The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXVIII, NO. 3 • JUNE 2023

Integrating Factor Models

DORON AVRAMOV, SI CHENG, LIOR METZKER, and STEFAN VOIGT*

ABSTRACT

This paper develops a comprehensive framework to address uncertainty about the correct factor model. Asset pricing inferences draw on a composite model that integrates over competing factor models weighted by posterior probabilities. Evidence shows that unconditional models record near-zero probabilities, while postearnings announcement drift, quality-minus-junk, and intermediary capital are potent factors in conditional asset pricing. Out-of-sample, the integrated model performs well, tilting away from subsequently underperforming factors. Model uncertainty makes equities appear *considerably* riskier, while model disagreement about expected returns spikes during crash episodes. Disagreement spans all return components involving mispricing, factor loadings, and risk premia.

FINANCIAL ECONOMISTS HAVE IDENTIFIED A PLETHORA of firm characteristics that predict future stock returns (e.g., Cochrane (2011) and Harvey, Liu, and Zhu (2016)). The literature has further proposed two major approaches to reduce the expanding number of predictors. The first invokes economic rationales, for example, plausible restrictions on the admissible Sharpe ratio, the present-value model, and the q-theory, to identify a small set of common factors, while the second approach formulates the dependence of average returns on common factors or firm characteristics through regression regularization techniques including deep learning extensions. However, the collection of factors that matter most remains subject to research controversy.[1] Signifi-

*Doron Avramov is at Reichman University (IDC). Si Cheng is at Syracuse University. Lior Metzker is at the Hebrew University of Jerusalem. Stefan Voigt is at the University of Copenhagen and the Danish Finance Institute. We thank John Chao, Stefan Nagel, two anonymous referees, and seminar participants at the Chinese University of Hong Kong (Shenzhen), Hong Kong University of Science and Technology, Norwegian University of Science and Technology, Southern University of Science and Technology (Shenzhen), Taiwan National University, and Tel Aviv University for useful comments and suggestions. Doron Avramov acknowledges the Israel Science Foundation (ISF) (Grant Number: 288/18) for financial support. We have read *The Journal of Finance* disclosure policy and have no conflicts of interest to disclose.

Correspondence: Doron Avramov, Reichman University (IDC), Herzliya, Israel; e-mail: davramov@huji.ac.il

[1] See, for example, Ross (1976), Fama and French (2015), and Hou, Xue, and Zhang (2015) for the first approach, and Green, Hand, and Zhang (2017a), Light, Maslov, and Rytchkov (2017),

DOI: 10.1111/jofi.13226

cant uncertainty also extends to the choice of macro variables that potentially govern time-varying investment opportunities. Moreover, even if the econometrician has prior information about the identity of asset pricing factors and macro predictors, there is still uncertainty about whether the underlying model holds exactly or instead admits the possibility of mispricing.

Surprisingly, a comprehensive analysis of Bayesian model uncertainty has not been accounted for in formulating expected returns or in deriving mean-variance efficient portfolios. When addressing model uncertainty, the researcher's core tasks are identifying a universe of competing factor models, assessing the probability that a candidate model generates the observed dynamics of asset returns, and then integrating over the vast model universe using posterior probabilities as weights. This approach, termed Bayesian model averaging (BMA), yields an integrated model that summarizes the various sources of uncertainty about the joint dynamics of asset returns. With model uncertainty taken into account, inferences about the cross-section are conditioned on the entire information set instead of relying on the information contained in a single model. The Bayesian approach can therefore temper the data-snooping concerns identified in the literature (e.g., Harvey, Liu, and Zhu (2016), Harvey (2017), Hou, Xue, and Zhang (2020)).

In this paper, we develop and apply a framework to study average returns, the covariance matrix, and efficient portfolios in the presence of model uncertainty. Candidate models differ with respect to the collection of cross-sectional factors, the set of macro predictors, and the factor model specification, which either holds exactly or admits various degrees of mispricing. A key challenge of the framework is the formulation of model posterior probability or the probability that a candidate model generates the observed dynamics of asset returns. In particular, motivating economically interpretable priors for all parameters underlying the factor model and the dynamics of factor risk premia is essential. Our informed priors are weighted against both time-varying moments and model mispricing. The resulting model posterior probability employs sound economic intuition and penalizes model complexity to the extent that an incremental factor beyond the market or a macro predictor is retained only if it considerably improves pricing ability.

In the presence of model uncertainty, expected returns are a mixture of model-implied expected returns, where mixture stands for weighting based on model posterior probabilities. The covariance matrix consists of three components. The first is a mixture of model-implied covariance, assuming that the model parameters are known. The other two components arise from uncertainty about the correct factor model and its underlying parameters.

The first of these two components reflects estimation risk, that is, the risk that the underlying model parameters are estimated with error. The second

Manresa, Peñaranda, and Sentana (2017), Feng, Giglio, and Xiu (2020), Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020), Kozak, Nagel, and Santosh (2020), Chen, Pelger, and Zhu (2023), and Cong et al. (2021) for the second. Notably, the various specifications could disagree on the set of factors that matter most.

summarizes model disagreement. Intuitively, a stock appears riskier when there is greater disagreement among candidate models about its expected return. Like the Ridge regression approach, the disagreement component could make an otherwise ill-conditioned covariance matrix of stock returns readily invertible. Through the Bayesian approach, the predictive distribution of future returns integrates out the within-model parameter space (parameter uncertainty) and the model space (model uncertainty). Thus, the resulting efficient portfolios do not depend on a particular model or underlying parameters.

We apply the framework to sample data that consist of 14 asset pricing factors and 13 macro predictors from 1977 to 2016. The model universe exceeds 52 million specifications. We first examine some stylized model features. For a reasonable prior Sharpe ratio, the 10 (100, 500) top-ranked individual models account for a cumulative posterior probability of 30% (76%, 93%), suggesting no clear winner across the whole space of candidate models.[2] Instead, many distinct models record a positive and meaningful probability of governing the joint distribution of stock returns. While model selection would narrow focus to a single factor model, or a few models, the Bayesian approach integrates across the dynamics of nonzero probability models.

Even when prior beliefs are weighted against time-varying moments, our procedure uniformly favors conditional models and indicates that both factor loadings and risk premia vary with macroeconomic conditions.[3] Remarkably, in the presence of conditional factor models, the cumulative probability of unconditional models is practically zero. Likewise, while prior beliefs are weighted against mispricing, the analysis shows that time-varying mispricing appears with a high probability. Hence, zero-alpha (or even constant-alpha) models selected from the collection of factors and macro predictors may not adequately explain cross-sectional and time-series effects in returns.

Several findings on the strength of factors and macro predictors in the integrated model are worth noting. First, for a reasonable prior Sharpe ratio, postearnings announcement drift (PEAD, Daniel, Hirshleifer, and Sun (2020)) and quality-minus-junk (QMJ, Asness, Frazzini, and Pedersen (2019)) display a posterior inclusion probability of close to 100%, followed by investment (CMA, Fama and French (2015)), size (SMB, Fama and French (1993)), intermediary capital (ICR, He, Kelly, and Manela (2017)), and management (MGMT, Stambaugh and Yuan (2017))—all of which offer a posterior inclusion probability of at least 90%, highlighting their promise in pricing other factors. Moreover, despite the expanding factor zoo, several new factors proposed after 2015, including fundamental and behavioral factors, have incremental ability to price the existing factors.

[2] In our baseline case, the prior Sharpe ratio for the tangency portfolio is set to be 50% higher than the market Sharpe ratio. The top-ranked models describe factor models that record the highest posterior probabilities based on the Bayesian procedure.

[3] In separate tests based on multivariate predictive regressions, the specifications that include nonlinearities and interactions between macro predictors uniformly dominate linear specifications. This result further supports the notion that both factor loadings and risk premiums vary with economic conditions.

Second, while PEAD, QMJ, and ICR stand out across different prior specifications, the inclusion probabilities for SMB, CMA, and MGMT diminish for a high prior Sharpe ratio. In contrast, betting-against-beta (BAB, Frazzini and Pedersen (2014)) exhibits high inclusion probability only when the prior is tilted toward a high Sharpe ratio. Thus, the pricing abilities of widely explored factors depend on one's views about how large the Sharpe ratio could be. Bounding the prior Sharpe ratio to sensible values reinforces one group of factors (e.g., CMA and MGMT) while challenging others (e.g., BAB).

Likewise, bounding the Sharpe ratio has implications for the inclusion of macro predictors in a factor model specification. For instance, the yield on Treasury bills and the term spread appear with almost zero probability for sensible values of prior Sharpe ratios. However, when the prior Sharpe ratio is considerably higher, both macro items record a 90% probability of inclusion. The increasing probability supports the notion that strong in-sample evidence on predictability by aggregate variables could be associated with inadmissibly large Sharpe ratios. The out-of-sample evidence on time-series predictability is therefore often weaker.

Overall, the probability analysis supports a conditional model with a handful of factors that originate from distinct economic foundations rather than an established, well-known paradigm. For instance, while PEAD, QMJ, and ICR are proposed by three independent works, their combination has not been examined in the previous literature.

We next assess the out-of-sample performance of the Bayesian approach using tangency portfolios that are based on the integrated predictive distribution. We first compute the Sharpe ratio and downside risk for the tangency portfolio. For comparison, we consider four benchmark models that are widely used by academics and practitioners, namely, the capital asset pricing model (CAPM), the Fama-French three-factor model (Fama and French (1993)), the Fama-French six-factor model (Fama and French (2018)), and the AQR six-factor model (Frazzini, Kabiller, and Pedersen (2018)). We further consider the three top-ranked individual models, particularly the three highest posterior probability models based on the Bayesian approach.

The integrated model outperforms the benchmark models out-of-sample. It generates an annualized Sharpe ratio of 1.240, indicating an 8% improvement over the best benchmark model. To ensure that the tangency portfolio relies on admissible long and short positions, we further impose the Regulation-T constraint on stock holdings.[4] We find that the integrated model produces an out-of-sample annualized Sharpe ratio of 0.979, outperforming the best benchmark model by 25%. The Bayesian approach also mitigates the downside risk. Relative to benchmark models, the tangency portfolio based on the integrated

[4] Regulation-T of the Federal Reserve Board mandates maximum two-to-one leverage (e.g., Jacobs, Levy, and Starer (1999)). See the Financial Industry Regulatory Authority (FINRA) website for details: https://www.finra.org/rules-guidance/key-topics/margin-accounts. Formally, accounting for Regulation-T, the sum of the absolute values of long and short positions is constrained to be less than or equal to two, where two is obtained by dividing one by the initial margin of 50%.

model exhibits less negative skewness, lower excess kurtosis, and lower maximum drawdown, to the extent that there are only modest declines in portfolio value when the overall market drops significantly. In addition, while the top-ranked individual models display similar posterior probabilities, we observe more variations in their performance and relative strength. Therefore, model selection may provide an unstable description of asset return dynamics, while model integration improves the stability of forecasts.

It is also important to assess the performance of the global minimum variance portfolio (GMVP), which relies only on the covariance matrix of returns. The covariance matrix accounts for model uncertainty through a mixture of model-implied covariance, estimation risk, and model disagreement about expected returns. Thus, if model uncertainty has meaningful asset pricing implications, the GMVP based on the integrated model should generate payoffs characterized by relatively low-risk measures.

Indeed, the GMVP based on the integrated model generates improved measures of realized volatility and maximum drawdown. For instance, monthly realized volatility for GMVPs based on the benchmark models ranges between 0.956% and 2.127%, while it appears to be only 0.756% for the integrated model, indicating a 21% to 64% reduction in volatility. In addition, the maximum drawdown (across the entire sample) for the benchmark models ranges between 6% and 27%, compared to 5% for the integrated model. The lower volatility characterizing the Bayesian approach translates into a higher out-of-sample Sharpe ratio. The GMVP based on the integrated model generates an annualized Sharpe ratio of 1.101, outperforming the best benchmark model by 35%. These results highlight the sizable impact of model uncertainty on the covariance matrix of returns.

We next explore the integrated model's portfolio tilts. We find a positive correlation between tangency portfolio weights and subsequently realized factor returns for 9 out of 14 factors out-of-sample, with an average correlation of 4.5%. These results suggest that the tangency portfolio is tilted toward subsequently outperforming factors and away from underperforming factors. In comparison, the equal-weighted portfolio indicates a zero correlation. Focusing on subperiods of negative factor returns, the correlation between portfolio weights and realized factor returns increases to 10.5%. Thus, the integrated model is instrumental in mitigating adverse investment outcomes through factor rotation.

We conduct four final experiments to shed further light on the implications of model uncertainty for the investment opportunity set. First, we compare the sample variance of factor returns with the sample average of the perceived variance based on the integrated model. Excluding model uncertainty, the sample variance should exceed the time-series average of the conditional variance, as the latter uses information from macro variables. With model uncertainty and estimation risk taken into account, however, we find conflicting forces underlying the variance comparison. Empirically, most of the factors display remarkably higher variance through the lens of the integrated model. For perspective, the integrated model variance is, on average, 53% higher than the sample variance across all factors. The findings suggest that the mixture of estimation risk and the model disagreement components jointly have a sizable impact on the ex-ante risk of equities. Thus, a Bayesian agent that accounts for uncertainty about the factor model specification perceives the traded factors to be *considerably* riskier than what would be implied by the sample volatility.

Second, we examine the time variation in model disagreement about expected returns. Following the literature on information theory, we use increasing entropy to measure the contribution of model uncertainty to the covariance matrix. While the increase in entropy is modest on average, it spikes dramatically during major market downturns, such as Black Monday in October 1987 and the recent financial crisis starting in September 2008. Compared to a benchmark value of one, indicating no entropy increase, the full sample average is 1.010 but increases to 1.069 at the 99$^{\text{th}}$ percentile and reaches a maximum of 1.379. We then estimate the contribution of each factor to the overall entropy increase. The time-varying model uncertainty component is driven primarily by the market, MGMT, and ICR factors. All three factors have a maximum contribution of at least 10% to the total increase in entropy in both the full sample and various subperiods.

Third, we analyze the underlying forces driving model disagreement over time. Specifically, we consider seven model-specific components that govern expected returns, including (i) fixed mispricing, (ii) time-varying mispricing, (iii) fixed factor loadings with fixed risk premia, (iv) fixed factor loadings with time-varying risk premia, (v) time-varying factor loadings with fixed risk premia, (vi) time-varying factor loadings with time-varying risk premia, and (vii) time-varying risk premia. We find that model disagreement appears in all components and is highly skewed. For instance, the maximum disagreement in time-varying factor loadings with time-varying risk premia (time-varying mispricing) is, on average, 9.27 (5.24) times its mean across all factors. Importantly, during crash episodes, candidate models significantly disagree more on mispricing, factor loadings, and risk premia, which all jointly contribute to the overall increase in entropy.

Finally, we investigate whether candidate models differ in their implied portfolio choice and performance. We find that candidate models display meaningful dispersion in tangency portfolio weights and investment returns, especially during major market downturns. Therefore, accounting for model uncertainty is highly relevant for academics and practitioners in portfolio construction and risk management.

Taken together, our approach identifies potent factors in conditional asset pricing. Out-of-sample, the integrated model delivers outperforming strategies, tilting away from underperforming factors. The Bayesian approach also reduces the downside risk and volatility of efficient portfolios. The set of findings is robust to imposing plausible constraints on the admissible Sharpe ratios and equity positions. In our setting, model disagreement can be mapped into disagreement among heterogeneous investors about expected stock returns. We show that such disagreement spikes around market downturns and is attributed to mispricing, factor loadings, and risk premia, both the fixed and time-varying components. Overall, in the presence of model uncertainty, equities appear *considerably* riskier from the perspective of a Bayesian agent.

To our knowledge, Avramov and Chao (2006) is the first study to formally compare asset pricing models, both nested and nonnested, using posterior probabilities. Subsequent studies include Anderson and Cheng (2016), Stambaugh and Yuan (2017), Barillas and Shanken (2018), Chib and Zeng (2019), Chib, Zeng, and Zhao (2020), Bryzgalova, Huang, and Julliard (2023), and Chib, Zhao, and Zhou (2023). Our study differs from these in four major respects. First, related work on model comparison and factor selection is typically based on rankings of posterior probabilities, while we propose a novel approach that integrates over the space of candidate models. Second, existing studies mainly focus on unconditional models, while we consider time-varying mispricing, factor loadings, and risk premia, and we provide evidence for nonlinear dependence between expected returns and macro items. Third, prior beliefs about the entire parameter space are economically interpretable in our setup, while statistically motivated training samples are often employed to formulate informed priors. Finally, our comprehensive examination of the integrated model implications for the investment opportunities set is novel.

The remainder of the paper proceeds as follows. Section I derives a general methodology for analyzing asset pricing with model uncertainty. Section II derives the posterior probabilities for factor models. Section III describes the data. Section IV presents a probability analysis of pricing models and individual factors and predictors. Section V assesses the out-of-sample performance of the integrated model through both tangency portfolios and GMVPs. Section VI presents evidence on the riskiness of equities in the presence of model uncertainty and dissects the time-series variation in model disagreement about expected returns. Section VII concludes the paper.

I. Asset Pricing with Model Uncertainty

A key challenge in our framework is the formulation of model posterior probabilities or the probability that a candidate factor model generates the observed dynamics of asset returns. For one, it is essential to formulate economically interpretable priors for all parameters underlying the factor model and the dynamics of factor risk premia. In a general context, combining an improper prior with a likelihood function yields a well-defined posterior distribution. In computing posterior probabilities, however, the prior density must be fully specified and avoid undefined constants characterizing a flat prior.[5] Otherwise, the posterior probability is not interpretable.

A large body of work in financial economics motivates economically meaningful priors on a subset of the parameter space. For instance, Pástor and Stambaugh (1999) account for prior information about mispricing, or alpha, which translates into a certain degree of belief in model pricing abilities. Kozak, Nagel, and Santosh (2020) impose an economically motivated prior on Stochas-

[5] See, for example, the discussions in Kass and Raftery (1995) and Poirier (1995).

tic Discout Factor (SDF) coefficients. They introduce a prior for mispricing that can be expressed in terms of the relation between risk premia of principal component factors and their eigenvalues.

In this paper, we propose economically interpretable priors for the *entire* parameter space underlying beta pricing specifications. The informed priors are weighted against both time-varying moments and model mispricing, or the priors favor an exact (zero-alpha) unconditional factor model. The resulting model posterior probability employs sound economic intuition. It penalizes model complexity to the extent that an incremental factor beyond the market or macro predictor is retained only if it considerably improves pricing abilities.

To set the stage, let r_t denote an N-vector of excess returns on test assets, let f_t denote a K-vector of factors that are return spreads, and let z_t denote an M-vector of macro variables that are potentially related to the distribution of future returns. The length of the time series is denoted by T, and the t subscript represents time t realizations.

