

# Real-time Portfolio Choice Implications of Asset Pricing Models

Francisco Barillas and Jay Shanken<sup>1</sup>

August 2018

## Abstract

A plethora of asset pricing factors have been proposed in the literature. We study the problem of an investor who is confronted with this “zoo of factors” and wishes to find an optimal portfolio. We propose a Bayesian asset allocation framework that accounts for uncertainty about the correct pricing model. This entails an optimal degree of economic shrinkage that is beneficial for portfolio performance. Under a wide range of beliefs about the extent of mispricing, we find that considering all asset pricing models that can be formed from a given set of factors leads to real-time performance that is superior to that of the sample tangency portfolio. The superiority in out-of-sample performance is even stronger when some of the factors are redundant, as might be the case when a factor has been data mined.

---

<sup>1</sup>Barillas is from Emory University. Shanken is from Emory University and the National Bureau of Economic Research. Corresponding author: Francisco Barillas, Goizueta Business School, Emory University, 1300 Clifton Road, Atlanta, GA 30322, USA; E-mail: [Francisco.barillas@emory.edu](mailto:Francisco.barillas@emory.edu).

# 1 Introduction

In the classic world of the Sharpe-Lintner-Treynor capital asset pricing model (CAPM), the investment decision of investors is simple. Each investor holds a mean-variance efficient portfolio, a combination of the value-weighted market portfolio of risky assets and a position (long or short) in the riskless asset. However, the CAPM restriction on expected excess returns, which requires that all security “alphas” are zero, is easily rejected by the data when the market factor is challenged to price various traded factors. It follows that positions in these factors, along with the market and the riskless asset, can be used to obtain portfolios that will be preferred by investors with mean-variance preferences.

Many models with traded factors have been proposed by asset-pricing researchers. For example, Fama and French (1993) suggest a three-factor model (henceforth FF3) that includes a size spread and a value spread along with the market. Some researchers, starting with Carhart (1997), add a fourth momentum factor based on the findings of Jegadeesh and Titman (1993). In more recent work, Fama and French (2015, 2016) and Hou, Xue and Zhang (2015), appealing to the discounted cash flow relation and q-theory, respectively, consider factor pricing models that include profitability and investment spreads. The papers differ in the precise implementation of these factors, however, and also on which factors are viewed as redundant.

A redundant factor is one that has zero alphas on a given models factors and, therefore, need not be included in the model. As discussed by Barillas and Shanken (2017, 2018), concerns about factor redundancy stem from the desire for parsimony expressed in the literature. The expected return predicted by a model is not altered by adding a redundant factor, so the simpler model that leaves that factor out is preferred. From a portfolio investment perspective, the issue of whether a factor is included might, at first glance, seem to be a matter of indifference. After all, in theory zero alphas in regressions on the model factors imply that the (squared) Sharpe ratio will not increase if we allow for a position in the redundant factor. Therefore, given a set of potential investment factors, failure to exclude a redundant factor from consideration simply means that the weight on this factor will be zero in the optimal portfolio. It also implies that taking a position in that factor must result

in a deviation from optimality. Consequently, in practice, as we will see, consideration of a redundant factor in estimating optimal portfolio weights introduces noise that can ultimately reduce out-of-sample performance.

Of course, one does not know for sure which factors are redundant or, equivalently, which is the best pricing model. Theory, rational or behavioral, might suggest that certain factors are redundant. But the theory could be wrong. Alternatively, a factor might appear to be important based on past data, but this could be a random type I error (false rejection of a zero null) or a product of data mining. In recent empirical work, both Fama and French (2015) and Hou, Xue and Zhang (2015) find that the FF3 value factor is redundant, while the latter study also concludes that momentum is redundant. However, building on the findings of Asness and Frazzini (2013), Barillas and Shanken (2018) conclude that both value and momentum are important when a more “timely” version of the value factor is considered, along with size, profitability, investment, and the market.

Thus, in practice, an investor faces material uncertainty as to which combinations of factors should be included in a model. In other words, an investor is confronted with “model uncertainty” in addition to “parameter uncertainty” and the fundamental uncertainty associated with risky financial securities. The Bayesian statistical framework incorporates this broad perspective on risk in the decision-making process, given prior probabilities for each model in question and prior beliefs about each model’s parameters. Portfolio decisions are then based on the “predictive distribution” of returns, which takes into account posterior beliefs about the model probabilities and model parameters, given the prior and the observed data.

Several Bayesian studies explore the effects of estimation risk in analyzing return models (see the review article by Avramov and Zhou (2010) and Brandt (2010)), but few address model uncertainty. Two papers that do are Avramov (2002) and Cremers (2002), which examine time-variation in asset expected returns as a function of economy-wide predictors using Bayesian model averaging to explore uncertainty about the true set of predictors. With regard to factor pricing models, a series of papers starting with the early work of Shanken (1987), Harvey and Zhou (1990) and McCulloch and Rossi (1991), and later papers by Pastor (2000) and Pastor and Stambaugh (2000) impose priors over asset alphas (or economically

relevant functions of alphas) in evaluating pricing models. The last two papers also consider implications for portfolio decisions, the focus of our study.

In particular, Pastor and Stambaugh (2000) compare a pair of pricing models, the FF3 factor pricing model and a related characteristics-based model. The research exercise here is not one of trying to determine which is the “better” model, however. Rather, they evaluate the utility implications of an investor’s belief about which is the true model. Specifically, they compute the reduction in certainty equivalent return an investor would experience if, rather than hold the optimal portfolio under the model believed to be valid, she were forced to hold the “optimal” portfolio implied by a belief in the other model. Pastor and Stambaugh do discuss a potential extension to the model uncertainty setting, but do not provide a method for obtaining posterior model probabilities. Barillas and Shanken (2018) develop such a methodology for comparing factor pricing models with traded factors. Although their model probabilities are of interest in themselves when a researcher has no explicit utility function, the aim of our paper is to explore the consequences of these model probabilities for real-time factor-based portfolio investment with model averaging.

In our empirical application, we use models that combine many prominent factors from the literature. In addition to the market excess return, Mkt, we include the book-to-market or “value” factor, HML (high-low), and a size factor, SMB (small-big) from the influential three-factor model of Fama and French (1993), hereafter FF3. We also consider the UMD (up minus down) momentum factor introduced by Carhart (1997) and examine factors from the recently proposed five-factor model of Fama and French (2015), hereafter FF5. These are RMW (robust minus weak), which is based on the profitability of firms, and CMA (conservative minus aggressive), related to firms’ net investments. Hou, Xue, and Zhang (2015a, 2015b), henceforth HXZ, propose their own versions of the size (ME), investment (IA), and profitability (ROE) factors, which we also examine. Finally, we consider the value factor HMLm from Asness and Frazzini (2013), which is based on book-to-market rankings that use the most recent monthly stock price in the denominator. In total, we have 10 factors in our analysis. As in Barillas and Shanken (2018), we structure the prior on the different models so as to recognize that several of the factors are different versions of the same underlying construct. Therefore, we only consider models that contain at most one of

the factors in each of the following categories: size (SMB or ME), profitability (RMW or ROE), value (HML or HMLm), and investment (CMA or IA).

Using a longer data set from 1967 to 2016, we find, like the earlier paper, that the model with the highest posterior probability is the six-factor model {Mkt HMLm UMD ROE SMB IA}. Interestingly, models that receive non-negligible posterior probability all include ROE, HMLm and UMD in addition to the market factor. Therefore, value is no longer a redundant factor when the more timely value factor HMLm is used. We also document that inferences based on posterior model probabilities are fairly stable across priors motivated by a market efficiency perspective and others that allow for large departures from efficiency. However, this is not the case for the portfolio implications which are significantly different across prior specifications.

Using the entire sample, the Sharpe ratio of the optimal allocation for a mean-variance investor who uses the Bayesian model averaging procedure (BMA) is just below 0.5 on a monthly basis, over a wide range of prior beliefs. As a reference point, the 10-factor sample tangency portfolio has a Sharpe ratio of 0.55 on a monthly basis, by construction the highest in-sample value with this data. Therefore, our Bayesian procedure achieves similar Sharpe ratios to those of the sample tangency portfolio, but it does so with much more reasonable portfolio characteristics. This is evident from the significant differences in the means and standard deviations of the optimal portfolio returns. For instance, with sample based means and variances-covariances, the optimal portfolio has a mean annual return of 144.3% with standard deviation 75.7% for an investor with relative risk aversion 2.53. In contrast, the portfolio of the Bayesian investor who believes a priori that the non-market factors can increase the Sharpe ratio of the market by 50% has a less extreme mean of 69.1% and standard deviation 40.7%.

In addition, the portfolio positions for the investor who averages across all models are substantially more moderate than the sample-based positions, which we find are driven primarily by the shrinkage of each model's factor alphas towards zero. An important property of the optimal portfolio is that the amount invested in the riskfree asset increases with the degree of shrinkage. In fact, with risk aversion of  $\gamma = 2.53$ , the sample-based optimal portfolio, which by definition imposes no shrinkage, is highly levered, borrowing 1.47 dollars

per unit of wealth invested. The Bayesian investor also uses leverage, but to a much smaller extent: 52 cents per unit of wealth.

Such high in-sample Sharpe ratios beg the question of whether an investor could have realized such remarkable performance in real-time. We therefore conduct an out-of-sample exercise in which, after an initial period of 20 years, the investor computes the optimal portfolio only using the data available at a given point in time. As expected, the out-of-sample Sharpe ratios are smaller than the in-sample values, but now the ratios for the sample-based portfolio and the BMA procedure are comparable. Similar to the in-sample results, the Bayesian optimal portfolio employs less leverage and has more moderate positions.

We also compute certainty equivalent returns (CER) to assess the economic significance of our portfolio selection rules. We find that, despite the similar out-of-sample Sharpe ratios, the Bayesian procedure substantially outperforms in terms of CERs. In fact, the sample-based optimal portfolio in real time has negative CERs and large volatility because of its extreme positions. Thus, a Bayesian investor who takes into account plausible beliefs about asset pricing models achieves greater utility than the naive investor who simply uses the sample means and variances-covariances to form optimal portfolios.

We also conduct an experiment that is meant to recreate a situation in which some of the factors are redundant. To do so, we modify some of the factor data to have zero alphas in the out-of-sample evaluation period. This exercise serves to highlight the advantages of our procedure in situations where the factors are likely to have been data mined or already exploited, arbitraging away the original alpha. In this experiment, the investor initially believes that all of the factors could be essential. Subsequently, we examine the extent to which whether our procedure allows the investor to gradually learn in real-time that alpha is now zero for some of the factors. As with the original data, the results in this exercise show that the Bayesian procedure substantially outperforms the sample-based portfolio. Thus, the BMA Sharpe ratios are much higher and the CERs are positive for the more conservative priors but very negative for the sample-based portfolio.

The rest of the paper is organized as follows. Section 2 outlines our portfolio choice framework as well as the Bayesian procedure that accomodates model uncertainty. Section 3

summarizes our empirical results. In Section 4, we describe the experiment that highlights the advantages of our procedure in situations where the factors are likely to have been data mined or already exploited. Section 5 summarizes our main conclusions. Proofs and additional material are provided in the Appendix.

## 2 Framework

This section is divided into three parts. In the first part, we describe the portfolio selection problem and the methodology used to obtain optimal allocations, taking into account model uncertainty. The second part describes the Bayesian inference procedure that we use to compute posterior probabilities for each of the models under consideration. This part describes the assumptions on the return generating process for the factor returns and the prior distribution on the model parameters, which we use to compute the marginal likelihood of each model. The third and last part of this section outlines our method for calculating the first and second moments of the predictive distribution of returns.

### 2.1 Portfolio Selection Problem

There is a riskless asset, with rate of return  $R_{f,t}$  and  $K$  risky factors with return  $f_t$  at time  $t$ . We denote the  $K \times 1$  mean vector and  $K \times K$  variance-covariance matrix of  $f_t$  as  $\mu$  and  $V$ . The investor is risk averse with mean-variance preferences and allocates funds between the riskless asset and an efficient portfolio of the factors. Let  $W_t$  denote the investor's wealth at time  $t$ . The investor allocates  $Z_{i,t}$  to each of the factors, which are zero-cost portfolios (either excess returns such as the market factor, Mkt, or long-short factors such as HML). Since the net investment in  $f_t$  is zero, we also have  $W_t$  invested in the riskless asset. Therefore, the total portfolio return is given by

$$R_{p,t+1} = \frac{Z_t f_{t+1} + W_t R_{f,t}}{W_t}$$

Letting  $w_t = Z_t/W_t$  denote the  $K \times 1$  vector of portfolio weights on the  $K$  factors, the investor chooses  $w$  to maximize the mean-variance objective function

$$U(w) = E(R_{p,t+1}) - \frac{1}{2}\gamma \text{var}(R_{p,t+1}) \tag{1}$$

where  $\gamma$  is the coefficient of relative risk aversion. The optimal portfolio of risky assets  $w$  is given by

$$w = \frac{1}{\gamma} V^{-1} \mu \quad (2)$$

which is the well-known solution to the tangency portfolio of risky assets, i.e.,  $w$  is proportional to the weights in the risky portfolio having the maximum Sharpe ratio.

As in Pastor and Stambaugh (2000), the asset corresponding to a factor spread in this context consists of \$1 on each side of the spread plus \$1 invested in the riskless asset. The factors then equal the excess returns on these assets. Since the overall portfolio weights must sum to one and the market factor is long the market and short the riskless asset, it follows that the net weight on the riskless asset, which we denote  $w_{rf}$ , is one minus the weight on Mkt,  $w_{mkt}$ .

In practice both  $\mu$  and  $V$  are unknown. A common approach is to simply substitute the sample estimates  $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T f_t$  and  $\hat{V} = \frac{1}{T} \sum_{t=1}^T (f_t - \hat{\mu})(f_t - \hat{\mu})'$ . This is sometimes referred to in the portfolio choice literature as the “plug-in” approach. However, this procedure ignores estimation risk since it assumes that the sample estimates equal the true parameters. As often posited in the literature, the factor returns are taken to be multivariate normally distributed with parameters  $\theta = (\mu, V)$ .

In the Bayesian approach,  $\theta$  is unknown and the investor chooses  $w$  to maximize expected utility

$$\int U(w) P(F_{t+1}|D) dF_{t+1}, \quad (3)$$

where  $D$  denotes the data and  $P(F_{t+1}|D)$  is the predictive density based on posterior beliefs about  $\theta$ . The posterior is proportional to the normal likelihood times the prior,  $P(\theta)$ , on these parameters.

$$P(\theta|D) \propto P(D|\theta)P(\theta) \quad (4)$$

The predictive density is then computed as

$$P(F_{t+1}|D) = \int P(F_{t+1}|\theta, D)P(\theta|D)d\theta \quad (5)$$

and the optimal portfolio weights are easily obtained by substituting the first two moments of the Bayesian predictive distributions in (2). Under the standard diffuse (noninformative)



multivariate prior on  $\mu$  and  $V$

$$P(\theta) \propto |V|^{\frac{K+1}{2}} \quad (6)$$

and assuming jointly normal returns, the predictive distribution follows a multivariate  $t$ -distribution with  $T - K$  degrees of freedom (see e.g. Brown (1976), Klein and Bawa (1976), and Stambaugh (1997)) and the tangency portfolio weights are essentially the same as the weights obtained under the “plug-in” approach that substitutes the sample estimates. We use this sample-based approach as a benchmark.

