Model Uncertainty in the Cross Section*

Jiantao Huang† Ran Shi‡

December 2023

Abstract

We develop a transparent Bayesian framework to measure uncertainty in asset pricing models. Our framework quantifies the tradeoff between mean-variance efficiency and parsimony for models to attain high posterior probabilities. Model uncertainty is defined as the entropy of these posterior probabilities, which is consistently interpretable even under misspecification due to omitted factors. Empirically, model uncertainty accumulates during major market events, carrying a significantly negative risk premium of approximately half the magnitude of the market. Positive shocks to model uncertainty predict persistent outflows from US equity funds and inflows to Treasury funds.

Keywords: Model Uncertainty, Asset Pricing Factor, Bayesian Inference, Model Selection Consistency, Omitted Factors, Mutual Fund Flows
JEL Classification Codes: C11, G11, G12.

* For helpful comments and discussions, we thank Svetlana Bryzgalova, Thummim Cho, Guanhao Feng, Shiyang Huang, Christian Julliard, Dong Lou, Ian Martin, Paul Schneider, Gustavo Schwenkler, Fabio Trojani, and conference participants at LSE, SFS Cavalcade 2022, Asian Meeting of Econometric Society 2022, 8th HK Joint Finance Research Workshop, NBER-NSF SBIES 2022, and SoFiE 2023 Conference. Any errors or omissions are the responsibility of the authors.

†Faculty of Business and Economics, the University of Hong Kong, huangjt@hku.hk.
‡University of Colorado Boulder, ran.shi@colorado.edu.

Conflict-of-interest disclosure statement

Jiantao Huang:

I have nothing to disclose

Ran Shi:

I have nothing to disclose

Introduction

Uncertainty is a crucial concern of market participants. Existing research, exemplified by works such as Bloom (2009) and Ludvigson, Ma, and Ng (2021), primarily employs implied or realized volatilities of economic variables to measure uncertainty in financial markets. However, another type of uncertainty arises when investors decide on the appropriate asset pricing models to use for estimating the expected returns of assets in their portfolios. This source of uncertainty, referred to as model uncertainty, raises several questions: To what extent does model uncertainty exist in the cross-section of asset returns? Does model uncertainty change over time? Is model uncertainty priced? Does it contain useful information related to investors' asset allocation decisions?

To answer these questions, we propose a new Bayesian approach to measuring uncertainty in asset pricing models. Applying this approach, we examine the US and international equity markets and confirm the prevalence of model uncertainty. The shocks to model uncertainty carry a negative risk premium of approximately half the magnitude of the market portfolio. These model uncertainty shocks are also informative about flows to both equity and government bond funds one to three years ahead. In contrast, volatility-based uncertainty measures do not convey such information.

In particular, we compute a closed-form (posterior) probability for each candidate asset pricing model and measure model uncertainty with the entropy of these probabilities. To illustrate the basic intuition, consider two candidate models. An extreme case is that one model strongly dominates the other with a probability of being "correct" near one. Under this low-uncertainty scenario, the entropy is close to its lower bound of zero. Conversely, if the two models' probabilities of being supported by the data are 50-50, choosing a model reduces to an exercise of tossing a fair coin. Model uncertainty is the highest in this case, and correspondingly, the entropy measure reaches its maximum.

There is a strong and commonly adopted—whenever *a* particular asset pricing model is chosen—null hypothesis about model uncertainty in the cross-section of returns. It is the case in which all the expected returns align with an "oracle" factor model governing the true data-generating process. No model uncertainty exists under this null, and a reliable measure should not falsely report any degree of uncertainty.

Our measure has this property and will report a zero entropy (of posterior model probabilities). When the set of candidate models includes the true model, our Bayesian inference guarantees model selection consistency and will tend to assign a posterior probability of one to this model (and zero to the others). The resulting model uncertainty will approach zero. Even when none of the candidate models are true—due to the initial failure to include all the true factors in our analysis—our measure will remain bounded within a specific limit, far from its theoretical maximum. This limit quantifies the highest degree of contamination that can occur due to model misspecification.[1]

We construct model uncertainty indices for the US, European, and Asia-Pacific stock markets based on commonly studied equity return factors and observe consistently high model uncertainty across these markets. Recall that our measures should stay remain low if a true model exists, even when the models are potentially misspecified. Therefore, the documented empirical patterns offer consistent evidence that there is a significant degree of model uncertainty when it comes to understanding the cross-section of equity returns.

Our uncertainty measures exhibit large variation over time and tend to spike during major market events. In the US equity market, model uncertainty increases before the recessions in the early 1980s and 1990s and reaches its (theoretical) upper limit during the technology bubble and the global financial crisis. Interestingly, model uncertainty in the Asia-Pacific market is uniquely high during the 1997 financial crisis in that region.

[1]To clarify, our method is not designed for correcting the biases induced by those misspecified models. Instead, by acknowledging the possibility that our proposed measure could be gauging model uncertainty plus model misspecification, we offer novel insights into the highest level of contamination that misspecified models can introduce to our measure.

As model uncertainty tends to escalate during market downturns, we expect that this source of uncertainty, if priced in the equity markets, will carry a negative risk premium—and we confirm this in the data. To do this, we extract AR(1) shocks to model uncertainty, normalized to have unit standard deviation. Then, we construct mimicking portfolios that track these model uncertainty shocks and calculate the average returns of these portfolios, which can be interpreted as the risk premium estimates of model uncertainty. The risk premia are highly significant and almost the same in the US and European stock markets, being at least -7.8% per year once we equalize the standard deviations of model uncertainty shocks to the market portfolio (for the ease of interpretation). The same risk premium estimate is -5.7% to -2.9% in the Asia-Pacific market depending on the construction of the mimicking portfolios, while not necessarily being significant. In addition, we apply the methodology of Giglio and Xiu (2021) to address omitted variable bias when estimating the risk premia of nontradable factors. In the US market, even after adjusting for such bias, the risk premium of model uncertainty shocks remains highly significant, although the magnitude drops to approximately -3.8% per year.

Moving from pricing implications to quantities, we also find that shocks to model uncertainty carry useful information about investors' asset allocation decisions. Precisely, these shocks predict *aggregate* flows to both equity and fixed-income funds, which we treat as proxies for investors' asset allocation decisions across different asset classes. We estimate the dynamic responses of fund flows to model uncertainty shocks under a simple vector autoregression (VAR) framework. Throughout our VAR exercises, we investigate two possibilities: Model uncertainty can be an exogenous cause or merely a propagation channel. Under both settings, positive model uncertainty shocks forecast sharp outflows from US equity funds and inflows to treasury funds, with effects persisting for approximately three years. The equity outflows only concern small-cap and style funds (funds that commit to certain investment styles such as growth or value), not large-cap or sector funds. These findings are consistent

with the "flight-to-quality" prediction:[2] When confronted with heightened model uncertainty in equity markets, investors tend to reduce their exposure to risky assets (by withdrawing from small-cap and style funds) and instead allocate their resources to safe assets such as government bonds.

While model uncertainty mildly correlates with other volatility-based uncertainty measures, such as the Cboe VIX index and the financial uncertainty measure proposed by Ludvigson et al. (2021), it stands out as a distinct and valuable predictor of aggregate mutual fund flows. Notably, we do not observe any significant responses in the flows of equity and fixed-income fund flows to these other uncertainty shocks. Our fund flow exercises resonate with the recent literature that examines asset pricing models using variables beyond returns.[3]

Our investigation into model uncertainty makes several methodological contributions. First and foremost, we lay out the basic criteria that any model uncertainty measurement or, more generally, Bayesian inference about asset pricing models, must satisfy. These criteria highlight the need to evaluate the *frequentist* properties—such as consistency and robustness—of Bayesian methods.

In particular, our method starts by adapting Zellner's g-priors (Zellner, 1986)—the "benchmark" priors for Bayesian inference in the context of model uncertainty[4]—to linear stochastic discount factor (SDF) models. We assign g-priors to the risk prices of factors entering the SDF, and let the risk prices of other factors be absolutely zero. Our focus on

[2]The explanations for the "flight-to-quality" phenomenon include institutional redemption pressures (Vayanos, 2004), preferences featuring robustness concerns (Caballero and Krishnamurthy, 2008), and asymmetric information (Guerrieri and Shimer, 2014).

[3]See Barber, Huang, and Odean (2016); Berk and Van Binsbergen (2016); Ben-David, Li, Rossi, and Song (2022) for the use of fund flows and Chinco, Hartzmark, and Sussman (2022) for the use of survey responses.

[4]See (Fernández, Ley, and Steel, 2001) for reference. The use of g-priors in the finance literature dates back to Kandel and Stambaugh (1996). As a class of proper priors, assigning g-priors to model-specific parameters is immune to the posterior indeterminacy issue elaborated in Chib, Zeng, and Zhao (2020) (in the finance literature, Cremers (2002), citing the seminal work of Kass and Raftery (1995), has also pointed out this issue). The g-priors are first introduced for regression analysis. We tailor them to make inferences about linear SDFs under model uncertainty. Our adaptation yields closed-form posteriors, which are also invariant to rotation transformations of the testing assets.

the SDF models (as opposed to the beta pricing models) is motivated by the observation that factors excluded from the SDF must have zero risk prices, but can still carry nonzero risk premia.

Our adapted g-priors for the risk prices yield closed-form posterior model probabilities that increase with model-implied maximal in-sample Sharpe ratios and decrease with model dimensions. This result crystallizes two contradicting criteria for an asset pricing model to be chosen by our Bayesian investor: higher mean-variance efficiency and model simplicity.

Despite its clean and intuitive economic implications, the g-priors suffer from model selection inconsistency. Even after observing a long history of return data generated from a fixed true model (*ex ante* the lowest uncertainty of being zero), the investor always fails to assign a high probability to this model *ex post*, leading to artificially inflated model uncertainty.

We propose a solution to this issue by resorting to economic restrictions. On the one hand, the SDF must be volatile enough to generate the sizable risk premia observed in the data. On the other hand, the SDF volatility should always be bounded from above by the maximal in-sample Sharpe ratios (Hansen and Jagannathan, 1991). We show that, to satisfy both conditions, the prior must assign large volatilities to the market prices of risk – equivalently, the SDF coefficients – and induce "mild" posterior shrinkage of these coefficients toward zero. These considerations are all reflected in the restrictions on the g parameter, which we incorporate into a (hyper-)prior.

Perhaps the most appealing outcome after incorporating these economic restrictions is that model selection consistency is restored. Moreover, even when every model we study is misspecified but some of the true factors are included in these models, our method can still consistently select the (subset of) true factors. The resulting model uncertainty measure will be bounded far from its theoretical maximum. The direct implication of this property is that extremely high model uncertainty, as observed in the data, cannot be caused by model misspecifications alone.

Additionally, we examine the normative question of how a Bayesian investor should manage model uncertainty. Specifically, we test whether combining models can deliver higher out-of-sample Sharpe ratios than selecting the highest probability model or dogmatically believing in canonical factor models. After equally splitting the US equity return sample according to our model uncertainty series, we find that the average SDF combining both "strong" and "weak" models significantly outperforms *only* in the high model uncertainty subsample. In this subsample, investors' updated beliefs regarding different models tend to equalize. Thus, ignoring weakly dominated models can be particularly detrimental to portfolio choices.

Literature. Our paper primarily contributes to the increasing interest in developing time-series uncertainty measures for real activities, financial markets, and economic policy (Bloom, 2009; Jurado et al., 2015; Baker et al., 2016; Manela and Moreira, 2017; Ludvigson et al., 2021). On the cross-sectional dimension, Dew-Becker and Giglio (2021), borrowing information across firms, present a uncertainty measure from option prices. Hassan et al. (2019) create a panel of firm-specific political uncertainty measures through textual analysis of earnings conference calls. Compared with these papers, our model uncertainty measure is conceptually new. It quantifies equity investors' uncertainty about which asset pricing models better describe the cross-section of expected returns.

Our paper also contributes to the literature on Bayesian inference about asset pricing models and Bayesian portfolio choices (Shanken (1987); Harvey and Zhou (1990); Pástor (2000); Pástor and Stambaugh (2000); Barillas and Shanken (2018); Kozak et al. (2020); Chib et al. (2020); Chib and Zeng (2020); Chib et al. (2023); Bryzgalova et al. (2023); Avramov et al. (2023)).[5] We contribute to this by proposing a new prior that incorporates meaningful economic restrictions to facilitate Bayesian inference *specifically* about asset pricing model

[5]Like our paper, Avramov, Cheng, Metzker, and Voigt (2023) also define and analyze uncertainty in the context of beta pricing models. Their paper studies the contribution of model uncertainty to better volatility and covariance estimates, while our primary focus is on developing a solid measure of uncertainty about asset pricing models and examining its economic implications.

uncertainty. We conduct thorough econometric analysis and simulation studies to highlight many novel and important properties of our approach, including the interpretable posterior probabilities, model selection consistency, and robustness to model misspecification.

The empirical findings of this paper offer new insights into drivers of mutual fund flows along the extensive margin. Ben-David et al. (2022) report a robust relationship between aggregate fund flows and (unadjusted) lagged fund returns, which implies that investors have simplistic return-chasing behavior. In our analysis, model uncertainty in the cross-section consistently emerges as a strong predictor of aggregate equity fund flows after controlling for past fund returns. Our findings suggest that investors, when losing the sense of a reliable benchmark model of equity markets, tend to cut equity positions, which is partially reflected in their mutual fund holdings.

1 Measuring Model Uncertainty: Theory and Method

Throughout our analysis, we focus on the cross-section of *excess* returns and their risk premia. Denote by $\boldsymbol{R}$ a random vector of dimension N, the excess returns under consideration. A subset of these excess returns would be regarded as asset pricing factors that drive the whole cross-section of $\mathbb{E}[\boldsymbol{R}]$. We use the notation $\boldsymbol{f}$, a random vector of dimension p ($p \leq N$), to represent these factors.[6] A linear factor model for excess returns in the SDF form can be written as (see Chapter 13 of Cochrane (2005) for a detailed exposition):

$$m = 1 - (\boldsymbol{f} - \mathbb{E}[\boldsymbol{f}])^\top \boldsymbol{b}, \tag{1}$$

where m is an SDF such that the prices of excess returns all equal zero, i.e., $\mathbb{E}[\boldsymbol{R} \cdot m] = \boldsymbol{0}$,[7] elements of $\boldsymbol{b}$ are market prices of risk for the factors, and the portfolio $\boldsymbol{b}^\top \boldsymbol{f}$ defines the

[6] We intentionally let the factors $\boldsymbol{f}$ be a subset of excess returns $\boldsymbol{R}$ to enforce that factors themselves are correctly priced.

[7] We are not assuming that m is *the* SDF for all security markets, acknowledging the possibility of market segmentation and cross-market arbitrage limits.

tangency portfolio.

The asset pricing model implied by the SDF is then

$$\mathbb{E}[\boldsymbol{R}] = \text{cov}[\boldsymbol{R}, \boldsymbol{f}]\boldsymbol{b}, \tag{2}$$

where the covariance term, $\text{cov}[\boldsymbol{R}, \boldsymbol{f}]$, is an $N \times p$ matrix.

We now formalize the concept of model uncertainty. Without knowing which of the p factors in $\boldsymbol{f} = (f_1, \ldots, f_p)^\top$ enter the SDF, a total number of 2^p models are possible candidates. To capture uncertainty regarding this collection of models, we index the whole set of 2^p models using a p-dimensional vector of indicator variables $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^\top$, with $\gamma_j = 1$ representing that factor f_j is included in the linear SDF and $\gamma_j = 0$ meaning that f_j is excluded. This vector $\boldsymbol{\gamma}$ uniquely defines a model for the SDF, denoted by $\mathcal{M}_{\boldsymbol{\gamma}}$: Under $\mathcal{M}_{\boldsymbol{\gamma}}$, the linear SDF is

$$m_{\boldsymbol{\gamma}} = 1 - \left(\boldsymbol{f}_{\boldsymbol{\gamma}} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\right)^\top \boldsymbol{b}_{\boldsymbol{\gamma}}, \tag{3}$$

and the resulting asset pricing model becomes

$$\mathbb{E}[\boldsymbol{R}] = \text{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}}. \tag{4}$$

The two equations above are counterparts of (1) and (2) after incorporating model uncertainty. We define $p_{\boldsymbol{\gamma}} = \sum_{j=1}^{p} I_{\{\gamma_j = 1\}}$, the number of factors that are included in model $\mathcal{M}_{\boldsymbol{\gamma}}$. $\boldsymbol{f}_{\boldsymbol{\gamma}}$ is a $p_{\boldsymbol{\gamma}}$-dimensional vector concatenating all factors that are included under $\mathcal{M}_{\boldsymbol{\gamma}}$; the elements of $\boldsymbol{b}_{\boldsymbol{\gamma}} \in \mathbb{R}^{p_{\boldsymbol{\gamma}}}$ are market prices of risk for the *included* factors — thus, by default, excluded factors have zero market prices of risk.

Remark. Another object of interest is factors' risk premia $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)^\top$. Under model $\mathcal{M}_{\boldsymbol{\gamma}}$, $\boldsymbol{\lambda} = \text{cov}[\boldsymbol{f}, \boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}}$. Clearly, factors that *do not* enter the SDF (their risk prices being zero) can carry nonzero risk premia. Knowing whether factors' risk premia equal zero does not help distinguish SDF models (Cochrane, 2005, Page 261).

1.1 Prior Specification and Bayesian Inference

We now present a Bayesian framework to define and analyze model uncertainty in the cross-section of expected stock returns. Conditional on model $\mathcal{M}_{\boldsymbol{\gamma}}$, we assume that the excess returns follow the data-generating process restricted by the moment condition (4):

$$\boldsymbol{R}_t = \boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{b}_{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon}_t, \tag{5}$$

where $\boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$; $\boldsymbol{C}_{\boldsymbol{\gamma}} = \text{cov}[\boldsymbol{R}_t, \boldsymbol{f}_{\boldsymbol{\gamma},t}]$ is a principal submatrix of $\boldsymbol{\Sigma}$ (since $\boldsymbol{f}_{\boldsymbol{\gamma}} \subseteq \boldsymbol{R}$, the factors are a subset of test assets).[8] The tuple of return data $\{\boldsymbol{R}_t\}_{t=1}^T$ is denoted by $\mathcal{D}$.

1.1.1 Baseline specification: the g-priors

Our prior specification for the risk prices $\boldsymbol{b}_{\boldsymbol{\gamma}}$ is motivated by Zellner's g-priors (Zellner, 1986), which have been proposed as the "benchmark priors" for Bayesian inference under model uncertainty (Fernández et al., 2001). We adapt the canonical g-priors for linear regressions to our asset pricing setting as follows: conditional on model $\mathcal{M}_{\boldsymbol{\gamma}}$, the factor risk prices

$$\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}} \sim \mathcal{N}\left(\mathbf{0}, \frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right), \quad g > 0, \tag{6}$$

where T is the sample size of the observed excess returns, and the parameter g reflects the prior level of uncertainty about the risk prices.[9]

Under our g-prior specification, we can integrate out the risk prices $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and calculate the

[8] We work under an empirical Bayes framework and treat $\boldsymbol{\Sigma}$ as known initially. We then replace $\boldsymbol{\Sigma}$ with its consistent estimator in the expressions of posterior probabilities. Section IA.1 in Internet Appendix provides additional details of this approach and its associated econometric properties.

