k)\Sigma_{zR}, \end{aligned} \tag{18}$$

where

$$\phi(k) = a_R'\left[((a_z')^{k-1} - I_M)(a_z' - I_M)^{-1}\right].$$

Of course, no predictability corresponds to $a_R = 0$. In that case, λ_{iid} and Υ_{iid} are simply equal to $K\alpha_R$ and $K\Sigma_{RR}$, respectively. Obviously, without accounting for estimation risk, the conditional mean and variance in an *iid* world increase linearly with the investment horizon.

C Decomposing the Predictive Variance of Cumulative Stock Returns

In the presence of both model uncertainty and parameter uncertainty within any forecasting model, the variance of the K-period-ahead cumulative stock returns can be expressed as a sum of three terms:[11]

$$\mathrm{Var}\{R_{T+K}|\Phi\} = \sum_{j=1}^{J} P\left(\mathcal{M}_j|\Phi\right)\left[\mathrm{E}_\Theta\{\Upsilon_{\mathcal{M}_j}\} + \mathrm{Var}_\Theta\{\lambda_{\mathcal{M}_j}\} + \left(\tilde{\lambda} - \mathrm{E}_\Theta\{\lambda_{\mathcal{M}_j}\}\right)\left(\tilde{\lambda} - \mathrm{E}_\Theta\{\lambda_{\mathcal{M}_j}\}\right)'\right], \tag{19}$$

where E_Θ and Var_Θ denote the expected value and variance operators taken with respect to the parameter space and[12]

$$\tilde{\lambda} = \sum_{j=1}^{J} P\left(\mathcal{M}_j|\Phi\right)\mathrm{E}_\Theta\{\lambda_{\mathcal{M}_j}\}. \tag{20}$$

That is, the predictive variance of cumulative stock returns can be decomposed into three components: i) a mixture of variances from any candidate model, where each variance is conditioned on the set of model-specific parameters; ii) a mixture of the within-model parameter uncertainty; and iii) the model-uncertainty component. The empirical section that follows quantifies each of these three components for various investment-horizon lengths.

One might suspect that uncertainty about the correct model for predicting asset returns could play a role in portfolio decisions, since expected future returns are likely to differ across the forecasting models under consideration. The third factor in (19) reveals that such differences produce additional uncertainty about stock returns. That uncertainty makes stocks look less attractive to risk-averse investors, who therefore could allocate less to equities.

[11]Formal details are given in part B of the appendix.

[12]Using quadratic predictive loss, the predictive mean $\tilde{\lambda}$ is the optimal point prediction, and it can differ substantially from the predictive mean of any particular model, including the highest-posterior-probability model.

D Prior, Posterior, Predictive Distribution, and Marginal Likelihood

1 Prior Distribution of the Predictive-Regression Parameters

Inherent in the Bayesian decision-making process is an individual who establishes a prior belief about model parameters. If several plausible models are entertained *a priori*, then the notion of unconditional prior density emerges. The unconditional prior averages over conditional priors, which, in turn, are denoted by $P_0\left(B, \Sigma | \mathcal{M}_j\right)$ for $j = 1, \ldots, J$, weighted by the prior probability in favor of the model:

$$P_0(B, \Sigma) = \sum_{j=1}^{J} P(\mathcal{M}_j) P_0\left(B, \Sigma | \mathcal{M}_j\right). \tag{21}$$

At this point, we must specify informative prior beliefs about parameters belonging to each of the models under consideration. It is known that non-informative improper priors do not work in hypothesis testing and model selection problems. Because in most cases, the arbitrary constants appearing in the improper prior specification do not vanish, the resulting marginal likelihood and posterior probabilities are not determined uniquely, but instead depend upon arbitrary normalizations.[13]

The literature proposes two common approaches to specify a prior density for the regression parameters. In one approach, as discussed, for example, by Chib and Greenberg (1996), B and Σ are independent and obey the multivariate normal and inverted Wishart densities, respectively. In the other approach, the so-called natural conjugate prior, the marginal prior probability density function (henceforth "pdf") for Σ is still inverted Wishart, but the marginal distribution of B obeys the Matric-Student t pdf.[14]

Let us employ the first approach and use a multivariate normal density as the marginal

[13]See, for example, Poirier, 1995, pp. 544-545.

[14]See, for example, Box and Tiao (1997).

prior pdf for the matrix B.[15] In so doing, we pay a price, for the joint posterior pdf does not obey an analytical form. That is, the general prior normal pdf does not combine so neatly with the likelihood function, as the natural conjugate counterpart. However, we show that applying a Markov chain Monte Carlo procedure facilitates drawing from the joint posterior in a straightforward manner.

The prior pdf of parameters belonging to $\mathcal{M}_j$ obeys the form (the notational dependence of the right-hand-side parameters on $\mathcal{M}_j$ is suppressed):

$$\begin{aligned} P_0\left(B, \Sigma | \mathcal{M}_j\right) &= N(\tilde{\beta}, C) \times W^{-1}(H_0, \nu_0), \\ &\propto \exp\left(-\frac{1}{2}(\beta - \tilde{\beta})' C^{-1} (\beta - \tilde{\beta})\right) \times |\Sigma|^{-\frac{(\nu_0 + N + M + 1)}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}[H_0 \Sigma^{-1}]\right), \end{aligned} \tag{22}$$

where $\beta = \mathrm{vec}(B)$. The regression parameters B and Σ are independent in the prior and $\mathrm{vec}(B)$ is normally distributed with mean and variance equal to $\tilde{\beta}$ and C, respectively. The marginal distribution of Σ obeys the form of the inverted Wishart pdf with parameter matrix H_0 and ν_0 degrees of freedom.

With as many as M instrumental variables, there are $J = 2^M$ distinct linear compositions of predictive variables, each of which retains some of the variables as valuable

[15]There is a non-trivial caveat in implementing the natural conjugate prior distribution in applied work. This caveat is discussed by Zellner (1971) pp. 238-240, who, in turn, refers to Rothenberg (1963). Rothenberg (1963) notes that using the natural conjugate prior involves placing restrictions on the parameters due to the fact that the matrix $(X'X)^{-1}$ enters the covariance structure in the following way $\Sigma \otimes (X'X)^{-1}$. Such a covariance matrix imposes equality upon the ratios of variances of corresponding coefficients for any pair of equations. In our setting, that restriction is particularly severe since, as shown below, within any forecasting model, a few entries in the covariance matrix of B are zeroed out to reflect a dogmatic belief about the uselessness of some variables in predicting asset returns. If we were to use the natural conjugate prior, we would be forced to zero out additional entries in the covariance matrix, which might result in a contradiction to those beliefs.

stock-return predictors and treats the others as useless. Following Box (1980), the combination of a prior pdf $P_0(B, \Sigma|\mathcal{M}_j)$ and the common likelihood function $\mathcal{L}(B, \Sigma; Y, x_0)$ in (16) is defined a statistical model $\mathcal{M}_j$. Although all the entertained forecasting models share a common likelihood function, their uniqueness stems from unique prior pdfs, as described below.

Values for the prior parameters $\tilde{\beta}$, C, H_0, and ν_0 are determined uniquely for each model and are obtained using two sources of information: the unique data-generating process implied by each forecasting model; and a training sample. Specifically, for any forecasting model, the prior belief about slopes in the regression of current stock returns on lagged useless variables are centered around zero or, more formally, for these slopes, there are discrete spikes of probability at zero. In doing so, we elicit prior values for a few entries in the mean vector $\tilde{\beta}$ and the variance matrix C.[16]

Computing the marginal likelihood also necessitates the elicitation of prior beliefs about the remaining parameters.[17] This task can be pursued by splitting the sample into two

[16]Centering a slope coefficient around zero, we explicitly treat that coefficient as a fixed non-random quantity, meaning that its variance and covariances with other parameters are zeroed out as well. Of course, zeroing out the variances of such slope coefficients causing their posterior distributions to be concentrated at zero.

[17]Leaving the prior as non-informative but still proper by specifying large prior variances is hopeless. For example, Pastor and Stambaugh (1999) note that doubling an already large variance keeps the prior non informative, hence having a minor effect on the posterior distribution; but the marginal likelihood might display a non-trivial sensitivity to that change. One way to avoid the difficulties arising from using a Bayesian technique in model selection is to employ the Schwarz criterion instead, which leads to appropriate conclusions in sufficiently large samples. However, the notion "sufficiently large" is rather obscure, and its operational meaning lacks any intuitive appeal. Another way of handling improper priors is the "imaginary training sample device." This consists of imagining that an additional data set is available. See Kass and

parts, the so-called training sample and the remaining observations constituting the primary sample.[18] The training sample is often combined with the improper prior distribution to produce a proper prior distribution. In particular, when a flat prior on regression parameters is combined with a normal likelihood function of data from the training sample, the natural conjugate prior emerges. (See, for example, Zellner, 1971.)

Our study undertakes an approach similar in spirit to the one described above, in that we set aside part of the data to be used as a training sample. However, we depart in two ways from the above-described methodology, which implicitly employs the natural-conjugate prior specification. First, the regression parameters B and Σ are kept independent in the prior, as noted above. Second, the training sample is used to establish only the remaining prior parameters, or those that have not been determined by beliefs about the correct return forecasting model. (For any forecasting model, such prior parameters include the intercepts and slope coefficients in the regression of current stock returns on lagged 'useful' predictive variables, their variances and covariances.)

Prior sensitivity will be investigated by entertaining three splitting points corresponding to $T_0 = 45$, $T_0 = 60$, and $T_0 = 75$ time series observations. The length of the training sample is mapped into our prior value for the inverted Wishart parameter ν_0 which, in turn, is equal to $T_0 - N - M - 1$. The choice of various splitting points was applied in other studies undertaking Bayesian approach for hypothesis testing and model selection, including McCulloch and Rossi (1991), who advocate a Bayesian approach to testing the arbitrage pricing theory. McCulloch and Rossi (1991) emphasize that ν_0 need not take too small or

Raftery (1995) and the reference therein.

[18]See, for example, Berger and Pericchi (1996), McCulloch and Rossi (1991), and Moreno, Bertolino, and Racugno (1998).

too high values; otherwise the prior would be extremely diffuse or too tight, respectively.

As noted earlier, each information variable does appear in the likelihood function (16) as part of the data whether or not it is believed to be useful in forecasting stock returns. The rationale is that in multi-period asset-allocation decisions, the stochastic evolution of a predictive variable plays a role in determining the distribution of future stock returns. (The first two moments of the cumulative excess log returns displayed in equation 18 serve as an illustrative example.) In particular, it might be the case that a variable is treated as useless in forecasting stock returns at the same time that it is treated as useful in forecasting other variables that are believed to be useful stock-return predictors.

2 Prior Probabilities of Return-Forecasting Models

Implementing Baysian model averaging, the decision-maker has to assign prior probabilities for the 2^M forecasting models. A conceivable way of dealing with the concern that "apparent predictability" is due to data mining is to associate higher prior probabilities to theoretically motivated predictive variables. (Section 3 discusses the choice of predictive variables in this study.) However, there are some practical difficulties with judging the relative strength of the theoretical arguments and, in particular, the plausibility of the underlying assumptions used to derive them. After all, theories are not formed in a vacuum. The theory-maker has already learned about many of the relationships among variables in the data. Furthermore, if a decision-maker suspects that some predictive variables are more likely to generate the evolution of expected stock returns than other predictors, or perhaps that a combination of more predictive variables is more favorable than combinations with fewer predictors, then assigning a higher prior probability to the more favorable model entails subjective judgment, which need not reflect a consensus view.

Obviously, we prefer not to get involved in determining which of the 2^M combinations is more favorable *a priori*. However, if we were to assign an equal prior probability to each candidate model being the "correct" one *ex ante*, as it is the implicit common practice using classical model-selection criteria, such as the adjusted R^2, we would weight the prior belief in favor of predictability by assigning a prior mass of only $1/2^M$ to the *iid* model.[19] Of course, that prior mass approaches zero with an increasing number of predictive variables.

Assuming that (i) investors consider predictability or lack thereof as equally likely *ex ante*, and (ii) conditional on having predictability, investors exhibit non-informative beliefs about the form that it takes, we associate a prior probability of 50% to the *iid* model and assign the remaining 50% equally across all other models. This specification reflects a substantial prior uncertainty about whether stock returns are predictable, and will be referred throughout as the "non-equal-prior-probability scenario." We also assign prior probabilities equally across models. This specification will be referred throughout as the "equal-prior-probability scenario." The latter scenario restricts potential implications of model uncertainty in stock-return predictability for optimal portfolio rules, because as noted earlier, in that scenario, investors exhibit an almost complete faith that stock returns are indeed predictable, and are only uncertain about the relevant variables that forecast these returns.

A convenient argument to combat the methodology developed here is the somewhat arbitrary fashion that prior probabilities are assigned to models. Taking this to the extreme, had the prior probability of a single model been unity, then the notion of model uncertainty would have completely disappeared. However, it is the nature of applied work that the

[19] It is common practice in Bayesian model averaging to take the prior probability of each model as equal. See, for example, Moulton (1991).

researcher reveals a prior with respect to the choice of model. For instance, Barberis (1999) implicitly assumes that the prior probability of the dividend yield model is unity. Employing a flat prior for the slope coefficient in the regression of excess returns on lagged dividend yield, Barberis assigns a zero probability to the event that the slope is equal to zero, thereby weighting his prior belief against no predictability. Bawa, Brown, and Klein (1979), among others, treat the *iid* model of stock returns as the "correct" one. Moreover, using model selection criteria, such as AIC (Akaike, 1974), SIC (Schwarz, 1978), or adjusted R^2, the researcher expresses her belief that one among several models fairly describes the state of the world. She implicitly assigns equal prior probabilities to all models under consideration and eventually chooses a single one to predict quantities of interest.

3 Posterior Distribution of the Predictive-Regression Parameters

Part C of the appendix details the derivation of the joint posterior pdf for the regression parameters. It is shown that the joint posterior does not resemble any well-known distribution. Nonetheless, each of the full conditional distributions obeys an analytical form. Naturally, we apply the Gibbs sampler technique to draw from the joint posterior.[20] The aforementioned posterior density is conditioned on a selected model and is denoted by $P\left(B,\Sigma|\Phi,\mathcal{M}_j\right)$. An unconditional posterior distribution for B and Σ is obtained by combining the likelihood function (16) and the unconditional prior density (21). The resulting posterior averages over the conditional posterior pdfs, weighted by the posterior probability in favor of the model:

$$P(B,\Sigma|\Phi) = \sum_{j=1}^{J} P(\mathcal{M}_j|\Phi) P\left(B,\Sigma|\Phi,\mathcal{M}_j\right). \tag{23}$$

[20]The Gibbs sampler is an iterative Monte Carlo method designed to extract marginal distributions from intractable joint distributions. The Gibbs sampler was introduced by Geman and Geman (1984). Chib and Greenberg (1996) present examples applying the Gibbs sampler to a variety of econometric models.

