ELSEVIER

Journal of Financial Economics 64 (2002) 423–458

JOURNAL OF Financial ECONOMICS

www.elsevier.com/locate/econbase

Stock return predictability and model uncertainty[☆]

Doron Avramov*

The Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA

Received 5 February 2001; accepted 22 May 2001

Abstract

We use Bayesian model averaging to analyze the sample evidence on return predictability in the presence of model uncertainty. The analysis reveals in-sample and out-of-sample predictability, and shows that the out-of-sample performance of the Bayesian approach is superior to that of model selection criteria. We find that term and market premia are robust predictors. Moreover, small-cap value stocks appear more predictable than large-cap growth stocks. We also investigate the implications of model uncertainty from investment management perspectives. We show that model uncertainty is more important than estimation risk, and investors who discard model uncertainty face large utility losses. © 2002 Elsevier Science B.V. All rights reserved.

JEL classification: G11; G12; C11

Keywords: Stock return predictability; Model uncertainty; Bayesian model averaging; Portfolio selection; Variance decomposition

[☆] The paper is based on my dissertation at the Wharton School, the University of Pennsylvania. I am grateful to my dissertation committee members, Michael Brandt, Craig MacKinlay, Andrew Metrick, and Frank Schorfheide for their useful comments and suggestions. I especially thank my committee chairman Rob Stambaugh for his guidance and encouragement. I also acknowledge the comments of Gurdip Bakshi, Simon Benninga, John Chao, Tarun Chordia, Jeremy Gold, Rick Green, Cam Harvey, Robert Hauswald, Chris Jones, Shmuel Kandel, Chris Kirby, Kai Li, Anthony Lynch, Lubos Pastor, Nagpurnanand Prabhala, Krishna Ramaswamy, Oded Sarig, William Schwert (the editor), Jay Shanken, Lucie Tepla, Alex Triantis, Amir Yaron, Shlomith Zuta, and seminar participants at Boston College, Cornell University, Deutsche Bank Asset Management, Goldman Sachs Asset Management, London Business School, University of Maryland, University of Pennsylvania, Vanderbilt University, the 2000 meetings of the Washington Area Finance Association, and the 2000 meetings of the Western Finance Association. Comments of an anonymous referee have greatly improved the paper. All remaining errors are mine.

*Tel.: +1-301-405-0400; fax: +1-301-405-0359.

E-mail address: davramov@rhsmith.umd.edu (D. Avramov).

PII: S0304-405X(02)00131-9

1. Introduction

Although financial economists have identified variables that predict stock returns through time, the ‘‘correct’’ predictive regression specification has remained an open issue for several reasons. First, existing equilibrium pricing theories are not explicit about which variables should enter the predictive regression. This aspect is undesirable, as it renders the empirical evidence subject to data overfitting concerns. In particular, Bossaerts and Hillion (1999) confirm in-sample return predictability, but fail to demonstrate out-of-sample predictability. Second, the multiplicity of potential predictors also makes the empirical evidence difficult to interpret. For example, one may find an economic variable statistically significant based on a particular collection of explanatory variables, but often not based on a competing specification. Given that the true set of predictive variables is virtually unknown, this paper proposes a Bayesian model averaging approach to analyze stock return predictability.

In the context of predictive regressions, the Bayesian methodology is attractive. For one, it explicitly incorporates model uncertainty, and is therefore robust to model misspecification, at least within the universe of linear forecasting models. To be precise, the Bayesian approach assigns posterior probabilities to a wide set of competing return-generating models; it then uses the probabilities as weights on the individual models to obtain a composite weighted model. This optimally weighted model is then employed (i) to investigate the sample evidence on return predictability and (ii) to explore implications of model uncertainty from investment management perspectives. In one particular application of economic interest, we investigate how model uncertainty affects asset allocation decisions.

When we apply our Bayesian characterizations to the post-war data, several conclusions emerge about stock return predictability. First, we show that variables could be significant based on individual predictive regressions, but need not be significant when one appeals to the weighted forecasting model. In essence, taking model uncertainty into account appears to substantially diminish the predictive power of some explanatory variables. One interpretation of this evidence is that ignoring model uncertainty could lead to erroneous inferences about the relevance of predictive variables.

Next, the Bayesian methodology reveals the existence of in-sample and out-of-sample predictability, even when commonly adopted model selection criteria (such as adjusted R^2) fail to demonstrate out-of-sample predictability (Bossaerts and Hillion, 1999). Moreover, the out-of-sample prediction errors generated by the weighted model satisfy certain desirable properties. As we show, these prediction errors have zero mean and are serially uncorrelated. In addition, the prediction errors are essentially uncorrelated with predicted returns. In contrast, the out-of-sample performance of prediction errors generated by model selection criteria is often unsatisfactory.

Our posterior analysis finds that term premium (defined as the rate of return differential between long-term and short-term treasuries) and the market premium are useful predictors of future stock returns. On the other hand, the dividend yield and book-to-market, among others, have relatively small posterior probabilities of being correlated with future returns. The posterior analysis also detects strong cross-sectional differences in predictability among size and book-to-market sorted portfolios. Posterior odds in favor of predictability are substantially higher for small-cap value stocks relative to large-cap growth stocks.

Based on model posterior probabilities, we find that trend-deviation-in-wealth (Lettau and Ludvigson, 2001) displays an impressive predictive power only when the shares of asset wealth and labor income (in total wealth) are based on data realized subsequent to the prediction period. However, trend-deviation-in-wealth has poor predictive power when constructed using quantities available at the time of prediction. In fact, this variable is dominated by the traditional valuation ratios such as book-to-market and earnings yield. The evidence thus suggests that the strong predictive power of trend-deviation-in-wealth documented by Lettau and Ludvigson (2001) could be due to a look-ahead bias.

Two metrics based on variance decompositions and certainty equivalent returns are developed to judge the statistical and economic implications of model uncertainty for investment management. In particular, in the presence of model uncertainty, we show that investment opportunities can be represented by a weighted Bayesian predictive distribution. This endogenously derived distribution has the appealing property that it integrates out both the uncertainty about the forecasting model, and the uncertainty about model parameters (estimation risk).

Based on variance decompositions, the analysis shows that model uncertainty is more important than estimation risk for short-horizon investors. Moreover, asset allocations in the presence of estimation risk display sensitivity to whether model uncertainty is incorporated or ignored. An investor who is forced to discard model uncertainty, and instead hold a suboptimal portfolio relying on model selection criteria, perceives a substantial loss in risk-free certainty equivalent returns.

This article is related to a strand of studies investigating mispricing uncertainty in stock returns. In a fundamental contribution, Pastor (2000) and Pastor and Stambaugh (1999, 2000) investigate the uncertainty about whether a given single asset pricing model is valid. Mispricing uncertainty is also at the core of the analysis in Brennan and Xia (2001) and Wang (2001). Our study departs along two important dimensions. First, the general paradigm developed here encompasses a vast set of return generating processes, and consolidates these processes into a composite optimal-weighted forecasting model. Second, our analysis designs a Bayesian model selection criterion, which exhibits robust out-of-sample forecasting properties.

In their innovations, Kandel and Stambaugh (1996) and Barberis (2000) explore an asset allocation problem when stock returns are potentially predictable and the investment universe consists of an equity portfolio and the risk-free Treasury bill. For example, Barberis studies multi-period asset allocations with future rebalancing. Relative to these studies, the buy-and-hold investor considered here allocates funds across multiple equity portfolios and incorporates the additional element of model risk. Our results therefore do not hinge on the validity of any single forecasting model.[1]

The remainder of the paper proceeds as follows. Section 2 derives an analytical result for posterior probabilities of competing return generating models. It also derives three statistics for investigating the robustness of predictive variables in the presence of model uncertainty. Section 3 develops an econometric framework (i) to study sources of uncertainty about predicted stock returns and (ii) to analyze an asset allocation problem under model uncertainty. In Section 4 we describe the sample data. Section 5 contains empirical results on stock return predictability, and Section 6 discusses variance decomposition and asset allocation. Conclusions and ideas for future research are offered in the final Section 7. Appendix A presents all mathematical derivations.

2. Predictability in the presence of model uncertainty

Suppose you want to predict future rate of returns on equity portfolios using a linear predictive regression. When M explanatory variables are suspected relevant, there are 2^M competing regression specifications. Each of these obeys the form

$$r_t' = x_{j,t-1}' B_j + \varepsilon_{j,t}', \tag{1}$$

where r_t is the $N \times 1$ vector of continuously compounded returns on N portfolios in excess of the continuously compounded Treasury bill rate and j is a model-specific indicator. In Eq. (1), $x_{j,t-1}' = (1, z_{j,t-1}')$, $z_{j,t-1}$ is a model-unique subset, which contains m_j variables observed at the end of $t-1$, and B_j is an $(m_j+1) \times N$ matrix of the regression intercept and slope coefficients. The parameter m_j ranges between zero and M. When $m_j = 0$, returns are independently and identically distributed (iid). The iid model discards all variables as worthless predictors. The all-inclusive specification corresponds to $m_j = M$. For tractability of analysis, we assume that $\varepsilon_{j,t}$ is normally distributed with conditional mean zero and variance–covariance matrix Σ_j.

In what follows, we term uncertainty about the true set of predictive variables as "model uncertainty" or "model risk". Since the available time-series is often limited model uncertainty is especially relevant. Indeed, in extremely large samples, all potential predictors can be included in an all-inclusive specification. In this predictive regression, irrelevant variables will have slope-coefficient estimates converging to zero, their true value. However, in applications studying stock return predictability there are many possible explanatory variables but only a limited number of

[1] After the writing of previous versions, I become aware of the work by Cremers (2000) who implements a Bayesian approach in return predictability. While sharing the Bayesian feature, our work differs in methodology and in the scope of empirical findings. Whereas he focuses on a posterior analysis only, we conduct an extensive posterior and predictive analysis including variance decomposition, long horizon asset allocation, and statistical tests of significance of variables in predictive regressions when model uncertainty is explicitly incorporated. Moreover, his focus is on a single equity portfolio, as opposed to our multi-asset paradigm.

observations. The traditional single predictive regression paradigm thus offers little help in identifying useful predictors.

Instead, we use Bayesian model averaging. This procedure computes posterior probabilities for the collection of all 2^M models. It then uses the probabilities as weights on the individual models to obtain one composite weighted forecasting model, which summarizes the dynamics of future stock returns. The weighted model is employed (i) to investigate the sample evidence on predictability and (ii) to analyze investment implications of model uncertainty.

Bayesian model averaging contrasts markedly with the traditional approach of model selection. In the heart of the model selection approach, one uses a specific criterion to select a single model and then operates as if the model is correct. Implementing model selection criteria, the econometrician views the selected model as the true one with a unit probability and discards the other competing models as worthless, thereby ignoring model uncertainty. In contrast, we average over the dynamics implied by the set of all 2^M models.

The posterior probability computation necessitates eliciting prior distributions of all the relevant parameters conditional on each possible model (e.g., Kass and Raftery, 1995; Poirier, 1995). Our prior distribution for each of the model-specific parameters (B_j, Σ_j) is based on a hypothetical prior sample weighted against predictability, as suggested by Kandel and Stambaugh (1996). In that sample, the slope coefficients in the regression of excess stock returns on a set of information variables are equal to zero, and the means and variances of stock returns and predictive variables are equal to the actual sample counterparts, which are given by:

$$\bar{r} = \frac{1}{T}\sum_{t=1}^{T} r_t, \tag{2}$$

$$\hat{V}_r = \frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})(r_t - \bar{r})', \tag{3}$$

$$\bar{z}_j = \frac{1}{T}\sum_{t=0}^{T-1} z_{j,t}, \tag{4}$$

$$\hat{V}_{j,z} = \frac{1}{T}\sum_{t=0}^{T-1}(z_{j,t} - \bar{z}_j)(z_{j,t} - \bar{z}_j)', \tag{5}$$

where T is the actual sample size.

Using statistics from the actual sample to specify some of the parameters of the prior distribution is commonly termed "empirical Bayes" (Robbins, 1955, 1964). Based on the hypothetical prior sample, the prior for the regression coefficient B_j conditional on Σ_j is given by the multivariate Normal distribution:

$$\mathrm{vec}(B_j)|\Sigma_j \sim N\left(\mathrm{vec}(B_{j,0}), \frac{1}{T_0}\Sigma_j \otimes \begin{bmatrix} 1 + \bar{z}_j'\hat{V}_{j,z}^{-1}\bar{z}_j & -\bar{z}_j'\hat{V}_{j,z}^{-1} \\ -\hat{V}_{j,z}^{-1}\bar{z}_j & \hat{V}_{j,z}^{-1} \end{bmatrix}\right), \tag{6}$$

where $B_{j,0} = [\bar{r}, \mathbf{0}_j]', \mathbf{0}_j$ is an $N \times m_j$ matrix of zeros reflecting the "no predictability" prior sample, T_0 is the size of the hypothetical sample, and vec($\bullet$) denotes the vector formed by stacking the successive transformed rows of the matrix. The marginal prior for Σ_j follows Kandel and Stambaugh (1996, Eq. (B.6)) and obeys the inverted Wishart distribution (Zellner, 1971):

$$\Sigma_j \sim \mathrm{IW}(T_0 \hat{V}_r, T_0 - m_j - 1). \tag{7}$$

The analysis depends upon T_0, which determines the strength of the informative prior. As an extreme, if T_0 approaches infinity, the investor displays dogmatic beliefs about no predictability. Any finite sample size cannot reverse such tight beliefs. Our task is, therefore, to pick a reasonable value for the prior sample size. Kandel and Stambaugh (1996) motivate such a value. Using Monte Carlo simulations, they show that the implied priors of R^2 in the regression of excess stock returns on lagged predictive variables are roughly invariant to the number of predictors if the number of hypothetical data entries per parameter is held fixed (50 observations per parameter) as the number of parameters changes. Our analysis relies primarily on this. Essentially, the hypothetical prior size increases as the model contains more predictive variables. Therefore, we will denote the prior sample size with the model-specific indicator, i.e., $T_{j,0}$.