[9] One might attempt to assign uninformative priors, such as the Jeffreys priors, to the risk prices. These priors, often being improper ones, can only be assigned to *common* parameters across models. Otherwise, posterior probabilities can be indeterminate. This has been noted in the finance literature (Cremers, 2002; Chib et al., 2020). With the intention to "transform" improper priors into proper ones, O'Hagan (1995) and Berger and Pericchi (1996) advocate the use of Zellner's g-priors to compute the so-called factional or intrinsic Bayes factors.

marginal likelihood in closed form, as summarized in the following proposition.

Proposition 1. *The marginal likelihood for the excess returns $\mathcal{D}$ under model $\mathcal{M}_{\boldsymbol{\gamma}}$ is*

$$\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] = \exp\left\{-\frac{T}{2}\left(\mathrm{SR}^2_{\max} - \frac{g}{1+g}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right)\right\}(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}} \times \text{constant} \tag{7}$$

where $\mathrm{SR}^2_{\max}$ is the maximal squared Sharpe ratio of all testing assets $\boldsymbol{R}$;[10] *$\mathrm{SR}^2_{\boldsymbol{\gamma}}$ is the maximal squared Sharpe ratio of the factors $\boldsymbol{f}_{\boldsymbol{\gamma}}$ included under model $\mathcal{M}_{\boldsymbol{\gamma}}$. These two Sharpe ratios are both in-sample quantities.*

1.1.2 Economic implications

To help build intuition, we will first discuss the asset pricing implications of the g-prior specification. We will focus on comparing asset pricing models and learning about the SDFs. Based on the implications, we next point out the issues with the g-priors if meaningful economic restrictions are imposed. Finally, we provide a solution to resolve these issues.

Comparing models. The closed-form marginal likelihood function in Proposition 1 enables straightforward interpretations when evaluating the support of return data $\mathcal{D}$ for different models. Under Proposition 1, we can calculate the Bayes factor for two linear factor models, namely $\mathcal{M}_{\boldsymbol{\gamma}}$ and $\mathcal{M}_{\boldsymbol{\gamma}'}$, as the ratio of their marginal likelihoods:

$$\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = \frac{\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}]}{\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}'}]} = \exp\left\{\frac{g}{2(1+g)}\left(T\mathrm{SR}^2_{\boldsymbol{\gamma}} - T\mathrm{SR}^2_{\boldsymbol{\gamma}'}\right) - \frac{\log(1+g)}{2}(p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}'})\right\}. \tag{8}$$

A large Bayes factor $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ indicates that the observed data lend stronger support for model $\mathcal{M}_{\boldsymbol{\gamma}}$. A special configuration of the Bayes factor is to compare $\mathcal{M}_{\boldsymbol{\gamma}}$ against the null model $\mathcal{M}_{\mathbf{0}}$ (the "numeraire" model), under which risk prices are all zeros, and the SDF is a

[10] Throughout the rest of the paper, we often use the term "Sharpe ratio" in short to refer to the maximal in-sample squared Sharpe ratio, whenever it does not cause confusion.

constant (characterizing a "risk-neutral" economy):

$$\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0}) = \exp\left\{\frac{g}{2(1+g)} T \mathrm{SR}_{\boldsymbol{\gamma}}^2 - \frac{\log(1+g)}{2} p_{\boldsymbol{\gamma}}\right\}. \tag{9}$$

Although the marginal likelihood in Proposition 1 depends on test assets through $\mathrm{SR}_{\max}^2$, the Bayes factors do not. They are determined solely by the set of factors entering the asset pricing model. This outcome is driven by the requirement that factors themselves must be correctly priced. It is reminiscent of the observation that, when estimating factor models, efficient GMM objective functions should assign zero weights to test assets that are not factors entering the SDF (see, for example, Cochrane (2005, Page 244-245)).

Bayes factors under the g-prior specification favor parsimonious factor models associated with large in-sample Sharpe ratios. Ignoring model uncertainty (i.e., considering a particular model of the predetermined dimension $p_{\boldsymbol{\gamma}}$), return samples will always favor models associated with larger Sharpe ratios. This echoes the intuitions behind the GRS tests (see, for example, Gibbons et al. (1989) and Barillas et al. (2020)), which interpret time-series tests of factor models as evaluating the mean-variance efficiency of factor portfolios.

Learning about the SDF. The g-prior captures investors' belief updating about the volatility of the SDF, under a given model $\mathcal{M}_{\boldsymbol{\gamma}}$, summarized in Proposition 2.

Proposition 2. *Conditional on model $\mathcal{M}_{\boldsymbol{\gamma}}$, the g-prior for factor risk prices implies*

$$\mathrm{var}\left[m_{\boldsymbol{\gamma}}\right] = \frac{g p_{\boldsymbol{\gamma}}}{T};$$

after observing the return sample, the posterior variance of the SDF is updated to

$$\mathrm{var}\left[m_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] = \left(\frac{g}{1+g}\right)^2 \mathrm{SR}_{\boldsymbol{\gamma}}^2 + \left(\frac{g}{1+g}\right)\frac{p_{\boldsymbol{\gamma}}}{T}.$$

The volatility of the SDF has clear economic implications. As the pricing model (4) is

equivalent to $\mathbb{E}[R] = \text{cov}[m_{\boldsymbol{\gamma}}, R]$ for any excess return R, we can see that[11]

$$\text{var}[m_{\boldsymbol{\gamma}}] \geq \max_{\{\text{all assets}\}} \frac{\mathbb{E}^2[R]}{\text{var}[R]} \equiv \text{SR}_\infty^2.$$

That is, the volatility of the SDF sets an upper bound on any achievable Shape ratios in the economy (Hansen and Jagannathan, 1991). In the meantime, assets with very high Sharpe ratios ought to be "deals" that are too good to be true (Cochrane and Saa-Requejo, 2000; Kozak et al., 2020). This motivates us to further impose the good-deal bound that $\text{var}[m_{\boldsymbol{\gamma}}] \leq h^2$, where h represents the Sharpe ratio above which a marginal investor will *always* take, in the language of Cochrane and Saa-Requejo (2000). Summing up, the volatility of the SDF should fall in the interval $[\text{SR}_\infty, h]$.

Under Proposition 2, the g-priors connect the *prior* volatility of the SDF to the model dimension. Thus, models with too many factors are unlikely to be realistic *a priori*.

After observing the return data, the *posterior* volatility of the SDF is a weighted average of the (observed) maximal in-sample squared Sharpe ratio and the prior volatility. An increased g, which reflects larger parameter uncertainty about the risk prices $\boldsymbol{b}_{\boldsymbol{\gamma}}$ according to (6), leads to a higher posterior weight on the in-sample Sharpe ratios.

Issues with the g-priors. The economic implications from Proposition 2 put a restriction on the parameter g. That is, even before observing the return data $\mathcal{D}$, we should expect the prior volatility of the SDF $gp_{\boldsymbol{\gamma}}/T \in [\text{SR}_\infty, h]$. As a result, we must have $g = O(T)$.

Let us now revisit our model comparison results. Under (8), the hurdle for including one more factor into the SDF must be such that the improvement in the squared Sharpe ratio, namely ΔSR^2, satisfies

$$\Delta\text{SR}^2 \geq \frac{\log(1+g)}{T}\frac{1+g}{g}.$$

The right-hand side of this inequality will approach zero for large T, as long as $g = O(T)$.

[11] Recall that we are using (re)normalized SDFs where $\mathbb{E}[m_{\boldsymbol{\gamma}}] = 1$ (instead of the inverse of the risk-free rate).

In other words, in large samples, this hurdle for including additional factors becomes trivial, and the g-priors will always lend support to the "denser" models.

In summary, to have a (mildly) volatile SDF, the parameter g needs to scale linearly with regard to the sample size. This economic restriction will artificially inflate the posterior probabilities of the models with more factors, leading to erroneous large sample properties.

1.1.3 The mixture of g-priors

A natural Bayesian approach to resolve the issue is by treating g as a random variable and assigning a (hyper)prior to it. Consider $g \sim \pi_a(g)$, where

$$\pi_a(g) = \frac{a-2}{2}(1+g)^{-\frac{a}{2}}, \quad g > 0,$$

as proposed in Liang et al. (2008). If the parameter $a = 4+2p_{\boldsymbol{\gamma}}/(TC^2)$ for some $C \in [\mathrm{SR}_\infty, h]$, it is easy to show that the prior volatility of the SDF will be exactly C, satisfying the economic restriction.

In general, the term $2p_{\boldsymbol{\gamma}}/(TC^2)$ tends to be small, i.e., $a \approx 4$. More formally, as $T \to \infty$, the prior $\pi_a(g)$ will converge in distribution to

$$\pi(g) = \frac{1}{(1+g)^2}, \quad g > 0, \tag{10}$$

which can be shown to be equivalent to $g/(1+g) \sim \mathrm{Uniform}(0, 1)$.

This representation allows a straightforward interpretation. From Proposition 2, the ratio $g/(1+g)$ decides how posterior beliefs about the SDF should be updated. This uniform prior reflects an agnostic view regarding this belief updating process due to the lack of theoretical guidance. We will adopt the specification of (10) throughout the rest of our paper, which encompasses all the meaningful economic restrictions discussed above.

Combining the baseline g-priors with (10), we have a new prior for the risk prices. Under

model $\mathcal{M}_{\boldsymbol{\gamma}}$, $\boldsymbol{b}_{\boldsymbol{\gamma}}$ follows a continuous scale mixture of normal distributions, that is,

$$\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}} \sim \frac{1}{G} \int_0^{\infty} \mathcal{N}\left(\mathbf{0},\, \frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top} \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right) \pi(g)\, \mathrm{d}g, \tag{11}$$

where $\pi(g)$ is defined in (10), and G is a normalizing constant. We follow Liang et al. (2008) and call this prior the (scale) mixture of g-priors. This specification still yields an analytical result for the Bayes factor, $\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})$, summarized in Proposition 3.

Proposition 3. *The Bayes factor for comparing model $\mathcal{M}_{\boldsymbol{\gamma}}$ against the null model $\mathcal{M}_{\mathbf{0}}$ under the mixture of g-priors is:*

$$\mathrm{BF}(\boldsymbol{\gamma},\, \mathbf{0}) = \exp\left(\frac{T}{2}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right)\left(\frac{T}{2}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right)^{-\frac{p_{\boldsymbol{\gamma}}+2}{2}} \underline{\Gamma}\left(\frac{p_{\boldsymbol{\gamma}}+2}{2},\; \frac{T}{2}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right),$$

where $\underline{\Gamma}(s, x) = \int_0^x t^{s-1} e^{-t}\, \mathrm{d}t$ is the lower incomplete Gamma function (Abramowitz and Stegun, 1965, Page 262). In addition, $\mathrm{BF}(\boldsymbol{\gamma},\, \mathbf{0})$ is increasing in $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ but decreasing in $p_{\boldsymbol{\gamma}}$.

The Bayes factor under the mixture of g-priors is related to the Sharpe ratio and model dimension similar to the one in equation (9), in which g is treated as a fixed number. However, as we will show later in Section 1.3, the mixture of g-priors enjoys favorable asymptotic properties that are crucial to credible estimations of model uncertainty. We begin the discussion by introducing our model uncertainty measure first.

1.2 Posterior Model Probability and Model Uncertainty

If we assign the same prior probability to every model, that is, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}}] = \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}'}]$ for any $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$, a direct outcome of the Bayes' theorem is that the posterior probability of model $\mathcal{M}_{\boldsymbol{\gamma}}$ equals

$$\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] = \frac{\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] \times \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}}]}{\sum_{\boldsymbol{\gamma}'} \mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}'}] \times \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}'}]} = \frac{\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})}{\sum_{\boldsymbol{\gamma}'} \mathrm{BF}(\boldsymbol{\gamma}', \mathbf{0})}, \tag{12}$$

in which the Bayes factor $\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})$ follows Equation (9).

We define the model uncertainty measure $\mathcal{E}$ as the entropy of the posterior model probabilities, normalized by its upper bound (achieved when all models have the same posterior probability of $1/2^p$): With p factors under consideration,

$$\mathcal{E} = -\frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}} \left(\log \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]\right) \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]. \tag{13}$$

Our model uncertainty measure is always between zero and one. From the perspective of a Bayesian investor, larger posterior entropy corresponds to higher model uncertainty. When $\mathcal{E} = 0$, there exists one model whose posterior probability equals one. When $\mathcal{E} = 1$, all models have the same posterior probability: they must be equally right or equally wrong.

1.3 Econometric Properties

To provide further justification for our prior choice and understand the resulting model uncertainty measure $\mathcal{E}$, we conduct frequentist assessment of our Bayesian inference in this section. The discussion concerns both asymptotic properties and finite-sample simulation results.

Under the g-prior specification, we have the following asymptotic results for posterior probabilities, summarized in Proposition 4.

Proposition 4. *(asymptotic analysis of the g-priors) Assume that the observed return data are generated from a true linear SDF model $\mathcal{M}_{\boldsymbol{\gamma}_0}$. If $\boldsymbol{\gamma}_0 \neq \mathbf{0}$ (the SDF is not a constant) and $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}$ (the set of factors under consideration include all true factors), under the g-prior specification with $g \in (0, \infty)$, as $T \to \infty$,*

1. *(factor selection consistency) if $\gamma_{0,j} = 1$, i.e., the true model includes factor j, the posterior marginal probability of choosing factor j converges to one in probability: $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] = \sum_{\{\boldsymbol{\gamma}:\ \gamma_j = 1\}} \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 1$;*[12]

[12] We will use the notation "$\xrightarrow{p}$" to denote "convergence in probability" throughout the paper. The corre-

2. *(model selection inconsistency) The posterior probability of the true model will always be strictly smaller than one, that is,* $\mathbb{P}[\mathcal{M}_{\gamma_0} \mid \mathcal{D}] < 1$ *with probability one.*

Model selection inconsistency of the g-priors is mainly due to their propensity to include factors that are not in the true model, namely, redundant factors. To be more accurate, under the g-priors, $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] > 0$ for js such that $\gamma_{0,j} = 0$. (This is a byproduct from our proof of Proposition 4 and validated through simulation studies tabulated in Table A3 and Table A4 of the Appendix). The g-priors can avoid discarding true factors at the cost of incorporating redundant ones.

Proposition 4 highlights the limitation of g-priors. Even if the observed return data are generated from a fixed true model — there is no model uncertainty, and $\mathcal{E}$ should be zero — an econometrician can never identify this model ($\mathcal{E} > 0$ as a result), regardless of how much return data have been accumulated. As a result, the model uncertainty measure defined in (13) exhibits upward bias under the g-priors.

The key methodological benefit of using the mixture of g-priors is to achieve posterior model selection consistency, as summarized in Proposition 5.

Proposition 5. *(asymptotic analysis of the mixture of g-priors) If all the assumptions of Proposition 4 hold, under the mixture of g-priors specification, as* $T \to \infty$, $\mathbb{P}[\mathcal{M}_{\gamma_0} \mid \mathcal{D}] \xrightarrow{p} 1$.

Accompanying the asymptotic theory presented in Propositions 4 and 5, we perform simulation studies examining the finite-sample behavior of our methods. Appendix Table A3 tabulates the results. Both g-priors and the mixture of g-priors consistently select true factors, even based on moderately sized return samples. However, inferences based on g-priors are plagued by redundant factors. The probability of mistakenly including factors in the SDF remains positive and largely unchanged as the sample size increases. In contrast, the same probability shrinks toward zero in larger samples under the mixture of g-priors.

sponding probability measure is always defined on the sample distribution under the true model. The same probability measure also applies to follow-up results arguing "convergence with probability one."

This family of newly proposed priors, guarding against redundant factors, assigns a posterior probability of approximately one to the true model.

With posterior model selection consistency, the econometrician can, in theory, identify the correct model (if there is one) after observing a long history of return data. The resulting model uncertainty measure $\mathcal{E}$ should converge in probability to zero when there is one linear factor model capturing the cross-section of expected returns. Simulation results in Table A3 of the Appendix show that $\mathcal{E}$ becomes statistically indistinguishable from zero under large samples (only) under the mixture of g-priors specification.

1.3.1 A misspecified set of factors

In real applications, to calculate our model uncertainty measure, we must focus on a predetermined set of factors, namely $\boldsymbol{f} = [f_1, \ldots, f_p]$, as in Equations (1) and (2). However, the belief that $\boldsymbol{f}$ under consideration includes *all* factors belonging to the true SDF is tenuous at best. In this section, we consider the setting under which the set of factors $\boldsymbol{f}$ is misspecified because it omits factors in the true SDF.

Under this setting, the econometrician can never identify the true model simply because the set of factors under her consideration is incomplete. As a result, no prior specification for the risk prices can deliver model selection consistency.

Even when there are true factors that are omitted from $\boldsymbol{f}$, the mixture of g-priors still maintains factor selection consistency: True factors that are under consideration can always be selected. Proposition 6 describes this property.

Proposition 6. *(posterior property with omitted factors) Assume that the observed return data are generated from a true linear SDF* $m_0 = 1 - (\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}_0$. *Let* $\boldsymbol{f}_{\boldsymbol{\gamma}_0} = \boldsymbol{f}_0 \cap \boldsymbol{f}$*; that is,* $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ *is the subset of* $\boldsymbol{f}$ *(factors under consideration) that includes the largest number of factors in the true model with no redundancy. As* $T \to \infty$*, for all* j *such that* $\gamma_{0,j} = 1$*,* $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] \xrightarrow{p} 1$.

Table A4 in the Appendix presents results from simulation studies under misspecification. Even in finite samples, true factors, as long as they are not omitted from $\boldsymbol{f}$, will always be identified because their posterior probabilities of entering the SDF approach one. The mixture of g-priors specification in general features the lowest rate of including redundant factors. Table A5 and A6 further confirm this pattern under a variety of settings (see sections of the two tables marked as the posterior probability of factors).

The model uncertainty measure $\mathcal{E}$ calculated from a misspecified set of factors, although not converging to zero due to omitted factors, still enjoys a good property, summarized in the following proposition.

Proposition 7. *(model uncertainty under misspecification) Under the assumption of Proposition 6, the model uncertainty measure $\mathcal{E}$ calculated based on the misspecified set of factors satisfies $\mathcal{E} \leq (p - p_{\gamma_0})/p$ with probability one, as $T \to \infty$.*

Proposition 7 indicates that if we observe very high model uncertainty $\mathcal{E}$, say $\mathcal{E} \approx 1$, only two possibilities exist: (1) $p_{\gamma_0} = 0$, which implies that the set of factors we are working with does not cover *any* of the true factors; (2) the assumptions in Propositions 6 and 7 are tenuous – the observed returns are not generated from a unique linear SDF model. By investigating factor probabilities $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$, we can rule out possibility (1) under the presence of "strong" factors ($\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] \approx 1$). Under the second possibility, the quest for one dominant linear factor model seems empirically quixotic.

The upper bound for model uncertainty in Proposition 7 tends to be loose. The proof of Proposition 7 in the Internet Appendix mandates that the upper limit is binding only when all models subsuming factors in $\boldsymbol{f}_{\gamma_0}$ (true factors that are *not* omitted) have exactly the same posterior probability. Simulation results under misspecified $\boldsymbol{f}$ (omitting the momentum factor) reported in Table A5 illustrate that the gap between $\mathcal{E}$ and $(p - p_{\gamma_0})/p$ becomes substantially larger as p increases. Table A6 further demonstrates that the gap is significant if other factors are ignored when calculating the model uncertainty measure.