Our objective is to use Bayesian methods to incorporate beliefs about the restrictions implied by different asset pricing models in the portfolio choice. This allows us to account for uncertainty about the true model in addition to parameter uncertainty. Pastor (2000) and Pastor and Stambaugh (2000) also incorporate restrictions implied by an asset pricing model via Bayesian methods. Both papers study optimal allocations with prior beliefs centered at a particular pricing model. As they recognize, while their approach takes into account parameter uncertainty, it does not reflect model uncertainty, implicitly assuming that the model used for portfolio choice is the “best” one. In contrast our investor, in arriving at an optimal portfolio, entertains several models a priori and explicitly takes into account uncertainty about the best model that remains after learning from the data.

Our investor considers  $J$  different asset pricing models ( $M_1, \dots, M_J$ ). Each model imposes different “alpha” restrictions for the factors under consideration. We show how to compute the predictive distribution of returns for model  $j$ ,  $P(F_{t+1}|M_j, D)$ , in subsection 2.3. The posterior probability of model  $j$ ,  $P(M_j|D)$ , is computed using Baye’s rule as:

$$P(M_i|D) = \frac{ML_i P(M_i)}{\sum_j ML_j P(M_j)}, \quad (7)$$

where  $P(M_j)$  is the prior probability assigned to model  $j$  and  $ML_j$  is the marginal likelihood of model  $j$ . The marginal likelihood is a measure of the evidence for a particular model and is computed as follows:

$$ML_j = \int P(F|\theta_j, M_j)P(\theta_j|M_j)d\theta_j \quad (8)$$

where  $P(F|\theta_j, M_j)$  is the likelihood function and  $P(\theta_j|M_j)$  the prior density, both for model  $j$ . Subsection 2.2 describes the prior specification and the methodology for computing these terms as developed in Barillas and Shanken (2018).

The predictive distribution that accounts for model uncertainty is

$$P(F_{t+1}|D) = \sum_{j=1}^J P(F_{t+1}|M_j, D)P(M_j|D) \quad (9)$$

, which is simply a weighted average of the predictive distributions for each model using as weights the posterior model probabilities. To compute optimal allocations, we substitute in (2) the first two moments of this predictive distribution. Leamer (1978) shows that these moments can be computed using the first two moments,  $\mu_j$  and  $V_j$ , of the predictive distribution for each model  $j$ , along with the posterior model probabilities as:

$$\mu = \sum_{j=1}^J \mu_j P(M_j|D) \quad (10)$$

and

$$V = \sum_{j=1}^J V_j P(M_j|D) + \sum_{j=1}^J (\mu_j - \mu_M)(\mu_j - \mu_M)' P(M_j|D), \quad (11)$$

where  $\mu$  and  $V$  are the predictive mean and variance that account for model uncertainty. The predictive mean is simply a weighted average of the predictive means from all the models. In contrast, the predictive variance has two terms. The first term is just a weighted average of the predictive variances of the models, while the second term is the weighted covariance matrix of the predictive means across models.

## 2.2 Comparing Asset Pricing Models

To compute the posterior model probabilities, we follow by and large the procedure in Barillas and Shanken (2018). The methodology relies on the Bayesian counterpart to the traditional “alpha” test of an asset pricing model (see Gibbons, Ross and Shanken (1989)), where the factors excluded from a given model play the role of the left-hand-side assets whose alphas should be zero under the model. We start by laying out the factor model notation and the assumption about our priors. Subsequently, we describe the restricted and unrestricted components of the marginal likelihoods for a given model. As noted in the earlier work, test-asset irrelevance holds in our setting with respect to the model probabilities. We then show how to obtain model probabilities both for the case in which we consider all models

based on a set of factors as well as the case of “categorical factors”. The latter framework is what we use in our empirical application.

### 2.2.1 Factor Models, Prior Specification and Marginal Likelihoods

A factor model is a multivariate linear regression with  $N$  test-asset excess returns,  $r_t$ , and  $K$  factors for each of  $T$  periods:

$$r_t = \alpha + \beta f_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \quad (12)$$

where  $r_t$ ,  $\varepsilon_t$ , and  $\alpha$  are  $N$  vectors,  $\beta$  is  $N \times K$ , and  $f_t$  is a  $K$  vector. The residuals are assumed to be normally distributed and independent over time. The null hypothesis is  $H_o : \alpha = 0$  when the model holds. Our Bayesian test of efficiency uses standard “diffuse” priors for the betas and residual covariance parameters as in Jeffreys (1961):

$$P(\beta, \Sigma) \propto |\Sigma|^{-\frac{N+1}{2}} \quad (13)$$

This way, we allow the data to dominate beliefs about these parameters. The informative prior for  $\alpha$  is centered at zero and is multivariate normal conditional on  $\beta$  and  $\Sigma$ :

$$P(\alpha|\beta, \Sigma) = \mathcal{N}(0, k\Sigma), \quad (14)$$

where  $k > 0$  is a parameter that reflects the prior beliefs about the magnitude of potential deviations from the expected return relation. The prior specification in (14) has been used, for example, by Harvey and Zhou (1990), Pastor (2000) and Pastor and Stambaugh (2000). There is an alternative way of thinking about this prior in terms of the maximum attainable (squared) Sharpe ratio. This makes use of the following portfolio relation (see Jobson and Korkie (1982) and Gibbons, Ross and Shanken (1989)):

$$\alpha' \Sigma \alpha = \text{Sh}(f, R)^2 - \text{Sh}(f)^2. \quad (15)$$

The term  $\text{Sh}(f)^2$  is the maximum squared Sharpe ratio of the factors and  $\text{Sh}(f, R)^2$  is the maximum squared Sharpe measure based on both  $f$  and  $R$ . These Sharpe ratios and associated tangency portfolios are equal when the asset pricing model holds and  $\alpha = 0$ . Conversely, the larger the magnitude of  $\alpha$ 's allowed under the prior, the larger is the Sharpe ratio improvement when allowing for the opportunity to invest in  $R$ . This investment perspective

provides an alternative to the practice of specifying  $k$  to match a particular standard deviation of  $\alpha$ , as in Pastor (2000).

Given our assumptions, the prior expected value for the increase in the squared Sharpe ratio,  $\alpha'\Sigma^{-1}\alpha$  is  $kN$ . Thus, to parameterize the prior we think of a target value for the Sharpe ratio,  $\text{Sh}_{\max}$ , as a multiple of  $\text{Sh}(f)$ , and set  $k$  as:

$$k = \frac{\text{Sh}_{\max}^2 - \text{Sh}(f)^2}{N}. \quad (16)$$

Using the priors in (13) and (14), Barillas and Shanken (2018) compute the marginal likelihoods in closed form under both the null hypothesis  $H_0 : \alpha = 0$  and the alternative  $H_1 : \alpha \neq 0$ .

Now we show how to adapt this framework to test one factor pricing model against another. As before, there are  $K$  factors and  $N$  test assets of interest. We analyze models corresponding to all subsets of the non-market factors while requiring that the Mkt be included in the models. For a given model  $M$ ,  $f_1$  is a  $K_1$  vector that contains the non-market factors of  $M$ .  $f_2$  is a  $K_2$  vector of the excluded factors. If model  $M$  holds, it prices both the excluded factor returns  $f_2$  as well as the test-asset returns  $r$ . Thus, the alphas of  $f_2$  and  $r$  regressed on  $\{\text{Mkt}, f_1\}$  equal zero under the model. Consider the following equivalent representation of model  $M$ . Let

$$f_2 = \alpha_2 + \beta_2[\text{Mkt}, f_1] + \varepsilon_2 \quad (17)$$

and

$$r = \alpha_r + \beta_r[\text{Mkt}, f_1, f_2] + \varepsilon_r \quad (18)$$

Notice that the model in (18), which we denote as  $M_{\text{all}}$  includes all  $K$  factors.

Barillas and Shanken (2017) show that model  $M$ , which is nested in  $M_{\text{all}}$ , holds if and only if  $\alpha_2 = 0$  and  $\alpha_r = 0$ , that is, if and only if  $M$  “prices” the excluded factors and  $M_{\text{all}}$  “prices” the test assets. Intuitively, since the restriction that  $\alpha_r = 0$  is common to both models, the test asset data are irrelevant for model comparison. Formally, factoring the joint density of returns in accordance with (18), the corresponding likelihood term cancels out in calculating the probabilities in (7). Thus, the test assets will be ignored in the analysis below.

### 2.2.2 The Model Comparison Methodology

The methodology exploits the fact that each model can be viewed as a restricted version of the model that includes all of the factors under consideration. We use braces to denote models, which correspond to subsets of the given factors. Next, we describe how to compute the model marginal likelihoods.

For a given model, the joint factor return process is determined by three regressions:

$$f_{2t} = \beta_2[mkt_t \ f_{1t}] + \varepsilon_{2t}, \quad \varepsilon_{2t} \sim \mathcal{N}(0, \Sigma_2), \quad (19)$$

$$f_{1t} = \alpha + \beta mkt_t + \varepsilon_{1t}, \quad \varepsilon_{1t} \sim \mathcal{N}(0, \Sigma_1), \quad (20)$$

$$mkt_t = \beta_{mkt} \mathbf{1} + \varepsilon_{mkt,t}, \quad \varepsilon_{mkt,t} \sim \mathcal{N}(0, \Sigma_{mkt}), \quad (21)$$

Notice that (19) is a restricted regression with no constant and therefore imposes the model restriction that the excluded factor alphas are zero. The included non-market returns have unrestricted regressions. Finally, the market regression (21) assumes that the market return is  $\mathcal{N}(\mu_{mkt}, \Sigma_{mkt})$ . All disturbances are assumed to be independent. The prior for all three regressions are independent of each other and of the form in (13) and (14).

Proposition 3 in Barillas and Shanken (2018) shows that under the distributional assumptions and independent priors discussed above, the marginal likelihood for a given model is of the form

$$\text{ML} = \text{ML}(\text{Mkt})\text{ML}_U(f_1|\text{Mkt})\text{ML}_R(f_2|\text{Mkt}, f_1) \quad (22)$$

where the unrestricted and restricted ( $\alpha = 0$ ) marginal likelihoods, indicated by the subscript U or R, are obtained using the closed form solutions in Barillas and Shanken (2018). The term  $\text{ML}(\text{Mkt})$  is the same across all asset pricing models considered and therefore drops out in model comparison. This term will play a role later in the predictive distribution, however.  $\text{ML}_U(f_1|\text{Mkt})$  is computed by letting  $f_1$  play the role of  $r$  and  $\text{Mkt}$  the role of  $f$  in (12). Similarly,  $\text{ML}_R(f_2|\text{Mkt}, f_1)$  is computed by letting  $f_2$  play the role of  $r$  and  $(\text{Mkt}, f_1)$  the role of  $f$ . To specify the prior on alpha in the unrestricted regressions we use equation (16) with  $K - 1$ , which is the number of non-Mkt factors, in the denominator. Therefore  $\text{Sh}_{\max}$  is the maximum Sharpe ratio achieved with all  $K$  factors in the model and  $\text{Sh}(\text{Mkt})$  substituted for  $\text{Sh}(f)$ .

To compute posterior model probabilities, we still need to specify prior probabilities for each of the factor models. Barillas and Shanken (2018) use uniform prior model probabilities, while noting that models with stronger theoretical foundations could be given higher probabilities. In this paper, we adopt an alternative approach which, while still ad hoc, assigns more probability to the CAPM. First, we assign equal prior probability to each set of models of a given size and then distribute that probability equally across models of the same size. For example, there are four models that include Mkt, which can be formed from the FF3 factors: CAPM, FF3, and the non-nested two-factor models {Mkt HML} and {Mkt SMB}. Therefore, we assign probability of 1/3 to the CAPM and to FF3. {Mkt HML} and {Mkt SMB} both receive 1/6 prior probability, so in total the two-factor models receive 1/3 probability. Had we given all models equal probability then the CAPM would only have received 1/4 prior probability. In an application with 6 factors, as in our empirical setting, there are  $2^5 = 32$  models that contain the market. Here the CAPM receives only 1/32 under the equal prior specification, but 1/6 under the refinement. Of course, many other assignments of prior probabilities could be considered, but we can report that the results are robust across several choices that we have explored.

In our actual empirical application, some of the factors are different implementations of the same underlying concept. For example, size or value. In this setting we structure the prior so that it only assigns positive probability to models that contain at most one version of the factors in each category. Barillas and Shanken (2018) describe the details on how to implement the calculations of model probabilities under these priors. The principles and the priors for a given model are, however, the same ones already described in this section.

### 2.3 The Predictive Distribution and its first two Moments

The optimal portfolio in our framework is a function of the first two moments of the predictive distribution that takes into account model uncertainty. These moments can be computed as in equations (10) and (11) and are functions of the predictive moments of  $f$  for each asset pricing model  $j$ ,  $\mu_j^*$  and  $V_j^*$ . We compute  $\mu_j^*$  and  $V_j^*$  by imposing the structure in equations (19) to (21). Therefore, for a given model we partition its factors into excluded factors  $f_2$ , non-market included factors  $f_1$ , and the market excess return factor Mkt. Under

this partition, the predictive moments are

$$\mu^* = \begin{bmatrix} \mu_2^* \\ \mu_1^* \\ \mu_{mkt}^* \end{bmatrix} \quad V^* = \begin{bmatrix} V_{22}^* & V_{21}^* & V_{2,mkt}^* \\ V_{12}^* & V_{11}^* & V_{1,mkt}^* \\ V_{mkt,2}^* & V_{mkt,1}^* & V_{mkt,mkt}^* \end{bmatrix},$$

where for simplicity we have suppressed the  $j$  subscripts. In the appendix, expressions for all the elements of  $\mu_j^*$  and  $V_j^*$  are obtained analytically. As an intermediate step toward obtaining the predictive distribution of (out-of-sample) returns, we derive the posterior distribution for all parameters in the factor return process. The posterior mean of  $\alpha$  in equation (20), which we denote as  $\tilde{\alpha}$ , is equal to a constant times the ordinary least squares estimate  $\hat{\alpha}$ :

$$\tilde{\alpha} = c\hat{\alpha}. \tag{23}$$

The constant  $c$  is between 0 and 1 and increases with the prior variance parameter,  $k$ . Thus, the posterior for  $\alpha$  shrinks the OLS estimate toward the prior mean of 0. If  $k = 0$ ,  $c = 0$  and both the prior and the posterior are concentrated at 0. As  $k$  approaches  $\infty$ , the prior becomes uninformative with  $c$  achieving the maximum value of 1. In this case there is no shrinkage. Since the prior variance of alpha can be expressed as a multiple of the Mkt Sharpe ratio, the higher the multiple the closer is the posterior Sharpe ratio to the sample Sharpe ratio for the tangency portfolio based on all the factors in the model.

The role of shrinkage estimators has been extensively studied in portfolio analysis (e.g. Jobson, Korkie, and Ratti (1979), Jorion (1985, 1991) and Frost and Savarino (1986)). These early studies find that shrinking the means reduces the estimation error in the expected return estimate, thereby enhancing portfolio performance. In this paper and those of Pastor (2000) and Pastor and Stambaugh (2000) shrinkage is determined endogenously using asset pricing model restrictions.

Our procedure selects the portfolio of the given factors that is optimal under pricing model uncertainty and the stated simplifying assumptions. As discussed earlier, the posterior model probabilities *for the models considered* would be unchanged if we were to condition on additional test asset returns as part of the data. However, the optimal portfolio might still be affected by including test assets unless the given factors actually span the tangency portfolio for the entire market, i.e., unless the corresponding pricing model is valid for all

assets. But it is by no means clear that deviations from this ideal could be exploited to enhance out-of-sample performance. Therefore, in this paper, we focus on investment in factor returns and explore the extent to which beliefs about different factor pricing models matter.