Having determined prior distributions for each of the competing models, we are ready to derive analytical expressions for the corresponding posterior probabilities. The posterior probability of model j (denoted $\mathscr{M}_j$) is given by

$$\mathrm{P}(\mathscr{M}_j|D) = \frac{\mathrm{P}(D|\mathscr{M}_j)\mathrm{P}(\mathscr{M}_j)}{\sum_{i=1}^{2^M} \mathrm{P}(D|\mathscr{M}_i)\mathrm{P}(\mathscr{M}_i)}, \tag{8}$$

where D stands for the data, $\mathrm{P}(\mathscr{M}_j)$ is the prior probability of $\mathscr{M}_j$, which is at the discretion of the decision-maker, and $\mathrm{P}(D|\mathscr{M}_j)$ is the corresponding marginal likelihood. The marginal likelihood of $\mathscr{M}_j$ is given by

$$\mathrm{P}(D|\mathscr{M}_j) = \frac{\mathscr{L}(\Sigma_j, B_j, D, \mathscr{M}_j)\mathrm{P}(\Sigma_j, B_j|\mathscr{M}_j)}{\mathrm{P}(\Sigma_j, B_j|D, \mathscr{M}_j)}, \tag{9}$$

where $\mathscr{L}(\Sigma_j, B_j; D, \mathscr{M}_j)$ is the likelihood function pertaining to $\mathscr{M}_j$ and $\mathrm{P}(\Sigma_j, B_j|\mathscr{M}_j)$ and $\mathrm{P}(\Sigma_j, B_j|D, \mathscr{M}_j)$ are the joint prior and posterior distributions of the model-specific parameters, respectively.

We show in Appendix A that the log marginal likelihood is given by

$$\begin{aligned} \ln[\mathrm{P}(D|\mathscr{M}_j)] = & -\frac{TN}{2}\ln(\pi) + \frac{T_{j,0} - m_j - 1}{2}\ln|T_{j,0}\hat{V}_r| \\ & - \frac{T_j^* - m_j - 1}{2}\ln|\tilde{S}_j| - \sum_{i=1}^{N}\ln\left\{\Gamma\left(\frac{T_{j,0} - m_j - i}{2}\right)\right\} \\ & + \sum_{i=1}^{N}\ln\left\{\Gamma\left(\frac{T_j^* - m_j - i}{2}\right)\right\} - \frac{N(m_j + 1)}{2}\ln\left(\frac{T_j^*}{T_{j,0}}\right), \end{aligned} \tag{10}$$

where

$$\tilde{S}_j = T_j^*(\hat{V}_r + \bar{r}\bar{r}') - \frac{T}{T_j^*}(T_{j,0}[\bar{r}, \bar{r}\bar{z}_j'] + R'X_j)(X_j'X_j)^{-1}(T_{j,0}[\bar{r}, \bar{r}\bar{z}_j']' + X_j'R), \tag{11}$$

$$X_j = [x_{j,0}, x_{j,1}, \ldots, x_{j,T-1}]', \tag{12}$$

$$R = [r_1, r_2, \ldots, r_T]', \tag{13}$$

$T_j^* = T + T_{j,0}$, $\Gamma(y)$ stands for the Gamma function evaluated at y, and $|x|$ is the determinant of x. For the iid model the marginal likelihood follows similarly except that $\tilde{S}_{\text{iid}} = T_{\text{iid}}^* \hat{V}_r$.

At this stage, the explanatory variables are deterministic, consistent with other studies computing marginal likelihood (Kass and Raftery, 1995) and traditional model selection criteria (Bossaerts and Hillion, 1999).

Having derived posterior probabilities, we propose three statistics to investigate the robustness of explanatory variables in predictive regressions. The first is cumulative posterior probabilities of the predictive variables. It is computed as $\mathscr{A}'\mathscr{P}$, where $\mathscr{A}$ is a $2^M \times M$ matrix representing all forecasting models by zeros and ones, designating exclusions and inclusions of predictors, respectively, and $\mathscr{P}$ is a $2^M \times 1$ vector containing model posterior probabilities. The resulting quantity indicates the probabilities that each of the predictive variables appears in the weighted forecasting model.

To illustrate, if the iid model receives a posterior probability equal to unity, the cumulative posterior probabilities are represented by an $M \times 1$ vector of zeros. On the other hand, if the all-inclusive model receives the entire posterior mass, the posterior probabilities are represented by an $M \times 1$ vector of ones. Let us also consider a more representative example. Suppose that some predictive variable, say dividend yield, receives a cumulative posterior probability of 45%. This suggests that dividend yield should appear in the weighted return-forecasting model with a probability of 45%.

The second statistic is a posterior t ratio. It is obtained by dividing the posterior mean of each of the slope coefficients in the weighted model by its corresponding posterior standard error. Based on Leamer (1978, pp. 117–118), the posterior mean and its corresponding variance are given by

$$\mathbb{E}(B|D) = \sum_{j=1}^{2^M} \mathrm{P}(\mathscr{M}_j|D)\tilde{B}_j, \tag{14}$$

$$\mathrm{Var}(B|D) = \sum_{i=1}^{2^M} \mathrm{P}(\mathscr{M}_j|D) \times \left\{ \frac{T\tilde{S}_j(X_j'X_j)^{-1}}{T_j^*(T_j^* - 4)} + [\tilde{B}_j - \mathbb{E}(B|D)][\tilde{B}_j - \mathbb{E}(B|D)]' \right\}, \tag{15}$$

where

$$\tilde{B}_j = \frac{T}{T_j^*}(X_j'X_j)^{-1}(T_{j,0}[\bar{r}, \bar{r}\bar{z}_j']' + X_j'R) \quad \text{for}\, j = 1, \ldots, 2^M \quad \text{and} \quad j \neq \text{iid}, \tag{16}$$

and $\tilde{B}_{\text{iid}} = \bar{r}$.[2] The mean in Eq. (14) follows by iterated expectations, conditioning first on the model space. The variance in Eq. (15) follows by using properties of the inverted gamma distribution (Zellner, 1971) and variance decomposition.

The posterior mean is merely a weighted average of slope estimates. The posterior variance incorporates both the estimated variance in every entertained model and the model-uncertainty component attributed to the dispersion in posterior means of the regression slopes across the models. The posterior t statistic differs from the well-known classical counterpart in that it explicitly accounts for model uncertainty.

Observe from Eq. (15) that the greater the uncertainty about the true forecasting model, the greater the variance of slope coefficients in the weighted model. Therefore, the greater model uncertainty the smaller the posterior t-ratio, or the less likely is the predictor to be statistically significant.

The third statistic is a posterior-odds ratio obtained by dividing the sum of posterior probabilities assigned to $2^M - 1$ models that retain at least one predictor by the posterior probability of the iid model. Posterior odds have also been found to be useful in testing portfolio efficiency, as noted by Shanken (1987). Specifically, Shanken (1987) shows that using posterior odds leads to a particular inference about mean variance efficiency that could substantially differ from the one obtained by the traditional p value.

3. Model uncertainty and investment perspectives

Kandel and Stambaugh (1996) and Barberis (2000) show that predictive regressions are useful in making portfolio decisions when investment opportunities are time-varying and the investment universe consists of an equity portfolio and the risk-free Treasury bill. Kandel and Stambaugh focus on a single period investor. Barberis extends their setting to a multi-period problem, in which the investor dynamically rebalances his portfolio. Both studies incorporate estimation risk but not model risk. This section develops a framework for analyzing buy-and-hold investment decisions in the presence of model uncertainty when the investment universe consists of multiple equity portfolios and a risk-free asset. Investment opportunities are expressed by the Bayesian weighted predictive distribution, as described below.

3.1. The Bayesian weighted predictive distribution

Let $y'_{j,t} = (r'_t, z'_{j,t})$ be the data-generating process corresponding to model j. We assume that the evolution of $y_{j,t}$ is governed by the stochastic process

$$y'_{j,t} = x'_{j,t-1}\Phi_j + u'_{j,t}, \tag{17}$$

[2] Here, we consider a multiple regression run separately for any risky asset. The multivariate student t distribution follows by integrating out Σ from the multivariate normal distribution $\mathrm{P}(B|\Sigma, D)$. Both $\tilde{B}_j$ and $\tilde{S}_j$ are of equal dimension for any entertained model since slope coefficients of excluded variables are zero. To illustrate, we rewrite $\tilde{B}_{\text{iid}}$ as $[\bar{r}, 0]'$ where 0 is a $1 \times M$ vector of zeros. Similarly $\tilde{S}_{\text{iid}}$ is an $(M+1) \times (M+1)$ matrix, all of which entries are zero, except for the first diagonal element, which is equal to $\sum_{t=1}^{T}(r_t - \bar{r})^2$.

where Φ_j is an $(m_j+1)\times(N+m_j)$ matrix whose first $(m_j+1)\times N$ columns is the matrix of predictive regression coefficients, B_j, and $u_{j,t}$ is an $(N+m_j)\times 1$ vector of forecast errors. We assume that $u_{j,t} \sim \text{iid N}(0,\Psi_j)$. The data-generating process in Eq. (17) nests a first order VAR for the dynamics of the predictive variables

$$z'_{j,t} = a'_j + z'_{j,t-1}A_j + \eta'_{j,t}. \tag{18}$$

The assumption of a first-order VAR is not restrictive since a higher-order VAR system can be re-expressed in a first order form (Campbell and Shiller, 1988a).

The Bayesian weighted predictive distribution of cumulative excess continuously compounded returns averages over the model space, and integrates over the posterior distribution that summarizes the within-model uncertainty about Φ_j and Ψ_j. It is given by

$$\mathrm{P}(R_{T+K}|D) = \sum_{j=1}^{2^M}\mathrm{P}(\mathscr{M}_j|D)\int_{\Psi_j,\Phi_j}\mathrm{P}(\Phi_j,\Psi_j|\mathscr{M}_j,D)\mathrm{P}(R_{T+K}|\mathscr{M}_j,\Phi_j,\Psi_j,D)\,\mathrm{d}\Phi_j\,\mathrm{d}\Psi_j, \tag{19}$$

where K is the investment horizon and $R_{T+K}=\sum_{k=1}^{K} r_{T+k}$. To our knowledge, an analytical solution for the integral in Eq. (19) is not feasible when $K>1$. Therefore, Monte Carlo integration is used in our empirical implementation.

Sampling from the Bayesian weighted predictive distribution is obtained by three steps, drawing first from the discrete distribution of models. To be precise, a model $\mathscr{M}_j$ is drawn with probability $\mathrm{P}(\mathscr{M}_j|D)$. Second, the model-specific parameters Φ_j and Ψ_j are drawn from their joint posterior distribution (derived in Appendix A). Third, given Φ_j and Ψ_j, an $N\times 1$ random vector of cumulative excess continuously compounded returns is drawn from the distribution of future stock returns conditioned upon the model, its specific parameters Φ_j and Ψ_j, and the sample data.

This conditional distribution is given by (see derivation in Appendix A)

$$R_{T+K}|\mathscr{M}_j,\Phi_j,\Psi_j,D \sim \mathrm{N}(\lambda_j,\Upsilon_j), \tag{20}$$

where

$$\lambda_j = Kb_j + C_j[((A'_j)^K - I_{m_j})(A'_j - I_{m_j})^{-1}]z_{j,T} \\ + C_j[A'_j((A'_j)^{K-1} - I_{m_j})(A'_j - I_{m_j})^{-1} - (K-1)I_{m_j}](A'_j - I_{m_j})^{-1}a_j, \tag{21}$$

$$\Upsilon_j = K\Sigma_j + \sum_{i=1}^{K}\delta_j(k)\Theta_j\delta_j(k)' + \sum_{k=1}^{K}\Lambda_j\delta_j(k)' + \sum_{k=1}^{K}\delta_j(k)\Lambda'_j, \tag{22}$$

$$\delta_j(k) = C_j[((A'_j)^{k-1} - I_{m_j})(A'_j - I_{m_j})^{-1}], \tag{23}$$

b_j and C_j are partitions of B_j corresponding to the intercept and slope coefficients in the regression of excess returns on lagged predictive variables, i.e., $B_j=[b_j,C_j]$, I_{m_j} is the identity matrix of order m_j, and Λ_j and Θ_j are partitions of the variance–covariance matrix Ψ_j:

$$\Psi_j = \begin{bmatrix} \Sigma_j & \Lambda_j \\ \Lambda_j' & \Theta_j \end{bmatrix}. \tag{24}$$

Essentially, no predictability corresponds to $C_{\text{iid}} = 0$, which yields $\lambda_{\text{iid}} = Kb_{\text{iid}}$ and $\Upsilon_{\text{iid}} = K\Sigma_{\text{iid}}$. The conditional mean and variance in an iid world increase linearly with the investment horizon.

The third step, i.e., drawing from the conditional distribution of future returns, differs from the algorithm proposed by Barberis (2000, Eqs. (18) and (19)) in that Barberis draws both returns and information variables from their joint conditional distribution, whereas we sample directly from the distribution of returns. Our algorithm is especially efficient when there is a large number of predictive variables and/or the investment universe contains multiple equity portfolios.

When investors are assumed to know the model and its specific parameters the only information from the sample relevant to drawing from the distribution of future stock returns would be the most recent observation of the predictive variables ($z_{j,T}$). Such an assumption is made in the classical approach for asset allocation. Accounting for both estimation and model risks, the perceived distribution of future returns departs from normality, and may be affected by higher-order moments such as skewness and fat tails.

It should be noted that the weighted predictive distribution makes use of posterior probabilities computed based on the return generating process in Eq. (1). There are several reasons to adhere to that practice. First, the full system in Eq. (17) requires eliciting informative priors on a much larger parameter space, making prior sensitivity more pronounced. Second, we attempt to investigate whether information variables predict future returns, not whether they are correlated with their own future realized values. Lastly, there is a potential feedback between model misspecification in the dynamics of predictive variables and model posterior probability.

