Integrating Factor Models*

Doron Avramov† Si Cheng‡ Lior Metzker§ Stefan Voigt¶

This Version: November 23, 2021

Abstract

This paper develops a comprehensive framework to address uncertainty about the correct factor model specification. Asset pricing inferences draw on a composite model that integrates over competing factor models weighted by posterior probabilities. The analyses show that models with time-invariant parameters record near-zero probabilities, and post-earnings announcement drift, quality-minus-junk, and intermediary capital factors are incremental to the market. From an investment perspective, augmenting the model disagreement about expected returns and the probability weighted uncertainty about model parameters, a Bayesian agent perceives equities as *considerably* riskier than sample estimates. In addition, the integrated model delivers stable strategies, even during market downturns.

Keywords: Factor models, Model integration, Posterior probability, Stock return predictability, Mispricing.

JEL classification: C11, C12, C52, G12

*We thank John Chao, Stefan Nagel, and two anonymous referees for useful comments and suggestions. Doron Avramov acknowledges the Israel Science Foundation (ISF) (Grant Number: 288/18) for financial support.

†Interdisciplinary Center (IDC), Herzliya, Israel. doron.avramov@idc.ac.il

‡Chinese University of Hong Kong, Hong Kong. sicheng@cuhk.edu.hk

§Hebrew University of Jerusalem, Jerusalem, Israel. lior.metzker@mail.huji.ac.il

¶University of Copenhagen and Danish Finance Institute, Copenhagen, Denmark. stefan.voigt@econ.ku.dk

1 Introduction

Financial economists have identified a plethora of firm characteristics that predict future stock returns (e.g., Cochrane (2011) and Harvey et al. (2016)). The literature has further proposed two major approaches to reduce the expanding dimension of cross-sectional predictors. The first invokes economic rationales, e.g., 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 the most remains subject to research controversy.[1] Significant uncertainty also extends to the choice of macro variables that potentially govern time-varying investment opportunities.[2] 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.[3]

Somewhat surprisingly, a comprehensive analysis of Bayesian model uncertainty has not been accounted for in formulating expected returns and second moments or in deriving mean-variance efficient portfolios. When addressing model uncertainty, the researcher's core tasks are to identify a universe of competing factor models, assess the probability that a candidate model generates the observed dynamics of asset returns, and then

[1]See, e.g., Ross (1976), Fama and French (2015), and Hou et al. (2015) for the first approach, and Green et al. (2017a), Light et al. (2017), Manresa et al. (2017), Messmer and Audrino (2017), Feng et al. (2020), Freyberger et al. (2020), Gu et al. (2020), Kozak et al. (2020), Chen et al. (2021), and Cong et al. (2021) for the second. Notably, the various specifications could disagree on the set of factors that matter the most.

[2]Cochrane (2011) argues that by virtue of the present value model, the price-to-dividend ratio must predict future returns. However, as the present value is based on unobserved expectations, the true set of predictors, beyond the price-to-dividend ratio, is unknown.

[3]Resorting to Instrumented Principal Component Analysis (IPCA), Kelly et al. (2019) extract common factors both including and excluding mispricing. Employing time-series asset pricing regressions, Ferson and Harvey (1999) and Avramov (2004) show that mispricing, with respect to the Fama-French 3-factor model, varies with macro variables; the former study focuses on statistical significance, while the latter analyzes investment performance.

integrate over the vast model universe using model posterior probabilities as weights. This approach, termed Bayesian model averaging (BMA), formulates a composite model that summarizes the uncertainties about the joint dynamics of stock returns.[4] Inferences about mean returns, the covariance matrix, and possibly even higher moments are made based on that integrated model. Hence, the analysis of the cross section is conditioned on the whole information set instead of relying only on information contained in a single model. In that way, the Bayesian approach could temper the data snooping concerns identified in the literature (e.g., Harvey et al. (2016), Harvey (2017), Hou et al. (2020)).

This paper develops and applies a Bayesian framework to study average returns and the covariance matrix 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 specifications, which either hold exactly or admit various degrees of mispricing. A key challenge in the analysis is the formulation of model posterior probability, or the probability that a candidate model generates the joint dynamics of returns. In particular, it is essential to motivate economically interpretable priors for all parameters underlying the factor model and the time-series evolution of factor risk premia.

Pástor and Stambaugh (1999), and a large body of work, propose economic priors on model mispricing in an unconditional setup. Instead, we suggest economically interpretable priors on the *entire* parameter space for models with fixed and time-varying parameters. Prior beliefs are weighted against predictability by macro variables and model mispricing. The posterior probability employs sound economic appeals and penalizes model complexity to the extent that an incremental factor or macro predictor is retained in the pricing model specification only if it considerably improves pricing abilities.

In the presence of model uncertainty, expected returns are a mixture of model-implied expected returns, where mixture stands for value weighting using model posterior prob-

[4] Avramov (2002) implements model averaging in the context of predictive regressions.

abilities as weights. The covariance matrix consists of three components. The first is a mixture of model-implied covariance, assuming the model parameters are known. The other two components evolve from uncertainty about the correct factor model and its underlying parameters. The first of the two components is a mixture of estimation risk, namely, the risk that underlying model parameters are estimated with errors. The second summarizes model disagreement. Intuitively, a stock appears riskier when there is greater disagreement among candidate models about its expected return. Similar to the Ridge regression approach, the disagreement component could make an otherwise ill-conditioned covariance matrix of stock returns readily invertible. Through BMA, the predictive distribution of future returns integrates out the within-model parameter space (parameter uncertainty) as well as the model space (model disagreement). Thus, the Bayesian 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 that differ in the inclusion of cross-sectional factors, macro predictors, and the presence of mispricing. We start by examining some stylized model features. First, 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 that there is no clear winner across the whole space of candidate models.[5] Instead, a plethora of distinct models record a positive and meaningful probability of governing the joint distribution of stock returns. While model selection would narrow down the focus to a single factor model, or a few ones, the Bayesian approach integrates the dynamics of nonzero probability models.

Second, even when prior beliefs are weighted against time-varying moments, our procedure uniformly favors conditional models and indicates that both factor loadings and

[5]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.

risk premiums vary with macroeconomic conditions.[6] Remarkably, in the presence of conditional factor models, the cumulative probability of unconditional models is near zero. Likewise, while prior beliefs are weighted against model mispricing, the composite model shows that time-varying mispricing appears with a probability that ranges between 58% and 69%. Hence, zero-alpha models selected from the collection of factors and macro predictors may not adequately explain cross-sectional and time-series effects in returns.

Analyzing the strength of factors in the integrated model, several findings are worth noting. First, for a reasonable prior Sharpe ratio, the post-earnings announcement drift (PEAD, from Daniel et al. (2020)) and quality-minus-junk (QMJ, from Asness et al. (2019)) display a posterior inclusion probability of close to 100%, followed by investment (CMA, from Fama and French (2015)), size (SMB, from Fama and French (1993)), intermediary capital (ICR, from He et al. (2017)), and management (MGMT, from Stambaugh and Yuan (2017))—which all offer a posterior inclusion probability of at least 90%, highlighting their promise in pricing other factors. Our findings also suggest that despite the expanding factor zoo, several new factors proposed after 2015, including both fundamental and behavioral factors, are incrementally competent in pricing the existing factors.

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., SMB, CMA, and MGMT) while challenging others (e.g., BAB). Likewise, bounding the Sharpe ratio has implications for the inclusion of macro predic-

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

tors. For instance, the net equity expansion and the Treasury Bill yield appear with almost zero probability for sensible values of prior Sharpe ratios. However, when prior Sharpe ratios get considerably higher, both the macro items record 90% probability of inclusion. This evidence reflects the notion that strong in-sample evidence on time-series predictability need not extend out-of-sample. Imposing bounds on the Sharpe ratio is advocated in prior work. For instance, Kozak et al. (2018) point out that in the presence of sentiment investors who cannot take extreme positions and a small number of arbitrageurs, extremely high Sharpe ratio investment strategies are unlikely to prevail.

Third, our results support a model with time-varying parameters and with five to seven factors recording high inclusion probabilities. The prominent factors originate from distinct economic foundations rather than an established well-known model. For instance, PEAD, QMJ, and ICR are proposed by three independent works, and this combined specification has not been examined in the previous literature.

We next assess the out-of-sample performance of the composite model through tangency portfolios that are based on a predictive distribution that integrates out the model space and the within-model parameter space. 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, i.e., the CAPM, the Fama-French 3-factor model (Fama and French (1993)), the Fama-French 6-factor model (Fama and French (2018)), and the AQR 6-factor model (Frazzini et al. (2018)). We further consider the three top-ranked individual models, namely, the three highest posterior probability models based on the Bayesian procedure.

The integrated model outperforms the benchmark models out of sample. For instance, the integrated model generates an annualized Sharpe ratio of 1.240, indicating a 8% improvement from 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.[7] Then, the integrated model produces an out-of-sample annualized Sharpe ratio of 0.979 and outperforms 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 model exhibits less negative skewness, lower excess kurtosis, and lower maximum drawdown, to the extent that there are only modest declines in the 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 based on a few top-ranked models could provide an unstable description of asset return dynamics, while model integration improves the stability of forecasts.

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

Indeed, the integrated model based GMVP 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 appears to be only 0.756% for the integrated model, indicating a 21% to 64% volatility reduction. In addition, the maximum drawdown (throughout the entire sample) for the benchmark models ranges between 6% and 27%, compared to 5% for the integrated model. Because expected returns across the specifications are not materially different, the lower volatility

[7] The Federal Reserve Board Regulation T mandates maximum two-to-one leverage (e.g., Jacobs et al. (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 smaller than 2, where 2 is obtained by dividing 1 by the initial margin of 50%.

characterizing the Bayesian approach translates into higher out-of-sample Sharpe ratio. Specifically, the integrated model based GMVP generates an annualized Sharpe ratio of 1.101 and outperforms the best benchmark model by 35%. The results highlight the sizable impact of model uncertainty on the covariance matrix of stock returns.

Finally, we conduct two experiments to further highlight 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 utilizes information from macro variables. With model uncertainty accounted for, however, there are two conflicting forces underlying the variance comparison. Empirically, we show that 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 14 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. In other words, a Bayesian agent who accounts for model uncertainty perceives equities 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 the entropy increase to measure the contribution of model uncertainty to the covariance matrix. While on average, the increase in entropy is modest, it spikes dramatically during major market downturns, e.g., Black Monday in October 1987 and the recent financial crisis starting in September 2008. Compared to a benchmark value of 1, indicating no entropy increase, the full sample average is 1.010 but increases to 1.069 at the 99^{th} percentile and reaches a maximum of 1.379. In addition, we estimate the contribution of every single factor to the overall entropy increase. Interestingly, the time-varying model uncertainty component is primarily driven by the market, MGMT, and ICR factors, and all three factors uniformly have a maximum contribution of at least 10% to the total entropy increase in the full sample and various subperiods. In sum, asset pricing models significantly disagree about expected stock returns during crash events in the financial market. Taking this finding along with the stable performance of the integrated model based efficient portfolios, we argue that accounting for model uncertainty effectively hedges against market crashes.

Taken together, the proposed Bayesian approach delivers stable and superior strategies. It further mitigates the downside risk and volatility of efficient portfolios. Our findings are robust to imposing economic restrictions on the admissible Sharpe ratios and on long and short stock positions. Notably, model uncertainty makes equities appear considerably riskier, while model disagreement about expected stock return especially spikes around market downturns.

To our knowledge, Avramov and Chao (2006) is the first work to formally compare asset pricing models, both nested and nonnested, through the metric of posterior probabilities. Follow-up studies include Anderson and Cheng (2016), Stambaugh and Yuan (2017), Barillas and Shanken (2018), Chib and Zeng (2019), Chib et al. (2019), Bryzgalova et al. (2020), and Chib et al. (2020). Our study differs from these in four major ways. First, related work on model comparison and factor selection is based on rankings of posterior probabilities, while we propose a novel model combination approach that integrates over the space of candidate models.[8] Second, existing studies typically focus on unconditional factor models, while we consider time-varying mispricing, factor loadings, and risk premia and provide supportive evidence for nonlinear dependence between expected returns and macro items. Third, prior beliefs about the entire parameter space

[8]In an independent work, Bryzgalova et al. (2020) develop a Bayesian estimator for linear stochastic discount factors (SDFs) and implement model averaging. We differ from their setup by (i) considering beta pricing specifications, (ii) proposing economically interpretable priors for the entire parameter space, (iii) implementing mean-variance portfolios based on the weighted model, (iv) studying the implications of model uncertainty for the riskiness of equities, and (v) dissecting time-series variation in model disagreement about expected returns.

are economically interpretable in our setup, while training samples are often considered to formulate statistically informed priors. Finally, examining the implications of the integrated model for the covariance matrix of returns is novel.

The remainder of the paper proceeds as follows. Section 2 derives a general methodology for analyzing asset pricing with model uncertainty. Section 3 derives the posterior probabilities for predictive regressions and factor models. Section 4 describes the data. Section 5 presents a probability analysis of factor models as well as the inclusion of individual factors and predictors. Section 6 assesses the out-of-sample performance of the integrated model through both tangency portfolios and GMVPs. Section 7 presents evidence on the riskiness of equities in the presence of model uncertainty and dissects the time-series variation in the implications of model disagreement about expected returns. Section 8 concludes the paper.

2 Asset Pricing with Model Uncertainty

This section develops an analytical framework for studying asset pricing in the presence of model uncertainty. A key challenge in the analysis is computing model probabilities, or the probability that a candidate factor model generates the joint dynamics of asset returns. To pursue that task, it is essential to formulate economically meaningful prior beliefs for the entire set of parameters underlying the factor model. This point merits further discussion. In a general context, combining an improper prior with a likelihood function yields a posterior distribution that is well defined and interpretable. In computing posterior probabilities, however, the prior density must be fully specified and avoid undefined constants characterizing a flat prior.[9]

In financial economics, a large body of work motivates economically meaningful priors, but on a subset of the parameter space. For instance, Pástor and Stambaugh (1999),

[9]See, e.g., the discussions in Kass and Raftery (1995) and Poirier (1995).

Pástor (2000), and Pástor and Stambaugh (2000) account for prior information about mispricing, or alpha, which translates into a certain degree of belief in model pricing abilities. Kozak et al. (2020) impose an economically motivated prior on stochastic discount factor (SDF) coefficients. They introduce a novel prior for mispricing that applies when factors are ordered eigenvectors. In this paper, we propose economically interpretable priors for the *entire* parameter space underlying beta pricing specifications when factors are prespecified. Moreover, we account for the possibility that model mispricing, factor loadings, and risk premiums (all or subsets) vary with business conditions.

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 a time series is denoted by T and the time t subscript represents time t realizations.

Excess returns are modeled through the time-series 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 modeled as $\alpha(z_t) = \alpha_0 + \alpha_1 z_t$ and $\beta(z_t) = \beta_0 + \beta_1 (I_K \otimes z_t)$, where $\otimes$ denotes the Kronecker product, and I_K is the identity matrix of size K. Then, excess stock returns can be re-expressed 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 are 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 $\alpha_0 \neq 0$, then time-invariant model mispricing is present, while when $\alpha_1 \neq 0$, model mispricing varies with macro conditions. 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) recognizes the possibility that risk premiums are also 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. In particular, does a prespecified factor model really explain the cross-sectional variation in average stock returns? Pástor and Stambaugh (1999) show that uncertainty about model pricing abilities could be substantial. Moreover, Gibbons et al. (1989), among others, derive classical asset pricing statistics to test zero-alpha restrictions (see Campbell et al. (1997) and Cochrane (2009) for a comprehensive coverage), while Harvey and Zhou (1990), McCulloch and Rossi (1991), Kandel and Stambaugh (1995), and Avramov and Chao (2006) develop Bayesian asset pricing tests.

Second, there is substantial uncertainty about the identity of asset pricing factors. Remarkably, Harvey et al. (2016) count 316 factors and Hou et al. (2020) cover 452 anomalies. To address the expanding dimension of the cross section, two major approaches have been proposed. The first identifies a small number of factors based on sound economic appeals. 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 et al. (2015) and Hou et al. (2021) propose q-factor models that draw on the q-theory of investment. Stambaugh and Yuan (2017) advocate two mispricing factors based on 11 anomalies studied in Stambaugh et al. (2012). The second approach proposes shrinkage methods such as Lasso, Ridge, and their extensions (e.g., Green et al. (2017b), DeMiguel et al. (2020), Feng et al. (2020), Freyberger et al. (2020), and Kozak et al. (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 is subject to research controversy.

The third type of uncertainty concerns the identity of macro variables that forecast changing investment opportunities. Past work has addressed this uncertainty through the predictive regression setup. In particular, when M macro variables are suspected to be relevant in predicting future returns, there are altogether 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 operates as if the model is correct with a 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 the Bayes rule and is justified from a decision-making perspective. The Bayesian method assigns posterior probabilities to each of the 2^M competing specifications and then uses the probabilities as weights on the individual models to obtain a composite weighted model. BMA displays robust out-of-sample predictive power relative to the model selection criteria (e.g., Avramov (2002)).

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, and then compute the posterior probability for each candidate model. Models differ with respect to the three sources of uncertainty described earlier. Panel A of Table

1 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 premiums. 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 formulated as

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

where D stands for the observed data consisting of a balanced panel of N test assets, K factors, and M macro predictors through T periods, l is a model-specific subscript, $\mathcal{M}_l$ represents a candidate factor model, $P\left(\mathcal{M}_l|D\right)$ is the model posterior probability, $E\left[r_{t+1}|\mathcal{M}_l, D\right]$ 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 two components as

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

where $V_t = \sum_{l=1}^{L} P\left(\mathcal{M}_l|D\right) \text{Var}\left[r_{t+1}|\mathcal{M}_l, D\right]$ is the weighted average of model-implied covariance (denoted $\text{Var}\left[r_{t+1}|\mathcal{M}_l, D\right]$), using posterior probabilities as weights. The model-implied covariance possibly varies over time with the factor loadings.

The V_t component can further be decomposed into two items. The first value weights (using posterior probabilities) the usual covariance matrices, similar to the classical approach but relying on posterior means rather than maximum likelihood estimates. The second value weights (again, using posterior probabilities) the estimation risk component. Estimation risk comes into play because the parameters in a Bayesian setting are stochastic.

In addition, Ω_t is given by

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

The Ω_t component summarizes the disagreement among candidate models about expected stock returns. The incremental variation is larger for an asset when candidate models disagree more about its expected returns. When restricting Ω_t to be diagonal, the matrix $\mathrm{Var}\left[r_{t+1}|D\right]$ can be readily invertible even when V_t is singular or ill-conditioned.[10] Thus, the addition of Ω_t resembles the Ridge regression penalty, but there are important differences. In Ridge regressions, the variance of returns takes the form $\mathrm{Var}\left[r_{t+1}|\mathcal{M}, D\right] = \tilde{V}_t + \gamma I_N$, 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.[11]

The covariance matrix decomposition in equation (5) has a close resemblance 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.

Taken together, the integrated model is associated with a three-component covariance

[10] In the empirical analyses, we keep Ω_t general enough to enable covariances due to cross-model disagreements.

[11] While the Ridge shrinkage has a single tuning parameter, it assigns more prominence to components with higher eigenvalues.

matrix, including (i) a mixture of model-implied covariance matrices assuming model parameters are known, (ii) a mixture of estimation risk, and (iii) the model disagreement about expected returns. As noted in the introduction, mixture stands for value weighting using posterior probabilities as weights.

Note that while BMA follows directly from the Bayes rule, as noted earlier, there are other approaches for model combination. Examples include decision-based model combinations per Billio et al. (2013) and optimal prediction pooling per Geweke and Amisano (2011).[12] BMA would, ex ante, be optimal under several loss functions, including the log loss and the squared error loss (Hoeting et al. (1999)).

3 Deriving Posterior Probabilities

3.1 General Formulation

Let θ denote the parameter space that is unique 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 $\mathcal{M}_l$.

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 likeli-

[12] Neither approach shows promise when applied to our setting. First, we have 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 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 at work. Narrowing the focus to a manageable universe of unconditional models leads to weak performance.

hood is computed (dropping the model-specific subscript to ease notation) as

$$\begin{aligned} m(D|\mathcal{M}) &= \int_\theta \mathcal{L}(D|\theta,\mathcal{M})\pi(\theta|\mathcal{M})d\theta \\ &= \frac{\mathcal{L}(D|\theta,\mathcal{M})\pi(\theta|\mathcal{M})}{\pi(\theta|D,\mathcal{M})}. \end{aligned} \tag{7}$$

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.