Despite some common elements, the previous analyses in Pastor (2000) and Pastor and Stambaugh (2000) are quite different in terms of the sorts of questions addressed. They take a particular factor model as given and introduce a prior belief about the extent to which other traded factors or test assets are mispriced by that model.<sup>1</sup> Portfolio allocations and utility are then analyzed. Our analysis of model uncertainty could be extended to allow for similar beliefs about model mispricing, though questions about the sensitivity to the choice of test assets and the handling of large sets of assets would inevitably arise. However, we think that a careful investigation of factor investing is a natural starting point for research into model uncertainty and portfolio choice.

## 3 Results

### 3.1 The Factors

Our empirical application considers a total of ten candidate factors analyzed previously in Barillas and Shanken (2018), though not from a portfolio perspective. First, there are the traditional FF3 factors Mkt, HML, and SMB plus the momentum factor UMD. To these we add the investment factor CMA and the profitability factor RMW of Fama and French (2015). We also include the size (ME), investment (IA), and profitability (ROE) factors of Hou, Xue, and Zhang (2015a). Finally, we have the value factor HMLm from Asness and Frazzini (2013). The size, profitability, and investment factors differ based on the type of stock sorts used in their construction. The ROE and HMLm factors also differ from the RMW and HML profitability and value factors in that more timely earnings (most-recent quarterly vs. annual) or market value (most-recent monthly vs. annually updated value)

---

<sup>1</sup>As alluded to earlier, they partially develop a framework for incorporating model uncertainty, but do not implement this analysis.



information is employed.<sup>2</sup> The sample period for our data is January 1967 to December 2016. Some factors are available at an earlier date, but the HXZ factors start in January of 1967 due to the limited coverage of earnings announcement dates and book equity in the Compustat quarterly files.

Rather than mechanically apply our methodology with all nine of the nonmarket factors treated symmetrically, we apply the categorical framework, which recognizes that several of the factors are just different ways of measuring the same underlying construct. Specifically, we only consider models that contain at most one version of the factors in each of the following categories: size (SMB or ME), profitability (RMW or ROE), value (HML or HMLm), and investment (CMA or IA). We refer to size, profitability, value, and investment as the categorical factors. The standard factors in this application are Mkt and UMD. Since each categorical model has up to six factors and Mkt is always included, there are 32 ( $2^5$ ) possible categorical models. Given all the possible combinations of UMD and the different types of size, profitability, value, and investment factors, there is a total of 162 models under consideration.

Regarding the prior for  $k$ , our benchmark scenario assumes that  $\text{Sh}_{\max} = 1.5 \times \text{Sh}(\text{Mkt})$ , that is, the square root of the prior expected squared Sharpe ratio for the tangency portfolio based on all six factors is 50% higher than the Sharpe ratio for the market only. We think of the 1.5 six-factor choice of multiple as a prior with a risk-based tilt, assigning relatively little probability to extremely large Sharpe ratios. Below, we examine the sensitivity of posterior beliefs to this assumption, as we also explore multiples corresponding to a more behavioral perspective (more mispricing) and one with a lower value.

### 3.2 Empirical Results on Model Comparison

We start by providing an historical perspective on the evolution of posterior beliefs. To do this, we use all monthly data from January 1967 up to a given point in time and plot the corresponding posterior model probabilities in Figure 1. The starting point here is five years earlier than in Barillas and Shanken (2018) and the prior model probabilities have

---

<sup>2</sup>See the summary in Barillas and Shanken (2018) for a detailed description of the factors.

been modified, but the results are fairly similar. The top panel in Figure 1 shows posterior probabilities for the individual models. The models are ordered in the legend from the highest probability at the end of the sample to the lowest. The highest-probability model is the six-factor model {Mkt HMLm UMD ROE SMB IA}. This model has ranked first since the late 1970s. The second-best model replaces ME with SMB, the third-best uses CMA instead of IA, and the fourth one uses CMA and ME, as opposed to IA and SMB. The fifth- and seventh-ranked are five-factor models that do not have a size factor and differ only in their investment factor choice, whereas the sixth-model is a five-factor model without an investment factor.

Interestingly, the top seven models all include ROE, HMLm, and UMD and account for the lion's share of the model probability. All of these models fare better than FF5 and the four-factor model of HXZ, as do several other four-factor models, all of which contain ROE. The bottom panel of Figure 1 gives cumulative factor probabilities, that is, the sum of the posterior probabilities for models that include that factor. The probabilities are close to one for ROE, UMD, and HMLm. The probabilities for IA and SMB are around 70% and those for the other investment and size factors, CMA and ME about 25%. The profitability and value factors in FF5, HML and RMW, have probabilities close to zero.

Figure 1 about here

Now we turn to the categorical findings, in which we aggregate results over the different versions of each categorical model. The top four models in Figure 1 are all versions of the six-factor categorical model {MKT MOM VAL PROF SIZE INV} and therefore this categorical model comes in first with over 90% of the probability at the end of the sample. The second best categorical model is the five-factor model that excludes size. The third one is also a five-factor model but without an investment factor. Overall, we observe that it is essential that the more timely versions of value and profitability are employed in the models. In fact, the cumulative probabilities across all models for ROE and HMLm were almost one while those of RMW and HML were close to zero. In terms of cumulative factor probabilities for the categorical factors, profitability, value and momentum have the highest posterior probabilities, followed closely by investment and size.

Figure 2 about here

Next, we examine the sensitivity of our inferences to different assumptions about the prior for the alphas (in terms of implied Sharpe ratios). The results above were based on the prior  $Sh_{\max} = 1.5 \times Sh(\text{Mkt})$  when all six factors are included in the model. Table 1 presents the results for the individual and categorical models for prior Sharpe multiples of 1.25, 1.5, 1.75, 2, 2.5, 3 and 5. For reference, the sample Sharpe ratio for the top-ranked model is 0.5 which is 4.3 times the  $Sh(\text{Mkt})$ . Panel A presents the posterior model probabilities.

The top four models under the 1.5 baseline specification, are also the four best under the more behavioral priors that allow for increases in the Sharpe ratio of anywhere from two to five times the market ratio. The probability of the top model rises from 47.2% to 67.8% and decreases slowly for the second and third models. Only when we look below the top four models does the ranking change slightly. However, the posterior probability received by these models is very low. Overall, the models that include timely value combined with momentum and timely profitability account for over 99% of the posterior probability across all priors that we would judge to be sensible (Sharpe multiple at most 2). Turning to the categorical model probabilities in Panel B of Table 1, consistent with Panel A, we see that the six-factor model always has probability greater than 90%. The other probabilities are less than 2% except for that of the five-factor model which deletes SIZE, with probability 4.7% under the prior multiple 1.25.

Table 1 about here

While the inferences based on posterior model probabilities are fairly stable across the different priors, this is not the case with regard to the portfolio implications, as we will document in the following subsections. A contrast between conclusions derived from classical p-value analysis and Bayesian inference has previously been emphasized, for example, by Shanken (1987) with regard to tests of portfolio efficiency and by Kandel and Stambaugh (1995) in connection with market return predictability and investor utility. Here, we have another interesting contrast this one between the prior-sensitivity of Bayesian conclusions about factor pricing models depending on whether the implications for portfolio investment are considered.

### 3.3 In-sample portfolio choice results

We start by describing statistics for portfolio performance computed using the entire sample from January 1967 to December 2016. Optimal allocations are computed using equation (2) and therefore differ only in what values of  $\mu$  and  $V$  are employed. Our Bayesian investor, who confronts model uncertainty, obtains the first two predictive moments of the factor returns using equations (10) and (11). These allocations incorporate estimation risk as well as model uncertainty, imposing the restrictions of the different asset pricing models that the investor considers. We compare these allocations to that based on the sample moments, i.e., computed by plugging in the sample mean and covariance matrix of the factors into equation (2). Given the mean-variance preferences of our investor, the optimal portfolio will always be a combination of the riskless asset and the tangency portfolio based on either the predictive or the sample moments.

For our baseline results, following Pastor (2000) and Pastor and Stambaugh (2000), the value of  $\gamma$  is set so the investor allocates all wealth to Mkt when that is the only risky asset available. Using data from January 1967 to December 2016, we have  $\gamma = 2.53$ . Building on the intuition of Barillas, Hansen and Sargent (2009), we also examine optimal allocations using  $\gamma = 5$  and  $\gamma = 10$ . We employ the same informative prior for  $\alpha$  that was used for inference about the models. Our benchmark scenario assumes that  $\text{Sh}_{\max} = 1.5 \times \text{Sh}(\text{Mkt})$ . We compute results for a wide range of priors with Sharpe multiples of 1.25, 1.5, 1.75, 2, 2.5, 3, and 5. For reference, the sample Sharpe ratio for the top-ranked model is 0.5 which is 4.3 times  $\text{Sh}(\text{Mkt})$ .

Table 2 reports the Sharpe ratios along with the sample means and standard deviations of return for the sample-based optimal portfolio as well as the Bayesian model averaging procedure (BMA). Of course, the sample-based tangency portfolio always possesses the maximum in-sample Sharpe ratio for any universe of investments considered. This Sharpe ratio is 0.55 on a monthly basis for the 10-factor tangency portfolio. Our BMA procedure has Sharpe ratios close to 0.5 for each prior, albeit slightly increasing in the prior Sharpe multiple. While the Sharpe ratios for BMA are of similar magnitude to those of the sample-based tangency portfolio, there are huge differences in the means and standard deviations of the optimal

portfolio returns for all levels of relative risk aversion considered. These portfolios are the optimal allocations allowing for investment in the riskless asset as well as the factor returns. For instance, the sample-based optimal portfolio has a mean annual return of 144.3% with a standard deviation of 75.7% for an investor with relative risk aversion 2.53. In contrast, the portfolio of the Bayesian investor who believes a priori that the non-market factors can increase the Sharpe ratio of the market by 50% has a mean of 69.1% and standard deviation of 40.7%.<sup>3</sup>

The BMA portfolios have mean returns and standard deviations that are increasing in the prior multiple. However, these statistics are substantially more moderate than those of the sample-based optimal portfolios for the more reasonable priors. This is due primarily to the greater shrinkage of each models factor alphas (in regressions on the market) toward zero for the lower prior Sharpe multiples. Increasing risk aversion makes the characteristics of all portfolios more moderate as well. To further understand these properties, we turn next to examining the individual asset allocations.

Table 2 about here

Table 3 reports the optimal allocations per \$1 of investment in the riskfree asset and the risky assets shown in the first column. The next column shows the weights based on sample moments of returns, followed by the weights derived from the Bayesian investors predictive moments for each prior multiple. We start by describing the positions in cash (Rf). With risk aversion of  $\gamma = 2.53$  the sample tangency portfolio is highly levered borrowing 1.47 dollars per unit of wealth invested. The Bayesian investor also uses leverage at this level of risk aversion, but to a smaller extent, 52 cents per unit of wealth. The amount borrowed decreases with the prior Sharpe multiple, however, and approaches that based on the sample moments for the most extreme prior. Naturally, these levels of leverage contribute to the extreme characteristics of the tangency portfolios observed in Table 2. Next, we examine the differences in the risky asset allocations for the different investors.

Table 3 reports the optimal allocations to the riskfree and the risky assets shown in the

---

<sup>3</sup>For reference, over the same time period the market excess return has a 6.25% mean and 15.6% standard deviation.

first column per \$1 of investment. The next column shows the optimal weights based on sample moments of returns, followed by the weights derived from the Bayesian investors predictive moments for each prior multiple. We start by describing the positions in cash (Rf). The allocation shown for the riskless asset is  $w_{rf}$ , as defined in Section 2.1. This is the net investment taking into account the short side of the excess return,  $Mkt = R_{mkt} - R_f$ . Thus, it equals one minus the allocation to Mkt. Accordingly, the market allocation is labelled RM, rather than Mkt.<sup>4</sup> With risk aversion of  $\gamma = 2.53$ , the sample-based optimal portfolio is highly levered, borrowing 1.47 dollars per unit of wealth invested. The Bayesian investor also uses leverage at this level of risk aversion, but to a lesser extent, 52 cents per unit of wealth. The amount borrowed increases with the prior Sharpe multiple, however, and approaches that based on the sample moments. Naturally, these levels of leverage contribute to the extreme characteristics of the tangency portfolios observed in Table 2.

Next, we examine the differences in the risky asset allocations for the different investors. The portfolio positions of the Bayesian investor reflect the posterior probabilities for the asset pricing models. For instance, in the extreme event that there were a dominant model that received all of the posterior probability, the optimal allocation of the BMA procedure would only invest in the factors of that model. While we consider models with at most 6 factors, our BMA portfolio invests in all 10 factors since the posterior model probabilities are spread over a variety of models. However, some factors receive little weight since the models that include them have low probability. Thus, the weights on the factors are related to the cumulative factor probabilities shown in the second panel of Figure 1.

For instance, the BMA investor allocates almost no wealth to the HML value and the RMW profitability factors. But, the sample-based investor takes extreme positions in the two highly correlated value factors, HML and HMLm. To exploit the differences between them, this investor plays each factor against the other by going short HML and long HMLm with weights of roughly -700% and 1,200%. Similar observations apply to the profitability factors RMW and ROE and the size factors ME and SMB, although the portfolio allocations are not as extreme. In contrast, the BMA portfolio employs essentially long-only positions in

---

<sup>4</sup>The optimal portfolio return would then be obtained as the sum of the allocations times the returns for the investments shown in the first column.

the factors and has much more moderate portfolio weights overall. Yet the BMA allocation achieves in-sample Sharpe ratios that are almost as high.

Table 3 about here

### 3.4 Real-time portfolio choice results

The in-sample Sharpe ratios are very large, four to five times that of the market factor for the BMA and sample-based investors, respectively. This begs the question of whether an investor could have achieved such levels in real-time. The following out-of-sample exercise is conducted by first allowing the investor to observe 20 years of data from January 1967 to December 1986 in forming predictive beliefs and then to recursively add observations for the next thirty years to the end of the sample, which is December 2016. Thus, the first asset allocation is computed using data from months 1 to  $Q = 240$ , and is then rebalanced using data from 1 to  $Q+1$  and so on. The last portfolio position is computed with data from 1 to  $T-1$ , where  $T$  is 600.

Table 4 summarizes the real-time performance for both the sample-based optimal portfolio and the BMA procedure under different prior beliefs about the degree of model mispricing of the excluded factors. Panel A presents the real-time Sharpe ratios together with the mean and standard deviation for the out-of-sample optimal portfolio returns. As expected, the out-of-sample Sharpe ratios are smaller than their in-sample counterparts. In most cases, the Bayesian investor who uses posterior beliefs about all of the models in computing optimal allocations still lags (in terms of Sharpe ratios) the naive investor who simply employs the sample mean and covariance of returns. However, as was the case with the in-sample results, the out-of-sample portfolio returns of the sample-based investor have substantially larger return variance. With the baseline risk aversion of 2.53 the standard deviation of the portfolio returns is 120.4% per year for this investor and less than 50% for the Bayesian investor with a reasonable prior multiple of 1.5. Increasing risk aversion to a value of 5 results in “acceptable” portfolio variances for the Bayesian investor at the lower prior multiples. With  $\gamma = 10$ , the Bayesian investor risk levels are reasonable even for the extreme prior multiple of 5.