Part D of the appendix provides a three-step algorithm for drawing cumulative excess returns from the weighted predictive distribution. Details on computing marginal likelihoods and posterior probabilities are given in part E of the appendix.

II The Data

For each of several return forecasting models, we compute the optimal portfolio choice as weights assigned to three size portfolios and the risk-free Treasury bill. Size portfolios are constructed from CRSP's capitalization file as equal averages of small, medium, and large capitalization firms belonging to decile portfolios 1-3, 4-7, and 8-10, respectively. We consider six information variables: the dividend yield (Div) on the value weighted NYSE index; the book-to-market (BM) on the Standard & Poor's Industrials; the default spread (Def); the term structure slope (Term); the trend-deviation-in-wealth (TDW) advocated by Lettau and Ludvigson (1999); and the relative bill rate (Tbill).[21]

The dividend yield is constructed as in Fama and French (1988, 1989) and others. Specifically, dividend yield is the total payment of dividends on the value-weighted NYSE portfolio over the recent twelve months divided by the contemporaneous level of the index. Inputs for calculating book-to-market is obtained from the Standard & Poor's publication: "Security Price Index Record - Statistical Service." Book-to-market is calculated for all firms in the Standard & Poors' Industrials. Year-end book value on Standard & Poor's Industrials is

[21] In previous versions of the paper Earnings yield on the Standard & Poor's Industrials was included as an additional predictive variable. Earnings yield possesses very high contemporaneous correlations with dividend yield and book-to-market and was, therefore, omitted.

available starting from December 1946. A quarterly book-to-market ratio is constructed by dividing the most recent year's book value by the contemporaneous level of the Standard & Poors' Industrials. For example, the book-to-market ratio in March 1990 is the book value for December 1989, divided by the level of the Standard & Poors' Industrials in March 1990.

The default spread is the difference in annualized yields of Moody's Baa and Aaa rated bonds. The term structure slope is the difference in annualized yields of ten-year and one-year Treasury bills. The relative bill rate is the difference between an annual yield on a three-month Treasury bill and its one year backward moving average. All the Treasury-bill yields are available at the Federal Reserve Board's web-site.

Trend-deviation-in-wealth is computed as $c_t - wa_t - (1 - w)y_t$, where c_t, a_t, and y_t denote log consumption, non-human wealth, and labor income, respectively.[22] The weight w equals the average share of non-human wealth in total wealth. TDW is in per-capita terms, measured in 1992 dollars. It is available quarterly starting from the first quarter of 1953. Inputs for computing TDW are released by the Federal Reserve Board. The data is published within two months of the end of the quarter. Therefore, the TDW realization at quarter t is made known to capital market participants at quarter $t + 1$ and, therefore, used

[22] Data on TDW were generously provided by Martin Lettau. Lettau and Ludvigson (1999) have shown that trend deviations in wealth are strong predictors of future returns at short and intermediate horizons. They consider a forward-looking model whose theoretical motivation goes back to Campbell and Mankiw (1989) and Campbell (1993). This model implies that the log consumption, labor income, and non-human wealth share a common trend. The deviations from this trend summarize investors' expectations about future returns on the market portfolio. In particular, when consumption is above its long-term trend with respect to labor income and non-human wealth, asset returns are expected to rise and *vice versa*. To learn about the variables used to compute TDW, its theoretical motivation, and econometric methodology used to estimate the unobserved parameter w, readers are referred to Lettau and Ludvigson (1999).

to predict stock returns to be realized at quarter $t + 2$.

Theoretical motivations have been advanced in the financial economics literature for each of the predictive variables used in this study. For example, the textbook treatment of Campbell, Lo, and MacKinlay (1997) describes a framework in which the dividend yield reflects investors' expectations about future returns on the market portfolio. Building on a similar framework, Lettau and Ludvigson (1999) argue that the trend-deviation-in-wealth summarizes such expectations. Furthermore, Pontiff and Schall (1998) advocate using the book-to-market as a predictor of equity market returns. In particular, they assert that book value proxies for future cash flows. Dividing such proxy by the contemporaneous level of the stock index produces a variable which is correlated with future stock returns. Bossaerts and Green (1989) describe a general equilibrium model in which conditional expected returns are inversely related to the price of an asset. Lastly, Boudoukh, Richardson, and Whitelaw (1997) exploit the pricing kernel representation to uncover a (non-linear) relationship between the conditional expected equity premium and the term structure slope.

Following the availability of data, as described above, our sample contains quarterly observations spanning the first quarter of 1953 to the fourth quarter of 1998. Table I presents summary statistics for the predictive variables for all the sample data, and Figure 1 plots the evolution of the predictive variables over the sample period. It is shown that book-to-market, dividend yield, and default spread display a fairly high persistency suggesting a long-term mean-reversion, whereas TDW, Tbill, and Term exhibit a lower auto-correlation and hence a faster mean-reversion. We also report slope coefficients and standard deviations obtained from running separately three multiple regressions of each of the size portfolios on lagged predictive variables.

III Results

A Posterior Probabilities of Return Forecasting Models

Consideration of all linear data-generating processes in the presence of six predictive variables necessitates the comparison of $2^6 = 64$ models. Tables II and III display posterior probabilities for several compositions of predictive variables, and each constitutes a linear forecasting model. Tables II and III correspond to the non-equal and equal prior-probability scenarios, respectively. Posterior probabilities of other forecasting models are fairly small and are, therefore, not displayed in the tables.

The compositions are uniquely identified by a combination of zeros and ones designating exclusions and inclusions of predictors, respectively. Posterior probabilities are computed with T_0 taking three distinct values representing three splitting points, as described above. The last column in both tables reports average posterior probabilities across the three values of T_0 within each forecasting model.[23]

Several features of the results displayed in Tables II and III merit closer attention. Focusing on the non-equal-prior-probability scenario, we observe that across the three prior specifications, posterior probabilities of the *iid* model (composition number 3 in Table II) are smaller than the pre-assigned prior probability of 50%, suggesting that, to some extent, the data favors predictability in stock returns.

Interestingly, the TDW is the most prominent predictor of returns on size-sorted portfolios. The model retaining that variable as a single predictor receives the highest-posterior

[23] The weighted model used to explore the implications of model uncertainty for optimal portfolio rules employs the average probabilities as weights on the corresponding individual compositions.

probability. Moreover, the last two rows in Tables II and III indicate that TDW appears as a valuable predictor for more combinations than its counterparts. In the same vein, the cumulative average posterior probability of models including the trend-deviation as a valuable predictor is 73% under the non-equal-prior-probability scenario, and is 87.2% under the equal-prior scenario.

Among the traditional market multipliers, the book-to-market and dividend yield, the latter seems to outperform the former. Dividend yield appears in two of the non-zero-probability models with a cumulative average probability of either 27.03% or 28.55%, depending on the prior scenario. At the same time, book-to-market does not appear as a valuable predictor in any of the non-zero-probability compositions. Following the same logic, the relative bill rate is proved to be especially dismal as stock-return predictor – it does not emerge in any of the non-zero probability models. The term structure slope appears in two compositions with cumulative posterior probabilities of either 24.8% or 27.88%, depending on the prior-probability scenario. Lastly, the default spread appears only once as a valuable predictor with low posterior mass, suggesting that default spread has, at best, only weak forecasting power.

The notion of selecting among models poses the question "which of the models under consideration is the most likely to have generated the data?" Having posterior probabilities at hand, the search for the "correct" model can be addressed by computing posterior-odds ratios or Bayes factors for each pair of models under consideration. The Bayes factor of $\mathcal{M}_j$ against $\mathcal{M}_i$ is defined as the ratio of the marginal likelihoods of these models: $\mathcal{B}_{ji} = \frac{P(\Phi|\mathcal{M}_j)}{P(\Phi|\mathcal{M}_i)}$. Multiplying the Bayes factor by the prior-odds ratio, $\frac{P(\mathcal{M}_j)}{P(\mathcal{M}_i)}$, yields the posterior odds in favor of $\mathcal{M}_j$ or against $\mathcal{M}_i$: $\frac{P(\mathcal{M}_j|\Phi)}{P(\mathcal{M}_i|\Phi)} = \mathcal{B}_{ji}\frac{P(\mathcal{M}_j)}{P(\mathcal{M}_i)}$. If $\mathcal{M}_i$ and $\mathcal{M}_j$ are equally likely *a priori*, then the posterior-odds ratio is equal to the Bayes factor.[24]

Using Bayes factors, the null and alternative hypotheses can be postulated in a non-nested structure as:

$$H_{0i}: \; P(B, \Sigma|\Phi) = P(B, \Sigma|\Phi, \mathcal{M}_i)$$

$$H_{1j}: \; P(B, \Sigma|\Phi) = P(B, \Sigma|\Phi, \mathcal{M}_j) \quad \text{for} \quad j \neq i, \text{ and } i, j = 1, \ldots, 64.$$

Simplistically, the data is said to favor H_{0i} relative to H_{1j} if the Bayes factor of $\mathcal{M}_i$ against $\mathcal{M}_j$ exceeds one. That is, the observed data is more likely to be generated under hypothesis H_{0i} than under H_{1j}. Following this logic, if the posterior probability of $\mathcal{M}_i$ is greater than that of $\mathcal{M}_j$, then $\mathcal{M}_i$ is superior in hypothesis testing. Insofar as we do not consider the *iid* model, the Bayes factor for $\mathcal{M}_i$ against $\mathcal{M}_j$ is computed as the posterior-odds ratio for the corresponding models. We assert that there is only a single sensible hypothesis test for the *iid* model, to wit: the *iid* against all other models (i.e., predictability versus no-predictability in stock returns). Under the non-equal-prior scenario, the prior odds for such hypothesis is unity, and the corresponding Bayes factor, which is equal to the posterior odds, takes either of the values 55.50, 14.70, or 1.75, depending on the training-sample length.

[24]The Bayes factor has the appealing feature of allowing hypothesis testing in a non-nested structure. Applying such structure in the classical approach in overly complicated, if not hopeless (see Leamer, 1978 p. 90). Jeffreys (1961) and Kass and Raftery (1995), among others, advocate the superiority of the Bayes factors in hypotheses testing versus resorting to the conventional P-values. Some practical difficulties with P-values are apparent in large samples, where P-values tend to reject null hypotheses even when the null model seems appealing theoretically, and close inspection of the data does not reveal any striking discrepancies. Poirier (1995), p. 614 describes the Bayesian moral in the following manner: *"Never make anything more than relative probability statement about the models explicitly entertained. Be suspicious of those who promise more!"*

What can we learn from the above-computed figures? Jeffreys (1961) suggests a qualitative interpretation for evidence in favor of the null or against the alternative hypotheses:[25]

$\log_{10}(\mathcal{B}_{i,j})$	$\mathcal{B}_{i,j}$	Evidence against H_j
0 to $\frac{1}{2}$	1 to 3.2	Does not justify more than a bare mention
$\frac{1}{2}$ to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Following Jeffreys, in two of three cases the Bayes factors/posterior odds provide strong statistical evidence in favor of predictability in asset returns. Economic implications of ignoring predictability while forming portfolio rules will be analyzed below.

For the sake of comparison, we also report the compositions selected by the AIC and SIC criteria:[26]

[25] Jeffreys is known as a philosopher and pioneer in Bayesian methods whose book on probability is considered to be of great importance for the philosophy of science.

[26] We compute the AIC and SIC criteria for all the 64 combinations. AIC and SIC are defined as:

$$\begin{aligned} \text{AIC} &= -2(\text{log maximized likelihood}) + 2(\tilde{N}), \\ \text{SIC} &= -2(\text{log maximized likelihood}) + (\log T)(\tilde{N}), \end{aligned}$$

where T denotes the sample size and $\tilde{N}$ is the number of parameters within any model. Akaike (1974) and Schwarz (1978) suggest that given a set of rival models, the one that should be selected minimizes the corresponding quantity. The first factors on the right-hand side of the equations are identical and measure the goodness of fit, whereas the second penalize for the complexity of the model. Notice that SIC penalizes more heavily for higher dimensional models and hence tends to select compositions with fewer predictors.

Predictor	BM	Div	Def	TDW	Tbill	Term
AIC	1	1	0	1	0	1
SIC	0	1	0	0	0	1

It is known that SIC leads to the same conclusion as the Bayes factor in sufficiently large samples, if *equal* prior probabilities are assigned. The posterior probabilities displayed in Table 3, suggest that the SIC criterion does indeed provide a reasonable indication of the evidence as the model selected by SIC is ranked as either the second, or third, or forth highest-posterior model, depending on the splitting point, whereas the model selected by AIC is among the zero-probability models.

B Model Uncertainty - The Variance Effect

As a first step in exploring the potential importance of differences across models, we measure the three components included in the variance of future stock returns as it emerges in equation (19). This step is pursued by drawing future returns from both the weighted predictive distribution, as explained in part D of the appendix, and from a hypothetical model that is similar to the weighted model, except that it does not account for the cross-model uncertainty. Future returns for that model are drawn using an algorithm described in the next paragraph. That algorithm is implemented separately for both prior-probability scenarios.

For any of the six (five) highest-probability models, displayed in Table 2 (Table 3), we draw cumulative future stock returns from their predictive distributions for investment horizons ranging from one to ten years. We then compute predictive means and variances of the returns using equation (18). Lastly, we draw a vector of future excess returns from a multivariate normal distribution whose mean and variance are obtained by averaging over means and variances computed for each of the six (five) models, weighted by their posterior probabilities.

Repeating the above-described algorithm, we obtain a random sample from the exact predictive distribution of the hypothetical model. Next, we take the difference between the variance based on the weighted model and the variance based on the hypothetical model to obtain the model-uncertainty component, as described in equation (19):

$$\sum_{j=1}^{J} P\left(\mathcal{M}_j|\Phi\right)\left(\tilde{\lambda} - \mathrm{E}_{\Theta}\{\lambda_{\mathcal{M}_j}\}\right)\left(\tilde{\lambda} - \mathrm{E}_{\Theta}\{\lambda_{\mathcal{M}_j}\}\right)'. \tag{24}$$

That algorithm enables us to compute the parameter-uncertainty component of model j, $\mathrm{Var}(\lambda_{\mathcal{M}_j})$, for each of the non-zero probability forecasting models. The mixture of parameter uncertainty is straightforwardly obtained by averaging across the parameter-uncertainty components associated with each model, using posterior probabilities as weights.

The three plots on the left in Figure 2 exhibit the decomposition of annualized variances of returns on small, medium, and large stock portfolios into the three components. The plots correspond to the non-equal-prior-probability scenario, in which investors display substantial prior uncertainty about whether stock returns are predictable. The predictive variance for each portfolio is annualized and equal to $4/K$ times the variance of R_{T+K}. (Recall that our sample contains quarterly observations.)