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

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

where $P(\mathcal{M}_l)$ is the prior probability that model $\mathcal{M}_l$ is correct. In the absence of a compelling reason to favor, ex ante, one model over another, we choose flat prior model probabilities.[13]

Given the general formulation, we next attempt to compute marginal likelihoods for competing models. Our informative prior distribution is based on a hypothetical sample of length T_0.[14] In that sample, the means and variances of stock returns, factors, and predictors are set equal to the actual sample counterparts given by

$$\begin{aligned} \bar{r} &= \frac{1}{T}\sum_{t=1}^{T} r_t & \hat{V}_r &= \frac{1}{T}\sum_{t=1}^{T}(r_t-\bar{r})(r_t-\bar{r})' \\ \bar{f} &= \frac{1}{T}\sum_{t=1}^{T} f_t & \hat{V}_f &= \frac{1}{T}\sum_{t=1}^{T}(f_t-\bar{f})(f_t-\bar{f})' \end{aligned}$$

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

[14]Our prior specification draws on Kandel and Stambaugh (1996), Pástor and Stambaugh (1999, 2000, 2002a,b), Avramov (2004), and Avramov and Wermers (2006).

$$
\begin{aligned}
\bar{z} &= \frac{1}{T}\sum_{t=0}^{T-1} z_t & \hat{V}_z &= \frac{1}{T}\sum_{t=0}^{T-1}(z_t - \bar{z})(z_t - \bar{z})' \\
\bar{y} &= \frac{1}{T}\sum_{t=1}^{T} y_t & \hat{V}_y &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})(y_t - \bar{y})',
\end{aligned} \tag{9}
$$

where $y_t = [r_t', f_t']'$. The hypothetical sample is also weighted against predictability by macro items and against model mispricing, while its size, T_0, is yet to be formulated. We develop the informed prior below.

Using statistics from the actual sample to specify some of the parameters of the prior distribution is commonly termed "empirical Bayes" (Robbins (1956, 1964)). Note, at this early point, that all our empirical out-of-sample experiments are conducted based on real-time information.

3.2 Posterior Probabilities for Predictive Regressions

To reinforce the case for time-varying parameters, we briefly depart from factor models and consider instead multivariate predictive regressions. The analysis of predictive regressions is motivated by taking expectations from both sides of equation (3) conditional on z_t and the underlying model parameters, as well as using the factor generating process in equation (2). In particular, expected returns are given by

$$
E\left[r_{t+1}|z_t,\theta\right] = \Big(\alpha_0 + \beta_0\alpha_f\Big) + \Big(\alpha_1 + \beta_0 a_F + \beta_1\left(\alpha_f \otimes I_M\right)\Big)z_t + \beta_1\left(a_F \otimes I_M\right)\left(z_t \otimes z_t\right). \tag{10}
$$

Thus, when both factor loadings and risk premiums are time varying, expected stock returns depend on macro predictors through their levels, squared values, and interactions.

Hence, a general formulation for predictive regressions should allow for nonlinear relations between future returns and predictors. To illustrate, focusing on the model that retains all M macro predictors, the predictive regression is formulated as

$$y_{t+1} = B_0 + B_1 z_t + B_2 \text{vech}\left(z_t z_t'\right) + \varepsilon_{t+1}, \tag{11}$$

where the operator $\text{vech}(z_t z_t')$ vectorizes the entries below and on the main diagonal of the matrix $z_t z_t'$ and is a $\frac{1}{2}(M+1) \times M$ vector that contains all interaction terms of the predictive variables, $B = [B_0, B_1, B_2]$ is a $(N+K) \times \left(1 + \frac{1}{2}M(M+3)\right)$ matrix of the regression intercept and slope coefficients, and ε_{t+1} is assumed to obey $\varepsilon_{t+1} \sim N(0, \Sigma)$.

The set of retained predictors is unique for every predictive regression. A candidate set of predictors is defined through a vector of binary variables of length M, indicating the inclusion or exclusion of a variable. The number of retained predictors is denoted by m, and it ranges between zero and M. In one extreme case, all predictors are excluded. Then, returns are independently and identically distributed (IID). The polar extreme is the all-inclusive specification that retains all M predictors.

We compute the marginal likelihoods for the collection of predictive regressions on the basis of three scenarios. The first excludes nonlinearities by setting $B_2 = 0$. The second allows for squared values and interactions. The third is a combination of the first two scenarios. We provide more details in the empirical section that follows. We can then assess the inclusion probability for both individual predictors and nonlinearities.

It is convenient to reformulate the data generating process in equation (11) using matrix notation

$$Y = XB + U, \tag{12}$$

where $X = [x_0, x_1, \ldots, x_{T-1}]'$, $Y = [y_1, y_2, \ldots, y_T]'$, $U = [\varepsilon_1, \ldots, \varepsilon_T]'$, and $x_t = [1, z_t']'$ if interaction terms are omitted or $x_t = [1, z_t', \text{vech}(z_t z_t')']'$ if interaction terms are included.

The marginal likelihood computation for the predictive regression is based on observing a hypothetical sample of length T_0 that is weighted against predictability. In that sample, the slope coefficients in the predictive regressions are centered around zero. The size of the hypothetical prior is set as a fixed number, i.e., 50, multiplied by the number of parameters underlying each of the predictive models, following Kandel and Stambaugh (1996).

We show in the Online Appendix A that the log marginal likelihood of any predictive regression is given by

$$\begin{aligned}
\ln\left[m(D|\mathcal{M})\right] &= -\frac{T(N+K)}{2}\ln(\pi) + \frac{T_0-m-1}{2}\ln|T_0\hat{V}_y| - \frac{T^*-m-1}{2}\ln|\tilde{S}| \\
&- \sum_{i=1}^{N+K}\ln\left\{\Gamma\left(\frac{T_0-m-i}{2}\right)\right\} + \sum_{i=1}^{N+K}\ln\left\{\Gamma\left(\frac{T^*-m-i}{2}\right)\right\} \\
&- \frac{(N+K)(m+1)}{2}\ln\left(\frac{T^*}{T_0}\right),
\end{aligned} \tag{13}$$

where

$$\tilde{S} = T^*\left(\hat{V}_y + \bar{y}\bar{y}'\right) - \frac{T}{T^*}\left(T_0\bar{y}\bar{x}' + Y'X\right)\left(X'X\right)^{-1}\left(T_0\bar{x}\bar{y}' + X'Y\right), \tag{14}$$

and where m is the number of retained predictors, $\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$, $T^* = T + T_0$, $\Gamma(\phi)$ stands for the Gamma function evaluated at ϕ, and $|A|$ is the determinant of matrix A. For the IID model, it follows that $\tilde{S}_{iid} = T^*\hat{V}_y$, where $\hat{V}_y$ is defined in equation (9).

We make two final notes for this subsection. First, the X matrix above is unique for every model, but we drop the model-specific subscript to ease notation. Second, while T_0 follows quite straightforwardly in a predictive regression setting, its formulation in an asset pricing context is more challenging. We provide details later in the text.

3.3 Posterior Probabilities for Factor Models

Returning to factor models, similar to Avramov and Chao (2006) and Barillas and Shanken (2018), all candidate models contain the market factor.[15] As noted earlier, the prior is based on a hypothetical sample of length T_0 with moments that are equal to the actual sample counterparts. In addition, the prior is weighted against both predictability by macro variables and model misspecification. Thus, on the basis of equation (3), regressing r_t on a constant term, z_{t-1}, f_t, and the interactions of $f_t \otimes z_{t-1}$ yields zero estimates for α_0, α_1, and β_1 in the prior sample. In other words, the prior densities of α_0, α_1, and β_1 are centered around zero.

We derive the marginal likelihood conditional on knowing T_0 and then provide an approach for computing T_0. In deriving the marginal likelihood, we adjust the collection of test assets for every individual model based on the included factors. Specifically, K denotes the maximal number of factors. When a candidate model $\mathcal{M}_l$ contains $k \leq K$ 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 only helpful if it prices the remaining assets correctly, including both test assets and traded factors. This specification also ensures that the marginal likelihoods for all models are conditioned on the same set of data.

Marginal likelihoods are first derived for models with time-varying parameters. Online Appendices B.1 and B.2 focus on unrestricted and restricted models, respectively. For the unrestricted case, the marginal likelihood is given by (again, we drop the model subscript

[15]Multifactor extensions have been inspired by Merton (1973) and Ross (1976). ICAPM factors should be correlated with the marginal utility of investors, while APT factors are extracted from the covariance matrix of asset returns. In the ICAPM, the market factor typically appears along with other state variables. In the APT, the first extracted factor typically has an almost unit correlation with the market portfolio (e.g., Geweke and Zhou (1996)).

to ease notation)

$$
\begin{aligned}
m(D|\mathcal{M}_C) = \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)(T_0+k)+\frac{1}{2}k(m+1)} \times \left[\frac{T}{T^*}\right]^{\frac{1}{2}(N+K-k)T} \times \\
\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^*-(k+1)m-1\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0-(k+1)m-1\right]\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^*+N+K-k-m-1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0+N+K-k-m-1\right]\right)}\right] \times \\
\left[\frac{|R'R-R'F\left(F'F\right)^{-1}F'R|^{\frac{1}{2}(T_0-(k+1)m-1)}}{|R'R-\tilde{\Phi}'W'W\tilde{\Phi}|^{\frac{1}{2}(T^*-(k+1)m-1)}}\right]\left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-m-1)}}{|S_F|^{\frac{1}{2}(T^*+N+K-k-m-1)}}\right],
\end{aligned}
\tag{15}
$$

where $\mathcal{M}_C$ stands for family of conditional models, $T^* = T + T_0$, $S_F = T^*\left(\hat{V}_f + \bar{f}\bar{f}'\right) - \frac{T}{T^*}\left(T_0[\bar{f}, \bar{f}\bar{z}'] + F'X\right)\left(X'X\right)^{-1}\left(T_0[\bar{f}, \bar{f}\bar{z}']' + X'F\right)$, $\hat{V}_y$, $\bar{f}$, and $\bar{z}$ are defined in equation (9), $W = [X, F, \Xi]$, and Ξ is the time-series collection of the vectorized matrix $\left(I_K \otimes z_t\right) f_{t+1}$ for all T periods. The last two terms in the marginal likelihood reflect the cross-sectional fit, while the remaining terms emerge from model complexity. As more factors or predictors are included, the model pricing abilities could improve. The potential improvement is associated with increasing complexity. Thus, the ultimate inclusion of a variable is subject to a rigorous trade-off.

The Online Appendix C derives the marginal likelihood for the case where unconditional models admit the possibility of mispricing (Online Appendix C.1) and when mispricing is excluded (Online Appendix C.2). The marginal likelihood for unconditional asset pricing models with mispricing is given by

$$
\begin{aligned}
m(D|\mathcal{M}_U) = \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)(T_0+k)+\frac{1}{2}k} \times \left[\frac{T}{T^*}\right]^{\frac{1}{2}(N+K-k)T} \times \\
\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^*-1\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0-1\right]\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^*+N+K-k-1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0+N+K-k-1\right]\right)}\right] \times \\
\left[\frac{|R'R-R'F\left(F'F\right)^{-1}F'R|^{\frac{1}{2}(T_0-1)}}{|R'R-\tilde{\Phi}'W'W\tilde{\Phi}|^{\frac{1}{2}(T^*-1)}}\right]\left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-1)}}{|T^*\hat{V}_f|^{\frac{1}{2}(T^*+N+K-k-1)}}\right],
\end{aligned}
\tag{16}
$$

where $\mathcal{M}_U$ stands for family of unconditional models.

Online Appendix D summarizes the marginal likelihoods for all the families of models. Notably, the marginal likelihood in the conditional case is invariant to a linear transformation of z_t in equation (3). The Online Appendix E provides a detailed proof.

3.4 Setting T_0

To complete the marginal likelihood derivation, it is essential to set T_0. To pursue that task, we link the variance of mispricing with the maximum admissible Sharpe ratio. First, as formulated in the Online Appendix F, the intercepts α_0 and α_1 in equation (3) have the joint prior distribution

$$\text{vec}\left([\alpha_0, \alpha_1]'\right) | \Sigma_{RR}, D \sim N\left(0, \Sigma_{RR} \otimes B_{11}\right), \tag{17}$$

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

$$B_{11} = \begin{pmatrix} 1 + \bar{z}'\hat{V}_z^{-1}\bar{z} + \bar{f}'\hat{V}_f^{-1}\bar{f} + \bar{z}'\hat{V}_z^{-1}\bar{z} \times \bar{f}'\hat{V}_f^{-1}\bar{f} & -\bar{z}'\hat{V}_z^{-1} \times (1 + \bar{f}'\hat{V}_f^{-1}\bar{f}) \\ -\hat{V}_z^{-1}\bar{z} \times (1 + \bar{f}'\hat{V}_f^{-1}\bar{f}) & \hat{V}_z^{-1} \times (1 + \bar{f}'\hat{V}_f^{-1}\bar{f}) \end{pmatrix}. \tag{18}$$

The unconditional variance of total mispricing is then equal to

$$\text{Var}\left(\alpha | \Sigma_{RR}, D\right) = \text{Var}\left(\alpha_0 + \alpha_1' z | \Sigma_{RR}, D\right) = \frac{\Sigma_{RR}}{T_0}\left(1 + SR^2_{max} + m(1 + SR^2_{max})\right), \tag{19}$$

where SR^2_{max} is the largest attainable Sharpe ratio based on investments in the benchmarks only and m is the number of predictors that 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 as

$$\alpha | \Sigma_{RR}, D \sim N\left(0, \eta\Sigma_{RR}\right), \tag{20}$$

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 et al. (1989) associate $\hat{\alpha}'\hat{\Sigma}_{RR}^{-1}\hat{\alpha}$ with the difference between two squared Sharpe ratios, i.e.,

$$\hat{\alpha}'\hat{\Sigma}_{RR}^{-1}\hat{\alpha} = \widehat{SR}^2(R,F) - \widehat{SR}^2(F), \tag{21}$$

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. By contrast, the second term $\widehat{SR}(F)$ varies across the models, while $\widehat{SR}(F)$ attains its minimum value for the CAPM and maximum when all K factors are retained. In the spirit of Barillas and Shanken (2018) and Chib et al. (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, i.e., $SR_{max} = SR(R,F) = \tau SR(Mkt)$, 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

$$E(\alpha'\Sigma_{RR}^{-1}\alpha|\Sigma_{RR}, D) = \eta\,(N+K-k) = \left(\tau^2 - 1\right) SR^2(Mkt). \tag{22}$$

The parameter η is then given by

$$\eta = \frac{(\tau^2-1)\,SR^2(Mkt)}{(N+K-k)}. \tag{23}$$

Finally, equating the variance of α in the hypothetical sample (equation (19)) with the prior variance in equation (20) and using η from equation (23), we obtain

$$T_0 = \frac{(N+K-k)\,(1+SR^2_{max}+m(1+SR^2_{max}))}{(\tau^2-1)\,SR^2(Mkt)}. \tag{24}$$

By setting T_0, the size of the hypothetical sample, we conclude the prior derivation.

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

Our prior specification for alpha resembles that of Pástor and Stambaugh (2000), namely, $\alpha|\Sigma \sim \mathrm{N}\left(0, \sigma^2_\alpha\left(\frac{1}{s^2}\Sigma_{RR}\right)\right)$, where σ^2_α 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 $\alpha'\Sigma^{-1}_{RR}\alpha$ could grow with the addition of more tests assets, while we bound that expression. The following relation is useful to map σ^2_α, the prior confidence in model pricing abilities, into the length of the hypothetical sample

$$T_0 = \frac{s^2}{\sigma^2_\alpha}\left(1+SR^2_{max}+m\left(1+SR^2_{max}\right)\right). \tag{25}$$

Derivations of T_0 for other asset pricing models, described in Table 1, are in Online Appendix F.

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

3.5 Incremental Remarks on the Methodology

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

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 substantially departs from normality. For one, integrating out the parameter space, the distribution of stock returns becomes Student's-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. The first is to draw a factor model by generating a uniform random variable to select a model based on cumulative model posterior probabilities. Second, conditioned on the model, underlying parameters are drawn from the joint normal-inverted-Wishart densities. Third, conditioned on the parameters, returns are drawn from a normal distribution. While the predictive distribution can be simulated by repeating these three steps, we have still been successful in deriving analytic expressions for the vector of mean returns and the covariance matrix. Model-specific first two moments are derived in Online Appendix G. Moments for the integrated model follow through equations (4), (5), and (6). We acknowledge that our findings are associated with conditional normality, while departing from that assumption could establish avenues for future research.

Third, while we develop a prior in the context of cross-sectional asset pricing, it would be useful, for completeness, to describe informed priors inspired by time-series econometrics. To start, Kandel and Stambaugh (1996) center the prior on the predictive regression R-squared 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 negative correlation between the innovations in predictive regressions and expected returns to maintain mean reversion; Avramov et al. (2017), who propose taking cues from various consumption-based models for understanding the riskiness of equities over the long run; and Giannone et al. (2015), who focus on coefficients in vector autoregressions.[16] 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.

4 Data

We focus on 14 representative asset pricing factors that are prominent in the asset pricing literature. We begin with the Fama-French five-factor model (Fama and French (2015))

[16]Note that time-series goodness of fit, even in asset pricing regressions, does not translate into cross-sectional pricing abilities, a point made by Chen et al. (1986). Thus, in our setup, it is the Sharpe ratio that plays a dominant role in formulating the prior, while the abovementioned papers focus on either R-squared or autocorrelation.

that consists of the market (MKT), size (SMB), value (HML), profitability (RMW), and investment (CMA), and augment it with momentum (MOM, from Carhart (1997) and Fama and French (2018)). We also include two behavioral factors, i.e., post-earnings announcement drift (PEAD) and financing (FIN) from Daniel et al. (2020). The additional factors include quality-minus-junk (QMJ, from Asness et al. (2019)), betting-against-beta (BAB, from Frazzini and Pedersen (2014)), mispricing factors related to management (MGMT) and performance (PERF) from Stambaugh and Yuan (2017), liquidity (LIQ, from Pástor and Stambaugh (2003)), and intermediary capital (ICR, from He et al. (2017)).[17]

We follow Welch and Goyal (2008) and employ 13 macro predictors, including the dividend price ratio (*dp*), the dividend yield (*dy*), the earnings price ratio (*ep*), the dividend payout ratio (*de*), the stock variance (*svar*), the book-to-market ratio (*bm*), the net equity expansion (*ntis*), the yield on Treasury bills (*tbl*), the long-term yield (*lty*), the long-term rate of returns (*ltr*), the term spread (*tms*), the default yield spread (*dfy*), and inflation (*infl*). Table 2 provides the detailed definitions of 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. In Table 3, Panel A reports the means, medians, and standard deviations 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.

[17] We consider the tradable version of the liquidity (LIQ) and intermediary capital (ICR) factors to facilitate model interpretation and comparison. In our setup, alpha indicates model mispricing only if the factors are tradable.

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.

Panel B reports the means, medians, standard deviations, and AR(1) coefficients of the monthly macro predictors. Most macro predictors are highly persistent with AR(1) coefficients above 0.94, except for *svar*, *ltr*, and *infl*. All AR(1) coefficients are less than one, indicating a slow mean reversion.

5 Probability Analysis

5.1 Predictive Regressions

We start by applying the BMA procedure to multivariate predictive regressions, per Section 3.2. We consider three scenarios based on equation (11): (i) include only macro predictors, i.e., $B_1 \neq 0$ and $B_2 = 0$, (ii) include macro predictors with interactions between predictors, i.e., $B_1 \neq 0$ and $B_2 \neq 0$, and (iii) include macro predictors with and without interactions. Note that 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 model universe to 5 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). Hence, we consider only five models. The remaining seven predictors contribute 2^7 combinations. Thus, the model space consists of 3,200 ($= 25 \times 2^7$) combinations for the first two scenarios. The third scenario consists of 6,399 ($= 2 \times 25 \times 2^7 - 1$) combinations, which obtains as the union of the first two scenarios while excluding the overlapping specification of an intercept-only model.

Posterior probability is assigned to each candidate model. The BMA routine then allows us to evaluate the relative importance of every individual predictor by its cumulative inclusion probability, computed as $\mathcal{A}'\mathcal{P}$, where 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.