Next, we proceed to compute the certainty equivalent returns for the optimal portfolios. This is a commonly used metric for assessing the economic performance of a portfolio selection rule. In the mean-variance framework with utility given by (1), this riskless (zero-variance) return is

$$CER = \bar{r}_p - \frac{1}{2}\gamma\sigma_p^2 \quad (24)$$

where  $\bar{r}_p$  and  $\sigma_p^2$  are the mean and variance of the out-of-sample returns.<sup>5</sup> Interestingly, the certainty equivalent results paint a very different picture than do the Sharpe ratios. We report annualized values of the CER. While the sample-based optimal portfolio had Sharpe ratios comparable to those of the model-averaged portfolio, its out-of-sample certainty equivalent return is -33.8% with  $\gamma = 2.53$ . Despite the high real-time average returns, the CER is large and negative for the sample optimal portfolio due to its extreme variance. This is a consequence of the substantial leverage employed, as indicated by the large average negative position of -164% in the riskfree return shown in Panel A.

Of course, had this out-of-sample risk been fully anticipated in the predictive distribution, the investor could have reduced leverage and lowered risk. Lower leverage is indeed what we see for Bayesian model averaging, especially with the more conservative priors. The resulting certainty-equivalents are large and positive with these priors. For example, with  $\gamma = 2.53$  or  $\gamma = 5$ , the CER exceeds 10% per year for prior Sharpe multiples of 2 or less. Only for the most extreme prior that allows the non-market factors to increase the Sharpe ratio to 5 times the market ratio does the CER turn negative. So, from an economic point of view the Bayesian investor who takes into account plausible beliefs about asset pricing models substantially outperforms the naive investor who simply uses the sample means and variances-covariances to form optimal portfolios. For comparison, the CER for the optimal combination of the market and the riskless asset is 4.44%

Table 4 about here

The real-time success of the BMA procedure is partly attributable to the fact that it shrinks a models sample alphas towards zero for the non-market included factors, leading

---

<sup>5</sup>Out-of-sample CERs have often been used in the literature and implicitly treat the out-of-sample returns as if they were a population from which returns are to be drawn at a given point in time.



to the more moderate optimal allocations. Table 3 shows the optimal allocations after observing all of the data. To give more color to our real-time results, Figure 3 plots the recursive optimal allocations for an investor with relative risk aversion of  $\gamma = 2.53$ . These are the allocations used to compute the real time results. The top panel shows the sample-based optimal portfolio weights and the bottom panel the weights of the Bayesian investor with a prior view that the non-market factors can increase the market Sharpe ratio by 50%. To facilitate comparison, the scale of both panels is the same. Not only are the sample-based allocations more extreme than those of the Bayesian investor, but they are also much more volatile. In addition, the Bayesian allocations are by and large long the factors, whereas the sample-based allocations take opposite positions in similar factors, e.g. HMLm’s position is positive whereas HML’s is negative, both with very large magnitudes. From a practical perspective, the portfolio turnover is much greater for the sample-based portfolio, which is likely to translate into higher trading costs that would make its real-time performance even worse.

Figure 3 about here

## 4 Experiment

A concern in the empirical asset pricing literature is that some of the factors have been discovered as the product of data mining. Another common dilemma from an investment perspective is that a factors discovery and the associated published research can lead to the factors demise: as more investors start employing strategies to exploit the factor, its alpha may be substantially reduced, e.g., Pontiff and Mclean (2015). While a full statistical treatment of these issues is beyond the scope of our paper, this section highlights some advantages of the Bayesian model averaging procedure in such situations. We explore these issues through an “experiment in which the original data is modified as follows. After the initial period for the recursive exercise of twenty years, some of the factor out-of-sample returns are generated in such a way that the alphas in the evaluation period are zero. This is meant to create an environment in which these factors are redundant going forward, even if they were not in the past.

There are other reasons why a factor might be redundant. An argument for the economic importance of a factor may be reasonable, yet still be wrong as a description of reality. Or the economic effect may already be captured by other factors. For example, Fama and French (2015) find that HML is redundant once they account for the four other factors in their model. In general, since a redundant factor has zero alphas on the other factors in a model, adding it to an investors portfolio will not increase the squared Sharpe ratio of the portfolio. Thus, from an investment perspective, it might appear that the choice of whether to include a redundant factor in the potential set of investment factors is of no consequence. However, any wealth that an investor directs toward that factor will result in a deviation from the optimal allocation and will likely reduce out-of-sample performance.

Empirically, we use as our benchmark model {Mkt SMB HML UMD}. Then we estimate the alphas of CMA, RMW, ME, HMLm, IA and ROE on this benchmark model and subtract the alpha estimates from the six excluded factors, but only for the observations after the initial period of 20 years. Therefore, the data is unchanged for the first twenty years. As a result, the investor initially believes that all of the factors are essential and then gradually learns in real-time that there is no longer alpha for some of the factors relative to the benchmark.

Table 5 presents the in-sample portfolio performance statistics from this experiment. Since six of the factors now have zero alphas on the benchmark for the majority of the sample, naturally the Sharpe ratio, 0.35, of the sample-based tangency portfolio is now lower than under the original data (0.55). Just as before, the BMA Sharpe ratios are increasing slightly in the prior Sharpe multiple and are relatively close to that for the sample tangency ratio. And again, the mean and standard deviation of the sample-based optimal portfolio are extreme at the lower levels of risk aversion, whereas those of the Bayesian investor are much more moderate, especially for the more reasonable priors.

Table 5 about here
--------------------

The optimal allocations at the end of the sample are reported in Table 6. At this point, the investor has observed 30 years of data in which CMA, RMW, ME, HMLm and IA have zero alphas. The alpha estimates over the sample are non-zero because these six

factors do have nonzero alphas on the benchmark over the first twenty years. However, the leverage of the sample-based optimal portfolio and the magnitude of its factor weights are considerably larger than those of the Bayesian investors. Since it is suboptimal to invest in zero-alpha assets (the true alphas are zero going out-of-sample), this helps explain the significant underperformance of the sample-based optimal portfolio, which we focus on next.

Table 6 about here

Table 7 reports the out-of-sample performance statistics. As earlier, investment begins after the twenty-year initial estimation period. The real-time Sharpe ratio of the sample-based optimal portfolio is now only 0.056. In contrast, the out-of-sample Sharpe ratios for the BMA procedure are significantly higher, especially for the more conservative priors. In the actual data, there was not much difference between the sample-based and Bayesian Sharpe ratios. We expect this to be the case when none of factors (other than categorical versions) is redundant, as appears to be the case in the actual data. On the other hand, when some of the factors are truly redundant, the sample-based portfolio does substantially worse in terms of out-of-sample Sharpe ratios. In fact, the large portfolio positions induce volatility into the optimal portfolio returns (85.3% standard deviation when  $\gamma = 2.53$ ) which is not compensated with commensurately high returns (only 16.5%).

Panel B of Table 7 reports the certainty equivalent returns. As before, the CERs are highly negative for the sample-based portfolio. For the Bayesian investor, they are positive with conservative priors. Thus, the skepticism inherent in the more conservative priors substantially enhances performance when, for any of the reasons alluded to above, some of the factors are, in fact, redundant. Of course, with complete skepticism, a prior multiple of one, the Bayesian investor will simply hold a classic position in the market portfolio. Our intent here is not to advocate a particular prior but, rather, to make the point that an investors prior beliefs about factor pricing models, as captured in Bayesian model averaging, can have a considerable impact on the ultimate performance of her portfolio.

Table 7 about here

## 5 Conclusion

The recent proliferation of anomalies places investors in a quandary. While the purported alphas of such anomalies are a priori attractive from an investment perspective, there are reasons why those alphas might not ultimately deliver benefits to investors. Anomalies might already be subsumed by existing factors, could be the product of data mining, or have already been exploited by financial market participants.

In this paper we have proposed a Bayesian asset allocation procedure that allows an investor to simultaneously consider all asset pricing models that can be formed with a given set of factors. The procedure incorporates both parameter uncertainty and model uncertainty. Our prior on the maximum squared Sharpe ratio achievable by adding factors to the market portfolio implies a degree of shrinkage that turns out to be beneficial for portfolio formation in real-time.

An experiment meant to recreate situations in which some of the factors are the product of data mining or in which a factor has already been exploited demonstrates how our procedure is advantageous in such situations, especially in an out-of-sample context.

We have allocated equal prior probability to models with the same number of factors and equal probability across model size. Other ways of allocating prior probability, perhaps by giving more prior weight to models with more economically meaningful factors could also be entertained, but we leave such extensions for future research.

# Appendix

## A1 Framework

We form models based on a collection of factors  $f$ . The market excess return,  $Mkt$ , is in all models. Then for a given model,  $f_1$  is a  $K_1$  vector that contains the non-market factors of the model.  $f_2$  is a  $K_2$  vector with the excluded factors. Therefore we consider in total  $K$  factors where  $K = K_1 + K_2 + 1$

For a given model the factor returns are governed by the following regressions:

$$f_{2t} = \beta_2[mkt_t \ f_{1t}] + \varepsilon_{2t}, \quad \varepsilon_{2t} \sim \mathcal{N}(0, \Sigma_2), \quad (\text{A1})$$

$$f_{1t} = \alpha + \beta mkt_t + \varepsilon_{1t}, \quad \varepsilon_{1t} \sim \mathcal{N}(0, \Sigma_1), \quad (\text{A2})$$

$$mkt_t = \beta_{mkt} \mathbf{1} + \varepsilon_{mkt,t}, \quad \varepsilon_{mkt,t} \sim \mathcal{N}(0, \Sigma_{mkt}), \quad (\text{A3})$$

Notice that (A1) is a restricted regression with no constant and therefore assumes that the asset pricing model holds and therefore produces zero alphas on the excluded factor returns. The non-market returns follow unrestricted regressions. Finally, the market regression (A3) assumes that the market return is  $\mathcal{N}(\mu_{mkt}, \Sigma_{mkt})$ . All disturbances are assumed to be independent.

For the derivations it is useful to write equations (A1)–(A3) in multivariate form as:

$$F_2 = X_1 B_2 + E_2, \quad (\text{A4})$$

where  $F_2$  is  $T \times K_2$ .  $X_1 = [MKT \ F_1]$  where  $MKT$  is a  $T$  vector and  $F_1$  is  $T \times K_1$ .  $B_2$  is therefore  $(1 + K_1) \times K_2$  and  $E_2$  is  $T \times K_2$  and matrix variate normal.

$$F_1 = X B_1 + E_1, \quad (\text{A5})$$

$F_1$  is  $T \times K_1$ .  $X = [\mathbf{1}_T \ MKT]$ ,  $F_1$  is  $T \times K_1$ .  $B_1$  is therefore  $2 \times K_1$  and  $E_1$  is  $T \times K_1$  and matrix-variate normal. Furthermore,  $B_1 = \begin{bmatrix} \alpha' \\ \beta' \end{bmatrix}$

$$MKT = \mathbf{1}_T \mu_{mkt} + E_{mkt}, \quad (\text{A6})$$

## A2 Posterior Distributions

Let  $\theta$  contain all of the parameters in equations (A1)-(A3). The first step towards deriving the predictive distribution is to get the posterior distribution of  $\theta$  which is proportional to the likelihood times the prior

$$p(\theta|F) \propto p(F|\theta)p(\theta)$$

Our assumptions about the prior distributions described below together with the fact that the disturbances in equations (A1)-(A3) are independent allow us to compute the posterior distribution of the parameters for each of the equations independently.

Except for  $\alpha$  in equation (A2) we use uninformative priors for all of the parameters. In particular for equation (A1) the prior for  $B_2$  and  $\Sigma_2$  are as in Jeffreys:

$$P(B_2, \Sigma_2) \propto |\Sigma_2|^{-\frac{K_2+1}{2}}.$$

The likelihood of  $F_2$  is

$$P(F_2|X_1, B_2, \Sigma_2) \propto |\Sigma_2|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\text{tr}[(F_2 - X_1B_2)'(F_2 - X_1B_2)\Sigma_2^{-1}]\right).$$

Note that

$$(F_2 - X_1B_2)'(F_2 - X_1B_2) = S_2 + (B_2 - \widehat{B}_2)'X_1'X_1(B_2 - \widehat{B}_2),$$

where  $S_2$  is the sum of squared residuals (SSR) from ordinary least squares (OLS) so that

$$P(F_2|X_1, B_2, \Sigma_2) \propto |\Sigma_2|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\text{tr}[S_2\Sigma_2^{-1}]\right) \exp\left(-\frac{1}{2}\text{tr}[(B_2 - \widehat{B}_2)'X_1'X_1(B_2 - \widehat{B}_2)\Sigma_2^{-1}]\right).$$

Using the fact that  $\text{tr}(A'BCD') = \text{vec}(A)'(D \otimes B)\text{vec}(C)$ , we can write

$$\text{tr}[(B_2 - \widehat{B}_2)'X_1'X_1(B_2 - \widehat{B}_2)\Sigma_2^{-1}] = (b_2 - \widehat{b}_2)'(\Sigma_2^{-1} \otimes (X_1'X_1))(b_2 - \widehat{b}_2),$$

where  $b_2 = \text{vec}(B_2)$  and  $\widehat{b}_2 = \text{vec}(\widehat{B}_2)$ . Furthermore,  $[\Sigma_2^{-1} \otimes (X_1'X_1)]^{-1} = \Sigma_2 \otimes (X_1'X_1)^{-1}$  and therefore the likelihood can be written as

$$P(F_2|X_1, B_2, \Sigma_2) \propto |\Sigma_2|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\text{tr}[S_2\Sigma_2^{-1}]\right) \exp\left(-\frac{1}{2}\text{tr}[(b_2 - \widehat{b}_2)'(\Sigma_2 \otimes (X_1'X_1)^{-1})(b_2 - \widehat{b}_2)]\right).$$

Next, multiplying the likelihood by the prior we obtain the posterior distribution which we expressed as  $P(B_2, \Sigma|F) = P(B_2|\Sigma_2, F)P(\Sigma_2|F)$ . The posterior follows a Normal-Inverse

Wishart distribution:<sup>1</sup>

$$P(B_2, \Sigma_2 | F) \propto |\Sigma_2|^{-\frac{K_1+1}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[(b_2 - \hat{b}_2)'(\Sigma_2 \otimes (X_1'X_1)^{-1})^{-1}(b_2 - \hat{b}_2)\right]\right) \times \\ |\Sigma_2|^{-\frac{T+K_2-K_1}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[\Sigma_2^{-1}S_2\right]\right)$$

Therefore,

$$P(b_2 | \Sigma_2, F) \propto \mathcal{N}(\hat{b}_2, \Sigma_2 \otimes (X_1'X_1)^{-1})$$

$$P(\Sigma_2 | F) \propto IW(S_2, T - K_1 - 1)$$

Denoting the posterior means with a  $\sim$ :

$$\tilde{b}_2 = E(b_2 | F) = \hat{b}_2 \tag{A7}$$

$$\tilde{\Sigma}_2 = E(\Sigma_2 | F) = \frac{S_2}{T - K_1 - K_2 - 2} \tag{A8}$$

and the posterior variance of  $b_2$  is

$$\text{var}(b_2 | F) = \tilde{\Sigma}_2 \otimes (X_1'X_1)^{-1}. \tag{A9}$$

Equation (A3) is just a special case of the derivations for (A1) in which  $B_2 = \mu_{mkt}$ ,  $X_1 = 1_T$ ,  $K_1 = 0$  and  $K_2 = 1$ . Therefore it follows that:

$$\tilde{\mu}_{mkt} = E(\mu_{mkt} | F) = \hat{\mu}_{mkt} \tag{A10}$$

$$\tilde{V}_{mkt,mkt} = \frac{T}{T-3} \hat{V}_{mkt,mkt} \tag{A11}$$

and the posterior variance of  $\mu_{mkt}$  is

$$\text{var}(\mu_{mkt} | F) = \frac{1}{T-3} \hat{V}_{mkt,mkt} \tag{A12}$$

Next we turn our attention to deriving the posterior distributions of the parameters in equation (A2). The likelihood of  $F_1$  is given by:

$$P(F_1 | X, \alpha, \beta, \Sigma_1) \propto |\Sigma_1|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[(F_1 - XB_1)'(F_1 - XB_1)\Sigma_1^{-1}\right]\right).$$

---

<sup>1</sup>An  $m \times m$  matrix  $G \sim IW(H, \nu)$  if  $P(G|H, \nu) \propto |G|^{-\frac{(\nu+m+1)}{2}} \exp(-\frac{1}{2}\text{tr}(G^{-1}H))$ . Furthermore  $E[G] = \frac{H}{\nu-m-1}$ .