2 Data

In our primary empirical implementation, we investigate 14 prominent factors from the past literature. First, we include Fama-French five factors (Fama and French, 2016) plus the momentum factor (Jegadeesh and Titman, 1993). In addition, we include the size, investment, and profitability factors from Hou, Xue, and Zhang (2015). We next consider the behavioral factor model of Daniel, Hirshleifer, and Sun (2020), including their short-term and long-term behavioral factors. Finally, we include the HML devil (Asness and Frazzini, 2013), quality-minus-junk (Asness et al., 2019), and betting-against-beta (Frazzini and Pedersen, 2014) factors from the AQR library. Appendix A presents a detailed description of these factors. Table A1 reports their annualized mean returns and Sharpe ratios. The sample starts from July 1972 to December 2020.

In addition to these 14 factors, we further include two mispricing factors in Stambaugh and Yuan (2017) as a robustness check. However, the sample of mispricing factors ends in December 2016. To ensure that we can measure model uncertainty until recent years, we exclude these two factors in the main analysis.

We obtain other uncertainty measures and economic variables from multiple sources. Specifically, we consider indices of economic policy uncertainty (EPU) in Baker et al. (2016) and three uncertainty measures developed in Jurado et al. (2015) and Ludvigson et al. (2021). All these uncertainty measures can be downloaded from the authors' websites. We download the VIX index from Wharton Research Data Services (WRDS). In addition, we use the term yield spread (the yield on ten-year government bonds minus the yield on three-month treasury bills) and the credit spread (the yield on BAA corporate bonds minus the yield on AAA corporate bonds). The bond yields are from the Federal Reserve Bank of St. Louis.

Finally, we obtain mutual fund data from the Center for Research in Security Prices (CRSP) survivorship-bias-free mutual fund database. In particular, we are interested in

monthly mutual fund flows, so we download the monthly total net assets, monthly fund returns, and the codes of fund investment objectives. In addition, we download the total market value of all US-listed stocks from CRSP.

3 Model Uncertainty in Equity Markets

We construct a monthly time series of model uncertainty in the US equity market based on the proposed framework. At the end of each month, we use daily returns in the past three years to compute the posterior model probabilities and model uncertainty measures based on equations (12) and (13).[13] Bayes factors in these equations are calculated under the mixture of g-priors as in Proposition 3. The behavioral factors in Daniel et al. (2020) are available only from July 1972, and we use 36-month data in the estimation, so the model uncertainty measure starts from June 1975.[14]

As certain pairs of factors are highly correlated, we consider models that contain at most one factor in each of the following categories: (a) size (SMB or ME); (b) profitability (RMW or ROE); (c) value (HML or HML Devil); (d) investment (CMA or IA). We refer to size, profitability, value, and investment as categorical factors. Therefore, there are ten effective factors, including market, size, profitability, value, investment, short-term and long-term behavioral factors, momentum, QMJ, and BAB. Under this setting, there are 5,184 different candidate models, and the possible range of our model uncertainty measure is [0, 1].

Figure 1 plots the time series of our model uncertainty measure. It exhibits several interesting features. Overall, we observe a surprisingly high level of model uncertainty. The average (median) model uncertainty is around 0.70 (0.75), with the first and third quartiles

[13] We repeat our calculation with four-year and five-year rolling windows. Figure IA.1 in the Internet Appendix confirms that our cross-sectional model uncertainty measure exhibits similar time-series patterns under alternative specifications of window widths.

[14] We also measure model uncertainty in the dataset with two mispricing factors of Stambaugh and Yuan (2017). The orange dotted line in Figure IA.2 of the Internet Appendix shows the corresponding time series, and we confirm that including mispricing factors almost does not change the main empirical patterns.

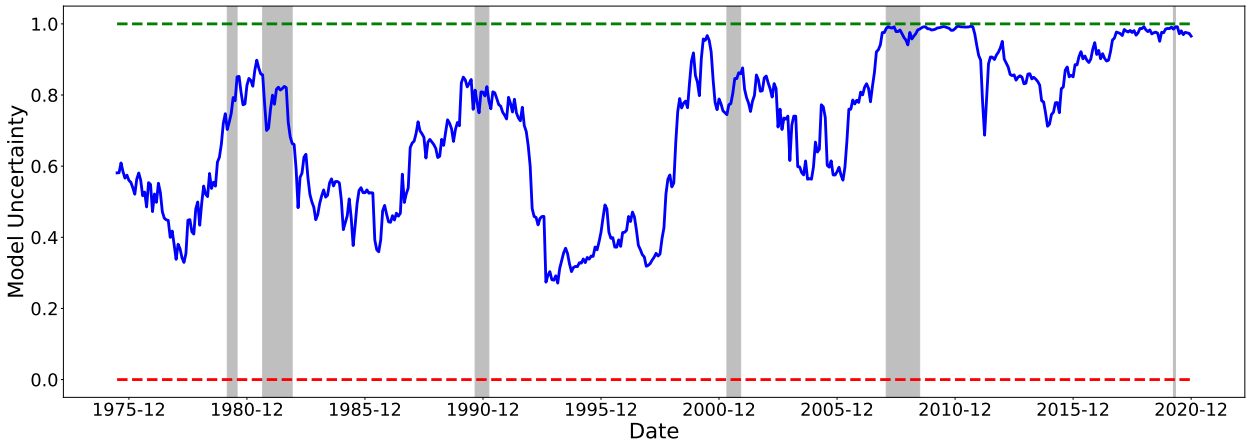


Figure 1: Time Series of Model Uncertainty in the US Equity Market

The figure plots the time series of model uncertainty in the linear factor model. We consider fourteen prominent factors in the literature and apply our framework to calculate uncertainty (see Section 2 and 3 for details). The red and green lines show the lower and upper bounds of model uncertainty. Shaded areas mark the NBER recession periods.

equal to 0.53 and 0.87, respectively. Model uncertainty in the cross-section is also a dynamic phenomenon: It fluctuates significantly over time. In particular, the index varies from the lowest value of 0.27 to the highest of 0.99, with a standard deviation of 0.21.

As shown in Figure 1, our model uncertainty measure is countercyclical. In particular, the 1990s, often remembered as a period of strong economic conditions and high stock returns, witnesses the lowest model uncertainty in our sample. As the orange diamonds in Figure 2 suggest, the posterior probabilities of the top two models are significantly larger than others. Hence, it is relatively straightforward to figure out the true SDF model in this period.

In addition, episodes of heightened model uncertainty tend to coincide with economic downturns and stock market crashes. For example, our model uncertainty measure touches the upper bound during the global financial crisis. The blue dots in Figure 2, showing the posterior probabilities of the top 50 models in December 2007, lie on a horizontal line – posterior probabilities of models are equalized. It is virtually infeasible to distinguish models

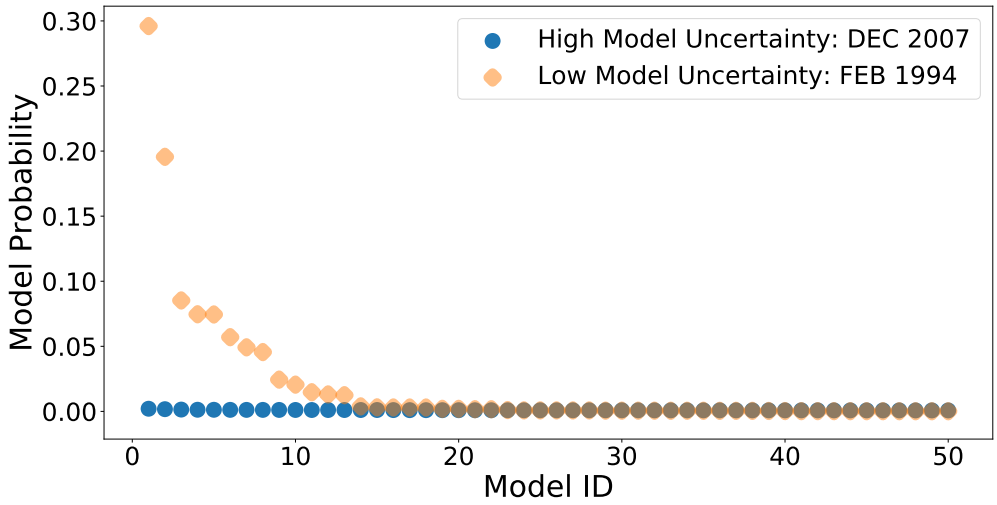


Figure 2: Posterior Probabilities of Top 50 models: High vs. Low Model Uncertainty

The figure plots the posterior probabilities of the top 50 models ranked by their posterior probabilities. At the end of each month, we compute the posterior model probabilities using the daily factor returns in the past three years. We use the entropy of model probabilities to quantify model uncertainty in the cross-section. We observe low model uncertainty in February 1994 (orange diamonds) but high model uncertainty in December 2007 (blue dots).

based on the observed data. The 2008 crisis is noteworthy also because model uncertainty stays at a high level for a prolonged period until recently. In the recent five years, model uncertainty has slowly increased from 0.7 to 1 at the end of 2020. In contrast, it declines shortly after other crisis periods.

3.1 The Informational Contents of Model Uncertainty

Our model uncertainty measure carries information that is orthogonal to (1) the mean, volatility, and correlation structure of factors, and (2) other macroeconomic variables.

Proposition 3 shows that posterior model probabilities increase in the squared Sharpe ratios SR_γ^2 and decrease in model dimensions p_γ. As the maximal Sharpe ratio is determined by the mean, volatility, and correlation of factor returns, we compare these three simple statistics with our model uncertainty measure in Figures A1 and 3. Figures A1a and A1b

show that model uncertainty is clearly different from the average and volatility of factor returns. Figure 3 plots the average pairwise correlations among factors under consideration. Factor correlations demonstrate a certain degree of association with model uncertainty, especially before the 2000s. This association breaks up in recent years. Overall, none of these three statistics can entirely convey the exact information in model uncertainty.

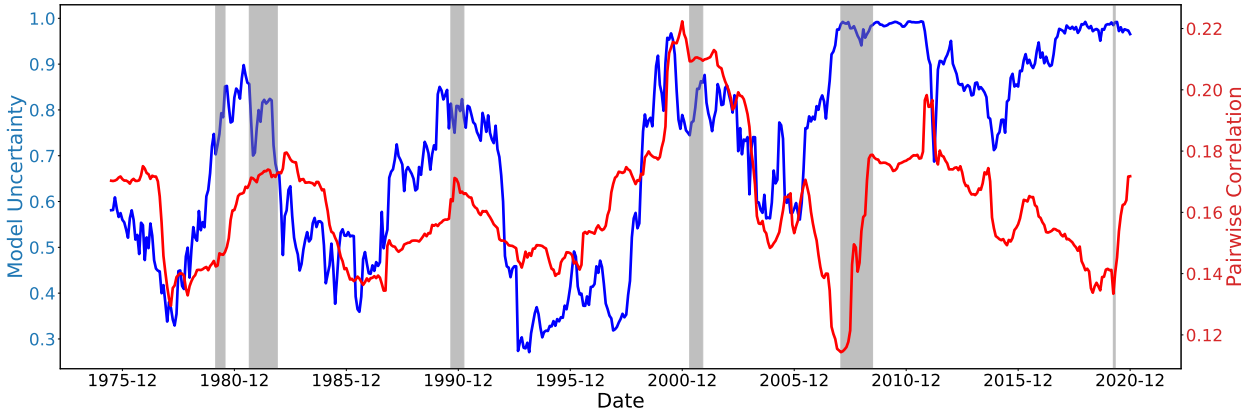


Figure 3: Time Series of Average Pairwise Correlation of 14 Factors

The figures plot the time series of average pairwise (absolute) correlation among daily factor returns. These statistics are estimated using the daily factor returns in the past 36 months.

We then examine whether our model uncertainty measure is related to other uncertainty indices in recent literature, including the VIX, economic policy uncertainty (EPU) in Baker et al. (2016), and three uncertainty measures from Jurado et al. (2015) and Ludvigson et al. (2021). Table 1 reports the results from regressing our model uncertainty measure $\mathcal{E}_t$ on these indices after controlling for a lagged term. The regression coefficients describe contemporaneous associations, with no intention to study causal relationships. Our model uncertainty measure is only significantly associated with the financial uncertainty index of Ludvigson et al. (2021) and the VIX, both of which are also constructed from asset prices.

We further compare model uncertainty with financial variables that are known to be related to aggregate fluctuations, including the term spread (the difference between 10-year and 3-month treasury yields) and the credit spread (the yield difference between BAA and

AAA bonds). The term spread is negatively associated with model uncertainty.

Overall, model uncertainty relates to other important economic variables. However, these variables only explain a small fraction of time-series variation in model uncertainty. As Figure A2 indicates, model uncertainty displays significant independent variation different from these variables from the existing literature.

Table 1: Regressions of Model Uncertainty on Contemporaneous Variables

X	Fin U.	Macro U.	Real U.	EPU I	EPU II	VIX	TS	DS
β	0.21	0.17	0.14	0.00	0.00	0.01	−0.03	−0.00
	(1.95)	(1.53)	(1.20)	(0.33)	(1.07)	(2.20)	(−3.44)	(−0.09)
# obs.	546	546	546	432	432	420	546	546

The table reports results from the following regression:

$$\mathcal{E}_t = \beta_0 + \beta X_t + \rho \mathcal{E}_{t-1} + \epsilon_t,$$

where the variable X_t represents a) macro, financial, and real uncertainty measures from Jurado et al. (2015) and Ludvigson et al. (2021) (Fin U, Macro U, and Real U); b) two economic policy uncertainty (EPU) indices from Baker et al. (2016) (EPU I and EPU II); c) the CBOE VIX index (VIX); d) the term spread between ten-year and three-month treasuries (TS), e) the default spread between BAA and AAA corporate bond yields (DS). We standardize all variables to have unit variances. The t-statistics in parenthesis are computed based on Newey-West standard errors with 36 lags.

3.2 International Evidence

This section presents the time series of model uncertainty in European and Asia-Pacific stock markets.[15] Instead of using all 14 factors as in the US stock market, we include only nine of them since the size (ME), profitability (ROE), and investment (IA) factors in Hou et al. (2015), and short-term and long-term behavioural factors are unavailable in international markets. Since the AQR library provides the QMJ factor from July 1993, and we use a three-year rolling window, our model uncertainty measure starts from June 1996.

[15]The European markets include the following countries: Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, UK, Greece, Ireland, Italy, Netherlands, Norway, Portugal, and Sweden. The Asia-Pacific markets refer to the stock markets in Australia, Hong Kong, New Zealand, and Singapore.

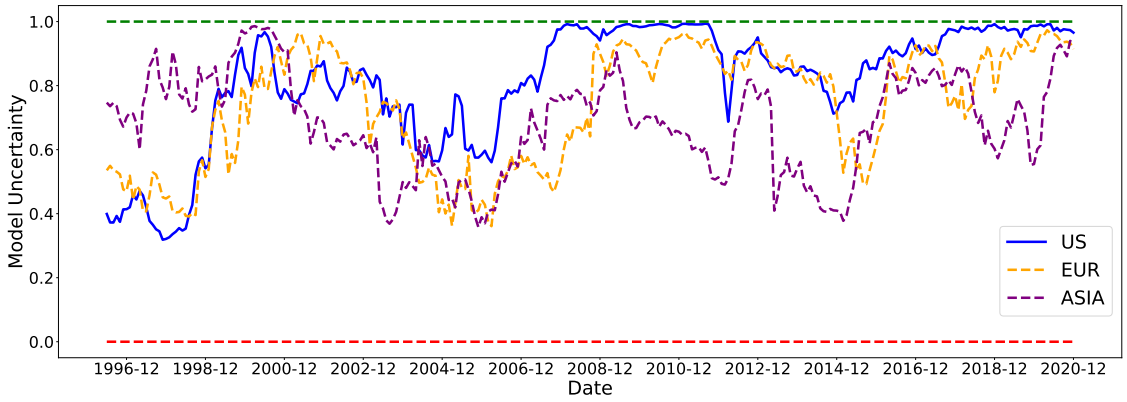


Figure 4: Model Uncertainty in European and Asia Pacific Markets

The figure plots the time series of model uncertainty about linear factor models in European and Asian-Pacific Stock Markets. The construction of model uncertainty is the same as in Figure 1 except that we use only nine factors to calculate the posterior model probabilities. Details about used factors could be found in section 3.2. The sample ranges from July 1993 to December 2020. Since we use 3-year rolling window, the model uncertainty index starts from June 1996. The red line and green lines in the figure show the lower (0) and upper bounds (1) of model uncertainty. As comparison, we include the US model uncertainty index (blue solid line) in Figure 1 during the same period.

The orange line in Figure 4 plots the time series of model uncertainty in the European stock market. Several patterns emerge. Model uncertainty in European markets is similar to the US market. In particular, model uncertainty increases from 1999 and reaches its first peak between 2000 and 2001 (the technology bubble burst). During these periods, model uncertainty almost touches its upper bound. After 2002, model uncertainty declines gradually and remains relatively low until the 2008 global financial crisis. Model uncertainty stays close to the upper bound from 2008 to 2012 and only declines gradually after 2012.

We observe some unique model uncertainty variation in the Asia-Pacific stock markets. According to the purple line in Figure 4, model uncertainty is heightened starting from 1997, probably due to the profound 1997 Asian financial crisis. The technology bubble in 2000 witnesses another peak in model uncertainty. Another steady increase in model uncertainty appears before and during the 2008 crisis, but the entropy is lower than that in the late 1990s and drops immediately from 2009. This pattern is unlike the US and European markets, where we observe the highest model uncertainty in the 2008 crisis.

In short, the international evidence lends further support to the time-varying nature of model uncertainty, highlighting the fact that heightened model uncertainty coincides with major events in corresponding asset markets. However, model uncertainty is not all alike. For example, Asian markets display unique behaviors that distinguish them from the US and European markets.

3.3 Factor Uncertainty

Our model uncertainty measure quantifies the difficulties of choosing asset pricing models. When the measure indicates that no model clearly dominates the others, one might postulate that there must be high uncertainty about which factors should enter the true model, namely, factor uncertainty. To quantify factor uncertainty, we compute the posterior (marginal) probability of selecting each individual factor according to definitions in Proposition 5.

According to Figure IA.3 in the Internet Appendix, posterior marginal probabilities of factor selection demonstrate pronounced time-series variation. During economic downturns, factors' chances of being selected tend to drop. The market, size, value, profitability, betting-against-beta (BAB), and short-term behavioral factors (PEAD) all undergo extended periods of being selected with probability one. Before 2000, only the market, value and profitability factors show consistently high probabilities of entering the true asset pricing model. After 2000, only the market and betting-against-beta factors pass the same scrutiny.

A surprising finding from plots in Figure IA.3 is that high model uncertainty does not preclude the existence of factors that have high probabilities of entering the asset pricing model (low factor uncertainty). For example, our model uncertainty measure almost reaches its maximum in late 1999; in the meantime, the posterior probabilities of including BAB and PEAD are both above 90%. High propensity factors are observed consistently during periods of heightened model uncertainty before 2008.

Our Proposition 7 calls for special attention to the phenomenon above. It is possible that

the observed high uncertainty is entirely driven by extreme misspecification, i.e., the set of factors under consideration does not include any true factors. Low uncertainty factors in the context of high model uncertainty help mitigate this concern.

4 The Market Price of Model Uncertainty

In this section, we investigate whether model uncertainty shocks are priced in the equity markets. To do so, we extract the AR(1) innovations, namely $\mathcal{E}_t^{ar1}$, to the model uncertainty measure and normalize these shocks to a unit standard deviation. We then construct the mimicking portfolios of model uncertainty shocks and test whether these portfolios command risk premia that are different from zero.