Excess returns are modeled by the asset pricing regression

$$r_{t+1} = \alpha(z_t) + \beta(z_t) f_{t+1} + u_{r,t+1}, \tag{1}$$

while factors are formulated using the time-series predictive regression

$$f_{t+1} = \alpha_f + a_F z_t + u_{f,t+1}. \tag{2}$$

The residuals $[u'_{r,t+1}, u'_{f,t+1}]'$ are orthogonal innovations assumed to obey the normal distribution: $u_{r,t+1} \sim N(0, \Sigma_{RR})$ and $u_{f,t+1} \sim N(0, \Sigma_{FF})$. The intercept $\alpha(z_t)$ and slope $\beta(z_t)$ coefficients are given by $\alpha(z_t) = \alpha_0 + \alpha_1 z_t$ and $\beta(z_t) = \beta_0 + \beta_1(I_K \otimes z_t)$, where $\otimes$ denotes the Kronecker product and I_K is an identity matrix of size K. Excess stock returns can then be rewritten as

$$r_{t+1} = \alpha_0 + \alpha_1 z_t + \beta_0 f_{t+1} + \beta_1(I_K \otimes z_t) f_{t+1} + u_{r,t+1}. \tag{3}$$

The intercepts α_0 and α_1 represent an N-vector and an $N \times M$ matrix reflecting fixed and time-varying model mispricing, respectively. When the factors are portfolio spreads, an asset pricing model implies that both alpha components are equal to zero. When only $\alpha_0 \neq 0$, time-invariant model mispricing is present, while when $\alpha_1 \neq 0$, model mispricing varies with macro conditions.[6] Next, $\beta(z_t)$ is an $N \times K$ matrix of potentially time-varying factor sensitivities, where β_0 is an $N \times K$ matrix, and β_1 is an $N \times (KM)$ matrix. Factor loadings are time-varying if $\beta_1 \neq 0$. The formulation in equation (2) allows risk premia to be time-varying ($a_F \neq 0$).

The asset pricing specification in equations (2) and (3) gives rise to multiple sources of uncertainty characterizing stock return dynamics. We start with mispricing uncertainty. Does a prespecified factor model explain the

[6] Avramov (2004) shows that mean-variance portfolios that account for time-varying mispricing outperform competing specifications. Ferson and Harvey (1999) support time-varying mispricing in asset pricing tests.

cross-sectional variation in average stock returns? Pástor and Stambaugh (1999) show that uncertainty about model pricing ability could be substantial. Gibbons, Ross, and Shanken (1989), among others, derive classical asset pricing statistics to test zero-alpha restrictions (see Campbell, Lo, and MacKinlay (1997) and Cochrane (2009) for a comprehensive review), while Harvey and Zhou (1990), McCulloch and Rossi (1991), Kandel and Stambaugh (1995), and Avramov and Chao (2006) develop Bayesian asset pricing tests.

There is also substantial uncertainty about the identity of asset pricing factors. Remarkably, Harvey, Liu, and Zhu (2016) count 316 factors, and Hou, Xue, and Zhang (2020) cover 452 anomalies. Two major approaches have been proposed to address the expanding dimension of the cross-section. The first identifies a small number of factors based on sound economic intuition. For instance, motivated by the dividend discount valuation model, Fama and French (2015) propose a five-factor model that augments the original market, size, and value factors with investment and profitability factors. Hou, Xue, and Zhang (2015) and Hou et al. (2021) propose q-factor models that draw on the q-theory of investment. Stambaugh and Yuan (2017) identify two mispricing factors based on 11 anomalies studied in Stambaugh, Yu, and Yuan (2012). The second approach proposes shrinkage methods such as Lasso, Ridge, and their extensions (e.g., Green, Hand, and Zhang (2017b), DeMiguel et al. (2020), Feng, Giglio, and Xiu (2020), Freyberger, Neuhierl, and Weber (2020), and Kozak, Nagel, and Santosh (2020)). Shrinkage methods employ a trade-off by reducing the variance of estimated parameters at the cost of introducing a bias. Nevertheless, the true set of asset pricing factors remains subject to debate.

A third type of uncertainty concerns the identity of macro variables that forecast changing investment opportunities. Prior work addresses this uncertainty through the predictive regression setup. When M macro variables are suspected to be relevant in predicting future returns, there are 2^M competing predictive regressions. In classical econometrics, model selection criteria are typically employed to select among competing models. At the heart of model selection, one applies a specific criterion (e.g., Bayesian information criterion) to select a single model and then continues as if the model is correct with unit probability. Using various model selection criteria, Bossaerts and Hillion (1999) and Welch and Goyal (2008) detect no out-of-sample return predictability even when the in-sample evidence is solid.

Counter to the classical approach, BMA is a comprehensive method that directly follows from Bayes rule and is justified from a decision-making perspective. The Bayesian method assigns posterior probabilities to each of the 2^M predictive regressions and then uses the probabilities as weights on the individual return forecasting models to obtain a composite weighted model. As shown by Avramov (2002), the Bayesian approach displays robust predictive power relative to model selection criteria. In addition, the Bayesian model integration approach does detect evidence on out-of-sample predictability by macro variables.

In this paper, we propose a novel Bayesian approach to study time-series and cross-sectional effects in asset returns when the true factor model and its underlying parameters are uncertain. We first consider a universe of candidate asset pricing factors and macro predictors. We then compute the posterior probability for each candidate model. Models differ with respect to the three sources of uncertainty described above. Table I, Panel A, lists the candidate models considered in the paper.

The symbols $\mathbb{M}_1$ and $\mathbb{M}_2$ represent the family of unconditional models without mispricing ($\mathbb{M}_1$) and with fixed mispricing ($\mathbb{M}_2$), while $\mathbb{M}_3$ and $\mathbb{M}_4$ represent the family of conditional models with time-varying factor loadings and risk premia. In particular, $\mathbb{M}_3$ excludes mispricing, while $\mathbb{M}_4$ allows for both fixed and time-varying mispricing. Within these families, models differ in their inclusion of asset pricing factors ($\mathbb{M}_1$ and $\mathbb{M}_2$) or their inclusion of both factors and predictors ($\mathbb{M}_3$ and $\mathbb{M}_4$).

In the presence of model uncertainty, expected stock returns are given by

$$E[r_{t+1}|D] = \sum_{l=1}^{L} P(\mathcal{M}_l|D)E[r_{t+1}|\mathcal{M}_l, D], \tag{4}$$

where D stands for the observed data, which consists of a balanced panel of N test assets, K factors, and M macro predictors through T periods, l is a model-specific subscript, $\mathcal{M}_l$ is a candidate factor model, $P(\mathcal{M}_l|D)$ is the model posterior probability, $E[r_{t+1}|\mathcal{M}_l, D]$ is the model-specific expected return, and L is the total number of candidate models.

The covariance matrix of stock returns can be decomposed into three components. We start with a two-component decomposition given by

$$\text{Var}[r_{t+1}|D] = V_t + \Omega_t, \tag{5}$$

where $V_t = \sum_{l=1}^{L} P(\mathcal{M}_l|D)\text{Var}[r_{t+1}|\mathcal{M}_l, D]$ and Ω_t reflects variation due to model disagreement about expected stock returns.

In particular, V_t is a mixture (probability-weighted average) of model-implied covariance matrices that takes into account the stochastic nature of the underlying model parameters in the Bayesian setting. We can decompose V_t further into two terms. The first is a mixture of model-implied covariance matrices, assuming that model parameters are known. The second is a mixture of model-implied estimation risks. Estimation risk, or parameter uncertainty, comes into play because the parameters in our setting are stochastic. Note that V_t can fluctuate over the business cycle due to time-varying factor loadings.

Next, Ω_t is given by

$$\begin{aligned}\Omega_t &= \text{Var}(E[r_{t+1}|\mathcal{M}_l, D]) \\ &= \sum_{l=1}^{L} P(\mathcal{M}_l|D)(E[r_{t+1}|\mathcal{M}_l, D] - E[r_{t+1}|D])(E[r_{t+1}|\mathcal{M}_l, D] - E[r_{t+1}|D])'. \end{aligned} \tag{6}$$

The Ω_t component summarizes the disagreement across candidate models about expected stock returns. An asset's incremental variation is larger when

Table I

Model Specifications and Parameter Values

Panel A lists the specifications of asset pricing models considered in the paper. $\mathbb{M}_1$ and $\mathbb{M}_2$ represent the unconditional models without mispricing ($\mathbb{M}_1$) and with fixed mispricing ($\mathbb{M}_2$), respectively. $\mathbb{M}_3$ and $\mathbb{M}_4$ represent the conditional models with time-varying factor loadings and risk premia, with $\mathbb{M}_3$ also being without mispricing and $\mathbb{M}_4$ allowing for both fixed and time-varying mispricing, respectively. Panel B presents the corresponding parameter values used for calculating the marginal likelihood in equation (IA.73) in the Internet Appendix III.

Panel A: Model Specifications

	Mispricing	Factor Loading	Risk Premia	Data-Generating Process	
Model	$\alpha(z_t)$	$\beta(z_t)$	a_F	Return	Factor
$\mathbb{M}_1$	$\alpha = 0$	$\beta_1 = 0$	$a_F = 0$	$r_{t+1} = \beta_0 f_{t+1} + u_{r,t+1}$	$f_{t+1} = \alpha_f + u_{f,t+1}$
$\mathbb{M}_2$	$\alpha \neq 0$	$\beta_1 = 0$	$a_F = 0$	$r_{t+1} = \alpha_0 + \beta_0 f_{t+1} + u_{r,t+1}$	$f_{t+1} = \alpha_f + u_{f,t+1}$
$\mathbb{M}_3$	$\alpha = 0$	$\beta_1 \neq 0$	$a_F \neq 0$	$r_{t+1} = \beta_0 f_{t+1} + \beta_1 (I_k \otimes z_t) f_{t+1} + u_{r,t+1}$	$f_{t+1} = \alpha_f + a_F z_t + u_{f,t+1}$
$\mathbb{M}_4$	$\alpha \neq 0$	$\beta_1 \neq 0$	$a_F \neq 0$	$r_{t+1} = \alpha_0 + \alpha_1 z_t + \beta_0 f_{t+1} + \beta_1 (I_k \otimes z_t) f_{t+1} + u_{r,t+1}$	$f_{t+1} = \alpha_f + a_F z_t + u_{f,t+1}$

Panel B: Parameter Values for the Marginal Likelihood

Model	Q_R	ν_R	Q_F	ν_F
$\mathbb{M}_1$	$(N+K-k)k$	0	k	$N+K-k-1$
$\mathbb{M}_2$	$(N+K-k)(1+k)$	-1	k	$N+K-k-1$
$\mathbb{M}_3$	$(N+K-k)(k+km)$	$-km$	$k(1+m)$	$N+K-k-m-1$
$\mathbb{M}_4$	$(N+K-k)(1+m+k+km)$	$-(k+1)m-1$	$k(1+m)$	$N+K-k-m-1$

candidate models disagree more about its expected returns. When restricting Ω_t to be diagonal, the matrix $\text{Var}[r_{t+1}|D]$ can be readily invertible even when V_t is singular or ill-conditioned.[7] Thus, adding Ω_t resembles the Ridge regression penalty, but there are important differences. In Ridge regressions, the variance of returns takes the form $\text{Var}[r_{t+1}|\mathcal{M}, D] = \tilde{V}_t + \gamma I_N$, where $\tilde{V}_t$ is a frequentist-based estimate of the covariance matrix of returns, in which each of its elements is smaller (in absolute values) than the Bayesian counterpart due to estimation risk, and γ corresponds to a homogeneous shrinkage intensity toward the identity matrix of order N, the number of test assets.

The covariance matrix decomposition in equation (5) is similar to the shrinkage methods proposed by Ledoit and Wolf (2003, 2004), which have been shown to improve volatility forecasting in high-dimensional setups. However, Ledoit and Wolf (2003, 2004) propose shrinkage toward a parsimonious target, whereas the posterior predictive variance imposes asset-specific shrinkage toward the grand mean, V_t, in proportion to the general agreement among candidate models about mean returns.

In sum, the integrated model has a three-component covariance matrix, which includes (i) a mixture of model-implied covariance matrices, assuming that model parameters are known, (ii) a mixture of model-implied estimation risks, and (iii) model disagreement about expected returns.

While BMA follows directly from Bayes' rule, there are other approaches to model combination. Examples include decision-based model combinations per Billio et al. (2013) and optimal prediction pooling per Geweke and Amisano (2011).[8] Ex-ante, BMA would be optimal under several loss functions, including log loss and squared error loss (Hoeting et al. (1999)).

II. Deriving Posterior Probabilities

A. General Formulation

Let θ denote the unique parameter space for every candidate model. The parameter space consists of the intercept and slope coefficients in equations (2) and (3) as well as the covariance matrices. Combining the prior density on the parameters, $\pi(\theta|\mathcal{M}_l)$, and the likelihood based on observing data, $\mathcal{L}(D|\theta, \mathcal{M}_l)$, yields the posterior distribution, $\pi(\theta|D, \mathcal{M}_l)$. The posterior reflects the distribution of unknown parameters θ given (i) prior views, (ii) the observed data D, and (iii) the particular factor model $\mathcal{M}_l$.

[7] In the empirical analyses, the investment universe is relatively small. We can therefore keep Ω_t general enough to enable covariances due to cross-model disagreements.

[8] Neither approach shows promise when applied to our setting. First, we examined utility-based combination weights as a function of the realized certainty equivalent or the Sharpe ratio as a suitable objective for a mean-variance investor. The resulting combination weights concentrate all mass on a single underperforming model. Second, the implementation of optimal prediction pools in the spirit of Geweke and Amisano (2011) requires sampling from a posterior predictive distribution for each model to evaluate log predictive scores. In our vast model universe, such a procedure is computationally infeasible even with supercomputing capacities available, and narrowing the focus to a manageable universe of unconditional models leads to weak performance.

An intermediate input in computing the model posterior probability is the model marginal likelihood, denoted by $m(D|\mathcal{M}_l)$. Following Chib (1995), the marginal likelihood is computed (dropping the model-specific subscript to ease notation) as

$$m(D|\mathcal{M}) = \int_\theta \mathcal{L}(D|\theta, \mathcal{M})\pi(\theta|\mathcal{M})d\theta$$

$$= \frac{\mathcal{L}(D|\theta, \mathcal{M})\pi(\theta|\mathcal{M})}{\pi(\theta|D, \mathcal{M})}. \quad (7)$$

The marginal likelihood in equation (7) does not depend on θ, as it integrates out the entire parameter space. By doing so, the marginal likelihood provides a consistent form to adjust for model complexity and thus guards against overfitting.

Next, the posterior probability of model $\mathcal{M}$ is given by

$$P(\mathcal{M}|D) = \frac{m(D|\mathcal{M})P(\mathcal{M})}{\sum_{l=1}^{L} m(D|\mathcal{M}_l)P(\mathcal{M}_l)}, \quad (8)$$

where $P(\mathcal{M}_l)$ is the prior probability that model $\mathcal{M}_l$ is correct. Without a compelling reason to favor, ex-ante, one model over another, we choose equal probabilities.[9]

B. Posterior Probabilities for Factor Models

Given the general formulation, we next attempt to compute marginal likelihoods for competing factor models. The prior distribution is based on a hypothetical sample of length T_0.[10] In that sample, the means and the covariance matrices of stock returns, factors, and predictors are set equal to the actual sample counterparts given by

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

$$\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t \qquad \hat{V}_f = \frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})',$$

$$\bar{z} = \frac{1}{T}\sum_{t=0}^{T-1} z_t \qquad \hat{V}_z = \frac{1}{T}\sum_{t=0}^{T-1}(z_t - \bar{z})(z_t - \bar{z})',$$

[9] Notably, for a Bayesian agent with a stronger prior tilt toward particular models or individual factors or predictors, our framework can be adjusted to accommodate an unequal prior allocation.

[10] Using statistics from the actual sample to specify some of the parameters of the prior distribution is commonly termed "empirical Bayes" (Robbins (1956, 1964)). Our prior specification draws on Kandel and Stambaugh (1996), Pástor and Stambaugh (1999, 2000, 2002a, 2002b), Avramov (2002), and Avramov and Wermers (2006).

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t \qquad \hat{V}_y = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y})', \tag{9}$$

where $y_t = [r_t', f_t']'$.

The prior sample is also weighted against predictability by macro variables and against mispricing. Thus, on the basis of equation (3), regressing r_t on a constant term, z_{t-1}, f_t, and the interaction terms $f_t \otimes z_{t-1}$ yields zero estimates for α_0, α_1, and β_1 in the prior sample. Hence, the prior densities of α_0, α_1, and β_1 are centered around zero. Moreover, as T_0 increases, the informed prior about underlying parameters becomes tighter. Below, we derive the marginal likelihood conditioned on T_0, and then provide steps for computing T_0.

Similar to Avramov and Chao (2006) and Barillas and Shanken (2018), all candidate models contain the market factor. The set of test assets is then adjusted for each model based on the included factors. Specifically, K denotes the maximal number of factors. When a candidate model $\mathcal{M}_l$ contains $k < K$ factors, the other $K - k$ "redundant" factors are included in the test assets in addition to the N base assets. This is reasonable because a parsimonious model is helpful only if it prices the remaining assets correctly, including test assets and traded factors. This specification also ensures that the marginal likelihoods for all models are conditioned on the same data set.

Marginal likelihoods are first formulated for models with time-varying parameters. See Internet Appendixes I.A and I.B for the unrestricted and the restricted cases, respectively.[11] The marginal likelihood for the most general family of models, $\mathbb{M}_4$, is given by (oppressing the model subscript to ease notation)

$$m(D|\mathbb{M}_4) = \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)(T_0+k)+\frac{1}{2}k(m+1)} \times \left[\frac{T}{T^*}\right]^{\frac{1}{2}(N+K-k)T} \times$$
$$\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^* - (k+1)m - 1\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0 - (k+1)m - 1\right]\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^* + N + K - k - m - 1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0 + N + K - k - m - 1\right]\right)}\right]\times$$
$$\left[\frac{|R'R - R'F\left(F'F\right)^{-1}F'R|^{\frac{1}{2}(T_0-(k+1)m-1)}}{|R'R - \tilde{\Phi}'W'W\tilde{\Phi}|^{\frac{1}{2}(T^*-(k+1)m-1)}}\right]\left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-m-1)}}{|S_F|^{\frac{1}{2}(T^*+N+K-k-m-1)}}\right]. \tag{10}$$

The marginal likelihood terms are all defined in Internet Appendix I.A.

The last two terms in the marginal likelihood reflect the cross-sectional fit, while the remaining terms describe the penalty factor due to model complexity. As more factors or predictors are included, model pricing ability could improve. However, the potential improvement is associated with increasing complexity. These two conflicting forces on the marginal likelihood indicate that the ultimate inclusion of a variable is subject to a rigorous trade-off.

[11] The Internet Appendix is available in the online version of the article on *The Journal of Finance* website.

Internet Appendix II derives the marginal likelihood for the case in which unconditional models admit the possibility of mispricing (Internet Appendix II.A) and the case in which mispricing is excluded (Internet Appendix II.B). The marginal likelihood for $\mathbb{M}_2$ models is given by

$$m(D|\mathbb{M}_2) = \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)(T_0+k)+\frac{1}{2}k} \times \left[\frac{T}{T^*}\right]^{\frac{1}{2}(N+K-k)T} \times$$
$$\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}[T^*-1]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}[T_0-1]\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}[T^*+N+K-k-1]\right)}{\Gamma_k\left(\frac{1}{2}[T_0+N+K-k-1]\right)}\right]\times$$
$$\left[\frac{|R'R-R'F\left(F'F\right)^{-1}F'R|^{\frac{1}{2}(T_0-1)}}{|R'R-\tilde{\Phi}'W'W\tilde{\Phi}|^{\frac{1}{2}(T^*-1)}}\right]\left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-1)}}{|T^*\hat{V}_f|^{\frac{1}{2}(T^*+N+K-k-1)}}\right]. \tag{11}$$

Internet Appendix III summarizes the marginal likelihoods for all families of models, including $\mathbb{M}_1$ and $\mathbb{M}_3$. We note that the marginal likelihood in the conditional case is invariant to a linear transformation of z_t in equation (3). Internet Appendix IV provides a detailed proof.