The solid lines show annualized variances for three size portfolios based on the weighted model. The dashed lines correspond to variances that include the within-model parameter uncertainty, but not the cross-model uncertainty. Those variances are the three diagonal entries in the 3×3 matrix $\sum_{j=1}^{J} P\left(\mathcal{M}_j|\Phi\right)\left[\mathrm{Var}(\lambda_{\mathcal{M}_j}) + \mathrm{E}(\Upsilon_{\mathcal{M}_j})\right]$. The dash/dot lines display mixtures of variances from any candidate model computed as though the model-specific parameters were known. These are the three diagonal elements in the 3×3 matrix $\sum_{j=1}^{J} P(\mathcal{M}_j|\Phi)\,\mathrm{E}(\Upsilon_{\mathcal{M}_j})$.

The difference between the solid and dashed lines constitutes the model-uncertainty component, and the difference between the dashed and dash/dot lines constitutes the mixture of parameter uncertainty component. The three plots on the right in Figure 2 display ratios of variances for each size portfolio obtained by dividing the cross-model uncertainty by the mixture of within-model parameter uncertainty for investment horizons ranging from one to ten years.

Figure 2 conveys an important insight into the role of model uncertainty in determining the perceived variance of future stock returns. In particular, we show that the fraction of *ex ante* variance attributable to model uncertainty can be significantly higher than the parameter-uncertainty component, especially for short and medium investment horizons and for large capitalization firms. For small stock portfolios the cross-model component ranges between 1.6% and 2.4%, whereas the within-model component ranges between 1.4% and 2.3%. Similarly, for large stocks the cross-model component ranges between 1.0% and 1.5%, whereas the within-model component ranges between 0.4% and 0.7%. Notice that the cross-model component accounts for 14.3% to 19.6% of the total predictive variance of small stock returns, for 19.2% to 26.7% for medium, and for 20.1% to 29.3% for large, depending on the investment horizon.

We repeat the analysis described above for the equal-prior-probability scenario, in which investors reveal a high prior confidence that stock returns are indeed predictable. Figure 3 contains the corresponding plots. We observe that under the equal-prior scenario, the cross-model component is fairly small for short and medium investment horizons and completely disappears for horizons exceeding five years. In contrast, the mixture of within-model parameter uncertainty is still prominent. That component increases with the horizon, displays a low sensitivity to the change in prior probabilities, and its magnitude ranges between 1.37% and 2.64% for small stocks, between 0.91% and 1.52% for medium, and between 0.54% and 0.60% for large stocks, depending on the investment horizon. An almost non-existent cross-model component implies that the primary source of variance attributable to model uncertainty emerges as investors exhibit a substantial prior uncertainty about whether returns are predictable. The realized data appear to have resolved the *ex ante* model uncertainty associated with the selection of instruments, which, in turn, becomes marginal.

Inspecting the right-hand-side plots in Figures 2 and 3, one can observe that both parameter uncertainty and model uncertainty display sensitivity to varying horizons. While the former increases with a lengthening horizon, the latter tends to decrease. Furthermore, for both prior-probability scenarios, the ratio obtained by dividing the cross-model component by the mixture of the within-model parameter uncertainty counterpart decreases almost monotonically with an increasing holding period.

The intuition behind the above-described phenomenon is as follows. Consider the case in which stock returns are *iid* and the prior belief about the parameters is non-informative. Part F of the appendix shows that in that case, the predictive distribution of stock returns, which accounts for parameter uncertainty, obeys the Student t pdf, and the corresponding predictive annualized variance increases linearly with the horizon length. (Obviously, without incorporating parameter uncertainty, future returns are normally distributed, and conditional variances remain constant with increasing horizons.) Note that with longer horizons there are simply more parameters to estimate, enhancing the role of estimation risk.

What is the horizon role in determining the variance attributable to model uncertainty? In longer horizons the predictive variables tend to revert to their long-term means, making conditional expected stock returns look similar across the various forecasting models. As a result, it is expected that the total predictive variance attributed to model uncertainty will converge to a fixed quantity and, subsequently, the annualized predictive variance, obtained by dividing that fixed quantity by the horizon length, will diminish with an increasing horizon.

Stock-return variances are computed under an additional scenario, in which the recent values of the predictive variables are set equal to their sample means. Figure 4 exhibits the decomposition of annualized variances of returns on small, medium, and large stock portfolios into the three components, noted earlier, as investors treat predictability or lack thereof as equally likely *ex ante*.

It emerges that the importance of both the cross-model uncertainty and within-model parameter uncertainty crucially depends on the most recent values of the predictive variables. To elaborate, centering the current values around their sample means virtually eliminates the differences in conditional means across models (and along the investment horizon within any given model), thus eliminating the cross-model component, even if there is a substantial prior uncertainty about whether returns are predictable, let alone in the absence of such uncertainty. Interestingly, Figure 4 shows that for recent values equal to the sample means, the within-model parameter uncertainty substantially diminishes as well.

In recent years, equity markets have been overwhelmingly bullish. Hence, the current values of variables that are perceived to have been indicators of fundamental values — such as book-to-market, dividend yield, and trend-deviation-in-wealth — have been substantially distant from their sample means. Some figures are presented below:

Predictive Variable	Level as of December 31, 1998	Sample Moments Mean	StDev
Div	0.0155	0.0363	0.0094
BM	0.1178	0.5078	0.1790
TDW	-0.0376	-0.0004	0.0112

The departures from long-term means suggest that perhaps it is not too surprising that model uncertainty has a relatively large impact on the distribution of future returns at the end of the sample period. Investors are uncertain about whether or not stock returns are predictable. Obviously, that uncertainty is more meaningful at the recent period compared to other periods when measures of perceived fundamental values were closer to their historical means.

C Model Uncertainty - Implications for the Optimal Portfolio Choice

Encompassing implications of model uncertainty for the investment opportunity set are assessed by comparing the portfolio selection based on the weighted model, which incorporates model uncertainty, with the same quantities based on single models that might have been selected otherwise by investors. The single models include the highest-posterior-probability model, displayed as the first composition in Tables 2 and 3, the all-inclusive model, and the model that drops all entertained predictors - the *iid* model of stock returns.

Figure 5 exhibits total allocations to equities across the above-mentioned models when the recent values of the predictive variables are equal to the actual realizations, as documented at the end-of-sample period. Asset allocations are derived for investment horizons ranging from one to ten years and relative risk-aversion coefficients of two, six, and ten. The solid and dotted lines denote allocations based on the weighted model - the equal and non-equal prior scenarios, respectively. The dashed lines correspond to the highest-posterior-probability model, the dash/dot lines correspond to the all-inclusive model, and the lines denoted by the symbol '*' correspond to the *iid* model.

The reader may note that leaving the most recent values of the predictive variables equal to actual realizations results in corner solutions, in which optimal allocations to equities are zero across short and medium horizons. Corner solutions emerge since the predicted equity premiums across models that include predictive variables are negative over such investment horizons. As noted earlier, the current values of predictive variables that are perceived to have been indicators of fundamental values are substantially smaller than their long-term means, suggesting that the equity market was believed to have been overvalued around the end of the sample period.

Figure 5 shows that the overall allocation to equities based on the weighted model, which accounts for both model uncertainty and parameter uncertainty within any non-zero posterior-probability model, increases with the investment horizon under both prior scenarios. This effect is due to both an increase in conditional expected returns towards long-run means and a reduction in the standard deviation due to such mean reversion. It is shown that asset allocations based on either of the single models and particularly on the *iid* model can considerably depart from those computed based on the weighted model. In particular, the differences display sensitivities to the investment horizon and the relative risk aversion coefficient.

Focusing on the equal-prior scenario (the non-equal-prior scenario), we show that an investor who is forced to ignore predictability and select instead the *iid* model over-allocates to equities by a fraction of wealth ranging from 70% to 99% (63% to 99%) for $\gamma = 2$, from 24% to 37% (23% to 37%) for $\gamma = 6$, and from 14% to 22% (13% to 22%) for $\gamma = 10$, depending on the holding period. Comparing asset allocations based on the weighted model with asset allocations based on the highest-posterior-probability model, we find that for $\gamma = 2$ the differences range between zero and 11% for the equal-prior scenario and between zero and 7% for the non-equal-prior scenario, depending, again, on the holding period. A similar comparison with the all-inclusive model reveals that for the equal-prior scenario the differences range between 3% and 21% for $\gamma = 2$, 1% and 8% for $\gamma = 6$, and 1% and 5% for $\gamma = 10$. For the non-equal-prior scenario, the differences range between 3% and 28% for $\gamma = 2$, 1% and 10% for $\gamma = 6$, and 1% and 6% for $\gamma = 10$.

Investors' risk tolerance displays an important role in determining the impact of model uncertainty in equity markets. It should be noted that such impact substantially diminishes as investors become less risk tolerant. That phenomenon is best explained through two polar examples. First, let us consider an infinitely risk-averse investor who does not allocate funds to equities or who prefers to hold all his wealth in a risk-free cash account. That investor is completely indifferent to the appearance of model uncertainty in equity markets since he does not approach these markets.

Let us consider now the opposite extreme — a risk-neutral investor who makes portfolio decisions based only upon the first moment of stock returns. Although a risk-neutral investor ignores the higher predictive variance in equity markets attributable to the uncertainty about the return forecasting model, he is not indifferent to model uncertainty, since expected stock returns do differ across the forecasting models. As a result, one would expect that the impact of model uncertainty will diminish as investors display a stronger aversion towards bearing risk.

The impact of model uncertainty is sensitive to the investment horizon as well. A close inspection at Figure 4 suggests that differences in portfolio allocations across the various models tend to decrease with a lengthening horizon. To understand that phenomenon we shall recall that expected returns display mean reversion and revert towards long-term means as the investment horizon increases. As a result, differences in expected returns across various models diminish. We would, therefore, expect that the impact of model uncertainty on equity markets will decrease with a lengthening horizon, as it is documented in Figure 5.

In several related studies, such as Barberis (1999), portfolio rules are derived when the current values of predictive variables are set equal to their sample means. We compute asset allocations based on the weighted model, the all-inclusive model, the highest-posterior-probability model, and the *iid* model under that scenario as well. The optimal portfolio choices are displayed in Figure 6. Asset allocations are derived for investment horizons ranging from one to ten years and relative risk-aversion coefficients of two, six, and ten. The solid lines denote allocations based on the weighted model, the dashed lines correspond to the highest-posterior-probability model, the dash/dot lines correspond to the all-inclusive model, and the lines denoted by the symbol '*' correspond to the *iid* model.[27]

Figure 6 shows that differences in optimal portfolios across models are still prominent. The differences are entirely attributed to conditional variances that differ across the forecasting models under consideration. Not surprisingly, allocations to equities based on models that include forecasting variables are very sensitive to the current values of these variables, and substantially increase when such values are set equal to their sample means.

As documented in Figure 5 — when the current values of predictive variables were equal to actual realizations — Figure 6 also shows that higher values of the relative risk aversion coefficient derive optimal portfolios to behave similarly across models. However, the horizon length appears to carry opposite implications for the role of model uncertainty on optimal portfolios. In particular, differences in portfolio allocations between the weighted and highest-posterior-probability models substantially increase with the horizon.

To explain such differences in optimal asset allocations, we essentially need to focus on predictive variances associated with each of the corresponding models. For this purpose, we take a step backward and look into equation (18). It should be noted that the 3×3 predictive variance covariance matrix of one- and two-period-ahead excess returns can be computed as:

$$\begin{aligned} \mathrm{Var}\{R_{T+1}|\Phi\} &= \Sigma_{RR}, \\ \mathrm{Var}\{R_{T+2}|\Phi\} &= 2\Sigma_{RR} + a_R'\Sigma_{zz}a_R + \Sigma_{Rz}a_R + a_R'\Sigma_{zR}. \end{aligned} \tag{25}$$

The 3×1 vector of differences between the two-period variance and two times the one-period

[27]Optimal allocations based on the weighted model across both prior-probability specifications are virtually identical.

variance of cumulative excess returns is given by:

$$\text{Diagonal}\left\{\text{Var}\{R_{T+2}|\Phi\} - 2 \times \text{Var}\{R_{T+1}|\Phi\}\right\} = \text{Diagonal}\left\{a_R' \Sigma_{zz} a_R + \Sigma_{Rz} a_R + a_R' \Sigma_{zR}\right\}, \tag{26}$$

where the notation 'Diagonal' stands for the three diagonal entries corresponding to the three size-sorted portfolios.

The first factor on the right hand side of equation (26) is a positive definite matrix, and hence its diagonal elements are positive. The diagonal elements of the remaining two factors depend on Σ_{Rz}, the correlation between contemporaneous shocks to predictive variables and stock returns. Such a correlation for the predictive variable trend-deviation-in-wealth, which governs the highest-posterior-probability model, is positive, suggesting that the annualized variance corresponding to the highest-posterior-probability model increases with the investment horizon. In contrast, the correlations for the other predictors are negative and, furthermore, the absolute values of the negative diagonal elements in the matrix $\Sigma_{Rz} a_R + a_R' \Sigma_{zR}$ exceed the positive diagonal elements in $a_R' \Sigma_{zz} a_R$. As a result, the predictive variance associated with the weighted model slightly decrease with the horizon for each size portfolio, as displayed in Figure 4. The contradicting patterns in the evolution of predictive variances corresponding to the weighted model and the highest-posterior-probability model account for the great divergence in portfolio allocations derived based on these models.

In a related study, Pastor and Stambaugh (1999) analyze optimal portfolios of investors whose prior beliefs are centered on either risk-based or characteristic-based asset pricing models. The authors show that the differences in the optimal portfolios across models are dramatically reduced by incorporating realistic margin requirements and modest uncertainty about the models' pricing abilities. In our study, investment constraints are also imposed in that short selling and buying on margin are precluded, since otherwise the expected utility would be equal to $-\infty$. As a result, the optimal portfolio choices, as shown in Figures 4 and 5, are often obtained as corner solutions in which the overall investment in equities is either zero or 99% of the wealth. If it were possible to eliminate those constraints and yet avoid the "explosion" of the expected utility, the role of model uncertainty in determining asset allocation decisions would probably be enhanced.

D Asset Allocations Across Size Portfolios

Portfolio allocations across size-sorted portfolios are computed for both prior-probability scenarios and for current values of predictive variables equal to their both sample means and actual realizations. Interestingly, with the actual recent levels, the allocation to equities is entirely attributed to small stocks, whereas investment in medium and large-size stocks is zero. Figure 7 displays asset allocations across size-sorted portfolios computed when the current values are set equal to their sample means.

The three plots on the left display total allocations to each of the size portfolios. The three plots on the middle and right display portfolio allocations to small and large capitalization firms, respectively. The optimal allocation to medium size stocks is not displayed simply because the overall investment in these stocks is zero. Within each graph, each line shows the percentage of wealth allocated to equity plotted against the investment horizon which ranges from one to ten years. The solid lines correspond to the weighted model. The dashed lines correspond to the *iid* model. Since portfolio allocations based on the weighted model are virtually identical across the two prior-probability scenarios, we plot optimal allocations to equities only for the non-equal-prior scenario.