To construct the mimicking portfolios, we project $\mathcal{E}_t^{ar1}$ onto the set of *tradable* factors $\boldsymbol{f}_t$ according to

$$\mathcal{E}_t^{ar1} = \eta_0 + \boldsymbol{\eta}^\top \boldsymbol{f}_t + e_t. \tag{14}$$

The risk premium of the model uncertainty shocks, denoted by $\lambda_{\mathcal{E}}$, can be calculated as the expected return of the mimicking portfolio: $\lambda_{\mathcal{E}} = \boldsymbol{\eta}^\top \mathbb{E}[\boldsymbol{f}_t]$.

To conduct proper statistical inference about the estimates of $\lambda_{\mathcal{E}}$, we need to account for two sources of parameter uncertainty: one from estimating the factor expected returns $\mathbb{E}[\boldsymbol{f}_t]$ using the time-series averages $1/T \sum_{t=1}^{T} \boldsymbol{f}_t$, and the other from the standard errors of the factor-loading estimator $\widehat{\boldsymbol{\eta}}$. The second adjustment is in the same spirit as the Shanken correction in two-pass regressions (Shanken, 1992). A hierarchical Bayesian framework is proposed in Internet Appendix IA.3 to make inferences about the estimates of $\lambda_{\mathcal{E}}$, adjusting for both types of estimation uncertainty.

Giglio and Xiu (2021) point out that risk premia estimates based on a smaller number of pre-specified factors will be subject to the canonical omitted variable bias. To overcome this issue, they propose to consider the space spanned by the principal components (PCs) of a

large cross-section of asset returns. We also adopt their methodology and estimate the risk premia of model uncertainty shocks $\lambda_{\mathcal{E}}$ in the US equity market using a large cross-section of 275 characteristic-sorted portfolios.[16]

Panel A of Table 2 reports the risk premia estimates of model uncertainty shocks in US, European, and Asia-Pacific markets based on the same set of factors used to compute the model uncertainty measure in Section 3. Since size, value, investment, and profitability have two versions of tradable factors, which are highly correlated, we allow only one of them to enter the right-hand side of equation (14).

Two observations in Panel A are noteworthy. First, model uncertainty shocks command significantly negative risk premia in the US and European equity markets, which implies that investors are willing to pay a premium to hedge against heightened model uncertainty. Risk premia estimates are almost identical in these two markets. This observation is consistent with model uncertainty measures being more similar in European and the US stock markets, as shown in Figure 4. Second, the magnitude of risk premia is smaller in the Asia-Pacific market. Although not consistently significant, the risk premium estimates are always negative in this market.

Panel B of Table 2 presents the risk premium estimates in the US market using a large cross-section of equity returns. The methodology follows Giglio and Xiu (2021). The number of PCs is estimated to be five, using the algorithm proposed by Giglio and Xiu (2021). The risk premium of model uncertainty shocks is -0.064. For robustness concerns, we also report results based on different numbers of PCs, which tend to be remarkably stable. Comparing the estimates in Panel B with those in columns (1) and (2) of Panel A, the magnitude of the risk premia is reduced by about a half—but remains significantly negative. Recall that we have normalized the model uncertainty shocks to a unit standard deviation;

[16] We consider 5×5 portfolios sorted by size versus (1) book-to-market ratio, (2) accrual, (3) market beta, (4) investment, (5) long-term reversal, (6) momentum, (7) net issuance, (8) profitability, (9) idiosyncratic volatility, (10) total volatility, (11) short-term reversal.

Table 2: Risk Premia of Model Uncertainty Shocks: Monthly Estimates

	Panel A. Small Cross-Section					
	US: 1975/07 - 2020/12		Europe: 1996/07 - 2020/12		Asia: 1996/07 - 2020/12	
$\lambda_{\mathcal{E}}$	-0.134^{***}	-0.168^{***}	-0.139^{***}	-0.148^{***}	-0.097^{***}	-0.050
s.e.	0.027	0.030	0.037	0.039	0.038	0.035
Time-series R^2	8.3%	10.3%	15.1%	14.1%	10.2%	6.5%
MKT	✓	✓	✓	✓	✓	✓
SMB	✓		✓	✓	✓	✓
HML	✓		✓		✓	
RMW	✓		✓	✓	✓	✓
CMA	✓		✓	✓	✓	✓
MOM	✓	✓	✓	✓	✓	✓
QMJ	✓	✓	✓	✓	✓	✓
BAB	✓	✓	✓	✓	✓	✓
HML devil		✓		✓		✓
ME		✓				
IA		✓				
ROE		✓				
FIN	✓	✓				
PEAD	✓	✓				
	Panel B. Large Cross-Section in the US: 1975/07 - 2020/12					
Number of PCs:	5	6	7	8	9	10
$\lambda_{\mathcal{E}}$	-0.064^{***}	-0.064^{***}	-0.064^{***}	-0.063^{***}	-0.062^{***}	-0.061^{***}
s.e.	0.017	0.017	0.017	0.017	0.018	0.018
Time-series R^2	5.8%	5.8%	5.8%	6.2%	6.2%	6.2%

This table reports the risk premia estimates of model uncertainty shocks. The risk premium is defined as the average return of the corresponding mimicking portfolio. In all estimations, we standardize $\mathcal{E}_t^{ar1}$ and factors to have unit variances. In Panel A, we project $\mathcal{E}_t^{ar1}$ onto the space of 14 observable factors. The confidence intervals are estimated using the algorithm in the Internet Appendix IA.3. We also show the standard errors of risk premia estimates based on the posterior distributions and which factors are used to construct the mimicking portfolio (✓means that the factor is included). In Panel B, we project $\mathcal{E}_t^{ar1}$ onto the space of large PCs of 275 Fama-French characteristic-sorted portfolios. The number of latent factors ranges from five to ten. We estimate the standard errors of risk premia estimates following Giglio and Xiu (2021). If the 90% (95%, 99%) confidence interval of the risk premium does not contain zero, the risk premium estimate will be highlighted by * (**, ***). We also report the time-series fit in each panel.

hence, the monthly risk premium $\lambda_{\mathcal{E}} = -0.064$ is effectively a Sharpe ratio. For ease of interpretation, if we equate the standard deviation of the model uncertainty shocks to that of the market portfolio (17% per year in our sample), the annualized risk premium equals $(\sqrt{12})\lambda_{\mathcal{E}} \times 17\% \approx -3.8\%$.

For clarification, we point out two caveats to the analysis in this section. First, we estimate the *unconditional* risk premia of model uncertainty shocks. As factor loadings and conditional factor returns are possibly time-varying, our estimates only serve as simple yet imperfect benchmarks. Second, we do not claim that model uncertainty shocks *explain* the cross-section of expected returns. As we have argued previously, factors' risk premia are distinct from their risk prices–and only the latter determines whether factors can enter the SDF or, equivalently, explain the cross-section.

5 Model Uncertainty and Investors' Portfolio Choices

If investors consider model uncertainty a crucial source of investment risk, a natural hypothesis is that model uncertainty will convey useful information about their portfolio choices. We test this hypothesis using aggregate flows into equity and fixed-income mutual funds as proxies for investors' asset allocation decisions.

Our data come from the CRSP survivor-bias-free US mutual fund database. The database includes investment styles or objective codes from three different sources over the whole life of the database.[17] Since high-quality investment objective data are unavailable until 1990, our sample begins afterwards.

To begin with, we define the aggregate mutual fund flows. Following the literature (Sirri and Tufano, 1998), we calculate the net fund flows to each fund i in period t as

$$\text{Flow}_{i,t} = \text{TNA}_{i,t} - \text{TNA}_{i,t-1} \times (1 + R_{i,t}), \tag{15}$$

where $\text{TNA}_{i,t}$ and $R_{i,t}$ are total net assets and gross returns of fund i in period t. Next, we

[17] From 1962 to 1993, Wiesenberger objective codes are used. Strategic insight objective codes are populated between 1993 and 1998. Lipper objective codes start in 1998. Instead of using the three measures mentioned above, CRSP builds its objective codes based on them. The CRSP style code consists of up to four letters. For example, a fund with the style "EDYG" means that i) this fund mainly invests in domestic equity markets (E = Equity, D = Domestic), and ii) it has a specific investment style "Growth" (Y = Style, G = Growth). More details are in the handbook of the CRSP survivor-bias-free US mutual fund database.

aggregate individual fund flows in each period across all funds in a specific group (e.g., all large-cap funds) and scale the aggregate flows by the lagged total market capitalization of all stocks in CRSP:

$$\text{Flow}_t^{\mathcal{O}} = \frac{\sum_{i\in\mathcal{O}} \text{Flow}_{i,t}}{\text{CRSP-Market-Cap}_{t-1}}, \tag{16}$$

where $\mathcal{O}$ specifies a certain investment objective, such as small-cap funds.

We use the vector autoregression (VAR) model to study the dynamic responses of fund flows to uncertainty shocks. Specifically, we consider the reduced-form VAR(l) model:

$$\boldsymbol{Y}_t = \boldsymbol{B}_0 + \boldsymbol{B}_1\boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{B}_l\boldsymbol{Y}_{t-l} + \boldsymbol{B}_{\boldsymbol{x}}\boldsymbol{X}_t + \boldsymbol{u}_t, \tag{17}$$

where l denotes the lag order, $\boldsymbol{Y}_t$ is a $k \times 1$ vector of economic variables, $\boldsymbol{X}_t$ is a vector of exogenous control variables, $\boldsymbol{u}_t$ is a $k \times 1$ vector of reduced-form innovations with the covariance matrix $\boldsymbol{\Sigma}_u$, and $(\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_l, \boldsymbol{B}_{\boldsymbol{x}})$ are the matrices of coefficients.

Past literature often relates reduced-form innovations to structural shocks, i.e., $\boldsymbol{u_t} = \boldsymbol{S}\boldsymbol{\epsilon_t}$, where $\boldsymbol{S}$ is a $k \times k$ non-singular matrix, and $\boldsymbol{\epsilon_t}$ is a $k \times 1$ vector of structural shocks, which are orthogonal to each other by definition. We use the Cholesky decomposition to identify the dynamic responses to uncertainty shocks, so the ordering of economic variables in $\boldsymbol{Y_t}$ determines the identification assumption, which will be specified below.

5.1 Aggregate Flow Responses: Equity and Fixed-Income Funds

In the baseline analysis, we consider the aggregate mutual fund flows to the entire US fixed-income (FI) and equity (EQ) markets, which means $\boldsymbol{Y_t}^\top = (\mathcal{E}_t,\ \text{Flow}_t^{FI},\ \text{Flow}_t^{EQ})$ in equation (17). We next use impulse response functions (IRFs) to better understand the dynamic effects and propagating mechanisms of uncertainty shocks.

IRFs heavily depend on the identification assumption, i.e., whether model uncertainty is an exogenous source of fluctuations in fund flows or an endogenous response. In the first

case, model uncertainty is a cause of fund flows, while it acts as a propagating mechanism in the latter case. Without taking a strong stance on the identification assumption, we aim to investigate the dynamic relationship between fund flows and several uncertainty measures, either as a cause or a propagating mechanism. To make as few assumptions as possible, we focus on the dynamic responses to uncertainty shocks and are silent on how innovations in fund flows affect model uncertainty. This simplification allows us to ignore the ordering of other economic variables beyond model uncertainty.

We present the empirical results based on two identification assumptions. In the first case, we place model uncertainty first in the VAR (exogenous assumption). Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to fund flows in the current period. We further consider a different identification assumption, in which we put model uncertainty as the last element in $\boldsymbol{Y_t}$ (endogenous assumption).

Table 3: VAR Estimation of Monthly Model Uncertainty, Flows to Domestic Equity Funds, and Flows to Domestic Fixed-Income Funds

	Flow_{t+1}^{FI}		Flow_{t+1}^{EQ}	
	est.	t-stat.	est.	t-stat.
$\mathcal{E}_t$	−0.005	−0.071	−0.344	−7.255
Flow_t^{FI}	0.246	3.750	−0.081	−1.337
Flow_t^{Equity}	−0.090	−1.450	0.240	3.976

This table reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (\mathcal{E}_t, \text{Flow}_t^{FI}, \text{Flow}_t^{EQ})$. $\mathcal{E}_t$ is the model uncertainty measure, and Flow_t^{FI} (Flow_t^{EQ}) is the aggregate flows to the domestic fixed-income (equity) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged market return, lagged fixed-income fund return, and VIX index in each regression. The sample spans from January 1991 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags.

Table 3 reports the results from the VAR regression. The sample ranges from January 1991 to December 2020. The lag is chosen by BIC and equals one. In addition, we standardize

all economic variables such that they have unit variances. We also include the lagged market return and VIX index as control variables. The reported t-statistics are based on the Newey-West estimates of the covariance matrix with 36 lags. Several results are noteworthy. First, model uncertainty relates only to its lag. Second, model uncertainty negatively forecasts equity fund flows, and the coefficient estimate is sizable in both economic and statistical senses. In particular, one standard deviation increase in model uncertainty implies 0.34 standard deviation equity fund outflows. Although we cannot interpret the regression results as causal, we still find that investors tend to decrease their exposures to domestic equity markets when model uncertainty increases.

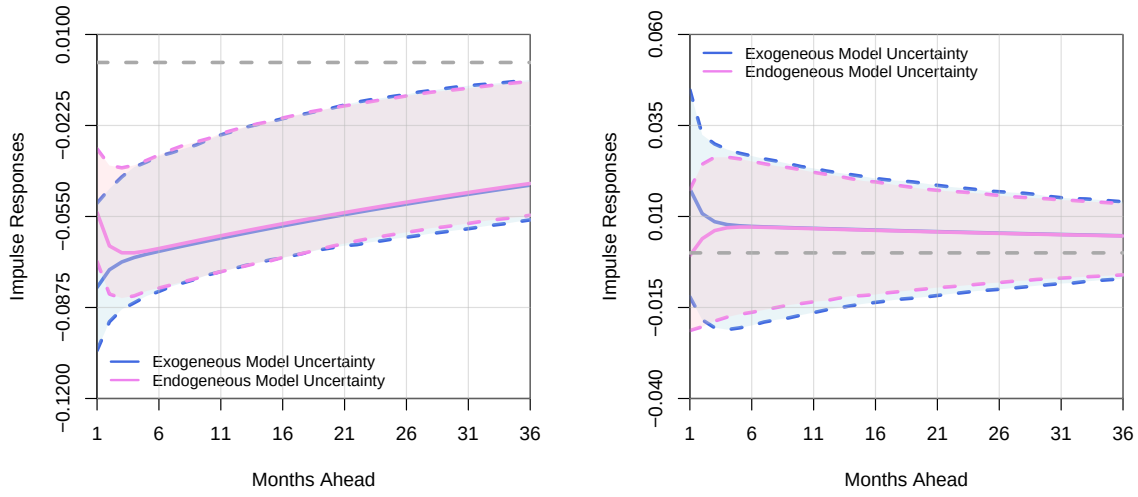
Figure 5 plots the dynamic responses of fund flows to model uncertainty shocks in VAR(1). Strikingly, model uncertainty innovations sharply induce fund outflows from the US equity market, with the effects persisting even after 36 months, as depicted in Panel (a). The IRFs start from around -0.08 in period zero and slowly decline to -0.05 in period 36, significantly negative based on the 90% standard error bands. In contrast, model uncertainty has negligible effects on fixed-income fund flows (see Panel (b)).

5.2 Heterogeneity in Flow Responses: Equity Mutual Funds

We further study the heterogeneous responses of different equity mutual funds to model uncertainty shocks. In particular, we split equity mutual funds into four categories: (a) style funds that specialize in factor investing, (b) sector funds that invest in specific industries (e.g., gold, oil, etc.), (c) small-cap funds that invest in small stocks,[18] and (d) large-cap funds.

Table 4 reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (\mathcal{E}_t, \text{Flow}_t^{style}, \text{Flow}_t^{sector}, \text{Flow}_t^{small}, \text{Flow}_t^{large})$. The lag of VAR is chosen by BIC and equals one. Since the cap-based investment objective code is available after 1997, the sample begins

[18] When we mention small funds, we refer to the funds with the CRSP investment objective codes equal "EDCM", "EDCS", and "EDCI".



(a) IRFs of Equity Fund Flows **(b)** IRFs of Fixed-Income Fund Flows

Figure 5: Impulse Responses of Equity and Fixed-Income Mutual Fund Flows to Model Uncertainty Shocks

This figure shows the dynamic impulse response functions (IRFs) of fund flows to model uncertainty shocks in VAR(1). The shaded area denotes the 90 percent standard error bands. We consider mutual fund flows to aggregate equity and fixed-income markets in the US. We consider two identification assumptions, (1) by placing model uncertainty first in the VAR (exogenous shock, highlighted in blue) and (2) by placing model uncertainty as the last variable in the VAR (endogenous response, highlighted in purple). The dotted gray line corresponds to the zero impulse response. The data are monthly and span the period 1991:01 - 2020:12.

in January 1998. First, after controlling its lag, model uncertainty is positively predicted by large-cap fund flows. Second, model uncertainty negatively forecasts style and small-cap fund flows, and the coefficients are sizable. Specifically, if model uncertainty rises by one standard deviation, style (small-cap) fund flows tend to drop by 0.26 (0.12) standard deviation over the next period. In contrast, we do not discover a significant relationship between model uncertainty and sector/large-cap fund flows.

Figure 6 shows the dynamic responses of four different types of equity fund flows to model uncertainty shocks in VAR(1). Consistent with Table 4, model uncertainty shocks reduce future style fund flows, and the effects are long-lasting (see Panel (a)). This observation is

Table 4: VAR Estimation of Monthly Model Uncertainty and Flows to Domestic Equity Funds with Different Investment Objectives

	$\text{Flow}_{t+1}^{style}$		$\text{Flow}_{t+1}^{sector}$		$\text{Flow}_{t+1}^{small}$		$\text{Flow}_{t+1}^{large}$	
	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
$\mathcal{E}_t$	−0.261	−5.672	−0.021	−0.525	−0.121	−1.967	−0.014	−0.209
Flow_t^{style}	0.211	2.936	−0.056	−1.034	−0.003	−0.054	0.003	0.034
Flow_t^{sector}	−0.056	−1.089	0.254	1.686	−0.059	−0.664	−0.123	−2.266
Flow_t^{small}	0.010	0.169	0.039	0.541	0.424	6.081	0.089	1.225
Flow_t^{large}	0.062	1.181	−0.043	−0.661	−0.107	−1.731	0.092	1.164

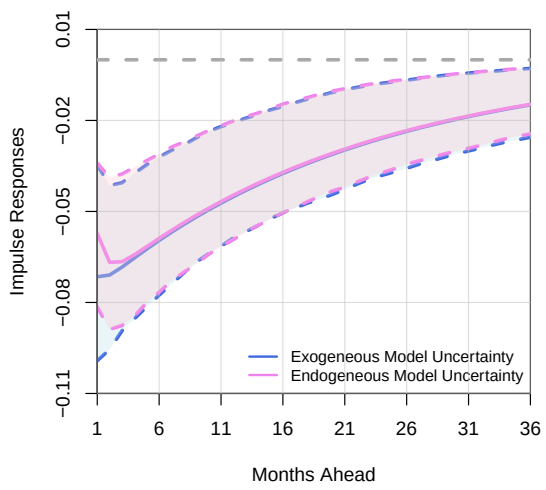
This table reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (\mathcal{E}_t, \text{Flow}_t^{style}, \text{Flow}_t^{sector}, \text{Flow}_t^{small}, \text{Flow}_t^{large})$. $\mathcal{E}_t$ is the model uncertainty measure, and Flow_t^{style} (Flow_t^{sector}, Flow_t^{small}, Flow_t^{large}) is the aggregate flows to the domestic style (sector, small-cap, large-cap) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged fund returns of each type and VIX index in each regression. The sample spans from January 1998 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags.

intuitive. Style funds refer to the growth, income, growth & income, and "hedged" funds, so they rely on factor strategies used in constructing model uncertainty. Therefore, outflows from style equity funds are remarkably enormous when model uncertainty is heightened.