C. Setting T_0

To complete the marginal likelihood derivation, we must set T_0. First, as shown in the Internet Appendix V, the parameters α_0 and α_1 in equation (3) obey the joint prior distribution

$$\text{vec}\left([\alpha_0, \alpha_1]'\right)|\Sigma_{RR}, D \sim N(0, \Sigma_{RR}\otimes B_{11}), \tag{12}$$

where B_{11} is a $(1+m)\times(1+m)$ matrix given by

$$B_{11} = \begin{pmatrix} 1+\bar{z}'\hat{V}_z^{-1}\bar{z}+\bar{f}'\hat{V}_f^{-1}\bar{f}+\bar{z}'\hat{V}_z^{-1}\bar{z}\times\bar{f}'\hat{V}_f^{-1}\bar{f} & -\bar{z}'\hat{V}_z^{-1}\times(1+\bar{f}'\hat{V}_f^{-1}\bar{f}) \\ -\hat{V}_z^{-1}\bar{z}\times(1+\bar{f}'\hat{V}_f^{-1}\bar{f}) & \hat{V}_z^{-1}\times(1+\bar{f}'\hat{V}_f^{-1}\bar{f}) \end{pmatrix}. \tag{13}$$

The unconditional variance of total mispricing is then equal to

$$\text{Var}(\alpha|\Sigma_{RR}, D) = \text{Var}\left(\alpha_0+\alpha_1'z|\Sigma_{RR}, D\right) = \frac{\Sigma_{RR}}{T_0}\left(1+SR^2_{\max}+m(1+SR^2_{\max})\right), \tag{14}$$

where $SR^2_{\max}$ is the largest attainable Sharpe ratio based on investments in the benchmarks only, and m is the number of predictors the model retains, ranging from zero for the IID model to M for the all-inclusive model.

Next, following Barillas and Shanken (2018), we formulate the prior on alpha according to

$$\alpha|\Sigma_{RR}, D \sim N(0, \eta\Sigma_{RR}), \tag{15}$$

where $\eta > 0$ controls for the prior spread. It then follows that $\alpha'(\eta\Sigma_{RR})^{-1}\alpha$ has a chi-square distribution with $N+K-k$ degrees of freedom. Hence,

$E(\alpha'\Sigma_{RR}^{-1}\alpha|\Sigma_{RR}, D) = \eta(N + K - k)$. Gibbons, Ross, and Shanken (1989) specify $\hat{\alpha}'\hat{\Sigma}_{RR}^{-1}\hat{\alpha}$ as the difference between two squared Sharpe ratios, that is

$$\hat{\alpha}'\hat{\Sigma}_{RR}^{-1}\hat{\alpha} = \widehat{SR}^2(R, F) - \widehat{SR}^2(F), \quad (16)$$

where $\widehat{SR}^2(F)$ is based on benchmark factors only and $\widehat{SR}^2(R, F)$ employs both benchmark factors and test assets. In our setup, due to the rotation between factors and "redundant" factors, (R, F) consists of the maximal number of factors and test assets. Hence, $\widehat{SR}^2(R, F)$ is identical across all considered models. In contrast, the second term $\widehat{SR}(F)$ varies across models, attaining its minimum value for the CAPM and its maximum when all K factors are retained. In the spirit of Barillas and Shanken (2018) and Chib, Zeng, and Zhao (2020), we set the expected value of the chi-squared distributed variable to the maximum value for the admissible Sharpe ratio relative to the market, that is, $SR_{\max} = SR(R, F) = \tau SR(Mkt)$, where τ refers to the prior Sharpe ratio multiple. To illustrate, for $\tau = 1.5$, the prior Sharpe ratio for the tangency portfolio based on a candidate model is 50% higher than the market Sharpe ratio.

It then follows that

$$E(\alpha'\Sigma_{RR}^{-1}\alpha|\Sigma_{RR}, D) = \eta\big(N + K - k\big) = \big(\tau^2 - 1\big)SR^2(Mkt). \quad (17)$$

The parameter η is thus given by

$$\eta = \frac{\big(\tau^2 - 1\big)SR^2(Mkt)}{\big(N + K - k\big)}. \quad (18)$$

Finally, equating the variance of α in the hypothetical sample (equation (14)) with the prior variance in equation (15) and using η from equation (18), we obtain

$$T_0 = \frac{\big(N + K - k\big)\big(1 + SR_{\max}^2 + m(1 + SR_{\max}^2)\big)}{\big(\tau^2 - 1\big)SR^2(Mkt)}. \quad (19)$$

By setting T_0, we conclude the derivations for the prior density. Thus, we also conclude the derivations for the posterior probability.

The resulting prior is sound. First, it is informed for the comprehensive parameter space. Moreover, model pricing ability can improve as more predictors are included. Hence, the prior is more strongly weighted against time-varying parameters because T_0 and m are positively related. Likewise, when the admissible squared Sharpe ratio increases, the prior is more strongly weighted against mispricing. Recall also from the marginal likelihood expressions that including more factors beyond the market or including more predictors leads to a higher penalty. Thus, the posterior probability is weighted against deviations from the unconditional CAPM.

Our prior specification for α resembles that of Pástor and Stambaugh (2000), $\alpha|\Sigma \sim \mathrm{N}(0, \sigma_\alpha^2(\frac{1}{s^2}\Sigma_{RR}))$, where σ_α^2 reflects the degree of beliefs in the pricing model and s^2 is the cross-sectional average of the test asset residual variances.

The prior on α is proportional to Σ_{RR} to avoid exploding Sharpe ratios. Note that in Pástor and Stambaugh (2000), the quantity $\alpha'\Sigma_{RR}^{-1}\alpha$ could grow with the addition of more test assets while we bound that expression. The following relation is useful to map σ_α^2, the prior confidence in model pricing ability, into the length of the hypothetical sample:

$$T_0 = \frac{s^2}{\sigma_\alpha^2}\left(1 + SR_{\max}^2 + m\left(1 + SR_{\max}^2\right)\right). \tag{20}$$

Internet Appendix V derives T_0 for other asset pricing model specifications, as described in Table I.

In the empirical experiments that follow, we exclusively use factors as test assets, as in Barillas and Shanken (2018) and Chib, Zeng, and Zhao (2020). That is, we consider the special case in which $N = 0$. At the same time, the developed setup is flexible enough to include incremental test assets. Thus, the number of test assets is $K - k$, the number of factors that are *not* included on the right-hand side of regression equation (3).

D. Additional Remarks on the Methodology

We make four additional remarks on the methodology. First, the literature acknowledges that asset pricing inferences could be sensitive to the set of test assets. As noted above, in the empirical analysis, when a model contains k factors, the remaining $14 - k$ "redundant" factors become the test assets. While our methodology allows us to include additional test assets, such as characteristic- and industry-sorted portfolios, we focus exclusively on factors as test assets to assess their *relative* performance. Our choice of test assets draws on Barillas and Shanken (2017), who suggest that the set of test assets is irrelevant for model comparison, that is, whether each model can price the factors in another model. Instead, only factor returns are required to conduct a relative test of model performance.

Second, we model stock return innovations as conditionally normal, while Arnott et al. (2019) show that the vast majority of factor returns are fat-tailed. However, the predictive distribution of stock returns in our setup departs substantially from normality. When integrating out the parameter space, the distribution of stock returns becomes Student-t. Further accounting for model uncertainty makes the predictive distribution even more fat-tailed due to mixing various t densities. In particular, one can draw returns from the predictive distribution in three steps: (i) draw a factor model by generating a uniform random variable to select a model based on cumulative model posterior probabilities, (ii) conditioned on the model, draw underlying parameters from the joint normal-inverted-Wishart densities, and (iii) conditioned on the parameters, draw returns from a normal distribution. While the predictive distribution can be simulated only by repeating these three steps, we have successfully derived analytic expressions for the vector of mean returns and the covariance matrix.[12] We acknowledge that departing from the assumption of normally distributed asset return innovations could possibly reduce the implicit penalty for models with skewed or fat-tailed factors. Hence, experimenting with suitable densities for asset returns could establish a promising avenue for future research.

Third, while we develop a prior in the context of cross-sectional asset pricing, for completeness, it would be useful to describe informed priors inspired by time-series econometrics. To start, Kandel and Stambaugh (1996) center the prior on the predictive regression R^2 around zero. Wachter and Warusawitharana (2009, 2015) further develop the Kandel–Stambaugh no-predictability prior. Innovative priors are also proposed by Pastor and Stambaugh (2009), who impose a negative correlation between the innovations in predictive regressions and expected returns to maintain mean reversion, Avramov, Cederburg, and Lučivjanská (2017), who propose taking cues from various consumption-based models for understanding the riskiness of equities over the long run, and Giannone, Lenza, and Primiceri (2015), who focus on coefficients in vector autoregressions.[13] The latter approach can motivate persistent factor risk premia and factor loadings modeled as latent variables. We leave this potentially interesting channel for future work.

Fourth, the prior specification requires the choice of τ, which is at the discretion of the econometrician. Following Barillas and Shanken (2018), the prior Sharpe ratio multiples take the values $\tau = 1.25, 1.5, 2$, and 3, while 1.5 is the baseline case. An alternative data-driven approach to choose τ, which is agnostic to economic restrictions, can be implemented by splitting the sample data into training, validation, and testing subsamples. The optimal τ is then chosen in the spirit of tuning machine-learning hyperparameters by considering various values for τ in the training sample, while the one recording optimal model ability in the validation is selected. Empirical experiments are then based on the test sample. We follow the first approach, which is more economically meaningful.

III. Data

We focus on 14 representative asset pricing factors that are prominent in the literature. We begin with the Fama-French five-factor model (Fama and French (2015)) that consists of the market (MKT), size (SMB), value (HML), profitability (RMW), and investment (CMA) factors, and augment it with momentum (MOM, Carhart (1997) and Fama and French (2018)). We also include two behavioral factors, namely, PEAD and financing (FIN) from Daniel, Hirshleifer, and Sun (2020). The additional factors include QMJ (Asness, Frazzini,

[12] Model-specific first two moments are derived in the Internet Appendix VIII. Moments for the integrated model follow through equations (4)–(6).

[13] Note that time-series goodness of fit, even in asset pricing regressions, does not translate into cross-sectional pricing ability, a point made by Chen, Roll, and Ross (1986). Thus, in our setup, it is the Sharpe ratio that plays a dominant role in formulating the prior, while the above-mentioned papers focus on either R^2 or autocorrelation.

and Pedersen (2019)), BAB (Frazzini and Pedersen (2014)), mispricing factors related to management (MGMT), and performance (PERF) from Stambaugh and Yuan (2017), liquidity (LIQ, Pástor and Stambaugh (2003)), and intermediary capital (ICR, He, Kelly, and Manela (2017)).[14]

We follow Welch and Goyal (2008) and employ 13 macro predictors, including dividend price ratio (*dp*), dividend yield (*dy*), earnings price ratio (*ep*), dividend payout ratio (*de*), stock variance (*svar*), book-to-market ratio (*bm*), net equity expansion (*ntis*), yield on Treasury bills (*tbl*), long-term yield (*lty*), long-term rate of returns (*ltr*), term spread (*tms*), default yield spread (*dfy*), and inflation (*infl*). Internet Appendix Table IA.I provides detailed definitions for each factor (Panel A) and macro predictor (Panel B).

The sample period ranges from June 1977 to December 2016 for a total of 475 monthly observations. Internet Appendix Table IA.II, Panel A, reports the mean, median, and standard deviation of monthly factor returns, as well as the monthly CAPM α and its corresponding t-statistics. All factors have positive average returns, ranging from 0.22% per month for SMB to 1.13% for ICR. While ICR has the highest average return, it also has the highest volatility, followed by MOM, while all other factors are less volatile than the market portfolio. All factors, except for SMB, display statistically significant and economically sizable CAPM α. BAB yields the highest CAPM α, followed by FIN and PERF.

In addition, the correlations between factor returns range between -0.55 (between MKT and QMJ) and 0.81 (between MKT and ICR). As expected, value- and investment-related factors such as HML, CMA, FIN, and MGMT are highly correlated. In addition, momentum- and profitability-related factors such as RMW and QMJ, MOM and PERF, and QMJ and PERF also exhibit high correlations.

Internet Appendix Table IA.II, Panel B, reports the mean, median, standard deviation, and AR(1) coefficient of the monthly macro predictors. Most predictive variables are highly persistent with AR(1) coefficients above 0.94, except *svar*, *ltr*, and *infl*. Nevertheless, all AR(1) coefficients are less than one.

IV. Probability Analysis

A. Predictive Regressions

To reinforce the case for time-varying parameters, we first apply the Bayesian approach to multivariate predictive regressions. Internet Appendix VII.A provides a detailed derivation of posterior probabilities in a predictive regression setting. Some combinations of macro predictors are jointly redundant. For instance, the dividend payout ratio (*de*) is the difference between the dividend-price ratio (*dp*) and the earnings-price ratio (*ep*). Therefore, among the 8 ($= 2^3$) possible inclusion/exclusion combinations, we restrict the

[14] We consider the tradable version of the LIQ and ICR factors to facilitate model interpretation and comparison.

model universe to five combinations: one without any predictor, three with only one predictor, and one with two predictors. Similarly, the term spread (*tms*) is the difference between the long-term yield (*lty*) and the yield on Treasury bills (*tbl*), and hence, we consider only five models. The remaining seven predictors contribute 2^7 combinations. The model space therefore consists of 3,200 ($= 25 \times 2^7$) combinations for (i) including macro predictors only and (ii) including macro predictors with possible interactions between predictors. A third scenario includes macro predictors with and without interactions. In that case, we consider 6,399 ($= 2 \times 25 \times 2^7 - 1$) combinations that are the union of the first two scenarios while excluding the overlapping intercept-only model.

A posterior probability is assigned to each candidate model. The Bayesian routine allows us to evaluate the relative importance of each individual predictor by its cumulative inclusion probability, given by $\mathcal{A}'\mathcal{P}$. For the first two scenarios, $\mathcal{A}$ is a $3,200 \times 13$ matrix representing all forecasting models by their unique combinations of zeros and ones, with zeros for the exclusion and ones for the inclusion of predictors, respectively, and $\mathcal{P}$ is a $3,200 \times 1$ vector including posterior probabilities for the models. For the third scenario, $\mathcal{A}$ is a $6,399 \times 13$ matrix and $\mathcal{P}$ is a $6,399 \times 1$ vector.

Internet Appendix Table IA.III presents the cumulative posterior probabilities for the macro items in predictive regressions. For the case of no interaction, the inclusion probability is approximately 100% for the dividend yield (*dy*), followed by stock variance (*svar*) at 95%, the earnings-price ratio (*ep*), the dividend-payout ratio (*de*), and the long-term rate of return (*ltr*) at 85%. Moving to the case with interactions, *dy*, *svar*, and the default yield spread (*dfy*) all have an inclusion probability close to 100%. The inclusion of two stock characteristics, namely, *dy* and *svar*, is strongly supported by the data regardless of the model specification. In contrast, the cumulative posterior probabilities for the book-to-market ratio (*bm*), the yield on Treasury bills (*tbl*), *ltr*, and the term spread (*tms*) drop significantly upon considering interactions. Moreover, the probability of including interactions is unity, suggesting that future returns depend on levels, squared values, and interactions between pairs of macro predictors. Overall, the stylized findings from predictive regressions motivate considering factor models with time-varying parameters.

B. Factor Models

We next apply our approach to conditional and unconditional asset pricing models, per Section II.B. The experiments are based on 14 asset pricing factors and 13 macro predictors. Panel A of Table I lists the families of candidate models considered in the paper. We restrict the model space by including the market as a factor (rather than a test asset) in all specifications except the single combination in which all factors are excluded (and only macro predictors serve as explanatory variables). Starting with unconditional models, the initial model space contains $2^{13} + 1$ ($= 2^{14-1} + 1$) combinations. We also discard the combination with all factors included and no factor as a test asset. Therefore, the final model space contains 2^{13} ($= 2^{13} + 1 - 1$) unconditional

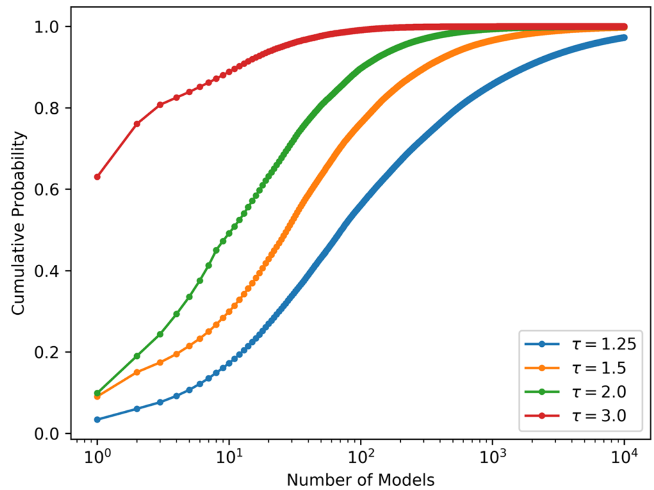


Figure 1. Cumulative posterior probabilities of asset pricing models. This figure plots the cumulative posterior probabilities for the universe of candidate models in a BMA framework for different values of τ. The candidate models $\mathbb{M}_1$ to $\mathbb{M}_4$ are specified in Table I. The cumulative posterior probabilities for models $\mathbb{M}_1$ to $\mathbb{M}_4$ are defined in equations (IA.72), (IA.64), (IA.41), and (IA.30) in the Internet Appendix, respectively. (Color figure can be viewed at wileyonlinelibrary.com)

combinations for both $\mathbb{M}_1$ and $\mathbb{M}_2$. For conditional models specified by $\mathbb{M}_3$ and $\mathbb{M}_4$, each includes $2^{13} \times (25 \times 2^7 - 1)$ combinations for inclusion/exclusion of the factors and predictors.[15] In sum, the integrated model accommodates a collection of over 52 million candidate models.[16]

We start by examining model probability. If a few models record sufficiently high posterior model probabilities, model uncertainty is not a primary concern, and model selection can deliver the right guidance about the factors and predictors that matter most. In contrast, if a large number of candidate models have meaningful probabilities, accounting for model uncertainty is essential and rationalized from Bayes' rule. We first compute the posterior probability for each candidate model. We then rank all models based on their probabilities from highest to lowest.

Figure 1 plots the cumulative posterior probabilities for the universe of candidate models under the various prior Sharpe multiples, namely, $\tau = 1.25$, 1.5, 2, and 3, where $\tau = 1.5$ is our baseline case. We find that the 10 (100, 500) top-ranked individual models account for a cumulative posterior probability of 30% (76%, 93%), suggesting that there is no clear winner across the entire

[15] As previously discussed, there are 2^{13} unconditional combinations. In addition, as noted in Section IV.A, some macro predictors are jointly redundant, resulting in 25×2^7 predictor combinations. We further exclude the single combination including no predictor, that is, the unconditional models specified in $\mathbb{M}_1$ and $\mathbb{M}_2$, resulting in $25 \times 2^7 - 1$ predictor combinations.