In terms of priors, we continue to assume an uninformative prior for  $\beta$  and  $\Sigma_1$  as in Jeffreys:

$$P(\beta, \Sigma_1) \propto |\Sigma_1|^{-\frac{K_1+1}{2}}.$$

The prior for  $\alpha$  is informative and reflects the belief on the investor about market efficiency:

$$P(\alpha|\Sigma_1) \sim N(0, k\Sigma_1) \propto |\Sigma_1|^{-\frac{1}{2}} \exp\left(-\frac{1}{2k}\alpha'\Sigma_1^{-1}\alpha\right)$$

Multiplying the likelihood by the prior we obtain the posterior density

$$P(B_1, \Sigma_1|F) \propto |\Sigma_1|^{-\frac{T+K_1+2}{2}} \exp\left(-\frac{1}{2k}\alpha'\Sigma_1^{-1}\alpha\right) \exp\left(-\frac{1}{2}\text{tr}[(F_1 - XB_1)'(F_1 - XB_1)\Sigma_1^{-1}]\right)$$

Let  $H = \begin{bmatrix} 1/k & 0 \\ 0 & 0 \end{bmatrix}$  then

$$\frac{1}{k}\alpha'\Sigma_1^{-1}\alpha = b_1'(\Sigma_1^{-1} \otimes H)b_1$$

where  $b_1 = \text{vec}(B_1)$  and as above

$$(F_1 - XB_1)'(F_2 - XB_1) = S_1 + (B_1 - \widehat{B}_1)'X'X(B_1 - \widehat{B}_1),$$

where  $S_1$  is the sum of squared residuals (SSR) from ordinary least squares (OLS). The posterior can then be written as

$$P(B_1, \Sigma_1|F) \propto |\Sigma_1|^{-\frac{T+K_1+2}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[b_1'(\Sigma_1^{-1} \otimes H)b_1 + (b_1 - \widehat{b}_1)'(\Sigma_1 \otimes (X'X)^{-1})^{-1}(b_1 - \widehat{b}_1)\right]\right) \times \exp\left(-\frac{1}{2}\text{tr}[S_1\Sigma_1^{-1}]\right)$$

Let  $G = H + X'X$  and  $D = X'X - X'XG^{-1}X'X$ . Completing the square for  $b_1$  the posterior can be written as:<sup>2</sup>

$$P(B_1, \Sigma_1|F) \propto |\Sigma_1|^{-\frac{2}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[(b_1 - \widetilde{b}_1)'(\Sigma_1 \otimes G^{-1})^{-1}(b_1 - \widetilde{b}_1)\right]\right) \times |\Sigma_1|^{-\frac{T+K_1}{2}} \exp\left(-\frac{1}{2}\text{tr}\left[(S_1 + \widehat{B}_1'D\widehat{B}_1)\Sigma_1^{-1}\right]\right)$$

where  $\widetilde{b}_1 = (I_{K_1} \otimes G^{-1}(X'X))\widehat{b}_1$ . It follows then that:

$$P(b_1|\Sigma_1, F) \propto \mathcal{N}(\widetilde{b}_1, \Sigma_1 \otimes G^{-1})$$

---

<sup>2</sup>To complete the square use the fact that:  $x'Ax + x'b + c = (x-h)'A(x-h) + k$  where  $h = -\frac{1}{2}A^{-1}b$  and  $k = c - \frac{1}{4}b'A^{-1}b$ .



$$P(\Sigma_1|F) \propto IW(S_1 + \widehat{B}'_1 D \widehat{B}_1, T - 1)$$

and

$$\widetilde{b}_1 = E(b_1|F) = (I_{K_1} \otimes G^{-1}(X'X)) \widehat{b}_1 \quad (\text{A13})$$

$$\widetilde{\Sigma}_1 = E(\Sigma_1|F) = \frac{S_1 + \widehat{B}'_1 D \widehat{B}_1}{T - K_1 - 2} \quad (\text{A14})$$

and the posterior variance of  $b_1$  is:

$$\text{var}(b_1|F) = \widetilde{\Sigma}_1 \otimes G^{-1}. \quad (\text{A15})$$

### A3 The Predictive Distribution

The first two moments of the predictive distribution of  $f$ ,  $\mu^*$  and  $V^*$  are needed to compute the optimal portfolio in our framework.

$$\mu^* = \begin{bmatrix} \mu_2^* \\ \mu_1^* \\ \mu_{mkt}^* \end{bmatrix}, \quad V^* = \begin{bmatrix} V_{22}^* & V_{21}^* & V_{2,mkt}^* \\ V_{12}^* & V_{11}^* & V_{1,mkt}^* \\ V_{mkt,2}^* & V_{mkt,1}^* & V_{mkt,mkt}^* \end{bmatrix}.$$

### A4 Mean of the Predictive Distribution

The predictive means for the factors  $\mu^*$  can be simply computed as the posterior mean since:

$$\mu^* = E(F_{T+1}|F) = E(E(F_{T+1}|\theta, F)|F) = E(\mu|F) = \widetilde{\mu}. \quad (\text{A16})$$

We then find that

$$\mu_{mkt}^* = \widetilde{\mu}_{mkt} = \widehat{\mu}_{mkt} \quad (\text{A17})$$

$$\mu_1^* = \widetilde{\mu}_1 = E(X B_1|F) = [1 \quad \widehat{\mu}_{mkt}] \widetilde{B}_1 \quad (\text{A18})$$

$$\mu_2^* = \widetilde{\mu}_2 = E(X_1 B_2|F) = [\widehat{\mu}_{mkt} \quad \widetilde{\mu}_1] \widetilde{B}_2 \quad (\text{A19})$$

### A5 Second Moment of the Predictive Distribution

To compute the different elements of  $V^*$  use the law of total variance which implies that

$$V^* = \text{var}(F_{T+1}|F) = E(V|F) + \text{var}(\mu|F) = \widetilde{V} + \text{var}(\mu|F) \quad (\text{A20})$$

### A5.1 $V_{mkt,mkt}^*$

Applying (A20) to  $mkt$  it follows that

$$V_{mkt,mkt}^* = \tilde{V}_{mkt,mkt} + \text{var}(\mu_{mkt}|F) = \frac{T+1}{T-3} \widehat{V}_{mkt,mkt} \quad (\text{A21})$$

### A5.2 $V_{1,1}^*$

We can represent  $V_{1,1}^*$  in terms of its  $(i, j)$  element. Let  $y_i$  be an element of  $F_1$  then from (A2)

$$\begin{aligned} y_{i,T+1} &= \alpha_i + \beta_i mkt_{T+1} + \varepsilon_{1,i,T+1} \\ y_{i,T+1} &= [1 \quad mkt_{T+1}] b_{1,i} + \varepsilon_{1,i,T+1} \end{aligned}$$

since  $b_{1,i} = (\alpha_i \quad \beta_i)'$ . The predictive covariance between  $y_{i,T+1}$  and  $y_{j,T+1}$ , the  $(i, j)$  element of  $V_{1,1}^*$ , can be obtained using the decomposition

$$\begin{aligned} \text{cov}(y_{i,T+1}, y_{j,T+1}|F) &= \text{cov}[\text{E}(y_{i,T+1}|b_1, F), \text{E}(y_{j,T+1}|b_1, F)|F] \\ &\quad + \text{E}(\text{cov}(y_{i,T+1}, y_{j,T+1})|b_1, F)|F) \end{aligned} \quad (\text{A22})$$

To compute the first term note that  $\text{E}(y_{i,T+1}|b_1, F) = [1 \quad \tilde{\mu}_{mkt}] b_{1,i}$  so

$$\text{cov}[\text{E}(y_{i,T+1}|b_1, F), \text{E}(y_{j,T+1}|b_1, F)|F] = [1 \quad \tilde{\mu}_{mkt}]' \text{cov}(b'_{1,j}, b_{1,i}|F) [1 \quad \tilde{\mu}_{mkt}].$$

For the second term

$$\text{E}[\text{cov}(y_{i,T+1}, y_{j,T+1}|b_1, F)|F] = \tilde{\beta}_i \tilde{\beta}_j V_{mkt,mkt}^* + \text{cov}(\beta_j, \beta_i|F) V_{mkt,mkt}^* + \tilde{\sigma}_{1,ij}$$

Therefore

$$\begin{aligned} \text{cov}(y_{i,T+1}, y_{j,T+1}|F) &= [1 \quad \tilde{\mu}_{mkt}]' \text{cov}(b'_{1,j}, b_{1,i}|F) [1 \quad \tilde{\mu}_{mkt}] + \tilde{\beta}_i \tilde{\beta}_j V_{mkt,mkt}^* \\ &\quad + \text{cov}(\beta_j, \beta_i|F) V_{mkt,mkt}^* + \tilde{\sigma}_{1,ij} \end{aligned} \quad (\text{A23})$$

### A5.3 $V_{1,mkt}^*$

We can compute  $V_{1,mkt}^*$  using the decomposition

$$\begin{aligned}
\text{cov}(F_{1,T+1}, mkt_{T+1}|F) &= \text{cov}([1 \ \mu_{mkt}]B_1, \mu_{mkt}|F) + \text{E}(\beta V_{mkt,mkt}|F) \\
&= \tilde{\beta} \text{var}(\mu_{mkt}|F) + \tilde{\beta} \tilde{V}_{mkt,mkt} \\
&= \tilde{\beta} \frac{\tilde{T} + 1}{\tilde{T} - 3} \hat{V}_{mkt,mkt} \\
&= \tilde{\beta} V_{mkt,mkt}^*
\end{aligned} \tag{A24}$$

### A5.4 $V_{2,2}^*$

We use similar arguments as in the derivations for  $V_{1,1}^*$ . Let  $y_i$  be an element of  $F_2$  then the predictive covariance between  $y_{i,T+1}$  and  $y_{j,T+1}$ , the  $(i, j)$  element of  $V_{2,2}^*$ , can be obtained using the decomposition

$$\begin{aligned}
\text{cov}(y_{i,T+1}, y_{j,T+1}|F) &= \text{cov}[\text{E}(y_{i,T+1}|b_2, b_1, F), \text{E}(y_{j,T+1}|b_2, b_1, F)|F] \\
&\quad + \text{E}(\text{cov}(y_{i,T+1}, y_{j,T+1})|b_2, b_1, F)|F)
\end{aligned} \tag{A25}$$

To compute the first term of (A25) note that  $\text{E}(y_{i,T+1}|b_2, b_1, F) = [\tilde{\mu}_{mkt} \ [1 \ \tilde{\mu}_{mkt}]B_1]b_{2,i} = [\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}$ . Then

$$\begin{aligned}
&\text{cov}[\text{E}(y_{i,T+1}|b_2, b_1, F), \text{E}(y_{j,T+1}|b_2, b_1, F)|F] = \\
&\text{cov}[\text{E}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}|b_1, F), \text{E}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,j}|b_1, F)|F] \\
&\quad + \text{E}(\text{cov}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}, [\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,j}|b_1, F)|F)
\end{aligned}$$

Now the first term of the expression above is

$$\begin{aligned}
&\text{cov}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]\tilde{b}_{2,i}, [\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]\tilde{b}_{2,j}|F) \\
&\quad \text{cov}([\alpha + \tilde{\mu}_{mkt}\beta]\tilde{b}_{2,1,i}, [\alpha + \tilde{\mu}_{mkt}\beta]\tilde{b}_{2,1,j}|F) \\
&= \tilde{b}'_{2,1,i} [\text{cov}(\alpha, \alpha|F) + 2\tilde{\mu}_{mkt}\text{cov}(\alpha, \beta|F) + \tilde{\mu}_{mkt}\text{cov}(\beta, \beta|F)] \tilde{b}_{2,1,j}
\end{aligned} \tag{A26}$$

and the second term

$$\begin{aligned}
& \mathbb{E}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]' \text{cov}(b_{2,i}, b_{2,j}|F) [\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta] b_{2,j} | b_1, F) | F) \\
&= [\tilde{\mu}_{mkt} \ \tilde{\mu}_1]' \text{cov}(b_{2,i}, b_{2,j}|F) [\tilde{\mu}_{mkt} \ \tilde{\mu}_1] + \text{tr}[\text{cov}(b_{2,1,i}, b_{2,1,j}|F) \text{var}(\alpha + \tilde{\mu}_{mkt}\beta|F)] \\
& \quad = [\tilde{\mu}_{mkt} \ \tilde{\mu}_1]' \text{cov}(b_{2,i}, b_{2,j}|F) [\tilde{\mu}_{mkt} \ \tilde{\mu}_1] \\
& \quad + \text{tr}[\text{cov}(b_{2,1,i}, b_{2,1,j}|F) ([\text{cov}(\alpha, \alpha|F) + 2\tilde{\mu}_{mkt}\text{cov}(\alpha, \beta|F) + \tilde{\mu}_{mkt}\text{cov}(\beta, \beta|F)])]
\end{aligned}$$

So the first term of (A25) is

$$\begin{aligned}
& \tilde{b}'_{2,1,i} [\text{cov}(\alpha, \alpha|F) + 2\tilde{\mu}_{mkt}\text{cov}(\alpha, \beta|F) + \tilde{\mu}_{mkt}\text{cov}(\beta, \beta|F)] \tilde{b}_{2,1,j} + [\tilde{\mu}_{mkt} \ \tilde{\mu}_1]' \text{cov}(b_{2,i}, b_{2,j}|F) [\tilde{\mu}_{mkt} \ \tilde{\mu}_1] \\
& \quad + \text{tr}[\text{cov}(b_{2,1,i}, b_{2,1,j}|F) ([\text{cov}(\alpha, \alpha|F) + 2\tilde{\mu}_{mkt}\text{cov}(\alpha, \beta|F) + \tilde{\mu}_{mkt}\text{cov}(\beta, \beta|F)])]
\end{aligned}$$

For the second term of (A25)

$$\mathbb{E}[\text{cov}(y_{i,T+1}, y_{j,T+1}) | b_2, b_1, F) | F] = \tilde{b}'_{2,i} V_{1mkt,1mkt}^* \tilde{b}_{2,j} + \text{tr}[V_{1mkt,1mkt}^* \text{cov}(b_{2,i}, b_{2,j}|F)] + \tilde{\sigma}_{2,ij}$$

where

$$V_{1mkt,1mkt}^* = \begin{bmatrix} V_{11}^* & V_{1,mkt}^* \\ V_{mkt,1}^* & V_{mkt,mkt}^* \end{bmatrix}.$$