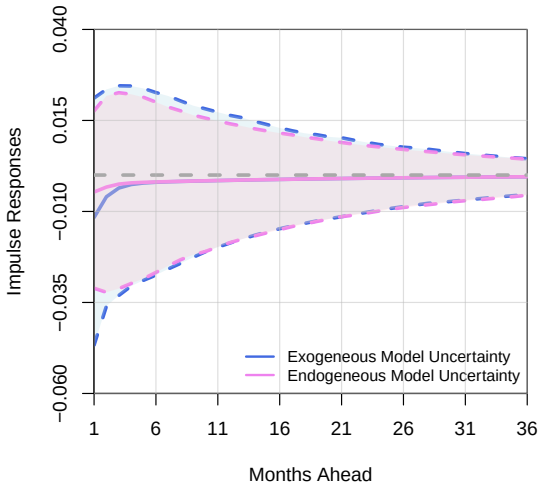
Moreover, we observe significantly negative IRFs of small-cap funds (see Panel (c)), although the effects are less persistent than style funds. On the contrary, sector and large-cap funds do not respond to model uncertainty shocks. One potential explanation is that these two types of funds are primarily passive-investing funds, but model uncertainty mainly affects actively-managed funds.

5.3 Heterogeneity in Flow Responses: Fixed-Income Mutual Funds

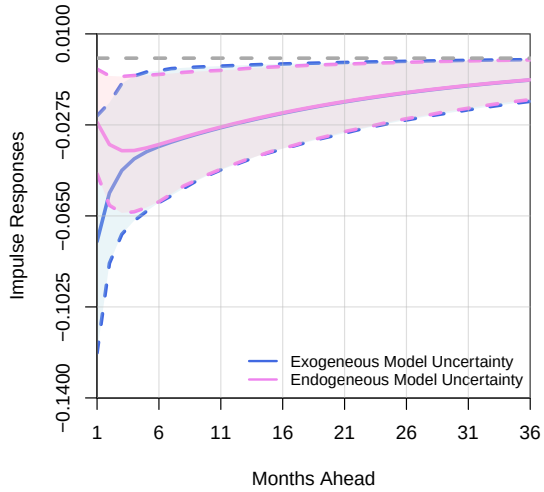
We further divide all fixed-income mutual funds into four categories: (a) government bond funds, (b) money market funds, (c) corporate bond funds, and (d) municipal bond funds. This subsection repeats a similar VAR estimation and investigates the dynamic responses of



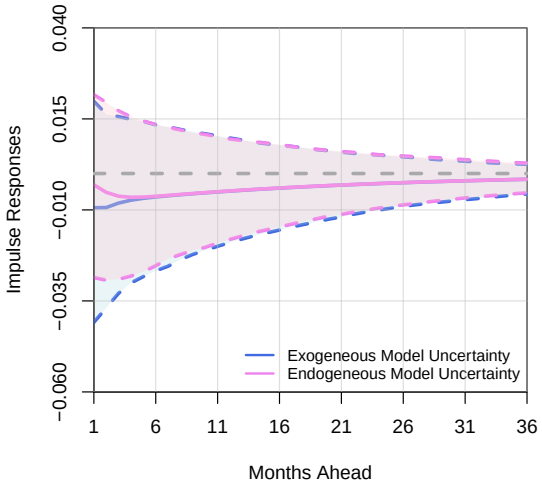
(a) Style Fund Flows



(b) Sector Fund Flows



(c) Small-Cap Fund Flows



(d) Large-Cap Fund Flows

Figure 6: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes to Model Uncertainty Shocks

This figure shows the dynamic impulse response functions (IRFs) of fund flows to model uncertainty shocks in VAR(1). The shaded area denotes the 90 percent standard error bands. We consider equity fund flows with different investment objective codes (style, sector, small-cap, and large-cap). We normalize the IRFs such that the model uncertainty shock increases one standard deviation model uncertainty. We consider two identification assumptions, (1) by placing model uncertainty first in the VAR (exogenous shock, highlighted in blue) and (2) by placing model uncertainty as the last variable in the VAR (endogenous response, highlighted in purple). The dotted gray line corresponds to the zero impulse response. The data are monthly and span the period 1998:01 - 2020:12.

fixed-income fund flows to model uncertainty shocks.

Table 5 shows the results from the VAR(1) regression. First, model uncertainty positively predicts the aggregate fund flows in US government bonds, which are notable for their superior safety over other asset classes. Hence, investors tend to allocate more wealth to safe assets when model uncertainty is more substantial. In contrast, model uncertainty negatively forecasts corporate fund flows, so mutual fund investors reduce their exposure to corporate bonds following high model uncertainty.

Table 5: VAR Estimation of Monthly Model Uncertainty and Flows to Domestic Fixed-Income Funds with Different Investment Objectives

	Flow_{t+1}^{gov}		$\text{Flow}_{t+1}^{money}$		Flow_{t+1}^{corp}		Flow_{t+1}^{muni}	
	est.	t-stat.	est.	t-stat.	est.	t-stat.	est.	t-stat.
$\mathcal{E}_t$	0.206	2.497	0.031	0.475	−0.135	−1.876	0.111	1.487
Flow_t^{gov}	0.326	4.322	0.074	0.990	0.087	1.912	0.137	2.397
Flow_t^{money}	−0.024	−0.427	0.185	2.609	−0.058	−1.276	0.033	0.563
Flow_t^{corp}	−0.015	−0.379	0.005	0.167	0.157	2.524	0.172	2.508
Flow_t^{muni}	0.135	1.732	−0.025	−0.614	0.092	0.858	0.115	0.853

This table reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (\mathcal{E}_t, \text{Flow}_t^{gov}, \text{Flow}_t^{money}, \text{Flow}_t^{corp}, \text{Flow}_t^{muni})$. $\mathcal{E}_t$ is the model uncertainty measure, and Flow_t^{gov} (Flow_t^{money}, Flow_t^{corp}, Flow_t^{muni}) is the aggregate flows to the domestic government bond (money market, corporate bond, and municipal bond) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged fund returns of each type and VIX index in each regression. The sample spans from January 1998 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

Next, we report the IRFs of different fixed-income funds in Figure 7. We document sharp dynamic inflows to government bond funds. As Panel (a) suggests, a unit model uncertainty shock corresponds to more than 0.05 standard deviation increase in government bond fund inflows over the next month, and the dynamic response persists for more than 36 periods. On the contrary, the IRFs of other fixed-income fund flows are insignificant.

It is also worth noting that we do not observe a significant relationship between model

uncertainty and money market funds. The difference between money market and government bond funds is that the first type has a smaller duration and is more liquid, while the latter consists of government bonds of different maturities.

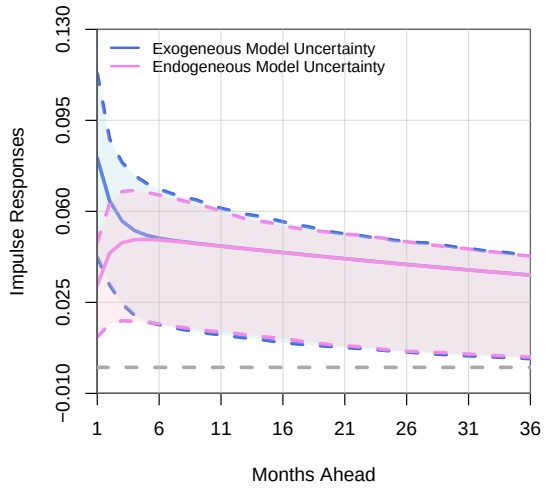
5.4 Fund Flow Responses to Other Uncertainty Measures

What is the uniqueness of model uncertainty? Previous analyses show that model uncertainty is significantly correlated with VIX and financial uncertainty, so these alternative uncertainty measures are likely to capture similar impulse responses of mutual fund flows. Hence, in this section, we study how other uncertainty measures affect mutual fund flows and compare their dynamic responses with the previous results.

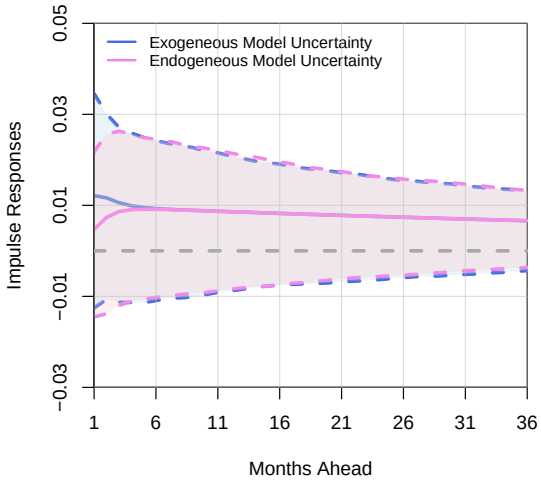
Figures IA.6 and IA.7 in the Internet Appendix plot the dynamic responses of four different types of equity fund flows to VIX and financial uncertainty shocks. Consistent with the previous analyses, we consider two identification assumptions, putting uncertainty measure as the first (exogenous shock) or last (endogenous response) variable in VAR. We also control the lagged model uncertainty in each regression. Unlike the responses to model uncertainty shocks, we do not observe significant IRFs of equity fund flows to VIX or financial uncertainty shocks. This evidence further supports that our model uncertainty measure captures significant investment risks confronted by equity investors.

We next consider the dynamic responses of fixed-income fund flows in Figures IA.8 and IA.9 of the Internet Appendix. Interestingly, we document significant inflows to money market funds after positive VIX shocks, but fixed-income fund flows do not respond to financial uncertainty shocks. In contrast, model uncertainty is not essential in money market funds. In other words, model uncertainty shocks primarily induce "flight to safety," while other volatility-based uncertainty measures are mainly related to "flight to liquidity."

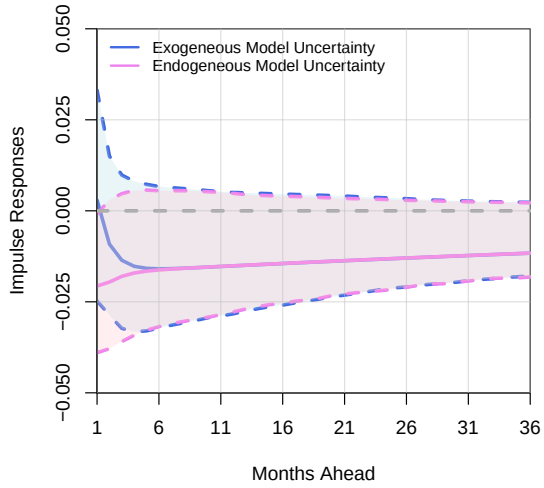
In summary, our model uncertainty measure captures some unique dynamic responses of fund flows, and notably, they are distinct from traditional volatility-based measures, in-



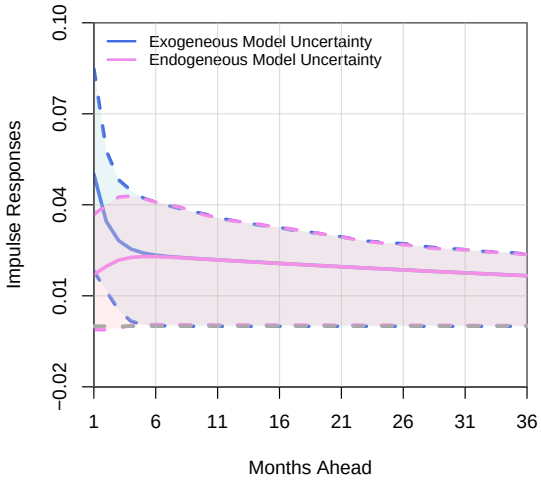
(a) Government-bond Fund Flows



(b) Money-market Fund Flows



(c) Corporate-bond Fund Flows



(d) Municipal-bond Fund Flows

Figure 7: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes to Model Uncertainty Shocks

This figure shows the dynamic impulse response functions (IRFs) of fund flows to model uncertainty shocks in VAR(1). The shaded area denotes the 90 percent standard error bands. We consider fixed-income fund flows with different investment objective codes (government bonds, money market, corporate bonds, and municipal bonds). We consider two identification assumptions, (1) by placing model uncertainty first in the VAR (exogenous shock, highlighted in blue) and (2) by placing model uncertainty as the last variable in the VAR (endogenous response, highlighted in purple). The dotted gray line corresponds to the zero impulse response. The data are monthly and span the period 1991:01 - 2020:12.

cluding VIX and financial uncertainty. In particular, we observe significant fund inflows to government bond funds and outflows from style and small-cap equity funds. Therefore, we conjecture that model uncertainty is a better measure of uncertainty to which mutual fund investors are risk averse.

6 Managing Model Uncertainty

How should investors manage model uncertainty in the cross-section? We propose a standard procedure to construct mean-variance efficient portfolios using Bayesian model averaging (BMA). When the SDF is $1-(\boldsymbol{f}-\mathbb{E}[\boldsymbol{f}])^\top \boldsymbol{b}$ (without model uncertainty), the tangency portfolio of the economy is $\boldsymbol{b}^\top \boldsymbol{f}$. Achieving mean-variance efficiency is equivalent to estimating risk prices $\boldsymbol{b}$. With model uncertainty, different models force different elements of $\boldsymbol{b}$ to become zeros ($\boldsymbol{b}_{-\boldsymbol{\gamma}}=\boldsymbol{0}$) and induce different posteriors for $\boldsymbol{b}_{\boldsymbol{\gamma}}$. Our proposed BMA estimator of $\boldsymbol{b}$ is then

$$\boldsymbol{b}_{bma} := \mathbb{E}[\boldsymbol{b} \mid \mathcal{D}] = \sum_{\boldsymbol{\gamma}} \mathbb{E}[\boldsymbol{b} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}] \times \mathbb{P}(\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}). \tag{18}$$

BMA takes the weighted average of the model-implied expectations, where the weights are posterior model probabilities. BMA deviates sharply from the traditional model selection, under which researchers always use a particular criterion (e.g., adjusted R^2s, information criteria, etc.) to select a single model and presume that the selected model is correct.

We evaluate the benefits of aggregating models by looking at the *out-of-sample* (OOS) Sharpe ratios of the tangency portfolio $\boldsymbol{b}_{bma}^\top \boldsymbol{f}$. Specifically, at the end of each month t, we estimate the risk prices $\boldsymbol{b}$ via BMA using the data from month $(t-35)$ to month t. We then update the tangency portfolio accordingly.

Column (1) of Table 6 presents the OOS Sharpe ratios of the tangency portfolio, constructed from the BMA estimates of risk prices. In comparison, we tabulate in columns

(2)–(7) the OOS Sharpe ratios from: (1) the model with the highest posterior probability (Top 1), (2) the full model that always includes all factors under consideration (All), (3) Carhart (1997) four-factor model (Carhart4), (4) Fama and French (2016) five-factor model (FF5), (5) Hou et al. (2015) q-factor model (HXZ4), and (6) Daniel et al. (2020) behavioral factor model (DHS3). We use the non-parametric Bootstrap to test the null hypothesis that BMA and the other model deliver an identical Sharpe ratio, i.e., $H_0 : \mathrm{SR}^2_{bma} = \mathrm{SR}^2_{\boldsymbol{\gamma}}$.[19]

Table 6: Out-of-Sample Model Performance

	(1) BMA	(2) Top 1	(3) All	(4) Carhart4	(5) FF5	(6) HXZ4	(7) DHS3
Full Sample: 07/1975 - 12/2020	1.818	1.750 **	1.772 -	0.736 ***	0.938 ***	1.135 ***	1.639 -
Subsample I: 07/1975 - 08/1990	2.327	2.226 **	2.293 -	1.014 ***	1.589 ***	1.853 *	2.142 -
Subsample II: 09/1990 - 10/2005	2.094	2.145 -	2.095 -	0.927 ***	0.916 ***	1.222 ***	2.072 -
Subsample III: 11/2005 - 12/2020	1.106	0.940 **	0.986 -	0.317 ***	0.452 ***	0.517 **	0.795 *
Low Model Uncertainty	2.572	2.565 -	2.568 -	1.288 ***	1.624 ***	1.829 ***	2.282 -
Middle Model Uncertainty	1.717	1.653 -	1.771 -	0.450 ***	0.677 ***	1.232 **	1.818 -
High Model Uncertainty	1.251	1.125 *	1.106 *	0.564 ***	0.584 ***	0.552 ***	0.897 **

This table reports the out-of-sample (annualised) Sharpe ratio of (1) BMA: the Bayesian model averaging of factor models, (2) Top 1: the top Bayesian model ranked by posterior model probabilities, (3) All: include all 14 factors, (4) Carhart4: Carhart (1997) four-factor model, (5) FF5: Fama and French (2016) five-factor model, (6) HXZ4: Hou et al. (2015) q-factor model, and (7) DHS3: the market factor plus two behavioural factors in Daniel et al. (2020). We also report the results on testing the null hypothesis that the Sharpe ratio of BMA is equal to the model $\boldsymbol{\gamma}$, i.e., $H_0 : \mathrm{SR}^2_{bma} = \mathrm{SR}^2_{\boldsymbol{\gamma}}$. We use the non-parametric Bootstrap to test the null hypothesis. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

We start with describing the full-sample performance, as shown in the first row of Table

[19] We draw 100,000 sample paths of $\{R_{\boldsymbol{\gamma},t^\star}, R_{bma,t^\star}\}_{t^\star=1}^{T}$ with replacement, where T is the sample size of the observed dataset. If the difference in Sharpe ratios between BMA and model $\boldsymbol{\gamma}$ in the observed dataset is larger than 90% (95%, 99%) of those in simulated datasets, we claim that H_0 is rejected by the data at 10% (5%, 1%) significance level.

6. First, BMA outperforms traditional factor models out-of-sample. The top Bayesian model (see column (2)) has an OOS Sharpe ratio of 1.75, which is virtually comparable to the model composed of all 14 factors (see column (3)). Second, BMA beats the top Bayesian model, but the distinction is marginal in the economic sense.

We further split the whole sample into three equal subsamples. Consistent with past literature, the performance of factor models tends to decline over time, and the drops in Sharpe ratios are particularly enormous from subsample II (September 1990 - October 2005) to subsample III (November 2005 - December 2020). Most interestingly, BMA is more valuable in the third subsample: its Sharpe ratio (1.106) is significantly higher than other models except for the one composed of all 14 factors.

The last three rows of Table 6 confirm that the performance of factor models, on average, declines as model uncertainty increases. Specifically, when model uncertainty is low, the top model and BMA have similar Sharpe ratios of around 2.57. In other words, when data overwhelmingly supports one dominant factor model, selecting models is equivalent to averaging them. On the contrary, it is particularly beneficial to incorporate model uncertainty into portfolio choice when model uncertainty is heightened. As the last row suggests, BMA has an OOS Sharpe ratio of 1.25, which is significantly more profitable than any other specifications.

In summary, Table 6 underscores the importance of considering model uncertainty when it is particularly high. In this scenario, BMA, which aggregates the information across all models, is salient for real-time portfolio choice.

7 Conclusions

In this paper, we introduce a Bayesian measure of model uncertainty in asset pricing models. The measure is defined based on the information entropy of asset pricing models' posterior probabilities. It is anchored within the range of zero to one; hence, we can detect when

model uncertainty is low or high in a transparent way.