We observe cross-sectional differences in asset allocations. First, investors avoid medium size stocks and invest instead in the risk-free Treasury bill and the other two size portfolios. Second, the allocation to large stocks increases with the horizon, whereas allocation to small stocks decreases. Conditioning on information variables, the annualized predictive standard deviation of the large stock portfolio exhibits a higher reduction in intermediate and long horizons due to predictability, making investment in large (small) stocks seem less (more) risky for risk-averse investors with an increasing horizon.

It is interesting to examine the asset allocations for the *iid* model in the presence of multiple investable risky assets. The overall investment in equities decreases with the horizon due to parameter uncertainty.[28] However, for $\gamma = 2$ the investment in large stocks displays an opposite pattern and increases. To understand this pattern, recall that in the presence of multiple risky assets, the contribution of each asset to the overall variation of the chosen portfolio is not restricted to its own variance. Rather, the covariation with the other investable assets should be accounted for. Put another way, investors care about portfolio returns, not about the behavior of any single included asset.

Moving from $\gamma = 2$ to $\gamma = 6$, we observe an interesting phenomena: The proportion of investment in large versus small stock portfolios, based on the weighted model, does not remain fixed. One might expect similar proportions due to the portfolio separation property, at least at short horizons in which the skewness and other higher-moment effect is relatively low. However, recall that portfolio weights are constrained to the unit interval. The constraints are binding since in their absence investors will sell short the portfolio of

[28]See part F of the appendix which explains the evolution of predictive variance along the horizon.

medium stocks. In turn, portfolio constraints break the separation property.

IV Conclusion

The primary objective of this study is to investigate the potential impact of model uncertainty in predictability on the investment opportunity set as perceived by a risk-averse long-horizon investor who allocates fund across three size-sorted portfolios and the risk-free Treasury bill. We have shown that model uncertainty accounts for a significant portion of the perceived variance of future stock-return, in some cases considerably larger than that attributed to parameter uncertainty or estimation risk. We have documented that with a lengthening investment horizon, the estimation risk becomes more prominent, whereas the impact of model uncertainty on the predictive variance of stock returns diminishes. The dominate fraction of the overall model uncertainty arises because investors are uncertain *ex ante* about whether asset returns are predictable. The remaining uncertainty — conditional on having predictability, which are the relevant forecasting variables — is almost entirely resolved by the realized data.

We find that model uncertainty can have a significant impact on optimal portfolio choices as well. For example, we have shown that investors who ignore the weighted model obtained by averaging over all entertained models, and consider instead a single model selected by a formal criterion perceive investment opportunities that can differ to economically significant degrees. However, differences across models are substantially reduced for long-horizon investors as well as for investors with high coefficients of relative risk-aversion.

With the calls emerging from both academics and practitioners to search for valuable predictors, we embrace the challenge by computing posterior probabilities for various data-generating processes. We measure the relative performance of predictors by Bayes factors and other model-selection criteria which are robust to some practical difficulties with interpreting the conventional P-values. We show that the posterior probability in favor of predictability exceeds the pre-assigned prior probability, suggesting that, to some extent, the data favors stock-return predictability.

Furthermore, we shed additional light on the predictive power of some *ex ante* variables. In particular, it is shown that the trend-deviation-in-wealth has a reliable power in forecasting returns on size portfolios. In contrast, the aggregate measure of book-to-market as well as the relative bill rate is found to be poor predictors of equity market returns. Lastly, the predictive power of an aggregate measure of the dividend yield, term structure slope, and default spread is found to be questionable.

Uncertainty about the model for predicting returns can be explored along several other dimensions that are not addressed in this study but present opportunities for future investigation. For example, a decision-maker who attempts to incorporate the restriction on predictability implied by rational asset pricing models might be concerned with choosing the "correct" pricing model or extracting factors that best explain the cross-sectional variation in expected stock returns.[29] Of course, having incorporated the asset pricing restriction,

[29] Asset pricing-based models include consumption models, models such as the CAPM in which the single factor is the excess return on the market portfolio, or the model proposed by Fama and French (1993), in which there are two additional factors along with the market portfolio: the difference in returns between small and large firms; and firms with high and low ratios of book-to-market. Factors can also be derived by the asymptotic principal components of Connor and Korajczyk (1986), or via factor analysis (see, for example, Chapter 6 in Campbell, Lo, and MacKinlay, 1997). Ferson and Harvey (1991), Ferson and Korajczyk (1995), and Kirby (1998) examine whether predictability in stock returns is consistent with rational asset pricing models.

investors still encounter the uncertainty in selecting the information set.

The research conducted here was confined to a relatively simple setting. The objective is to convey the basic concept of combining and selecting among models in financial economics using a Bayesian model-selection based criterion with application to predictability in asset returns. A straightforward extension is to focus on predictability in asset returns without taking into account the multi-period portfolio decisions. In that case, posterior probabilities of various models can be obtained analytically, and therefore, one does not have to resort to simulation techniques which substantially restrict the number of pre-assigned variables that can be included in the predictive regression.

The above-described extension is of great interest since the current evidence on the importance of several variables in forecasting asset returns does not reflect a concensus view. Building on our setting, one can include in a predictive regression a large number of information variables such as macroeconomic variables, liquidity variables, several lagged returns, etc. Computing posterior probabilities for all feasible linear forecasting models might shed significant light on the ability of various variables to track time variation in expected stock returns.

The analysis can also be extended by including returns on portfolios sorted on other equity characteristics or by incorporating conditional heteroscedasticity for stock returns. The nature of the underlying data-generating process might change over time, and hence, regression parameters might be treated as time-varying rather than fixed. The linearity of forecasting models can also be questioned. A challenging task would, therefore, be to compute posterior probabilities for forecasting models that depart from normality, linearity, constant variance, and constant predictive-regression parameters. Finally, the methodology developed here can be applied to other domains in financial economics where model uncertainty is suspected of playing role in deriving quantities of interest.

A Predictive Moments of the K-Period Excess Log Returns

Part A of the appendix derives the first two moments of the distribution of the K-period-ahead cumulative excess log returns conditional on the regression parameters B and Σ and the investor's data set Φ.

Partitioning equation (11) yields

$$(r_t', z_t') = (1, z_{t-1}') \begin{bmatrix} \alpha_R' & \alpha_z' \\ a_R & a_z \end{bmatrix} + \begin{pmatrix} U_{R,t} \\ U_{z,t} \end{pmatrix}', \tag{A.1}$$

where

$$\begin{pmatrix} U_{R,t} \\ U_{z,t} \end{pmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_{RR} & \Sigma_{Rz} \\ \Sigma_{zR} & \Sigma_{zz} \end{bmatrix}\right). \tag{A.2}$$

It follows from equation (A.1) that:

$$r_{T+1} = \alpha_R + a_R' z_T + U_{R,T+1}, \tag{A.3}$$

$$z_{T+1} = \alpha_z + a_z' z_T + U_{z,T+1}. \tag{A.4}$$

R_{T+K}, the cumulative excess return over the investment horizon, is computed as

$$R_{T+K} = \sum_{k=1}^{K} r_{t+k} = K\alpha_R + a_R' \left(\sum_{j=1}^{K} z_{T+j-1} \right) + \sum_{j=1}^{K} U_{R,T+j}, \tag{A.5}$$

where z_{T+J} is obtained by iterating equation (A.4):

$$z_{T+J} = [(a_z')^J - I_M](a_z' - I_M)^{-1}\alpha_z + (a_z')^J z_T + \sum_{j=1}^{J} (a_z')^{J-j} U_{z,T+j}. \tag{A.6}$$

Substituting equation (A.6) into equation (A.5) for $J = 1, \ldots, K-1$ yields:

$$\begin{aligned} R_{T+K} \;&=\; K\alpha_R + a_R' \left[a_z' \left((a_z')^{K-1} - I_M \right) (a_z' - I_M)^{-1} - (K-1) I_M \right] (a_z' - I_M)^{-1} \alpha_z \\ &+\; a_R' \left((a_z')^K - I_M \right) (a_z' - I_M)^{-1} z_T + \sum_{j=2}^{K} \sum_{i=1}^{j-1} a_R' (a_z')^{j-i-1} U_{z,T+i} + \sum_{j=1}^{K} U_{R,T+j}, \end{aligned} \tag{A.7}$$

for $K \geq 2$. The results follow immediately.

B The Decomposition of the Predictive Variance of Cumulative Stock Returns

Decomposing the predictive variance $\mathrm{Var}\{R_{T+K}|\Phi\}$ with respect to the model space and using the law of iterated expectation, we obtain

$$\mathrm{Var}\{R_{T+K}|\Phi\} = \sum_{j=1}^{J} P\left(\mathcal{M}_j|\Phi\right)\left[\mathrm{Var}\{R_{T+K}|\mathcal{M}_j,\Phi\} + (\tilde{\lambda} - \mathrm{E}_\Theta\{\lambda_{\mathcal{M}_j}\})(\tilde{\lambda} - \mathrm{E}_\Theta\{\lambda_{\mathcal{M}_j}\})'\right], \tag{B.1}$$

where $\tilde{\lambda} = \sum_{j=1}^{J} P\left(\mathcal{M}_j|\Phi\right)\mathrm{E}_\Theta\{\lambda_{\mathcal{M}_j}\}$ and E_Θ is the expected value operator taken with respect to the parameter space.

The total uncertainty about future returns is composed of the within-model (the first factor on the right-hand side of equation B.1) and cross-model (the second factor) uncertainty. The former is merely a mixture of variances from each candidate model. The latter reflects the model uncertainty in forecasting stock returns, in that it measures the overall variance of future returns attributable to the uncertainty about which of the forecasting models investors should use.

Estimation risk or parameter uncertainty is an additional source of the *ex ante* stock-return variance. The within-model parameter uncertainty can be obtained by decomposing the within-model variance as follows:

$$\begin{aligned} &\mathrm{Var}\{R_{T+K}|\mathcal{M}_j,\Phi\} && \text{(B.2)} \\ &= E_\Theta\{\mathrm{Var}\left[R_{T+K}|\mathcal{M}_j,\Phi,B,\Sigma\right]\} + \mathrm{Var}_\Theta\{E\left[R_{T+K}|\mathcal{M}_j,\Phi,B,\Sigma\right]\}, \\ &= E_\Theta\{\Upsilon_{\mathcal{M}_j}\} + \mathrm{Var}_\Theta\{\lambda_{\mathcal{M}_j}\}. \end{aligned}$$

The result follows immediately by substituting equation (B.2) into the first factor on the right-hand side of equation (B.1).

C The Posterior Distribution of Regression Parameters

This part of the appendix derives the full conditional pdfs of the regression parameters $P\left(B|\Sigma,\tilde{\beta},C,\Phi\right)$ and $P\left(\Sigma|B,H_0,\nu_0,\Phi\right)$.

Multiplying the joint prior density (22) by the likelihood function (16) yields:

$$\begin{aligned} & P\left(B,\Sigma|\tilde{\beta},C,H_0,\nu_0,\Phi\right) \\ \propto \ & |\Sigma|^{-\frac{\nu_0+T+N+M+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[S+(B-\hat{B})'X'X(B-\hat{B})+H_0\right]\Sigma^{-1}\right) \\ \times \ & \exp\left(-\frac{1}{2}(\beta-\tilde{\beta})'C^{-1}(\beta-\tilde{\beta})\right). \end{aligned} \tag{C.1}$$

The conditional posterior pdf for Σ is easily obtained as:

$$\Sigma|B,H_0,\nu_0,\Phi \sim W^{-1}\left[\Psi,\nu\right], \tag{C.2}$$

where

$$\begin{aligned} \Psi &= H_0+(Y-XB)'(Y-XB), \\ \nu &= \nu_0+T. \end{aligned} \tag{C.3}$$

Following Zellner (1971) p. 227, equation (C.1) can be rewritten as:

$$\begin{aligned} & P\left(B,\Sigma|\tilde{\beta},C,H_0,\nu_0,\Phi\right) \\ \propto \ & |\Sigma|^{-\frac{\nu_0+T+N+M+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left[(S+H_0)\Sigma^{-1}\right]-\frac{1}{2}\left[(\beta-\hat{\beta})'\Sigma^{-1}\otimes X'X(\beta-\hat{\beta})\right]\right) \\ \times \ & \exp\left(-\frac{1}{2}(\beta-\tilde{\beta})'C^{-1}(\beta-\tilde{\beta})\right), \end{aligned} \tag{C.4}$$

where $\hat{\beta}=\mathrm{vec}(\hat{B})$.

By completing the square on β it follows that the conditional posterior pdf for β given Σ is multivariate normal with mean E and variance F which are given by

$$\begin{aligned} E &= \left(\Sigma^{-1}\otimes X'X+C^{-1}\right)^{-1}\left[(\Sigma^{-1}\otimes X'X)\hat{\beta}+C^{-1}\tilde{\beta}\right], \\ F &= \left(\Sigma^{-1}\otimes X'X+C^{-1}\right)^{-1}. \end{aligned} \tag{C.5}$$

At this stage we can apply the Gibbs sampler technique. A Gibbs sampling chain is formed as follows (the notational dependence on prior parameters is suppressed):

1. Specify starting values $B^{(0)}$ $\Sigma^{(0)}$ and set $i = 1$.

2. Draw from the full conditional distributions:

   - Draw $B^{(i)}$ from the conditional pdf $P(B|\Sigma^{(i-1)}, \Phi)$
   - Draw $\Sigma^{(i)}$ from the conditional pdf $P(\Sigma|B^{(i)}, \Phi)$

3. Set $i = i + 1$ and go to step 2.

After m iterations the sample $B^{(m)}, \Sigma^{(m)}$ is obtained. Under mild regularity conditions (see, for example, Tierney ,1994), the pairs $(B^{(m)}, \Sigma^{(m)})$ converges in distribution to the relevant marginal and joint distributions. That is, $P(B^{(m)}|\Phi) \to P(B|\Phi)$, $P(\Sigma^{(m)}|\Phi) \to P(\Sigma|\Phi)$, and $P(B^{(m)}, \Sigma^{(m)}|\Phi) \to P(B, \Sigma|\Phi)$. When m is large enough, the G values $(B^{(g)}, \Sigma^{(g)})_{g=m+1}^{m+G}$ are a sample from the joint posterior.

D Drawing from the Predictive Distribution of Long-Horizon Cumulative Excess Log Returns

The predictive density of the K-period cumulative excess log returns that incorporates model uncertainty and parameter uncertainty within any forecasting model belonging to $\mathcal{M}$ is given by (the prior-specific parameters are suppressed):

$$P\left(R_{T+K}|\Phi\right) = \sum_{j=1}^{J} \int_B \int_\Sigma P(\mathcal{M}_j|\Phi) P\left(R_{T+K}|B, \Sigma, \mathcal{M}_j, \Phi\right) P\left(B, \Sigma|\mathcal{M}_j, \Phi\right) d\Sigma dB.$$

The joint posterior pdf, $P\left(B, \Sigma|\mathcal{M}_j, \Phi\right)$, has been analyzed in part C of the appendix. The distribution of future excess returns given the regression parameters and the data $P\left(R_{T+K}|B, \Sigma, \mathcal{M}_j, \Phi\right)$ has been shown to obey the multivariate normal distribution whose mean and variance are displayed in equation (18).