[16] The total number of competing models in $\mathbb{M}_1$ to $\mathbb{M}_4$ is computed as $2 \times 2^{13} + 2 \times 2^{13} \times (25 \times 2^7 - 1) = 52,428,800$.

space of potential factor models. Instead, a plethora of models that differ in the inclusion of factors, predictors, and mispricing record a positive and meaningful probability of governing the joint distribution of stock returns. Only when we adopt a prior Sharpe multiple of 3 do the best 10 models achieve a nontrivial cumulative posterior probability of 88%. From a practical investment management perspective, extremely high Sharpe ratios relative to the market are unlikely. The evidence therefore suggests that multiple distinct models could govern the joint dynamics of stock returns, which reinforces the role of model uncertainty.

Beyond probabilities for factor models, we next compute the cumulative posterior probabilities of individual factors and macro predictors. In particular, the posterior inclusion probability of a factor is given by

$$P\big(k \text{ included}|D\big) = \sum_{l=1}^{L} P(\mathcal{M}_l|D)\mathbb{1}_{\{k \text{ included in } \mathcal{M}_l\}}. \tag{21}$$

Similarly, the posterior inclusion probability of a predictor is given by

$$P\big(m \text{ included}|D\big) = \sum_{l=1}^{L} P(\mathcal{M}_l|D)\mathbb{1}_{\{m \text{ included in } \mathcal{M}_l\}}. \tag{22}$$

The results are reported in Table II. Panel A presents the cumulative posterior probabilities for the factors under different prior Sharpe ratio multiples. Several findings are worth noting. First, consider the baseline case of $\tau = 1.5$. We find that PEAD and QMJ display a posterior inclusion probability of close to 100%, followed by CMA, SMB, ICR, and MGMT—all of which achieve a posterior inclusion probability of at least 90%, indicating their prominence in pricing other factors beyond the market. For perspective, among the six factors proposed prior to 2015, only SMB displays high inclusion probability, while five of the eight factors proposed after 2015 exhibit high inclusion probability, suggesting that despite the expanding factor zoo, several new factors, both fundamental and behavioral, offer incremental ability to price the existing factors.

Second, PEAD, QMJ, and ICR stand out across different priors, with an inclusion probability of at least 93% in all cases. Moving to SMB, CMA, and MGMT, the inclusion probability is high for low τ values but diminishes for high τ values. In contrast, BAB exhibits high inclusion probability only for $\tau = 3$, when the prior is tilted toward rather extreme Sharpe ratios. It would thus be difficult for the BAB factor to clear prior asset pricing thresholds, such as reasonable Sharpe ratios (Ross (1976)). Taken together, the pricing abilities of widely explored factors depend on one's prior views about how large the Sharpe ratio could be.

Third, across all τ values, five to seven factors have a posterior inclusion probability of at least 90%, although the identified factors could vary. Our findings support a parsimonious model advocated by the empirical literature, while factors with high inclusion probability originate from distinct economic

Table II

Posterior Probabilities of Factors and Predictors in Asset Pricing Models

This table presents results based on the universe of candidate models using the BMA procedure for different values of τ. The candidate models $\mathbb{M}_1$ to $\mathbb{M}_4$ are specified in Table I. Panel A presents the cumulative posterior probabilities for the factors, as defined in equation (21). Panel B presents similar statistics for the macro predictors, as defined in equation (22). Panel C reports other model features, including: (i) the conditional model probability, defined as the sum of posterior probabilities for all models included in $\mathbb{M}_3$ and $\mathbb{M}_4$, (ii) the mispricing probability, defined as the sum of posterior probabilities for all models included in $\mathbb{M}_2$ and $\mathbb{M}_4$, and (iii) the equal-weighted average of hypothetical sample size T_0, as defined in equation (19), and (iv) the equal-weighted average of shrinkage, defined as $\frac{T_0}{T_0+T}$. Internet Appendix Table IA.I provides detailed definitions for each variable.

Panel A: Posterior Probabilities of Factors

	$\tau = 1.25$	$\tau = 1.5$	$\tau = 2$	$\tau = 3$
MKT	1.00	1.00	1.00	1.00
SMB	0.98	0.94	0.97	0.11
HML	0.30	0.17	0.03	0.00
RMW	0.01	0.00	0.00	0.00
CMA	0.97	0.97	0.35	0.00
MOM	0.65	0.39	0.01	0.00
PEAD	1.00	1.00	1.00	1.00
FIN	0.68	0.51	0.17	0.00
QMJ	1.00	1.00	1.00	1.00
BAB	0.19	0.15	0.50	0.95
MGMT	0.98	0.90	0.21	0.00
PERF	0.67	0.76	0.89	0.96
LIQ	0.89	0.57	0.77	0.97
ICR	0.97	0.93	0.94	0.97

Panel B: Posterior Probabilities of Macropredictors

	$\tau = 1.25$	$\tau = 1.5$	$\tau = 2$	$\tau = 3$
dp	0.35	0.22	0.06	0.01
dy	0.67	0.68	0.75	0.95
ep	0.35	0.40	0.75	0.84
de	0.28	0.30	0.46	0.29
svar	0.17	0.33	0.27	0.08
bm	0.02	0.00	0.04	0.06
ntis	0.14	0.15	0.21	0.96
tbl	0.10	0.03	0.06	0.90
lty	0.89	0.97	0.94	0.10
ltr	0.01	0.00	0.00	0.00
tms	0.02	0.00	0.04	0.90
dfy	0.03	0.01	0.00	0.00
infl	0.00	0.00	0.00	0.00

(*Continued*)

Table II—*Continued*

Panel C: Other Model Features				
	$\tau = 1.25$	$\tau = 1.5$	$\tau = 2$	$\tau = 3$
Conditional Model Probability	1.000	1.000	1.000	1.000
Mispricing Probability	0.641	0.686	0.579	0.057
Average T_0	4,693	2,112	880	330
Average Shrinkage $\frac{T_0}{T_0+T}$	0.897	0.799	0.630	0.398

foundations rather than an established asset pricing model. For instance, PEAD, QMJ, and ICR are proposed by three independent works, and their combination has not been examined in the previous literature.

Finally, RMW appears redundant. This could be due to the high correlation between RMW and QMJ (0.75 in Panel A of Internet Appendix Table IA.II), as profitability is also one of the quality characteristics in QMJ.[17] Empirically, QMJ dominates RMW in pricing other factors, and thus, we observe persistent inclusion for QMJ and exclusion for RMW.

Panel B of Table II implements a similar analysis for the macro predictors. Perhaps not surprisingly, in the presence of asset pricing factors, the average inclusion probability is considerably lower for macro predictors than for factors. Taking the baseline case of $\tau = 1.5$ as an example, the long-term yield (*lty*) has an inclusion probability of 97%, followed by the dividend yield (*dy*) with an inclusion probability of 68%. Moving to $\tau = 3$, more macro predictors display high inclusion probability, with net equity expansion (*ntis*), *dy*, the yield on Treasury bills (*tbl*), and the term spread (*tms*) having an inclusion probability of at least 90%. The rising inclusion probability with practically infeasible Sharpe ratios provides important evidence suggesting strong in-sample predictive power of macro items could be associated with only mild forecasting power out-of-sample. Last, the book-to-market ratio (*bm*), long-term rate of return (*ltr*), default yield spread (*dfy*), and inflation (*infl*) are always discarded, regardless of the prior. Evidence from Panel B therefore reinforces the view that asset pricing factors should be augmented with macro predictors to better capture cross-sectional and time-series effects in average returns.

In addition to the cumulative inclusion probabilities for asset pricing factors and macro predictors, we explore several other model features. The results are tabulated in Panel C of Table II. We start with the probability of factor models with time-varying parameters, defined as the sum of posterior probabilities for all models included in $\mathbb{M}_3$ and $\mathbb{M}_4$. The conditional models display an aggregate posterior probability of 100%, implying that our Bayesian approach uniformly favors models with time-varying parameters, even when prior beliefs are weighted against the inclusion of macro predictors.[18] Our find-

[17] The QMJ factor goes long high-quality stocks and short low-quality stocks, where high-quality stocks are those with high profitability, growth, and safety.

[18] Note that by construction, the sum of the posterior probabilities for all models included in $\mathbb{M}_1$ to $\mathbb{M}_4$ equals one. As previously noted, there are 2×2^{13} unconditional models in $\mathbb{M}_1$ and $\mathbb{M}_2$ and

ings further highlight the importance of incorporating nonlinearities in asset pricing models, especially by conditioning on the macroeconomic states—a point also emphasized by Chen, Pelger, and Zhu (2023) in a nonparametric setup. Furthermore, our results complement prior work that focuses on the nonlinear relationship between firm characteristics and returns (e.g., Freyberger, Neuhierl, and Weber (2020)) and that employs a conditional factor model in which the factor loadings are nonlinear in firm characteristics (e.g., Gu, Kelly, and Xiu (2021)).

Another essential feature in the BMA framework is the probability of model mispricing, defined as the sum of posterior probabilities for all models included in $\mathbb{M}_2$ and $\mathbb{M}_4$. For $\tau = 1.25$, 1.5, and 2, the mispricing probability varies between 58% and 69%. Even for sensible prior Sharpe ratios, the findings clearly highlight a prominent mispricing component in factor models. This evidence indicates that zero-alpha models selected from the set of factors and predictors that we analyze may not adequately explain cross-sectional and time-series effects in stock returns. Additionally, note that the probability of mispricing evolves only from conditional models, as the unconditional counterparts record near-zero probability. Overall, the evidence suggests that factor loadings, risk premia, and mispricing all vary with macroeconomic conditions.

We also report the (equal-weighted) average of (i) hypothetical sample size T_0, which is inversely related to τ as defined in equation (19), and (ii) the shrinkage intensity, defined as $\frac{T_0}{T^*} = \frac{T_0}{T_0+T}$. The amount of shrinkage increases when T_0 increases or, equivalently, when τ declines.[19] Intuitively, when the prior Sharpe ratio multiple is low, more shrinkage is applied to penalize mispricing and time-varying parameters. For $\tau = 1.5$ ($\tau = 3$), the average weight of the actual sample is approximately 20% (60%), and the remaining 80% (40%) is assigned to the hypothetical sample, where α_0, α_1, and β_1 are set to zero in equation (3).

In sum, a plethora of models that differ in the inclusion of factors, predictors, and mispricing record a positive and meaningful probability of governing the joint dynamics of stock returns. We show that the Bayesian approach provides a reasonable setup for analyzing the cross-section of expected stock returns in the presence of model uncertainty.

$2 \times 2^{13} \times (25 \times 2^7 - 1)$ conditional models in $\mathbb{M}_3$ and $\mathbb{M}_4$. Therefore, the cumulative probability for unconditional models is $\frac{1}{25\times 2^7} = 3.125\text{e-}4$, a priori. However, the cumulative posterior probability for unconditional models is 9.15e-50 for $\tau = 1.5$, suggesting that conditional models uniformly dominate their unconditional counterparts. In addition, we assess the presence of unconditional models among the top-ranked individual models. While we expect to see, a priori, 3,125 unconditional models among the top 10 million candidate models, we fail to detect any unconditional model based on posterior probabilities. The results are similar for alternative τ values.

[19] Specifically, the posterior regression means are a weighted average of estimates in the actual sample (with a weight of $\frac{T}{T_0+T}$) and the hypothetical sample (with a weight of $\frac{T_0}{T_0+T}$), as shown in equations (IA.23) and (IA.24) for unrestricted models (Internet Appendix I.A) and equation (IA.39) for restricted models (Internet Appendix I.B). Therefore, higher T_0 implies more shrinkage toward the hypothetical sample, that is, the model estimates are weighted against mispricing.

V. Out-of-Sample Performance

A. Efficient Portfolios: Sharpe Ratio

In this subsection, we assess the out-of-sample performance of the integrated model. Our analysis is based on mean-variance efficient portfolios that are derived form the predictive distribution that integrates out the within-model parameter space (estimation risk) and the model space (model disagreement). We study performance through Sharpe ratios and downside risk measures. For comparison, we consider four benchmark models that are widely used by academics and practitioners: (i) the CAPM, which adjusts only for MKT, (ii) the Fama-French three-factor model (FF3), which consists of MKT, SMB, and HML (Fama and French (1993)), (iii) the Fama-French six-factor model (FF6) that comprises MKT, SMB, HML, RMW, CMA, and MOM (Fama and French (2018)), and (iv) the AQR six-factor model (AQR6) that consists of the MKT, SMB, HML, MOM, BAB, and the QMJ factors (Frazzini, Kabiller, and Pedersen (2018)). We also consider the three highest posterior probability models based on the Bayesian approach. Our prior is that the Bayesian approach could deliver stable out-of-sample performance, given (i) its conceptual foundation from Bayes' rule and (ii) its empirical merits in identifying competent models.

Our first experiment examines the Sharpe ratio of the tangency portfolio. We divide the full sample into two periods, namely, the in-sample period and the out-of-sample performance period. Following Barillas and Shanken (2018), we consider two in-sample periods that correspond to half of the sample (denoted as $\frac{T}{2}$) and two-thirds of the sample (denoted as $\frac{2T}{3}$).[20] For the benchmark models, we use the in-sample period returns to derive the tangency portfolio weights and apply the optimal weights to the out-of-sample returns. We can then compute the out-of-sample Sharpe ratios. In the Bayesian setup, we use all data in the in-sample period to compute posterior probabilities and predictive moments based on the integrated model. A detailed description of the computation of the model-specific predictive moments is provided in the Internet Appendix VIII.[21]

We tabulate the in-sample and out-of-sample annualized Sharpe ratios in Table III, where Panel A corresponds to the four benchmark models and Panel B to the models based on the Bayesian approach with a prior Sharpe multiple of $\tau = 1.5$. The columns "EST" report the in-sample Sharpe ratio, and the columns "OOS" report the out-of-sample Sharpe ratio. For perspective, take $\frac{2T}{3}$ to be the in-sample period. The integrated model (denoted as BMA)

[20] The in-sample period that corresponds to $\frac{T}{2}$ ($\frac{2T}{3}$) ranges from June 1977 to December 1997 (December 2013), for a total of 247 (319) monthly observations.

[21] The derivation builds on Avramov (2004) and Avramov and Chordia (2006) with several important modifications to account for economically informed prior beliefs and model integration. The tangency portfolio for all models is constructed using the 14 benchmark assets that rotate between factors and test assets, as noted earlier, depending on the factor model. The predictive moments are computed based on equations (IA.115) and (IA.116) for factors and equations (IA.117), (IA.118), (IA.119), and (IA.120) for test assets. Moments for the integrated model follow from equations (4)–(6).

Table III

Out-of-Sample Sharpe Ratio

Panel A presents the in-sample and out-of-sample annualized Sharpe ratio for the tangency portfolio based on four benchmark models CAPM, FF3, FF6, and AQR6. The columns "EST" report the in-sample Sharpe ratio computed in the full sample (T), as well as in the in-sample periods that correspond to half of the sample ($\frac{T}{2}$) and two-thirds of the sample ($\frac{2T}{3}$). The columns "OOS" report the out-of-sample Sharpe ratio. We use the in-sample period returns to determine the tangency portfolio weights and apply the optimal weights to the out-of-sample returns. Panel B presents similar statistics for the three top-ranked individual models based on the Bayesian procedure (denoted Top 1, Top 2, and Top 3) and the integrated model (denoted BMA). The investment universe consists of 14 factors as listed in Panel A of the Internet Appendix Table IA.I, and we employ a prior Sharpe multiple of $\tau = 1.5$. In the Bayesian setup, we use all data in the in-sample period to compute posterior probabilities and predictive moments based on the integrated model. Panels C and D report similar statistics as Panels A and B, where we further impose the Regulation-T constraint. In particular, the sum of the absolute tangency portfolio weights is set to be less than or equal to two, $\sum_{i=1}^{14} |w_i| \leq 2$. Panels E and F report similar statistics as Panels A and B, where we replace the tangency portfolio with the global minimum variance portfolio.

	T	$\frac{T}{2}$		$\frac{2T}{3}$	
Model	EST	EST	OOS	EST	OOS
Panel A: Tangency Portfolio Based on Benchmark Models					
CAPM	0.489	0.601	0.375	0.468	0.540
FF3	0.729	1.111	0.468	0.960	0.431
FF6	1.317	2.180	0.676	1.518	0.798
AQR6	1.679	2.803	0.954	1.829	1.152
Panel B: Tangency Portfolio Based on Bayesian Models					
Top 1	2.249	3.305	1.009	2.616	1.226
Top 2	2.233	3.280	1.027	2.699	1.425
Top 3	2.100	3.337	1.019	2.567	1.163
BMA	2.212	3.228	0.968	2.542	1.240
Panel C: Tangency Portfolio Based on Benchmark Models with Regulation-T					
CAPM	0.489	0.601	0.375	0.468	0.540
FF3	0.706	1.057	0.456	0.872	0.465
FF6	1.017	1.272	0.430	1.094	0.367
AQR6	1.168	1.699	0.491	1.240	0.785
Panel D: Tangency Portfolio Based on Bayesian Models with Regulation-T					
Top 1	1.673	2.150	0.872	1.884	1.425
Top 2	1.688	2.161	0.884	1.840	1.332
Top 3	1.251	2.223	0.860	1.581	1.013
BMA	1.621	2.137	0.617	1.772	0.979
Panel E: Global Minimum Variance Portfolio Based on Benchmark Models					
FF3	0.662	0.994	0.483	0.923	0.401
FF6	1.254	2.038	0.672	1.423	0.818
AQR6	1.507	2.358	0.988	1.593	0.700

(*Continued*)

Table III—*Continued*

Panel F: Global Minimum Variance Portfolio Based on Bayesian Models					
Top 1	1.897	2.850	0.998	2.502	0.924
Top 2	1.897	2.849	0.994	2.391	0.896
Top 3	1.817	2.858	1.002	2.382	0.870
BMA	1.925	2.923	1.040	2.433	1.101

outperforms the benchmark models both in-sample and out-of-sample. For instance, the integrated model generates an in-sample annualized Sharpe ratio of 2.542, while the best benchmark model AQR6 delivers an annualized Sharpe ratio of 1.829. The integrated model also continues to deliver superior out-of-sample performance, with an annualized Sharpe ratio of 1.240, which offers an improvement of 8% compared to the best benchmark model AQR6, which has an annualized Sharpe ratio of 1.152.[22]

In addition, the three top-ranked individual models (denoted as Top 1, Top 2, and Top 3), the three models with the highest posterior probability, also deliver sound out-of-sample performance. The annualized Sharpe ratio ranges from 1.163 to 1.425 out-of-sample, indicating a 1% to 24% improvement from the best benchmark model AQR6. Note that the in-sample posterior probabilities of the top-ranked models are indistinguishable, suggesting that they are virtually equally likely to govern the joint distribution of stock returns. However, we observe more variation in their out-of-sample performance. For instance, the second-ranked (i.e., Top 2) model turns out to be the best performing and significantly outperforms AQR6, while the third-ranked model only edges out AQR6. Importantly, the integrated model does not rely on the crucial assumption that a single or a few top-ranked models must be correct while all other specifications should be discarded. For perspective, the integrated model outperforms two of the three top-ranked individual models.

Panels C and D have the same layout as Panels A and B, but we further impose the Regulation-T constraint. In particular, to ensure that the tangency portfolio does not rely on extreme, possibly infeasible long and short positions in real time, we require that the sum of absolute tangency portfolio weights less than or equal to two, that is, $\sum_{i=1}^{14} |w_i| \leq 2$. As expected, the Regulation-T constraint reduces the Sharpe ratio for nearly all models, both in-sample and out-of-sample.