Our Bayesian approach enjoys two favorable properties. First, under the null hypothesis of zero model uncertainty, our measure converges to zero in large samples. This is the conventional setting in which *a* true asset pricing model governing the cross-section of returns exists. More importantly, even when all models under study are misspecified due to omitted factors, our measure remains bounded far below one—its theoretical maximum—under the same null. Thus, we can interpret the high levels of model uncertainty observed in both the US and global equity markets as reliable evidence against the existence of a fixed "true" asset pricing model (instead of merely reflecting model misspecification).

Model uncertainty displays procyclical time-series variation and carries a significantly negative risk premium in equity markets. The magnitude of this risk premium is about half the size of the market portfolio. Thus, investors are willing to pay a premium to hedge against heightening model uncertainty.

Model uncertainty shocks carry meaningful information regarding investors' portfolio choices, proxied by aggregate fund flows into different asset classes. We document that positive model uncertainty shocks predict significant and persistent outflows from the US (small-cap and style) equity funds and inflows to treasury funds. The latter is consistent with the conventional wisdom of "flight to quality."

Departing from the existing literature on volatility-based uncertainty measures, we emphasize investors' uncertainty about their models in the context of cross-sectional asset pricing. The high degree of model uncertainty in the cross-section has implications beyond investment and portfolio choices. This phenomenon is also noteworthy in areas such as capital budgeting, performance evaluation, and risk attribution, where the asset pricing models under consideration can significantly impact decision-makers' choices.

References

Abramowitz, M. and I. A. Stegun (1965). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, Volume 55. Courier Corporation.

Asness, C. and A. Frazzini (2013). The devil in hml's details. *Journal of Portfolio Management 39*(4), 49–68.

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

Avramov, D., S. Cheng, L. Metzker, and S. Voigt (2023). Integrating factor models. *Journal of Finance, forthcoming*.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics 131*(4), 1593–1636.

Barber, B. M., X. Huang, and T. Odean (2016). Which factors matter to investors? evidence from mutual fund flows. *Review of Financial Studies 29*(10), 2600–2642.

Barillas, F., R. Kan, C. Robotti, and J. Shanken (2020). Model comparison with sharpe ratios. *Journal of Financial and Quantitative Analysis 55*(6), 1840–1874.

Barillas, F. and J. Shanken (2018). Comparing asset pricing models. *Journal of Finance 73*(2), 715–754.

Ben-David, I., J. Li, A. Rossi, and Y. Song (2022). What do mutual fund investors really care about? *Review of Financial Studies 35*(4), 1723–1774.

Berger, J. O. and L. R. Pericchi (1996). The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association 91*(433), 109–122.

Berk, J. B. and J. H. Van Binsbergen (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics 119*(1), 1–23.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica 77*(3), 623–685.

Bryzgalova, S., J. Huang, and C. Julliard (2023). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Journal of Finance 78*(1), 487–557.

Caballero, R. J. and A. Krishnamurthy (2008). Collective risk management in a flight to quality episode. *Journal of Finance 63*(5), 2195–2230.

Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance 52*(1), 57–82.

Chib, S. and X. Zeng (2020). 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. *Journal of Finance 75*(1), 551–577.

Chib, S., L. Zhao, and G. Zhou (2023). Winners from winners: A tale of risk factors. *Management Science*.

Chinco, A., S. M. Hartzmark, and A. B. Sussman (2022). A new test of risk factor relevance. *Journal of Finance 77*(4), 2183–2238.

Cochrane, J. H. (2005). *Asset pricing: Revised edition*. Princeton University Press.

Cochrane, J. H. and J. Saa-Requejo (2000). Beyond arbitrage: Good-deal asset price bounds in incomplete markets. *Journal of Political Economy 108*(1), 79–119.

Cremers, M. (2002). Stock return predictability: A bayesian model selection perspective. *Review of Financial Studies 15*(4), 1223–1249.

Daniel, K., D. Hirshleifer, and L. Sun (2020). Short-and long-horizon behavioral factors. *Review of Financial Studies 33*(4), 1673–1736.

Dew-Becker, I. and S. Giglio (2021+). Cross-sectional uncertainty and the business cycle: evidence from 40 years of options data. *American Economic Jounral: Macroeconomics forthcoming*.

Efron, B. (2012). *Large-scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*, Volume 1. Cambridge University Press.

Fama, E. F. and K. R. French (2016). Dissecting anomalies with a five-factor model. *Review of Financial Studies 29*(1), 69–103.

Fernández, C., E. Ley, and M. F. Steel (2001). Benchmark priors for bayesian model averaging. *Journal of Econometrics 100*(2), 381–427.

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

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica 57*(5), 1121–1152.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Guerrieri, V. and R. Shimer (2014). Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality. *American Economic Review 104*(7), 1875–1908.

Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*(2), 225–262.

Harvey, C. R. and G. Zhou (1990). Bayesian inference in asset pricing tests. *Journal of Financial Economics 26*(2), 221–254.

Hassan, T. A., S. Hollander, L. Van Lent, and A. Tahoun (2019). Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics 134*(4), 2135–2202.

Hou, K., C. Xue, and L. Zhang (2015). Digesting anomalies: An investment approach. *Review of Financial Studies 28*(3), 650–705.

James, W. and C. Stein (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, Berkeley, Calif., pp. 361–379. University of California Press.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Johnson, N. L., S. Kotz, and N. L. Johnson (1995). *Continuous univariate distributions, Second Edition*, Volume 2. John Weiley& Sons, Inc.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. *American Economic Review 105*(3), 1177–1216.

Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: an asset-allocation perspective. *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.

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

Liang, F., R. Paulo, G. Molina, M. A. Clyde, and J. O. Berger (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association 103*(481), 410–423.

Ludvigson, S. C., S. Ma, and S. Ng (2021). Uncertainty and business cycles: exogenous impulse or endogenous response? *American Economic Journal: Macroeconomics 13*(4), 369–410.

Manela, A. and A. Moreira (2017). News implied volatility and disaster concerns. *Journal of Financial Economics 123*(1), 137–162.

O'Hagan, A. (1995). Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 99–118.

Pástor, L. (2000). Portfolio selection and asset pricing models. *Journal of Finance 55*(1), 179–223.

Pástor, L. and R. F. Stambaugh (2000). Comparing asset pricing models: an investment perspective. *Journal of Financial Economics 56*(3), 335–381.

Shanken, J. (1987). A bayesian approach to testing portfolio efficiency. *Journal of Financial Economics 19*(2), 195–215.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial Studies 5*(1), 1–33.

Sirri, E. R. and P. Tufano (1998). Costly search and mutual fund flows. *Journal of Finance 53*(5), 1589–1622.

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

Vayanos, D. (2004). Flight to quality, flight to liquidity, and the pricing of risk.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press.

Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. In P. K. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, Chapter 29, pp. 233–243. Amsterdam: North-Holland/Elsevier.

Appendix A Factor Details

Fama-French five-factor model. Fama and French (2016) introduce a five-factor model that includes the market (MKT), size (SMB), value (HML), profitability (RMW), and investment (CMA) factors. The data come from Ken French's website.

Momentum. Jegadeesh and Titman (1993) find that stocks that perform well or poorly in the past three to 12 months continue their performance in the next three to 12 months. We download the momentum (MOM) factor from Ken French's data library.

The q-factor model. Hou et al. (2015) introduce a four-factor model that includes market excess return (MKT), size (ME), investment (IA), and profitability (ROE) factors. We download the last three factors from the authors' website.

Behavioral factors. Daniel et al. (2020) propose two behavioral factors. The short-term behavioral factor is based on the post-earnings announcement drift (PEAD) signal and captures the underreaction to earnings news in short horizons. The long-term behavioral factor (FIN) is based on the one-year net and five-year composite share issuance. We download these factors from the authors' website.

Quality-minus-junk. Asness et al. (2019) define high-quality stocks as ones with higher profits, faster growth, lower betas/volatilities, and larger payout ratio. We download the QMJ factor from the AQR data library.

Betting-against-beta. Frazzini and Pedersen (2014) constructs market-neutral betting-against-beta (BAB) factor that longs the low-beta stocks and shorts high-beta assets. We download the BAB factor from the AQR data library.

HML devil. Asness and Frazzini (2013) construct the value factor using more timely market value information. We download the HML Devil factor from the AQR data library.

Mispricing factors. Stambaugh and Yuan (2017) propose two mispricing factors that aggregate information across 11 prominent anomalies discovered in past literature. We download the data from Robert Stambaugh's website.

Appendix B Additional Tables

Table A1: Summary Statistics of 14 Factors

	Full Sample		Subsample I		Subsample II	
	Mean (%)	SR	Mean (%)	SR	Mean (%)	SR
MKT	7.36	0.43	5.54	0.40	9.18	0.47
ME	1.97	0.22	1.79	0.23	2.16	0.21
IA	3.92	0.66	6.36	1.38	1.48	0.21
ROE	6.21	0.91	8.50	1.72	3.92	0.47
SMB	1.24	0.14	0.89	0.12	1.58	0.16
HML	3.39	0.37	6.30	1.03	0.48	0.04
RMW	3.26	0.52	2.77	0.73	3.74	0.47
CMA	3.42	0.59	4.76	1.05	2.07	0.30
MOM	6.89	0.55	8.94	1.22	4.85	0.30
QMJ	4.31	0.63	3.76	0.94	4.85	0.55
BAB	10.10	1.00	11.99	1.81	8.21	0.65
HML devil	3.03	0.30	5.80	0.90	0.27	0.02
FIN	8.47	0.73	11.67	1.36	5.28	0.38
PEAD	7.57	1.30	9.34	2.00	5.80	0.85

This table reports the annualised mean returns and Sharpe ratios of 14 factors listed in Appendix A. The full sample starts from July 1972 and ends in December 2020. Subsample I starts from July 1972 and ends in September 1986. Subsample II covers the remaining.

Table A2: The Autocorrelation of Different Uncertainty Measures

	(1) $\mathcal{E}$	(2) Fin U	(3) Macro U	(4) Real U	(5) EPU I	(6) EPU II	(7) VIX
AR(1)	0.986	0.977	0.985	0.984	0.844	0.700	0.812
	(158.08)	(98.78)	(73.92)	(46.84)	(24.64)	(14.30)	(23.40)
Sample size	546	546	546	546	431	431	419

The table reports autocorrelations of uncertainty measures. $\mathcal{E}$ is our model uncertainty measure in the cross section. Financial, macro, and real uncertainty measures come from Jurado et al. (2015) and Ludvigson et al. (2021). EPU I and EPU II are two economic policy uncertainty sequences from Baker et al. (2016). VIX is CBOE volatility index. The t-statistics are computed using Newey-West standard errors with 36 lags.

Table A3: Simulation Study: Posterior Properties Without Misspecification

	$T = 750$				$T = 1500$				$T = 15000$			
scenario	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:												
$\gamma_{j,0} = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{j,0} = 0$	0.48	0.44	0.33	0.19	0.48	0.44	0.33	0.15	0.47	0.43	0.32	0.06
	(0.13)	(0.16)	(0.18)	(0.17)	(0.14)	(0.16)	(0.18)	(0.16)	(0.13)	(0.15)	(0.18)	(0.11)
Posterior Probabilities of Models $\mathbb{P}[\mathcal{M}_\gamma \mid \mathcal{D}]$:												
$\mathcal{M}_{\gamma_0} = \mathcal{M}_\gamma$	0.28	0.32	0.45	0.66	0.28	0.32	0.45	0.73	0.28	0.33	0.46	0.88
	(0.10)	(0.13)	(0.17)	(0.19)	(0.10)	(0.13)	(0.18)	(0.20)	(0.10)	(0.13)	(0.17)	(0.15)
$\mathcal{M}_{\gamma_0} \subset \mathcal{M}_\gamma$	0.24	0.23	0.18	0.11	0.24	0.23	0.18	0.09	0.24	0.22	0.18	0.04
	(0.09)	(0.11)	(0.13)	(0.13)	(0.10)	(0.11)	(0.13)	(0.12)	(0.09)	(0.11)	(0.13)	(0.09)
$\mathcal{M}_{\gamma_0} \not\subseteq \mathcal{M}_\gamma$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Model Uncertainty Measure $\mathcal{E}$:												
	0.38	0.37	0.32	0.24	0.38	0.36	0.32	0.20	0.38	0.37	0.32	0.11
	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)	(0.06)	(0.03)	(0.03)	(0.04)	(0.06)

This table presents simulation results on posterior probabilities of models being correct and factors entering the SDF. The full set of factors under consideration $\boldsymbol{f}$ are the Carhart four factors. We follow the assumptions in Proposition 4 and 5, treating the true SDF as $1 - (\boldsymbol{f}_{\boldsymbol{\gamma}_0} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}_0}])^\top \boldsymbol{b}$ where $\boldsymbol{b} = 0.5 \times \mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$). The true factors $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ are the Fama-French three factors.

We simulate 1,000 return samples according to $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}})$, where $\widehat{\boldsymbol{\Sigma}}$ is estimated from our sample of US equity returns (July 1972–December 2020), to mimic our empirical exercises. The mean vector, as determined by the true SDF, is $\boldsymbol{\mu}_0 = \widehat{\text{cov}}(\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}_0})\boldsymbol{b}$, where $\widehat{\text{cov}}(\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}_0})$ is a submatrix of $\widehat{\boldsymbol{\Sigma}}$. The simulated sample sizes are $T = 750$, 1,500, and 15,000 days.

For each of the 1,000 Monte Carlo samples, we calculate the posterior probability of each factor $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}], j = 1, \ldots, 4$, the posterior probability of each model $\mathbb{P}[\mathcal{M}_\gamma \mid \mathcal{D}]$, and the corresponding entropy measure $\mathcal{E}$. The table reports results from both the g-priors ($g = 2, 4, 16$) and the mixture of g-priors. Numbers without parenthesis are average across Monte Carlo samples; numbers in parenthesis are standard deviations across Monte Carlo samples.

Table A4: Simulation Study: Posterior Properties With Omitted Factors

	$T = 750$				$T = 1500$				$T = 15000$			
scenario	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g
Panel (A): true model (Carhart4)/ factors under consideration (FF5)												
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:												
$\gamma_{0,j} = 1$	0.91	0.91	0.89	0.88	0.95	0.95	0.94	0.92	1.00	1.00	1.00	1.00
	(0.17)	(0.18)	(0.22)	(0.25)	(0.13)	(0.13)	(0.16)	(0.21)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{0,j} = 0$	0.59	0.56	0.48	0.42	0.65	0.63	0.56	0.46	0.75	0.74	0.69	0.57
	(0.21)	(0.24)	(0.28)	(0.31)	(0.24)	(0.26)	(0.32)	(0.36)	(0.27)	(0.29)	(0.35)	(0.45)
Model Uncertainty Measure $\mathcal{E}$:												
	0.46	0.44	0.40	0.37	0.38	0.36	0.34	0.31	0.18	0.18	0.16	0.08
	(0.07)	(0.07)	(0.07)	(0.08)	(0.06)	(0.07)	(0.07)	(0.08)	(0.03)	(0.03)	(0.03)	(0.05)
Panel (B): true model (FF5)/ factors under consideration (Carhart4)												
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:												
$\gamma_{j,0} = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{j,0} = 0$	0.52	0.49	0.39	0.25	0.57	0.55	0.46	0.27	0.96	0.97	0.96	0.88
	(0.17)	(0.19)	(0.23)	(0.23)	(0.19)	(0.22)	(0.26)	(0.27)	(0.08)	(0.08)	(0.10)	(0.23)
Model Uncertainty Measure $\mathcal{E}$:												
	0.23	0.22	0.20	0.16	0.22	0.21	0.20	0.15	0.04	0.03	0.03	0.06
	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)

This table presents simulation results on factors' posterior probabilities of entering the SDF, under model misspecification. We follow the assumptions in Proposition 6, treating the true SDF as $1 - (\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}$ where $\boldsymbol{b} = 0.5 \times \mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_0$). True factors $\boldsymbol{f}_0$ and factors under consideration $\boldsymbol{f}$ are described accordingly in the titles of table panels. $\boldsymbol{f}_{\boldsymbol{\gamma}_0} = \boldsymbol{f} \cap \boldsymbol{f}_0$ is the subset of factors under consideration that enter the SDF.

We simulate 1,000 return samples according to $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}})$, where $\widehat{\boldsymbol{\Sigma}}$ is estimated from our sample of US equity returns (July 1972–December 2020). The mean vector, as determined by the true SDF, is $\boldsymbol{\mu}_0 = \widehat{\text{cov}}(\boldsymbol{R}, \boldsymbol{f}_0)\boldsymbol{b}$, which is also estimated from factor and testing asset returns, to mimic our empirical exercises. The simulated sample sizes are $T = 750$, 1,500, and 15,000 days.

For each of the 1,000 Monte Carlo sample, we calculate the posterior probability of each factor $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}], j = 1, \ldots, p$ ($p =$ in Panel (A); $p = 5$ in Panel (B)), and the corresponding entropy measure $\mathcal{E}$. The table reports results from both the g-priors ($g = 2, 4, 16$) and the mixture of g-priors. Numbers without parenthesis are average across Monte Carlo samples; numbers in parenthesis are standard deviations across Monte Carlo samples.

Table A5: Simulation Study: the Model Uncertainty Measure Under Misspecification (I)

	$p=4$	$p=5$	$p=6$	$p=7$	$p=8$	$p=9$
Panel (A): true model: market + momentum						
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:						
$\gamma_{0,j}=1$	1.00	1.00	1.00	1.00	1.00	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{0,j}=0$	0.47	0.48	0.47	0.44	0.48	0.50
	(0.37)	(0.36)	(0.33)	(0.32)	(0.33)	(0.33)
Model Uncertainty Measure $\mathcal{E}$:						
	0.37	0.43	0.50	0.54	0.53	0.53
	(0.09)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
$(p-p_{\boldsymbol{\gamma}_0})/p$	0.75	0.80	0.83	0.86	0.88	0.89
Panel (B): true model: Carhart4						
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:						
$\gamma_{0,j}=1$	0.96	0.88	0.89	0.88	0.92	0.94
	(0.13)	(0.25)	(0.23)	(0.24)	(0.19)	(0.17)
$\gamma_{0,j}=0$	0.24	0.42	0.41	0.38	0.46	0.49
	(0.16)	(0.31)	(0.29)	(0.26)	(0.31)	(0.33)
Model Uncertainty Measure $\mathcal{E}$:						
	0.25	0.37	0.43	0.48	0.46	0.46
	(0.10)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)
$(p-p_{\boldsymbol{\gamma}_0})/p$	0.25	0.40	0.50	0.57	0.62	0.67

This table presents simulation results on the model uncertainty measure under model misspecification. We follow the assumptions in Proposition 6 and 7, treating the true SDF as $1-(\boldsymbol{f}_0-\mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}$ where $\boldsymbol{b}=0.5\times\mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_0$). True factors $\boldsymbol{f}_0$ are described accordingly in the titles of table panels. When $p=4$, four factors are under consideration, namely $\boldsymbol{f}=\{\text{market}, \text{SMB}, \text{HML}, \text{RMW}\}$. For larger p, five additional factors are sequentially added into $\boldsymbol{f}$, in the order of $\{\text{CMA}, \text{QMJ}, \text{FIN}, \text{PEAD}, \text{BAB}\}$.