Sampling from the predictive distribution of excess cumulative future returns is obtained by first drawing a model belonging to the set $\mathcal{M}$. Conditional upon the selected model, B and Σ are drawn from the joint posterior distribution using the Gibbs sampler algorithm, as explained in part C of the appendix. Given B and Σ, we generate an $N \times 1$ random vector from the conditional distribution $P\left(R_{T+K}|B, \Sigma, \Phi\right)$. Repeating this three-stage algorithm yields a random sample from the exact weighted predictive distribution.

E Computing Marginal Likelihood and Posterior Probabilities

In deriving the marginal likelihood, we closely follow the approach advocated by Chib (1995). First, according to Bayes rule, the marginal likelihood function can be expressed as (the notational dependence on prior parameters is suppressed):

$$P(Y, x_0) = \frac{P\left(Y, x_0|\Sigma, B\right) P\left(\Sigma, B\right)}{P\left(\Sigma, B|Y, x_0\right)}. \tag{E.1}$$

In our setting, the initial observation is assumed to be fixed and therefore $P\left(Y, x_0|\Sigma, B\right) = P\left(Y|\Sigma, B, x_0\right)$. Note that $P(Y, x_0)$ does not depend on model parameters. Therefore, for any (Σ, B) belonging to the parameter space, Θ, (take without loss of generality $\Sigma^*, B^* \in \Theta$), the proposed estimate of the marginal likelihood, $P(Y, x_0)$, satisfies:

$$\ln\{\hat{P}(Y, x_0)\} = \ln P\left(Y|\Sigma^*, B^*, x_0\right) + \ln P\left(\Sigma^*, B^*\right) - \ln P\left(\Sigma^*, B^*|Y, x_0\right). \tag{E.2}$$

where $P\left(Y|\Sigma^*, B^*, x_0\right)$, $P\left(\Sigma^*, B^*\right)$, and $P\left(\Sigma^*, B^*|Y, x_0\right)$ are the likelihood function, the joint prior, and the joint posterior estimated at Σ^* and B^*, respectively.

At this stage most of the quantities needed to derive the marginal likelihood can be evaluated. In particular, equations (22) and (16) specify the joint prior distribution and likelihood function, respectively. The conditional posterior pdf for β given Σ is multivariate normal with mean E and variance F. It remains to evaluate the marginal pdf for Σ at Σ^*. We elaborate on this step below.

We denote the output from the Gibbs algorithm, as discussed in part C of the appendix, by $\{B^{(g)}, \Sigma^{(g)}\}_{g=m+1}^{m+G}$. Integrating out B from the joint posterior pdf yields

$$P(\Sigma|Y) = \int_B P(\Sigma|Y, B) P(B|Y) \, dB. \tag{E.3}$$

A Monte Carlo estimate of $P(\Sigma|Y)$ at Σ^* is obtained by

$$\hat{P}(\Sigma^*|Y) = \sum_{g=1}^{G} \left(\Sigma^*|Y, B^{(g)}\right). \tag{E.4}$$

This technique is known in the literature as the Rao-Blackwellization.[30] Under mild regularity conditions $\hat{P}(\Sigma^*|Y) \overset{a.s.}{\to} P(\Sigma^*|Y)$ as $G \to \infty$. (The notation a.s. denotes an almost sure convergence.)

The marginal likelihood of the data will be utilized in computing Bayes factors for hypotheses testing and calculating posterior probabilities, according to which models are weighted. In the following, we provide all the quantities used in calculating the marginal likelihood. The column-wise vectorization of B^* is denoted by β^*. For notational convenience, the sum $N + M$ is denoted by $\overline{N}$.

A The Prior Pdfs for the Regression Parameters:

$$P\left(\beta^*|\tilde{\beta}, C\right) = (2\pi)^{-\frac{\overline{N}(M+1)}{2}} |C|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\beta^* - \tilde{\beta})' C^{-1} (\beta^* - \tilde{\beta})\right),$$

$$P(\Sigma^*|H_0, \nu_0) = \left(2^{\frac{\nu_0 \overline{N}}{2}} \pi^{\frac{\overline{N}(\overline{N}-1)}{4}} \prod_{i=1}^{\overline{N}} \Gamma\left[\tfrac{\nu_0+1-i}{2}\right]\right)^{-1} |H_0|^{\frac{\nu_0}{2}} |\Sigma^*|^{-\frac{\nu_0+\overline{N}+1}{2}} \exp\left(-\tfrac{1}{2}\mathrm{tr} H_0 \Sigma^{*-1}\right).$$

B The Likelihood Function of the Matrix Y:

$$P(Y|B^*, \Sigma^*, x_0) = (2\pi)^{-\frac{\overline{N}T}{2}} |\Sigma^*|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}(Y - XB^*)'(Y - XB^*)\Sigma^{*-1}\right).$$

[30] See, for example, Gelfand and Smith (1990).

C The Full Conditional Posterior Pdfs for the Regression Parameters:

$$P\left(\beta^*|\Sigma^*,\tilde{\beta},C,\Phi\right) = (2\pi)^{-\frac{\overline{N}(M+1)}{2}}|F^*|^{-\frac{1}{2}}\exp\left(-\tfrac{1}{2}(\beta^*-E^*)'F^{*-1}(\beta^*-E^*)\right),$$

$$P\left(\Sigma^*|B^{(g)},H_0,\nu_0,\Phi\right) = \left(2^{\frac{\nu\overline{N}}{2}}\pi^{\frac{\overline{N}(\overline{N}-1)}{4}}\prod_{i=1}^{\overline{N}}\Gamma\left[\tfrac{\nu+1-i}{2}\right]\right)^{-1}|\Psi^{(g)}|^{\frac{\nu}{2}}|\Sigma^*|^{-\frac{\nu+\overline{N}+1}{2}}\exp\left(-\tfrac{1}{2}\mathrm{tr}\Psi^{(g)}\Sigma^{*-1}\right),$$

where

$$\begin{aligned} E^* &= \left(\Sigma^{*-1}\otimes X'X + C^{-1}\right)^{-1}\left[(\Sigma^{*-1}\otimes X'X)\hat{\beta} + C^{-1}\tilde{\beta}\right], \\ F^* &= \left(\Sigma^{*-1}\otimes X'X + C^{-1}\right)^{-1}, \\ \Psi^{(g)} &= H_0 + (Y - XB^{(g)})'(Y - XB^{(g)}). \end{aligned}$$

Recall that g denotes the output from the Gibbs algorithm. The estimate of the marginal likelihood is obtained as:

$$\begin{aligned} \ln\left\{\hat{P}(Y,x_0)\right\} &= \ln\left\{P\left(Y|\Sigma^*,B^*,x_0\right)\right\} + \ln\left\{P\left(\beta^*|\tilde{\beta},C\right)\right\} + \ln\left\{P\left(\Sigma^*|H_0,\nu_0\right)\right\} \\ &- \ln\left\{P\left(\beta^*|\Sigma^*,C,\tilde{\beta},Y,x_0\right)\right\} - \ln\left\{\frac{1}{G}\sum_{g=1}^{G}P\left(\Sigma^*|B^{(g)},H_0,\nu_0,Y,x_0\right)\right\}. \end{aligned}$$

An estimate for the Bayes factor of any two models $\mathcal{M}_i$ and $\mathcal{M}_j$ is obtained by

$$\hat{\mathcal{B}}_{i,j} = \exp\left\{\ln(\hat{P}(Y,x_0|\mathcal{M}_i) - \ln(\hat{P}(Y,x_0|\mathcal{M}_j)\right\}, \qquad \text{(E.5)}$$

and the posterior probability of $\mathcal{M}_i$ is estimated as:

$$\hat{P}(\mathcal{M}_i|\Phi) = \frac{\hat{P}(Y,x_0|\mathcal{M}_i)P(\mathcal{M}_i)}{\sum_{j=1}^{J}\hat{P}(Y,x_0|\mathcal{M}_j)P(\mathcal{M}_j)} = \frac{P(\mathcal{M}_i)}{\sum_{j=1}^{J}\hat{\mathcal{B}}_{j,i}P(\mathcal{M}_j)}. \qquad \text{(E.6)}$$

F The Predictive Distribution Associated with Independently and Identically Distributed Stock Returns

Part F of the appendix derives the predictive distribution associated with the *iid* model of stock returns taking the form $r_t = \mu + \epsilon_t$, where prior beliefs about the regression parameters are non-informative:

$$P(\mu, \sigma_\epsilon^2) \propto (\sigma_\epsilon^2)^{-1}, \tag{F.1}$$

and the likelihood function of excess returns is proportional to

$$(\sigma_\epsilon^2)^{-\frac{T}{2}} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\left[(r-\mu\iota)'(r-\mu\iota)\right]\right), \tag{F.2}$$

with ι a $T \times 1$ vector of ones and $r = [r_1, r_2, \dots, r_T]'$.

The posterior distribution is obtained by multiplying the non-informative prior (F.1) by the likelihood function (F.2) which results in the normal-inverted gamma posterior:

$$\mu|\Phi \sim N\left(\hat{\mu}, \frac{\sigma_\epsilon^2}{T}\right), \tag{F.3}$$
$$\frac{\psi}{\sigma_\epsilon^2}|\Phi \sim \chi^2_{\tilde{\nu}},$$

where $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} r_t$, $\psi = \sum_{t=1}^{T} r_t^2 + T\hat{\mu}^2$, and $\tilde{\nu} = T-1$.

The predictive distribution integrates the conditional distribution of future cumulative stock returns over the joint posterior density of the parameters μ and σ_ϵ, which, in turn, summarizes all the uncertainty about those parameters after observing the data. Specifically, the predictive distribution is obtained as:

$$\begin{aligned}
P\left(R_{T+K}|\Phi\right) &= \int_{\mu,\sigma_\epsilon^2} P\left(R_{T+K}|\mu,\sigma_\epsilon^2,\Phi\right) P\left(\mu,\sigma_\epsilon^2|\Phi\right) d\sigma_\epsilon^2 d\mu, \\
&\propto \int_{\mu,\sigma_\epsilon^2} \sigma_\epsilon^{-(T+1)} \exp\left(-\frac{1}{2\sigma_\epsilon^2}\left[\psi + \frac{(R_{T+K}-K\mu)^2}{K} + T(\mu-\hat{\mu})^2\right]\right) d\sigma_\epsilon^2 d\mu, \\
&\propto \int_{\mu} \left|\left[\psi + \frac{(R_{T+K}-K\mu)^2}{K} + T(\mu-\hat{\mu})^2\right]\right|^{-\frac{T}{2}} d\mu \\
&\propto \int_{\mu} \left|(\mu-\tilde{\mu})^2(T+K) + \frac{R_{T+K}^2}{K} + \psi - \tilde{\mu}^2 K\right|^{-\frac{T}{2}} d\mu,
\end{aligned}$$

where

$$\tilde{\mu} = (T+K)^{-1}\left[R_{T+K} + \hat{\mu}T\right].$$

To integrate out σ_ϵ^2, we used the properties of the inverted gamma probability density function.

Integrating with respect to μ and completing the square on R_{T+K}, we show that the predictive distribution under the *iid* model obeys the form of a Student-t distribution:

$$\frac{\Gamma[\frac{\tilde{\nu}+1}{2}]}{\Gamma[\frac{\tilde{\nu}}{2}]\Gamma[\frac{1}{2}]}\sqrt{\frac{F}{\tilde{\nu}}}\left[1+\frac{F}{\tilde{\nu}}\left(R_{T+K}-\tilde{R}_{T+K}\right)^2\right]^{-\frac{\tilde{\nu}+1}{2}}, \tag{F.4}$$

where

$$\tilde{R}_{T+K} = K\hat{\mu},$$
$$F = \frac{T\tilde{\nu}}{K(K+T)\psi}.$$

Using the properties of the Student-t probability density function, it can be shown that the first two moments of the cumulative excess future return given the available data at time T are:

$$\begin{aligned} \mathrm{E}\{R_{T+K}|\Phi\} &= \tilde{R}_{T+K}, \\ \mathrm{Var}\{R_{T+K}|\Phi\} &= \frac{\psi}{(\tilde{\nu}-2)}\frac{K(K+T)}{T}. \end{aligned} \tag{F.5}$$

The reader may note that the annualized variance of cumulative returns is simply equal to the sample variance scaled by the factor $\frac{T+K}{T}$. That factor includes a component at the magnitude of $\frac{K}{T}$ corresponding to parameter uncertainty. That is, in the presence of parameter uncertainty, the annualized predictive variance increases linearly with the investment horizon.

Table I
Descriptive Statistics of Predictive Variables and Excess Returns

The table shows descriptive statistics for the predictive variables. The statistics are computed based on 184 quarterly observations spanning the first quarter of 1953 to the last quarter of 1998. The book-to-market (BM) is computed as the sum of fiscal year-end per share book-values of all stocks included in the S&P Industrials divided by the contemporaneous level of the S&P Industrials price index. Dividend yield (Div) is the total payments of dividends on the value-weighted NYSE portfolio over the recent twelve months divided by the contemporaneous level of the index. The default spread (Def) is the difference in annualized yields of Moody's Baa and Aaa rated bonds. The trend-deviation-in-wealth (TDW) is computed as $c_t - wa_t - (1-w)y_t$, where c_t, a_t, and y_t denote log of consumption, non-human wealth, and labor income, respectively. The weight w equals the average share of non-human wealth in total wealth. TDW is in per-capita terms, measured in 1992 dollars. The relative bill rate (Tbill) is the difference between an annual yield on a three-month Treasury bill and its one-year backward-moving average. Lastly, the term structure slope (Term) is the difference in annualized yield of ten-year and one-year Treasury bills. The parameter ρ_t is the sample autocorrelation at lag t months. The bottom part of the table exhibits the slope coefficients in the multivariate regression of size portfolios on lagged instruments. We report slope coefficients of those regressions. Standard deviations are displayed in parenthesis below the slopes.