[22] While we focus on four observable factor models as benchmarks, we further consider the tangency portfolio based on the unconditional model with 14 factors. Taking $\frac{2T}{3}$ to be the in-sample period, the integrated model continues to outperform the unconditional model (annualized Sharpe ratio of 1.019) by 22%. When we examine the equal-weighted portfolio with 14 factors in the full sample, we find an annualized Sharpe ratio of 1.705 and a monthly FF6-adjusted (AQR6-adjusted) return of 0.215% (0.121%). For perspective, the integrated model delivers an annualized Sharpe ratio of 2.212 and a monthly FF6-adjusted (AQR6-adjusted) return of 0.324% (0.220%), indicating a 30% to 82% improvement across different performance metrics. Our results highlight that the strong out-of-sample performance of the integrated model goes beyond the positive alphas of the asset pricing factors in the full sample.

Taking $\frac{2T}{3}$ as the in-sample period, the integrated model produces an out-of-sample annualized Sharpe ratio of 0.979 and outperforms all benchmark models by a significant margin. For instance, the annualized Sharpe ratio is 0.785 for the best benchmark model AQR6, indicating that the integrated model outperforms by 25% after applying sensible economic restrictions. Thus, when efficient portfolios are admissible, the performance gap between the integrated model and benchmark models widens. Furthermore, we continue to find superior performance among top-ranked individual models, which outperform AQR6 by 29% to 82%. Taken together, the results suggest that the Bayesian approach is able to detect outperforming models in the presence of economic restrictions.

With a shorter in-sample period ($\frac{T}{2}$), we observe much lower out-of-sample Sharpe ratios as well as larger gaps between in-sample and out-of-sample performance across nearly all model specifications with and without economic restrictions. This is possibly due to overfitting attributable to the short in-sample period (247 months).[23] Importantly, all models based on the Bayesian approach deliver higher Sharpe ratios than the best benchmark model AQR6 with and without economic restrictions, especially in the former case. For instance, the integrated model outperforms AQR6 by 26%, and the top-ranked individual models outperform AQR6 by 75% to 80% after imposing the Regulation-T constraint. Overall, while we focus on the $\frac{2T}{3}$ case to interpret the findings, the Bayesian approach continues to deliver superior and more admissible out-of-sample performance than all benchmark models for the shorter in-sample period.

Our next experiment focuses on the GMVP, which relies exclusively on the covariance matrix formulated in equation (5). If model uncertainty has meaningful asset pricing implications, the GMVP based on the integrated model should provide investment payoffs characterized by lower risk measures compared to benchmarks.[24] We report the results in Panels E and F of Table III, with Panel E corresponding to the benchmark models and Panel F to models based on the Bayesian approach.

While in the following subsection, we analyze risk measures associated with the minimum variance portfolio, and we briefly describe the out-of-sample Sharpe ratios. Taking $\frac{2T}{3}$ to be the in-sample period, the integrated model generates an annualized out-of-sample Sharpe ratio of 1.101 and outperforms all competing models by a considerable margin. For instance, the integrated

[23] There are 14 factors and 11 predictors in total because some macro predictors are jointly redundant, as previously discussed. The total number of estimated parameters is given by $(K-k)[(1+m)(1+k)+\frac{K-k+1}{2}]+k(1+m+\frac{k+1}{2})$, where K denotes the maximal number of factors (14), and k and m denote the number of included factors and predictors, respectively. The number of estimated parameters varies between 119 (when $m=0$) and 812 (when $m=11$ and $k=7$).

[24] Garlappi, Uppal, and Wang (2007) document that in the presence of a stable and significant degree of ambiguity aversion, the GMVP could play an important role in optimal portfolio choice because it is not subject to ambiguity about expected returns. While this is not the focus of our work, our findings extend to ambiguity-averse investors.

model delivers a 35% higher Sharpe ratio than the best benchmark model (FF6, annualized Sharpe ratio of 0.818) and 19% higher Sharpe ratio than the best individual model (Top 1, annualized Sharpe ratio of 0.924). Taken together, our findings highlight a sizable impact of model uncertainty on the covariance matrix of stock returns, a novel feature in our BMA setup.

B. Efficient Portfolios: Downside Risk

It is worth evaluating risk and downside risk measures associated with trading the tangency portfolio and the GMVP.[25] Using $\frac{2T}{3}$ as the in-sample period, we report the out-of-sample mean, standard deviation, skewness, and excess kurtosis of the monthly excess returns and the maximum drawdown. Following Gu, Kelly, and Xiu (2020), we define the maximum drawdown across the entire sample as

$$MDD = \max_{0 \le t_1 \le t_2 \le T} \left(Y_{t_1} - Y_{t_2} \right), \tag{23}$$

where Y_{t_1} and Y_{t_2} refer to the cumulative log return from month 0 to t_1 and t_2, respectively.

We tabulate the results in Table IV, where Panels A and B report the results for tangency portfolios constructed from benchmark models and models based on the Bayesian approach with $\tau = 1.5$, respectively, and Panels C and D report similar statistics after imposing the Regulation-T constraint. Given their economic relevance, we focus on Panels C and D to interpret our findings. First, when compared to benchmark models, the higher Sharpe ratio of the integrated model can be attributed to a combination of higher returns and lower return volatility. Second, the integrated model exhibits less negative skewness, lower excess kurtosis, and a lower maximum drawdown. For instance, the maximum drawdown across the entire sample is 44% for the integrated model, while the benchmark models experience a larger drawdown of 51% to 79%. Notably, the top-ranked individual models are more volatile, and the maximum drawdown varies over a wide range between 32% and 57%.

Panels E and F of Table IV report similar statistics for GMVP, with Panel E reporting results for GMVP constructed from benchmark models and Panel F for GMVP constructed from models based on the Bayesian approach with $\tau = 1.5$. Notably, risk metrics are particularly relevant in the context of GMVP, as GMVP relies exclusively on the covariance matrix. Several findings are noteworthy. First, the GMVP based on the integrated model displays less risky payoffs than the benchmark models. For instance, monthly realized volatility for GMVPs based on the benchmark models ranges between 0.956% and 2.127%, while it appears to be only 0.756% for the integrated model, indicating a 21% to 64% volatility reduction. Because expected returns across the various

[25] Related work shows that individual anomaly payoffs are prone to large drawdowns. For instance, Daniel and Moskowitz (2016) document that momentum strategies are characterized by occasional large crashes.

Table IV

Out-of-Sample Downside Risk

Panel A reports the out-of-sample mean, standard deviation, skewness, and excess kurtosis of the monthly excess returns, the annualized Sharpe ratio, and the maximum drawdown for the tangency portfolio based on four benchmark models CAPM, FF3, FF6, and AQR6. We employ the in-sample period that corresponds to two-thirds ($\frac{2T}{3}$) of the sample. We use the in-sample period returns to determine the tangency portfolio weights and apply the optimal weights to the out-of-sample returns. Panel B presents similar statistics for the three top-ranked individual models based on the Bayesian procedure (denoted as Top 1, Top 2, and Top 3) and the integrated model (denoted as BMA). The investment universe consists of 14 factors as listed in Panel A of the Internet Appendix Table IA.I, and we employ a prior Sharpe multiple of $\tau = 1.5$. In the Bayesian setup, we use all data in the in-sample period to compute posterior probabilities and predictive moments based on the integrated model. Panels C and D report similar statistics as Panels A and B, where we further impose the Regulation-T constraint. That is, the sum of the absolute tangency portfolio weights is set to be less than or equal to two, $\sum_{i=1}^{14} |w_i| \leq 2$. Panels E and F report similar statistics as Panels A and B, where we replace the tangency portfolio with the global minimum variance portfolio.

Model	Mean	Std.Dev.	Sharpe Ratio	Skewness	Excess Kurtosis	Maximum Drawdown
Panel A: Tangency Portfolio Based on Benchmark Models						
CAPM	0.640	4.110	0.540	−0.699	2.103	51.511
FF3	0.282	2.266	0.431	−0.622	2.797	31.065
FF6	0.223	0.968	0.798	−0.258	0.721	8.778
AQR6	0.305	0.917	1.152	−0.441	2.816	6.577
Panel B: Tangency Portfolio Based on Bayesian Models						
Top 1	0.274	0.774	1.226	0.348	2.339	3.860
Top 2	0.338	0.821	1.425	0.396	1.251	2.487
Top 3	0.270	0.804	1.163	0.347	2.129	4.008
BMA	0.277	0.775	1.240	0.001	1.344	5.149
Panel C: Tangency Portfolio Based on Benchmark Models with Regulation-T						
CAPM	1.281	8.219	0.540	−0.699	2.103	78.807
FF3	0.708	5.272	0.465	−0.835	3.014	63.947
FF6	0.414	3.911	0.367	−2.017	10.046	50.631
AQR6	0.989	4.361	0.785	−1.213	4.800	58.583
Panel D: Tangency Portfolio Based on Bayesian Models with Regulation-T						
Top 1	1.543	3.749	1.425	0.022	2.121	32.254
Top 2	1.482	3.855	1.332	0.395	2.271	37.544
Top 3	1.208	4.134	1.013	−0.589	3.536	57.364
BMA	1.035	3.664	0.979	−0.518	0.928	43.743
Panel E: Global Minimum Variance Portfolio Based on Benchmark Models						
FF3	0.246	2.127	0.401	−0.442	2.537	26.997
FF6	0.226	0.956	0.818	0.013	0.813	5.771
AQR6	0.244	1.207	0.700	−0.376	5.156	7.491

(*Continued*)

Table IV—*Continued*

Panel F: Global Minimum Variance Portfolio Based on Bayesian Models						
Top 1	0.232	0.871	0.924	0.257	4.180	4.392
Top 2	0.221	0.855	0.896	0.337	3.453	5.823
Top 3	0.205	0.818	0.870	0.289	5.217	4.977
BMA	0.240	0.756	1.101	0.155	3.607	4.988

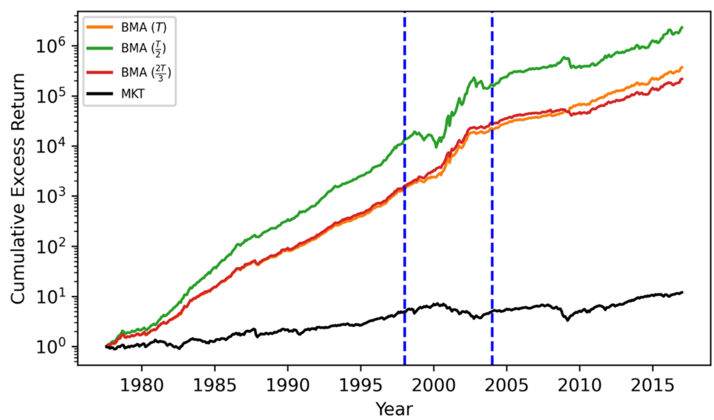


Figure 2. BMA model performance. This figure plots the cumulative excess returns on an initial \$1 investment for the market portfolio (MKT) and three tangency portfolios based on the integrated model. We employ a prior Sharpe multiple of $\tau = 1.5$ and consider three in-sample periods that correspond to the full sample (T), half of the sample ($\frac{T}{2}$), and two-thirds of the sample ($\frac{2T}{3}$). The tangency portfolios are levered to the extent that their in-sample return volatility is the same as that of the market portfolio. The leverage ratios are carried forward to the out-of-sample periods. The blue dashed lines mark the end of the in-sample periods for $\frac{T}{2}$ and $\frac{2T}{3}$. (Color figure can be viewed at wileyonlinelibrary.com)

specifications are not materially different, the lower volatility characterizing the Bayesian approach translates into a substantially higher out-of-sample Sharpe ratio. Second, while most benchmark models are negatively skewed, the integrated model displays positive skewness. Third, the integrated model exhibits a lower maximum drawdown than all benchmark models. For perspective, the maximum drawdown for the benchmark models ranges between 6% and 27%, compared to 5% for the integrated model. The BMA approach therefore mitigates the downside risk of both the tangency portfolio and the GMVP.

To better understand integrated model performance over time, Figure 2 plots the cumulative excess returns of an initial \$1 investment for the market portfolio (MKT) and three BMA tangency portfolios with $\tau = 1.5$. The BMA tangency portfolios vary in the in-sample period as previously discussed. We derive a leverage ratio for each tangency portfolio by equalizing its in-sample return volatility to the market return volatility during the same period. The leverage ratio is then carried forward to the out-of-sample period.[26] This allows us to compare the performance of BMA portfolios with the market directly. We find that BMA portfolios significantly outperform the market and experience a more stable advance over time. Importantly, we observe only modest declines for BMA portfolios when the overall market drops significantly, consistent with the high Sharpe ratio and low downside risk out-of-sample. Our findings are similar across all three in-sample periods.

C. Portfolio Tilts

In addition to the integrated model's performance as measured by Sharpe ratio and downside risk, we explore its portfolio tilts. Specifically, we examine the portfolio weights for each of the 14 factors in the tangency portfolio based on the integrated model and, more importantly, whether and how the portfolio choice varies with subsequently realized factor returns over time. Intuitively, if the tangency portfolio bets on the correct factors, we expect it to tilt toward subsequently outperforming factors and away from underperforming factors. Such factor rotation could be particularly valuable on the downside, underweighting factors when they subsequently record negative returns.

Table V tabulates the results. We employ the in-sample period that corresponds to two-thirds of the sample ($\frac{2T}{3}$) and a prior Sharpe multiple of 1.5. Denote by $w_{k,t}$ the weight of factor k in the tangency portfolio at portfolio formation time t, $f_{k,t+1}$ the realized return of factor k during the holding period $t+1$, $\text{Corr}(w_{k,t}, f_{k,t+1})$ the correlation between the tangency portfolio weights and subsequently realized factor returns. The columns "EST" and "OOS" report the in-sample and out-of-sample results, and the column "OOS$^-$" focuses on the subperiod with negative realized factor returns out-of-sample.[27]

Several findings are worth noting. First, the tangency portfolio significantly overweights PEAD, QMJ, and MGMT factors both in-sample and out-of-sample compared with an equal-weighted benchmark portfolio. Second, we observe positive correlations between tangency portfolio weights and realized factor returns for 9 out of 14 factors out-of-sample, with an average correlation of 4.5%. From the perspective of economic significance, this is a considerable improvement relative to the equal-weighted portfolio, which indicates a zero correlation between portfolio weights and realized factor returns. Remarkably, when focusing on the downside, the correlation between portfolio weights and subsequently realized factor returns increases to 10.5% during subperiods of negative realized factor returns. This suggests that the Bayesian setup is instrumental in tempering adverse investment outcomes through factor rotation.

[26] The leverage ratio is 4.87, 6.37, and 5.00 for the in-sample periods that correspond to the full sample (T), half of the sample ($\frac{T}{2}$), and two-thirds of the sample ($\frac{2T}{3}$), respectively.

[27] The subperiod is factor-specific, depending on the realized returns for each factor during the out-of-sample period.

Table V

BMA Tangency Portfolio Choice

This table presents the in-sample and out-of-sample tangency portfolio weights (denoted as $w_{k,t}$) and realized factor returns (denoted as $f_{k,t+1}$) for the 14 factors listed in Panel A of the Internet Appendix Table IA.I. $w_{k,t}$ is the weight for factor k in the tangency portfolio at portfolio formation time t, and $f_{k,t+1}$ is the holding-period return of factor k at time $t+1$. The tangency portfolio is based on the integrated model with a prior Sharpe multiple of $\tau = 1.5$. The columns "EST" report the in-sample results based on two-thirds of the sample ($\frac{2T}{3}$). The columns "OOS" report the out-of-sample results. We use all data in the in-sample period to compute posterior probabilities and predictive moments based on the integrated model. In addition, we report the correlation between the tangency portfolio weights and realized factor returns ($\text{Corr}(w_{k,t}, f_{k,t+1})$) during the in-sample and out-of-sample periods, as well as the subperiods with negative realized factor returns out-of-sample (column "OOS$^-$").

	$w_{k,t}$		$f_{k,t+1}$		$\text{Corr}(w_{k,t}, f_{k,t+1})$		
	EST	OOS	EST	OOS	EST	OOS	OOS$^-$
MKT	0.102	0.110	0.006	0.006	0.220	0.006	−0.337
SMB	0.102	0.109	0.002	0.002	0.311	−0.021	−0.115
HML	0.006	−0.002	0.004	0.001	0.166	0.035	0.156
RMW	0.039	0.034	0.004	0.003	0.052	0.027	0.194
CMA	0.049	0.051	0.004	0.000	0.128	0.004	−0.003
MOM	0.042	0.039	0.009	0.000	0.132	0.318	0.607
PEAD	0.223	0.215	0.008	0.002	0.097	−0.206	−0.304
FIN	0.013	0.009	0.009	0.004	0.054	-0.161	0.014
QMJ	0.181	0.184	0.005	0.004	0.004	0.280	0.536
BAB	0.021	0.022	0.011	0.006	0.125	0.053	−0.089
MGMT	0.148	0.152	0.008	0.001	0.096	−0.027	−0.309
PERF	−0.005	−0.002	0.008	0.005	0.190	0.182	0.576
LIQ	0.047	0.045	0.005	0.003	0.163	0.135	0.069
ICR	0.032	0.032	0.014	0.006	0.087	−0.001	0.469

D. Robustness Analyses

Thus far, we have assessed the out-of-sample performance based on the baseline prior Sharpe multiple of 1.5. As a robustness check, we examine the sensitivity of our findings to alternative prior Sharpe multiples. Table VI has a similar layout as Table III, with Panel A corresponding to the unconstrained tangency portfolio, Panel B to the tangency portfolio with the Regulation-T constraint, and Panel C to the GMVP. Taking $\frac{2T}{3}$ as the in-sample period by way of example, the integrated model continues to outperform the best benchmark model from Table III across all τ values. As shown in Panel A (Panel B), the out-of-sample annualized Sharpe ratio of the integrated model is 1.208 (0.976), 1.271 (1.018), and 1.253 (0.819) when $\tau = 1.25$, 2, and 3, while the best benchmark model AQR6 delivers an out-of-sample annualized Sharpe ratio of 1.152 (0.785) before (after) applying economic restrictions. The integrated model outperforms AQR6 by 5% to 10% without economic restrictions and outperforms AQR6 by 4% to 30% with economic restrictions. Moving to the GMVP in Panel C, the integrated model delivers a higher Sharpe ratio than the best benchmark model FF6 across all τ values, and the improvement

Table VI

Out-of-Sample Sharpe Ratio: Alternative Prior Sharpe Multiple

Panel A presents the in-sample and out-of-sample annualized Sharpe ratio of the three top-ranked individual models based on the Bayesian procedure (denoted as Top 1, Top 2, and Top 3) and the integrated model (denoted as BMA). The investment universe consists of the 14 factors listed in Panel A of the Internet Appendix Table IA.I, and we employ alternative prior Sharpe multiples of $\tau = 1.25$, 2, and 3. The columns "EST" report the in-sample Sharpe ratio computed in the full sample (T), as well as in the in-sample periods that correspond to half ($\frac{T}{2}$) and two-thirds ($\frac{2T}{3}$) of the sample. The columns "OOS" report the out-of-sample Sharpe ratio. We use all data in the in-sample period to compute posterior probabilities and predictive moments based on the integrated model. Panel B reports similar statistics with the Regulation-T constraint. That is, the sum of the absolute tangency portfolio weights is set to be smaller than or equal to two, $\sum_{i=1}^{14} |w_i| \leq 2$. Panel C reports similar statistics, where we replace the tangency portfolio with the global minimum variance portfolio.