We simulate $1,000$ return samples according to $\mathcal{D}=\{\boldsymbol{R}_t\}_{t=1}^T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}})$, where $T=750$ days, and $\widehat{\boldsymbol{\Sigma}}$ is estimated from our sample of US equity returns (July 1972–December 2020). The mean vector, as determined by the true SDF, is $\boldsymbol{\mu}_0=\widehat{\text{cov}}(\boldsymbol{R}, \boldsymbol{f}_0)\boldsymbol{b}$, which is also estimated from factor and testing asset returns, to mimic our empirical exercises.

For each of the $1,000$ Monte Carlo sample, we calculate the posterior probability of each factor $\mathbb{P}[\gamma_j=1\mid\mathcal{D}]$, and the corresponding entropy measure $\mathcal{E}$, under the mixture of g-priors specification. Numbers without parenthesis are average across Monte Carlo samples; numbers in parenthesis are standard deviations across Monte Carlo samples.

Table A6: Simulation Study: the Model Uncertainty Measure Under Misspecification (II)

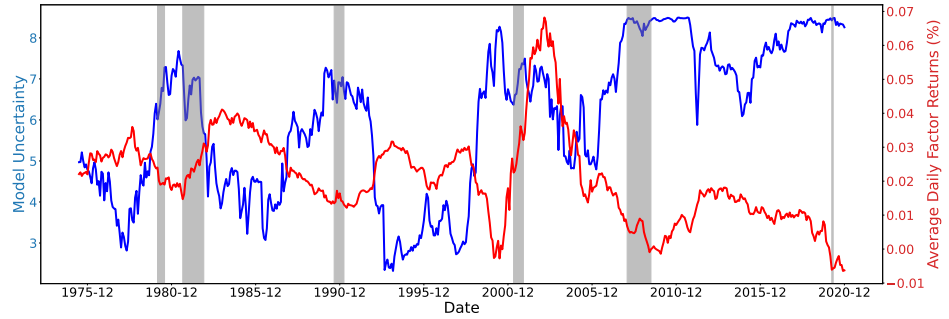
	MOM	RMW	CMA	FIN	PEAD	QMJ	BAB
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:							
$\gamma_{0,j} = 1$	0.94	1.00	1.00	1.00	1.00	0.99	1.00
	(0.17)	(0.02)	(0.00)	(0.01)	(0.01)	(0.06)	(0.01)
$\gamma_{0,j} = 0$	0.49	0.35	0.27	0.43	0.29	0.36	0.38
	(0.33)	(0.27)	(0.21)	(0.31)	(0.22)	(0.25)	(0.27)
Model Uncertainty Measure $\mathcal{E}$:							
	0.46	0.46	0.46	0.43	0.46	0.49	0.47
	(0.08)	(0.06)	(0.05)	(0.07)	(0.05)	(0.06)	(0.06)
$(p - p_{\boldsymbol{\gamma}_0})/p$	0.67	0.67	0.67	0.67	0.67	0.67	0.67

This table presents additional simulation results on the model uncertainty measure under model misspecification. We follow the assumptions in Proposition 6 and 7, treating the true SDF as $1-(\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}$ where $\boldsymbol{b} = 0.5 \times \mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_0$). True factors are the Fama-French three factors plus one additional factor (the name of which is at the top of each column). This additional factor will always be omitted. (The rationale behind only considering one omitted factor is due to the *linear* SDF assumption. If there are multiple omitted factors, a linear combination of them with their risk prices as coefficients create one omitted factors.) Factors under consideration, namely $\boldsymbol{f}$, are the factors in $\{\text{market}, \text{SMB}, \text{HML}, \text{MOM}, \text{RMW}, \text{CMA}, \text{QMJ}, \text{FIN}, \text{PEAD}, \text{BAB}\}$ *excluding* the factor defined by each column name.

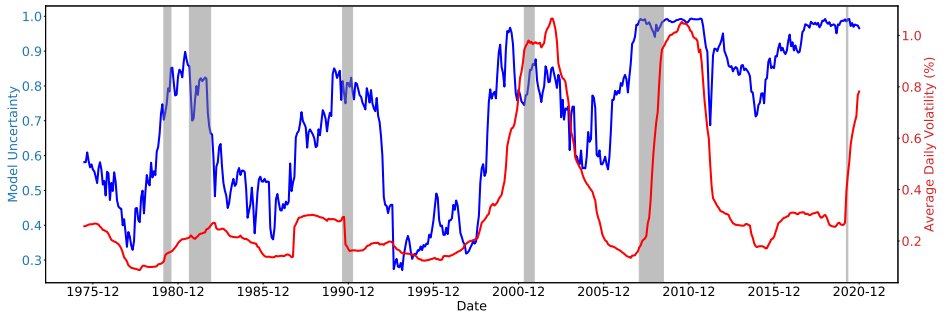
We simulate $1,000$ return samples according to $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}})$, where $T = 750$ days, and $\widehat{\boldsymbol{\Sigma}}$ is estimated from our sample of US equity returns (July 1972–December 2020). The mean vector, as determined by the true SDF, is $\boldsymbol{\mu}_0 = \widehat{\text{cov}}(\boldsymbol{R}, \boldsymbol{f}_0)\boldsymbol{b}$, which is also estimated from factor and testing asset returns, to mimic our empirical exercises.

For each of the $1,000$ Monte Carlo sample, we calculate the posterior probability of each factor $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$, and the corresponding entropy measure $\mathcal{E}$, under the mixture of g-priors specification. Numbers without parenthesis are average across Monte Carlo samples; numbers in parenthesis are standard deviations across Monte Carlo samples.

Appendix C Additional Figures



(a) Average (Daily) Return



(b) Average (Daily) Volatility

Figure A1: Time-Series of Average (Daily) Mean Return and Volatility of 14 Factors

The figures plot the time-series of (a) average daily returns of factors and (b) average daily factor volatility, and these statistics are estimated using the daily factor returns in the past 36 months.

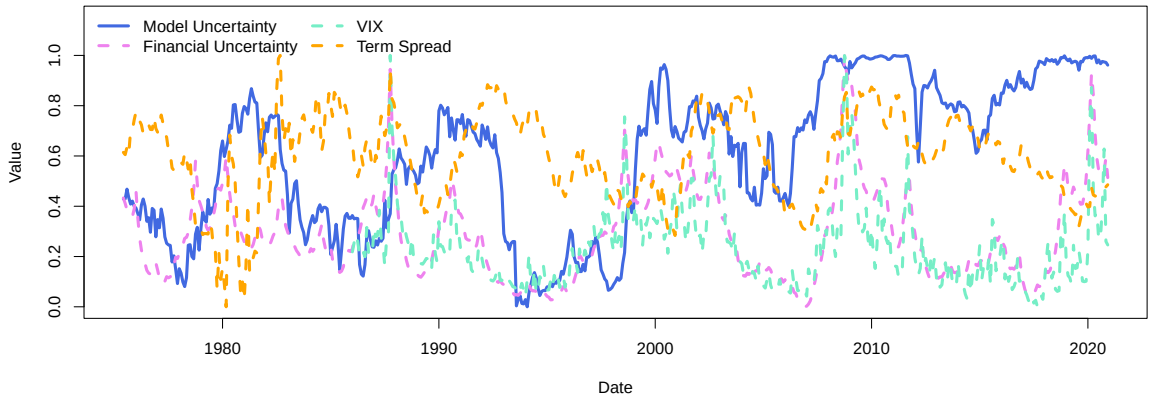


Figure A2: Time-Series of Model Uncertainty, Financial Uncertainty, VIX, and Term Spread

Internet Appendix for:

Model Uncertainty in the Cross Section

Jiantao Huang[a] Ran Shi[b,*]

[a] *University of Hong Kong*

[b] *University of Colorado Boulder*

Abstract

The Internet Appendix provides additional propositions, proofs, tables, figures, and empirical results supporting the main text.

[*] *Email addresses*: huangjt@hku.hk (Jiantao Huang) and ran.shi@colorado.edu (Ran Shi).

IA.1 Empirical Bayes and the Sharpe Ratios

Throughout our Bayesian inference exercises, we only assign priors to $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and treat the variance-covariance matrix $\boldsymbol{\Sigma}$ as known when deriving the posterior probabilities $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$ and calculating our model uncertainty measures. Then we substitute $\boldsymbol{\Sigma}$ with its in-sample consistent estimator[2]

$$\boldsymbol{S} = 1/T \sum_{t=1}^{T} (\boldsymbol{R}_t - \overline{\boldsymbol{R}})(\boldsymbol{R}_t - \overline{\boldsymbol{R}})^\top,$$

where $\overline{\boldsymbol{R}} = 1/T \sum_{t=1}^{T} \boldsymbol{R}_t$, when calculating Bayes factors (and the model uncertainty $\mathcal{E}$). This substitution drives our posterior inference through the in-sample Sharpe ratios.

Denote by $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}$ the maximal squared in-sample Sharpe ratio under model $\mathcal{M}_{\boldsymbol{\gamma}}$ after plugging in estimators for $\boldsymbol{\Sigma}$. Define $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ and $\boldsymbol{S}_{\boldsymbol{\gamma}}$ the population and sample variance-covariance matrices of $\boldsymbol{f}_{\boldsymbol{\gamma}}$ (which are submatrices of $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$), then

$$\mathrm{SR}^2_{\boldsymbol{\gamma}} = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}},$$

$$\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{S}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}.$$

Proposition IA.1 covers the econometric property of $\mathrm{SR}^2_{\boldsymbol{\gamma}}$, treating $\boldsymbol{\Sigma}$ as known.

Proposition IA.1. *(the expectations of the Sharpe ratios) If there exists a true linear factor SDF model* $\mathcal{M}_{\boldsymbol{\gamma}_0} : m_{\boldsymbol{\gamma}_0} = 1 - (\boldsymbol{f}_{\boldsymbol{\gamma}_0} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}_0}])^\top \boldsymbol{b}_{\boldsymbol{\gamma}_0}$ *generating the observed return data, that is,* $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T \overset{\text{iid}}{\sim} \mathcal{N}(\mathrm{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}_0}]\boldsymbol{b}_{\boldsymbol{\gamma}_0}, \boldsymbol{\Sigma})$, *then the maximal in-sample Sharpe ratio of* $\boldsymbol{f}_{\boldsymbol{\gamma}}$*, namely factors under consideration for model* $\mathcal{M}_{\boldsymbol{\gamma}}$*, satisfies*

$$\mathbb{E}\left[\mathrm{SR}^2_{\boldsymbol{\gamma}}\right] = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \left(\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0}\right] - \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\right) \boldsymbol{b}_{\boldsymbol{\gamma}_0} + \frac{p_{\boldsymbol{\gamma}}}{T}$$

Proof. When the data are generated from model $\mathcal{M}_{\boldsymbol{\gamma}_0}$, for any $\boldsymbol{\gamma}$,

$$\boldsymbol{f}_{\boldsymbol{\gamma},t} \overset{\text{iid}}{\sim} \mathcal{N}\left(\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0},\ \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\right), \quad t = 1, \ldots, T,$$

where $\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} = \mathrm{cov}[\boldsymbol{f}_{\boldsymbol{\gamma}}, \boldsymbol{f}_{\boldsymbol{\gamma}_0}]$ is a $p_{\boldsymbol{\gamma}} \times p_{\boldsymbol{\gamma}_0}$ matrix. As a result,

$$\boldsymbol{z}_{\boldsymbol{\gamma}} = \sqrt{T} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-\frac{1}{2}} \left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0}\right) \sim \mathcal{N}\left(\boldsymbol{0},\ \boldsymbol{I}_{p_{\boldsymbol{\gamma}}}\right).$$

[2]The use of point estimators to replace parameters in posterior distributions dates back to the seminal James-Stein estimator (James and Stein, 1961). For a monograph on modern empirical Bayes methods, see Efron (2012).

Define $\delta_{\boldsymbol{\gamma}} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0}$, then

$$\begin{aligned} T\mathrm{SR}_{\boldsymbol{\gamma}}^2 - T\delta_{\boldsymbol{\gamma}} =& T\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0} \\ =& T\left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0}\right)^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0}\right) + 2T\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0}\right) \\ =& \underbrace{\boldsymbol{z}_{\boldsymbol{\gamma}}^\top \boldsymbol{z}_{\boldsymbol{\gamma}}}_{\sim\chi^2(p_{\boldsymbol{\gamma}})} + \underbrace{\left(2\sqrt{T}\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}{}^{-\frac{1}{2}}\right) \boldsymbol{z}_{\boldsymbol{\gamma}}}_{\sim\mathcal{N}(0,\,4T\delta_{\boldsymbol{\gamma}})}. \end{aligned} \tag{IA.1}$$

Since the expectation of $\boldsymbol{z}_{\boldsymbol{\gamma}}^\top \boldsymbol{z}_{\boldsymbol{\gamma}}$ equals $p_{\boldsymbol{\gamma}}$,

$$\mathbb{E}[\mathrm{SR}_{\boldsymbol{\gamma}}^2] = \delta_{\boldsymbol{\gamma}} + \frac{p_{\boldsymbol{\gamma}}}{T}. \tag{IA.2}$$

Now define $\delta_{\boldsymbol{\gamma}_0} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}_0} \boldsymbol{b}_{\boldsymbol{\gamma}_0}$. By definition,

$$\delta_{\boldsymbol{\gamma}_0} - \delta_{\boldsymbol{\gamma}} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_0} - \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0}\right) \boldsymbol{b}_{\boldsymbol{\gamma}_0} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right] \boldsymbol{b}_{\boldsymbol{\gamma}_0}. \tag{IA.3}$$

As a result,

$$\delta_{\boldsymbol{\gamma}} = \delta_{\boldsymbol{\gamma}_0} - \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right] \boldsymbol{b}_{\boldsymbol{\gamma}_0} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \left(\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0}\right] - \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\right) \boldsymbol{b}_{\boldsymbol{\gamma}_0}.$$

Plugging this expression for $\delta_{\boldsymbol{\gamma}}$ into equation (IA.2), the proof is completed. □

According to Proposition IA.1, in addition to including more factors (increasing $p_{\boldsymbol{\gamma}}$), the maximal in-sample Sharpe ratio under model $\mathcal{M}_{\boldsymbol{\gamma}}$ is expected to become larger when (1) factors in the true model have large risk prices ($\|\boldsymbol{b}_{\boldsymbol{\gamma}_0}\|$ is large); (2) factors defined by $\mathcal{M}_{\boldsymbol{\gamma}}$ are close to span the true set of factors ($\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]$ is small). If $\mathcal{M}_{\boldsymbol{\gamma}}$ includes all true factors in $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$, the conditional variance $\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]$ is zero. Under this scenario, $\mathbb{E}\left[\mathrm{SR}_{\boldsymbol{\gamma}}^2\right] = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \left(\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0}\right]\right) \boldsymbol{b}_{\boldsymbol{\gamma}_0} + p_{\boldsymbol{\gamma}}/T \geq \mathbb{E}\left[\mathrm{SR}_{\boldsymbol{\gamma}_0}^2\right]$ (for $p_{\boldsymbol{\gamma}} \geq p_{\boldsymbol{\gamma}_0}$). Without the penalty on model sizes, model comparison based on Sharpe ratios will artificially favor dense models.

Proposition IA.2 summarizes properties of the Sharpe ratio $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ under empirical Bayes.

Proposition IA.2. *When the covariance matrix of returns $\boldsymbol{\Sigma}$ is replaced by its method of moments estimator, the corresponding maximal squared in-sample Sharpe ratio under model $\mathcal{M}_{\boldsymbol{\gamma}}$, denoted by $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$, satisfies*

1. *if $T > p_{\boldsymbol{\gamma}} + 2$,*

$$\frac{\mathbb{E}\left[\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2\right] - \mathbb{E}\left[\mathrm{SR}_{\boldsymbol{\gamma}}^2\right]}{\mathbb{E}\left[\mathrm{SR}_{\boldsymbol{\gamma}}^2\right]} = \frac{p_{\boldsymbol{\gamma}} + 2}{T - p_{\boldsymbol{\gamma}} - 2};$$

2. *There exist sequences $l_T = O(1/\sqrt{T})$ and $u_T = O(\sqrt{T})$, such that for all ψ satisfying $l_T < \psi + \sqrt{p_{\boldsymbol{\gamma}}} <$*

u_T, with probability at least $1-2e^{-\psi^2/2}$,

$$\frac{\left|\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2-\mathrm{SR}_{\boldsymbol{\gamma}}^2\right|}{\mathrm{SR}_{\boldsymbol{\gamma}}^2}\leq 3\left(\psi+\sqrt{p_{\boldsymbol{\gamma}}}\right)l_T.$$

Based on Proposition IA.2, the expected value of $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ converges to that of $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ (presented in Proposition IA.1), as the sample size T becomes large (and the model dimension does not scale with T). According to the second part of Proposition IA.2, the distribution of $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ concentrates heavily around that of $\mathrm{SR}_{\boldsymbol{\gamma}}^2$. A byproduct of this result is that $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ must be a consistent estimator. Without further clarification, $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ will be replaced with $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ throughout our empirical studies; all theoretical results involving $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ will be derived accounting for this replacement (i.e., we only present sample-based instead of population-based results.)

The proof of Proposition IA.2 is as follows.

Proof. Let $\boldsymbol{S}_{\boldsymbol{\gamma}}$ be the sample counterpart of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$, indexed from the estimator $\boldsymbol{S}$. Then $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2=\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top\boldsymbol{S}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}$, and $T\boldsymbol{S}_{\boldsymbol{\gamma}}\sim\mathcal{W}_{p_{\boldsymbol{\gamma}}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}},T-1\right)$, a Wishart distribution with $(T-1)$ degrees of freedom and a scale matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Recall from the proof of Proposition IA.1, $\sqrt{T}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}\sim\mathcal{N}\left(\sqrt{T}\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0}\boldsymbol{b}_{\boldsymbol{\gamma}_0},\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\right)$, and it is well-known that $\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}\perp\boldsymbol{S}_{\boldsymbol{\gamma}}$, thus

$$(T-1)\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2=(T-1)\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top\boldsymbol{S}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}\sim\mathcal{T}^2(p_{\boldsymbol{\gamma}},T-1;T\delta_{\boldsymbol{\gamma}}),$$

a non-central Hotelling's $\mathcal{T}^2$ distribution with degree-of-freedom parameters $p_{\boldsymbol{\gamma}}$ and $(T-1)$, as well as a non-centrality parameter $T\delta_{\boldsymbol{\gamma}}$.[3] Equivalently,

$$\frac{T-p_{\boldsymbol{\gamma}}}{p_{\boldsymbol{\gamma}}}\times\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2\sim\mathcal{F}(p_{\boldsymbol{\gamma}},T-p_{\boldsymbol{\gamma}};T\delta_{\boldsymbol{\gamma}}),\tag{IA.4}$$

in which $\mathcal{F}(p_{\boldsymbol{\gamma}},T-p_{\boldsymbol{\gamma}},T\delta_{\boldsymbol{\gamma}})$ denotes a non-central F-distribution with 1) $p_{\boldsymbol{\gamma}}$ degrees of freedom and a non-centrality parameter $\delta_{\boldsymbol{\gamma}}$ for the numerator and 2) $(T-p_{\boldsymbol{\gamma}})$ degrees of freedom for the denominator. According to Johnson et al. (1995, Page 481), when $T>p_{\boldsymbol{\gamma}}+2$,

$$\mathbb{E}\left[\frac{T-p_{\boldsymbol{\gamma}}}{p_{\boldsymbol{\gamma}}}\times\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2\right]=\frac{(T-p_{\boldsymbol{\gamma}})(p_{\boldsymbol{\gamma}}+T\delta_{\boldsymbol{\gamma}})}{p_{\boldsymbol{\gamma}}(T-p_{\boldsymbol{\gamma}}-2)}.$$

[3] $\delta_{\boldsymbol{\gamma}}$ is defined in the proof of Proposition IA.1. We will use this notation without referencing back to the previous proofs from now on.