3.2. Variance decomposition

Based on the weighted predictive distribution, one can show that future stock returns over the investment horizon are subject to three sources of uncertainty: (i) model uncertainty; (ii) a mixture of estimation risk; and (iii) a mixture of the within-model forecast error. Appendix A shows that the variance of predicted stock returns can be decomposed as follows:

$$\operatorname{var}\{R_{T+K}|D\} = \sum_{j=1}^{2^M} \mathrm{P}(\mathcal{M}_j|D)[\mathbb{E}\{\Upsilon_j\} + \operatorname{var}\{\lambda_j\} + (\mathbb{E}\{\lambda_j\} - \tilde{\lambda})(\mathbb{E}\{\lambda_j\} - \tilde{\lambda})'], \tag{25}$$

where $\mathbb{E}\{\Upsilon_j\}$ and $\operatorname{var}\{\lambda_j\}$ are two variance components corresponding to the forecast error and parameter uncertainty, respectively. By standard results (Leamer, 1978), the model uncertainty component is given by

$$\sum_{j=1}^{2^M} \mathrm{P}(\mathcal{M}_j|D)(\mathbb{E}\{\lambda_j\} - \tilde{\lambda})(\mathbb{E}\{\lambda_j\} - \tilde{\lambda})', \tag{26}$$

where $\tilde{\lambda} = \sum_{j=1}^{2^M} \mathrm{P}(\mathcal{M}_j|D)\mathbb{E}\{\lambda_j\}$ is the predicted mean of cumulative stock returns that averages across model-specific predicted means using posterior probabilities as weights. The empirical section quantifies each of the three risk components.

3.3. Portfolio choice in the presence of model uncertainty

What are the implications of model uncertainty for asset allocation decisions? The optimization problem of a buy-and-hold investor with iso-elastic preferences who allocates funds across N risky portfolios and the risk-free Treasury bill and who does not know a priori the true set of predictors is given by:

$$\omega^* = \arg\max_{\omega} \int_{R_{T+K}} \frac{[(1-\omega'\iota_N)\exp(r_f K) + \omega' \exp(r_f K \iota_N + R_{T+K})]^{1-\gamma}}{1-\gamma} \times \mathrm{P}(R_{T+K}|D)\, \mathrm{d}R_{T+K}, \tag{27}$$

where the integral is taken over the Bayesian weighted predictive distribution derived earlier. In Eq. (27), γ is the relative risk aversion parameter, ω is an $N \times 1$ vector denoting portfolio weights chosen for N risky portfolios 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, assumed constant over the investment horizon. Portfolio weights are restricted to the unit interval, meaning that short selling and buying on margin are precluded; otherwise, the expected utility would be equal to $-\infty$ (as explained, for example, by Barberis, 2000).

Our expected utility maximization is a version of the general Bayesian control problem developed by Zellner and Chetty (1965). Bawa et al. (1979), Jobson and Korkie (1980), Jorion (1985), Frost and Savarino (1986), Pastor (2000), and Pastor and Stambaugh (2000) compute optimal portfolios in a one-period framework in which returns are assumed iid. On the other hand, Kandel and Stambaugh (1996), Barberis (2000), and Avramov (2001) analyze a portfolio decision when returns are potentially predictable. In these studies the conditional distribution of stock returns is integrated over the parameter space to account for estimation risk. Integrating over both the model space and the within-model parameter space extends existing frameworks.

The integral in Eq. (27) is approximated by generating 400,000 independent draws for $\{R_{T+K}^{(g)}\}_{g=1}^{G}$ from the weighted predictive distribution using the algorithm described above. A constrained optimization code is then used to maximize the quantity

$$\mathbb{E}[U(W_{T+K}(\omega))] = \frac{1}{G}\sum_{g=1}^{G} \frac{\{(1-\omega'\iota_N)\exp(r_f K) + \omega' \exp(r_f K\iota_N + R^{(g)}_{T+K}\}^{1-\gamma}}{1-\gamma} \quad (28)$$

subject to ω being nonnegative, where G denotes the number of draws.

4. Data

The empirical examination uses monthly and quarterly observations on stock returns and information variables over April 1953 through December 1998. The investment universe consists of the six Fama and French (1993) portfolios, formed as the intersection of two size (S,B) and three book-to-market (L,M,H) groups. Each of the 2^M competing models considered in the study retains a unique subset of the following $M = 14$ information variables (taking one lag):

1. dividend yield on the value-weighted NYSE index (Div),
2. book-to-market on the Standard & Poor's Industrials (BM),
3. earnings yield on the Standard & Poor's Composite (EY),
4. the winners-minus-losers one-year momentum in stock returns (WML),
5. default risk spread, formed as the difference in annualized yields of Moody's Baa and Aaa rated bonds (Def),
6. monthly rate of a three-month Treasury bill (Tbill),
7. excess return on the CRSP value-weighted index with dividends (RET),
8. default risk premium, formed as the difference between return on long-term corporate bonds and return on long-term government bond (DEF),
9. term premium, formed as the difference between the monthly return on long-term government bond and the one-month Treasury bill rate (TERM),
10. January Dummy (Jan),
11. monthly inflation rate (Inf),
12. size premium (SMB),
13. value premium (HML),
14. term spread, formed as the difference in annualized yield of ten-year and one-year Treasuries (Term).

The source of data is provided in Appendix A.

In deciding which predictors to include, attention was given to those variables found important in previous studies as well as those popular business cycle variables for which there exist some theoretical motivations.[3] The reasoning for including the variables HML, SMB, and WML, mostly notable as economy-wide factors in asset

[3] Studies using subsets of the above-listed predictors include Brandt (1999), Ait-Sahalia and Brandt (2001), Campbell (1987), Campbell and Shiller (1988b), Carhart (1997), Chen et al. (1986), Fama and French (1988, 1989, 1993), Fama and Schwert (1977), Ferson and Harvey (1991, 1999), French et al. (1987), Goetzmann and Jorion (1993), Keim and Stambaugh (1986), Kirby (1997), Kothari and Shanken (1997), Lo and MacKinlay (1997), Pesaran and Timmermann (1995), Schwert, (1990), and Shanken (1990).

pricing models, follows Merton (1973) and Campbell (1996) whose intertemporal CAPM does not distinguish between variables that predict market returns and variables that explain the cross-section variation in expected return. Moreover, Liew and Vassalou (2000) show that SMB and HML are useful in predicting economic growth, making the inclusion of these variables of interest while examining stock return predictability.

Table 1 exhibits slope coefficients (first rows) and their corresponding *t*-ratios (second rows) obtained by regressing excess monthly returns on size book-to-market sorted portfolios on an intercept and 14 lagged predictive variables described above.

Table 1

Multiple regressions of monthly excess continuously compounded returns on predictive variables

The table exhibits slope coefficients (first rows) and their corresponding *t*-ratios (second rows) obtained by regressing excess returns on each of the size book-to-market portfolios on an intercept and 14 lagged predictive variables described below. Also reported (third rows) are covariances between unexpected returns and innovations in predictive variables, $\mathrm{cov}_t(\varepsilon_{t+1}, \eta'_{t+1})$. (January dummy does not evolve stochastically and therefore such covariances are, by definition, equal to zero.) Excess returns are on six portfolios formed as the intersection of two size (S,B) and three book-to-market (L,M,H) groups. The set of predictors includes: dividend yield (Div); book-to-market (BM); earnings yield (EY); the one-year momentum portfolio (WML); default risk spread (Def); the monthly rate of a three-month Treasury bill (Tbill); excess return on the value-weighted index (RET); default risk premium (DEF); term premium (TERM); January Dummy (Jan); inflation rate (Inf); size premium (SMB); value premium (HML); and the term spread (Term).

Portfolio	Predictive variables													
	Div	BM	EY	WML	Def	Tbill	RET	DEF	TERM	Jan	Inf	SMB	HML	Term
SL	−0.01	0.00	−0.07	0.08	−0.02	0.02	−6.80	0.02	0.43	0.32	0.02	−0.01	0.19	−0.11
	−0.99	0.00	−1.64	2.55	−0.27	1.47	−3.11	0.25	1.60	2.55	2.43	−1.35	1.74	−0.96
	−0.08	−1.15	−1.32	−0.32	0.06	−0.01	2.33	0.07	0.22	1.21	−2.21	1.05	−0.60	1.78
SM	−0.01	−0.16	−0.03	0.06	−0.02	0.01	−3.97	0.04	0.33	0.34	0.03	−0.01	0.13	0.00
	−1.01	−0.21	−1.09	2.32	−0.35	1.11	−2.28	0.72	1.51	3.46	3.69	−1.52	1.45	−0.02
	−0.07	−0.95	−1.12	−0.34	0.11	−0.01	1.90	0.07	0.21	0.82	−1.80	0.84	−0.22	2.31
SH	−0.01	−0.49	−0.01	0.05	−0.05	0.01	−3.12	0.08	0.25	0.29	0.04	−0.01	0.11	0.09
	−0.73	−0.65	−0.42	2.21	−0.82	0.71	−1.80	1.22	1.14	2.99	5.50	−1.49	1.25	0.92
	−0.07	−0.91	−1.10	−0.40	0.00	−0.01	1.84	0.08	0.21	0.59	−1.47	0.84	−0.04	2.52
BL	0.00	0.33	−0.07	0.04	−0.01	0.01	−4.08	−0.08	0.41	0.26	0.00	−0.02	0.07	−0.08
	0.19	0.47	−2.20	1.95	−0.16	1.55	−2.53	−1.39	2.08	2.87	−0.60	−2.40	0.88	−0.91
	−0.07	−0.99	−0.98	0.00	0.14	−0.01	1.87	0.02	0.30	0.70	−1.75	0.23	−0.52	1.48
BM	0.00	0.08	−0.05	0.04	0.03	0.01	−4.63	−0.10	0.35	0.30	0.01	−0.01	0.08	0.01
	−0.09	0.12	−1.80	2.27	0.60	1.62	−3.28	−2.07	2.03	3.72	1.25	−1.36	1.07	0.16
	−0.06	−0.84	−0.81	−0.05	0.24	0.00	1.63	0.03	0.29	0.71	−1.32	0.20	−0.20	1.81
BH	0.00	−0.13	−0.01	0.03	−0.03	0.01	−2.74	−0.06	0.21	0.22	0.03	−0.01	0.03	0.02
	−0.30	−0.19	−0.41	1.41	−0.49	1.01	−1.80	−1.05	1.13	2.52	3.92	−1.74	0.34	0.30
	−0.06	−0.85	−0.83	−0.18	0.13	−0.01	1.63	0.04	0.28	0.64	−1.41	0.28	0.08	2.04

Also reported (third rows) are covariances between unexpected returns and innovations in predictive variables, $\mathrm{cov}(\varepsilon_{t+1}, \eta'_{t+1})$. Such covariances are important determinants of asset allocation decisions when investment opportunities are time-varying (see Barberis, 2000). Table 1 exhibits ample evidence supporting return predictability, as many of the information variables are significant at conventional significance levels.

5. Empirical results on stock return predictability

Consideration of all linear data-generating processes in the presence of 14 predictive variables necessitates the comparison of $2^{14} = 16{,}384$ models. Eq. (10) computes the marginal likelihood for every model, and Eq. (8) weights the marginal likelihood by the model prior probability and normalizes the result to obtain the model posterior probability. Prior probabilities are allocated equally across models.

5.1. Monthly observations

Table 2 reports results for the monthly sample. First rows display cumulative posterior probabilities $\mathscr{A}'\mathscr{P}$ for 14 predictors, as noted earlier. Second rows denote the highest-probability compositions, represented by combinations of zeros and ones, designating exclusions and inclusions of predictive variables, respectively. Several aspects of results merit closer attention.

First, only one or at most two predictors are retained as useful in the highest-probability models. Other predictive variables are discarded as worthless. Second, the highest-cumulative-probability predictors are the market premium, term premium, January Dummy, and inflation. The market premium is prominent in predicting small-cap stocks regardless of their book-to-market classification. However, it poorly predicts large-cap stocks. January Dummy better predicts small-cap than large-cap stocks. This is consistent with Blume and Stambaugh (1983) and Keim (1983), who trace much of the evidence on the size effect to the month of January. January Dummy also better predicts high book-to-market relative to low book-to-market stocks.

Next, among the traditional market ratios, i.e., dividend yield, book-to-market, and earnings yield, the last receives the highest cumulative probabilities. Lastly, although previous evidence has shown that SMB and HML are robust in predicting contemporaneous stock returns (Fama and French, 1993) and future economic growth (Liew and Vassalou, 2000), they are correlated only marginally with future monthly returns.

Table 3 exhibits posterior means of slope coefficients in the weighted model, as computed in Eq. (14), and two t-ratios. The first is obtained by dividing the posterior mean by the posterior standard error corresponding to the first component in Eq. (15), thereby ignoring model uncertainty. The second, the posterior t-ratio, divides the posterior mean by the two sources of uncertainty, including model

Table 2

Posterior probabilities of forecasting models based on a prior sample weighted against predictability

The first rows display cumulative posterior probabilities computed as $\mathscr{A}'\mathscr{P}$, where $\mathscr{A}$ is a $2^{14} \times 14$ matrix representing all forecasting models by their unique combinations of zeros and ones and $\mathscr{P}$ is a $2^{14} \times 1$ vector including posterior probabilities for all models. The second rows denote the highest-posterior-probability compositions represented by a combination of zeros and ones designating exclusions and inclusions of predictive variables, respectively. The stock universe comprises six portfolios identified by two letters designating increasing values of size (S,B) and book-to-market (L,M,H). Following are the predictors spanning the information set: dividend yield (Div); book-to-market (BM); earnings yield (EY); the momentum portfolio (WML); the difference in annualized yields of Moody's Baa and Aaa rated bonds (Def); the monthly rate of a three-month Treasury bill (Tbill); excess return on the value-weighted index (RET); the difference between the return on long-term corporate bonds and the return on long-term government bond (DEF); the difference between the monthly return on long-term government bond and the one-month Treasury bill rate (TERM); January Dummy (Jan); inflation rate (Inf); size premium (SMB); value premium (HML); and the difference in annualized yield of ten-year and one-year Treasury bills (Term). Figures displayed below are computed when investors encounter a hypothetical sample weighted against predictability.