	Predictive Variables					
Statistic	BM	Div	Def	TDW	Tbill	Term
Means	0.5078	0.0363	0.9488	-0.0040	0.0012	0.0073
Standard Deviations	0.1790	0.0094	0.4436	0.0112	0.0203	0.0098
Contemporaneous Correlation with						
BM	1					
Div	0.9075	1				
Def	0.4800	0.5089	1			
TDW	0.2293	0.2864	0.1500	1		
Tbill	0.0143	0.0196	-0.2115	-0.1575	1	
Term	-0.2118	-0.0115	0.1309	0.3309	-0.4262	1
Autocorrelations:						
ρ_3	0.9683	0.9474	0.9098	0.8325	0.4402	0.8549
ρ_6	0.9261	0.8715	0.8386	0.6554	-0.0295	0.7458
ρ_9	0.8853	0.8056	0.7763	0.4879	-0.0742	0.6455
ρ_{12}	0.8545	0.7467	0.7077	0.3494	-0.0935	0.5370
ρ_{60}	0.5293	0.3260	0.3904	0.0159	0.0638	0.0151
Regression Coefficients and their Standard Deviations						
Small	0.0369	0.8348	0.0060	1.6391	-0.4459	1.5019
	(0.1393)	(2.6546)	(0.0276)	(1.0239)	(0.5556)	(1.2761)
Mid	0.0182	1.6970	-0.0065	1.4721	-0.3170	1.6226
	(0.1162)	(2.2146)	(0.0230)	(0.8542)	(0.4635)	(1.0646)
Large	-0.0301	2.2122	-0.0051	0.9996	-0.0956	1.6289
	(0.0899)	(1.7125)	(0.0178)	(0.6605)	(0.3584)	(0.8232)

Table II

The Highest-Posterior-Probability Models when Investors Consider the Events of Predictability versus no Predictability as Equally Likely

The table shows estimated posterior probabilities for six non-zero-probability return forecasting models computed for three prior specifications corresponding to training samples that include 45, 60, and 75 quarterly observations. (The training-sample length is denoted by the notation T_0.) A posterior probability is estimated as:

$$\hat{P}\left(\mathcal{M}_i|\Phi\right) = \frac{\hat{P}\left(Y, x_0|\mathcal{M}_i\right) P(\mathcal{M}_i)}{\sum_{j=1}^{J} \hat{P}\left(Y, x_0|\mathcal{M}_j\right) P(\mathcal{M}_j)},$$

where $\hat{P}\left(Y, x_0|\mathcal{M}_i\right)$ denotes the estimated marginal likelihood of the data conditioning upon $\mathcal{M}_i$, and $P(\mathcal{M}_i)$ denotes the pre-assigned prior probability. Assuming that (i) investors consider predictability or lack thereof as equally likely *ex ante*, and (ii) conditional on having predictability, investors exhibit non-informative beliefs about the form that it takes, we associate a prior probability of 50% to the *iid* model and assign equally the remaining 50% to all other models.

The forecasting models presented below are distinguished by a unique combination of zeros and ones designating exclusions and inclusions of predictive variables from the model, respectively. For example, the first row corresponds to a model that retains only the trend-deviation-in-wealth and discards the remaining predictors. The last column displays posterior probabilities obtained as simple averages of posterior probabilities computed for the three training-sample lengths. $Total^1$ counts the appearance and $Total^2$ computes the cumulative average probability for each predictive variable. To illustrate, Div appears in 2 combinations whose cumulative average posterior probability is 27.03%.

	Predictive Variables						Probability (%)			
Model	BM	Div	Def	TDW	Tbill	Term	$T_0 = 45$	$T_0 = 60$	$T_0 = 75$	**Average**
1	0	0	0	1	0	0	11.03	65.23	46.64	42.25
2	0	1	0	1	0	0	47.04	1.67	0.03	15.13
3	0	0	0	0	0	0	1.77	6.37	36.56	15.10
4	0	0	0	1	0	1	13.81	16.12	8.67	12.90
5	0	1	0	0	0	1	26.22	9.09	1.73	11.90
6	0	0	1	1	0	0	0.13	1.52	6.37	2.72
$Total^1$	0	2	1	4	0	2	100.00	100.00	100.00	100.00
$Total^2$	0%	27.03%	2.72%	73.00%	0.00%	24.80%				

Table III

The Highest-Posterior-Probability Models when Investors Allocate Prior Probabilities Equally Across Models

The table shows estimated posterior probabilities for five non-zero-probability return forecasting models computed for three prior specifications corresponding to training samples that include 45, 60, and 75 quarterly observations. (The training-sample length is denoted by the notation T_0.) A posterior probability is estimated as:

$$\hat{P}\left(\mathcal{M}_i|\Phi\right) = \frac{\hat{P}\left(Y, x_0|\mathcal{M}_i\right) P(\mathcal{M}_i)}{\sum_{j=1}^{J} \hat{P}\left(Y, x_0|\mathcal{M}_j\right) P(\mathcal{M}_j)},$$

where $\hat{P}\left(Y, x_0|\mathcal{M}_i\right)$ denotes the estimated marginal likelihood of the data conditioning upon $\mathcal{M}_i$ and $P(\mathcal{M}_i)$ denotes the pre-assigned prior probability. Assuming that investors consider the various return-forecasting models as equally likely *ex ante*, we allocate prior probabilities equally across models.

The forecasting models presented below are distinguished by a unique combination of zeros and ones designating exclusions and inclusions of predictive variables from the model, respectively. For example, the first row corresponds to a model that retains only the trend-deviation-in-wealth and discards the remaining predictors. The last column displays posterior probabilities obtained as simple averages of posterior probabilities computed for the three training-sample lengths. $Total^1$ counts the appearance and $Total^2$ computes the cumulative average probability for each predictive variable. To illustrate, Div appears in 2 combinations whose cumulative average posterior probability is 28.55%.

	Predictive Variables						Probability (%)			
Model	BM	Div	Def	TDW	Tbill	Term	$T_0 = 45$	$T_0 = 60$	$T_0 = 75$	**Average**
1	0	0	0	1	0	0	11.23	69.67	73.52	52.50
2	0	1	0	1	0	0	47.89	1.78	0.05	15.75
3	0	0	0	1	0	1	14.05	17.22	13.67	15.08
4	0	1	0	0	0	1	26.70	9.70	2.72	12.80
5	0	0	1	1	0	0	0.13	1.63	10.04	3.87
$Total^1$	0	2	1	4	0	2	100.00	100.00	100.00	100.00
$Total^2$	0%	28.55%	3.87%	87.20%	0.00%	27.88%				

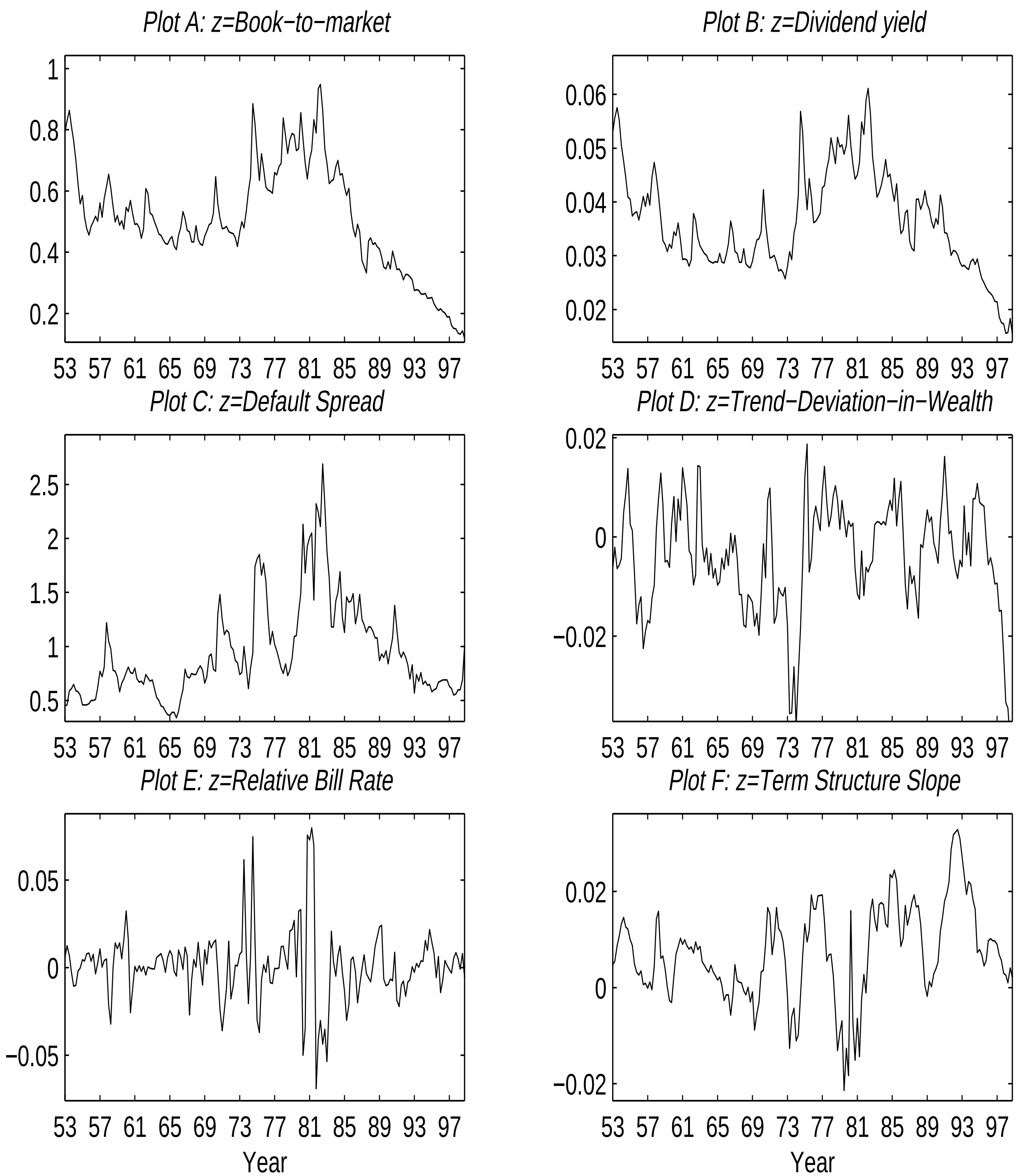

Figure 1 - The Evolution of the Predictive Variables Over the 1953-1998 Sample Period.
Figure 1 plots book-to-market, dividend yield, default spread, trend-deviation-in-wealth, relative Treasury bill, and term structure slope for quarterly observations spanning the first quarter of 1953 through the fourth quarter of 1998.

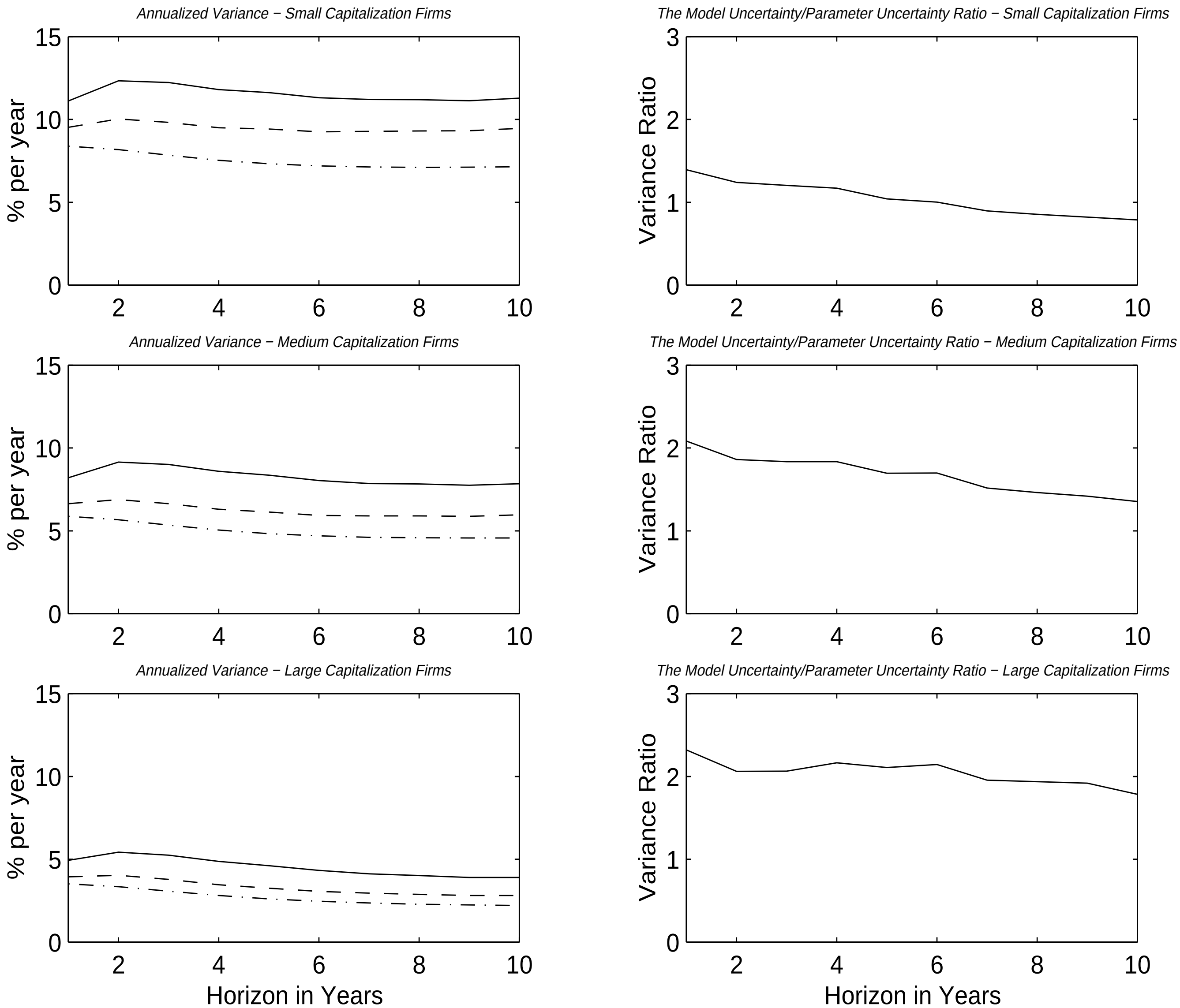

Figure 2 - The Decomposition of Future Stock-Return Variances Plotted against the Investment Horizon in Years when Investors Consider the Events of Predictability versus no Predictability as Equally Likely.

The three plots on the left display the decompositions of annualized variances of small, medium, and large capitalization firms into three components, including the cross-model uncertainty and the mixture of within-model parameter uncertainty. The solid lines show the annualized variances computed based on the weighted model. The dashed lines correspond to the variances that include the within-model parameter uncertainty, but not the cross-model uncertainty. The dash/dot lines display a mixture of variances from any candidate model such that within each model, the variances are computed as though the model-specific parameters were known. The difference between the solid and dashed lines constitutes the model-uncertainty component, and the difference between the dashed and dash/dot lines constitutes the mixture of parameter uncertainty component. The three plots on the right exhibit ratios of variances obtained by dividing the model-uncertainty component by the parameter-uncertainty counterpart for each of the three capitalization firms.