Table 4 presents the cumulative posterior probabilities for the macro predictors in predictive regressions. For the case of no interaction, the inclusion probability is approximately 100% for the dividend yield (dy), followed by the 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%. On the one hand, the inclusion of two stock characteristics, namely, dy and $svar$, is strongly supported by the data regardless of the model specification. On the other hand, the cumulative posterior probabilities significantly drop for the book-to-market ratio (bm), the Treasury bill yield (tbl), ltr, and the term spread (tms). Moreover, the probability of including interactions ($B_2 \neq 0$) is unity. That is, future returns depend on levels, squared values, and interactions between pairs of macro predictors. Overall, the stylized findings from predictive regressions motivate us to consider time-varying factor loadings and factor risk premia in the subsequent assessment of factor models.

5.2 Factor Models

We apply the BMA procedure to conditional and unconditional asset pricing models, per Section 3.3. Our model space includes 14 asset pricing factors and 13 macro predictors. Panel A of Table 1 lists the 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 for the single combination when 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 single 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 combinations for both $\mathbb{M}_1$ and $\mathbb{M}_2$. For conditional models specified through $\mathbb{M}_3$ and $\mathbb{M}_4$, each includes $2^{13} \times (25 \times 2^7 - 1)$ combinations for inclusion/exclusion of the factors and predictors.[18] Collectively, the integrated model accommodates a collection of over 52 million candidate models.[19]

We start by examining the 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 the most. In contrast, if a large number of candidate models have meaningful probabilities, accounting for model uncertainty is essential and rationalized from the Bayes rule. We first compute the posterior probability for each candidate model and 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 different prior Sharpe multiples, i.e., $\tau = 1.25$, 1.5, 2, and 3. We follow

[18] As previously discussed, there are 2^{13} unconditional combinations. In addition, as noted in Section 5.1, some macro predictors are jointly redundant, resulting in 25×2^7 predictor combinations. We further exclude the single combination including no predictor, i.e., the unconditional models specified in $\mathbb{M}_1$ and $\mathbb{M}_2$, resulting in $25 \times 2^7 - 1$ predictor combinations.

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

Barillas and Shanken (2018) to consider $\tau = 1.5$ as the 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 whole 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. Thus, evidence suggests that multiple distinct models could govern the joint dynamic 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

$$P\left(k \text{ included}|D\right) = \sum_{l=1}^{L} P\left(\mathcal{M}_l|D\right) \mathbb{1}_{\{k \text{ included in } \mathcal{M}_l\}}. \tag{26}$$

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

$$P\left(m \text{ included}|D\right) = \sum_{l=1}^{L} P\left(\mathcal{M}_l|D\right) \mathbb{1}_{\{m \text{ included in } \mathcal{M}_l\}}. \tag{27}$$

The results are reported in Table 5. 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 $\tau = 1.5$. We find that post-earnings announcement drift (PEAD) and quality-minus-junk (QMJ) display a posterior inclusion probability of close to 100%, followed by investment (CMA), size (SMB), intermediary capital (ICR), and management (MGMT)—which all 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 a high inclusion probability. On the other hand, five out 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 competence in pricing 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, betting-against-beta (BAB) exhibits high inclusion probability only for $\tau = 3$, when the prior is tilted toward rather extreme Sharpe ratios. It would thus be challenging for the BAB factor to clear prior asset pricing thresholds, such as reasonable Sharpe ratios (Ross (1976)). Collectively, 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, there are five to seven factors with 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 probabilities originate from distinct economic foundations rather than an established asset pricing model. For instance, PEAD, QMJ, and ICR are proposed by three independent works, and this combined specification has not been examined in the previous literature.

Finally, profitability (RMW) appears redundant. This could be due to the high correlation between RMW and QMJ (0.75 from Panel A of Table 3), as profitability is also one of the quality characteristics in QMJ.[20] Empirically, QMJ dominates RMW in pricing other factors; hence, we observe persistent inclusion for QMJ and exclusion for RMW.

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

Panel B of Table 5 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 $\tau = 1.5$ 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 the net equity expansion ($ntis$), dy, the yield 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 an important clue that 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), the long-term rate of return (ltr), the default yield spread (dfy), and inflation ($infl$) are always discarded, regardless of the prior. Evidence from Panel B thus reinforces the notion 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 5. 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 procedure uniformly favors models with time-varying parameters, even when prior beliefs are weighted against the inclusion of macro predictors.[21] Our findings further highlight the importance of incorporating nonlinearities in asset pricing models, especially by conditioning on the macroeconomic states—a point also emphasized by Chen et al. (2021) in a nonparametric setup. Furthermore, our results complement prior

[21]By construction, the sum of the posterior probabilities for all models included in $\mathbb{M}_1$ to $\mathbb{M}_4$ equals one. Our findings indicate that the aggregate posterior probability of the unconditional models included in $\mathbb{M}_1$ and $\mathbb{M}_2$ is virtually zero.

work that focuses on the nonlinear relationship between firm characteristics and returns (e.g., Freyberger et al. (2020)) and that employs a conditional factor model where the factor loadings are nonlinear in firm characteristics (e.g., Gu et al. (2021)).

Another essential feature in our 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 indicates that zero-alpha models selected from the collection 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 premiums, 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 (24), 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.[22] Intuitively, when the prior Sharpe ratio multiple is low, more shrinkage is applied to penalize mispricing and time-varying factor risk premiums. 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).

Collectively, we show that a plethora of models that differ in the inclusion of factors, predictors, and mispricing record a positive and meaningful probability of governing the

[22]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 (B.23) and (B.24) for unrestricted models and equation (B.39) for restricted models. Therefore, higher T_0 implies more shrinkage toward the hypothetical sample, i.e., the model estimates are weighted against mispricing.

joint dynamics of stock returns. Meanwhile, our BMA approach provides a reasonable setup for analyzing the cross section of expected stock returns.

6 Out-of-Sample BMA Model Performance

6.1 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, i.e., only adjusting for the market factor (MKT), (ii) the Fama-French 3-factor model (FF3) consisting of the market factor (MKT), the size factor (SMB), and the value factor (HML) (Fama and French (1993)), (iii) the Fama-French 6-factor model (FF6) consisting of the market factor (MKT), the size factor (SMB), the value factor (HML), the profitability factor (RMW), the investment factor (CMA), and the momentum factor (MOM) (Fama and French (2018)), and (iv) the AQR 6-factor model (AQR6) consisting of the market factor (MKT), the size factor (SMB), the value factor (HML), the momentum factor (MOM), the betting-against-beta factor (BAB), and the quality-minus-junk factor (QMJ) (Frazzini et al. (2018)). We further consider the three highest posterior probability models based on the Bayesian procedure. Our prior is that the Bayesian approach could deliver stable out-of-sample performance, given (i) its empirical merits in identifying single competent models and (ii) its conceptual foundation from the Bayes rule.

Our first experiment examines the Sharpe ratio of the tangency portfolio. We divide the full sample into two periods, i.e., 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 $\frac{T}{2}$) and two-thirds of the sample (denoted $\frac{2T}{3}$).[23] For 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 computing model-specific predictive moments is provided in the Online Appendix G.[24]

We tabulate the in-sample and out-of-sample annualized Sharpe ratio in Table 6, with Panel A for the four benchmark models and Panel B for models based on the Bayesian procedure 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, consider $\frac{2T}{3}$ as the in-sample period. First, the integrated model (denoted BMA) 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. In addition, the integrated model continues to deliver superior out-of-sample performance, with an annualized Sharpe ratio of 1.240, which offers a 8% improvement compared to the best benchmark model AQR6, which has an annualized Sharpe ratio of 1.152.[25]

[23]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.

[24]The derivation builds on 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 (G.1) and (G.2) for factors and equations (G.3) and (G.4) for test assets. Moments for the integrated model follow through equations (4), (5), and (6).

[25]While we focus on four observable factor models as benchmarks, untabulated analyses further consider the tangency portfolio based on the unconditional model with 14 factors. Taking $\frac{2T}{3}$ as the in-sample period, the integrated model continues to outperform the unconditional model (annualized Sharpe ratio at 1.019) by 22%. In addition, we examine the equal-weighted portfolio with 14 factors in the full

Second, the three top-ranked individual models (denoted Top 1, Top 2, and Top 3), namely, the three highest posterior probability models, 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 variations in their out-of-sample performance. For instance, the second-ranked (i.e., Top 2) model turns out to be the best performing and it 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 out 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 set the sum of absolute tangency portfolio weights to be smaller than or equal to 2, i.e., $\sum_{i=1}^{14} |w_i| \leq 2$. As expected, the Regulation T constraint reduces the Sharpe ratio for nearly all models both in sample and out of sample.

Taking $\frac{2T}{3}$ as the in-sample period for an example, first, 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

sample. It delivers 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 2.214 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 inherent positive alphas of the asset pricing factors in the full sample.

applying sensible economic restrictions. Thus, when efficient portfolios are admissible, the performance gap between the integrated model and benchmark models widens. Second, we continue to find superior performance among top-ranked individual models, and they outperform AQR6 by 29% to 82%. Collectively, 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).[26] Importantly, all models based on the Bayesian procedure provide 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.[27] We report the results in Panels E and F of Table 6, with Panel E for the benchmark models and Panel F for

[26]There are 14 factors and 11 predictors in total because some macro predictors are jointly redundant, as previously described. The total number of estimated parameters is given by $(K-k)\left[(1+m)(1+k)+\frac{K-k+1}{2}\right]+k\left(1+m+\frac{k+1}{2}\right)$, where K stands for the maximal number of factors (14), and k and m stand for 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 $k=7$).

[27]Garlappi et al. (2007) document that in the presence of a stable and significant degree of ambiguity aversion, the GMVP could play an important role in the 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.

models based on the Bayesian procedure.

Taking $\frac{2T}{3}$ as the in-sample period for an example, 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 model delivers a 35% higher Sharpe ratio than the best benchmark model (FF6, annualized Sharpe ratio at 0.818) and 19% higher Sharpe ratio than the best individual model (Top 1, annualized Sharpe ratio at 0.924).[28] 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.

6.2 Efficient Portfolios: Downside Risk

Beyond the out-of-sample Sharpe ratio, it is worth evaluating other potential risks, especially the downside risk in trading the tangency portfolio.[29] 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 for the tangency portfolio.[30] We follow Gu et al. (2020) to define the maximum drawdown throughout the entire sample as

$$MDD = \max_{0 \leq t_1 \leq t_2 \leq T} \left(Y_{t_1} - Y_{t_2} \right), \tag{28}$$

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 7, where Panels A and B show the results for tangency portfolios constructed from benchmark models and models based on the Bayesian

[28]While the primary purpose of examining GMVP is to understand how model uncertainty affects portfolio risk, it could still have a meaningful impact on the Sharpe ratio, depending on the risk-return trade-off. We will shed more light on the risk implications later.

[29]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.

[30]For perspective, skewness and excess kurtosis are equal to zero under a normal distribution.

procedure 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/similar return volatility. Second, the integrated model exhibits less negative skewness, lower excess kurtosis, and a lower maximum drawdown. For instance, the maximum drawdown for the integrated model is 44%, while the benchmark models experience a larger drawdown between 51% and 79%. Notably, the top-ranked individual models are more volatile and the maximum drawdown varies in a wide range between 32% and 57%.

Panels E and F of Table 7 report similar statistics for GMVPs, with Panel E showing results for GMVPs constructed from benchmark models and Panel F for those constructed from models based on the Bayesian procedure with $\tau = 1.5$. Notably, risk metrics are particularly relevant in the context of GMVP, as GMVP relies exclusively on the covariance matrix and if model uncertainty plays a significant role in asset pricing, we expect the GMVP based on the integrated model to be less risky. Several findings are noteworthy. First, we find that the GMVP based on the integrated model is considerably less volatile than the benchmark models. For instance, monthly realized volatility for GMVPs based on the benchmark models ranges between 0.956% and 2.127%, while appears to be only 0.756% for the integrated model, indicating a 21% to 64% volatility reduction. Because expected returns across the various 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. Collectively, the BMA approach mitigates the downside risk of both the tangency portfolio and the GMVP.

6.3 Additional Analyses

To obtain a complete outlook of integrated model performance over time, Figure 2a plots the cumulative excess returns of an initial investment of $1 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. While the ultimate investment outcome appears similar for the market and BMA portfolios, the BMA portfolios experience a more stable increase over time. Their performance is much less volatile than the market portfolio. Importantly, we observe only modest declines for BMA portfolios when the overall market often 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.

Figure 2b plots similar cumulative excess returns but focuses on the out-of-sample periods: one starts from January 1998 ($\frac{T}{2}$) and another starts from January 2004 ($\frac{2T}{3}$). We further consider BMA tangency portfolios with and without the Regulation T constraint and the GMVP. Taking $\frac{2T}{3}$ as the in-sample period for an example, the BMA tangency portfolio with the Regulation T constraint consistently outperforms its unconstrained counterpart and the market portfolio over time. We also confirm that a longer in-sample period helps improve the performance of the Bayesian approach.

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 8 has a similar layout to Table 6, with Panel A for the unconstrained tangency portfolio, Panel B for the tangency portfolio with the Regulation T constraint, and Panel C for the GMVP. Taking $\frac{2T}{3}$ as the in-sample period for an example, the integrated model continues to outperform the best benchmark model from Table 6 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 in Sharpe ratio ranges from 17% to 37%. This more decisive evidence on the GMVP 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 out-of-sample performance in general, we observe considerable variations 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 out of 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 through BMA achieves a rather stable, superior, and admissible out-of-sample Sharpe ratio and mitigates the downside risk of the investment. Our findings are robust to imposing economic restrictions on the prior Sharpe ratio and stock positions as well as using alternative in-sample periods. The analyses of the GMVP further highlight the impact of model uncertainty on the covariance matrix of stock returns. In addition, the Bayesian approach is also 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.

7 Dissecting Model Uncertainty

7.1 Variance Decomposition

We provide additional evidence to highlight the importance of model uncertainty in shaping the investment opportunity set. Our first experiment compares the sample variance of factor returns with the variance based on the integrated model. In particular, by variance decomposition, we have

$$\mathrm{Var}\left(r_{t+1}\right) = E\left[\mathrm{Var}\left(r_{t+1}|z_t\right)\right] + \mathrm{Var}\left[E\left(r_{t+1}|z_t\right)\right], \tag{29}$$

where $\mathrm{Var}\left(r_{t+1}\right)$ is the unconditional variance and $E\left[\mathrm{Var}\left(r_{t+1}|z_t\right)\right]$ 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 the population, the inequality $\mathrm{Var}\left(r_{t+1}\right) > E\left[\mathrm{Var}\left(r_{t+1}|z_t\right)\right]$ is binding. However, $\mathrm{Var}\left(r_{t+1}|z_t\right)$ does not incorporate model disagreement and the mixture of estimation risk, emphasized by our approach. Thus, the variance perceived by a Bayesian investor who is perceptive of model uncertainty is higher than $\mathrm{Var}\left(r_{t+1}|z_t\right)$.

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, i.e., $\mathrm{Var}\,[r_{t+1}|D]$ 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) and half ($\frac{T}{2}$) and two-thirds ($\frac{2T}{3}$) of the sample under a prior Sharpe multiple of 1.5, and compute the in-sample and out-of-sample variance for each factor.

Table 9 presents the results, with the columns "EST" and "OOS" corresponding 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 $\bar{V}_t + \bar{\Omega}_t$) than the sample variance (denoted OBS). Using $\frac{2T}{3}$ of the sample as 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 14 factors. The integrated model variance is also more than doubled of the sample variance for the profitability (RMW), investment (CMA), financing (FIN), management (MGMT), and betting-against-beta (BAB) factors.

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

7.2 Time-Varying Model Disagreement

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 weighted average 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 a standard measure of information in information theory, i.e., entropy. 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, and 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 and uncertainty reduction.

Similar to the entropy reduction due to additional information, we can view the Ω_t component as an entropy extension arising from model disagreement. In other words, we measure the contribution of the model disagreement component to the covariance matrix as an increase in entropy relative to the V_t component (the weighted average of model-implied covariance),

$$EI_t = \frac{|V_t + \Omega_t|}{|V_t|}. \tag{30}$$

We compute the relative increase in entropy for the three in-sample periods that correspond to the full sample (T), half of the sample ($\frac{T}{2}$), and two-thirds of the sample ($\frac{2T}{3}$) under a prior Sharpe multiple of 1.5. Panel A of Table 10 reports the mean, the 95^{th} percentile, the 99^{th} percentile, and the 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, i.e., 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 3a 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, e.g., Black Monday in October 1987 and the recent financial crisis starting in September 2008. Our findings support the notion that asset pricing models significantly disagree about expected stock returns at times of crash events in the financial market, which makes stocks appear riskier. Hence, accounting for model uncertainty could hedge against extremely negative market conditions, 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 every single factor to the entropy increase. In this experiment, we zero out the off-diagonal elements of Ω_t in equation (6) for simplicity.[31] Let $\Omega_{i,t}$ be a matrix with only the i^{th} diagonal element being equal to the corresponding diagonal element of Ω_t and with other elements equal to zero, where $i \in 1, 2, ..., K$, and K refers to the maximal number of factors. Similar to the definition in equation (30), we define

[31]In all previous analyses, Ω_t is a standard symmetric covariance matrix that captures the correlations between factors. Unreported results employing a diagonal Ω_t also confirm our main findings.

the entropy increase attributed to factor i at time t as

$$EI_{i,t} = \frac{|V_t + \Omega_{i,t}|}{|V_t|}. \tag{31}$$

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

$$REI_{i,t} = \frac{\frac{log(EI_{i,t})}{log(EI_t)}}{\sum_{j=1}^{K} \frac{log(EI_{j,t})}{log(EI_t)}}. \tag{32}$$

We 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 ($\frac{2T}{3}$) 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 tabulated in Table 10, with Panels B and C for the average and maximum factor contributions, respectively. As shown in Panel B, the liquidity (LIQ) factor stands out, as it contributes to 21% of the total entropy increase in the full sample, followed by the size (SMB) and betting-against-beta (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, management (MGMT), and intermediary capital (ICR) factors carry a sizable disagreement component and jointly contribute to 40% of the overall entropy increase.

Since the model disagreement component in the overall covariance matrix can be low at 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, i.e., 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 diminishes 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 et al. (2017)). Collectively, 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 further confirm this finding in Figure 3b, where for each factor, we plot the contribution to the overall entropy increase over time using the aforementioned three in-sample periods.

Overall, we find that asset pricing models significantly disagree about expected stock returns during market crashes. Hence, accounting for model uncertainty effectively hedges against downside risk and enhances out-of-sample performance. In addition, the market, management, and intermediary capital factors stand out in explaining the time-varying model disagreement component.

8 Conclusion

This paper develops a comprehensive Bayesian framework to study the cross section of average returns and the covariance matrix in the presence of model uncertainty. The framework combines a large universe of candidate asset pricing models into an integrated model based on model probabilities. In addition, prior beliefs about the entire parameter space are economically interpretable and weighted against deviations from the unconditional CAPM. The integrated model is used to assess the strength of asset pricing factors and macro predictors in explaining the joint distribution 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 collection of factors and predictors while some factor models hold exactly and others admit mispricing. We first document that there is a fairly large number of positive probability models. Hence, integrating factor models follows directly from the Bayes rule. We further show that the underlying return generating process exhibits considerable mispricing and is uniformly dominated by models with time varying parameters. Then, the post-earnings announcement drift, quality-minus-junk, and intermediary capital factors are competent in pricing other factors beyond the market factor. From an investment perspective, the integrated model delivers a stable, superior, and admissible out-of-sample Sharpe ratio and mitigates the downside risk for both the tangency portfolio and the GMVP. The Bayesian approach is also instrumental in identifying competent individual models, while model selection based solely on top-ranked individual models could provide unstable forecasts. Finally, asset pricing models significantly disagree about expected stock returns at times of market crash events, while the spikes in model disagreement about expected returns are primarily driven by market, management, and intermediary capital factors.

References

Anderson, E. W. and A.-R. M. Cheng (2016). Robust Bayesian Portfolio Choices. *The Review of Financial Studies 29*(5), 1330–1375.

Arnott, R., C. R. Harvey, V. Kalesnik, and J. Linnainmaa (2019). Alice's Adventures in Factorland: Three Blunders That Plague Factor Investing. *The Journal of Portfolio Management 45*(4), 18–36.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2019). Quality Minus Junk. *Review of Accounting Studies 24*(1), 34–112.