Portfolio	Predictive variables													
	Div	BM	EY	WML	Def	Tbill	RET	DEF	TERM	Jan	Inf	SMB	HML	Term
SL	0.20	0.08	0.38	0.02	0.14	0.28	0.48	0.04	0.12	0.21	0.31	0.16	0.05	0.08
	0	0	**1**	0	0	0	0	0	0	0	**1**	0	0	0
SM	0.12	0.06	0.16	0.02	0.10	0.09	0.40	0.06	0.54	0.77	0.15	0.12	0.02	0.19
	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0	0
SH	0.06	0.05	0.07	0.03	0.06	0.04	0.49	0.03	0.35	1.00	0.06	0.08	0.02	0.22
	0	0	0	0	0	0	**1**	0	0	**1**	0	0	0	0
BL	0.12	0.05	0.14	0.04	0.14	0.13	0.05	0.07	0.20	0.03	0.69	0.05	0.05	0.15
	0	0	0	0	0	0	0	0	0	0	**1**	0	0	0
BM	0.15	0.06	0.15	0.03	0.20	0.34	0.03	0.09	0.54	0.07	0.23	0.04	0.03	0.25
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
BH	0.07	0.06	0.06	0.03	0.09	0.09	0.02	0.03	0.17	0.92	0.21	0.02	0.02	0.47
	0	0	0	0	0	0	0	0	0	**1**	0	0	0	**1**

uncertainty that summarizes the dispersion in posterior means of slope coefficients across the models.

The extra variance of slope coefficients in predictive regressions attributed to (ex post) model uncertainty calls into question the apparent predictive power of many economic variables. The market premium, term premium, January dummy, inflation, and term spread can be significant based on t-ratios that ignore model uncertainty, but often not when such uncertainty is taken into account. This shows that after observing the sample data there is still a large amount of uncertainty about the true return-generating model, leading to considerable uncertainty about the true values of slope coefficients in the weighted model.

Table 3

Slope coefficients in the weighted model and posterior t-ratios

The first rows denote posterior means of slope coefficients obtained by averaging slope estimates across models:

$$\mathbb{E}(B|D) = \sum_{j=1}^{2^M} \mathrm{P}(\mathcal{M}_j|D)\tilde{B}_j.$$

Second and third rows denote t-ratios unadjusted and adjusted to account for model uncertainty, respectively. In particular, the former is obtained by dividing the posterior mean of each of the slope coefficients by its posterior standard error corresponding to the first variance component in the following equation:

$$\mathrm{Var}(B|D) = \sum_{j=1}^{2^M} \mathrm{P}(\mathcal{M}_j|D)\left\{\frac{T\tilde{S}_j(X_j'X_j)^{-1}}{T_j^*(T_j^*-4)} + [\tilde{B}_j - \mathbb{E}(B|D)][\tilde{B}_j - \mathbb{E}(B|D)]'\right\}.$$

The latter divides the posterior mean by the posterior standard error corresponding to the overall variance, including model uncertainty that summarizes the dispersion in slopes across models. The statistics are computed separately for each of six equity portfolios formed as the intersection of two size (S,B) and three book-to-market (L,M,H) groups. Following are the predictors spanning the information set: dividend yield (Div); book-to-market (BM); earnings yield (EY); the momentum portfolio (WML); the difference in annualized yields of Moody's Baa and Aaa rated bonds (Def); the monthly rate of a three-month Treasury bill (Tbill); excess return on the value-weighted index (RET); the difference between the return on long-term corporate bonds and the return on long-term government bond (DEF); the difference between the monthly return on long-term government bond and the one-month Treasury bill rate (TERM); January Dummy (Jan); inflation rate (Inf); size premium (SMB); value premium (HML); and the difference in annualized yield of ten-year and one-year Treasury bills (Term). Figures displayed below are computed when investors encounter a hypothetical sample weighted against predictability.

Portfolio	Predictive variables													
	Div	BM	EY	WML	Def	Tbill	RET	DEF	TERM	Jan	Inf	SMB	HML	Term
SL	0.13	0.00	0.01	0.00	0.00	−0.96	0.08	0.01	0.02	0.00	−0.01	0.03	−0.01	0.00
	0.99	0.22	1.79	−0.05	0.96	−1.53	2.01	0.23	0.72	1.05	−1.60	0.87	−0.28	0.50
	0.43	0.11	0.69	−0.05	0.34	−0.55	0.85	0.14	0.32	0.47	−0.60	0.38	−0.17	0.25
SM	0.06	0.00	0.00	0.00	0.00	−0.22	0.05	0.02	0.12	0.02	0.00	0.02	0.00	0.00
	0.75	0.38	1.04	−0.12	0.81	−0.77	1.93	0.46	2.37	2.89	−1.05	0.80	−0.04	1.10
	0.32	0.19	0.39	−0.09	0.28	−0.28	0.74	0.21	0.97	1.54	−0.38	0.34	−0.04	0.44
SH	0.02	0.00	0.00	0.00	0.00	−0.09	0.07	0.01	0.07	0.03	0.00	0.01	0.00	0.00
	0.44	0.39	0.57	−0.24	0.53	−0.47	2.22	0.24	1.77	4.69	−0.49	0.61	0.11	1.25
	0.21	0.19	0.24	−0.14	0.20	−0.18	0.89	0.12	0.67	4.48	−0.21	0.26	0.08	0.48
BL	0.04	0.00	0.00	0.00	0.00	−0.24	0.00	0.02	0.03	0.00	−0.01	0.00	0.00	0.00
	0.57	0.04	0.65	−0.10	0.77	−0.79	0.19	0.39	0.94	−0.06	−2.53	0.24	−0.18	0.74
	0.30	0.02	0.33	−0.09	0.34	−0.32	0.14	0.21	0.43	−0.05	−1.26	0.17	−0.14	0.35
BM	0.06	0.00	0.00	0.00	0.00	−0.73	0.00	0.02	0.08	0.00	0.00	0.00	0.00	0.00
	0.76	0.21	0.79	0.05	1.16	−1.68	0.00	0.51	2.04	0.36	−1.16	0.16	0.02	1.19
	0.36	0.13	0.35	0.05	0.43	−0.62	0.00	0.25	0.93	0.21	−0.48	0.12	0.02	0.51
BH	0.02	0.00	0.00	0.00	0.00	−0.15	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.00
	0.44	0.35	0.37	−0.10	0.63	−0.65	0.08	0.09	0.88	3.39	−1.16	0.09	0.06	1.91
	0.23	0.20	0.19	−0.08	0.25	−0.26	0.07	0.07	0.40	2.36	−0.46	0.08	0.06	0.84

As suspected, the cumulative posterior probabilities ($\mathscr{A}'\mathscr{P}$) are related to the posterior t-ratios. Focusing on small-cap portfolios, high cumulative posterior probabilities for the market premium, January Dummy, term premium, and inflation (Table 2) are followed by higher values of posterior t-ratios (Table 3). Focusing on large-cap portfolios, smaller cumulative probabilities for RET, WML, and HML are followed by smaller posterior t-ratios.

The third statistic undertaken to assess the sample evidence on predictability is the posterior-odds ratio of predictability versus no predictability. Posterior odds for the six size book-to-market portfolios are 550 (SL), 10,886 (SM), 1,040,000 (SH), 74 (BL), 121 (BM), and 1,249 (BH). These figures reveal prominent cross-sectional differences in predictability. The evidence in favor of predictability is the strongest for small value stocks (SH), the weakest for large growth stocks (BL).

Holding book-to-market fixed, posterior odds in favor of predictability are substantially higher for small-cap stocks relative to large-cap stocks (550 versus 74 for low book-to-market stocks and 1,040,000 versus 1,249 for high book-to-market stocks). Similarly, controlling for size, posterior odds are higher for high-versus-low book-to-market stocks (1,040,000 versus 550 for small stocks and 1,249 versus 74 for large stocks).

5.2. Quarterly observations

In a recent study, Lettau and Ludvigson (2001) introduce the trend-deviation-in-wealth (henceforth *cay*) as a powerful predictor of quarterly returns at short and intermediate horizons. We examine the predictive power of *cay* and the overall evidence on predictability of three-month holding period returns using our Bayesian framework. We first construct an additional set of information variables in which *cay* replaces January Dummy.

Drawing on Campbell and Shiller (1988a), Lettau and Ludvigson (2001) argue that *cay* summarizes expectations about future stock returns. The variable *cay* is computed as $c_t - wa_t - (1-w)y_t$, where c_t, a_t, and y_t denote log consumption, nonhuman wealth (or asset wealth), and labor income, respectively and w equals the average share of nonhuman wealth in total wealth. Consumption, wealth, and income data are released by the Federal Reserve Board within two months of the end of a quarter, suggesting that the *cay* realization at quarter t is made known to capital market participants at the subsequent quarter and hence can be used to predict returns realized at or after quarter $t+2$.

Here are two points that must be noted about *cay*. First, the share of nonhuman wealth in total wealth, w, is computed based on data realized after the time future returns are predicted. This raises some difficulties in interpreting the trend-deviation-in-wealth as a purely ex ante variable, at least from an investment perspective. Second, the estimated weights on asset wealth (w) and labor income ($1-w$) do not sum up to unity. Rather, the former is equal to 0.3054 and the latter to 0.5891, thereby summing up to 0.8945 (Lettau and Ludvigson, 2001).

Using quarterly observations, we entertain a new benchmark value for the prior sample size, T_0. Leaving the prior sample size unchanged amounts to weighting priors against predictability to a stronger degree as the ratio T_0/T would increase three times. To maintain this ratio fixed across the monthly and quarterly experiments, posterior probabilities for the new model space are computed with T_0 taking values equivalent to 17 prior observations per parameter.

Panel A of Table 4 exhibits cumulative posterior probabilities for the new set of predictors. The variable *cay* indeed dominates dividend yield, the market premium, default-risk spread, and term spread, predictive variables studied by Lettau and Ludvigson (2001). The variable *cay* outperforms book-to-market, WML, HML, and inflation as well. However, its notable challenger is the term premium, which, in some cases (portfolios SH, BL, and BH) receives higher posterior probabilities. In fact, averaging cumulative posterior probabilities equally across the six portfolios, one finds that the average cumulative probability pertaining to *cay* is approximately 31%, smaller than 34%, the counterpart pertaining to term premium. Interestingly, SMB better predicts quarterly versus-monthly returns on large-cap stocks, whereas HML remains dismal for quarterly observations as well.

Lettau and Ludvigson (2001) address the concern arising due to the fact that the shares of asset wealth and labor income in total wealth are estimated using the whole sample. They run an analysis estimating *cay* every period, using only data available at the time the forecast is made, thereby generating 122 out-of-sample observations. We repeat our analysis incorporating these out-of-sample observations. Cumulative posterior probabilities and highest-posterior-probability compositions are displayed in Panel B of Table 4.

The analysis shows that the apparent forecasting power of *cay* crucially depends upon whether its coefficients are estimated with or without the look-ahead bias. Based on out-of-sample estimates, the predictive power of *cay* is dominated by many of the other variables, including dividend yield, book-to-market, and earnings yield. None of the highest-posterior-probability compositions retains the out-of-sample *cay*. In contrast, term premium appears in four highest-posterior-probability compositions (portfolios SH, BL, BM, and BH) and possesses the highest cumulative posterior probabilities, ranging between 25% and 38%.

In the analysis that follows, results for the quarterly sample are based upon the in-sample *cay*. In particular, Table 5 exhibits *t*-ratios unadjusted (second rows) and adjusted (third rows) to account for model uncertainty. Again, model uncertainty questions the relevance of economic variables in forecasting future returns. Term premium and *cay* are close to significant or significant in forecasting quarterly returns based on *t*-ratios that ignore model uncertainty, but not when such uncertainty is incorporated.

5.3. The market and term premia

Why should term and market premia be correlated with future stock returns? The ability of the market premium, essentially biased towards large-cap stocks, to predict small-cap stock returns is closely related to the lead–lag pattern uncovered by Lo and MacKinlay (1990). The notion is that current returns on small stocks are correlated with past returns of large stocks.

Table 4

Posterior probabilities of forecasting models using quarterly observations

In both panels, first rows display cumulative posterior probabilities computed as $\mathscr{A}'\mathscr{P}$, where $\mathscr{A}$ is a $2^{14} \times 14$ matrix representing all forecasting models by their unique combinations of zeros and ones and $\mathscr{P}$ is a $2^{14} \times 1$ vector including posterior probabilities for all models. Second rows denote highest-posterior-probability compositions represented by a combination of zeros and ones designating exclusions and inclusions of predictive variables, respectively. The stock universe comprises six portfolios identified by two letters designating increasing values of size (S,B) and book-to-market (L,M,H). Following are the predictive variables: dividend yield (Div); book-to-market (BM); earnings yield (EY); momentum (WML); the difference in annualized yields of Moody's Baa and Aaa rated bonds (Def); the monthly rate of a three-month Treasury bill (Tbill); excess return on the value-weighted index (RET); the difference between the return on long-term corporate bonds and the return on long-term government bond (DEF); the difference between the monthly return on long-term government bond and the one-month Treasury bill rate (TERM); trend-deviation-in-wealth (*cay*), inflation rate (Inf); size premium (SMB); value premium (HML); and the difference in annualized yield of ten-year and one-year Treasury bills (Term). The *cay* is first estimated with the look-ahead bias using the full sample (Panel A) and then using one-quarter ahead out-of-sample forecasts (Panel B).