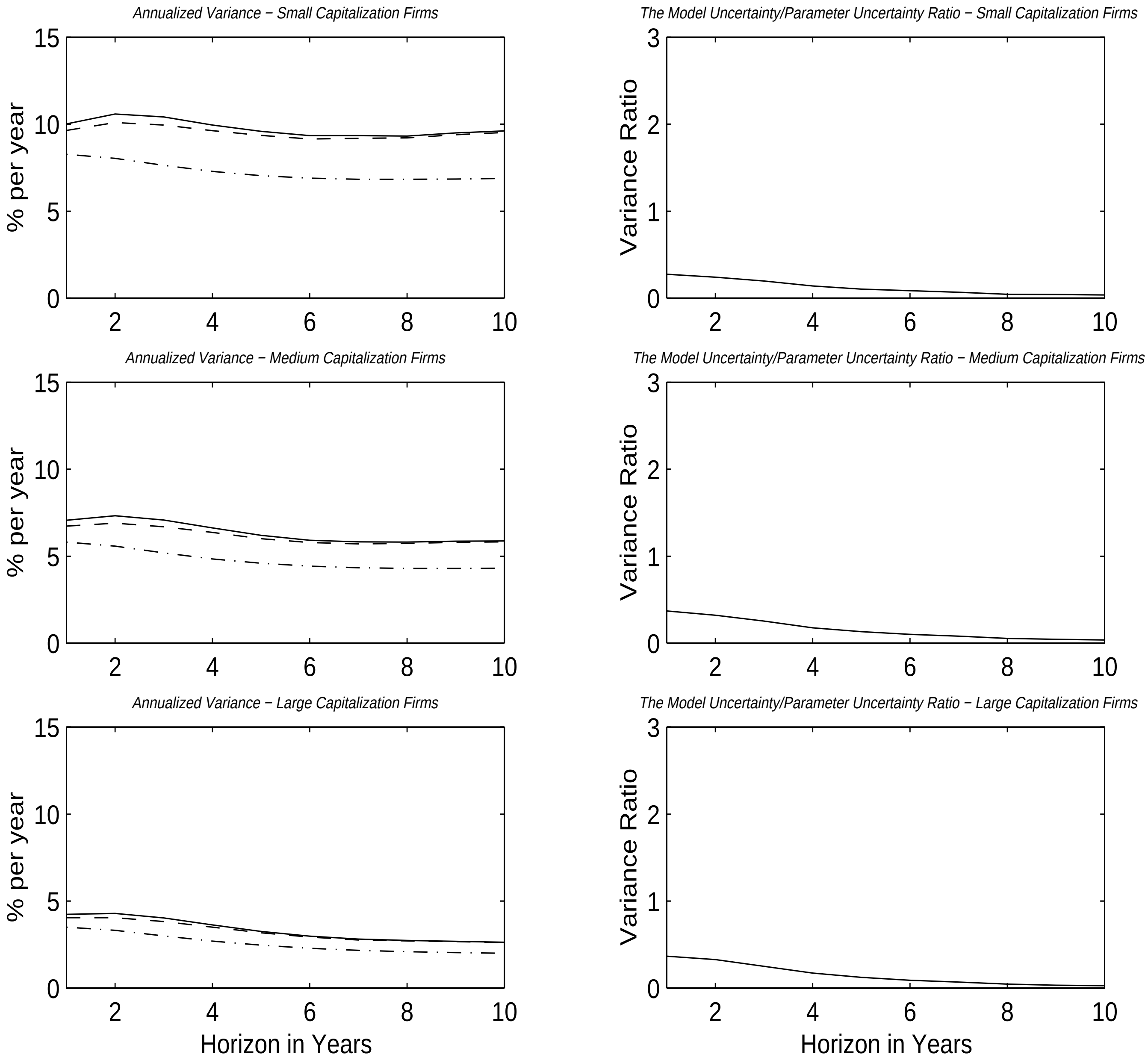

Figure 3 - The Decomposition of Future Stock-Return Variances Plotted against the Investment Horizon in Years when Investors Allocate Prior Probabilities Equally Across Models.

The three plots on the left display the decompositions of annualized variances of small, medium, and large capitalization firms into three components, including the cross-model uncertainty and the mixture of within-model parameter uncertainty. The solid lines show the annualized variances computed based on the weighted model. The dashed lines correspond to the variances that include the within-model parameter uncertainty, but not the cross-model uncertainty. The dash/dot lines display a mixture of variances from any candidate model such that within each model, the variances are computed as though the model-specific parameters were known. The difference between the solid and dashed lines constitutes the model-uncertainty component, and the difference between the dashed and dash/dot lines constitutes the mixture of parameter uncertainty component. The three plots on the right exhibit ratios of variances obtained by dividing the model-uncertainty component by the parameter-uncertainty counterpart for each of the three capitalization firms.

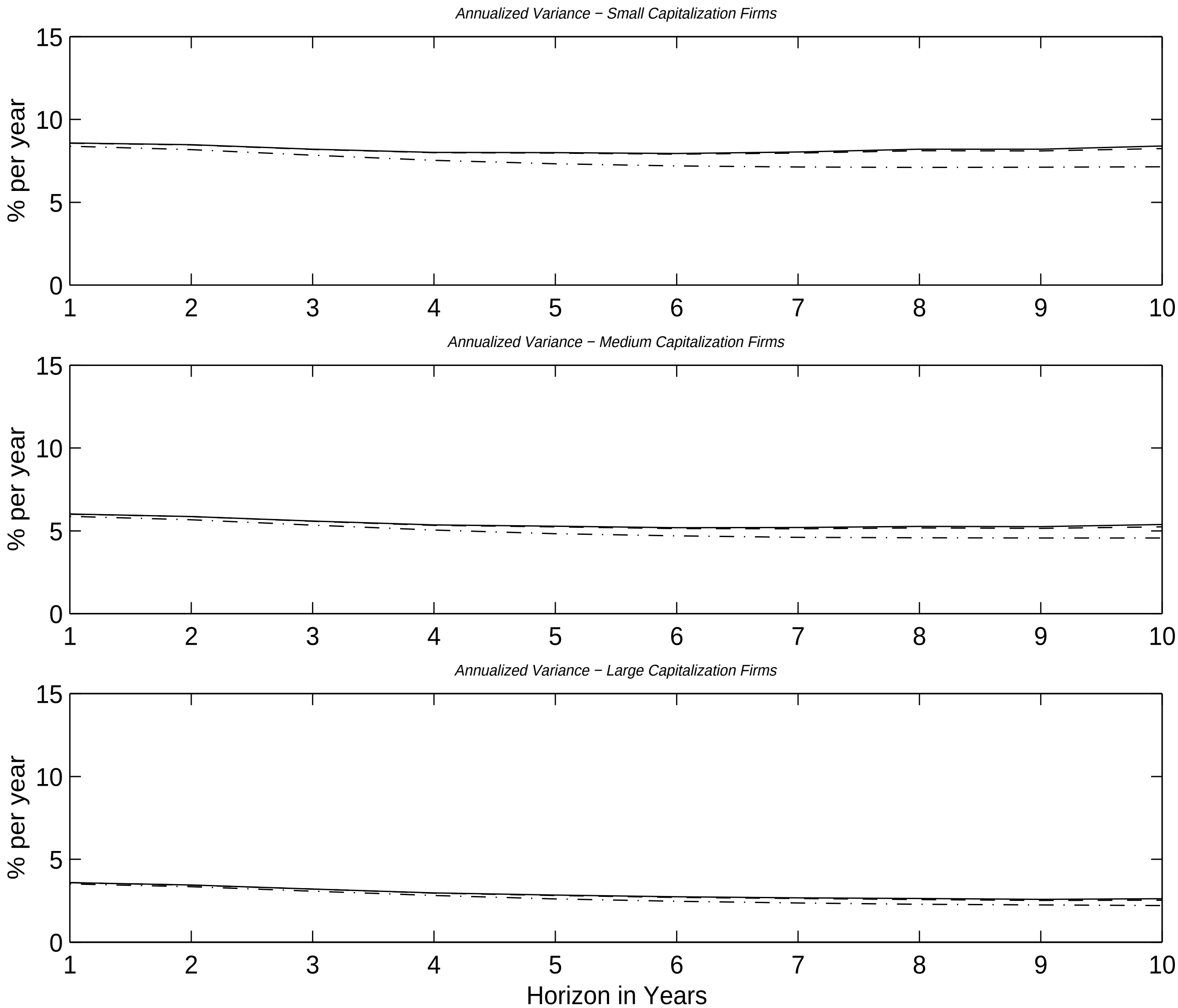

Figure 4 - The Decomposition of Future Stock-Return Variances Plotted against the Investment Horizon in Years when the Current Values of the Predictive Variables are Set Equal to their Sample Means.

The plots display the decompositions of annualized variances of small, medium, and large capitalization firms into three components, including the cross-model uncertainty and the mixture of within-model parameter uncertainty. The predictive annualized variance is computed when the current values of the predictive variables are set equal to their sample means. The solid lines show the annualized variances computed based on the weighted model. The dashed lines correspond to the variances that include the within-model parameter uncertainty, but not the cross-model uncertainty. The dash/dot lines display a mixture of variances from any candidate model such that within each model, the variances are computed as though the model-specific parameters were known. The difference between the solid and dashed lines constitutes the model-uncertainty component, and the difference between the dashed and dash/dot lines constitutes the mixture of parameter uncertainty component.

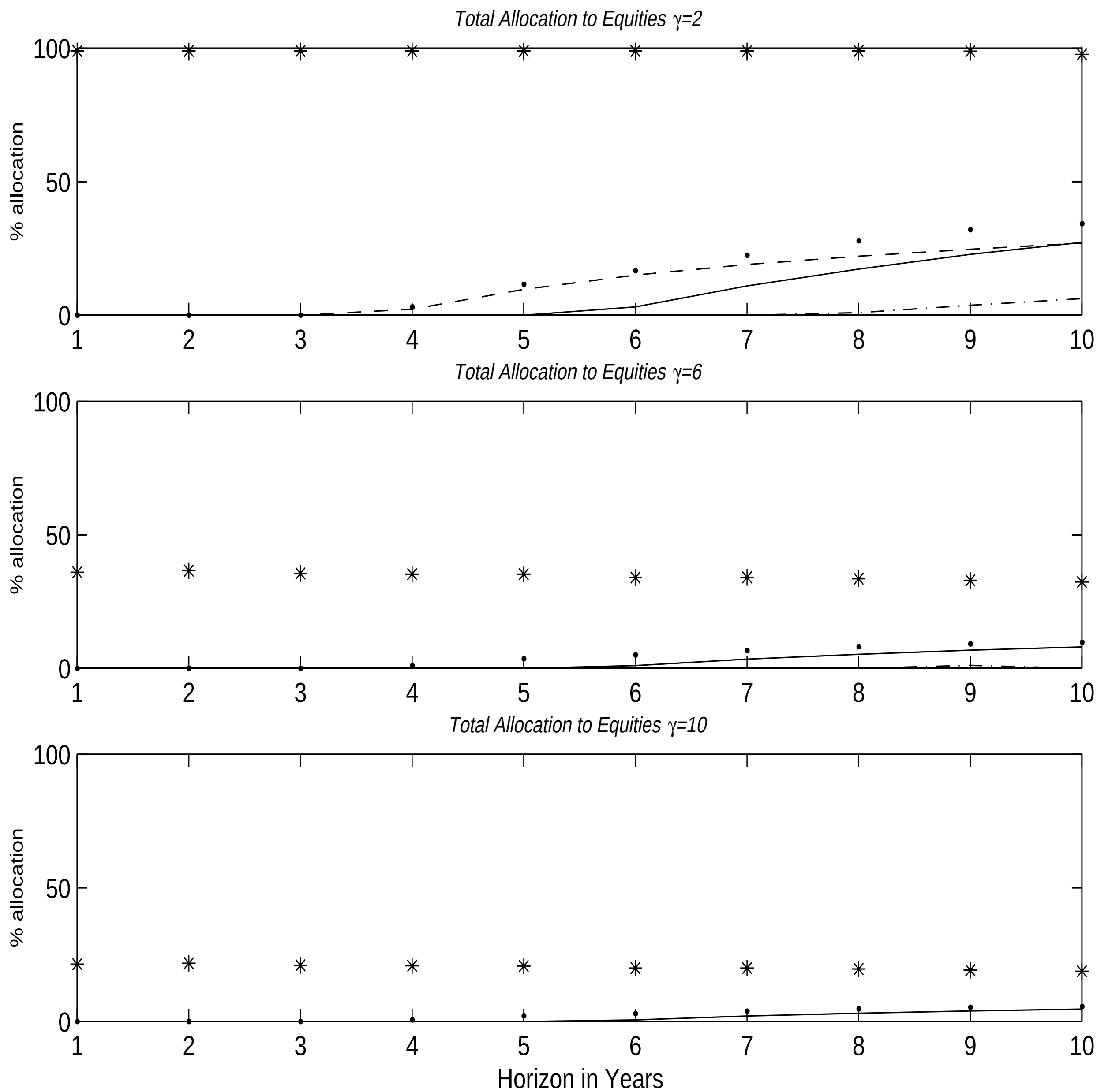

Figure 5 - Total Asset Allocations to Equities Based on Different Models Plotted against the Investment Horizon in Years for an Investor Following a Buy-and-Hold Strategy with a Power Utility Function and a Relative Risk-Aversion Coefficient Equal to γ. The Current Values of the Predictive Variables are Equal to Actual Realizations. Figure 5 displays total allocations to equities based on the weighted model, for the equal-prior-probability scenario (solid lines) and the non-equal-prior scenario (dotted lines), the all-inclusive model (dash/dot lines), the highest-posterior-probability model (dashed lines), and the *iid* model (the lines denoted by the symbol '*'). Asset allocations are computed for investment horizons ranging from one to ten years and relative risk-aversion coefficients of two, six, and ten.