Since  $\text{var}(\alpha + \tilde{\mu}_{mkt}\beta|F) = \text{cov}(\alpha, \alpha|F) + 2\tilde{\mu}_{mkt}\text{cov}(\alpha, \beta|F) + \tilde{\mu}_{mkt}\text{cov}(\beta, \beta|F)$  we can write the  $(i, j)$  elements of  $V_{2,2}^*$  as

$$\begin{aligned}
\text{cov}(y_{i,T+1}, y_{j,T+1}|F) &= \tilde{b}'_{2,1,i} \text{var}(\alpha + \tilde{\mu}_{mkt}\beta|F) \tilde{b}_{2,1,j} + [\tilde{\mu}_{mkt} \ \tilde{\mu}_1]' \text{cov}(b_{2,i}, b_{2,j}|F) [\tilde{\mu}_{mkt} \ \tilde{\mu}_1] \\
& \quad + \text{tr}[\text{cov}(b_{2,1,i}, b_{2,1,j}|F) \text{var}(\alpha + \tilde{\mu}_{mkt}\beta|F)] \\
& \quad + \tilde{b}'_{2,i} V_{1mkt,1mkt}^* \tilde{b}_{2,j} + \text{tr}[V_{1mkt,1mkt}^* \text{cov}(b_{2,i}, b_{2,j}|F)] + \tilde{\sigma}_{2,ij}. \quad (\text{A27})
\end{aligned}$$

### A5.5 $V_{2,mkt}^*$

Let  $B_{2,mkt}$  be the  $mkt$  betas in  $B_2$  and  $B_{2,1}$  the loadings of  $F_2$  on  $F_1$  in (A1). Recall that  $X_1 = [MKT \ F_1]$ . We can compute  $V_{2,mkt}^*$  using the decomposition

$$\begin{aligned}
\text{cov}(F_{2,T+1}, mkt_{T+1}|F) &= \text{cov}[\mathbb{E}(F_{2,T+1}|B_2, B_1, \mu_{mkt}, F), \mathbb{E}(mkt_{T+1}|B_2, B_1, \mu_{mkt}, F)|F] \\
& \quad + \mathbb{E}(\text{cov}(F_{2,T+1}, mkt_{T+1}|B_2, B_1, \mu_{mkt}, F)|F)
\end{aligned}$$

For the first term note that

$$\begin{aligned}
\mathbb{E}(F_{2,T+1}|B_2, B_1, \mu_{mkt}, F) &= \mathbb{E}([\mu_{mkt} \ \mu_1]B_2|F) \\
&= \mathbb{E}([\mu_{mkt} \ [1 \ \mu_{mkt}]B_1]B_2|F) \\
&= \mathbb{E}([\mu_{mkt}B_{2,mkt} \ [1 \ \mu_{mkt}]B_1B_{2,1}]|F)
\end{aligned}$$

and therefore the first term is

$$\text{cov}[\mathbb{E}(F_{2,T+1}|B_2, B_1, \mu_{mkt}, F), \mathbb{E}(mkt_{T+1}|B_2, B_1, \mu_{mkt}, F)|F] = \tilde{B}_{2,mkt} \text{var}(\mu_{mkt}|F) + \tilde{B}_{2,1} \tilde{\beta} \text{var}(\mu_{mkt}|F)$$

The second term equals

$$\begin{aligned}
\mathbb{E}(\text{cov}(F_{2,T+1}, mkt_{T+1}|B_2, B_1, \mu_{mkt}, F)|F) &= \mathbb{E}(\text{cov}([mkt_{T+1} \ \alpha + \beta mkt_{T+1} + \varepsilon_{1,T+1}]B_2 + \varepsilon_{2,T+1}, mkt_{T+1})|F) \\
&= \mathbb{E}(B_{2,mkt}V_{mkt,mkt} + B_{2,1}\beta V_{mkt,mkt}|F) \\
&= \tilde{B}_{2,mkt} \tilde{V}_{mkt,mkt} + \tilde{B}_{2,1} \tilde{\beta} \tilde{V}_{mkt,mkt}
\end{aligned}$$

Therefore

$$V_{2,mkt}^* = (\tilde{B}_{2,mkt} + \tilde{B}_{2,1} \tilde{\beta}) V_{mkt,mkt}^* \quad (\text{A28})$$

### A5.6 $V_{2,1}^*$

We can represent  $V_{2,1}^*$  in terms of its  $(i, j)$  element. Let  $y_i$  be an element of  $F_2$  and  $x_j$  be an element of  $F_1$  then from (A1)

$$\begin{aligned}
y_{i,T+1} &= B_{2,i}[mkt_{T+1} \ f_{1,T+1}] + \varepsilon_{2,i,T+1} \\
&= B_{2,i}[mkt_{T+1} \ \alpha + \beta mkt_{T+1} + \varepsilon_{1,T+1}] + \varepsilon_{2,i,T+1}
\end{aligned}$$

and from (A2)

$$\begin{aligned}
x_{j,T+1} &= \alpha_j + \beta_j mkt_{T+1} + \varepsilon_{1,j,T+1} \\
&= [1 \ mkt_{T+1}] b_{1,j} + \varepsilon_{1,j,T+1}
\end{aligned}$$

since  $b_{1,j} = (\alpha_j \ \beta_j)'$ .

The predictive covariance between  $y_{i,T+1}$  and  $x_{j,T+1}$ , the  $(i, j)$  element of  $V_{2,1}^*$ , can be obtained using the decomposition

$$\begin{aligned} \text{cov}(y_{i,T+1}, x_{j,T+1}|F) &= \text{cov}[\text{E}(y_{i,T+1}|b_2, b_1, F), \text{E}(x_{j,T+1}|b_2, b_1, F)|F] \\ &\quad + \text{E}(\text{cov}(y_{i,T+1}, x_{j,T+1})|b_2, b_1, F)|F) \end{aligned} \quad (\text{A29})$$

To compute the first term note that  $\text{E}(y_{i,T+1}|b_2, b_1, F) = [\tilde{\mu}_{mkt} \ [1 \ \tilde{\mu}_{mkt}]B_1]b_{2,i} = [\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}$ . Also  $\text{E}(x_{j,T+1}|b_2, b_1, F) = [1 \ \tilde{\mu}_{mkt}]b_{1,j} = \alpha_j + \beta_j\tilde{\mu}_{mkt}$ .

$$\text{cov}[\text{E}(y_{i,T+1}|b_2, b_1, F), \text{E}(x_{j,T+1}|b_2, b_1, F)|F] = \text{cov}[[\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}, \tilde{\mu}_{mkt}\beta_j|F] \quad (\text{A30})$$

Now to compute this covariance we can condition on  $b_1$

$$\begin{aligned} \text{cov}[[\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}, \tilde{\mu}_{mkt}\beta_j|F] &= \text{cov}[\text{E}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}|b_1, F), \text{E}(\alpha_j + \tilde{\mu}_{mkt}\beta_j|b_1, F)|F] \\ &\quad + \text{E}(\text{cov}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]b_{2,i}, \alpha_j + \tilde{\mu}_{mkt}\beta_j|b_1, F)|F) \\ &= \text{cov}([\tilde{\mu}_{mkt} \ \alpha + \tilde{\mu}_{mkt}\beta]\tilde{b}_{2,i}, \alpha_j + \tilde{\mu}_{mkt}\beta_j|F) \\ &= \text{cov}[(\alpha + \beta\tilde{\mu}_{mkt})\tilde{b}_{2,i,1}, \alpha_j + \tilde{\mu}_{mkt}\beta_j|F] \\ &= \tilde{b}'_{2,i,1}[\text{cov}(\alpha, \alpha_j|F) + \tilde{\mu}_{mkt}\text{cov}(\alpha, \beta_j|F) \\ &\quad + \tilde{\mu}_{mkt}\text{cov}(\beta, \alpha_j|F) + \tilde{\mu}_{mkt}^2\text{cov}(\beta, \beta_j|F)]. \end{aligned} \quad (\text{A31})$$

For the second term

$$\text{cov}(y_{i,T+1}, x_{j,T+1}|b_2, b_1, F) = B_{2,i,mkt}\beta_j V_{mkt,mkt}^* + B'_{2,i,1}V_{1,mkt}^*\beta_j + B'_{2,i,1}\tilde{\sigma}_{1,j} \quad (\text{A32})$$

where  $B_{2,mkt}$  and  $B_{2,i,1}$  are the elements of  $B_{2,i}$  that load on  $mkt$  and  $F_1$  respectively and  $\sigma_{1,j}$  is the  $j$ th column of  $\sigma_1$ . Taking the expectation

$$\text{E}[\text{cov}(y_{i,T+1}, x_{j,T+1}|b_2, b_1, F)|F] = \tilde{B}_{2,i,mkt}\tilde{\beta}_j V_{mkt,mkt}^* + \tilde{B}'_{2,i,1}V_{1,mkt}^*\tilde{\beta}_j + \tilde{B}'_{2,i,1}\tilde{\sigma}_{1,j} \quad (\text{A33})$$

Therefore

$$\begin{aligned} \text{cov}(y_{i,T+1}, x_{j,T+1}|F) &= \tilde{b}'_{2,i,1}[\text{cov}(\alpha, \alpha_j|F) + \tilde{\mu}_{mkt}\text{cov}(\alpha, \beta_j|F) \\ &\quad + \tilde{\mu}_{mkt}\text{cov}(\beta, \alpha_j|F) + \tilde{\mu}_{mkt}^2\text{cov}(\beta, \beta_j|F)] \\ &\quad + \tilde{B}_{2,i,mkt}\tilde{\beta}_j V_{mkt,mkt}^* + \tilde{B}'_{2,i,1}V_{1,mkt}^*\tilde{\beta}_j + \tilde{B}'_{2,i,1}\tilde{\sigma}_{1,j}. \end{aligned} \quad (\text{A34})$$

## References

- Asness, C., Frazzini, C., 2013. The devil in HML's details. *Journal of Portfolio Management* 39, 49–68.
- Ball, R., Gerakos, J., Linnainmaa, J. T., Nikolaev, V., 2016. Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics* 121, 28–45.
- Barillas, F., Shanken, J., 2017. Which alpha?. *Review of Financial Studies* 30, 1316–1338.
- Barillas, F., Shanken, J., 2018. Comparing asset pricing models. *Journal of Finance* 73(2), 715–754.
- Bawa, V., Brown, S., Klein, R., 1979. *Estimation Risk and Optimal Portfolio Choice*. North-Holland, Amsterdam.
- Brandt, Michael W., Portfolio Choice Problems, in Y. Ait-Sahalia and L.P. Hansen, *Handbook of Financial Econometrics, Volume 1: Tools and Techniques*, North Holland, 2010.
- Brown, Stephen J., 1976, *Optimal portfolio choice under uncertainty: A Bayesian approach*, Ph.D. dissertation, University of Chicago.
- Carhart, M. M., 1997. On persistence in mutual fund performance. *Journal of Finance* 52, 57–82.
- Cochrane, J. H., 2011. Presidential address: Discount rates. *The Journal of Finance*, 66(4), 1047–1108.
- Cremers, M., 2002, Stock return predictability: A Bayesian model selection perspective, *Review of Financial Studies* 15, 1223–1249.
- Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3–56.

- Fama, E. F., French, K. R., 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1–22.
- Fama, E. F., French, K. R., 2017. Choosing factors. Working Paper.
- Frost, Peter A., and James E. Savarino, 1986, An empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* 21, 293–305.
- Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. *Econometrica* 57, 1121–1152.
- Harvey, Campbell R., and Guofu Zhou, 1990, Bayesian inference in asset pricing tests, *Journal of Financial Economics* 26, 221–254.
- Harvey, C. R., Y. Liu, and H. Zhu (2015) ... and the cross-section of expected returns. *Review of Financial Studies* 29(1), 5–68.
- Hou, K., Xue, C., Zhang, L., 2015a. Digesting anomalies: An investment approach. *Review of Financial Studies* 28, 650–705.
- Hou, Kewei, Chen Xue, and Lu Zhang, 2015b, A comparison of new factor models, Working paper, Ohio State University.
- Hou, K., Xue, C., and Zhang, L. 2016. Replicating anomalies. Technical report, National Bureau of Economic Research.
- Jeffreys, Harold, 1961, *Theory of Probability* (Oxford University Press, New York).
- Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance* 48, 65–91.
- Jobson, J. D., B. Korkie, and V. Ratti, 1979, Improved estimation for Markowitz portfolios using JamesStein type estimators, in *Proceedings of the American Statistical Association, Business and Economic Statistics Section* (American Statistical Association, Washington)



- Jorion, Philippe, 1985, International portfolio diversification with estimation risk, *Journal of Business* 58, 259–278.
- Jorion, Philippe, 1986, BayesStein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* 21, 279–292.
- Jorion, Philippe, 1991, Bayesian and CAPM estimators of the means: Implications for portfolio selection, *Journal of Banking and Finance* 15, 717–727.
- Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385–424.
- Klein, Roger W., and Vijay S. Bawa, 1976, The effect of estimation risk on optimal portfolio choice, *Journal of Financial Economics* 3, 215–231.
- Leamer, E.E., 1978. *Specification Searches: Ad Hoc inference with Non-experimental data.* Wiley, New York.
- Linnainmaa, Juhani, and Michael Roberts, 2017, The history or the cross section of stock returns, Working paper. 49, 141–168.
- MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the capital asset pricing model, *Journal of Financial Economics* 38, 3–28.
- McLean, R. D. and J. Pontiff, 2016. Does academic research destroy stock return predictability? *Journal of Finance* 71(1), 5–32.
- McCulloch, Robert, and Peter E. Rossi, 1991, A Bayesian approach to testing the arbitrage pricing theory, *Journal of Econometrics* 49, 141–168.
- Pastor, L., 2000. Portfolio selection and asset pricing models. *Journal of Finance* 55, 179–223.
- Pastor, L., Stambaugh, R.F., 2000. Comparing asset pricing models: an investment perspective. *Journal of Financial Economics* 56, 335–381.

- Shanken, Jay, 1987, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics* 19, 195–215.
- Stambaugh, Robert F., 1997, Analyzing investments whose histories differ in length, *Journal of Financial Economics* 45, 285–331.
- Tu, J., and G. Zhou, 2004, Data-generating process uncertainty: What difference does it make in portfolio decisions? *Journal of Financial Economics* 72, 385–421.
- Wahal, S. 2016. The profitability and investment premium: Pre-1963 evidence. *Journal of Financial Economics*. Forthcoming

Table 1

## Posterior Model Probabilities

Panel A reports posterior model probabilities (in percent) for the seven models with highest probability (ranked at the end of the sample) when the prior multiple is set to 1.5 times the sample market Sharpe ratio,  $Sh(\text{Mkt})$ . Panel B reports posterior model probabilities (in percent) for the seven categorical models with highest probability (ranked at the end of the sample) when the prior multiple is set to 1.5 times the sample market Sharpe ratio,  $Sh(\text{Mkt})$ . Seven prior multiples are considered: 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, and 5.0. The sample period is January 1967 to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm, and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (Mkt or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). For the categorical models in Panel B the factors consist of the market (Mkt) and momentum (UMD) factors, along with the categorical factors (two versions of each) size, value (VAL), investment (INV), and profitability (PROF). The prior is set so that  $Sh_{\max} = \text{prior multiple} \times Sh(\text{Mkt})$ , where  $Sh_{\max}$  is the square root of the expectation of the maximum squared Sharpe ratio with six factors included, taken with respect to the prior under the alternative that the nonmarket factor alphas are nonzero.  $Sh(\text{Mkt})$  is 0.115 and the sample Sharpe ratio for the top-ranked model is 0.500 or 4.3 times  $Sh(\text{Mkt})$ .