Portfolio	Predictive variables													
	Div	BM	EY	WML	Def	Tbill	RET	DEF	TERM	*cay*	Inf	SMB	HML	Term
Panel A: cay is estimated using the full sample, i.e., with the look-ahead bias														
SL	0.22	0.09	0.26	0.07	0.07	0.30	0.04	0.12	0.35	0.36	0.04	0.08	0.04	0.07
	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0
SM	0.18	0.08	0.15	0.07	0.09	0.14	0.06	0.14	0.40	0.47	0.04	0.11	0.04	0.13
	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0
SH	0.16	0.10	0.15	0.08	0.11	0.14	0.10	0.14	0.47	0.25	0.05	0.10	0.04	0.16
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
BL	0.12	0.06	0.08	0.05	0.08	0.22	0.06	0.13	0.33	0.21	0.06	0.23	0.04	0.13
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
BM	0.19	0.07	0.16	0.04	0.17	0.41	0.05	0.06	0.21	0.31	0.06	0.15	0.04	0.17
	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0
BH	0.12	0.08	0.09	0.04	0.15	0.21	0.15	0.09	0.26	0.25	0.08	0.21	0.04	0.22
	0	0	0	0	0	0	0	0	0	**1**	0	0	0	0
Panel B: cay is estimated using one-quarter-ahead out-of-sample forecasts														
SL	0.31	0.12	0.37	0.13	0.12	0.36	0.08	0.14	0.32	0.12	0.08	0.20	0.07	0.20
	0	0	**1**	0	0	**1**	0	0	0	0	0	0	0	0
SM	0.28	0.12	0.32	0.12	0.18	0.21	0.08	0.14	0.32	0.11	0.08	0.21	0.07	0.35
	0	0	**1**	0	0	0	0	0	0	0	0	0	0	**1**
SH	0.24	0.12	0.28	0.13	0.18	0.18	0.09	0.14	0.38	0.10	0.08	0.21	0.07	0.30
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
BL	0.14	0.11	0.15	0.09	0.13	0.15	0.10	0.18	0.30	0.18	0.09	0.21	0.09	0.16
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
BM	0.19	0.11	0.20	0.09	0.16	0.26	0.10	0.13	0.25	0.12	0.10	0.15	0.09	0.24
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0
BH	0.14	0.10	0.14	0.10	0.23	0.14	0.11	0.13	0.29	0.10	0.10	0.26	0.08	0.23
	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0

Table 5

Slope coefficients in the weighted model and posterior t-ratios: the case of quarterly observations

The first rows denote posterior means of slope coefficients obtained by averaging slope estimates across models:

$$\mathbb{E}(B|D) = \sum_{j=1}^{2^M} \mathrm{P}(\mathcal{M}_j|D)\tilde{B}_j.$$

Second and third rows denote t-ratios unadjusted and adjusted to account for model uncertainty, respectively. In particular, the former is obtained by dividing the posterior mean of each of the slope coefficients obtained by averaging slope estimates across models by the posterior standard error corresponding to the first variance component in the following equation:

$$\mathrm{Var}(B|D) = \sum_{i=1}^{2^M} \mathrm{P}(\mathcal{M}_j|D)\left\{\frac{T\tilde{S}_j(X_j'X_j)^{-1}}{T_j^*(T_j^*-4)} + [\tilde{B}_j - \mathbb{E}(B|D)][\tilde{B}_j - \mathbb{E}(B|D)]'\right\}.$$

The latter divides the posterior mean by the posterior standard error corresponding to the overall variance, including model uncertainty that summarizes the dispersion in slopes across models. The hypothetical no-predictability informative sample takes values equivalent to 17 observations per parameter. The statistics are computed separately for each of six equity portfolios formed as the intersection of two size (S,B) and three book-to-market (L,M,H) groups. Following are the predictors constituting the information set: dividend yield (Div); book-to-market (BM); earnings yield (EY); momentum (WML); default risk spread (Def); the three-month rate of a three-month Treasury bill (Tbill); a quarterly excess return on the value-weighted index (RET); default risk premium (DEF); term premium (TERM); trend deviation in wealth (*cay*); the three-month inflation rate (Inf); size premium (SMB); value premium (HML); and the term spread (Term). Figures displayed below are computed when investors perceive the events of predictability versus no predictability as equally likely prior to encountering a hypothetical sample weighted against predictability.

Portfolio	Predictive variables													
	Div	BM	EY	WML	Def	Tbill	RET	DEF	TERM	*cay*	Inf	SMB	HML	Term
SL	0.45	0.00	0.02	0.01	0.00	−0.82	0.00	0.08	0.13	0.86	0.00	−0.01	0.00	0.00
	1.03	0.30	1.24	0.28	0.31	−1.27	0.10	0.53	1.39	1.76	−0.09	−0.33	−0.08	0.29
	0.42	0.16	0.42	0.16	0.14	−0.45	0.07	0.24	0.57	0.73	−0.06	−0.18	−0.07	0.16
SM	0.23	0.00	0.01	0.01	0.00	−0.22	0.00	0.08	0.13	0.98	0.00	−0.01	0.00	0.00
	0.80	0.34	0.68	0.29	0.53	−0.66	0.21	0.64	1.63	2.15	−0.12	−0.45	−0.09	0.66
	0.34	0.18	0.27	0.16	0.20	−0.25	0.12	0.27	0.66	0.92	−0.08	−0.22	−0.07	0.28
SH	0.18	0.00	0.01	0.01	0.00	−0.23	0.01	0.10	0.19	0.39	0.00	−0.01	0.00	0.00
	0.68	0.43	0.62	0.32	0.62	−0.66	0.45	0.68	1.94	1.22	−0.12	−0.38	−0.07	0.79
	0.31	0.22	0.27	0.18	0.23	−0.25	0.21	0.29	0.82	0.51	−0.08	−0.20	−0.06	0.33
BL	0.08	0.00	0.00	0.00	0.00	−0.29	0.00	0.06	0.08	0.20	0.00	−0.04	0.00	0.00
	0.40	0.02	0.23	−0.02	0.31	−0.81	0.15	0.51	1.27	0.88	−0.20	−0.81	−0.04	0.58
	0.21	0.01	0.13	−0.02	0.15	−0.35	0.11	0.25	0.56	0.41	−0.13	−0.38	−0.04	0.29
BM	0.18	0.00	0.00	0.00	0.00	−0.79	0.00	0.01	0.04	0.39	0.00	−0.02	0.00	0.00
	0.82	0.16	0.74	0.04	0.97	−1.70	0.20	0.23	0.92	1.51	−0.24	−0.58	−0.08	0.87
	0.34	0.09	0.29	0.04	0.34	−0.59	0.12	0.13	0.40	0.63	−0.13	−0.28	−0.07	0.37
BH	0.08	0.00	0.00	0.00	0.00	−0.36	0.02	0.03	0.06	0.28	0.00	−0.03	0.00	0.00
	0.45	0.28	0.30	0.06	0.85	−0.99	0.67	0.40	1.18	1.16	−0.33	−0.77	0.01	1.08
	0.23	0.16	0.16	0.05	0.29	−0.35	0.28	0.20	0.50	0.50	−0.17	−0.34	0.01	0.46

Several studies attempt to explain that cross-autocorrelation pattern in stock returns. For example, Chordia and Swaminathan (2000) find that high volume stocks (mostly large-cap stocks) respond rapidly whereas low volume stocks (mostly small-cap stocks) respond slowly to marketwide information. It should be mentioned that those studies focus on short-horizon returns, e.g., weekly. Indeed, we show that the market premium is robust in forecasting monthly but not quarterly returns.

The term premium captures exposures related to shifts in interest rates and economic conditions that change the likelihood of default. This variable is related to the degree of risk aversion embedded in pricing, and is nonexistent in a risk-neutral world. In particular, under a rational equilibrium pricing with iso-elastic preferences, the degree of risk aversion enters security prices via the approximation (e.g., Cochrane, 2001)

$$\mathbb{E}_t[r_{i,t+1}] \approx \gamma_t \sigma_t(\Delta c_{t+1}) \rho_t(\xi_{t+1}, r_{i,t+1}), \tag{29}$$

where $\sigma_t(\Delta c_{t+1})$ and $\rho_t(\xi_{t+1}, \gamma_{i,t+1})$ are the conditional standard deviation of consumption growth and the conditional coefficient of correlation between the pricing kernel (ξ) and excess return, respectively.

Observe from Eq. (29) that expected future excess return could depend upon a time-varying risk-aversion parameter that the term premium proxies. Notice that unlike the market premium, the term premium is robust in forecasting both monthly and quarterly returns. Furthermore, it is expected to do a good job in forecasting longer horizon returns since the investor attitude towards risk (γ_t) is more likely to shift over longer horizons.

5.4. *Bayesian model averaging: out-of-sample performance*

Thus far, the analysis exhibits evidence supporting in-sample predictability of monthly and quarterly returns on size book-to-market sorted portfolios. In a related study, Bossaerts and Hillion (1999) confirm the presence of predictability using several model selection criteria. However, they discover that those criteria perform poorly out-of-sample. Is there out-of-sample stock return predictability based upon Bayesian model averaging?

This section analyzes the properties of forecast errors generated by the weighted model and other individual models that may have been selected otherwise. A good forecasting model produces out-of-sample prediction errors satisfying several important properties, including zero mean, zero serial correlation (if the prediction is one-step-ahead), and zero correlation with the predicted values (efficiency). These properties are tested using statistics advocated by West and McCracken (1998). They develop robust regression-based tests corresponding to different schemes often adopted in the forecasting literature in testing hypothesis about out-of-sample prediction error.

Our examination is based upon two schemes. The first, the *rolling* (Akgiray, 1989), fixes the estimation window size and drops distant observations as recent ones are added. To illustrate, the model parameters are first estimated with data from 1 to P (our P corresponds to the first one-third sample observations), next with data from 2 to $P+1, \ldots$, and finally with data from $T-P$ to $T-1$. The second scheme, the *recursive* (Fair and Shiller, 1990), uses all available data in the sense that the model parameters are first estimated based on data from 1 to P, next with data from 1 to $P+1, \ldots$, and finally with data from 1 to $T-1$. Due to the high dimensionality of the model space, the out-of-sample examination focuses on a single risky asset, the value-weighted CRSP index.

Table 6 displays several statistics, as explained below. We use these statistics to analyze the properties of out-of-sample forecast errors generated by several

Table 6

Bayesian model averaging: external validity

The table displays several statistics examining the properties of out-of-sample monthly forecast errors generated by several return-generating processes and the weighted forecasting model. The former set includes the all-inclusive model (All), the iid model (iid), and five models selected by adjusted R^2, AIC, SIC, FIC, and PIC, all of which are described by Bossaerts and Hillion (1999). We examine three prior specifications corresponding to a hypothetical sample size equal to 50, 100, and 25 observations per parameter. MPE is the mean forecast error. Efficiency stands for the estimated slope in the regression of forecast errors on predicted one-period-ahead returns. Serial correlation expresses the estimated slope in the regression of current on lagged forecast errors. The quantities t-statistic's are the corresponding statistics testing the equality of the forecast errors, of the correlation between forecast errors and future predicted returns (efficiency), and of serial correlations to zero. MSE is the mean squared error in percent. We adopt two different schemes having distinct asymptotic properties. The *rolling* scheme fixes the estimation window size and drops distant observations as recent ones are added. The *recursive* scheme uses all available data. The bottom part of the table displays mean squared errors for the quarterly sample corresponding to three prior scenarios, in which the number of hypothetical observations is equal to one third of the monthly counterparts, i.e., $T_0 = 17, 33$, and 8.

	$T_0=50$	$T_0=100$	$T_0=25$	All	iid	Adj R^2	AIC	SIC	FIC	PIC
The rolling scheme—monthly sample										
MPE	0.0006	0.0007	0.0003	−0.0006	0.0007	−0.0002	0.0000	−0.0023	0.0001	−0.0003
t-Statistic	0.4225	0.4944	0.2368	−0.3874	0.5126	−0.1551	0.0176	−1.5588	0.0365	−0.2117
Efficiency	−0.0563	−0.0287	−0.2335	−0.7874	−0.4371	−0.7642	−0.7919	−0.9454	−0.8709	−0.7926
t-Statistic	−0.1788	−0.0863	−0.8557	−7.8065	−1.3761	−7.4691	−6.9512	−7.9715	−7.4193	−7.2763
Serial correlation	0.0397	0.0499	0.0323	−0.0284	0.0684	−0.0043	−0.0051	0.0274	−0.0185	−0.0269
t-Statistic	0.6676	0.8326	0.5494	−0.4856	1.1288	−0.0738	−0.0895	0.5024	−0.3167	−0.4659
MSE	0.2137	0.2141	0.2139	0.2333	0.2155	0.2309	0.2298	0.2319	0.2339	0.2312
The recursive scheme—monthly sample										
MPE	−0.0003	−0.0004	−0.0003	0.0005	−0.0010	0.0007	0.0012	0.0013	0.0028	0.0020
t-Statistic	−0.1049	−0.1659	−0.1421	0.1847	−0.3924	0.3018	0.5103	0.5099	1.1455	0.8152
Efficiency	−0.2357	−0.1047	−0.4018	−0.6708	−0.4500	−0.5959	−0.5804	−0.7953	−0.7319	−0.7300
t-Statistic	−0.6675	−0.2572	−1.3455	−3.0407	−0.9504	−2.8158	−2.7292	−3.7994	−3.0175	−2.9242
Serial correlation	0.0401	0.0489	0.0372	0.0036	0.0706	0.0144	0.0143	0.0417	0.0144	0.0120
t-Statistic	0.6728	0.8079	0.6320	0.0655	1.1414	0.2597	0.2572	0.7386	0.2663	0.2194
MSE	0.2133	0.2133	0.2143	0.2231	0.2155	0.2197	0.2189	0.2260	0.2237	0.2239
MSEs for the quarterly sample										
Rolling	0.7546	0.7577	0.7651	0.9333	0.7777	0.9041	0.8629	0.8570	0.9286	0.9347
Recursive	0.7757	0.7678	0.7930	0.8312	0.7781	0.8163	0.8233	0.8952	0.8170	0.8337

individual models and by the weighted forecasting model. We study three prior scenarios corresponding to sample size equal to 25, 50, and 100 hypothetical observations per parameter. The set of individual forecasting models consists of the all-inclusive model (All), the iid model, and five models selected by adjusted R^2, AIC, SIC, FIC, and PIC, all of which are described by Bossaerts and Hillion (1999).