Avramov, D. (2002). Stock Return Predictability and Model Uncertainty. *Journal of Financial Economics 64*(3), 423–458.

Avramov, D. (2004). Stock Return Predictability and Asset Pricing Models. *The Review of Financial Studies 17*(3), 699–738.

Avramov, D., S. Cederburg, and K. Lučivjanská (2017). Are Stocks Riskier over the Long Run? Taking Cues from Economic Theory. *The Review of Financial Studies 31*(2), 556–594.

Avramov, D. and J. C. Chao (2006). An Exact Bayes Test of Asset Pricing Models with Application to International Markets. *The Journal of Business 79*(1), 293–324.

Avramov, D. and T. Chordia (2006). Predicting Stock Returns. *Journal of Financial Economics 82*(2), 387–415.

Avramov, D. and R. Wermers (2006). Investing in Mutual Funds When Returns Are Predictable. *Journal of Financial Economics 81*(2), 339–377.

Baker, M. and J. Wurgler (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance 61*(4), 1645–1680.

Ball, R. (1978). Anomalies in Relationships between Securities' Yields and Yield-Surrogates. *Journal of Financial Economics 6*(2-3), 103–126.

Barillas, F. and J. Shanken (2017). Which Alpha? *The Review of Financial Studies 30*(4), 1316–1338.

Barillas, F. and J. A. Shanken (2018). Comparing Asset Pricing Models. *The Journal of Finance 73*(2), 715–754.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2013). Time-Varying Combinations of Predictive Densities Using Nonlinear Filtering. *Journal of Econometrics 177*(2), 213–232.

Bossaerts, P. and P. Hillion (1999). Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn? *The Review of Financial Studies 12*(2), 405–428.

Bryzgalova, S., J. Huang, and C. Julliard (2020). Bayesian Solutions for the Factor Zoo: We Just Ran Two Quadrillion Models. *Working Paper*.

Campbell, J. Y. (1987). Stock Returns and the Term Structure. *Journal of Financial Economics 18*(2), 373–399.

Campbell, J. Y., J. Hilscher, and J. A. Szilagyi (2008). In Search of Distress Risk. *The Journal of Finance 63*(6), 2899–2939.

Campbell, J. Y., A. W.-C. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Campbell, J. Y. and R. J. Shiller (1988). Stock Prices, Earnings, and Expected Dividends. *The Journal of Finance 43*(3), 661–676.

Campbell, J. Y. and T. Vuolteenaho (2004). Inflation Illusion and Stock Prices. *American Economic Review 94*(2), 19–23.

Campbell, J. Y. and M. Yogo (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics 81*(1), 27–60.

Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance 52*(1), 57–82.

Chen, L., M. Pelger, and J. Zhu (2021). Deep Learning in Asset Pricing. *Working Paper*.

Chen, N.-F., R. Roll, and S. A. Ross (1986). Economic Forces and the Stock Market. *The Journal of Business 59*(3), 383–403.

Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association 90*(432), 1313–1321.

Chib, S. and X. Zeng (2019). Which Factors are Risk Factors in Asset Pricing? A Model Scan Framework. *Journal of Business & Economic Statistics 38*(4), 771–783.

Chib, S., X. Zeng, and L. Zhao (2020). On Comparing Asset Pricing Models. *The Journal of Finance 75*(1), 551–577.

Chib, S., L. Zhao, D. Huang, and G. Zhou (2019). Winners from Winners: A Tale of Risk Factors. *Working Paper*.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton University Press.

Cochrane, J. H. (2011). Presidential Address: Discount Rates. *The Journal of Finance 66*(4), 1047–1108.

Cong, L., K. Tang, J. Wang, and Y. Zhang (2021). AlphaPortfolio for Investment and Economically Interpretable AI. *Working Paper*.

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short- and Long-Horizon Behavioral Factors. *The Review of Financial Studies 33*(4), 1673–1736.

Daniel, K. and T. J. Moskowitz (2016). Momentum Crashes. *Journal of Financial Economics 122*(2), 221–247.

DeMiguel, V., A. Martín-Utrera, F. J. Nogales, and R. Uppal (2020). A Transaction-Cost Perspective on the Multitude of Firm Characteristics. *The Review of Financial Studies 33*(5), 2180–2222.

Fama, E. F. and K. R. French (1989). Business Conditions and Expected Returns on Stocks and Bonds. *Journal of Financial Economics 25*(1), 23–49.

Fama, E. F. and K. R. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics 33*(1), 3–56.

Fama, E. F. and K. R. French (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and K. R. French (2018). Choosing Factors. *Journal of Financial Economics 128*(2), 234–252.

Feng, G., S. Giglio, and D. Xiu (2020). Taming the Factor Zoo: A Test of New Factors. *The Journal of Finance 75*(3), 1327–1370.

Ferson, W. E. and C. R. Harvey (1999). Conditioning Variables and the Cross Section of Stock Returns. *The Journal of Finance 54*(4), 1325–1360.

Frazzini, A., D. Kabiller, and L. H. Pedersen (2018). Buffett's Alpha. *Financial Analysts Journal 74*(4), 35–55.

Frazzini, A. and L. H. Pedersen (2014). Betting Against Beta. *Journal of Financial Economics 111*(1), 1–25.

Freyberger, J., A. Neuhierl, and M. Weber (2020). Dissecting Characteristics Nonparametrically. *The Review of Financial Studies 33*(5), 2326–2377.

Garlappi, L., R. Uppal, and T. Wang (2007). Portfolio Selection with Parameter and Model Uncertainty: A Multi-Prior Approach. *The Review of Financial Studies 20*(1), 41–81.

Geweke, J. and G. Amisano (2011). Optimal Prediction Pools. *Journal of Econometrics 164*(1), 130–141.

Geweke, J. and G. Zhou (1996). Measuring the Pricing Error of the Arbitrage Pricing Theory. *The Review of Financial Studies 9*(2), 557–587.

Giannone, D., M. Lenza, and G. E. Primiceri (2015). Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics 97*(2), 436–451.

Gibbons, M. R., S. A. Ross, and J. A. Shanken (1989). A Test of the Efficiency of a Given Portfolio. *Econometrica*, 1121–1152.

Green, J., J. Hand, and X. Zhang (2017a). The Remarkable Multidimensionality in the Cross-Section of Expected U.S. Stock Returns. *Working Paper*.

Green, J., J. R. M. Hand, and X. F. Zhang (2017b). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies 30*(12), 4389–4436.

Gu, S., B. Kelly, and D. Xiu (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies 33*(5), 2223–2273.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder Asset Pricing Models. *Journal of Econometrics 222*(1, Part B), 429–450.

Guo, H. (2006). On the Out-of-Sample Predictability of Stock Market Returns. *The Journal of Business 79*(2), 645–670.

Harvey, C. R. (2017). Presidential Address: The Scientific Outlook in Financial Economics. *The Journal of Finance 72*(4), 1399–1440.

Harvey, C. R., Y. Liu, and H. Zhu (2016). . . . and the Cross-Section of Expected Returns. *The Review of Financial Studies 29*(1), 5–68.

Harvey, C. R. and G. Zhou (1990). Bayesian Inference in Asset Pricing Tests. *Journal of Financial Economics 26*(2), 221–254.

He, Z., B. Kelly, and A. Manela (2017). Intermediary Asset Pricing: New Evidence from Many Asset Classes. *Journal of Financial Economics 126*(1), 1–35.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science 14*(4), 382–417.

Hou, K., H. Mo, C. Xue, and L. Zhang (2021). An Augmented q-Factor Model with Expected Growth. *Review of Finance 25*(1), 1–41.

Hou, K., C. Xue, and L. Zhang (2015). Digesting Anomalies: An Investment Approach. *The Review of Financial Studies 28*(3), 650–705.

Hou, K., C. Xue, and L. Zhang (2020). Replicating Anomalies. *The Review of Financial Studies 33*(5), 2019–2133.

Jacobs, B. I., K. N. Levy, and D. Starer (1999). Long-Short Portfolio Management: An Integrated Approach. *The Journal of Portfolio Management 25*(2), 23–32.

Kandel, S. and R. F. Stambaugh (1995). Portfolio Inefficiency and the Cross-Section of Expected Returns. *The Journal of Finance 50*(1), 157–184.

Kandel, S. and R. F. Stambaugh (1996). On the Predictability of Stock Returns: An Asset-Allocation Perspective. *The Journal of Finance 51*(2), 385–424.

Kass, R. E. and A. E. Raftery (1995). Bayes Factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics Are Covariances: A Unified Model of Risk and Return. *Journal of Financial Economics 134*(3), 501–524.

Kothari, S. P. and J. A. Shanken (1997). Book-to-Market, Dividend Yield, and Expected Market Returns: A Time-Series Analysis. *Journal of Financial Economics 44*(2), 169–203.

Kozak, S., S. Nagel, and S. Santosh (2018). Interpreting Factor Models. *The Journal of Finance 73*(3), 1183–1223.

Kozak, S., S. Nagel, and S. Santosh (2020). Shrinking the Cross-Section. *Journal of Financial Economics 135*(2), 271–292.

Lamont, O. (1998). Earnings and Expected Returns. *The Journal of Finance 53*(5), 1563–1587.

Ledoit, O. and M. Wolf (2003). Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance 10*(5), 603–621.

Ledoit, O. and M. Wolf (2004). Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management 30*(4), 110–119.

Light, N., D. Maslov, and O. Rytchkov (2017). Aggregation of Information About the Cross Section of Stock Returns: A Latent Variable Approach. *The Review of Financial Studies 30*(4), 1339–1381.

Lintner, J. (1965). Security Prices, Risk, and Maximal Gains from Diversification. *The Journal of Finance 20*(4), 587–615.

Manresa, E., F. Peñaranda, and E. Sentana (2017). Empirical Evaluation of Overspecified Asset Pricing Models. *Working Paper*.

McCulloch, R. and P. E. Rossi (1991). A Bayesian Approach to Testing the Arbitrage Pricing Theory. *Journal of Econometrics 49*(1), 141–168.

Merton, R. C. (1973). An Intertemporal Capital Asset Pricing Model. *Econometrica 41*(5), 867–887.

Messmer, M. and F. Audrino (2017). The (Adaptive) Lasso in the Zoo-Firm Characteristic Selection in the Cross-Section of Expected Returns. *Working Paper*.

Pástor, Ľ. (2000). Portfolio Selection and Asset Pricing Models. *The Journal of Finance 55*(1), 179–223.

Pástor, Ľ. and R. F. Stambaugh (1999). Costs of Equity Capital and Model Mispricing. *The Journal of Finance 54*(1), 67–121.

Pástor, Ľ. and R. F. Stambaugh (2000). Comparing Asset Pricing Models: An Investment Perspective. *Journal of Financial Economics 56*(3), 335–381.

Pástor, Ľ. and R. F. Stambaugh (2002a). Investing in Equity Mutual Funds. *Journal of Financial Economics 63*(3), 351–380.

Pástor, Ľ. and R. F. Stambaugh (2002b). Mutual Fund Performance and Seemingly Unrelated Assets. *Journal of Financial Economics 63*(3), 315–349.

Pástor, Ľ. and R. F. Stambaugh (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy 111*(3), 642–685.

Pastor, L. and R. F. Stambaugh (2009). Predictive Systems: Living with Imperfect Predictors. *The Journal of Finance 64*(4), 1583–1628.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach.* Cambridge, Mass. and London: MIT Press.

Robbins, H. (1956). An Empirical Bayes Approach to Statistics. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1 of *Third Berkeley Symposium on Mathematical Statistics and Probability*, pp. 157–164. University of California Press.

Robbins, H. (1964). The Empirical Bayes Approach to Statistical Decision Problems. *The Annals of Mathematical Statistics 35*(1), 1–20.

Ross, S. A. (1976). The Arbitrage Theory of Capital Asset Pricing. *Journal of Economic Theory 13*(3), 341–360.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance 19*(3), 425–442.

Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The Short of It: Investor Sentiment and Anomalies. *Journal of Financial Economics 104*(2), 288–302.

Stambaugh, R. F. and Y. Yuan (2017). Mispricing Factors. *The Review of Financial Studies 30*(4), 1270–1315.

Tu, J. and G. Zhou (2004). Data-Generating Process Uncertainty: What Difference Does It Make in Portfolio Decisions? *Journal of Financial Economics 72*(2), 385–421.

Van Nieuwerburgh, S. and L. Veldkamp (2010). Information Acquisition and Under-Diversification. *Review of Economic Studies 77*(2), 779–805.

Wachter, J. A. and M. Warusawitharana (2009). Predictable Returns and Asset Allocation: Should A Skeptical Investor Time the Market? *Journal of Econometrics 148*(2), 162–178.

Wachter, J. A. and M. 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*(1), 74–93.

Welch, I. and A. Goyal (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies 21*(4), 1455–1508.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics.* Wiley classics library. New York and Chichester: John Wiley.

Table 1: 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 premiums, 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 the Online Appendix equation (D.1).

Panel A: Model Specifications

Model	Mispricing	Factor Loading	Risk Premium	Data Generating Process	
	$\alpha(z_t)$	$\beta(z_t)$	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	$N+K-k-1$
$\mathbb{M}_3$	$(N+K-k)(k+km)$	$-km$	$k(1+m)$	$N+K-k-m-1$
$\mathbb{M}_4$	$(N+K-k)(1+m+k+km)$	$-(k+1)m-1$	$k(1+m)$	$N+K-k-m-1$

Table 2: Variable Definitions

Variable	Abbreviation	Definition
Panel A: Asset Pricing Factors		
Market	MKT	The excess return on the value-weighted CRSP market index over the one-month Treasury bill rate (Sharpe (1964), Lintner (1965)).
Size	SMB	The small minus big firm return premium (Fama and French (1993)).
Value	HML	The high book-to-market minus the low book-to-market return premium (Fama and French (1993)).
Profitability	RMW	The robust minus the weak return premium (Fama and French (2015)).
Investment	CMA	The conservative minus the aggressive return premium (Fama and French (2015)).
Momentum	MOM	The winner minus the loser return premium (Carhart (1997)).
Post-earnings Announcement Drift	PEAD	The positive earnings surprise minus negative earnings surprise premium (Daniel et al. (2020)).
Financing	FIN	The low-issuance minus the high-issuance return premium (Daniel et al. (2020)).
Quality-minus-junk	QMJ	The quality minus the junk return premium (Asness et al. (2019)).
Betting-against-beta	BAB	The low-beta minus the high-beta return premium (Frazzini and Pedersen (2014)).
Management	MGMT	The underpriced minus the overpriced return premium based on six anomaly variables related to firms' management, including net stock issues, composite equity issues, accruals, net operating assets, asset growth, and investment to assets (Stambaugh and Yuan (2017)).
Performance	PERF	The underpriced minus the overpriced return premium based on five anomaly variables related to firms' performance, including distress, O-score, momentum, gross profitability, and return on assets (Stambaugh and Yuan (2017)).
Liquidity	LIQ	The high-liquidity-beta minus the low-liquidity-beta return premium (Pástor and Stambaugh (2003)).
Intermediary Capital	ICR	The value-weighted equity return for the primary dealer sector (He et al. (2017)).
Panel B: Macro Predictors		
Dividend Price Ratio	dp	The difference between the log of dividends and the log of prices, where dividends are 12-month moving sums of dividends paid on the S&P 500 index, and prices are monthly averages of daily closing prices (Campbell and Shiller (1988), Campbell and Yogo (2006)).
Dividend Yield	dy	The difference between the log of dividends and the log of lagged prices (Ball (1978)).
Earnings Price Ratio	ep	The difference between the log of earnings and the log of prices, where earnings are 12-month moving sums of earnings on the S&P 500 index (Campbell and Shiller (1988)).
Dividend Payout Ratio	de	The difference between the log of dividends and the log of earnings (Lamont (1998)).
Stock Variance	svar	The sum of squared daily returns on the S&P 500 index (Guo (2006)).
Book-to-market Ratio	bm	The ratio of book value to market value for the Dow Jones Industrial Average (Kothari and Shanken (1997)).
Net Equity Expansion	ntis	The ratio of 12-month moving sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stocks (Campbell et al. (2008)).
Treasury Bills	tbl	The *3-Month Treasury Bill: Secondary Market Rate* from the economic research data base at the Federal Reserve Bank at St. Louis (Campbell (1987)).
Long Term Yield	lty	The long-term government bond yield from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook (Welch and Goyal (2008)).
Long Term Rate of Returns	ltr	The long-term government bond returns from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook (Welch and Goyal (2008)).
Term Spread	tms	The difference between the long term yield on government bonds and the Treasury bill (Campbell (1987)).
Default Yield Spread	dfy	The difference between BAA and AAA-rated corporate bond yields (Fama and French (1989)).
Inflation	infl	The *Consumer Price Index (All Urban Consumers)* from the Bureau of Labor Statistics (Campbell and Vuolteenaho (2004)).

Table 3: Summary Statistics

This table presents the summary statistics for the data used in the paper during the period from June 1977 to December 2016. Panel A reports the means, medians, and standard deviations of monthly factor returns and the correlations between them. We also report the monthly CAPM α and its corresponding t-statistics. Panel B reports the means, medians, standard deviations, and AR(1) coefficients of the monthly macro predictors. Table 2 provides the detailed definitions of each variable.

Panel A: Summary Statistics for Factors

	MKT	SMB	HML	RMW	CMA	MOM	PEAD	FIN	QMJ	BAB	MGMT	PERF	LIQ	ICR
Mean	0.637	0.219	0.289	0.363	0.290	0.625	0.576	0.715	0.466	0.952	0.572	0.710	0.443	1.132
Median	1.070	0.110	0.120	0.330	0.160	0.770	0.621	0.621	0.457	1.129	0.531	0.648	0.380	1.480
Std.Dev.	4.441	2.912	2.895	2.352	1.972	4.458	1.890	3.910	2.363	3.633	2.827	3.987	3.411	6.680
CAPM α		0.121	0.406	0.464	0.401	0.689	0.596	0.990	0.650	1.039	0.779	0.869	0.448	0.357
t-statistics		0.921	3.145	4.458	4.752	3.346	6.806	6.264	7.070	6.245	6.902	4.882	2.830	1.958
MKT	1.000	0.233	-0.281	-0.301	-0.390	-0.101	-0.074	-0.492	-0.546	-0.168	-0.511	-0.277	-0.010	0.810
SMB		1.000	-0.148	-0.406	-0.079	0.049	0.065	-0.417	-0.441	-0.016	-0.331	-0.101	-0.035	0.108
HML			1.000	0.214	0.690	-0.221	-0.175	0.637	0.021	0.385	0.715	-0.299	0.022	0.026
RMW				1.000	0.120	0.115	-0.076	0.653	0.752	0.377	0.344	0.431	-0.043	-0.226
CMA					1.000	-0.011	-0.034	0.599	0.120	0.352	0.767	-0.019	0.002	-0.216
MOM						1.000	0.462	0.039	0.273	0.224	0.003	0.728	0.024	-0.256
PEAD							1.000	-0.067	0.146	0.067	-0.023	0.396	0.049	-0.170
FIN								1.000	0.586	0.495	0.789	0.184	0.007	-0.293
QMJ									1.000	0.248	0.399	0.643	-0.024	-0.528
BAB										1.000	0.369	0.187	0.069	-0.047
MGMT											1.000	0.026	-0.015	-0.261
PERF												1.000	0.016	-0.483
LIQ													1.000	-0.022
ICR														1.000

Panel B: Summary Statistics for Macro Predictors

	dp	dy	ep	de	svar	bm	ntis	tbl	lty	ltr	tms	dfy	infl
Mean	-3.670	-3.663	-2.878	-0.792	0.003	0.444	0.006	0.047	0.069	0.007	0.022	0.011	0.003
Median	-3.790	-3.785	-2.898	-0.853	0.001	0.337	0.009	0.049	0.065	0.008	0.024	0.009	0.003
Std.Dev.	0.439	0.440	0.483	0.355	0.005	0.273	0.020	0.036	0.030	0.032	0.015	0.005	0.004
AR(1)	0.994	0.994	0.988	0.985	0.456	0.993	0.980	0.993	0.996	0.049	0.948	0.963	0.625

Table 4: Posterior Probabilities of Predictors in Predictive Regressions

This table presents the cumulative posterior probabilities for the macro predictors in predictive regressions, computed as $\mathcal{A}'\mathcal{P}$, where $\mathcal{A}$ is a matrix representing all forecasting models by their unique combinations of zeros and ones, and $\mathcal{P}$ is a vector including posterior probabilities for all models. We consider three scenarios described in equation (11): (i) no interaction ($B_1 \neq 0$ and $B_2 = 0$), (ii) with interaction ($B_1 \neq 0$ and $B_2 \neq 0$), and (iii) combined, including both (i) and (ii). Table 2 provides the detailed definitions of each variable.