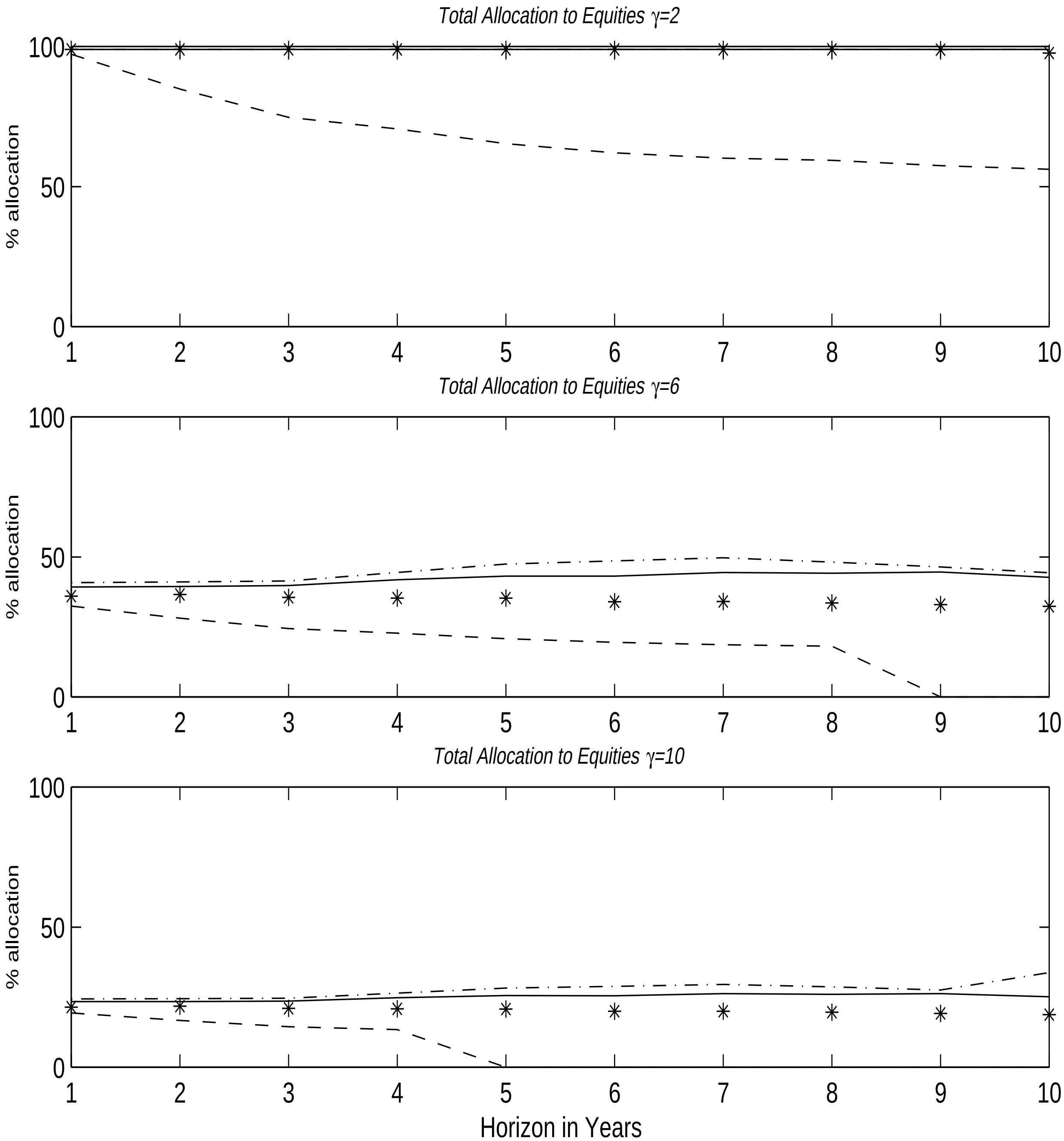

Figure 6 - Total Asset Allocations to Equities Based on Different Models Plotted against the Investment Horizon in Years for an Investor Following a Buy-and-Hold Strategy with a Power Utility Function and a Relative Risk-Aversion Coefficient Equal to γ. The Current Values of the Predictive Variables are Equal to their Sample Means.
Figure 6 displays total allocations to equities based on the weighted model (solid lines), the all-inclusive model (dash/dot lines), the highest-posterior-probability model (dashed lines), and the *iid* model (the lines denoted by the symbol '*'). Asset allocations are computed for investment horizons ranging from one to ten years and relative risk-aversion coefficients of two, six, and ten.

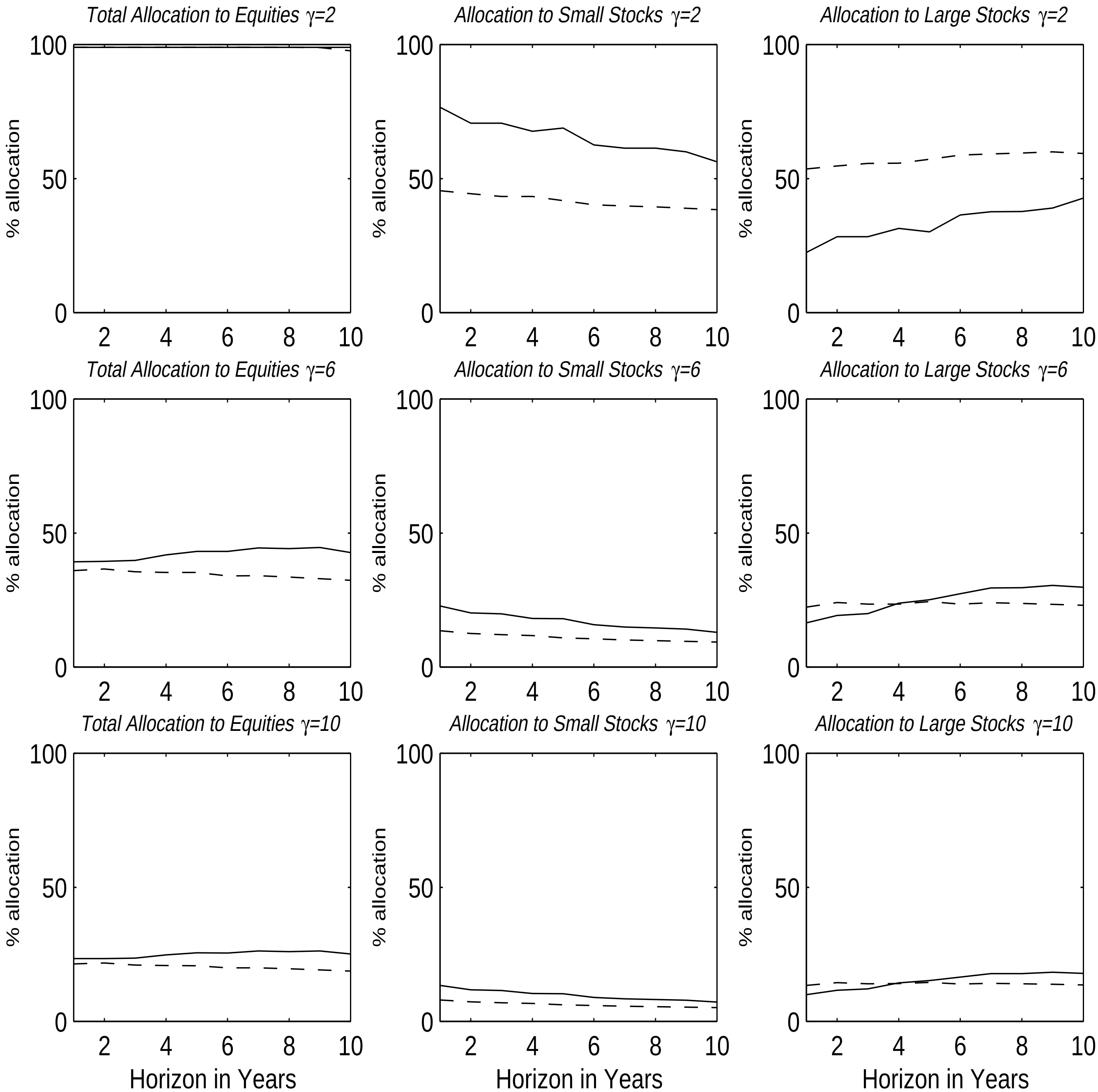

Figure 7 - Optimal Allocations to Size-Sorted Portfolios Plotted against the Investment Horizon in Years for an Investor Following a Buy-and-Hold Strategy, with a Power Utility Function and a Relative Risk-Aversion Coefficient Equal to γ.

Figure 7 displays the optimal allocations to stocks for the weighted model (solid lines) and the *iid* (dashed lines) model. Asset allocations associated with the weighted model are computed when the current values of the predictive variables are set equal to their sample means. The plots on the left display total allocations to equities. The decomposition of the total allocations is given by the allocation to small and large capitalization stocks: the three plots on the center and right, respectively. The investment in medium size stocks for every length of horizon and relative risk aversion coefficient is zero.

References

Akaike, H., 1974, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control* AC-19, 716-723.

Barberis, N., 1999, "Investing for the Long Run when Returns are Predictable," *Journal of Finance*, forthcoming.

Bawa Vijay, Stephen Brown, and Roger Klein, 1979. *Estimation Risk and Optimal Portfolio Choice* (North Holland, Amsterdam.).

Berger, J.O., and L.R. Pericchi, 1996, "The Intrinsic Bayes Factor for Model Selection and Prediction," *Journal of the American Statistical Association* 91, 109-122.

Bossaerts, P., and P. Hillion, 1999, "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?", *The Review of Financial Studies* 12, 405-428.

Bossaerts, P., and R. Green, 1989, "A General Equilibrium Model of Changing Risk Premia: Theory and Tests," *The Review of Financial Studies* 2, 231-263.

Boudoukh, J., M. Richardson, and R. Whitelaw, 1997, "Nonlinearities in the Relation Between the Equity Risk Premium and the Term Structure," *Management Science* 3, 371-385.

Box, G.E.P., 1980, "Sampling and Bayes' Inference in Scientific Modeling," *Journal of the Royal Statistical Society* [A] 143, 383-430.

Box, G.E.P., and G.C. Tiao, 1997. *Bayesian Inference in Statistical Analysis* (Addison-Wesely, Reading, MA.).

Brandt, M., 1999, "Estimating Portfolio and Consumption Choice: A Conditional Euler Equations Approach," *Journal of Finance*, forthcoming.

Brennan, M., E. Schwartz, and R. Lagnado, 1997, "Strategic Asset Allocation," *Journal of Economic Dynamic and Control* 21, 1377-1403.

Campbell, J., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics* 18, 373-399.

Campbell, J., 1993, "Intertemporal Asset Pricing without Consumption Data," *American Economic Review* 83, 487-512.

Campbell, J., and R., Shiller, 1988a, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies* 1, 195-227.

Campbell, J., and R., Shiller, 1988b, "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance* 43, 661-676.

Campbell, J., and G., Mankiw, 1989, "Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence," in Olivier J. Blanchard and Stanley Fischer (Eds.), NBER Macroeconomics Annual Cambridge, Mass, MIT Press, 185-216.

Campbell, J., A. Lo, and A.C. MacKinlay, 1997. *The Econometrics of Financial Markets* (Princeton.).

Campbell, J., and L. Viceira, 1999, "Consumption and Portfolio Decisions when Expected Returns are Time Varying," Working paper, Harvard University.

Chen, N., 1991, "Financial Investment Opportunities and the Macroeconomy," *Journal of Finance* 46, 529-554.

Chib, S., 1995, "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association* 90, 1313-1321.

Chib, S., and E. Greenberg, 1996, "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory* 12, 409-431.

Connor, G., and R. Korajczyk, 1986, "Performance Measurement with the Arbitrage Pricing Theory: A new Framework for Analysis," *Journal of Financial Economics* 15, 373-394.

Fama, E., and K. French, 1988, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy* 96, 246-273.

Fama, E., and K. French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* 19, 3-29.

Fama, E., and K. French, 1993, "Common Risk Factors in the Returns on Stock and Bonds," *Journal of Financial Economics* 33, 3-56.

Ferson, W., and C. Harvey, 1991, "The Variation in Economic Risk Premiums," *Journal of Political Economy* 99, 385-415.

Ferson, W., and R. Korajczyk, 1995, "Do Arbitrage Pricing Models Explain the Predictability of Stock Returns?," *Journal of Business* 68, 309-349.

Frost, P., and J. Savarino, 1986, "An Empirical Bayes Approach to Efficient Portfolio Selection," *Journal of Financial and Quantitative Analysis* 21, 293-305.

Gelfand, A.E., and A.F.M., Smith, 1990, "Sampling-Based approaches to Calculating Marginal Densities," *Journal of the American Statistical Association* 85, 398-409.

Geman, S., and D. Geman, 1984, "Stochastic Relaxation, Gibbs Distribution and the Bayesian Restoration of Images," *IEEE Transaction on Pattern Analysis and Machine Intelligence* 12, 609-628.

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

Jobson, J.D., and R. Korkie, 1980, "Estimation for Markowitz Efficient Portfolios," *Journal of the American Statistical Association* 75, 544-554.

Kandel, S., and R. 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.

Keim, D., and R. Stambaugh, 1986, "Predicting Returns in the Stock and the Bond Markets," *Journal of Financial Economics* 17, 357-390.

Kirby, C., 1998, "The Restriction on Predictability Implied by Rational Asset Pricing Models", *The Review of Financial Studies* 11, 343-382.

Kothari, S.P., and J. Shanken, 1997, "Book-to-Market, Dividend Yield, and Expected Market Returns: A Time Series Analysis," *Journal of Financial Economics* 44, 169-203.

Lettau, M., and S., Ludvigson, 1999, "Consumption, Aggregate Wealth and Expected Stock Returns", Federal Reserve Bank of New York, Staff Report No. 77.

Leamer, E.E., 1978. *Specification Searches: Ad Hoc Inference with Non-Experimental Data* (John Wiley, New York, NY.).

Leamer, E.E., 1983, "Let's Take the Con out of Econometrics," *American Economic Review* 73, 31-43.

Lo, A., and A.C. MacKinalay, 1997, "Maximizing Predictability in the Stock and Bond Markets," *Macroeconomic Dynamics* 1, 131-170.

McCulloch R., and P.E. Rossi, 1991, "A Bayesian Approach to Testing the Arbitrage Pricing Theory," *Journal of Econometrics* 49, 141-168.

Moreno, E., F. Bertolino, and W. Racugno, 1998, "An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing," *Journal of the American Statistical Association* 93, 1451-1460.

Moulton, B.R., 1991, "A Bayesian Approach to Regression Selection and Estimation, with Application to a Price Index for Ratio Services," *Journal of Econometrics* 49, 169-193.

Pastor, L., and R. Stambaugh, 1999, "Comparing Asset Pricing Models: An Investment Perspective," *Journal of Financial Economics*, forthcoming.

Pesaran, M.H., and A. Timmermann, 1995, "Predictability of Stock Returns: Robustness and Economic Significance," *Journal of Finance*, 50, 1201-1228.

Poirier, D.J., 1995. *Intermediate Statistics and Econometrics* (MIT Press, Cambridge, MA.).

Pontiff, J., and L.D. Schall, 1998, "Book-to-Market Rations as Predictors of Market Returns," *Journal of Financial Economics* 49, 141-160.

Raftery, A.E., D. Madigan, and J. Hoeting, 1997, "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association* 92, 179-191.

Rothenberg, T.J., 1963, "A Bayesian Analysis of Simultaneous Equation Systems," Report 6315, Econometric Institute, Netherlands School of Economics, Rotterdam.

Schwarz, G., 1978, "Estimating the Dimension of Model," emAnnals of Statistics 6, 416-464.

Stambaugh, R., 1986, "Bias in Regressions with Lagged Stochastic Regressors," working paper, University of Chicago, Chicago, IL.

Stambaugh, R., 1999, "Predictive Regressions," *Journal of Financial Economics*, forthcoming.

Tierney, L., 1994, "Markov Chains for Exploring Posterior Distributions," *Annals of Statistics* 22, 1701-1762.

Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics* (John Wiley and Sons, New-York.).

Zellner, A., and V.K. Chetty, "Prediction and Decision Problem in Regression Models from the Bayesian Point of View," 1965, *Journal of the American Statistical Association* 60, 608-615.