The primary focus is on monthly observations. The quarterly sample produces virtually identical results, and conveys no additional insights. We report only mean squared errors for the quarterly sample corresponding to three prior scenarios, in which T_0 is equal to one third of the monthly counterparts, i.e., $T_0 = 8, 17$, and 34.

The out-of-sample statistics are MPE, efficiency, serial correlation, and MSE. MPE is the mean prediction error. Efficiency stands for the estimated slope in the regression of forecast errors on predicted one-period-ahead returns. Serial correlation expresses the estimated slope in the regression of current on lagged forecast errors. MSE is the mean squared forecasting error in percent. Mean square errors for a one-period forecast are also computed by Keim and Stambaugh (1986). The quantity "*t*-statistic" is the corresponding statistic testing the equality of forecast errors, of correlations with future predicted returns, and of serial correlations to zero.

The out-of-sample forecast errors of models selected by the traditional criteria (along with the all-inclusive model) display undesirable properties. Forecasts are not efficient, in the sense that the coefficient in the regression of forecast errors on forecasted future returns is negative and statistically significant. Moreover, focusing on the monthly sample, the MSEs for the iid model based on the rolling and recursive schemes are both equal to 0.2155. The MSE associated with the optimally selected models is higher. It ranges between 0.2298 and 0.2339 based on the rolling scheme and between 0.2189 and 0.2260 based on the recursive scheme. Similar results are obtained for quarterly observations as well. In sum, model selection criteria detect no out-of-sample predictability and display poor out-of-sample performance, consistent with Bossaerts and Hillion (1999).

In contrast, the analysis shows that Bayesian model averaging has an impressive out-of-sample performance. In most cases, it produces zero mean forecasting errors, zero correlations between forecast errors and predicted future returns, and zero serial correlations. Perhaps the most striking result in Table 6 is that for every prior scenario, for every scheme examined, and for both the monthly and quarterly samples, Bayesian model averaging produces mean squared errors smaller than those corresponding to the iid model. These superior results are consistent with out-of-sample return predictability. Specifically, focusing on the monthly sample, mean squared errors generated by Bayesian model averaging range between 0.2137 and 0.2141 based on the rolling scheme and between 0.2133 and 0.2143 based on the recursive scheme. Those are the smallest mean squared errors across the various forecasting models examined.

6. Variance decomposition and asset allocation

6.1. Variance decomposition

We perform the variance decomposition of predicted future stock returns into the three components: model risk, estimation risk, and uncertainty attributed to forecast errors. The decomposition is based on current values of predictive variables (z_T) equal to actual end-of-sample realizations. We examine both the monthly and quarterly samples. For each sample, we analyze three prior specifications noted earlier. Table 7 displays results.

Based on monthly observations and $T_0 = 50$, we show that for a single-period investor, the average (across portfolios) contributions of the three components to the overall uncertainty about predicted returns are 3.5%, 2.7%, and 93.8%, respectively. Based on quarterly observations and $T_0 = 17$, such results are 9.5%, 6%, and 84.5%, respectively. Model uncertainty is larger than parameter uncertainty. Similar results are obtained for the other prior specifications.

Our conclusion about the importance of model uncertainty differs from the one suggested by Pastor and Stambaugh (1999). They estimate cost of equity capital and show that uncertainty about which asset pricing model to use is less important, on average, than within-model parameter uncertainty. In their exercise, Pastor and Stambaugh (1999) identify a large uncertainty about the equity premium, which inflates their within-model uncertainty. In this study no premium is estimated. The uncertainty about the equity premium can play a potentially important role in explaining the difference in conclusions.

What drives the magnitude of model uncertainty? As shown in Eq. (25), model uncertainty becomes more prominent the greater the dispersion of forecasted conditional expected returns across the models. This dispersion crucially depends upon the deviation of most recent values of the predictive variables from their historical means. As an extreme example, if such values are equal to their historical means, conditional expected returns are identical across models, and the model uncertainty component becomes nonexistent.

At the end-of-sample period the current values of variables perceived as indicators of fundamental values, such as book-to-market, dividend yield, earnings yield, and trend-deviation-in-wealth, deviate substantially from their sample means, giving rise to the greater impact of model uncertainty. To be precise, the current values of those variables are between 1.57 (earnings yield) and 3.03 (*cay*) standard deviations away from their corresponding sample means.

What are the implications of the sample size for both model and parameter risks? Higher frequency data provides substantially more information about the variance. Therefore, with fewer observations both parameter and model risks are expected to increase, and they do.[4] To illustrate, based on monthly observations, model and parameter risks account altogether (on average using $T_0 = 50$) for only 6.2% of the total

[4] To illustrate the impact on parameter uncertainty, consider the case where stock returns follow the iid process, i.e., $r_t = \mu + u_t$ with $u_t \sim \mathrm{N}(0, \sigma^2)$. The total variance of each period return is given by $\mathrm{var}(r_t) = \mathrm{var}(\mu) + \sigma^2$. Standard results imply that the parameter uncertainty component, $\mathrm{var}(\mu)$, is equal to σ^2/T. Parameter uncertainty increases at increasing rates with any reduction in the sample size.

Table 7
Variance decompositions
The table exhibits the marginal contribution of each source of uncertainty about predicted stock returns, i.e., model risk, estimation risk, and uncertainty attributed to forecast errors (denoted For. error), to the overall uncertainty about predicted returns. The variance components are given by

$$\mathrm{var}\{R_{T+1}|D\} = \sum_{j=1}^{2^M} \mathrm{P}(\mathscr{M}_j|D)[\mathbb{E}\{\Upsilon_j\} + \mathrm{var}\{\lambda_j\} + (\tilde{\lambda} - \mathbb{E}\{\lambda_j\})(\tilde{\lambda} - \mathbb{E}\{\lambda_j\}')].$$

where R_{T+1} is the next-period excess return, $\mathrm{P}(\mathscr{M}_j|D)$ is the posterior probability of model $j, \mathbb{E}\{\Upsilon_i\}$ and $\mathrm{var}\{\lambda_j\}$ are the forecast error and parameter uncertainty components corresponding to model j, respectively. The model uncertainty component is given by $\sum_{j=1}^{2^M} \mathrm{P}(\mathscr{M}_j|D)(\tilde{\lambda} - \mathbb{E}\{\lambda_j\})(\tilde{\lambda} - \mathbb{E}\{\lambda_j\})'$ where $\tilde{\lambda} = \sum_{j=1}^{2^M} \mathrm{P}(\mathscr{M}_j|D)\mathbb{E}\{\lambda_j\}$ is the predicted mean of the next-period excess return that incorporates model uncertainty. The decompositions are performed separately for each of six equity portfolios formed as the intersection of two size (S,B) and three book-to-market (L,M,H) groups, and are presented for both monthly and quarterly samples. For each sample, we examine three specifications of the prior sample size T_0.

Portfolio	*Monthly observations*			*Quarterly observations*		
	Estimation risk	Model risk	For. error	Estimation risk	Model risk	For. Error
	$T_0 = 50$ observations per parameter			$T_0 = 17$ observations per parameter		
SL	0.02	0.05	0.93	0.06	0.09	0.85
SM	0.03	0.08	0.89	0.07	0.11	0.82
SH	0.04	0.02	0.94	0.06	0.10	0.84
BL	0.02	0.01	0.97	0.05	0.06	0.89
BM	0.02	0.02	0.96	0.06	0.10	0.84
BH	0.03	0.03	0.94	0.05	0.11	0.84
	$T_0 = 100$ observations per parameter			$T_0 = 34$ observations per parameter		
SL	0.03	0.05	0.92	0.08	0.09	0.83
SM	0.03	0.09	0.88	0.08	0.12	0.80
SH	0.04	0.03	0.93	0.08	0.10	0.82
BL	0.02	0.02	0.96	0.07	0.07	0.86
BM	0.03	0.02	0.95	0.07	0.10	0.83
BH	0.04	0.05	0.91	0.08	0.11	0.81
	$T_0 = 25$ observations per parameter			$T_0 = 8$ observations per parameter		
SL	0.02	0.04	0.94	0.05	0.09	0.86
SM	0.03	0.07	0.90	0.06	0.11	0.83
SH	0.03	0.02	0.95	0.05	0.08	0.87
BL	0.01	0.01	0.98	0.04	0.04	0.92
BM	0.01	0.02	0.97	0.05	0.10	0.85
BH	0.03	0.02	0.95	0.04	0.09	0.87

uncertainty about future stock returns, whereas based on the quarterly counterpart, they account (on average using $T_0 = 17$) for a considerably larger fraction, 15.5%.

What are the implications of the investment horizon for model-versus-parameter risks? Parameter uncertainty increases with the investment horizon, as shown by Barberis (2000). In longer horizons, predictive variables revert to their long-run means, making conditional expected stock returns look similar across the various forecasting models. The total predictive variance attributed to model uncertainty will, therefore, converge to a fixed quantity. Consequently, the annualized predictive variance, obtained by dividing the fixed quantity by the horizon length, will diminish with an increasing horizon. One can therefore expect that short investment horizons provide a lower bound on the ratio obtained by dividing parameter uncertainty by model uncertainty.

6.2. Asset allocation

Using the framework developed in Section 3, we compute asset allocations when the current values of predictive variables (z_T) are equal to the actual end-of-sample realizations. We focus on the monthly sample (the quarterly provides no additional insights) and study three prior scenarios corresponding to 25, 50, and 100 hypothetical observations per parameter. We also examine asset allocations when the current values of predictive variables are equal to historical means, focusing on $T_0 = 50$. The buy-and-hold investment horizons range between one and ten years. As in Stambaugh (1999), the relative risk-aversion coefficient is equal to seven. Table 8 exhibits allocation to six size book-to-market portfolios, total allocation to equities (Total), and a utility loss.

This utility loss provides an economic metric for gauging the effect of ignoring model uncertainty. This metric is inspired by several recent studies, including Kandel and Stambaugh (1996) and Pastor and Stambaugh (2000). It summarizes the loss perceived by investors who are forced to ignore model uncertainty and, instead, allocate funds based on individual models that are selected otherwise. The set of individual models consists of the all-inclusive model and five other models selected by adjusted R^2, AIC, SIC, FIC, and PIC, all of which are described by Bossaerts and Hillion (1999). A utility loss is computed as the difference between two risk-free certainty equivalent returns $\mathrm{CE}^* - \mathrm{CE}^{\mathcal{M}_j}$, where

$$\mathrm{CE}^* = \{(1-\gamma)\mathbb{E}[U(W_{T+K}(\omega^*))]\}^{1/H(1-\gamma)}, \tag{30}$$

$$\mathrm{CE}^{\mathcal{M}_j} = \{(1-\gamma)\mathbb{E}[U(W_{T+K}(\omega^{\mathcal{M}_j}))]\}^{1/H(1-\gamma)}, \tag{31}$$

H is the horizon length in years, $\mathbb{E}$ is the expected value operator taken with respect to the weighted predictive distribution, ω^* and $\omega^{\mathcal{M}_j}$ are optimal allocations to equities based on the weighted model and each of the above-described single models, respectively.

Asset allocations under model uncertainty deliver three interesting observations. First, given short-sell constraints investors allocate funds only to small-cap value stocks and large-cap value stocks. Second, investors do not allocate more to equities the longer their horizons. Focusing on z_T equal to sample means, one-year and ten-year buy-and-hold investment horizons correspond to 65% and 64% of wealth invested in equities, respectively. Third, the utility loss is economically significant. Focusing on $T_0 = 50$ and z_T equal to end-of-sample actual realizations, it ranges between 1.75% and 4.37% based on the all-inclusive model, between 1.23%and 2.13% based on adjusted R^2, and between 1.71% to 3.71% based on SIC.

Table 8
Asset allocation and the economic loss of ignoring model uncertainty
The table exhibits asset allocations to six size book-to-market portfolios as percentages of the total invested wealth using three prior scenarios corresponding to a hypothetical prior sample size equal to 25, 50, and 100 observations per parameter. Asset allocations are derived for investment horizons of one, two, four, six, eight, and ten years, for relative risk-aversion coefficient (γ) equal to seven, and for current values of predictive variables (z_T) equal to actual-end-of-sample realizations. We also examine asset allocation when the current values are equal to historical means focusing on $T_0 = 50$. The table exhibits allocation to individual portfolios, total allocation to equities (Total), and a utility loss. Utility loss is computed as the loss in an annual certainty equivalent risk-free rate perceived by investors who are forced to ignore model uncertainty and, instead, allocate funds based upon several return-generating processes. The latter includes the all-inclusive model (All), and models selected by adjusted R^2, AIC, SIC, FIC, and PIC, all of which are described by Bossaerts and Hillion (1999).