	No Interaction	With Interaction	Combined
dp	0.15	0.35	0.35
dy	1.00	1.00	1.00
ep	0.85	0.65	0.65
de	0.85	0.65	0.65
svar	0.95	1.00	1.00
bm	0.72	0.04	0.04
ntis	0.74	0.69	0.69
tbl	0.76	0.36	0.36
lty	0.24	0.63	0.63
ltr	0.85	0.31	0.31
tms	0.67	0.24	0.24
dfy	0.72	1.00	1.00
infl	0.24	0.24	0.24

Table 5: 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, i.e., $\mathbb{M}_1$ to $\mathbb{M}_4$, are specified in Table 1. Panel A presents the cumulative posterior probabilities for the factors, as defined in equation (26). Panel B presents similar statistics for the macro predictors, as defined in equation (27). Panel C reports a list of 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 $\mathbb{M}_4$, (iii) the equal-weighted average of hypothetical sample size T_0, as defined in equation (24), and (iv) the equal-weighted average of shrinkage, defined as $\frac{T_0}{T_0+T}$. Table 2 provides the detailed definitions of 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 Macro Predictors

	$\tau = 1.25$	$\tau = 1.5$	$\tau = 2$	$\tau = 3$
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

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

Table 6: 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, i.e., 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, Top 3) and the integrated model (denoted BMA). The investment universe consists of 14 factors as listed in Panel A of Table 2, 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. Namely, the sum of the absolute tangency portfolio weights is set to be smaller than or equal to 2, i.e., $\sum_{i=1}^{14} |w_i| \leq 2$. 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	T	$\frac{T}{2}$		$\frac{2T}{3}$	
	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
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

Table 7: 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, i.e., 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 Top 1, Top 2, Top 3) and the integrated model (denoted BMA). The investment universe consists of 14 factors as listed in Panel A of Table 2, 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. Namely, the sum of the absolute tangency portfolio weights is set to be smaller than or equal to 2, i.e., $\sum_{i=1}^{14} |w_i| \leq 2$. 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
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

Table 8: 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 Top 1, Top 2, Top 3) and the integrated model (denoted BMA). The investment universe consists of 14 factors as listed in Panel A of Table 2, and we employ a list of 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; namely, the sum of the absolute tangency portfolio weights is set to be smaller than or equal to 2, i.e., $\sum_{i=1}^{14} |w_i| \leq 2$. Panel C reports similar statistics, where we replace the tangency portfolio with the global minimum variance portfolio.

τ	Model	T	$\frac{T}{2}$		$\frac{2T}{3}$	
		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
$\tau = 2$	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

Table 9: BMA Model Variance and Sample Variance

This table presents the in-sample and out-of-sample variance for each factor. We report (i) the sample average of the variance based on the integrated model (denoted $\bar{V}_t + \bar{\Omega}_t$), defined as the time-series average of the diagonal elements of the covariance matrix, i.e., $\text{Var}\left[r_{t+1}|D\right]$ in equation (5), and (ii) the sample variance computed from realized factors returns (denoted 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 14 factors as listed in Panel A of Table 2, 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

Table 10: BMA Model Uncertainty

Panel A presents the in-sample and out-of-sample entropy increase, i.e., EI_t in equation (30). We report the mean, the 95^{th} percentile, the 99^{th} percentile, and the 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 14 factors as listed in Panel A of Table 2, 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, i.e., $EI_{i,t}$ in equation (31).

	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 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, i.e., $\mathbb{M}_1$ to $\mathbb{M}_4$, are specified in Table 1. The cumulative posterior probabilities for models in $\mathbb{M}_1$ to $\mathbb{M}_4$ are defined in equations (C.32), (C.24), (B.41), and (B.30), respectively.

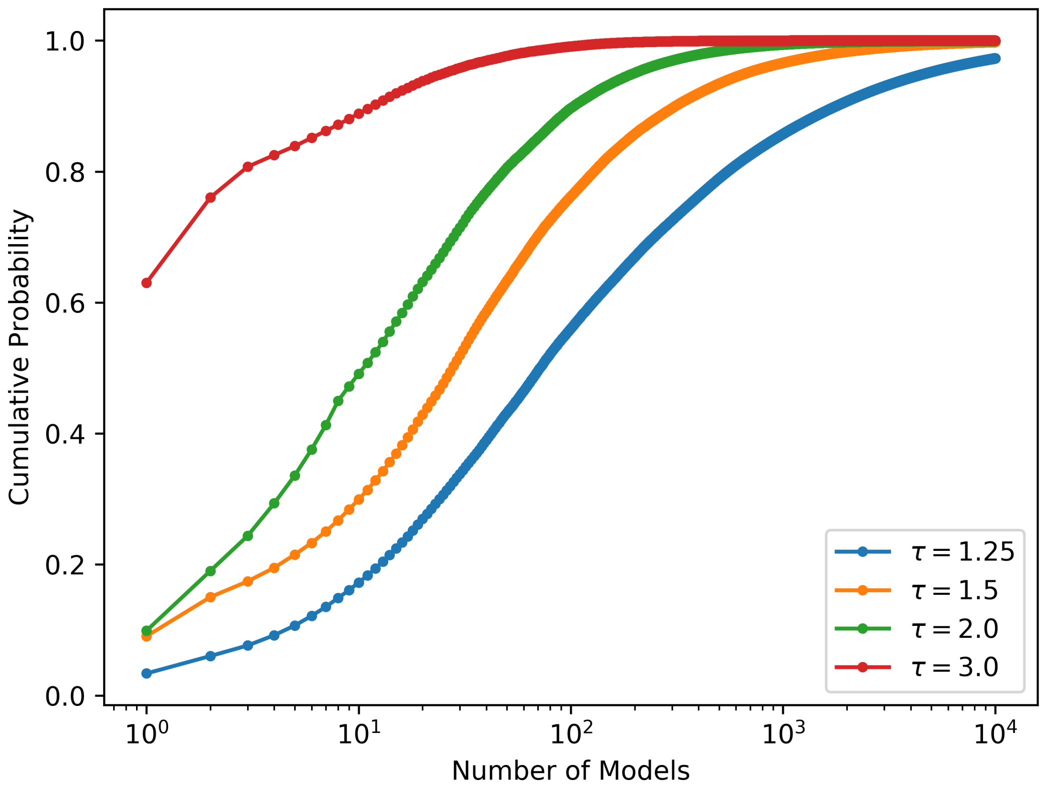
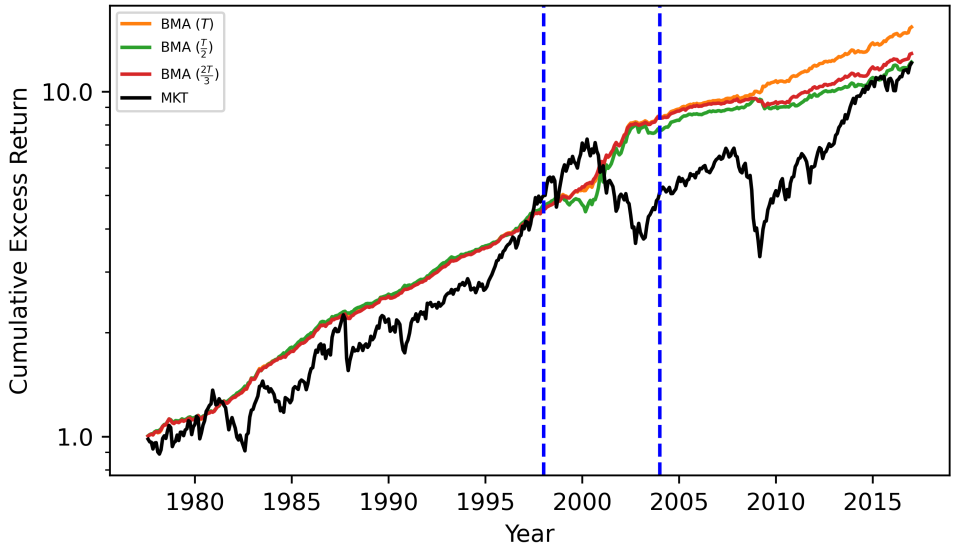
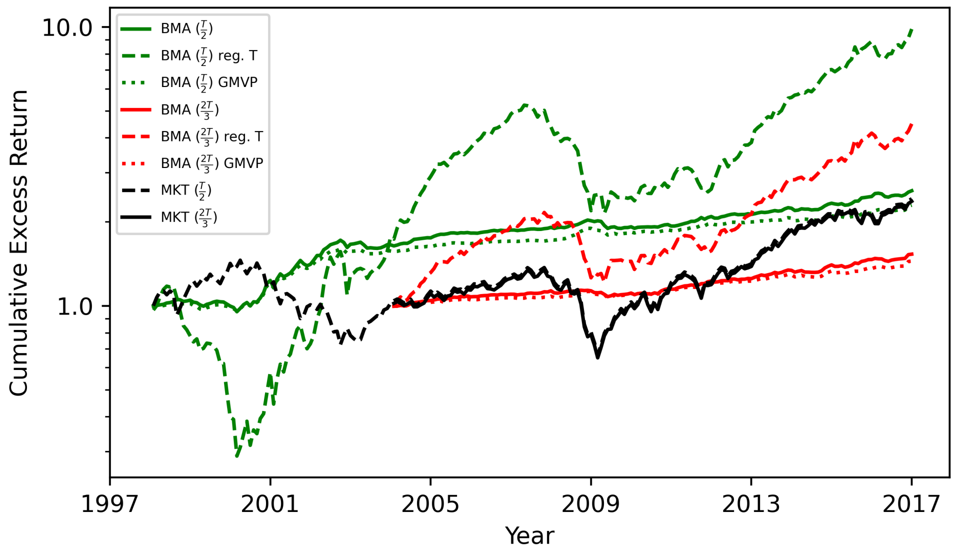


Figure 2: BMA Model Performance Over Time

This figure plots the BMA model performance 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 ($\frac{2T}{3}$). Figure 2a plots the cumulative excess returns on an initial investment of \$1 for the market portfolio (MKT) and three tangency portfolios based on the integrated model. The blue dashed lines mark the end of the in-sample periods for $\frac{T}{2}$ and $\frac{2T}{3}$. Figure 2b plots similar statistics for the market portfolio (MKT), tangency portfolios with and without Regulation T constraint, and the GMVP based on the integrated model. We only plot the out-of-sample periods: one starts from January 1998 ($\frac{T}{2}$) and another starts from January 2004 ($\frac{2T}{3}$).



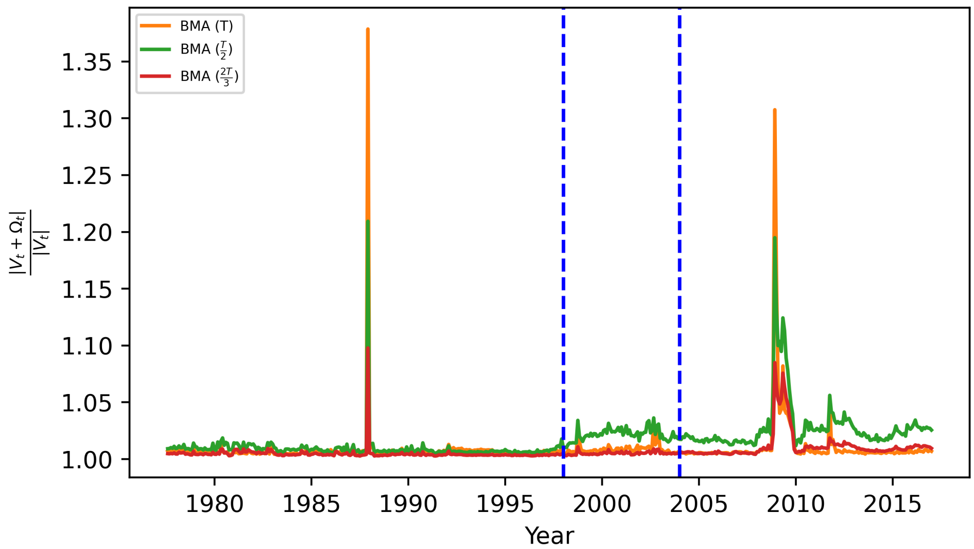
(a) Cumulative Excess Returns on a \$1 Investment: Full Sample



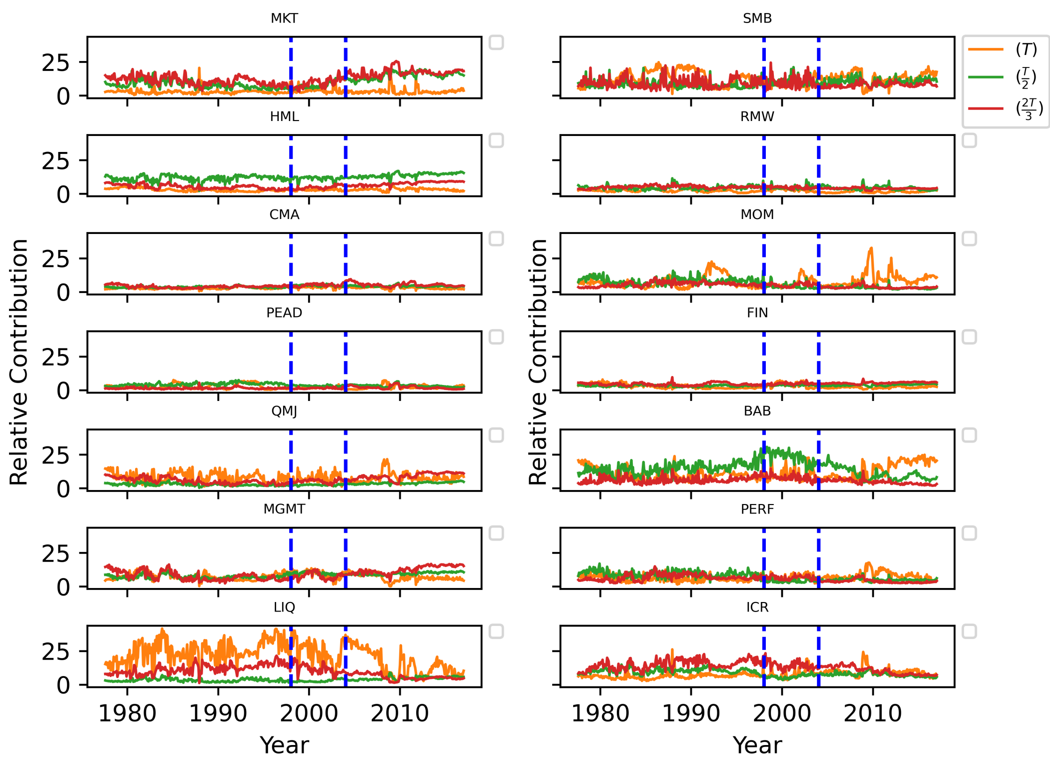
(b) Cumulative Excess Returns on a \$1 Investment: Out-of-Sample

Figure 3: Model Disagreement

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 ($\frac{2T}{3}$). Figure 3a plots the time series of the relative increase in entropy, i.e., EI_t in equation (30). Figure 3b plots, for each factor, the time series of the contribution to the overall entropy increase, i.e., $EI_{i,t}$ in equation (31). The blue dashed lines mark the end of the in-sample periods for $\frac{T}{2}$ and $\frac{2T}{3}$.



(a) Model Disagreement Over Time



(b) Factor Contribution to Model Disagreement Over Time

Integrating Factor Models

Online Appendix

Contents

A Marginal Likelihood of Predictive Regressions

The marginal likelihood derivation for predictive regressions, per equation (11), is based on Avramov (2002), with essential modifications to account for nonlinearities in the predictors as well as interactions. First, it is convenient to reformulate the data generating process in equation (11) in matrix notation $Y = XB + U$, where $X = [x_0, x_1, \ldots, x_{T-1}]'$, $Y = [y_1, y_2, \ldots, y_T]'$, $U = [\varepsilon_1, \ldots, \varepsilon_T]'$, and $x_t = [1, z_t']'$ if interaction terms are omitted or $x_t = [1, z_t', \text{vech}(z_t z_t')']'$ if interaction terms are included. The set of selected predictors differs across models, but we refrain from model-specific subscripts to keep the notation lightweight.

We conduct Bayesian inference for the primary sample based on the joint posterior distribution of B and Σ from the hypothetical sample equipped with a noninformative prior. Multiplying the likelihood of the hypothetical sample with the noninformative prior $\pi(B, \Sigma) \propto |\Sigma|^{-\frac{N+K+1}{2}}$ yields the informative prior, formulated as

$$\pi(B, \Sigma|D_0) \propto |\Sigma|^{-\frac{T_0+N+K+1}{2}} \exp\left(-\frac{1}{2}\text{Tr}\left[\left(S_0 + (B - B_0)'X_0'X_0(B - B_0)\right)\Sigma^{-1}\right]\right), \quad \text{(A.1)}$$

where

$$S_0 = (Y_0 - X_0B_0)'(Y_0 - X_0B_0) = T_0\hat{V}_y \quad \text{(A.2)}$$

$$B_0 = \begin{bmatrix} \bar{y}' \\ 0 \end{bmatrix} \quad \text{(A.3)}$$

$$(X_0'X_0) = \frac{T_0}{T}(X'X) = T_0\overline{xx'}, \quad \text{(A.4)}$$

and $\overline{xx'}$ is the sample mean of xx'.

Standard results imply that the priors for B and Σ can be expressed as

$$\mathrm{vec}\,(B)\,|\Sigma, D_0 \sim f_{\mathrm{MV}}\left(\mathrm{vec}\,(B_0)\,, \Sigma \otimes (X_0' X_0)^{-1}\right), \tag{A.5}$$

$$\Sigma | D_0 \sim f_{\mathrm{IW}}^{N+K}\,(S_0, T_0 - m - 1)\,, \tag{A.6}$$

where $\mathrm{vec}\,(\cdot)$ denotes the vector formed by stacking the successive transformed rows of a matrix, m is the number of retained predictors, MV stands for the multivariate normal distribution, IW stands for the inverted Wishart density, and $N+K$ denotes the degrees of freedom of the inverted Wishart density. We refer to Zellner (1971) for further technical details.

Combining the likelihood of the actual sample D and the priors in equations (A.5) and (A.6) yields the posterior density

$$\mathrm{vec}\,(B)\,|\Sigma, D \sim N\left(\mathrm{vec}\left(\tilde{B}\right), \Sigma \otimes (X_0' X_0 + X'X)^{-1}\right) \tag{A.7}$$

$$\Sigma | D \sim f_{IW}^{N+K}\left(\tilde{S}, T^* - m - 1\right), \tag{A.8}$$

where $T^* = T + T_0$,

$$\tilde{B} = \frac{T}{T^*}\,(X'X)^{-1}\,(T_0 \bar{x}\bar{y}' + X'Y)\,, \tag{A.9}$$

$$\tilde{S} = T^*\left(\hat{V}_y + \bar{y}\bar{y}'\right) - \frac{T}{T^*}\,(T_0 \bar{y}\bar{x}' + Y'X)\,(X'X)^{-1}\,(T_0 \bar{x}\bar{y}' + X'Y)\,. \tag{A.10}$$

Then, the log marginal likelihood for the predictive regression setup formulated in equation (11) takes the form

$$\begin{aligned}\ln\left[m(D|\mathcal{M})\right] \;&=\; -\frac{T(N+K)}{2}\ln(\pi) + \frac{T_0 - m - 1}{2}\ln|T_0 \hat{V}_y| - \frac{T^* - m - 1}{2}\ln|\tilde{S}| \\ &-\; \sum_{i=1}^{N+K}\ln\left\{\Gamma\left(\frac{T_0 - m - i}{2}\right)\right\} + \sum_{i=1}^{N+K}\ln\left\{\Gamma\left(\frac{T^* - m - i}{2}\right)\right\}\end{aligned}$$

$$- \quad \frac{(N+K)(m+1)}{2} \ln \left(\frac{T^*}{T_0} \right), \tag{A.11}$$

where $\Gamma(\phi)$ stands for the Gamma function evaluated at ϕ, and $|x|$ is the determinant of x.