Panel A: Posterior Model Probabilities

Model	1.25	1.5	1.75	2	2.5	3	5
Mkt HMLm UMD SMB ROE IA	47.2	56.0	60.2	62.6	65.2	66.4	67.8
Mkt HMLm UMD ME ROE IA	21.5	19.6	18.0	16.9	15.7	15.1	14.1
Mkt HMLm UMD SMB ROE CMA	16.4	15.9	15.2	14.7	14.0	13.7	13.0
Mkt HMLm UMD ME ROE CMA	7.7	5.8	4.8	4.2	3.6	3.3	2.9
Mkt HMLm UMD ROE IA	3.1	1.2	0.7	0.5	0.4	0.4	0.4
Mkt HMLm UMD SMB ROE	0.5	0.6	0.6	0.7	0.8	0.9	1.4
Mkt HMLm UMD ROE CMA	1.2	0.4	0.2	0.2	0.1	0.1	0.1

Panel B: Posterior Categorical Model Probabilities

Model	1.25	1.5	1.75	2	2.5	3	5
Mkt UMD SIZE VAL INV PROF	93.4	97.3	98.2	98.4	98.5	98.4	97.8
Mkt UMD VAL INV PROF	4.3	1.6	0.9	0.7	0.5	0.5	0.5
Mkt UMD SIZE VAL PROF	0.8	0.8	0.8	0.9	1.0	1.1	1.7
Mkt SIZE VAL INV PROF	1.2	0.2	0.1	0.1	0.0	0.0	0.0
Mkt UMD VAL PROF	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Mkt SIZE INV PROF	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Mkt UMD SIZE INV PROF	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 2  
In-Sample results

This table reports the sample monthly Sharpe ratios of the optimal mean variance sample portfolio as well as those of the Bayesian Model Averaged procedure. Seven prior multiples are considered: 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, and 5.0. The sample period is January 1967 to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm ,and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (Mkt or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). We also report the annualized mean and standard deviation of the resulting portfolio for three different levels of risk aversion:  $\gamma=2.53$ ,  $\gamma=5$ , and  $\gamma=10$ .

		Sample Optimal	BMA						
			1.25	1.5	1.75	2	2.5	3	5
	SR	0.5526	0.4742	0.4929	0.4962	0.4973	0.4980	0.4982	0.4984
$\gamma=2.53$	Mean	144.31	47.41	69.11	81.75	89.79	99.09	104.08	111.22
	Std	75.67	28.97	40.65	47.76	52.35	57.70	60.57	64.71
$\gamma=5$	Mean	73.19	24.16	35.14	41.53	45.60	50.31	52.83	56.44
	Std	38.28	14.66	20.57	24.16	26.48	29.19	30.64	32.74
$\gamma=10$	Mean	36.77	12.26	17.75	20.95	22.98	25.34	26.60	28.40
	Std	19.14	7.33	10.28	12.08	13.24	14.59	15.32	16.37

Table 3  
Optimal Allocations

Panel A reports optimal allocations per \$1 of wealth for a mean variance investors. The first column shows the allocations for the sample-based optimal portfolio. Other columns report the optimal allocations of the Bayesian Model Averaged procedure. Seven prior multiples are considered: 1.25, 1.5, 1.75, 2, 2.5, 3, and 5. The sample period is January 1967 to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm, and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (Mkt or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). Panel A reports the results for a risk aversion of 2.53, Panel B uses  $\gamma = 5$  and Panel C  $\gamma = 10$ .

		Panel A: $\gamma = 2.53$						
	Sample	BMA						
		1.25	1.5	1.75	2	2.5	3	5
RF	-1.469	-0.518	-0.805	-0.979	-1.093	-1.227	-1.299	-1.404
Mkt	2.469	1.518	1.805	1.979	2.093	2.227	2.299	2.404
UMD	4.984	1.123	1.709	2.044	2.256	2.502	2.633	2.824
SMB	5.554	0.595	1.001	1.255	1.421	1.618	1.724	1.877
HML	-7.033	0.002	-0.001	-0.001	-0.001	-0.001	-0.002	-0.002
CMA	0.699	0.124	0.156	0.171	0.178	0.185	0.187	0.189
RMW	-1.518	0.006	0.001	0.001	0.001	0.001	0.001	0.001
ME	-3.033	0.240	0.309	0.331	0.340	0.346	0.347	0.346
HMLm	11.527	1.950	2.976	3.561	3.932	4.360	4.589	4.922
IA	3.083	0.671	1.041	1.277	1.432	1.614	1.711	1.845
ROE	8.322	2.220	3.391	4.067	4.497	4.994	5.260	5.641

		Panel B: $\gamma = 5$						
	Sample	BMA						
		1.25	1.5	1.75	2	2.5	3	5
RF	-0.249	0.232	0.087	-0.001	-0.059	-0.127	-0.163	-0.216
Mkt	1.249	0.768	0.913	1.001	1.059	1.127	1.163	1.216
UMD	2.521	0.568	0.865	1.034	1.142	1.266	1.332	1.429
SMB	2.810	0.301	0.507	0.635	0.719	0.818	0.872	0.950
HML	-3.558	0.001	0.000	-0.001	-0.001	-0.001	-0.001	-0.001
CMA	0.354	0.063	0.079	0.087	0.090	0.093	0.095	0.096
RMW	-0.768	0.003	0.001	0.000	0.001	0.001	0.001	0.001
ME	-1.534	0.121	0.156	0.168	0.172	0.175	0.175	0.175
HMLm	5.831	0.986	1.506	1.802	1.989	2.206	2.322	2.490
IA	1.560	0.339	0.527	0.646	0.724	0.816	0.866	0.933
ROE	4.210	1.123	1.715	2.058	2.275	2.526	2.661	2.854

Panel C:  $\gamma = 10$

	Sample	BMA						
		1.25	1.5	1.75	2	2.5	3	5
RF	0.375	0.616	0.543	0.499	0.471	0.437	0.418	0.392
Mkt	0.625	0.384	0.457	0.501	0.529	0.563	0.582	0.608
UMD	1.261	0.284	0.432	0.517	0.571	0.633	0.666	0.714
SMB	1.405	0.151	0.253	0.317	0.359	0.409	0.436	0.475
HML	-1.779	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CMA	0.177	0.031	0.040	0.043	0.045	0.047	0.047	0.048
RMW	-0.384	0.002	0.000	0.000	0.000	0.000	0.000	0.000
ME	-0.767	0.061	0.078	0.084	0.086	0.087	0.088	0.088
HMLm	2.916	0.493	0.753	0.901	0.995	1.103	1.161	1.245
IA	0.780	0.170	0.263	0.323	0.362	0.408	0.433	0.467
ROE	2.105	0.562	0.858	1.029	1.138	1.263	1.330	1.427

Table 4  
Real time results

The table reports statistics for the returns generated by the real-time dynamic trading strategie. The estimation window starts after observing 20 years of data. Seven prior multiples are considered: 1.25, 1.5, 1.75, 2, 2.5, 3, and 5. The sample period is January 1967 to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm ,and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (Mkt or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). Panel A reports Sharpe ratios, the average position in the risk free asset, and annualized mean and standard deviations of portfolio returns. Panel B reports certainty equivalent returns. All statistics are reported for three different levels of risk aversion:  $\gamma=2.53$ ,  $\gamma=5$ , and  $\gamma=10$ .

		Panel A: Real Time Sharpe Ratios							
		Sample	BMA						
			1.25	1.5	1.75	2	2.5	3	5
$\gamma=2.53$	SR	0.3597	0.2995	0.3428	0.3401	0.3392	0.3392	0.3398	0.3411
	$\bar{w}_{rf}$	-1.64	-0.29	-0.56	-0.76	-0.91	-1.11	-1.23	-1.43
	Mean	149.97	28.82	57.20	72.74	84.02	98.97	108.02	122.62
	Std	120.36	27.78	48.17	61.74	71.51	84.21	91.77	103.76
$\gamma=5$	SR	0.3669	0.3316	0.3610	0.3543	0.3514	0.3496	0.3493	0.3495
	$\bar{w}_{rf}$	-0.34	0.35	0.21	0.11	0.03	-0.07	-0.13	-0.23
	Mean	77.44	16.15	30.51	38.37	44.08	51.64	56.22	63.61
	Std	60.92	14.06	24.40	31.26	36.21	42.64	46.47	52.53
$\gamma=10$	SR	0.3815	0.3956	0.3975	0.3828	0.3760	0.3704	0.3684	0.3664
	$\bar{w}_{rf}$	0.33	0.67	0.61	0.55	0.52	0.47	0.44	0.39
	Mean	40.31	9.67	16.85	20.78	23.63	27.41	29.70	33.39
	Std	30.50	7.05	12.23	15.67	18.14	21.36	23.27	26.31

		Panel B: Real Time Certainty Equivalents							
		Sample	BMA						
			1.25	1.5	1.75	2	2.5	3	5
$\gamma=2.53$		-33.76	19.03	27.77	24.40	19.17	9.02	1.19	-13.94
$\gamma=5$		-15.61	11.19	15.59	13.87	11.21	6.06	2.09	-5.58
$\gamma=10$		-6.33	7.17	9.34	8.47	7.13	4.53	2.54	-1.31

Table 5  
 In Sample results for the redundant factors experiment

This table reports the sample monthly Sharpe ratios of the sample-optimal mean variance portfolio as well as those of the Bayesian Model Averaged procedure. Seven prior multiples are considered: 1.25, 1.5, 1.75, 2.0, 2.5, 3.0, and 5.0. The sample period is January 1967 to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm ,and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (Mkt or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). The benchmark model for this experiment is {Mkt SMB HML UMD}. The alphas of CMA, RMW, ME, HMLm, IA and ROE on this benchmark model are made equal to zero after the initial period of 20 years. We also report the annualized mean and standard deviation of the resulting portfolio for three different levels of risk aversion:  $\gamma=2.53$ ,  $\gamma=5$ , and  $\gamma=10$ .

		Tangency	BMA						
			1.25	1.5	1.75	2	2.5	3	5
	SR	0.3521	0.3200	0.3428	0.3471	0.3485	0.3494	0.3496	0.3494
$\gamma=2.53$	Mean	58.45	25.23	35.75	41.47	45.02	48.99	51.03	53.57
	Std	48.22	22.88	30.28	34.70	37.51	40.73	42.40	44.55
$\gamma=5$	Mean	29.75	12.94	18.26	21.16	22.95	24.96	25.99	27.28
	Std	24.39	11.58	15.32	17.56	18.98	20.60	21.45	22.54
$\gamma=10$	Mean	15.06	6.65	9.31	10.76	11.66	12.66	13.18	13.82
	Std	12.20	5.79	7.66	8.78	9.49	10.30	10.72	11.27



Table 6  
Optimal Allocations for the Redundant Factors Experiment

Panel A reports optimal allocations per \$1 of wealth for a mean variance investors. The first column shows the allocations for the sample tangency portfolio. Other columns report the optimal allocations of the Bayesian Model Averaged procedure. Seven prior multiples are considered: 1.25, 1.5, 1.75, 2, 2.5, 3, and 5. The sample period is January 1967 to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm, and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (Mkt or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). The benchmark model for this experiment is {Mkt SMB HML UMD}. The alphas of CMA, RMW, ME, HMLm, IA and ROE on this benchmark model are made equal to zero after the initial period of 20 years. Panel A reports the results for a risk aversion of 2.53, Panel B uses  $\gamma = 5$  and Panel C  $\gamma = 10$ .

	Tangency	BMA						
		1.25	1.5	1.75	2	2.5	3	5
RF	-1.003	-0.382	-0.579	-0.693	-0.765	-0.849	-0.895	-0.964
Mkt	2.003	1.382	1.579	1.693	1.765	1.849	1.895	1.964
UMD	3.187	1.130	1.814	2.168	2.386	2.634	2.767	2.974
SMB	2.252	0.248	0.386	0.474	0.528	0.583	0.604	0.587
HML	-0.536	0.227	0.098	0.073	0.061	0.049	0.045	0.040
CMA	0.418	0.241	0.215	0.240	0.257	0.277	0.286	0.294
RMW	-0.404	0.165	0.131	0.123	0.118	0.110	0.106	0.098
ME	-0.890	0.148	0.254	0.300	0.326	0.350	0.357	0.341
HMLd	4.053	0.999	1.893	2.323	2.585	2.879	3.032	3.243
IA	0.990	0.362	0.564	0.652	0.705	0.763	0.792	0.821
ROE	2.861	0.561	1.082	1.353	1.519	1.700	1.784	1.832

	Tangency	BMA						
		1.25	1.5	1.75	2	2.5	3	5
RF	-0.013	0.301	0.201	0.144	0.107	0.065	0.042	0.006
Mkt	1.013	0.699	0.799	0.856	0.893	0.935	0.958	0.994
UMD	1.612	0.572	0.918	1.097	1.207	1.333	1.400	1.504
SMB	1.139	0.125	0.195	0.240	0.267	0.295	0.305	0.297
HML	-0.271	0.115	0.050	0.037	0.031	0.025	0.023	0.020
CMA	0.212	0.122	0.109	0.121	0.130	0.140	0.145	0.149
RMW	-0.204	0.084	0.066	0.062	0.059	0.056	0.053	0.049
ME	-0.450	0.075	0.129	0.152	0.165	0.177	0.181	0.172
HMLd	2.050	0.505	0.958	1.175	1.308	1.456	1.534	1.641
IA	0.501	0.183	0.285	0.330	0.357	0.386	0.401	0.415
ROE	1.448	0.284	0.547	0.684	0.769	0.860	0.903	0.927

Panel C:  $\gamma = 10$

	Tangency	BMA						
		1.25	1.5	1.75	2	2.5	3	5
RF	0.493	0.650	0.601	0.572	0.554	0.532	0.521	0.503
Mkt	0.507	0.350	0.399	0.428	0.446	0.468	0.479	0.497
UMD	0.806	0.286	0.459	0.548	0.604	0.666	0.700	0.752
SMB	0.570	0.063	0.098	0.120	0.133	0.147	0.153	0.148
HML	-0.135	0.057	0.025	0.018	0.015	0.012	0.011	0.010
CMA	0.106	0.061	0.054	0.061	0.065	0.070	0.072	0.074
RMW	-0.102	0.042	0.033	0.031	0.030	0.028	0.027	0.025
ME	-0.225	0.037	0.064	0.076	0.082	0.089	0.090	0.086
HMLd	1.025	0.253	0.479	0.588	0.654	0.728	0.767	0.820
IA	0.250	0.092	0.143	0.165	0.178	0.193	0.200	0.208
ROE	0.724	0.142	0.274	0.342	0.384	0.430	0.451	0.463