Horizon	*Asset allocation*							*Utility loss*					
	SL	SM	SH	BL	BM	BH	Total	All	Adj R^2	AIC	SIC	FIC	PIC
$T_0 = 50$ observations per parameter, and z_T = end-of-sample realizations													
1	0.00	0.00	0.36	0.00	0.00	0.31	0.67	4.37	1.23	1.71	1.71	3.71	2.33
2	0.00	0.00	0.32	0.00	0.00	0.33	0.65	4.07	1.95	2.56	2.61	3.41	3.13
4	0.00	0.00	0.30	0.00	0.00	0.33	0.63	3.90	2.09	2.63	2.65	2.55	2.73
6	0.00	0.00	0.29	0.00	0.00	0.34	0.63	3.00	1.83	2.27	2.25	1.79	2.08
8	0.00	0.00	0.28	0.00	0.00	0.35	0.62	2.23	1.56	1.89	1.91	1.29	1.63
10	0.00	0.00	0.27	0.00	0.00	0.35	0.62	1.75	1.37	1.59	3.71	0.98	1.35
$T_0 = 100$ observations per parameter, and z_T = end-of-sample realizations													
1	0.00	0.00	0.31	0.00	0.00	0.34	0.65	3.99	1.00	1.45	1.45	3.35	2.03
2	0.00	0.00	0.29	0.00	0.00	0.35	0.64	3.83	1.78	2.36	2.41	3.19	2.91
4	0.00	0.00	0.27	0.00	0.00	0.36	0.63	3.68	1.94	2.46	2.47	2.38	2.55
6	0.00	0.00	0.27	0.00	0.00	0.36	0.63	2.94	1.79	2.22	2.20	1.75	2.04
8	0.00	0.00	0.28	0.00	0.00	0.34	0.62	2.20	1.54	1.87	1.89	1.27	1.61
10	0.00	0.00	0.27	0.00	0.00	0.34	0.61	1.72	1.34	1.56	3.64	0.95	1.32
$T_0 = 25$ observations per parameter, and z_T = end-of-sample realizations													
1	0.00	0.00	0.41	0.00	0.00	0.27	0.69	4.81	1.49	2.00	2.00	4.11	2.66
2	0.00	0.00	0.34	0.00	0.00	0.31	0.65	4.26	2.07	2.69	2.75	3.58	3.28
4	0.00	0.00	0.31	0.00	0.00	0.32	0.63	3.97	2.14	2.68	2.70	2.60	2.78
6	0.00	0.00	0.29	0.00	0.00	0.33	0.62	3.05	1.86	2.30	2.28	1.82	2.12
8	0.00	0.00	0.28	0.00	0.00	0.34	0.61	2.22	1.54	1.88	1.90	1.27	1.62
10	0.00	0.00	0.27	0.00	0.00	0.33	0.60	1.69	1.31	1.53	3.66	0.92	1.29
$T_0 = 50$ observations per parameter, and z_T = sample means													
1	0.00	0.00	0.29	0.00	0.00	0.36	0.65	0.15	0.17	0.27	0.27	0.10	0.15
2	0.00	0.00	0.29	0.00	0.00	0.35	0.64	0.18	0.20	0.32	0.33	0.08	0.14
4	0.00	0.00	0.29	0.00	0.00	0.36	0.64	0.17	0.21	0.28	0.26	0.05	0.08
6	0.00	0.00	0.28	0.00	0.00	0.34	0.63	0.14	0.19	0.25	0.25	0.04	0.06
8	0.00	0.00	0.28	0.00	0.00	0.33	0.62	0.15	0.19	0.22	0.25	0.04	0.09
10	0.00	0.00	0.29	0.00	0.00	0.35	0.64	0.24	0.30	0.33	0.32	0.08	0.21

Centering z_T around sample means should considerably mitigate the impact of model uncertainty, and it does. Based on this scenario, expected returns are forced to be constant across models and along the investment horizon within any model. However, even when current values are equal sample means, the annual utility loss is nonnegligible. It ranges between 0.15% and 0.24% based on the all-inclusive model and between 0.27% and 0.32% based on SIC. When expected future returns are forced to be equal across models the utility loss is attributed primarily to second moments, which differ across the models.

We document no horizon effect, whereas Barberis (2000) shows that investors do allocate substantially more to stocks the longer their horizon. Holding expected returns constant over the horizon, Barberis (2000, pp. 244–245) shows that a necessary (not sufficient) condition for the horizon effect is a negative covariance between unexpected returns and innovations in dividend yield. This could lead to a diminishing predicted variance over the investment horizon, thereby making stocks more attractive. Barberis proposes a strong economic intuition for such mean reversion when returns are predictable.

To understand the absence of horizon effect in our study, we refer the readers to Table 1. The evidence shows that covariances between unexpected returns and innovations in predictive variables are negative for dividend yield and several other economic variables. However, these "negative covariance" variables are not among the highest-posterior-probability predictors. Rather, the weighted model is biased in favor of term premium and market premium. Both variables exhibit positive covariances, leading to an increasing perceived variance of future returns over the investment horizon, thereby making equities appear less attractive for longer-horizon investors.

The disappearance of the horizon effect documented here is consistent with Heaton and Lucas (2000) and Ameriks and Zeldes (2000) who show that older people (probably shorter horizon investors) could hold more in stocks than younger cohorts. Interestingly, Ameriks and Zeldes (2000) also show that almost half of their sample members made no active changes to their portfolio allocation, i.e., those are buy-and-hold investors similar to the ones examined in our study.

7. Conclusion

In this article, we implement a Bayesian model averaging approach to analyze the sample evidence on return predictability incorporating model uncertainty. Furthermore, we study the implications of model uncertainty from investment perspectives. We obtain the following general results. First, a model that averages across various return generating processes displays robust properties. Specifically, it produces zero mean out-of-sample forecast errors that are serially uncorrelated over time.

Second, the evidence supports both in-sample and out-of-sample return predictability. However, our results show that incorporating model uncertainty can substantially weaken the predictive power of economic variables. Third, our analysis suggests that the predictive power of term and market premia is superior to that of other predictors. In particular, trend-deviation-in-wealth is a robust predictor of quarterly returns only if the shares of nonhuman wealth and labor income are computed based on quantities observed after the prediction is made.

In contrast, the feasible version of trend-deviation-in-wealth (calculated using data available only at the time of prediction) works poorly, suggesting that the impressive predictive power of that variable could be due to a look-ahead bias. Fourth, model uncertainty appears more significant than estimation risk. Finally, model uncertainty is also shown to be important to a Bayesian investor, whose utility losses are large if such uncertainty is ignored.

There are several natural extensions to our analysis. First, it can be used to examine return predictability in the context of international pricing models. It can be modified to study predictability of future economic activity. One can also examine the extent to which bond market returns are predictable. Second, our Bayesian approach is sufficiently flexible to examine the performance of nonnested models. For example, one can compute the posterior probabilities for the GARCH model and the stochastic volatility model, and arrive at an optimal composite model. Third, while not done here, the normative implications of asset allocation decisions under model uncertainty deserves further research. The Bayesian framework is especially suited for addressing this issue. Finally, it can be argued that investors will require an extra premium to compensate for model uncertainty. Studying the equity premium in a framework that accommodates model uncertainty can be a worthy objective. Much more work needs to be done on model uncertainty.

Appendix A. Proof of results and data description

A.1. Marginal likelihood

All quantities based on the hypothetical sample, denoted by the subscript 0, are expressed in terms of quantities observed from the actual sample. Specifically (the model-specific-subscript is suppressed for notational clarity)

$$\frac{1}{T_0}(X_0'X_0) = \frac{1}{T}(X'X) = \begin{bmatrix} 1 & \bar{z}' \\ \bar{z} & \bar{z}\bar{z}' + \hat{V}_z \end{bmatrix}, \tag{A.1}$$

$$\begin{aligned} X_0'R_0 &= (X_0'X_0)B_0, \\ &= \frac{T_0}{T}(X'X)\begin{bmatrix} \bar{r}' \\ 0 \end{bmatrix} \\ &= T_0\begin{bmatrix} \bar{r}' \\ \bar{z}\bar{r}' \end{bmatrix}. \end{aligned} \tag{A.2}$$

The joint posterior distribution of B and Σ based on the hypothetical sample constitutes the prior distribution for the primary sample. Assuming the standard noninformative prior $\mathrm{P}(B,\Sigma) \propto |\Sigma|^{-(N+1)/2}$ before observing the hypothetical sample, we obtain

$$\mathscr{L}(B,\Sigma;D_0) \propto |\Sigma|^{-(T_0+N+1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}[S_0+(B-B_0)'X_0'X_0(B-B_0)]\Sigma^{-1}\right), \tag{A.3}$$

where

$$\begin{aligned} S_0 &= (R_0 - X_0B_0)'(R_0 - X_0B_0) \\ &= (R_0 - \imath_{T_0}\bar{r})'(R_0 - \imath_{T_0}\bar{r}) \\ &= T_0\hat{V}_r, \end{aligned} \tag{A.4}$$

and $\imath_{T_0}$ is a $T_0 \times 1$ vector of ones. Standard results (e.g., Zellner, 1971) imply that Σ obeys the inverted Wishart distribution with a parameter matrix S_0 and $T_0 - m - 1$ degrees of freedom. Conditional on Σ, the vector $b=\mathrm{vec}(B)$ is multivariate normally distributed with mean $b_0=\mathrm{vec}(B_0)$ and variance $\Sigma \otimes (X_0'X_0)^{-1}$. The priors for B and Σ can be expressed as

$$\begin{aligned} \mathrm{P}(b|\Sigma,D_0) = {} & (2\pi)^{-N(m+1)/2}|\Sigma \otimes (X_0'X_0)^{-1}|^{-1/2} \\ & \times \exp\left(-\frac{1}{2}(b-b_0)'[\Sigma^{-1} \otimes X_0'X_0](b-b_0)\right), \end{aligned} \tag{A.5}$$

$$\mathrm{P}(\Sigma|D_0) = \psi_0|S_0|^{(T_0-m-1)/2}|\Sigma|^{-(T_0+N-m)/2}\exp\left(-\frac{1}{2}\mathrm{tr}[S_0\Sigma^{-1}]\right), \tag{A.6}$$

where

$$\psi_0 = \left(2^{(T_0-m-1)N/2}\pi^{N(N-1)/4}\prod_{i=1}^{N}\Gamma\left[\frac{T_0-m-i}{2}\right]\right)^{-1}. \tag{A.7}$$

The likelihood function (the one that integrates to unity) of normally distributed data constituting the actual sample obeys the form

$$\mathscr{L}(B,\Sigma;D) = (2\pi)^{-TN/2}|\Sigma|^{-T/2}\exp\left(-\frac{1}{2}\mathrm{tr}[S+(B-\hat{B})'X'X(B-\hat{B})]\Sigma^{-1}\right), \tag{A.8}$$

where

$$S = (R - X\hat{B})'(R - X\hat{B}), \tag{A.9}$$

$$\hat{B} = (X'X)^{-1}X'R. \tag{A.10}$$

Combining the likelihood in Eq. (A.8) and the prior in Eq. (A.3) and completing the square on b yield

$$\begin{aligned} \mathrm{P}(b|\Sigma,D) = {} & (2\pi)^{-N(m+1)/2}|\Sigma \otimes (X_0'X_0 + X'X)^{-1}|^{-1/2} \\ & \times \exp\left(-\frac{1}{2}(b-\tilde{b})'[\Sigma^{-1} \otimes (X_0'X_0 + X'X)](b-\tilde{b})\right), \end{aligned} \tag{A.11}$$

$$\mathrm{P}(\Sigma|D) = \psi|\tilde{S}|^{v/2}|\Sigma|^{-(v+N+1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}[\tilde{S}\Sigma^{-1}]\right), \tag{A.12}$$

where

$$\tilde{b} = \text{vec}(\tilde{B}), \tag{A.13}$$

$$\tilde{B} = (X_0'X_0 + X'X)^{-1}(X_0'R_0 + X'R), \tag{A.14}$$

$$\tilde{S} = R'R + S_0 + R_0'X_0(X_0'X_0)^{-1}X_0'R_0 - \tilde{B}'(X_0'X_0 + X'X)\tilde{B}, \tag{A.15}$$

$$\psi = \left(2^{vN/2}\pi^{N(N-1)/4}\prod_{i=1}^{N}\Gamma\left[\frac{v+1-i}{2}\right]\right)^{-1}, \tag{A.16}$$

$$v = T_0 + T - m - 1. \tag{A.17}$$

Computing the log marginal likelihood is as follows. Take logs from both sides of Eq. (9) and then substitute for the prior, likelihood, and posterior densities the corresponding quantities given in Eqs. (A.5)–(A.17).

A.2. The joint posterior distribution of Φ and Ψ

To derive the posterior distribution of Φ and Ψ, we follow Kandel and Stambaugh (1996) and assume that the prior sample produces the same values as the actual counterpart for the statistics corresponding to ρ and $\tilde{z}$, where

$$\rho = \frac{1}{T}\sum_{t=0}^{T-1}(z_t - \bar{z})(z_{t+1} - \bar{z})' \tag{A.18}$$

is the matrix of autocorrelation and cross autocorrelation of m predetermined variables and $\tilde{z} = (1/T)\sum_{t=1}^{T} z_t$. The joint posterior distribution of Φ and Ψ is given by

$$\text{vec}(\Phi)|\Psi \sim \text{N}(\text{vec}(\Phi_0), \Psi \otimes (X_0'X_0)^{-1}), \tag{A.19}$$

$$\Psi \sim \text{IW}(\Psi_0, T_0 - (N+m) - 1), \tag{A.20}$$

where

$$\begin{aligned}\Phi_0 &= (X_0'X_0)^{-1}(X_0'Y_0), \\ &= [B_0, (X_0'X_0)^{-1}(X_0'Z_0)],\end{aligned} \tag{A.21}$$

$$X_0'Z_0 = T_0\begin{bmatrix} \tilde{z}' \\ \rho + \bar{z}\tilde{z}' \end{bmatrix}, \tag{A.22}$$

$$\Psi_0 \approx T_0\begin{bmatrix} \hat{V}_r & V \\ V & \hat{V}_z + \tilde{z}\tilde{z}' - \frac{1}{T_0}Z_0'X_0(X_0'X_0)^{-1}X_0'Z_0 \end{bmatrix}. \tag{A.23}$$

The approximation becomes equality if the first and last observations of the predictive variables are equal. The off-diagonal matrix V is assumed zero. Combining the joint prior distribution in Eqs. (A.19) and (A.20) with normally distributed data constituting the primary sample, the posterior distributions for $\phi = \mathrm{vec}(\Phi)$ and Ψ are obtained as

$$\phi|\Psi, D \sim \mathrm{N}(\tilde{\phi}, \Psi \otimes (X_0'X_0 + X'X)^{-1}), \tag{A.24}$$

$$\Psi|D \sim \mathrm{IW}(\tilde{\Psi}, T + T_0 - (N+m) - 1), \tag{A.25}$$

where

$$\tilde{\phi} = \mathrm{vec}(\tilde{\Phi}), \tag{A.26}$$

$$\tilde{\Phi} = (X_0'X_0 + X'X)^{-1}(X_0'Y_0 + X'Y), \tag{A.27}$$

$$\tilde{\Psi} = Y'Y + \Psi_0 + Y_0'X_0(X_0'X_0)^{-1}X_0'Y_0 - \tilde{\Phi}'(X_0'X_0 + X'X)\tilde{\Phi}. \tag{A.28}$$