B Conditional Asset Pricing Models

B.1 Marginal Likelihood for the Unrestricted Case

The multivariate representation of the beta pricing equations (1)-(2) is given by

$$\begin{aligned} R &= \iota_T \alpha_0' + Z_{-1}\alpha_1' + F\beta_0' + \Xi\beta_1' + U_R \\ &= W\Phi + U_R \end{aligned} \tag{B.1}$$

$$F = XA_F + U_F, \tag{B.2}$$

where $R = [r_1, \ldots, r_T]'$, $F = [f_1, \ldots, f_T]'$, $X = [\iota_T, Z_{-1}]$, $Z_{-1} = [z_0, \ldots, z_{T-1}]'$, ι_T is a T-vector of ones, $\Xi = [\xi_1, \ldots, \xi_T]' \in \mathbb{R}^{T\times(km)}$ is defined through $\xi_t = (I_k \otimes z_{t-1})\, f_t$, $W = [X, F, \Xi]$, $\Phi = [\alpha_0, \alpha_1, \beta_0, \beta_1]'$, $U_R \in \mathbb{R}^{T\times(N+K-k)}$ and $U_F \in \mathbb{R}^{T\times k}$ are the matrices of residuals, N is the number of test assets, K is the total number of factors, and $K-k$ is the number of factors that are not benchmark factors in the asset pricing specification. These factors are added to the test assets on the left-hand side of equation (B.1). The k remaining factors are on the right-hand side of equation (B.1) and the left-hand side of equation (B.2). The number of retained predictors is denoted by m, and it ranges, in the conditional setting, between one and M.

We assume that the error terms $U_R = [u_{r,1}, \ldots, u_{r,T}]'$ and $U_F = [u_{f,1}, \ldots, u_{f,T}]'$ conditional on the parameter space obey the normal distribution

$$\begin{pmatrix} u_{r,t} \\ u_{f,t} \end{pmatrix} \sim \text{ i.i.d.} f_{\text{MV}}(0, \Psi), \text{ where } \Psi = \begin{pmatrix} \Sigma_{RR} & 0 \\ 0 & \Sigma_{FF} \end{pmatrix}. \tag{B.3}$$

The factor and return innovations are orthogonal through equation (B.1). The unconditional distribution of the error terms (upon integrating out the parameter space) essentially departs from normality. Notably, Tu and Zhou (2004) show that certainty equivalent losses associated with ignoring fat tails are small, suggesting that the normality assumption could well characterize the stock return distribution from a decision-making perspective. Then, based on the multivariate representation of the beta pricing equations, the likelihood of the hypothetical sample $D_0 = (R_0, F_0, X_0)$ can be written as

$$\begin{aligned} &\mathcal{L}(D_0|Z_0, \Sigma_{RR}, \Sigma_{FF}, \Phi, A_F, \mathcal{M}) \propto \\ &\quad |\Sigma_{RR}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\text{Tr}\left[\Sigma_{RR}^{-1}\left(R_0' Q_{W_0} R_0 + \left(\Phi - \hat{\Phi}_0\right)' W_0' W_0 \left(\Phi - \hat{\Phi}_0\right)\right)\right]\right\} \times \\ &\quad |\Sigma_{FF}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\text{Tr}\left[\Sigma_{FF}^{-1}\left(F_0' Q_{X_0} F_0 + \left(A_F - \hat{A}_{F_0}\right)' X_0' X_0 \left(A_F - \hat{A}_{F_0}\right)\right)\right]\right\} \end{aligned} \tag{B.4}$$

where

$$\hat{\Phi}_0 = \left[\mathbf{0}_{(N+K-k)\times(m+1)}, \hat{\beta}_0, \mathbf{0}_{(N+K-k)\times km}\right]' \tag{B.5}$$

$$\hat{A}_{F_0} = \left[\hat{\bar{f}}, \mathbf{0}_{k\times m}\right]'. \tag{B.6}$$

Here, $\hat{\beta}_0 = (F'F)^{-1} F'R$ is the slope coefficient in a zero-intercept regression of the returns r_t on the factors f_t, $\mathbf{0}_{i\times j}$ is a matrix of zeros with i rows and j columns, and the operator $Q_J = I_T - J(J'J)^{-1} J'$ is defined for a matrix J with full column-rank and T rows. The hypothetical sample is weighted against both mispricing and time variation in the alpha, beta, and risk premia.

Combining the likelihood of the hypothetical sample with the uninformative prior $\pi(\Sigma_{RR}, \Sigma_{FF}, \Phi, A_F) \propto |\Psi|^{-\frac{N+K+1}{2}}$ yields the informed prior, formulated as

$$\mathrm{vec}\,(A_F)\,|\Sigma_{FF}, D_0 \sim f_{\mathrm{MV}}\left(\mathrm{vec}\left(\hat{A}_{F_0}\right), \Sigma_{FF} \otimes \left(X_0{}'X_0\right)^{-1}\right) \tag{B.7}$$

$$\Sigma_{FF}|D_0 \sim f_{\mathrm{IW}}^{k}\left(F_0' Q_{X_0} F_0, T_0 + N + K - k - m - 1\right) \tag{B.8}$$

$$\mathrm{vec}\,(\Phi)\,|\Sigma_{RR}, D_0 \sim f_{\mathrm{MV}}\left(\mathrm{vec}\left(\hat{\Phi}_0\right), \Sigma_{RR} \otimes \left(W_0' W_0\right)^{-1}\right) \tag{B.9}$$

$$\Sigma_{RR}|D_0 \sim f_{\mathrm{IW}}^{N+K-k}\left(R_0' Q_{W_0} R_0, T_0 - (k+1)m - 1\right). \tag{B.10}$$

All quantities based on the hypothetical sample in equations (B.7)-(B.10) can be expressed in terms of quantities observed from the actual sample. Specifically, note that

$$X_0' X_0 = \frac{T_0}{T}\left(X'X\right) = T_0 \begin{pmatrix} 1 & \bar{z}' \\ \bar{z} & \hat{V}_z + \bar{z}\bar{z}' \end{pmatrix} \tag{B.11}$$

$$\begin{aligned} F_0' Q_{X_0} F_0 &= \left(F_0 - X_0 \hat{A}_{F_0}\right)' \left(F_0 - X_0 \hat{A}_{F_0}\right) \\ &= \left(F_0 - \iota_{T_0} \bar{f}'\right)' \left(F_0 - \iota_{T_0} \bar{f}'\right) = T_0 \hat{V}_f \end{aligned} \tag{B.12}$$

$$\begin{aligned} R_0' Q_{W_0} R_0 &= \left(R_0 - W_0 \hat{\Phi}_0\right)' \left(R_0 - W_0 \hat{\Phi}_0\right) \\ &= \frac{T_0}{T}\left(R'R - \hat{\Phi}_0' W' W \hat{\Phi}_0\right) = \frac{T_0}{T}\left(R'R - \hat{\beta}_0' F' F \hat{\beta}_0\right). \end{aligned} \tag{B.13}$$

We next express $W_0' W_0$ in terms of quantities observed from the actual sample. First note that

$$X_0' F_0 = \left(X_0' X_0\right) \hat{A}_{F_0} = T_0 \begin{pmatrix} 1 & \bar{z}' \\ \bar{z} & \hat{V}_z + \bar{z}\bar{z}' \end{pmatrix} \begin{pmatrix} \bar{f}' \\ \mathbf{0}_{m \times k} \end{pmatrix} = T_0 \begin{pmatrix} \bar{f}' \\ \bar{z}\bar{f}' \end{pmatrix}. \tag{B.14}$$

Further, we exploit the prior independence between F_0 and Z_0 to compute $Z_0' \Xi_0$, $F_0' \Xi_0$

and $\Xi_0'\Xi_0$ as

$$Z_0'\Xi_0 = \begin{pmatrix} \sum_{t=0}^{T_0-1}\left(z_{0,t}^1 \cdot \left(f_{0,t+1}' \otimes z_{0,t}'\right)\right) \\ \vdots \\ \sum_{t=0}^{T_0-1}\left(z_{0,t}^m \cdot \left(f_{0,t+1}' \otimes z_{0,t}'\right)\right) \end{pmatrix} = T_0 \begin{pmatrix} \bar{f}' \otimes \overline{z^1 \cdot z'} \\ \vdots \\ \bar{f}' \otimes \overline{z^m \cdot z'} \end{pmatrix} = T_0\left(\bar{f}' \otimes \mathit{\Gamma}\right) \tag{B.15}$$

$$F_0'\Xi_0 = \begin{pmatrix} \sum_{t=1}^{T_0}\left(f_{0,t}^1\left(f_{0,t}' \otimes z_{0,t-1}'\right)\right) \\ \vdots \\ \sum_{t=1}^{T_0}\left(f_{0,t}^k\left(f_{0,t}' \otimes z_{0,t-1}'\right)\right) \end{pmatrix} = T_0 \begin{pmatrix} \overline{f^1 \cdot f'} \otimes \bar{z}' \\ \vdots \\ \overline{f^k \cdot f'} \otimes \bar{z}' \end{pmatrix} = T_0\left(\Delta \otimes \bar{z}'\right) \tag{B.16}$$

where $\overline{z^i \cdot z'} = \frac{1}{T}\sum_{t=0}^{T-1}(z_t^i \cdot z_t')$, $\overline{f^i \cdot f'} = \frac{1}{T}\sum_{t=1}^{T}(f_t^i \cdot f_t')$, $\Gamma = \begin{pmatrix} \overline{z^1 \cdot z'} \\ \vdots \\ \overline{z^m \cdot z'} \end{pmatrix}$, and $\Delta = \begin{pmatrix} \overline{f^1 \cdot f'} \\ \vdots \\ \overline{f^k \cdot f'}. \end{pmatrix}$.

Moreover,

$$\Xi_0'\Xi_0 = \sum_{t=1}^{T_0}(f_t \otimes z_{t-1}) \cdot \left(f_t' \otimes z_{t-1}'\right) = \sum_{t=1}^{T_0}\left(f_t f_t' \otimes z_{t-1} z_{t-1}'\right) = T_0\left(\overline{ff'} \otimes \overline{zz'}\right) \quad . \tag{B.17}$$

Then, it follows that

$$W_0'W_0 = \begin{pmatrix} \iota_{T_0}'\iota_{T_0} & \iota_{T_0}'Z_0 & \iota_{T_0}'F_0 & \iota_{T_0}'\Xi_0 \\ Z_0'\iota_{T_0} & Z_0'Z_0 & Z_0'F_0 & Z_0'\Xi_0 \\ F_0'\iota_{T_0} & F_0'Z_0 & F_0'F_0 & F_0'\Xi_0 \\ \Xi_0'\iota_{T_0} & \Xi_0'Z_0 & \Xi_0'F_0 & \Xi_0'\Xi_0 \end{pmatrix}$$

$$= T_0 \begin{pmatrix} 1 & \bar{z}' & \bar{f}' & \bar{\Xi}' \\ \bar{z} & \hat{V}_z + \bar{z}\bar{z}' & \bar{z}\bar{f}' & \bar{f}' \otimes \Gamma \\ \bar{f} & \bar{f}\bar{z}' & \hat{V}_f + \bar{f}\bar{f}' & \Delta \otimes \bar{z}' \\ \bar{\Xi} & \bar{f} \otimes \Gamma' & \Delta' \otimes \bar{z} & \left(\hat{V}_f + \bar{f}\bar{f}'\right) \otimes \left(\hat{V}_z + \bar{z}\bar{z}'\right) \end{pmatrix}. \tag{B.18}$$

Combining the likelihood of the observed sample $D = (R, F, Z)$ with the prior distributions based on the hypothetical sample (equations (B.7)-(B.10)) yields the following posterior distribution:

$$\mathrm{vec}\,(A_F)\,|\Sigma_{FF}, D \sim f_{\mathrm{MV}}\left(\mathrm{vec}\left(\tilde{A}_F\right), \Sigma_{FF} \otimes \frac{T}{T^*}\,(X'X)^{-1}\right) \tag{B.19}$$

$$\Sigma_{FF}|D \sim f^{k}_{\mathrm{IW}}\,(S_F, T^* + N + K - k - m - 1) \tag{B.20}$$

$$\mathrm{vec}\,(\Phi)\,|\Sigma_{RR}, D \sim f_{\mathrm{MV}}\left(\mathrm{vec}\left(\tilde{\Phi}\right), \Sigma_{RR} \otimes (W_0'W_0 + W'W)^{-1}\right) \tag{B.21}$$

$$\Sigma_{RR}|D \sim f^{N+K-k}_{\mathrm{IW}}\,(S_R, T^* - (k+1)m - 1) \tag{B.22}$$

where

$$\tilde{A}_F = \frac{T}{T^*}\,(X'X)^{-1}\left(X'F + T_0\left[\bar{f}, \bar{f}\bar{z}'\right]'\right) = \frac{1}{T^*}\left(T\,(X'X)^{-1}X'F + T_0\hat{A}_{F_0}\right) \tag{B.23}$$

$$\tilde{\Phi} = (W_0{}'W_0 + W'W)^{-1}\left(W'R + W_0'W_0\hat{\Phi}_0\right) = \frac{1}{T^*}\left(T\,(W'W)^{-1}W'R + T_0\hat{\Phi}_0\right) \tag{B.24}$$

$$S_F = T^*\left(\hat{V}_f + \bar{f}\bar{f}'\right) - \frac{T}{T^*}\left(T_0[\bar{f}, \bar{f}\bar{z}'] + F'X\right)(X'X)^{-1}\left(T_0[\bar{f}, \bar{f}\bar{z}']' + X'F\right) \tag{B.25}$$

$$\begin{aligned} S_R &= R_0'Q_{W_0}R_0 + \left(R - W\tilde{\Phi}\right)'\left(R - W\tilde{\Phi}\right) + \left(\tilde{\Phi} - \hat{\Phi}_0\right)'(W_0'W_0)\left(\tilde{\Phi} - \hat{\Phi}_0\right) \\ &= R_0'R_0 + R'R - \tilde{\Phi}'\,(W_0'W_0 + W'W)\,\tilde{\Phi} = \frac{T^*}{T}\left(R'R - \tilde{\Phi}'W'W\tilde{\Phi}\right), \end{aligned} \tag{B.26}$$

where we express $W_0'W_0$ in terms of quantities observed from the actual sample, $\frac{W_0'W_0}{T_0} = \frac{W'W}{T}$ and similarly, $\frac{R_0'R_0}{T_0} = \frac{R'R}{T}$.

To compute the marginal likelihood, we employ equation (7) and substitute the corresponding quantities given in equations (B.4)-(B.22) for the prior, likelihood, and posterior densities. This yields

$$m(D|\mathcal{M}_C) = \pi^{-\frac{1}{2}(T)(N+K)} \times \left[\frac{|W_0'W_0|}{|W_0'W_0 + W'W|}\right]^{\frac{1}{2}(N+K-k)} \left[\frac{T_0}{T+T_0}\right]^{\frac{1}{2}k(m+1)} \times$$
$$\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^* - (k+1)m - 1\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0 - (k+1)m - 1\right]\right)}\right] \left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^* + N + K - k - m - 1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0 + N + K - k - m - 1\right]\right)}\right] \times$$
$$\left[\frac{|R_0'Q_{W_0}R_0|^{\frac{1}{2}(T_0-(k+1)m-1)}}{|S_R|^{\frac{1}{2}(T^*-(k+1)m-1)}}\right] \left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-m-1)}}{|S_F|^{\frac{1}{2}(T^*+N+K-k-m-1)}}\right], \tag{B.27}$$

where $\Gamma_p(\cdot)$ is the multivariate gamma function, a generalization of the gamma function. As $W'W$ is $(1+m+k+km) \times (1+m+k+km)$ matrix, we obtain

$$|W_0'W_0| = \left(\frac{T_0}{T}\right)^{(1+m+k+km)} |W'W| \tag{B.28}$$
$$|W_0'W_0 + W'W| = \left(\frac{T^*}{T}\right)^{(1+m+k+km)} |W'W|. \tag{B.29}$$

Therefore, the marginal likelihood from equation (B.27) is

$$m(D|\mathcal{M}_C) = \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)(T_0+k)+\frac{1}{2}k(m+1)} \times \left[\frac{T}{T^*}\right]^{\frac{1}{2}(N+K-k)T} \times$$
$$\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^* - (k+1)m - 1\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0 - (k+1)m - 1\right]\right)}\right] \left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^* + N + K - k - m - 1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0 + N + K - k - m - 1\right]\right)}\right] \times$$
$$\left[\frac{|R'R - R'F\left(F'F\right)^{-1}F'R|^{\frac{1}{2}(T_0-(k+1)m-1)}}{|R'R - \tilde{\Phi}'W'W\tilde{\Phi}|^{\frac{1}{2}(T^*-(k+1)m-1)}}\right] \left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-m-1)}}{|S_F|^{\frac{1}{2}(T^*+N+K-k-m-1)}}\right]. \tag{B.30}$$

B.2 Marginal Likelihood for the Restricted Case

The derivation of the marginal likelihood when restricting the parameters α_0 and α_1 to zero is closely related to the unrestricted case. The likelihood of the hypothetical sample

$D_0 = (R_0, F_0, X_0)$ can be rewritten as

$$\begin{aligned}
&\mathcal{L}(D_0|Z_0, \Sigma_{RR}, \Sigma_{FF}, \Phi^R, A_F, \mathcal{M}_R) \propto \\
&\qquad |\Sigma_{RR}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left[\Sigma_{RR}^{-1}\left(R_0' Q_{W_0^R} R_0 + \left(\Phi^R - \hat{\Phi}_0^R\right)' W_0^{R\prime} W_0^R \left(\Phi^R - \hat{\Phi}_0^R\right)\right)\right]\right\} \times \\
&\qquad |\Sigma_{FF}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left[\Sigma_{FF}^{-1}\left(F_0' Q_{X_0} F_0 + \left(A_F - \hat{A}_{F_0}\right)' X_0' X_0 \left(A_F - \hat{A}_{F_0}\right)\right)\right]\right\}
\end{aligned} \tag{B.31}$$

where

$$\hat{\Phi}_0^R = \left[\hat{\beta}_0, \mathbf{0}_{(N+K-k)\times km}\right]' \tag{B.32}$$

$$W_0^R = \left[F_0, \Xi_0\right]. \tag{B.33}$$

Then, the prior distributions for the parameters Σ_{FF} and A_F remain the same as in equations (B.7) and (B.8). The priors for Σ_{RR} and Φ^R are given by

$$\mathrm{vec}\left(\Phi^R\right)|\Sigma_{RR}, D_0 \sim f_{\mathrm{MV}}\left(\mathrm{vec}\left(\hat{\Phi}_0^R\right), \Sigma_{RR} \otimes \left(W_0^{R\prime} W_0^R\right)^{-1}\right) \tag{B.34}$$

$$\Sigma_{RR}|D_0 \sim f_{\mathrm{IW}}^{N+K-k}\left(R_0' Q_{W_0^R} R_0, T_0 - km\right). \tag{B.35}$$

Equivalent to equation (B.18), one can replace all quantities based on the hypothetical sample in terms of moments of the observed data

$$W_0^{R\prime} W_0^R = T_0 \begin{pmatrix} \hat{V}_f + \bar{f}\bar{f}' & \Delta \otimes \bar{z}' \\ \Delta' \otimes \bar{z} & \hat{V}_\Xi + \bar{\Xi}\bar{\Xi}' \end{pmatrix}. \tag{B.36}$$