		T	$\frac{T}{2}$		$\frac{2T}{3}$	
τ	Model	EST	EST	OOS	EST	OOS
		Panel A: Tangency Portfolio				
$\tau = 1.25$	Top 1	2.307	3.201	0.975	2.631	1.293
	Top 2	2.159	3.247	1.013	2.611	1.277
	Top 3	2.283	3.188	0.982	2.569	1.177
	BMA	2.184	3.175	0.985	2.527	1.208
$\tau = 2$	Top 1	2.124	3.370	0.961	2.771	1.550
	Top 2	1.888	3.333	1.034	2.608	0.791
	Top 3	1.929	3.339	0.971	2.628	1.311
	BMA	2.163	3.338	0.946	2.613	1.271
$\tau = 3$	Top 1	0.583	3.619	0.980	2.719	1.046
	Top 2	0.806	3.687	0.952	2.729	1.291
	Top 3	0.784	3.709	0.908	2.649	1.203
	BMA	0.744	3.634	0.982	2.737	1.253
		Panel B: Tangency Portfolio with Regulation T				
$\tau = 1.25$	Top 1	1.761	2.152	0.681	1.699	1.195
	Top 2	1.395	2.252	0.711	1.650	1.103
	Top 3	1.702	2.148	0.711	1.727	0.942
	BMA	1.626	2.029	0.621	1.700	0.976
$\tau = 2$	Top 1	1.628	2.355	0.733	1.788	1.628
	Top 2	1.344	2.198	0.700	1.679	0.526
	Top 3	1.309	2.309	0.912	1.720	1.452
	BMA	1.569	2.392	0.652	1.841	1.018
$\tau = 3$	Top 1	1.410	2.626	0.404	1.796	0.695
	Top 2	1.407	2.604	0.455	1.839	0.933
	Top 3	1.679	2.827	0.482	1.688	0.961
	BMA	1.481	2.802	0.586	1.878	0.819
		Panel C: Global Minimum Variance Portfolio				
$\tau = 1.25$	Top 1	1.865	2.605	1.068	2.396	0.909
	Top 2	1.918	2.617	1.109	2.394	0.907
	Top 3	1.911	2.606	1.066	2.153	0.961
	BMA	1.927	2.902	1.062	2.424	1.121

(*Continued*)

Table VI—*Continued*

Panel C: Global Minimum Variance Portfolio						
$\tau = 2$	Top 1	1.730	2.832	0.937	2.510	0.903
	Top 2	1.737	2.837	0.973	2.392	0.876
	Top 3	1.746	2.827	0.930	2.392	0.869
	BMA	1.908	2.922	1.026	2.499	1.015
$\tau = 3$	Top 1	1.734	2.949	0.963	2.359	0.901
	Top 2	1.734	2.951	0.970	2.388	0.940
	Top 3	1.720	2.913	0.976	2.353	0.963
	BMA	1.763	2.981	1.083	2.376	0.961

in Sharpe ratio ranges from 17% to 37%. The decisive evidence on GMVPs confirms the meaningful impact of model uncertainty in asset pricing.

In addition, while the top-ranked individual models display similar in-sample posterior probabilities and deliver promising performance, we observe considerable variation in their performance and relative rankings. As shown in Panel B, when $\tau = 2$, the first-ranked (second-ranked) model generates an annualized Sharpe ratio of 1.628 (0.526) after applying economic restrictions, and it significantly outperforms (underperforms) the best benchmark model AQR6 with an annualized Sharpe ratio of 0.785 and the integrated model with an annualized Sharpe ratio of 1.018. Moving to the GMVP in Panel C, the integrated model outperforms all top-ranked individual models for $\tau = 1.25$ and 2 and outperforms two of the three top-ranked individual models for $\tau = 3$. Furthermore, the third-ranked model has the highest out-of-sample Sharpe ratio for $\tau = 1.25$ and 3, while the first-ranked model yields the highest out-of-sample Sharpe ratio for $\tau = 2$.

Overall, accounting for model uncertainty yields a rather stable, superior, and admissible out-of-sample Sharpe ratio and mitigates the downside risk of the investment. The proposed investment strategy based on the integrated model benefits from factor rotation, especially by tilting away from the subsequently underperforming factors. Our findings are robust to imposing economic restrictions on the prior Sharpe ratio and stock positions as well as using alternative in-sample periods. Analyses of the GMVP further highlight the impact of model uncertainty on the covariance matrix of stock returns. In addition, the Bayesian approach is instrumental in identifying competent models, while we should remain cautious that model selection based on a single or a few top-ranked models could provide an unstable description of asset return dynamics.

VI. Dissecting Model Uncertainty

In this section, we conduct four experiments to highlight the role of model uncertainty in shaping the investment opportunity set. We first implement a variance decomposition to shed light on how model uncertainty affects the perceived risk of equities. We then analyze the relative contribution of model disagreement to the covariance matrix, as measured by the increase in entropy. Third, we investigate how candidate models disagree on mispricing, factor loadings, and risk premia over time. Finally, we examine the dispersion in portfolio choice and performance to further assess the economic consequence of model disagreement.

A. Variance Decomposition

The first experiment compares the sample variance of factor returns with the variance based on the integrated model. In particular, from variance decomposition, we obtain

$$\mathrm{Var}(r_{t+1}) = E[\mathrm{Var}(r_{t+1}|z_t)] + \mathrm{Var}[E(r_{t+1}|z_t)], \tag{24}$$

where $\mathrm{Var}(r_{t+1})$ is the unconditional variance and $E[\mathrm{Var}(r_{t+1}|z_t)]$ is the (time-series) average of conditional variance. The variance decomposition is conditioned on a particular factor model and the parameter space underlying that model. For notational convenience, we drop such dependencies.

Resorting to sample estimates, the variance of each factor should be higher than the mean of the conditional variance. This is because, in population, the inequality $\mathrm{Var}(r_{t+1}) > E[\mathrm{Var}(r_{t+1}|z_t)]$ is binding. However, $\mathrm{Var}(r_{t+1}|z_t)$ does not incorporate model disagreement and the mixture of estimation risk emphasized by our approach. Thus, the variance perceived by a Bayesian investor that accounts for model uncertainty is higher than $\mathrm{Var}(r_{t+1}|z_t)$.

Taken together, the difference between the sample analog of $\mathrm{Var}(r_{t+1})$ and the sample average of $\mathrm{Var}[r_{t+1}|D]$ depends on the net effect of the two conflicting forces and remains an empirical question. If model uncertainty plays a significant role in asset pricing, we expect the sample average of the variance based on the integrated model to exceed the sample (unconditional) variance.

To proceed, we compute (i) the sample average of the variance based on the integrated model, defined as the time-series average of the diagonal elements of the covariance matrix, that is, $\mathrm{Var}[r_{t+1}|D]$ in equation (5), and (ii) the sample variance computed from realized factors returns. We consider three in-sample periods that correspond to the full sample (T) half the sample ($\frac{T}{2}$) and two-thirds of the sample ($\frac{2T}{3}$) under a prior Sharpe multiple of 1.5, and compute the in-sample and out-of-sample variance of each factor.

Table VII presents the results, where the columns “EST” and “OOS” correspond to the in-sample and out-of-sample results, respectively. In the full sample, 8 out of 14 factors display higher variance based on the integrated model (denoted as $\bar{V}_t + \bar{\Omega}_t$) than the sample variance (denoted as OBS). Likewise, using $\frac{2T}{3}$ of the sample as an in-sample period, 8 out of 14 factors display higher variance based on the integrated model than the sample variance during the out-of-sample period. Notably, the gap between the integrated model variance and sample variance widens considerably out-of-sample, and the integrated model variance is on average 53% higher than the sample variance across all

Table VII

BMA Model Variance and Sample Variance

This table presents the in-sample and out-of-sample variance of each factor. We report (i) the sample average of the variance based on the integrated model (denoted as $\bar{V}_t + \bar{\Omega}_t$), defined as the time-series average of the diagonal elements of the covariance matrix, $\text{Var}[r_{t+1}|D]$ in equation (5), and (ii) the sample variance computed from realized factors returns (denoted as OBS). The columns "EST" report the in-sample variance computed in the full sample (T), as well as in the in-sample periods that correspond to half of the sample ($\frac{T}{2}$) and two-thirds of the sample ($\frac{2T}{3}$). The columns "OOS" report the out-of-sample variance. The investment universe consists of the 14 factors listed in Panel A of the Internet Appendix Table IA.I, and we employ a prior Sharpe multiple of $\tau = 1.5$.

	T		$\frac{T}{2}$				$\frac{2T}{3}$			
	EST		EST		OOS		EST		OOS	
	$\bar{V}_t + \bar{\Omega}_t$	OBS	$\bar{V}_t + \bar{\Omega}_t$	OBS	$\bar{V}_t + \bar{\Omega}_t$	OBS	$\bar{V}_t + \bar{\Omega}_t$	OBS	$\bar{V}_t + \bar{\Omega}_t$	OBS
MKT	19.875	19.771	18.912	18.797	19.188	20.873	21.325	21.242	21.444	16.889
SMB	8.576	8.507	6.779	6.743	6.885	10.438	9.968	9.928	10.018	5.653
HML	8.067	8.417	6.432	6.376	6.506	10.643	9.182	9.362	8.880	6.491
RMW	5.070	5.553	2.086	2.004	2.057	9.402	6.520	6.963	6.793	2.695
CMA	3.923	3.901	2.999	2.941	3.124	4.954	4.814	4.843	4.799	1.904
MOM	19.800	19.955	10.380	10.117	10.157	30.524	19.253	19.325	20.080	20.871
PEAD	3.602	3.587	2.114	2.093	2.148	5.174	3.294	3.281	3.309	3.971
FIN	15.047	15.319	8.114	8.058	8.551	23.180	18.732	18.998	19.238	7.684
QMJ	5.633	5.605	2.579	2.561	2.614	8.914	5.280	5.259	5.306	6.330
BAB	12.982	13.255	7.957	7.619	8.629	19.166	15.680	15.939	16.189	7.673
MGMT	8.056	8.021	6.854	6.788	6.993	9.370	9.765	9.725	9.817	4.290
PERF	15.932	15.961	7.725	7.594	7.759	25.047	13.247	13.366	13.466	21.297
LIQ	11.977	11.681	10.287	10.023	10.508	13.476	11.203	11.100	11.237	12.902
ICR	45.186	44.746	40.582	40.182	41.258	49.493	43.998	43.363	44.240	47.397

14 factors. The integrated model variance is also more than double the sample variance for the RMW, CMA, FIN, MGMT, and BAB factors.

Overall, the mixture of estimation risk and model disagreement components in the covariance matrix jointly have a sizable impact on perceived risk, especially during the recent out-of-sample period. The rationale is that because the investor does not know the correct factor model or the correct values of underlying model parameters, equities are perceived to be *considerably* riskier than historical sample estimates.

B. Time-Varying Model Disagreement: Entropy Increase

Our second experiment focuses on the model disagreement component in the covariance matrix. In the BMA setup, the covariance matrix of stock returns is defined in equation (5), where V_t is the mixture of model-implied covariance, and Ω_t summarizes the disagreement among candidate factor models about expected stock returns, as defined in equation (6). While both components account for model uncertainty, Ω_t is particularly informative for understanding the implications of model disagreement.

To measure the relative contribution of the model disagreement component to the covariance matrix, we rely on entropy, a standard measure from information theory. For instance, Van Nieuwerburgh and Veldkamp (2010) model the amount of information transmitted as the reduction in entropy achieved by conditioning on that additional information. Let Σ ($\Sigma|D$) be the covariance matrix before (after) the information is revealed. The entropy reduction is given by the ratio $\frac{|\Sigma|}{|\Sigma|D|}$, where $|\Sigma|$ is the determinant of matrix Σ. Since learning information D can reduce payoff uncertainty, a higher ratio indicates more information acquisition leading to uncertainty reduction.

Similar to the entropy reduction due to additional information, Ω_t can be perceived as an entropy extension arising from model disagreement. Hence, we measure the contribution of the model disagreement component to the covariance matrix as the increase in entropy relative to the V_t component (the mixture of model-implied covariance),

$$EI_t = \frac{|V_t + \Omega_t|}{|V_t|}. \tag{25}$$

We compute the relative increase in entropy for the three in-sample periods corresponding to the full sample (T), half of the sample ($\frac{T}{2}$), and two-thirds of the sample ($\frac{2T}{3}$) under a prior Sharpe multiple of 1.5. Panel A of Table VIII reports the mean, 95^{th} percentile, 99^{th} percentile, and maximum of the entropy increase, with the columns "EST" and "OOS" corresponding to the in-sample and out-of-sample results, respectively. The increase in entropy is modest, on average, but positively skewed, that is, the full-sample average is 1.010 but increases to 1.069 at the 99^{th} percentile and reaches a maximum of 1.379. Using $\frac{2T}{3}$ ($\frac{T}{2}$) as the in-sample period, we observe a significant entropy increase of 1.069 (1.121) at the 99^{th} percentile during the out-of-sample period, and the maximum entropy increase is even more prominent at 1.085 (1.195).

Figure 3(a), Panel A, plots the time series of the entropy increase for the three in-sample periods. The blue dashed lines mark the end of the in-sample periods for $\frac{T}{2}$ and $\frac{2T}{3}$. While the average increase in entropy is small, it spikes dramatically during major market downturns, such as Black Monday in October 1987 and the recent financial crisis starting in September 2008. Our findings support the view that asset pricing models disagree significantly about expected stock returns at times of crash events in the financial market, which

Table VIII

BMA Model Uncertainty: Entropy Increase

Panel A presents the in-sample and out-of-sample entropy increase, EI_t in equation (25). We report the mean, 95th percentile, 99th percentile, and maximum of the entropy increase. The columns "EST" report the in-sample entropy increase computed in the full sample (T), as well as in the in-sample periods that correspond to half of the sample ($\frac{T}{2}$) and two-thirds of the sample ($\frac{2T}{3}$). The columns "OOS" report the out-of-sample entropy increase. The investment universe consists of the 14 factors listed in Panel A of the Internet Appendix Table IA.I, and we employ a prior Sharpe multiple of $\tau = 1.5$. Panels B and C report the average and maximum contribution of each factor to the entropy increase, $EI_{k,t}$ in equation (26).

	T	$\frac{T}{2}$		$\frac{2T}{3}$	
	EST	EST	OOS	EST	OOS
Panel A: Entropy Increase					
Mean	1.010	1.009	1.026	1.005	1.012
95^{th} Pctl.	1.013	1.014	1.053	1.007	1.049
99^{th} Pctl.	1.069	1.017	1.121	1.009	1.069
Max	1.379	1.209	1.195	1.098	1.085
Panel B: Average Contribution to the Entropy Increase					
MKT	3.077	8.173	12.842	10.877	17.246
SMB	12.686	8.619	10.484	9.562	7.957
HML	3.080	11.610	13.226	5.116	7.723
RMW	2.175	4.531	4.325	5.127	4.027
CMA	3.020	3.630	4.491	4.298	5.505
MOM	7.936	8.169	3.568	5.656	3.573
PEAD	2.464	4.349	3.239	1.452	2.021
FIN	2.784	3.274	3.791	4.572	5.462
QMJ	7.961	2.826	3.517	5.513	9.170
BAB	11.505	14.687	14.761	6.568	4.663
MGMT	7.430	7.193	9.694	7.810	12.623
PERF	6.704	9.268	4.775	6.764	3.889
LIQ	21.493	3.356	4.263	11.724	5.715
ICR	7.685	10.316	7.024	14.960	10.426
Panel C: Maximum Contribution to the Entropy Increase					
MKT	20.671	19.363	19.763	19.727	25.747
SMB	24.881	22.289	22.957	24.587	13.217
HML	6.503	15.252	17.011	9.172	9.592
RMW	4.890	11.541	10.632	8.138	6.339
CMA	6.112	5.690	6.813	8.290	9.562
MOM	32.882	15.881	11.341	10.937	8.715
PEAD	7.559	7.513	6.585	5.527	5.450
FIN	7.946	5.034	5.083	9.633	8.474
QMJ	21.714	6.075	5.750	10.985	12.617
BAB	25.010	29.716	30.836	14.740	13.094
MGMT	15.934	12.338	11.775	16.400	16.836
PERF	17.779	17.075	9.781	15.012	8.435
LIQ	41.740	10.512	9.023	21.592	10.389
ICR	26.268	15.651	12.681	23.430	16.664

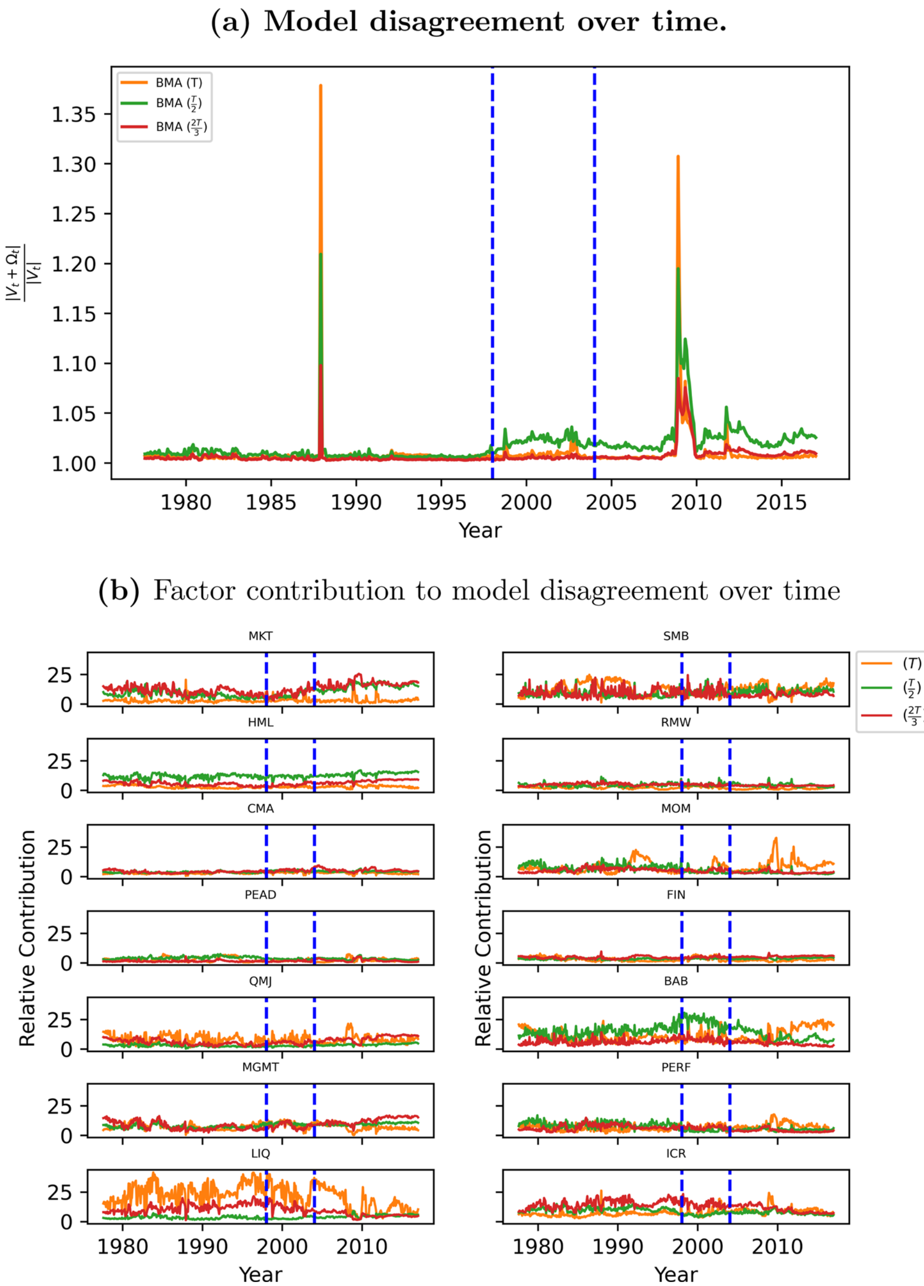


Figure 3. Model disagreement: Entropy increase. This figure plots the contribution of the model disagreement component to the covariance matrix over time. We employ a prior Sharpe multiple of $\tau = 1.5$, and consider three in-sample periods that correspond to the full sample (T), half of the sample ($\frac{T}{2}$), and two-thirds of the sample ($\frac{2T}{3}$). Panel A plots the time series of the relative increase in entropy, EI_t, in equation (25). Panel B plots, for each factor, the time series of the contribution to the overall entropy increase, $EI_{k,t}$ in equation (26). The blue dashed lines mark the end of the in-sample periods for $\frac{T}{2}$ and $\frac{2T}{3}$. (Color figure can be viewed at wileyonlinelibrary.com)

makes stocks appear riskier. Hence, accounting for model uncertainty becomes more important during crash periods, consistent with our previous finding that the BMA procedure mitigates the downside risk of the tangency portfolio and the GMVP.