A.3. The conditional distribution of future stock returns

Partitioning Eq. (17) yields

$$(r_t', z_t') = (1, z_{t-1}')\begin{bmatrix} b' & a' \\ C' & A \end{bmatrix} + \begin{pmatrix} \varepsilon_t \\ e_t \end{pmatrix}', \tag{A.29}$$

where

$$\begin{pmatrix} \varepsilon_t \\ e_t \end{pmatrix} \sim \mathrm{N}\left(0, \quad \begin{bmatrix} \Sigma & \Lambda \\ \Lambda' & \Phi \end{bmatrix}\right). \tag{A.30}$$

It follows from Eq. (A.29) that

$$r_{T+1} = b + Cz_T + \varepsilon_{T+1}, \tag{A.31}$$

$$z_{T+1} = a + A'z_T + e_{T+1}. \tag{A.32}$$

The cumulative excess log return over the investment horizon is computed as

$$\begin{aligned} R_{T+K} &= \sum_{k=1}^{K} r_{t+k} \\ &= Kb + C\left(\sum_{j=1}^{K} z_{T+j-1}\right) + \sum_{j=1}^{K} \varepsilon_{T+j}, \end{aligned} \tag{A.33}$$

where z_{T+j} is obtained by iterating over Eq. (A.32). In particular,

$$z_{T+J} = [(A')^J - I_m](A' - I_m)^{-1}a + (A')^J z_T + \sum_{j=1}^{J} (A')^{J-j} e_{T+j}. \tag{A.34}$$

Substituting Eq. (A.34) into Eq. (A.33) for $J = 1, \ldots, K-1$ yields

$$\begin{aligned} R_{T+K} = {} & Kb \\ & + C[A'((A')^{K-1} - I_m)(A' - I_m)^{-1} - (K-1)I_m](A' - I_m)^{-1}a \\ & + C((A')^K - I_m)(A' - I_m)^{-1}z_T + \sum_{j=2}^{K}\sum_{i=1}^{j-1} C(A')^{j-i-1}e_{T+i} + \sum_{j=1}^{K}\varepsilon_{T+j}, \end{aligned} \tag{A.35}$$

for $K \geqslant 2$. The desired result follows immediately.

A.4. Variance decomposition

Based on Leamer (1978), decomposing the predictive variance $\mathrm{Var}\{R_{T+K}|D\}$ with respect to the model space and using the law of iterated expectations yield

$$\mathrm{Var}\{R_{T+K}|D\} = \sum_{j=1}^{2^M} \mathrm{P}(\mathscr{M}_j|D)[\mathrm{Var}\{R_{T+K}|\mathscr{M}_j, D\} + (\mathbb{E}\{\lambda_j\} - \tilde{\lambda})(\mathbb{E}\{\lambda_j\} - \tilde{\lambda})']. \tag{A.36}$$

Decomposing the within-model variance yields

$$\begin{aligned} \mathrm{Var}\{R_{T+K}|\mathscr{M}_j, D\} & = \mathbb{E}\{\mathrm{Var}[R_{T+K}|\mathscr{M}_j, D, \Phi, \Psi]\} + \mathrm{Var}\{\mathbb{E}[R_{T+K}|\mathscr{M}_j, D, \Phi, \Psi]\} \\ & = \mathbb{E}\{\Upsilon_j\} + \mathrm{Var}\{\lambda_j\}. \end{aligned} \tag{A.37}$$

The three variance components are obtained by substituting the two components of within-model uncertainty into the corresponding quantity in Eq. (A.36).

A.5. Data

Data used to compute dividend yield, Treasury bill rate, and market premium are from the Center for Research in Security Prices (CRSP) at the University of Chicago. Inputs for calculating book-to-market are obtained from the Standard & Poor's publication: "Security Price Index Record—Statistical Service". Inputs for computing default risk spread are obtained form Citibase. Data on term premium and default risk premium are from Ibbotson and associates. Returns on size book-to-market portfolios, size premium, and value premium are from Kenneth French. The winners-minus-losers portfolio is from Mark Carhart. Earnings and inflation data are from Robert Shiller. Earnings yield is formed by dividing the most recent twelve-month earnings by the contemporaneous value of the S&P 500 index. Treasury yields are taken from the Federal Reserve Board. In the quarterly sample, the trend deviation in wealth replaces January Dummy. In-sample and out-of-sample trend-deviation-in-wealth are from Martin Lettau and Sydney Ludvigson.

Table 9 presents descriptive statistics based on the actual sample spanning 549 months from April 1953 to December 1998 for continuously compounded returns on six equity portfolios and 13 predictors. One can see that dividend yield, book-to-market, earnings yield, default spread, Treasury-bill rate, and term spread display

Table 9
Descriptive statistics
The table shows descriptive statistics based on the actual sample spanning 549 months from April 1953 to December 1998 for continuously compounded returns on six equity portfolios and 13 predictors. The portfolios are identified by a combination of two letters designating increasing values of size (S,B) and book-to-market (L,M,H). The 13 predictors are: dividend yield on the value-weighted NYSE index (Div); book-to-market (BM) on the Standard & Poor's Industrials; earnings yield on the Standard & Poor's Composite index (EY); the one-year momentum portfolio (WML); the difference in annualized yields of Moody's Baa and Aaa rated bonds (Def); the monthly rate of a three-month Treasury bill (Tbill); excess return on the value-weighted index (RET); the difference between the return on long-term corporate bonds and the return on long-term government bond (DEF); the difference between the monthly return on long-term government bond and the one-month Treasury bill rate (TERM); inflation rate (Inf); size premium (SMB); value premium (HML); and the difference in annualized yield of ten-year and one-year Treasury bills (Term). Std.Dev. denotes the standard deviation. The parameter ρ_t is the sample autocorrelation at lag t months.

Statistic	Mean	Std.Dev.	ρ_1	ρ_3	ρ_6	ρ_{12}	ρ_{60}
Predictive variables							
Div	0.0362	0.0091	0.9828	0.9478	0.8847	0.7620	0.3276
BM	0.5048	0.1735	0.9889	0.9674	0.9304	0.8572	0.4912
EY	0.8531	0.2936	0.9929	0.9679	0.9162	0.7981	0.3638
WML	0.0097	0.0357	−0.0377	−0.1016	0.0706	0.2347	0.2293
Def	0.9476	0.4385	0.9738	0.9106	0.8360	0.6941	0.3859
Tbill	0.0044	0.0024	0.9565	0.9113	0.8638	0.7818	0.4258
RET	0.0063	0.0423	0.0655	0.0041	−0.0650	0.0312	−0.0504
DEF	0.0003	0.0115	−0.1881	−0.0493	−0.0434	0.0054	0.0088
TERM	0.0011	0.0263	0.0662	−0.1037	0.0452	−0.0107	−0.0242
Inf	0.3330	0.3334	0.5541	0.4755	0.4416	0.5152	0.2929
SMB	0.0009	0.0262	0.1659	−0.0134	0.0708	0.1871	0.0305
HML	0.0039	0.0244	0.1483	−0.0077	0.0430	0.1013	0.0063
Term	0.7195	0.9908	0.9589	0.8368	0.7033	0.5071	0.0217
Equity portfolios							
SL	0.0098	0.0614	0.1722	−0.0242	−0.0237	0.0085	−0.0401
SM	0.0130	0.0501	0.1854	−0.0122	−0.0010	0.0694	0.0034
SH	0.0149	0.0509	0.1795	−0.0275	−0.0132	0.1272	0.0482
BL	0.0108	0.0451	0.0571	0.0013	−0.0665	0.0535	−0.0770
BM	0.0110	0.0399	0.0129	0.0127	−0.0660	0.0057	−0.0262
BH	0.0132	0.0434	0.0443	0.0244	−0.0214	0.0544	0.0026

persistence, whereas WML, excess return, default risk premium, term premium, inflation, size premium, and value premium possess lower or no autocorrelation.

References

Ait-Sahalia, Y., Brandt, M., 2001. Variable selection for portfolio choice. Journal of Finance 56, 1297–1351.

Akgiray, V., 1989. Conditional heteroskedasticity in time series of stock returns: evidence and forecasts. Journal of Business 62, 55–80.

Ameriks, J., Zeldes, S., 2000. How do household portfolio shares vary with age? Unpublished working paper. Columbia University.

Avramov, D., 2001. Stock return predictability and asset pricing models. Unpublished working paper. University of Maryland.

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

Bawa, V., Brown, S., Klein, R., 1979. Estimation Risk and Optimal Portfolio Choice. North-Holland, Amsterdam.

Blume, M., Stambaugh, R., 1983. Biases in computed returns: an application to the size effect. Journal of Financial Economics 12, 387–404.

Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: what do we learn? Review of Financial Studies 12, 405–428.

Brandt, M., 1999. Estimating portfolio and consumption choice: a conditional Euler equations approach. Journal of Finance 54, 1609–1645.

Brennan, M.J., Xia, Y., 2001. Assessing asset pricing anomalies. The Review of Financial Studies, 14, 905–942.

Campbell, J., 1987. Stock returns and the term structure. Journal of Financial Economics 18, 373–399.

Campbell, J., 1996. Understanding risk and return. Journal of Political Economy 104, 572–621.

Campbell, J., Shiller, R., 1988a. The dividend–price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–227.

Campbell, J., Shiller, R., 1988b. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

Carhart, M.M., 1997. On the persistence in mutual fund performance. Journal of Finance 52, 57–82.

Chen, N.F., Roll, R., Ross, S., 1986. Economic forces and the stock market. Journal of Business 59, 383–404.

Chordia, T., Swaminathan, B., 2000. Trading volume and cross-autocorrelations in stock returns. Journal of Finance 55, 913–935.

Cochrane, J.H., 2001. Asset Pricing. Princeton University Press, Princeton, NJ.

Cremers, M., 2000. Stock return predictability: a Bayesian model selection perspective. Unpublished working paper, New York University.

Fair, R.C., Shiller, R., 1990. Comparing information in forecasts from econometric models. American Economic Review 80, 375–389.

Fama, E., French, K., 1988. Permanent and temporary components of stock prices. Journal of Political Economy 96, 246–273.

Fama, E., French, K., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 19, 3–29.

Fama, E., French, K., 1993. Common risk factors in the returns on stock and bonds. Journal of Financial Economics 33, 3–56.

Fama, E., Schwert, G.W., 1977. Asset returns and inflation. Journal of Financial Economics 5, 115–146.

Ferson, W., Harvey, C., 1991. The variation in economic risk premiums. Journal of Political Economy 99, 385–415.

Ferson, W., Harvey, C., 1999. Conditioning variables and the cross section of stock returns. Journal of Finance 4, 1325–1360.

French, K., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Frost, P., Savarino, J., 1986. An empirical Bayes approach to efficient portfolio selection. Journal of Financial and Quantitative Analysis 21, 293–305.

Goetzmann, W.N., Jorion, P., 1993. Testing the predictive power of dividend yields. Journal of Finance 48, 663–679.

Heaton, J., Lucas, D., 2000. Portfolio choice and asset prices: the importance of en-trepreneurial risk. Journal of Finance 55, 1163–1198.

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

Jorion, P., 1985. International portfolio diversification with estimation risk. Journal of Business 58, 259–278.
Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset allocation perspective. Journal of Finance 51, 385–424.
Kass, R., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.
Keim, D., 1983. Size related anomalies and stock return seasonality: further empirical evidence. Journal of Financial Economics 1, 13–32.
Keim, D., Stambaugh, R.F., 1986. Predicting returns in the stock and the bond markets. Journal of Financial Economics 17, 357–390.
Kirby, C., 1997. Measuring the predictable variation in stock and bond returns. Review of Financial Studies 10, 579–630.
Kothari, S.P., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: a time series analysis. Journal of Financial Economics 44, 169–203.
Leamer, E.E., 1978. Specification Searches: Ad Hoc inference with Non-experimental data. Wiley, New York.
Lettau, M., Ludvigson, S., 2001. Consumption, aggregate wealth, and expected stock returns. Journal of Finance 56, 815–849.
Liew, J., Vassalou, M., 2000. Can book-to-market, size, and momentum be risk factors that predict economic growth? Journal of Financial Economics 57, 221–245.
Lo, A., MacKinlay, A.C., 1990. When are contrarian profits due to stock market overreaction? Review of Financial Studies 3, 175–205.
Lo, A., MacKinlay, A.C., 1997. Maximizing predictability in the stock and bond markets. Macroeconomic Dynamics 1, 131–170.
Merton, R., 1973. An intertemporal asset pricing model. Econometrica 41, 867–887.
Pastor, L., 2000. Portfolio selection and asset pricing models. Journal of Finance 55, 179–223.
Pastor, L., Stambaugh, R.F., 1999. Costs of equity capital and model mispricing. Journal of Finance 54, 67–121.
Pastor, L., Stambaugh, R.F., 2000. Comparing asset pricing models: an investment perspective. Journal of Financial Economics 56, 335–381.
Pesaran, M.H., Timmermann, A., 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.
Robbins, H., 1955. An empirical Bayes approach to statistics. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 157–164.
Robbins, H., 1964. The empirical Bayes approach to statistical problems. Annals of Mathematical Statistics 35, 1–20.
Schwert, G.W., 1990. Stock returns and real activity: a century of evidence. Journal of Finance 45, 1237–1257.
Shanken, J., 1987. A Bayesian approach to testing portfolio efficiency. Journal of Financial Economics 19, 195–215.
Shanken, J., 1990. Intertemporal asset pricing: an empirical investigation. Journal of Econometrics 45, 99–120.
Stambaugh, R.F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.
Wang, Z., 2001. A shrinkage theory of asset allocation using asset pricing models. Unpublished working paper, Columbia University.
West, K.D., McCracken, M.W., 1998. Regression-based tests of predictive ability. International Economic Review 39, 817–840.
Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. Wiley, New York.
Zellner, A., Chetty, A.V., 1965. Prediction and decision problems in regression models from the Bayesian point of view. Journal of the American Statistical Association 60, 608–615.