Combining the likelihood of the observed sample $D = (R, F, Z)$ with the prior distributions based on the hypothetical sample yields the following posterior distributions for Φ^R

and Σ_{RR}

$$\operatorname{vec}\left(\Phi^{R}\right)|\Sigma_{RR}, D \propto f_{\mathrm{MV}}\left(\operatorname{vec}\left(\tilde{\Phi}^{R}\right), \Sigma_{RR} \otimes\left(W_{0}^{R\prime} W_{0}^{R}+W^{R\prime} W^{R}\right)^{-1}\right) \tag{B.37}$$

$$\Sigma_{RR}|D \propto f_{\mathrm{IW}}^{N+K-k}\left(\tilde{S}_{R}, T^{*}-km\right) \tag{B.38}$$

where

$$\begin{aligned}
\tilde{\Phi}^{R} &=\left(W_{0}^{R\prime} W_{0}^{R}+W^{R\prime} W^{R}\right)^{-1}\left(W^{R\prime} R+W_{0}^{R\prime} W_{0}^{R} \hat{\Phi}_{0}^{R}\right) \\
&=\frac{1}{T^{*}}\left(T\left(W^{R\prime} W^{R}\right)^{-1} W^{R\prime} R+T_{0} \hat{\Phi}_{0}^{R}\right)
\end{aligned} \tag{B.39}$$

$$\begin{aligned}
\tilde{S}_{R} &=R_{0}^{\prime} Q_{W_{0}^{R}} R_{0}+\left(R-W^{R} \tilde{\Phi}^{R}\right)^{\prime}\left(R-W^{R} \tilde{\Phi}^{R}\right)+\left(\tilde{\Phi}^{R}-\hat{\Phi}_{0}^{R}\right)^{\prime}\left(W_{0}^{R\prime} W_{0}^{R}\right)\left(\tilde{\Phi}^{R}-\hat{\Phi}_{0}^{R}\right) \\
&=R_{0}^{\prime} R_{0}+R^{\prime} R-\tilde{\Phi}^{R\prime}\left(W_{0}^{R\prime} W_{0}^{R}+W^{R\prime} W^{R}\right) \tilde{\Phi}^{R}=\frac{T^{*}}{T}\left(R^{\prime} R-\tilde{\Phi}^{R\prime} W^{R\prime} W^{R} \tilde{\Phi}^{R}\right).
\end{aligned} \tag{B.40}$$

Thus, the marginal likelihood is

$$\begin{aligned}
m(D|\mathcal{M}_{C}^{R}) = \pi^{-\frac{1}{2}T(N+K)} &\times \left[\frac{T_{0}}{T^{*}}\right]^{\frac{1}{2}(N+K-k)(k+km)+\frac{1}{2}k(m+1)} \times \\
&\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^{*}-km\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_{0}-km\right]\right)}\right]\left[\frac{\Gamma_{k}\left(\frac{1}{2}\left[T^{*}+N+K-k-m-1\right]\right)}{\Gamma_{k}\left(\frac{1}{2}\left[T_{0}+N+K-k-m-1\right]\right)}\right] \times \\
&\left[\frac{|R_{0}^{\prime} Q_{W_{0}^{R}} R_{0}|^{\frac{1}{2}(T_{0}-km)}}{|\tilde{S}_{R}|^{\frac{1}{2}(T^{*}-km)}}\right]\left[\frac{|T_{0} \hat{V}_{f}|^{\frac{1}{2}(T_{0}+N+K-k-m-1)}}{|S_{F}|^{\frac{1}{2}(T^{*}+N+K-k-m-1)}}\right] \\
&= \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_{0}}{T^{*}}\right]^{\frac{1}{2}(N+K-k)(T_{0}+k)+\frac{1}{2}k(m+1)} \times \left[\frac{T}{T^{*}}\right]^{\frac{1}{2}(N+K-k)T} \times \\
&\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^{*}-km\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_{0}-km\right]\right)}\right]\left[\frac{\Gamma_{k}\left(\frac{1}{2}\left[T^{*}+N+K-k-m-1\right]\right)}{\Gamma_{k}\left(\frac{1}{2}\left[T_{0}+N+K-k-m-1\right]\right)}\right] \times \\
&\left[\frac{|R^{\prime} R-R^{\prime} F\left(F^{\prime} F\right)^{-1} F^{\prime} R|^{\frac{1}{2}(T_{0}-km)}}{|R^{\prime} R-\tilde{\Phi}^{R\prime} W^{R\prime} W^{R} \tilde{\Phi}^{R}|^{\frac{1}{2}(T^{*}-km)}}\right]\left[\frac{|T_{0} \hat{V}_{f}|^{\frac{1}{2}(T_{0}+N+K-k-m-1)}}{|S_{F}|^{\frac{1}{2}(T^{*}+N+K-k-m-1)}}\right].
\end{aligned} \tag{B.41}$$

C Unconditional Asset Pricing Models

C.1 Marginal Likelihood for the Unrestricted Case

Starting from the multivariate representation of the beta pricing equations (1)-(2), in the case where α, β and the risk premiums are time invariant, we obtain:

$$\begin{aligned} R &= \iota_T \alpha' + F\beta' + U_R \\ &= W\Phi + U_R \end{aligned} \tag{C.1}$$

$$F = \iota_T A_F + U_F \tag{C.2}$$

where $W = [\iota_T, F]$, $\Phi = [\alpha, \beta]'$, $U_R \in \mathbb{R}^{T\times(N+K-k)}$ and $U_F \in \mathbb{R}^{T\times k}$ are the matrices of residuals. We assume that the error terms U_R and U_F conditional on the parameter space obey the normal distribution

$$\begin{pmatrix} u_{r,t} \\ u_{f,t} \end{pmatrix} \sim \text{ i.i.d.} f_{\text{MV}}(0, \Psi), \text{ where } \Psi = \begin{pmatrix} \Sigma_{RR} & 0 \\ 0 & \Sigma_{FF} \end{pmatrix}. \tag{C.3}$$

Note that the residuals U_R and U_F in equations (C.1) and (C.2) and the corresponding covariance matrices Σ_{RR} and Σ_{FF} in equation (C.3) are different from their counterparts in appendix B. To ease notation, we leave these quantities unchanged.

Based on the multivariate representation of the beta pricing equations, the likelihood of the hypothetical sample $D_0 = (R_0, F_0)$, denoted by the subscript 0, can be written as

$$\begin{aligned} &\mathcal{L}(D_0|\Sigma_{RR}, \Sigma_{FF}, \Phi, A_F, \mathcal{M}) \propto \\ &\qquad |\Sigma_{RR}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\text{Tr}\left[\Sigma_{RR}^{-1}\left(R_0'Q_{W_0}R_0 + \left(\Phi - \hat{\Phi}_0\right)' W_0'W_0\left(\Phi - \hat{\Phi}_0\right)\right)\right]\right\} \times \\ &\qquad |\Sigma_{FF}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\text{Tr}\left[\Sigma_{FF}^{-1}\left(F_0'Q_{\iota_{T_0}}F_0 + \left(A_F - \hat{A}_F\right)' T_0\left(A_F - \hat{A}_F\right)\right)\right]\right\} \end{aligned} \tag{C.4}$$

where

$$\hat{\Phi}_0 = \left[\mathbf{0}_{(N+K-k)\times 1}, \hat{\beta}_0\right]' \tag{C.5}$$

$$\hat{A}_{F_0} = \hat{\bar{f}}'. \tag{C.6}$$

Note that the hypothetical sample is exclusively weighted against mispricing.

Combining the likelihood of the hypothetical sample with the uninformative prior $\pi(\Sigma_{RR}, \Sigma_{FF}, \Phi, A_F) \propto |\Psi|^{-\frac{N+K+1}{2}}$, the prior distributions for the parameters $\Sigma_{RR}, \Sigma_{FF}, \Phi$, and A_F are given by

$$\mathrm{vec}\,(A_F)\,|\Sigma_{FF}, D_0 \sim f_{\mathrm{MV}}\left(\mathrm{vec}\left(\hat{A}_{F_0}\right), \frac{1}{T_0}\Sigma_{FF}\right) \tag{C.7}$$

$$\Sigma_{FF}|D_0 \sim f^{k}_{\mathrm{IW}}\left(F_0' Q_{\iota_{T_0}} F_0, T_0 + N + K - k - 1\right) \tag{C.8}$$

$$\mathrm{vec}\,(\Phi)\,|\Sigma_{RR}, D_0 \sim f_{\mathrm{MV}}\left(\mathrm{vec}\left(\hat{\Phi}_0\right), \Sigma_{RR} \otimes (W_0' W_0)^{-1}\right) \tag{C.9}$$

$$\Sigma_{RR}|D_0 \sim f^{N+K-k}_{\mathrm{IW}}\left(R_0' Q_{W_0} R_0, T_0 - 1\right). \tag{C.10}$$

All quantities based on the hypothetical sample in equations (C.7)-(C.10) can be expressed in terms of quantities observed from the actual sample. Specifically, note that

$$\begin{aligned} F_0' Q_{\iota_{T_0}} F_0 &= \left(F_0 - \iota_{T_0}\hat{A}_{F_0}\right)'\left(F_0 - \iota_{T_0}\hat{A}_{F_0}\right) \\ &= \left(F_0 - \iota_{T_0}\bar{f}'\right)'\left(F_0 - \iota_{T_0}\bar{f}'\right) = T_0 \hat{V}_f \end{aligned} \tag{C.11}$$

$$\begin{aligned} R_0' Q_{W_0} R_0 &= \left(R_0 - W_0\hat{\Phi}_0\right)'\left(R_0 - W_0\hat{\Phi}_0\right) \\ &= \frac{T_0}{T}\left(R'R - \hat{\Phi}_0' W' W \hat{\Phi}_0\right) = \frac{T_0}{T}\left(R'R - \hat{\beta}_0' F' F \hat{\beta}_0\right), \end{aligned} \tag{C.12}$$

where we express $W_0' W_0$ in terms of quantities observed from the actual sample, $\frac{W_0' W_0}{T_0} = \frac{W'W}{T}$, and similarly, $\frac{R_0' R_0}{T_0} = \frac{R'R}{T}$.

Then, combining the likelihood of the observed sample $D = (R, F)$ with the prior distributions based on the hypothetical sample (equations (C.7)-(C.10)) yields the following posterior distribution:

$$\operatorname{vec}(A_F)\,|\Sigma_{FF}, D \sim f_{\mathrm{MV}}\left(\operatorname{vec}\left(\tilde{A}_F\right), \frac{1}{T^*}\Sigma_{FF}\right) \tag{C.13}$$

$$\Sigma_{FF}|D \sim f_{\mathrm{IW}}^{k}\left(S_F, T^* + N + K - k - 1\right) \tag{C.14}$$

$$\operatorname{vec}(\Phi)\,|\Sigma_{RR}, D \sim f_{\mathrm{MV}}\left(\operatorname{vec}\left(\tilde{\Phi}\right), \Sigma_{RR} \otimes \left(W_0'W_0 + W'W\right)^{-1}\right) \tag{C.15}$$

$$\Sigma_{RR}|D \sim f_{\mathrm{IW}}^{N+K-k}\left(S_R, T^* - 1\right) \tag{C.16}$$

where $T^* = T_0 + T$,

$$\tilde{A}_F = \bar{f}' \tag{C.17}$$

$$\tilde{\Phi} = \left(W_0{}'W_0 + W'W\right)^{-1}\left(W'R + W_0'W_0\hat{\Phi}_0\right) = \frac{1}{T^*}\left(T\left(W'W\right)^{-1}W'R + T_0\hat{\Phi}_0\right) \tag{C.18}$$

$$S_F = T^*\hat{V}_f \tag{C.19}$$

$$\begin{aligned} S_R =& R_0'Q_{W_0}R_0 + \left(R - W\tilde{\Phi}\right)'\left(R - W\tilde{\Phi}\right) + \left(\tilde{\Phi} - \hat{\Phi}_0\right)'\left(W_0'W_0\right)\left(\tilde{\Phi} - \hat{\Phi}_0\right) \\ =& R_0'R_0 + R'R - \tilde{\Phi}'\left(W_0'W_0 + W'W\right)\tilde{\Phi} = \frac{T^*}{T}\left(R'R - \tilde{\Phi}'W'W\tilde{\Phi}\right). \end{aligned} \tag{C.20}$$

To compute the marginal likelihood, we employ equation (7) and substitute the corresponding quantities given in equations (C.4)-(C.16) for the prior, likelihood, and posterior densities. This yields

$$\begin{aligned} m(D|\mathcal{M}_U) = \pi^{-\frac{1}{2}(T)(N+K)} \times \left[\frac{|W_0'W_0|}{|W_0'W_0 + W'W|}\right]^{\frac{1}{2}(N+K-k)}\left[\frac{T_0}{T+T_0}\right]^{\frac{1}{2}k} \times \\ \left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^*-1\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0-1\right]\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^*+N+K-k-1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0+N+K-k-1\right]\right)}\right] \times \\ \left[\frac{|R_0'Q_{W_0}R_0|^{\frac{1}{2}(T_0-1)}}{|S_R|^{\frac{1}{2}(T^*-1)}}\right]\left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N-1)}}{|S_F|^{\frac{1}{2}(T^*+N-1)}}\right]. \end{aligned} \tag{C.21}$$

As $W'W$ is $(1+k) \times (1+k)$ matrix, we obtain

$$|W_0'W_0| = \left(\frac{T_0}{T}\right)^{(1+k)} |W'W| \tag{C.22}$$

$$|W_0'W_0 + W'W| = \left(\frac{T^*}{T}\right)^{(1+k)} |W'W|. \tag{C.23}$$

Thus, the marginal likelihood from equation (C.21) is

$$\begin{aligned} m(D|\mathcal{M}_U) = \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)(T_0+k)+\frac{1}{2}k} \times \left[\frac{T}{T^*}\right]^{\frac{1}{2}(N+K-k)T} \times \\ \left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^*-1\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0-1\right]\right)}\right] \left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^*+N+K-k-1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0+N+K-k-1\right]\right)}\right] \times \\ \left[\frac{|R'R - R'F\left(F'F\right)^{-1}F'R|^{\frac{1}{2}(T_0-1)}}{|R'R - \tilde{\Phi}'W'W\tilde{\Phi}|^{\frac{1}{2}(T^*-1)}}\right] \left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N-1)}}{|T^*\hat{V}_f|^{\frac{1}{2}(T^*+N-1)}}\right]. \end{aligned} \tag{C.24}$$

Note that for the marginal likelihood for the unconditional and unrestricted case, equation (C.24) is equal to the conditional unrestricted case in equation (B.30) when the number of predictors m is zero.

C.2 Marginal Likelihood for the Restricted Case

The derivation of the marginal likelihood when restricting α to zero is closely related to the unrestricted case. The likelihood of the hypothetical sample $D_0 = (R_0, F_0)$ can be rewritten as

$$\begin{aligned} &\mathcal{L}(D_0|\Sigma_{RR}, \Sigma_{FF}, \Phi^R, A_F, \mathcal{M}_R) \propto \\ &\quad |\Sigma_{RR}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left[\Sigma_{RR}^{-1}\left(R_0'Q_{W_0^R}R_0 + \left(\Phi^R - \hat{\Phi}_0^R\right)' W_0^{R\prime}W_0^R\left(\Phi^R - \hat{\Phi}_0^R\right)\right)\right]\right\} \times \\ &\quad |\Sigma_{FF}|^{-\frac{1}{2}T_0} \exp\left\{-\frac{1}{2}\mathrm{Tr}\left[\Sigma_{FF}^{-1}\left(F_0'Q_{\iota_{T_0}}F_0 + \left(A_F - \hat{A}_F\right)' T_0\left(A_F - \hat{A}_F\right)\right)\right]\right\} \end{aligned} \tag{C.25}$$

where $\hat{\Phi}_0^R = \hat{\beta}_0$ and $W_0^R = F_0$.

The prior distributions for the parameters Σ_{FF} and A_F remain the same as in equations (C.7) and (C.8). The priors for Σ_{RR} and Φ^R are given by

$$\operatorname{vec}\left(\Phi^R\right)|\Sigma_{RR}, D_0 \sim f_{\text{MV}}\left(\operatorname{vec}\left(\hat{\Phi}_0^R\right), \Sigma_{RR} \otimes \left(W_0^{R\prime} W_0^R\right)^{-1}\right) \tag{C.26}$$

$$\Sigma_{RR}|D_0 \sim f_{\text{IW}}^{N+K-k}\left(R_0' Q_{W_0^R} R_0, T_0\right). \tag{C.27}$$

Combining the likelihood of the observed sample $D = (R, F)$ with the prior distributions based on the hypothetical sample yields the following posterior distributions for Φ^R and Σ_{RR}:

$$\operatorname{vec}\left(\Phi^R\right)|\Sigma_{RR}, D \propto f_{\text{MV}}\left(\operatorname{vec}\left(\tilde{\Phi}^R\right), \Sigma_{RR} \otimes \left(W_0^{R\prime} W_0^R + W^{R\prime} W^R\right)^{-1}\right) \tag{C.28}$$

$$\Sigma_{RR}|D \propto f_{\text{IW}}^{N+K-k}\left(\tilde{S}_R, T^*\right) \tag{C.29}$$

where

$$\tilde{\Phi}^R = \left(W_0^{R\prime} W_0^R + W^{R\prime} W^R\right)^{-1}\left(W^{R\prime} R + W_0^{R\prime} W_0^R \hat{\Phi}_0^R\right) = \hat{\Phi}_0^R = \hat{\beta}_0 \tag{C.30}$$

$$\begin{aligned}\tilde{S}_R =& R_0' Q_{W_0^R} R_0 + \left(R - W^R \tilde{\Phi}^R\right)'\left(R - W^R \tilde{\Phi}^R\right) + \left(\tilde{\Phi}^R - \hat{\Phi}_0^R\right)'\left(W_0^{R\prime} W_0^R\right)\left(\tilde{\Phi}^R - \hat{\Phi}_0^R\right) \\ =& \frac{T^*}{T_0} R_0' Q_{W_0^R} R_0.\end{aligned} \tag{C.31}$$

The marginal likelihood takes the form

$$\begin{aligned} m(D|\mathcal{M}_U^R) = \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)k + \frac{1}{2}k} \times \\ \left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}T^*\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}T_0\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^* + N + K - k - 1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0 + N + K - k - 1\right]\right)}\right] \times \\ \left[\frac{|R_0' Q_{W_0^R} R_0|^{\frac{1}{2}T_0}}{|\tilde{S}_R|^{\frac{1}{2}T^*}}\right]\left[\frac{|T_0 \hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-1)}}{|S_F|^{\frac{1}{2}(T^*+N+K-k-1)}}\right] \end{aligned}$$

$$
\begin{aligned}
&= \pi^{-\frac{1}{2}T(N+K)} \times \left[\frac{T}{T^*}\right]^{\frac{1}{2}(N+K-k)(T_0+k)+\frac{1}{2}k} \times \left[\frac{T_0}{T^*}\right]^{\frac{1}{2}(N+K-k)T} \times \\
&\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}T^*\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}T_0\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^*+N+K-k-1\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0+N+K-k-1\right]\right)}\right] \times \\
&\left[\frac{|R'R-R'F\left(F'F\right)^{-1}F'R|^{\frac{1}{2}T_0}}{|R'R-\tilde{\Phi}^{R\prime}W^{R\prime}W^{R}\tilde{\Phi}^{R}|^{\frac{1}{2}T^*}}\right]\left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+N+K-k-1)}}{|S_F|^{\frac{1}{2}(T^*+N+K-k-1)}}\right].
\end{aligned}
\tag{C.32}
$$

Note that for the marginal likelihood for the unconditional and restricted case, equation (C.32) is equal to the conditional and restricted case in equation (B.41) when the number of predictors m is zero.

D Summary of the Marginal Likelihood Calculations

The general formula for calculating the marginal likelihood is given by

$$
\begin{aligned}
m(D|\mathcal{M}) = \pi^{-\frac{1}{2}(T)(N+K)} &\times \left[\frac{T_0}{T^*}\right]^{\frac{Q_R}{2}}\left[\frac{T_0}{T^*}\right]^{\frac{Q_F}{2}} \times \\
&\left[\frac{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T^*+\nu_R\right]\right)}{\Gamma_{N+K-k}\left(\frac{1}{2}\left[T_0+\nu_R\right]\right)}\right]\left[\frac{\Gamma_k\left(\frac{1}{2}\left[T^*+\nu_F\right]\right)}{\Gamma_k\left(\frac{1}{2}\left[T_0+\nu_F\right]\right)}\right] \times \\
&\left[\frac{|R_0'Q_{W_0}R_0|^{\frac{1}{2}(T_0+\nu_R)}}{|S_R|^{\frac{1}{2}(T^*+\nu_R)}}\right]\left[\frac{|T_0\hat{V}_f|^{\frac{1}{2}(T_0+\nu_F)}}{|S_F|^{\frac{1}{2}(T^*+\nu_F)}}\right].
\end{aligned}
\tag{D.1}
$$

The parameters Q_R, Q_F, ν_R, and ν_F are determined based on the model specifications in Panel A of Table 1. The parameter values for each model are specified in Panel B of Table 1. For models $\mathbb{M}_1-\mathbb{M}_2$, we obtain

$$
S_F = T^*\hat{V}_f \tag{D.2}
$$

and for models $\mathbb{M}_3-\mathbb{M}_4$

$$
S_F = T^*\left(\hat{V}_f+\bar{f}\bar{f}'\right) - \frac{T}{T^*}\left(T_0[\bar{f},\bar{f}\bar{z}']+F'X\right)\left(X'X\right)^{-1}\left(T_0[\bar{f},\bar{f}\bar{z}']'+X'F\right). \tag{D.3}
$$

For all models

$$R_0' Q_{W_0} R_0 = \frac{T_0}{T}\left(R'R - \hat{\beta}_0' F' F \hat{\beta}_0\right) \tag{D.4}$$

$$S_R = \frac{T^*}{T}\left(R'R - \tilde{\Phi}' W' W \tilde{\Phi}\right), \tag{D.5}$$

where R, F, S_R, S_F and $\hat{V}_f$ vary according to the exact model specification and depend on the subgroup of the included/excluded factors and predictors.