Beyond the aggregate entropy increase resulting from the entire investment universe, we estimate the contribution of each factor to the entropy increase. In this experiment, we zero out the off-diagonal elements of Ω_t in equation (6). Let $\Omega_{k,t}$ be a matrix with only the k$^{\text{th}}$ diagonal element equal to the corresponding diagonal element of Ω_t and with other elements equal to zero, where $k \in 1, 2, \ldots, K$, with K denoting the maximal number of factors. Similar to the definition in equation (25), we define the entropy increase attributed to factor k at time t as

$$EI_{k,t} = \frac{|V_t + \Omega_{k,t}|}{|V_t|}. \tag{26}$$

When each factor is associated with a small entropy increase, that is, $EI_{k,t} \approx 1$ for every k, a first-order approximation holds—$\prod_{k=1}^{K} EI_{k,t} \approx EI_t$. However, when the entropy spikes, the first-order approximation no longer holds due to a large component of higher order mutual factor interactions. We therefore normalize the measure and define factor k's first-order contribution to the entropy increase as

$$REI_{k,t} = \frac{\frac{\log(EI_{k,t})}{\log(EI_t)}}{\sum_{j=1}^{K} \frac{\log(EI_{j,t})}{\log(EI_t)}}. \tag{27}$$

We consider three in-sample periods corresponding to the full sample (T), half of the sample ($\frac{T}{2}$), and two-thirds of the sample ($\frac{2T}{3}$) under a prior Sharpe multiple of 1.5 and compute the in-sample and out-of-sample contribution of each factor to the increase in entropy. The results are reported in Table VIII, with Panels B and C corresponding to the average and maximum factor contributions, respectively. As shown in Panel B, the LIQ factor stands out, as it contributes 21% of the total entropy increase in the full sample, followed by the SMB and BAB factors. Jointly, the top three factors account for 46% of the total entropy increase. Moving to the out-of-sample test using $\frac{2T}{3}$ as the in-sample period, the market, MGMT, and ICR factors carry a sizable disagreement component and jointly contribute 40% of the overall entropy increase.

Since the model disagreement component in the overall covariance matrix can be low in normal times but spike occasionally, we are also interested in extreme scenarios. As shown in Panel C, the market, SMB, BAB, MGMT, and ICR factors display drastic entropy increases: all five factors uniformly have a maximum contribution of at least 10% across all in-sample and out-of-sample periods. A possible underlying mechanism is that in addition to the market factor, the other factors also vary significantly with market conditions. For instance, the SMB factor is stronger after periods of low sentiment because small stocks are more likely to be overpriced during high-sentiment periods, and the subsequent correction reduces the size effect (Baker and Wurgler (2006)), the BAB factor is exposed to funding liquidity risk and exhibits lower realized returns following periods with more binding funding constraints (Frazzini and Pedersen (2014)), the MGMT factor is significantly higher following high sentiment episodes due to the correction of overpriced stocks in the short leg (Stambaugh and Yuan (2017)), and the ICR factor is strongly procyclical and low intermediary capital growth coincides with adverse economic shocks (He, Kelly, and Manela (2017)). Together, the market, MGMT, and ICR factors play a critical role in driving the time-varying model disagreement component in the covariance matrix both on average and in the extreme, especially for the recent out-of-sample performance. We confirm this finding in Figure 3(b), Panel B, where for each factor, we plot the contribution to the overall entropy increase over time using the three aforementioned in-sample periods.

C. Disagreement about Mispricing, Loadings, and Risk Premia

We next analyze how mispricing, factor loadings, and risk premia affect model disagreement over time. Per equations (IA.115) and (IA.117) in the Internet Appendix VIII, expected returns are determined by the following seven model-specific components: (i) fixed mispricing (α_0), (ii) time-varying mispricing ($\alpha_1 z_t$), (iii) fixed factor loadings with fixed risk premia ($\beta_0 \alpha_f$), (iv) fixed factor loadings with time-varying risk premia ($\beta_0 a_F z_t$), (v) time-varying factor loadings with fixed risk premia ($\beta_1(I \otimes z_t)\alpha_f$), (vi) time-varying factor loadings with time-varying risk premia ($\beta_1(I \otimes z_t)a_F z_t$), and (vii) time-varying risk premia ($a_F z_t$).

For each of the 14 asset pricing factors, we assess the dispersion in the seven return components across positive probability models.[28] To illustrate, the dispersion corresponding to factor k's fixed mispricing component, α_0, is computed as $\sigma(\alpha_0)_{k,t} = \sqrt{\sum_{l=1}^{L} P_l \times (\alpha_{0,k,l,t} - \alpha_{0,k,t})^2}$, where P_l is the posterior probability for model l, $\alpha_{0,k,l,t}$ is the fixed mispricing component for factor k based on model l, $\alpha_{0,k,t} = \sum_{l=1}^{L} P_l \times \alpha_{0,k,l,t}$ is the average fixed mispricing component for factor k across all candidate models, L is the total number of models, $k \in 1, 2, \ldots, K$, and K is the maximal number of factors.

Table IX presents the results, with Panels A and B corresponding to the average and maximum of each model disagreement component, respectively. We employ the in-sample period that corresponds to two-thirds of the sample ($\frac{2T}{3}$) and a prior Sharpe multiple of 1.5. The columns "EST" and "OOS" report the in-sample and out-of-sample results, respectively. We find that model disagreement, reflected through the dispersion measure, appears in all components and is highly skewed. Focusing on the out-of-sample period, the maximum dispersion is 2.94 to 9.27 times its mean across all factors. For instance, the maxi-

[28] For computational efficiency, we take the top one million models to compute the mean, variance, and dispersion. The cumulative probability of these models equals 0.995488. We renormalize their posterior probabilities, so they sum to one.

Table IX

Model Disagreement in Mispricing, Factor Loadings, and Risk Premia

This table presents the in-sample and out-of-sample model disagreement in fixed mispricing (α_0), fixed factor loadings with fixed risk premia ($\beta_0\alpha_f$), time-varying mispricing ($\alpha_1 z_t$), fixed factor loadings with time-varying risk premia ($\beta_0 a_F z_t$), time-varying factor loadings with fixed risk premia ($\beta_1(I \otimes z_t)\alpha_f$), time-varying factor loadings with time-varying risk premia ($\beta_1(I \otimes z_t)a_F z_t$), and time-varying risk premia ($a_F z_t$), as defined in equations (IA.115) and (IA.117) in the Internet Appendix VIII. The model disagreement is computed as the standard deviation across all candidate models using model posterior probabilities as weights. Panels A and B report, for each factor, the average and maximum of each model disagreement component, respectively. We employ the in-sample period that corresponds to two-thirds of the sample ($\frac{2T}{3}$) and a prior Sharpe multiple of $\tau = 1.5$. The columns "EST" and "OOS" report the in-sample and out-of-sample model disagreement, respectively. The investment universe consists of 14 factors as listed in Panel A of the Internet Appendix Table IA.I.

	$\sigma(\alpha_0)$	$\sigma(\beta_0\alpha_f)$	$\sigma(\alpha_1 z_t)$		$\sigma(\beta_0 a_F z_t)$		$\sigma(\beta_1(I \otimes z_t)\alpha_f)$		$\sigma(\beta_1(I \otimes z_t)a_F z_t)$		$\sigma(a_F z_t)$	
	EST	EST	EST	OOS	EST	OOS	EST	OOS	EST	OOS	EST	OOS
Panel A: Average Model Disagreement												
MKT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.079	0.179
SMB	0.024	0.054	0.045	0.090	0.039	0.082	0.045	0.087	0.007	0.019	0.065	0.108
HML	0.013	0.022	0.026	0.046	0.033	0.073	0.020	0.039	0.003	0.012	0.045	0.110
RMW	0.008	0.017	0.017	0.033	0.025	0.057	0.016	0.033	0.003	0.008	0.024	0.047
CMA	0.008	0.009	0.021	0.041	0.024	0.054	0.016	0.034	0.002	0.010	0.030	0.050
MOM	0.020	0.061	0.099	0.187	0.043	0.060	0.071	0.122	0.007	0.026	0.051	0.089
PEAD	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.037
FIN	0.015	0.030	0.033	0.064	0.055	0.133	0.027	0.058	0.004	0.013	0.048	0.106
QMJ	0.022	0.061	0.024	0.046	0.024	0.045	0.023	0.031	0.004	0.008	0.032	0.071
BAB	0.020	0.042	0.057	0.096	0.040	0.060	0.038	0.069	0.005	0.020	0.066	0.105
MGMT	0.014	0.033	0.027	0.047	0.046	0.097	0.022	0.044	0.003	0.009	0.047	0.112
PERF	0.014	0.029	0.054	0.096	0.028	0.048	0.043	0.067	0.005	0.016	0.048	0.071
LIQ	0.017	0.026	0.045	0.083	0.013	0.025	0.048	0.072	0.006	0.025	0.038	0.079
ICR	0.034	0.039	0.064	0.130	0.089	0.173	0.050	0.091	0.007	0.017	0.085	0.180

(Continued)

Table IX—*Continued*

Panel B: Maximum Model Disagreement												
MKT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.287	0.566
SMB	0.024	0.054	0.191	0.266	0.102	0.182	0.143	0.180	0.092	0.103	0.390	0.342
HML	0.013	0.022	0.249	0.215	0.089	0.239	0.221	0.184	0.157	0.117	0.106	0.300
RMW	0.008	0.017	0.142	0.115	0.145	0.135	0.127	0.108	0.080	0.072	0.174	0.148
CMA	0.008	0.009	0.330	0.263	0.049	0.158	0.179	0.147	0.077	0.087	0.077	0.157
MOM	0.020	0.061	0.947	0.791	0.319	0.291	0.407	0.400	0.203	0.199	0.536	0.430
PEAD	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.166	0.161
FIN	0.015	0.030	0.371	0.298	0.224	0.297	0.347	0.301	0.130	0.115	0.268	0.304
QMJ	0.022	0.061	0.363	0.305	0.127	0.120	0.118	0.102	0.128	0.090	0.148	0.143
BAB	0.020	0.042	1.078	0.863	0.144	0.152	0.622	0.505	0.300	0.261	0.258	0.333
MGMT	0.014	0.033	0.211	0.183	0.126	0.293	0.124	0.113	0.065	0.078	0.144	0.282
PERF	0.014	0.029	0.932	0.733	0.182	0.176	0.394	0.316	0.121	0.123	0.347	0.294
LIQ	0.017	0.026	0.391	0.345	0.045	0.061	0.300	0.254	0.133	0.190	0.073	0.159
ICR	0.034	0.039	0.885	0.691	0.355	0.527	0.332	0.337	0.334	0.243	0.495	0.677

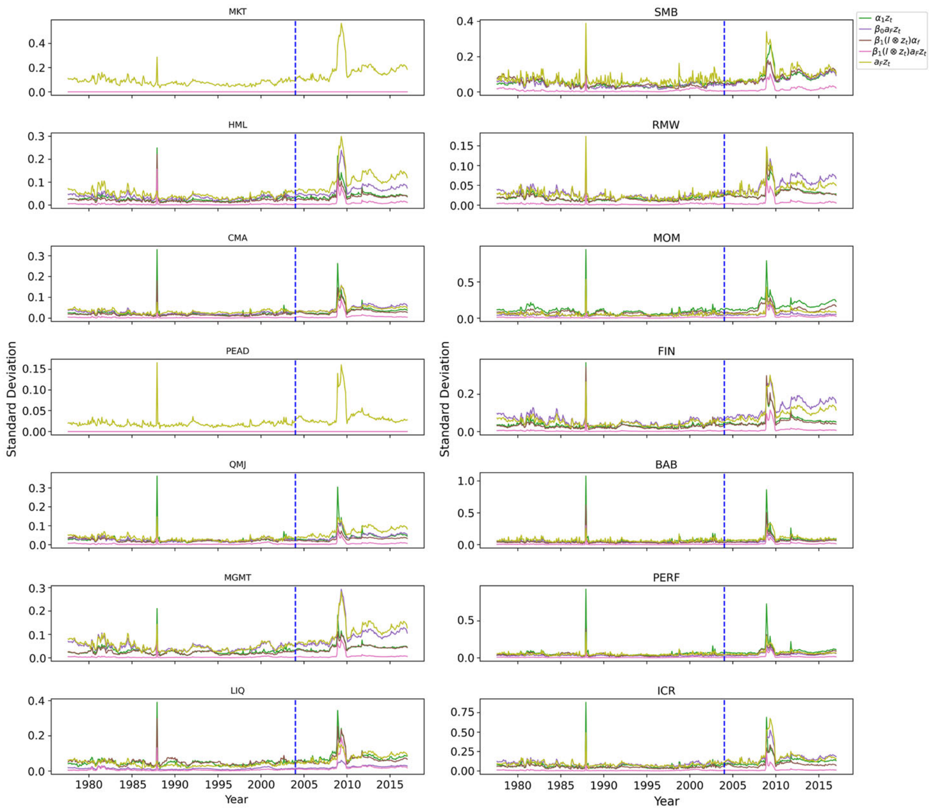


Figure 4. Model disagreement in mispricing, factor loadings, and risk premia. This figure plots, for each factor, the time series of the model disagreement in time-varying mispricing ($\alpha_1 z_t$), fixed factor loadings with time-varying risk premia ($\beta_0 a_F z_t$), time-varying factor loadings with fixed risk premia ($\beta_1(I \otimes z_t)\alpha_f$), time-varying factor loadings with time-varying risk premia ($\beta_1(I \otimes z_t)a_F z_t$), and time-varying risk premia ($a_F z_t$), as defined in equations (IA.115) and (IA.117) in the Internet Appendix VIII. We employ a prior Sharpe multiple of $\tau = 1.5$, and consider the in-sample period that corresponds to two-thirds of the sample ($\frac{2T}{3}$). The blue dashed lines mark the end of the in-sample period for $\frac{2T}{3}$. (Color figure can be viewed at wileyonlinelibrary.com)

mum dispersion in time-varying factor loadings with time-varying risk premia ($\beta_1(I \otimes z_t)a_F z_t$) in Panel B is on average 9.27 times its mean in Panel A across all factors, and the maximum dispersion in time-varying mispricing ($\alpha_1 z_t$) is 5.24 times its mean.

Figure 4 plots the model disagreement components for each factor over time. We observe spikes in all components during crash periods for most factors. We further find that model disagreement regarding the time-varying factor loadings with time-varying risk premia ($\beta_1(I \otimes z_t)a_F z_t$) is 5.84 times its in-sample mean around October 1987 and 2.10 times its out-of-sample mean around September 2008.[29] In addition, the model disagreement regarding time-varying mispricing ($\alpha_1 z_t$), time-varying factor loadings with fixed risk premia ($\beta_1(I \otimes z_t)\alpha_f$), time-varying risk premia ($a_F z_t$), and fixed factor loadings with time-varying risk premia ($\beta_0 a_F z_t$) is 88%, 65%, 27%, and 21% higher during crash events out-of-sample. These experiments thus suggest that during crash episodes, candidate models significantly disagree more on mispricing, factor loadings, and risk premia. They could jointly contribute to the overall entropy increase in Figure 3(a), Panel A.

D. Dispersion in Portfolio Choice and Performance

Finally, we investigate whether individual models differ in their portfolio choice and performance. This provides yet another way to understand the merits of model integration. For each factor k, we compute the dispersion in portfolio choice across positive probability models as $\sigma(w)_{k,t} = \sqrt{\sum_{l=1}^{L} P_l \times (w_{k,l,t} - w_{k,t})^2}$, where P_l is the posterior probability for model l, $w_{k,l,t}$ is the weight for factor k in the tangency portfolio based on model l, $w_{k,t} = \sum_{l=1}^{L} P_l \times w_{k,l,t}$ is the average weight for factor k in the tangency portfolio across all candidate models, L is the total number of candidate models, $k \in 1, 2, \ldots, K$, and K is the maximal number of factors.

Similarly, we compute the dispersion in individual model performance as $\sigma(r)_{t+1} = \sqrt{\sum_{l=1}^{L} P_l \times (r_{l,t+1} - r_{t+1})^2}$, where $r_{l,t+1} = \sum_{k=1}^{K} w_{k,l,t} \times f_{k,t+1}$, $r_{t+1} = \sum_{l=1}^{L} P_l \times r_{l,t+1}$, $f_{k,t+1}$ is factor k's return at time $t+1$, and all other variables are defined as in $\sigma(w)_{k,t}$.

We present the findings in Table X. Panel A reports the mean, 95$^{\text{th}}$ percentile, 99$^{\text{th}}$ percentile, and the maximum of the dispersion in the tangency portfolio weights for each factor. We employ the in-sample period that corresponds to two-thirds of the sample ($\frac{2T}{3}$) and a prior Sharpe multiple of 1.5. The columns "EST" and "OOS" report the in-sample and out-of-sample results, respectively. Focusing on the out-of-sample period, we find that the average dispersion in tangency portfolio weights ranges between 0.008 and 0.045 across all factors. The distribution is also skewed in general, with the dispersion ranging between 0.014 and 0.099 at the 99$^{\text{th}}$ percentile and between 0.017 and 0.103 at the maximum. When compared with the average weights in the tangency portfolio based on the integrated model (Table V), the dispersion among candidate models accounts for 11% to 29 times of the absolute average weight across all factors, with a mean (median) of four times (35%).

Panel B reports similar statistics for the dispersion in the tangency portfolio returns. We find a sizable dispersion in model performance with an average of 0.206 out-of-sample, and it increases sharply to 0.921 at the 99$^{\text{th}}$ percentile and 1.670 at the maximum. For perspective, the average tangency portfolio return (r_{t+1}) is 0.555 during the same period.