Table 7

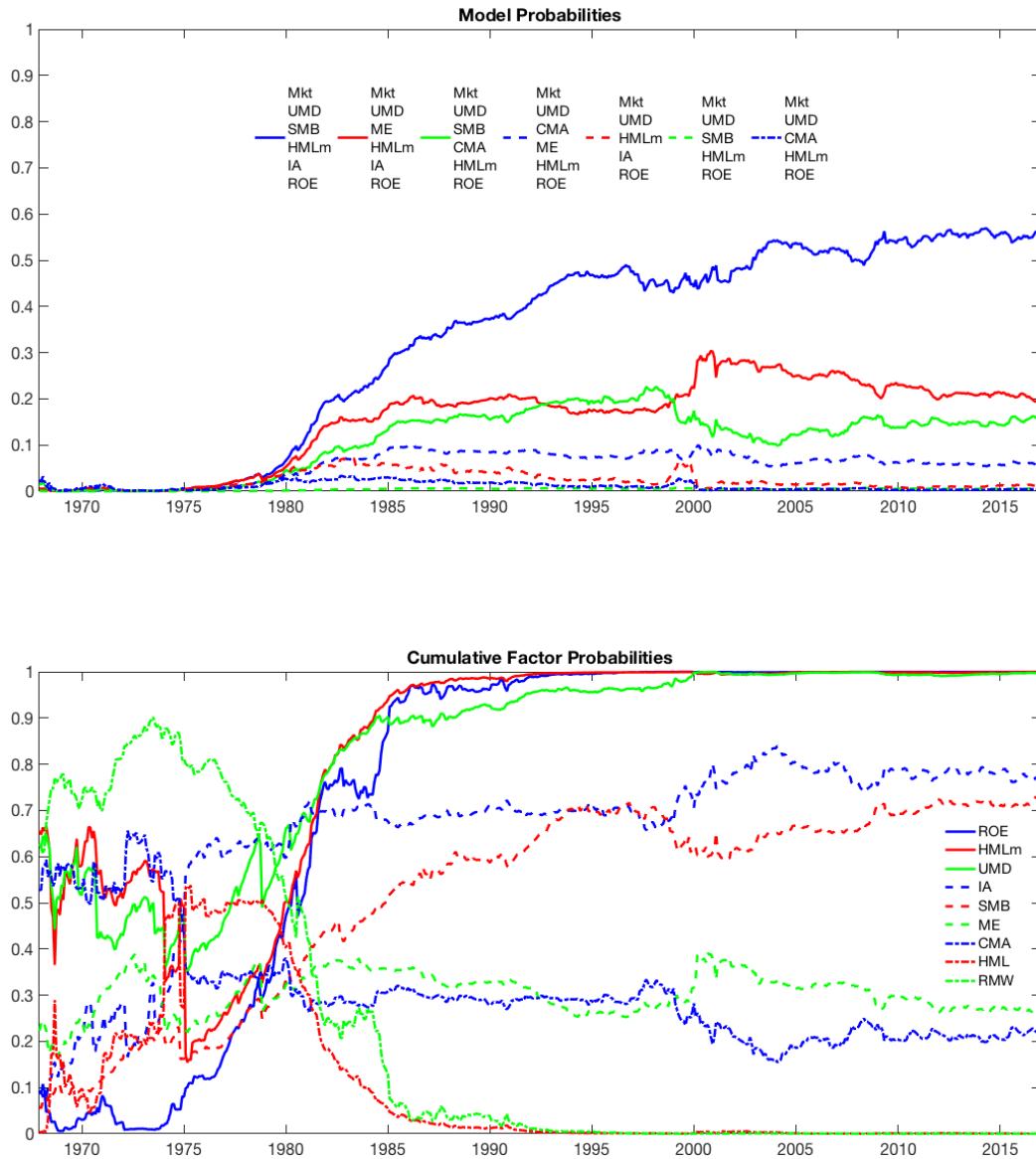
## Real Time results for the redundant factors experiment

The table reports statistics for the returns generated by the real-time dynamic trading strategy for the experiment. The estimation window starts after observing 20 years of data. The benchmark model for this experiment is {Mkt SMB HML UMD}. The alphas of CMA, RMW, ME, HMLm, IA and ROE on this benchmark model are made equal to zero after the initial period of 20 years. Seven prior multiples are considered: 1.25, 1.5, 1.75, 2, 2.5, 3, and 5. The sample period is January 1967 to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm, and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (Mkt or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). Panel A reports Sharpe ratios, the average position in the risk free asset, and annualized mean and standard deviations of portfolio returns. Panel B reports certainty equivalent returns. All statistics are reported for three different levels of risk aversion:  $\gamma=2.53$ ,  $\gamma=5$ , and  $\gamma=10$ .

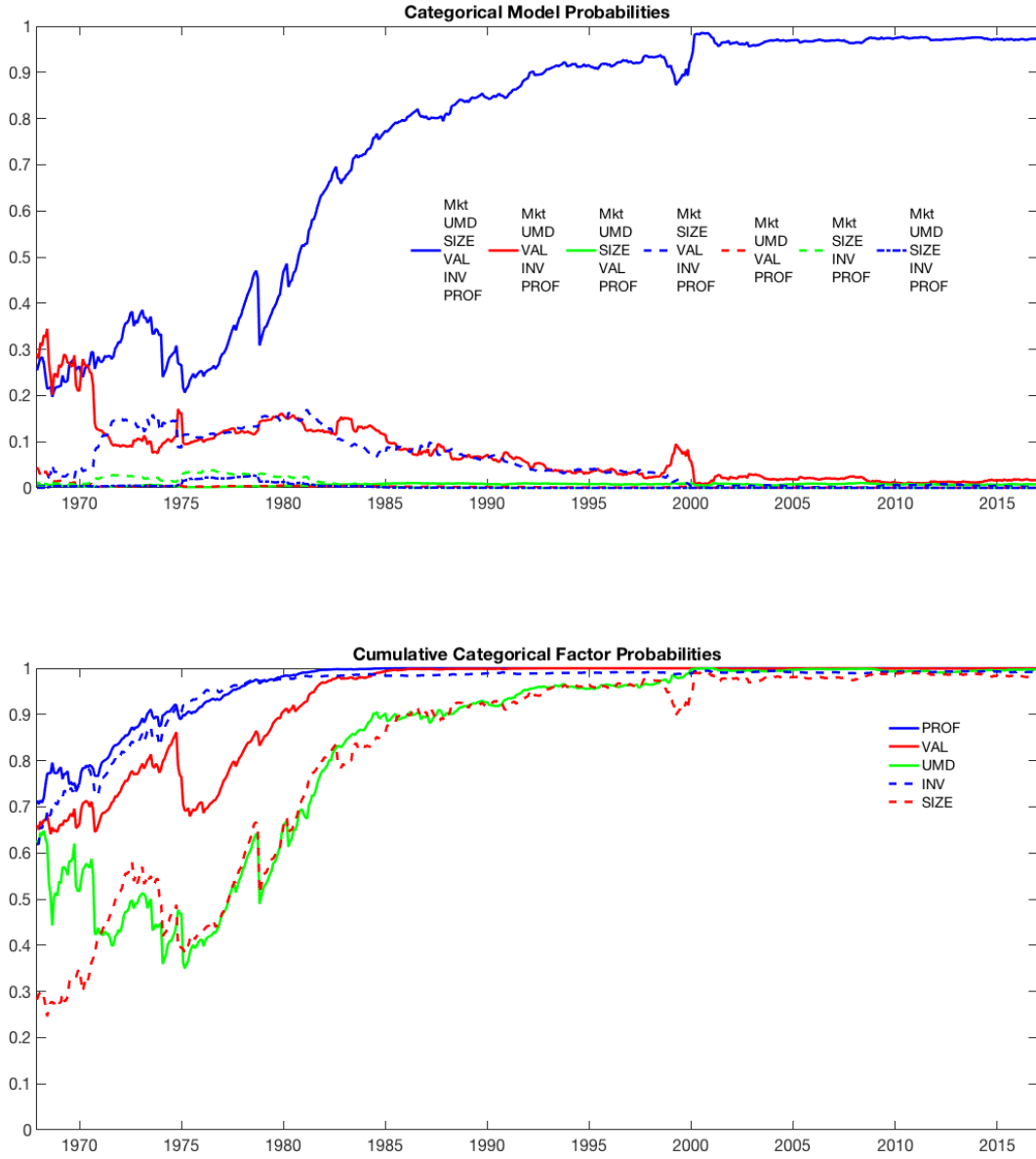
Panel A: Real Time Sharpe Ratios									
		Tangency	BMA						
			1.25	1.5	1.75	2	2.5	3	5
$\gamma=2.53$	SR	0.0558	0.1776	0.1314	0.1228	0.1190	0.1157	0.1141	0.1122
	Mean	16.49	13.82	17.23	20.43	22.73	25.70	27.44	30.13
	Std	85.33	22.47	37.87	48.02	55.12	64.14	69.40	77.56
$\gamma=5$	SR	0.0663	0.2172	0.1549	0.1414	0.1353	0.1296	0.1270	0.1237
	Mean	9.91	8.56	10.29	11.91	13.07	14.57	15.45	16.82
	Std	43.17	11.38	19.17	24.31	27.89	32.46	35.12	39.25
$\gamma=10$	SR	0.0875	0.2960	0.2024	0.1789	0.1679	0.1577	0.1530	0.1470
	Mean	6.55	5.87	6.73	7.54	8.12	8.88	9.32	10.00
	Std	21.59	5.73	9.61	12.17	13.97	16.25	17.58	19.64

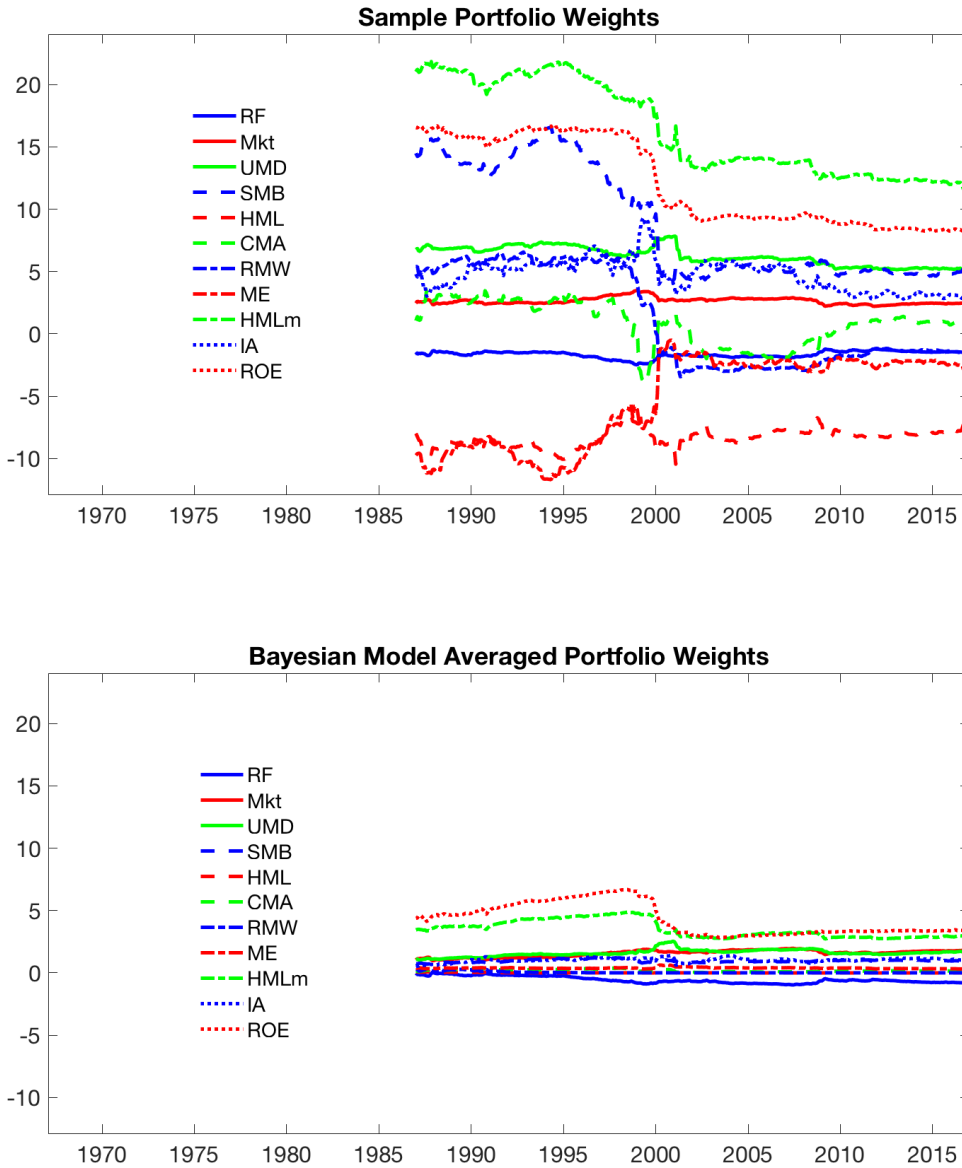
Panel B: Real Time Certainty Equivalents									
		Tangency	BMA						
			1.25	1.5	1.75	2	2.5	3	5
$\gamma=2.53$		-75.85	7.42	-0.96	-8.82	-15.80	-26.48	-33.65	-46.17
$\gamma=5$		-36.81	5.32	1.08	-2.90	-6.44	-11.84	-15.47	-21.80
$\gamma=10$		-16.83	4.23	2.11	0.11	-1.65	-4.36	-6.17	-9.34



**Figure 1. Model probabilities and cumulative factor probabilities.** The top panel plots the time series of posterior model probabilities for the seven models with highest probability (ranked at the end of the sample). The sample periods are recursive, beginning in January 1967 and ending each month up to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm, and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (SMB or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). The bottom panel plots the time series of cumulative posterior probabilities for each of the 10 factors. The prior is set so that  $Sh_{\max} = 1.5 \times Sh(\text{Mkt})$ , where  $Sh(\text{Mkt})$  is the sample Mkt Sharpe ratio and  $Sh_{\max}$  is the square root of the expectation of the maximum squared Sharpe ratio with six factors included, taken with respect to the prior under the alternative that the nonmarket factor alphas are nonzero.



**Figure 2. Categorical model probabilities and cumulative categorical factor probabilities.** The top panel plots the times series of posterior model probabilities for the seven categorical models with highest probability (ranked at the end of the sample). The sample periods are recursive, beginning in January 1967 and ending each month up to December 2016. The factors consist of the market (Mkt) and momentum (UMD) factors, along with the categorical factors (two versions of each) size, value (VAL), investment (INV), and profitability (PROF). The bottom panel plots the time series of cumulative factor probabilities for each of the five categorical-model factors. The prior is set so that  $Sh_{\max} = 1.5 \times Sh(\text{Mkt})$ , where  $Sh(\text{Mkt})$  is the sample Mkt Sharpe ratio and  $Sh_{\max}$  is the square root of the expectation of the maximum squared Sharpe ratio with six factors included, taken with respect to the prior under the alternative that the alphas are nonzero.



**Figure 3. Optimal allocations.** The figures plot the time series of optimal allocations for a mean-variance-optimizing investor with relative risk aversion equal to 2.53. The top panel computes the allocation using the sample mean and covariance up to a given point in time. The bottom panel shows the allocations for the Bayesian investor that computes the predictive mean and variance of returns by combining the asset pricing models. The sample periods are recursive, beginning in January 1967 and ending each month up to December 2016. Models are based on 10 prominent factors: the Fama and French (2015) factors (Mkt, HML, SMB, CMA, RMW), the HXZ (2015a) factors (Mkt, ME, ROE, IA), the Asness and Frazzini (2013) value factor HMLm, and the Carhart (1997) momentum factor UMD. Each model contains at most one factor from the following categories: size (SMB or ME), value (HML or HMLm), investment (CMA or IA), and profitability (RMW or ROE). The prior is set so that  $Sh_{\max} = 1.5 \times Sh(\text{Mkt})$ , where  $Sh(\text{Mkt})$  is the sample Mkt Sharpe ratio and  $Sh_{\max}$  is the square root of the expectation of the maximum squared Sharpe ratio with six factors included, taken with respect to the prior under the alternative that the nonmarket factor alphas are nonzero.