E Invariance of the Marginal Likelihood under Linear Transformation of the Predictors

In this section, we show that given the following data generating process

$$r_{t+1} = \beta_0 f_{t+1} + \beta_1 z_t f_{t+1} + u_{r,t+1}, \tag{E.1}$$

the marginal likelihood given in equation (D.1) is invariant under a linear transformation of z_t and, in particular, standardization. To show this, we prove in the following that a linear transformation of z_t does not change S_R and S_F in equation (D.1).

We start with the term S_R. Without loss of generality and for ease of notation we assume that z_t and f_{t+1} in equation (E.1) are scalars. We rewrite equation (E.1) in matrix notation

$$\begin{aligned} R &= F\beta_0' + \Xi\beta_1' + U_R \\ &= W\Phi + U_R \end{aligned} \tag{E.2}$$

where $R = [r_1, \ldots, r_T]' \in \mathbb{R}^{T\times N}$, $F = [f_1, \ldots, f_T]' \in \mathbb{R}^T$, $\Xi = [\xi_1, \ldots, \xi_T]' \in \mathbb{R}^T$ is defined through $\xi_t = z_{t-1} f_t$, $W = [F, \Xi]$, and $\Phi = [\beta_0, \beta_1]'$. $U_R = [u_{r,1}, \ldots, u_{r,T}]' \in \mathbb{R}^{T\times N}$ is the matrix of residuals. The covariance matrix of the residuals is

$$S_R = R' \left(I_T - W \left(W'W \right)^{-1} W' \right) R, \tag{E.3}$$

where

$$\left(W'W \right)^{-1} = \frac{1}{T \left(\bar{f^2}\bar{\xi^2} - \overline{f\xi}^2 \right)} \begin{bmatrix} \bar{\xi^2} & -\overline{f\xi} \\ -\overline{f\xi} & \bar{f^2} \end{bmatrix}. \tag{E.4}$$

We have

$$\begin{aligned} S_R &= R' \left(I - W \left(W'W \right)^{-1} W' \right) R, \\ &= R' \left(I - \frac{1}{T \left(\bar{f^2}\bar{\xi^2} - \overline{f\xi}^2 \right)} \begin{bmatrix} F & \Xi \end{bmatrix} \begin{bmatrix} \bar{\xi^2} & -\overline{f\xi} \\ -\overline{f\xi} & \bar{f^2} \end{bmatrix} \begin{bmatrix} F' \\ \Xi' \end{bmatrix} \right) R, \\ &= R' \left(I - \frac{\left[\bar{\xi^2} F'F - 2\overline{f\xi}\Xi F' + \bar{f^2}\Xi\Xi' \right]}{T \left(\bar{f^2}\bar{\xi^2} - \overline{f\xi}^2 \right)} \right) R \end{aligned} \tag{E.5}$$

To set the stage, we start with a shift in z_t. In particular, let $\tilde{z}_t = z_t - a$, where a is a constant. Then

$$\tilde{\xi}_t = \tilde{z}_{t-1} f_t = \xi - af \tag{E.6}$$

$$\bar{\tilde{\xi}} = \bar{\xi} - a\bar{f} \tag{E.7}$$

$$\overline{f\tilde{\xi}} = \overline{f\xi} - a\bar{f^2} \tag{E.8}$$

$$\overline{\tilde{\xi}^2} = \bar{\xi^2} - 2a\overline{f\xi} + a^2\bar{f^2}. \tag{E.9}$$

For the transformed system, it follows that

$$\widetilde{S_R} = R' \left(I - \frac{\left[\overline{\tilde{\xi}^2} F'F - 2\overline{f\tilde{\xi}}\tilde{\Xi}F' + \bar{f}^2 \tilde{\Xi}\tilde{\Xi}' \right]}{T\left(\bar{f}^2 \overline{\tilde{\xi}^2} - \overline{f\tilde{\xi}}^2 \right)} \right) R, \tag{E.10}$$

where $\tilde{\Xi} = \Xi - aF$. Substituting equations (E.7), (E.8), and (E.9) into equation (E.10), we obtain

$$\widetilde{S_R} = R' \left(I - \frac{\left[\left(\bar{\xi^2} - 2a\overline{f\xi} + a^2 \bar{f^2} \right) F'F - 2\left(\overline{f\xi} - a\bar{f^2} \right) \left(\Xi - aF \right) F' + \bar{f^2} \left(\Xi - aF \right) \left(\Xi' - aF' \right) \right]}{T\left(\bar{f^2} \left(\bar{\xi^2} - 2a\overline{f\xi} + a^2 \bar{f^2} \right) - \left(\overline{f\xi} - a\bar{f^2} \right)^2 \right)} \right) R. \tag{E.11}$$

After some algebra and the cancellation of common terms, equation (E.11) equals equation (E.5).

Second, consider a scale transformation of z_t, namely $\tilde{z}_t = cz_t$ where c is a scalar. The coefficients β_1 in equation (E.1) change to $\tilde{\beta}_1 = \frac{1}{c}\beta_1$ and the covariance matrix of the residuals, S_R remains unchanged. This completes the proof for S_R.

The proof for S_F boils down from the above proof for S_R. It is a reduction of the data generating process in equation (E.1) that is achieved by setting $f_t \equiv 1$ and changing the notation r_t to f_t. This completes the proof that a linear transformation of the predictors z_t has no effect on the marginal likelihood in equation (D.1).

F Derivation of T_0

To address the choice of T_0, we establish an exact link between the variance of mispricing and T_0. This link has different forms depending on the time-varying nature of α and β. We distinguish between two scenarios: (i) both α and β are constant, namely both $\alpha_1 = 0$ and $\beta_1 = 0$ in equation (3), and (ii) both α and β are time varying. Given the distribution of Φ in the hypothetical sample (equation (B.9)), one needs to invert the full matrix $(W_0'W_0)$ given in equation (B.18) for the second scenario and a submatrix of for the first scenario. In the following, we provide the derivations for the two scenarios.

F.1 α and β are Constant

When $\alpha = \alpha_0$ and $\beta = \beta_0$, the matrix $W_0'W_0$ takes the form

$$W_0'W_0 = T_0 \begin{pmatrix} 1 & \bar{f}' \\ \bar{f} & \hat{V}_f + \bar{f}\bar{f}' \end{pmatrix} \tag{F.1}$$

and

$$(W_0'W_0)^{-1} = \frac{1}{T_0} \begin{pmatrix} 1 + \bar{f}'\hat{V}_f^{-1}\bar{f} & -\bar{f}'\hat{V}_f^{-1} \\ -\hat{V}_f^{-1}\bar{f} & \hat{V}_f^{-1} \end{pmatrix} \tag{F.2}$$

Therefore,

$$\begin{aligned} \mathrm{Var}(\alpha|\Sigma_{RR}, D_0) &= \mathrm{Var}\left(\alpha_0|\Sigma_{RR}, D_0\right) \\ &= \frac{\Sigma_{RR}}{T_0}\left(1 + \bar{f}'\hat{V}_f^{-1}\bar{f}\right) = \frac{\Sigma_{RR}}{T_0}\left(1 + SR_{max}^2\right), \end{aligned} \tag{F.3}$$

where SR_{max} is the Sharpe ratio of the tangency portfolio constructed by the factors.

Equating the variance of α in the hypothetical sample (equation (F.3)) with the prior's variance in equation (20) with the formulation of η from equation (23), we obtain

$$T_0 = \frac{(N + K - k)\left(1 + SR_{max}^2\right)}{\left(\tau^2 - 1\right) SR^2(Mkt)}, \tag{F.4}$$

where N is the number of test assets, $K - k$ is the number of redundant factors and

$SR^2(Mkt)$ is the squared Sharpe ratio of the market.

This exact relation assigns our prior specification the mispricing uncertainty interpretation. This prior specification is economically sound. In particular, when more factors are included in an asset pricing specification, the admissible squared Sharpe ratio essentially increases. Thus, the prior is more strongly weighted against mispricing and thereby tames the squared Sharpe ratio (based on the formulation in equation (21)). On the other hand, as τ increases, we allow for more arbitrage opportunities that translate into achieving a higher Sharpe ratio or a lower hypothetical sample.

F.2 α and β are Time Varying

When α and β are both time varying, we invert the matrix $(W_0'W_0)$ given in equation (B.18). We partition the matrix $(W_0'W_0)^{-1}$ into four blocks,

$$(W_0'W_0)^{-1} = \frac{1}{T_0}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{F.5}$$

where B_{11} is a $(1+m)\times(1+m)$ matrix, B_{21} is a $(1+m)\times(k+km)$, $B_{12} = B_{21}'$ and B_{22} is a $(k+km)\times(k+km)$ matrix. Specifically,

$$B_{11} = \begin{pmatrix} 1 + \bar{z}'\hat{V}_z^{-1}\bar{z} + \bar{f}'\hat{V}_f^{-1}\bar{f} + \bar{z}'\hat{V}_z^{-1}\bar{z} \times \bar{f}'\hat{V}_f^{-1}\bar{f} & -\bar{z}'\hat{V}_z^{-1} \times (1 + \bar{f}'\hat{V}_f^{-1}\bar{f}) \\ -\hat{V}_z^{-1}\bar{z} \times (1 + \bar{f}'\hat{V}_f^{-1}\bar{f}) & \hat{V}_z^{-1} \times (1 + \bar{f}'\hat{V}_f^{-1}\bar{f}), \end{pmatrix} \tag{F.6}$$

$$B_{21} = -\begin{pmatrix} (1 + \bar{z}'\hat{V}_z^{-1}\bar{z}) \times \hat{V}_f^{-1}\bar{f} & -\hat{V}_f^{-1}\bar{f} \otimes \bar{z}'\hat{V}_z^{-1} \\ -\hat{V}_f^{-1}\bar{f} \otimes \hat{V}_z^{-1}\bar{z} & \hat{V}_f^{-1}\bar{f} \otimes \hat{V}_z^{-1} \end{pmatrix} \tag{F.7}$$

$$B_{22} = \begin{pmatrix} \hat{V}_f^{-1} \times (1 + \bar{z}'\hat{V}_z^{-1}\bar{z}) & -\hat{V}_f^{-1} \otimes \bar{z}'\hat{V}_z^{-1} \\ -\hat{V}_f^{-1} \otimes \hat{V}_z^{-1}\bar{z} & \hat{V}_f^{-1} \otimes V_z^{-1} \end{pmatrix}. \tag{F.8}$$

If follows that the unconditional variance of total mispricing (sum of fixed and time varying) is then equal to

$$\begin{aligned}
\mathrm{Var}(\alpha|\Sigma_{RR}, D_0) &= \mathrm{Var}\left(\alpha_0 + \alpha_1' z|\Sigma_{RR}, D_0\right) \\
&= \frac{\Sigma_{RR}}{T_0} \mathrm{Tr}\left[B_{11} \begin{pmatrix} 1 & \bar{z}' \\ \bar{z} & \hat{V}_z + \bar{z}\bar{z}' \end{pmatrix} \right] \\
&= \frac{\Sigma_{RR}}{T_0} \mathrm{Tr}\left[\begin{pmatrix} 1 + \bar{f}'\hat{V}_f^{-1}\bar{f} & \dots \\ \dots & I_{mxm}\left(1 + \bar{f}'\hat{V}_f^{-1}\bar{f}\right) \end{pmatrix} \right] \\
&= \frac{\Sigma_{RR}}{T_0} \left(1 + \bar{f}'\hat{V}_f^{-1}\bar{f} + m(1 + \bar{f}'\hat{V}_f^{-1}\bar{f})\right) \\
&= \frac{\Sigma_{RR}}{T_0} \left(1 + SR_{max}^2 + m(1 + SR_{max}^2)\right),
\end{aligned} \tag{F.9}$$

where Tr stands for the trace operator and SR_{max} is the Sharpe ratio of the tangency portfolio constructed by the factors and m is the number of predictors in the model.

Equating the unconditional variance of α in the hypothetical sample (equation (F.9)) with the prior's variance in equation (20) and with the formulation of η from equation (23), we obtain

$$T_0 = \frac{(N + K - k)\left(1 + SR_{max}^2 + m(1 + SR_{max}^2)\right)}{(\tau^2 - 1)\, SR^2(Mkt)}, \tag{F.10}$$

where N is the number of test assets, $K - k$ is the number of redundant factors and $SR^2(Mkt)$ is the squared Sharpe ratio of the market. This exact relation assigns our prior specification the mispricing uncertainty interpretation.

G Bayesian Predictive Moments

Starting from equation (2) and denoting the posterior means of the parameters θ with $\tilde{\theta}$, the mean and variance of the predicted factor returns for model $\mathcal{M}_l$ are

$$E\left[f_{t+1}|D_t, \mathcal{M}_l\right] = \tilde{\alpha}_{l,f} + \tilde{a}_{l,F} z_t = \hat{f}_{l,t+1}, \tag{G.1}$$

$$\begin{aligned}
\operatorname{Var}\left[f_{t+1}|D_t, \mathcal{M}_l\right] &= \left[I_k \otimes x_t'\right] \left[\tilde{\Sigma}_{FF} \otimes \frac{t}{t^*} \left(X'X\right)^{-1}\right] \left[I_k \otimes x_t'\right]' + \tilde{\Sigma}_{FF} \\
&= \frac{1}{t^*} \left[I_k \otimes x_t'\right] \left[\tilde{\Sigma}_{FF} \otimes \begin{pmatrix} 1 + \bar{z}'\hat{V}_z^{-1}\bar{z} & -\bar{z}'\hat{V}_z^{-1} \\ -\hat{V}_z^{-1}\bar{z} & \hat{V}_z^{-1} \end{pmatrix}\right] \left[I_k \otimes x_t\right] + \tilde{\Sigma}_{FF} \\
&= \frac{1}{t^*} \tilde{\Sigma}_{FF} \otimes x_t' \begin{pmatrix} 1 + \bar{z}'\hat{V}_z^{-1}\bar{z} & -\bar{z}'\hat{V}_z^{-1} \\ -\hat{V}_z^{-1}\bar{z} & \hat{V}_z^{-1} \end{pmatrix} x_t + \tilde{\Sigma}_{FF} \\
&= \tilde{\Sigma}_{FF} \left(1 + \frac{\left(1 + \left(\bar{z} - z_t\right)' \hat{V}_z^{-1} \left(\bar{z} - z_t\right)\right)}{t^*}\right) \\
&= \tilde{\Sigma}_{FF} \left(1 + \delta_{t,t^*}\right),
\end{aligned} \tag{G.2}$$

where $x_t = [1, z_t']'$, t is the sample size, t_0 is the hypothetical sample size corresponding to the sample t, $t^* = t + t_0$, $\delta_{t,t^*} = \frac{\left(1+(\bar{z}-z_t)'\hat{V}_z^{-1}(\bar{z}-z_t)\right)}{t^*}$, $\tilde{\Sigma}_{FF} = \frac{S_F}{t^*+N+K-2k-m-2}$ and the matrix S_F is defined in equation (B.25).

From equation (3), the mean and variance of the predicted future returns for the test assets for model $\mathcal{M}_l$ are

$$E\left[r_{t+1}|D_t, \mathcal{M}_l\right] = \tilde{\alpha}_{l,0} + \tilde{\alpha}_{l,1} z_t + \left(\tilde{\beta}_{l,0} + \tilde{\beta}_{l,1}(I_k \otimes z_t)\right) \left(\tilde{\alpha}_{l,f} + \tilde{a}_{l,F} z_t\right), \tag{G.3}$$

$$\begin{aligned}
\operatorname{Var}\left[r_{t+1}|D_t, \mathcal{M}_l\right] &= \left[I_{N+K-k} \otimes \hat{w}_{t+1}'\right] \left[\tilde{\Sigma}_{RR} \otimes \frac{t}{t^*} \left(W'W\right)^{-1}\right] \left[I_{N+K-k} \otimes \hat{w}_{t+1}'\right]' \\
&\quad + \left(1 + \delta_{t,t^*}\right) \left(\tilde{\beta}_{l,0} + \tilde{\beta}_{l,1}(I_k \otimes z_t)\right) \tilde{\Sigma}_{FF} \left(\tilde{\beta}_{l,0} + \tilde{\beta}_{l,1}(I_k \otimes z_t)\right)' + \tilde{\Sigma}_{RR}
\end{aligned}$$

$$
\begin{aligned}
&= \tilde{\Sigma}_{RR}\left(1+\frac{t}{t^*}\hat{w}'_{t+1}\left(W'W\right)^{-1}\hat{w}_{t+1}\right) \\
&+ (1+\delta_{t,t^*})\left(\tilde{\beta}_{l,0}+\tilde{\beta}_{l,1}(I_k\otimes z_t)\right)\tilde{\Sigma}_{FF}\left(\tilde{\beta}_{l,0}+\tilde{\beta}_{l,1}(I_k\otimes z_t)\right)', \qquad \text{(G.4)}
\end{aligned}
$$

where $\hat{w}_{t+1} = \left[1, z'_t, \hat{f}'_{t+1}, \hat{f}'_{t+1}\otimes z'_t\right]$, $\tilde{\Sigma}_{RR} = \frac{\tilde{S}_R}{t^*-(k+1)m-N-K+k-2}$ and the matrix $\tilde{S}_R$ is defined in equation (B.26).

In the restricted setup, both $\alpha_0 = 0$ and $\alpha_1 = 0$ in equation (3). The mean and variance of the predicted future returns for the test assets for model $\mathcal{M}_l$ are

$$
E\left[r_{t+1}|D_t,\mathcal{M}_l\right] = \left(\tilde{\beta}_{l,0}+\tilde{\beta}_{l,1}(I_k\otimes z_t)\right)\left(\tilde{\alpha}_{l,f}+\tilde{a}_{l,F}z_t\right), \qquad \text{(G.5)}
$$

$$
\begin{aligned}
\operatorname{Var}\left[r_{t+1}|D_t,\mathcal{M}_l\right] &= \left[I_{N+K-k}\otimes\hat{w}^{R\prime}_{t+1}\right]\left[\tilde{\Sigma}_{RR}\otimes\frac{t}{t^*}\left(W'W\right)^{-1}\right]\left[I_{N+K-k}\otimes\hat{w}^{R\prime}_{t+1}\right]' \\
&+ (1+\delta_{t,t^*})\left(\tilde{\beta}_{l,0}+\tilde{\beta}_{l,1}(I_k\otimes z_t)\right)\tilde{\Sigma}_{FF}\left(\tilde{\beta}_{l,0}+\tilde{\beta}_{l,1}(I_k\otimes z_t)\right)' + \tilde{\Sigma}_{RR} \\
&= \tilde{\Sigma}_{RR}\left(1+\frac{t}{t^*}\hat{w}^{R\prime}_{t+1}\left(W^{R\prime}W^R\right)^{-1}\hat{w}^{R\prime}_{t+1}\right) \\
&+ (1+\delta_{t,t^*})\left(\tilde{\beta}_{l,0}+\tilde{\beta}_{l,1}(I_k\otimes z_t)\right)\tilde{\Sigma}_{FF}\left(\tilde{\beta}_{l,0}+\tilde{\beta}_{l,1}(I_k\otimes z_t)\right)', \qquad \text{(G.6)}
\end{aligned}
$$

where $\hat{w}^R_{t+1} = \left[\hat{f}'_{t+1}, \hat{f}'_{t+1}\otimes z'_t\right]'$, $\tilde{\Sigma}_{RR} = \frac{\tilde{S}_R}{t^*-km-N-K+k-1}$ and the matrix $\tilde{S}_R$ is defined in equation (B.40).