[29] We consider a six-month window around October 1987 (July to December 1987) and September 2008 (June to November 2008) as crash periods.

Table X

Dispersion in Portfolio Choice and Performance

Panel A presents the in-sample and out-of-sample dispersion in tangency portfolio weights for the 14 factors listed in Panel A of the Internet Appendix Table IA.I. The dispersion is computed as the standard deviation of tangency portfolio weights across all candidate models using model posterior probabilities as weights. We report the mean, 95^{th} percentile, 99^{th} percentile, and maximum of the dispersion. We employ the in-sample period that corresponds to two-thirds of the sample ($\frac{2T}{3}$) and a prior Sharpe multiple of $\tau = 1.5$, and construct the tangency portfolio for each individual model. The columns "EST" and "OOS" report the in-sample and out-of-sample results, respectively. Panel B presents similar statistics for the dispersion in tangency portfolio returns.

	EST				OOS			
	Mean	95^{th} Pctl.	99^{th} Pctl.	Max	Mean	95^{th} Pctl.	99^{th} Pctl.	Max
Panel A: Dispersion in Tangency Portfolio Weights								
MKT	0.024	0.028	0.029	0.030	0.024	0.029	0.032	0.033
SMB	0.011	0.015	0.018	0.069	0.014	0.034	0.048	0.056
HML	0.028	0.030	0.033	0.059	0.032	0.051	0.060	0.064
RMW	0.046	0.053	0.057	0.116	0.045	0.051	0.078	0.095
CMA	0.034	0.041	0.044	0.071	0.033	0.044	0.053	0.059
MOM	0.014	0.016	0.018	0.031	0.017	0.027	0.029	0.031
PEAD	0.015	0.018	0.021	0.043	0.023	0.064	0.099	0.103
FIN	0.031	0.037	0.040	0.046	0.031	0.036	0.040	0.041
QMJ	0.037	0.042	0.046	0.127	0.036	0.056	0.082	0.101
BAB	0.036	0.039	0.040	0.041	0.038	0.041	0.042	0.042
MGMT	0.017	0.019	0.021	0.039	0.020	0.033	0.038	0.043
PERF	0.042	0.047	0.048	0.065	0.045	0.066	0.076	0.081
LIQ	0.009	0.010	0.011	0.012	0.008	0.013	0.014	0.017
ICR	0.008	0.010	0.011	0.021	0.008	0.014	0.017	0.019
Panel B: Dispersion in Tangency Portfolio Returns								
Return	0.170	0.339	0.538	0.605	0.206	0.496	0.921	1.670

Figure 5(a), Panels A and B, plots the time series of the dispersion in tangency portfolio weights for each factor and of model performance, respectively. Consistent with the spikes in entropy during major market downturns (Figure 3[a], Panel A), candidate models significantly disagree more in their portfolio choice for almost all factors at the same time. While the dispersion in portfolio choice is small on average but spikes during market crashes, the dispersion in model performance is more volatile over time due to the interaction of portfolio weights with realized factor returns. The evidence thus suggests that candidate models vary in their portfolio choice and model performance, with dispersion rising significantly at times of crash events in the financial market. Therefore, accounting for model uncertainty is highly relevant for practitioners in portfolio construction and risk management.

Overall, we find that asset pricing models disagree significantly about expected stock returns during market crashes. The spikes in model disagreement are attributed to the various return components involving mispricing, factor loadings, and risk premia. We also observe similar spikes in the dispersion in portfolio choice and model performance during crash episodes. Our findings support the view that accounting for model uncertainty effectively mitigates downside risk and enhances performance.

VII. Conclusion

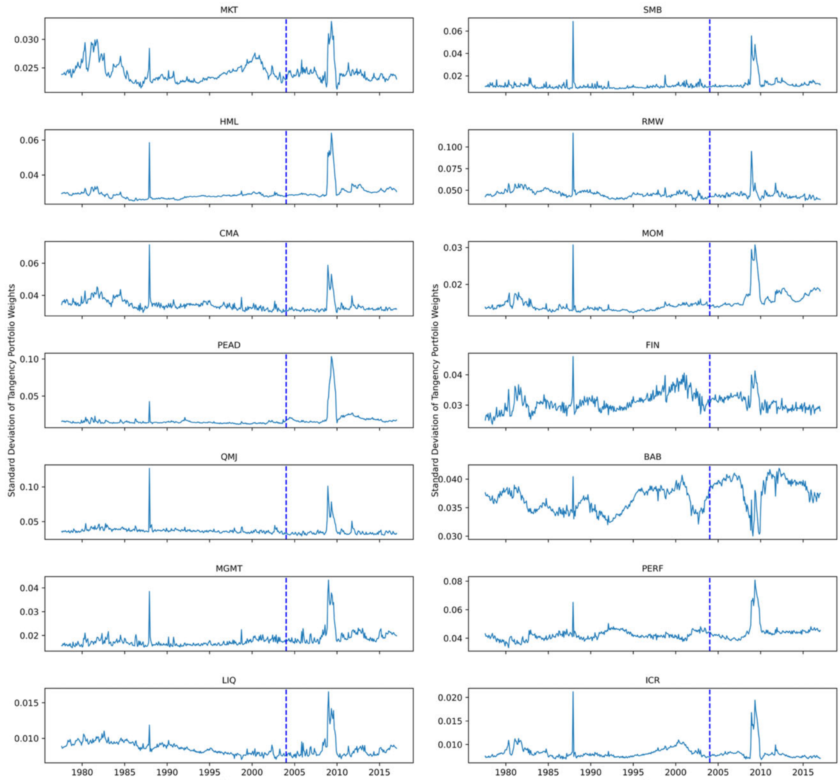
This paper develops a comprehensive Bayesian framework to study average stock returns and the covariance matrix in the presence of model uncertainty. The framework combines a large universe of candidate factor models into an integrated model. Prior beliefs about the entire parameter space are economically interpretable and weighted against deviations from unconditional models. The integrated model is used to assess the strength of factors and predictors in explaining the joint dynamics of stock returns. The empirical analyses apply to a set of 14 factors and 13 macro predictors. The model space exceeds 52 million models that differ with respect to the set of factors and predictors, while some factor models hold exactly and others admit mispricing.

We first document that a fairly large number of models record a positive and meaningful probability with no clear winner. Furthermore, the underlying return-generating process exhibits considerable mispricing and is uniformly dominated by models with time-varying parameters. We next show that the PEAD, QMJ, and intermediary capital factors are potent factors in conditional asset pricing.

From an investment perspective, the integrated model delivers a stable, superior, and admissible out-of-sample Sharpe ratio and mitigates downside risk for both the tangency portfolio and the GMVP. The integrated model mitigates adverse investment outcomes by tilting away from the subsequently underperforming factors. In addition, the Bayesian approach is instrumental in identifying competent individual models, while model selection based solely on top-ranked individual models could provide unstable forecasts.

We finally show that a Bayesian agent who accounts for model uncertainty could perceive equities to be considerably riskier due to model disagreement about expected returns. During adverse market conditions, competing factor

(a) Dispersion in Tangency Portfolio Weights Over Time



(b) Dispersion in Tangency Portfolio Returns Over Time

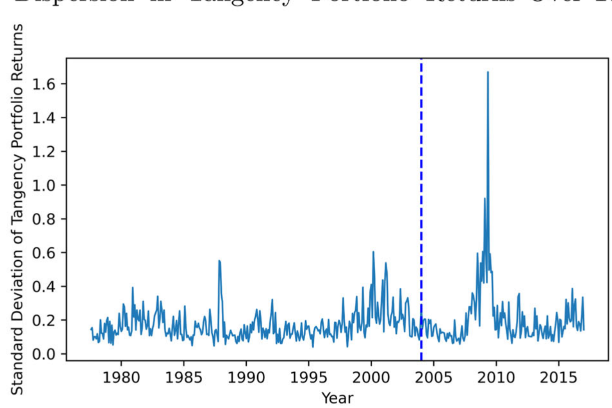


Figure 5. Dispersion in portfolio choice and performance. Panel A plots, for each factor, the time series of the dispersion in tangency portfolio weights. Panel B plots the time series of the dispersion in tangency portfolio returns. We employ a prior Sharpe multiple of $\tau = 1.5$, and consider the in-sample period that corresponds to two-thirds of the sample ($\frac{2T}{3}$). The blue dashed lines mark the end of the in-sample period for $\frac{2T}{3}$. (Color figure can be viewed at wileyonlinelibrary.com)

models with substantial posterior probabilities considerably disagree on all return components involving mispricing, factor loadings, and risk premia.

Initial submission: September 15, 2021; Accepted: January 13, 2022
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

REFERENCES

Anderson, Evan W., and Ai-Ru (Meg) Cheng, 2016, Robust bayesian portfolio choices, *Review of Financial Studies* 29, 1330–1375.

Arnott, Rob, Campbell R. Harvey, Vitali Kalesnik, and Juhani Linnainmaa, 2019, Alice's adventures in factorland: Three blunders that plague factor investing, *Journal of Portfolio Management* 45, 18–36.

Asness, Clifford S., Andrea Frazzini, and Lasse Heje Pedersen, 2019, Quality minus junk, *Review of Accounting Studies* 24, 34–112.

Avramov, Doron, 2002, Stock return predictability and model uncertainty, *Journal of Financial Economics* 64, 423–458.

Avramov, Doron, 2004, Stock return predictability and asset pricing models, *Review of Financial Studies* 17, 699–738.

Avramov, Doron, Scott Cederburg, and Katarína Lučivjanská, 2017, Are stocks riskier over the long run? Taking cues from economic theory, *Review of Financial Studies* 31, 556–594.

Avramov, Doron, and John C. Chao, 2006, An exact Bayes test of asset pricing models with application to international markets, *Journal of Business* 79, 293–324.

Avramov, Doron, and T. Chordia, 2006, Predicting stock returns, *Journal of Financial Economics* 82, 387–415.

Avramov, Doron, and Ross Wermers, 2006, Investing in mutual funds when returns are predictable, *Journal of Financial Economics* 81, 339–377.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Barillas, Francisco, and Jay Shanken, 2017, Which alpha?, *Review of Financial Studies* 30, 1316–1338.

Barillas, Francisco, and Jay A. Shanken, 2018, Comparing asset pricing models, *Journal of Finance* 73, 715–754.

Billio, Monica, Roberto Casarin, Francesco Ravazzolo, and Herman K. van Dijk, 2013, Time-varying combinations of predictive densities using nonlinear filtering, *Journal of Econometrics* 177, 213–232.

Bossaerts, Peter, and Pierre Hillion, 1999, Implementing statistical criteria to select return forecasting models: What do we learn?, *Review of Financial Studies* 12, 405–428.

Bryzgalova, Svetlana, Jiantao Huang, and Christian Julliard, 2023, Bayesian solutions for the factor zoo: We just ran two quadrillion models, *Journal of Finance* 78, 487–557.

Campbell, John Y., Andrew Wen-Chuan Lo, and Archie Craig MacKinlay, 1997, The Econometrics of Financial Markets (Princeton University Press, Princeton, NJ).

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

Chen, Luyang, Markus Pelger, and Jason Zhu, 2023, Deep learning in asset pricing, *Management Science*, forthcoming.

Chen, Nai-Fu, Richard Roll, and Stephen A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383–403.

Chib, Siddhartha, 1995, Marginal likelihood from the Gibbs output, *Journal of the American Statistical Association* 90, 1313–1321.

Chib, Siddhartha, and Xiaming Zeng, 2019, Which factors are risk factors in asset pricing? A model scan framework, *Journal of Business & Economic Statistics* 38, 771–783.

Chib, Siddhartha, Xiaming Zeng, and Lingxiao Zhao, 2020, On comparing asset pricing models, *Journal of Finance* 75, 551–577.

Chib, Siddhartha, Lingxiao Zhao, and Guofu Zhou, 2023, Winners from winners: A tale of risk factors, *Management Science*, forthcoming.

Cochrane, John H., 2011, Presidential address: Discount rates, *Journal of Finance* 66, 1047–1108.

Cochrane, John H., 2009, Asset Pricing:(Revised Edition) (Princeton University Press, Princeton, NJ).

Cong, Lin, Ke Tang, Jingyuan Wang, and Yang Zhang, 2021, Alphaportfolio for investment and economically interpretable ai, Working Paper .

Daniel, Kent, David Hirshleifer, and Lin Sun, 2020, Short- and long-horizon behavioral factors, *Review of Financial Studies* 33, 1673–1736.

Daniel, Kent, and Tobias J. Moskowitz, 2016, Momentum crashes, *Journal of Financial Economics* 122, 221–247.

DeMiguel, Victor, Alberto Martín-Utrera, Francisco J. Nogales, and Raman Uppal, 2020, A transaction-cost perspective on the multitude of firm characteristics, *Review of Financial Studies* 33, 2180–2222.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 2015, A five-factor asset pricing model, *Journal of Financial Economics* 116, 1–22.

Fama, Eugene F., and Kenneth R. French, 2018, Choosing factors, *Journal of Financial Economics* 128, 234–252.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2020, Taming the factor zoo: A test of new factors, *Journal of Finance* 75, 1327–1370.

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross section of stock returns, *Journal of Finance* 54, 1325–1360.

Frazzini, Andrea, David Kabiller, and Lasse Heje Pedersen, 2018, Buffett's alpha, *Financial Analysts Journal* 74, 35–55.

Frazzini, Andrea, and Lasse Heje Pedersen, 2014, Betting against beta, *Journal of Financial Economics* 111, 1–25.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber, 2020, Dissecting characteristics nonparametrically, *Review of Financial Studies* 33, 2326–2377.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang, 2007, Portfolio selection with parameter and model uncertainty: A multi-prior approach, *Review of Financial Studies* 20, 41–81.

Geweke, John, and Gianni Amisano, 2011, Optimal prediction pools, *Journal of Econometrics* 164, 130–141.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri, 2015, Prior selection for vector autoregressions, *Review of Economics and Statistics* 97, 436–451.

Gibbons, Michael R., Stephen A. Ross, and Jay A. Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121–1152.

Green, Jeremiah, J. Hand, and X. Zhang, 2017a, The remarkable multidimensionality in the cross-section of expected U.S. stock returns, Working Paper .

Green, Jeremiah, John R. M. Hand, and X. Frank Zhang, 2017b, The characteristics that provide independent information about average U.S. monthly stock returns, *Review of Financial Studies* 30, 4389–4436.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical asset pricing via machine learning, *Review of Financial Studies* 33, 2223–2273.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2021, Autoencoder asset pricing models, *Journal of Econometrics* 222, 429–450.

Harvey, Campbell R., 2017, Presidential address: The scientific outlook in financial economics, *Journal of Finance* 72, 1399–1440.

Harvey, Campbell R., Yan Liu, and Heqing Zhu, 2016, ... and the cross-section of expected returns, *Review of Financial Studies* 29, 5–68.

Harvey, Campbell R., and Guofu Zhou, 1990, Bayesian inference in asset pricing tests, *Journal of Financial Economics* 26, 221–254.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

Hoeting, Jennifer A., David Madigan, Adrian E Raftery, and Chris T Volinsky, 1999, Bayesian model averaging: A tutorial, *Statistical Science* 14, 382–417.

Hou, Kewei, Haitao Mo, Chen Xue, and Lu Zhang, 2021, An Augmented *q*-Factor Model with Expected Growth, *Review of Finance* 25, 1–41.

Hou, Kewei, Chen Xue, and Lu Zhang, 2015, Digesting anomalies: An investment approach, *Review of Financial Studies* 28, 650–705.

Hou, Kewei, Chen Xue, and Lu Zhang, 2020, Replicating anomalies, *Review of Financial Studies* 33, 2019–2133.

Jacobs, Bruce I., Kenneth N. Levy, and David Starer, 1999, Long-short portfolio management: An integrated approach, *Journal of Portfolio Management* 25, 23–32.

Kandel, Shmuel, and Robert F. Stambaugh, 1995, Portfolio inefficiency and the cross-section of expected returns, *Journal of Finance* 50, 157–184.

Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset-allocation perspective, *Journal of Finance* 51, 385–424.

Kass, Robert E., and Adrian E. Raftery, 1995, Bayes factors, *Journal of the American Statistical Association* 90, 773–795.

Kozak, Serhiy, Stefan Nagel, and Shrihari Santosh, 2020, Shrinking the cross-section, *Journal of Financial Economics* 135, 271–292.

Ledoit, Olivier, and Michael Wolf, 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* 10, 603–621.

Ledoit, Olivier, and Michael Wolf, 2004, Honey, i shrunk the sample covariance matrix, *Journal of Portfolio Management* 30, 110–119.

Light, Nathaniel, Denys Maslov, and Oleg Rytchkov, 2017, Aggregation of information about the cross section of stock returns: A latent variable approach, *Review of Financial Studies* 30, 1339–1381.

Manresa, Elena, Francisco Peñaranda, and Enrique Sentana, 2017, Empirical evaluation of over-specified asset pricing models, Working Paper.

McCulloch, Robert, and Peter E. Rossi, 1991, A Bayesian approach to testing the arbitrage pricing theory, *Journal of Econometrics* 49, 141–168.

Pástor, Ľuboš, and Robert F. Stambaugh, 1999, Costs of equity capital and model mispricing, *Journal of Finance* 54, 67–121.

Pástor, Ľuboš, and Robert F. Stambaugh, 2000, Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* 56, 335–381.

Pástor, Ľuboš, and Robert F. Stambaugh, 2002a, Investing in equity mutual funds, *Journal of Financial Economics* 63, 351–380.

Pástor, Ľuboš, and Robert F. Stambaugh, 2002b, Mutual fund performance and seemingly unrelated assets, *Journal of Financial Economics* 63, 315–349.

Pástor, Ľuboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Pastor, Lubos, and Robert F. Stambaugh, 2009, Predictive systems: Living with imperfect predictors, *Journal of Finance* 64, 1583–1628.

Poirier, Dale J., 1995, Intermediate Statistics and Econometrics: A Comparative Approach (MIT Press, Cambridge, MA, and London).

Robbins, Herbert, 1956, An empirical bayes approach to statistics, in Jerzy Neyman, ed.; Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, volume 1 of *Third Berkeley Symposium on Mathematical Statistics and Probability*, 157–164 (University of California Press, Berkeley, CA).

Robbins, Herbert, 1964, The empirical Bayes approach to statistical decision problems, *Annals of Mathematical Statistics* 35, 1–20.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Stambaugh, Robert F., Jianfeng Yu, and Yu Yuan, 2012, The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* 104, 288–302.

Stambaugh, Robert F., and Yu Yuan, 2017, Mispricing factors, *Review of Financial Studies* 30, 1270–1315.

Van Nieuwerburgh, Stijn, and Laura Veldkamp, 2010, Information acquisition and under-diversification, *Review of Economic Studies* 77, 779–805.

Wachter, Jessica A., and Missaka Warusawitharana, 2009, Predictable returns and asset allocation: Should a skeptical investor time the market?, *Journal of Econometrics* 148, 162–178.

Wachter, Jessica A., and Missaka Warusawitharana, 2015, What is the chance that the equity premium varies over time? Evidence from regressions on the dividend-price ratio, *Journal of Econometrics* 186, 74–93.

Welch, Ivo, and Amit Goyal, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication Code.