Solving for $\mathbb{E}[\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2]$ and comparing the formula with equation (IA.2),

$$\mathbb{E}\left[\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2\right] = \frac{T\delta_{\boldsymbol{\gamma}} + p_{\boldsymbol{\gamma}}}{T - p_{\boldsymbol{\gamma}} - 2} = \mathbb{E}\left[\mathrm{SR}_{\boldsymbol{\gamma}}^2\right]\left(\frac{T}{T - p_{\boldsymbol{\gamma}} - 2}\right).$$

Simple algebra gives the formula in bullet point one.

Now we prove the second part of the proposition. To begin with, as $T\boldsymbol{S}_{\boldsymbol{\gamma}} \sim \mathcal{W}_{p_{\boldsymbol{\gamma}}}(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}, T-1)$, we can rewrite $\boldsymbol{S}_{\boldsymbol{\gamma}}$ as $\boldsymbol{S}_{\boldsymbol{\gamma}} = \boldsymbol{\Omega}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Omega}_{\boldsymbol{\gamma}}/T$, in which the t-th row of matrix $\boldsymbol{\Omega}_{\boldsymbol{\gamma}} \in \mathbb{R}^{(T-1)\times p_{\boldsymbol{\gamma}}}$, denoted by $\boldsymbol{\omega}_{\boldsymbol{\gamma},t}$, is an i.i.d. draw from $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$. We can then write $\boldsymbol{\Omega}_{\boldsymbol{\gamma}} = \boldsymbol{Z}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{\frac{1}{2}}$, and every element of the random matrix $\boldsymbol{Z}_{\boldsymbol{\gamma}} \in \mathbb{R}^{(T-1)\times p_{\boldsymbol{\gamma}}}$ are i.i.d. standard normal random variables. Thus, $\boldsymbol{S}_{\boldsymbol{\gamma}} = \boldsymbol{\Omega}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Omega}_{\boldsymbol{\gamma}}/T = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{\frac{1}{2}} \boldsymbol{Z}_{\boldsymbol{\gamma}}^\top \boldsymbol{Z}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{\frac{1}{2}}/T$, and

$$\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{S}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} = T\left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-\frac{1}{2}} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}\right)^\top \left(\boldsymbol{Z}_{\boldsymbol{\gamma}}^\top \boldsymbol{Z}_{\boldsymbol{\gamma}}\right)^{-1} \left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-\frac{1}{2}} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}\right).$$

From Theorem 6.1 of Wainwright (2019), if $\sigma_{\min}$ and $\sigma_{\max}$ are the smallest and the largest singular values of $\boldsymbol{Z}_{\boldsymbol{\gamma}}$, for all $\varepsilon > 0$,

$$1 - \varepsilon - \sqrt{\frac{p_{\boldsymbol{\gamma}}}{T-1}} \leq \frac{\sigma_{\min}}{\sqrt{T-1}} \leq \frac{\sigma_{\max}}{\sqrt{T-1}} \leq 1 + \varepsilon + \sqrt{\frac{p_{\boldsymbol{\gamma}}}{T-1}}$$

with a probability at least $1 - 2e^{-(T-1)\varepsilon^2/2}$. Taking $\varepsilon = \psi(T-1)^{-1/2}$, and defining $\eta = (\psi + \sqrt{p_{\boldsymbol{\gamma}}})(T-1)^{-1/2}$, if $\eta \in (-3/2,\, 3/2)$,

$$\left(\frac{1}{1 + \varepsilon + \sqrt{p_{\boldsymbol{\gamma}}/(T-1)}}\right)^2 = \frac{1}{(1+\eta)^2} \geq 1 - 2\eta, \quad \left(\frac{1}{1 - \varepsilon - \sqrt{p_{\boldsymbol{\gamma}}/(T-1)}}\right)^2 = \frac{1}{(1-\eta)^2} \leq 1 + 2\eta.$$

As a result, with probability at least $1 - 2e^{-\psi^2/2}$, both

$$\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{S}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} \leq \frac{T}{\sigma_{\min}^2} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} \leq \left(\frac{T}{T-1}\right) \frac{\mathrm{SR}_{\boldsymbol{\gamma}}^2}{(1-\eta)^2} \leq \left(\frac{T}{T-1}\right)(1+2\eta)\mathrm{SR}_{\boldsymbol{\gamma}}^2 < (1+3\eta)\mathrm{SR}_{\boldsymbol{\gamma}}^2,$$

(the last "<" holds when $\eta > 1/(T-3)$) and

$$\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{S}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} \geq \frac{T}{\sigma_{\max}^2} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} \geq \left(\frac{T}{T-1}\right) \frac{\mathrm{SR}_{\boldsymbol{\gamma}}^2}{(1+\eta)^2} \geq \left(\frac{T}{T-1}\right)(1-2\eta)\mathrm{SR}_{\boldsymbol{\gamma}}^2 > (1-3\eta)\mathrm{SR}_{\boldsymbol{\gamma}}^2.$$

Now, to sum up, for all $1/(T-3) < \eta < 3/2$, with a probability at least $1 - 2e^{-\psi^2/2}$,

$$\frac{|\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 - \mathrm{SR}_{\boldsymbol{\gamma}}^2|}{\mathrm{SR}_{\boldsymbol{\gamma}}^2} < 3\eta = 3(\psi + \sqrt{p_{\boldsymbol{\gamma}}})\sqrt{\frac{1}{T-1}} < 3(\psi + \sqrt{p_{\boldsymbol{\gamma}}})\frac{\sqrt{T-1}}{T-3}$$

The condition $1/(T-3) < \eta < 3/2$ implies $\sqrt{T-1}/(T-3) < \psi + \sqrt{p_{\boldsymbol{\gamma}}} < 3/2\sqrt{T-1}$. Define $l_T = \sqrt{T-1}/(T-3)$ and $u_T = 3/2\sqrt{T-1}$, we arrive at the stated result. □

IA.2 Proofs

Of note, all notations defined early on in this part will carry through for later proofs.

IA.2.1 Proof of Proposition 1

We begin the proof of Proposition 1 with the following lemma.

Lemma IA.1. *Define* $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \mathrm{var}[\boldsymbol{f}_{\boldsymbol{\gamma}}]$, $\boldsymbol{C}_{\boldsymbol{\gamma}} = \mathrm{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]$, *and* $\boldsymbol{\Sigma} = \mathrm{var}[\boldsymbol{R}]$, *then*

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \begin{pmatrix} \boldsymbol{I}_{p_{\boldsymbol{\gamma}}} \\ \boldsymbol{0}_{(N-p_{\boldsymbol{\gamma}})} \end{pmatrix}, \qquad \boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\gamma}}, \qquad \boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}.$$

Proof. Recall that under our specification, it is always that $\boldsymbol{f}_{\boldsymbol{\gamma}} \subseteq \boldsymbol{f} \subseteq \boldsymbol{R}$. Without loss of generality, the vector $\boldsymbol{R}$ can be arranged as

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{f}_{\boldsymbol{\gamma}} \\ \boldsymbol{f}_{-\boldsymbol{\gamma}} \\ \boldsymbol{r}^e \end{pmatrix}$$

where $\boldsymbol{r}^e$ is a vector of test assets that are excess returns themselves but are excluded from factors under consideration (i.e., $\boldsymbol{f}$). Then

$$\boldsymbol{\Sigma} = \mathrm{var}[\boldsymbol{R}] = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} & \boldsymbol{U}_{\boldsymbol{\gamma}}^{\top} \\ \boldsymbol{U}_{\boldsymbol{\gamma}} & \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} \end{pmatrix}, \quad \boldsymbol{C}_{\boldsymbol{\gamma}} = \mathrm{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}] = \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \\ \boldsymbol{U}_{\boldsymbol{\gamma}} \end{pmatrix},$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \mathrm{var}[\boldsymbol{f}_{\boldsymbol{\gamma}}], \quad \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} = \mathrm{var}\left[\begin{pmatrix} \boldsymbol{f}_{-\boldsymbol{\gamma}} \\ \boldsymbol{r}^e \end{pmatrix}\right], \quad \boldsymbol{U}_{\boldsymbol{\gamma}} = \mathrm{cov}\left[\begin{pmatrix} \boldsymbol{f}_{-\boldsymbol{\gamma}} \\ \boldsymbol{r}^e \end{pmatrix}, \boldsymbol{f}_{\boldsymbol{\gamma}}\right].$$

Inverting $\boldsymbol{\Sigma}$ blockwise, we have

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} (\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} & -\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top (\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \\ -\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}} (\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} & (\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \end{pmatrix}.$$

or exchanging the two off-diagonal blocks and taking transposes,

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} (\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} & -(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \\ -(\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} & (\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \end{pmatrix}.$$

Thus

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}} &= \begin{pmatrix} (\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} & -(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \\ -(\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} & (\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \\ \boldsymbol{U}_{\boldsymbol{\gamma}} \end{pmatrix} \\ &= \begin{pmatrix} (\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - (\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}})^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}} \\ -(\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{-\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}} + (\boldsymbol{\Sigma}_{-\boldsymbol{\gamma}} - \boldsymbol{U}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}}^\top)^{-1} \boldsymbol{U}_{\boldsymbol{\gamma}} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{I}_{p_{\boldsymbol{\gamma}}} \\ \boldsymbol{0}_{(N-p_{\boldsymbol{\gamma}})} \end{pmatrix}, \end{aligned}$$

which directly implies that $\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\gamma}}$ and that $\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. □

Under this lemma, we prove Proposition 1 as follows.

Proof. Under our distributional assumption

$$[\boldsymbol{R}_t \mid \boldsymbol{b}_{\boldsymbol{\gamma}}, \mathcal{M}_{\boldsymbol{\gamma}}] \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{b}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}), \quad t = 1, \ldots, T,$$

and under our g-prior specification,

$$[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g] \sim \mathcal{N}\left(\boldsymbol{0}, \frac{g}{T} \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right).$$

Integrating out $\boldsymbol{b}_{\boldsymbol{\gamma}}$, we have

$$[\boldsymbol{R}_1^\top, \ldots, \boldsymbol{R}_T^\top]^\top \triangleq \boldsymbol{R}_{[1:T]} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_T \otimes \boldsymbol{\Sigma} + \frac{g}{T} (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}}) \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1} (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^\top\right),$$

where $\otimes$ performs the matrix Kronecker product. As a result,

$$\begin{aligned}\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] = &\exp\left\{-\frac{1}{2}\boldsymbol{R}_{[1:T]}^{\top}\left[\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1} + \frac{g}{T}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^{\top}\right]^{-1}\boldsymbol{R}_{[1:T]}\right\} \\ &\times \left|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1} + \frac{g}{T}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^{\top}\right|^{-\frac{1}{2}}(2\pi)^{-\frac{NT}{2}}. \qquad \text{(IA.5)}\end{aligned}$$

To simplify (IA.5), first, by the Sherman-Morrison-Woodbury formula,[4]

$$\begin{aligned}&\left[\boldsymbol{I}_T \otimes \boldsymbol{\Sigma} + \frac{g}{T}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^{\top}\right]^{-1} \\ =&\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1} - \\ &[\mathbf{1}_T \otimes (\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}})]\left(\frac{T}{g}\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} + (\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^{\top}(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})\right)^{-1}[\mathbf{1}_T^{\top} \otimes (\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1})] \\ =&\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1} - \frac{g}{(1+g)T}[\mathbf{1}_T \otimes (\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}})]\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}[\mathbf{1}_T^{\top} \otimes (\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1})].\end{aligned}$$

Second, by the generalized Sylvester's theorem for determinants,[5]

$$\begin{aligned}&\left|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma} + \frac{g}{T}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^{\top}\right| \\ =&\frac{|Tg^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} + (\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^{\top}(\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})|}{|Tg^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}| \times |\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}|} \\ =&\frac{|(g^{-1}+1)\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}|}{|g^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}| \times |\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}|} \qquad (\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \text{ according to Lemma IA.1}) \\ =&\frac{(1+g)^{p_{\boldsymbol{\gamma}}}}{|\boldsymbol{\Sigma}^{-1}|^{T}}.\end{aligned}$$

[4] $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1}$, whenever the matrix multiplication and inverse are well-defined for the matrices A, U, C, V.

[5] $|X+ACB| = |X|\times|C|\times|C^{-1}+BX^{-1}A|$, whenever the matrix multiplication and inverse are well-defined for the matrices A, B, C, X.

Plugging these two results above back to equation (IA.5), we get

$$
\begin{aligned}
&\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] \\
&= \exp\left\{-\frac{1}{2}\boldsymbol{R}_{[1:T]}^{\top}\left[\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1} - \frac{g}{(1+g)T}[\mathbf{1}_T \otimes (\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}})]\left(\boldsymbol{C}_{\gamma}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}[\mathbf{1}_T^{\top} \otimes (\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1})]\right]\boldsymbol{R}_{[1:T]}\right\} \\
&\quad \times \frac{|\boldsymbol{\Sigma}^{-1}|^{\frac{T}{2}}}{(1+g)^{\frac{p_{\boldsymbol{\gamma}}}{2}}(2\pi)^{\frac{NT}{2}}} \\
&= \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{R}_t^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_t + \frac{g}{1+g}\frac{T}{2}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_{\boldsymbol{\gamma},t}\right)^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{f}_{\boldsymbol{\gamma},t}\right)\right\}\frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}} \\
&= \exp\left\{-\frac{1}{2}\left[T\mathrm{tr}\left(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right) + T\overline{\boldsymbol{R}}^{\top}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}\right] + \frac{g}{1+g}\frac{T}{2}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}\right\}\frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}} \\
&= \exp\left\{-\frac{T}{2}\mathrm{tr}\left(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\right) - \frac{T}{2}\Big(\underbrace{\overline{\boldsymbol{R}}^{\top}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}}_{\mathrm{SR}^2_{\max}} - \frac{g}{1+g}\underbrace{\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}}_{\mathrm{SR}^2_{\boldsymbol{\gamma}}}\Big)\right\}\frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}}
\end{aligned}
$$

where $\overline{\boldsymbol{R}} = \left(\sum_{t=1}^{T}\boldsymbol{R}_t\right)/T$, $\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} = \left(\sum_{t=1}^{T}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma},t}\right)/T$, $\boldsymbol{S} = \sum_{t=1}^{T}(\boldsymbol{R}_t - \overline{\boldsymbol{R}})(\boldsymbol{R}_t - \overline{\boldsymbol{R}})^{\top}/T$; the second equation above relies on the fact that $\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\gamma}}$ and that $\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ according to Lemma IA.1. □

IA.2.2 Proof of Proposition 2

Proof. According to our distributional assumption for the returns and the g-prior specification (as well as Lemma IA.1),

$$
[\boldsymbol{R}_t \mid \boldsymbol{b}_{\boldsymbol{\gamma}}, \mathcal{M}_{\boldsymbol{\gamma}}] \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{C}_{\boldsymbol{\gamma}}\boldsymbol{b}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}), \quad t = 1, \ldots, T; \qquad [\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g] \sim \mathcal{N}\left(\mathbf{0}, \frac{g}{T}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\right).
$$

The posterior distribution of $\boldsymbol{b}_{\boldsymbol{\gamma}}$ then satisfies

$$
\mathbb{P}[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}, g] \propto \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T} 2\underbrace{\boldsymbol{R}_t^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}}_{=\boldsymbol{f}_{\boldsymbol{\gamma},t}^{\top},\ \text{Lemma IA.1}}\boldsymbol{b}_{\boldsymbol{\gamma}} + T\boldsymbol{b}_{\boldsymbol{\gamma}}^{\top}\underbrace{\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}}_{=\boldsymbol{\Sigma}_{\boldsymbol{\gamma}},\ \text{Lemma IA.1}}\boldsymbol{b}_{\boldsymbol{\gamma}} + \frac{T}{g}\boldsymbol{b}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}b_{\boldsymbol{\gamma}}\right]\right\},
$$

thus, it is trivial to show that

$$
[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}, g] \sim \mathcal{N}\left(\frac{g}{1+g}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}, \frac{g}{(1+g)T}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\right). \tag{IA.6}
$$

The volatility of the SDF can be computed as

$$\begin{aligned}
\text{var}[m_{\boldsymbol{\gamma}}] &= \mathbb{E}\left[\text{var}\left[\left(\boldsymbol{f}_{\boldsymbol{\gamma}} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\right)^{\top} \boldsymbol{b}_{\boldsymbol{\gamma}} \mid \boldsymbol{b}_{\boldsymbol{\gamma}}\right]\right] + \text{var}\left[\mathbb{E}\left[1 - \left(\boldsymbol{f}_{\boldsymbol{\gamma}} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\right)^{\top} \boldsymbol{b}_{\boldsymbol{\gamma}} \mid \boldsymbol{b}_{\boldsymbol{\gamma}}\right]\right] \\
&= \mathbb{E}\left[\boldsymbol{b}_{\boldsymbol{\gamma}}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \boldsymbol{b}_{\boldsymbol{\gamma}}\right] + \text{var}\left[1 - \mathbf{0}^{\top} \boldsymbol{b}_{\boldsymbol{\gamma}}\right] \\
&= \text{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \mathbb{E}\left[\boldsymbol{b}_{\boldsymbol{\gamma}} \boldsymbol{b}_{\boldsymbol{\gamma}}^{\top}\right]\right) \\
&= \text{tr}\left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \text{var}[\boldsymbol{b}_{\boldsymbol{\gamma}}]\right) + \mathbb{E}\left[\boldsymbol{b}_{\boldsymbol{\gamma}}\right]^{\top} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \mathbb{E}[\boldsymbol{b}_{\boldsymbol{\gamma}}]
\end{aligned}$$

Under the prior distribution $[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g] \sim \mathcal{N}\left(\mathbf{0}, g/T\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\right)$,

$$\text{var}[m_{\boldsymbol{\gamma}}] = \frac{g p_{\boldsymbol{\gamma}}}{T}$$

Under the posterior distribution according to equation (IA.6),

$$\text{var}[m_{\boldsymbol{\gamma}}] = \frac{g p_{\boldsymbol{\gamma}}}{(1+g)T} + \left(\frac{g}{1+g}\right)^2 \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^{\top} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} = \left(\frac{g}{1+g}\right)^2 \text{SR}_{\boldsymbol{\gamma}}^2 + \left(\frac{g}{1+g}\right) \frac{p_{\boldsymbol{\gamma}}}{T}$$

□

IA.2.3 Proof of Proposition 3

Proof. Since we now have a prior on g as $\pi(g) = 1/(1+g)^2 I_{\{g>0\}}$, we can calculate the new marginal likelihood by integrating out g in equation (7). Noticing that SR_{max}^2, $\boldsymbol{\Sigma}$, and $\boldsymbol{S}$ in (7) are all independent of model $\mathcal{M}_{\boldsymbol{\gamma}}$, when $\boldsymbol{\gamma} \neq \mathbf{0}$:

$$\begin{aligned}
\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] &= \mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}] \exp\left(\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right) \int_0^{\infty} (1+g)^{-\frac{p_{\boldsymbol{\gamma}}+4}{2}} \exp\left\{-\frac{1}{1+g}\left[\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right]\right\} \mathrm{d}g \\
&= \mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}] \exp\left(\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right) \int_0^{1} k^{\frac{p_{\boldsymbol{\gamma}}+4}{2}-2} \exp\left\{-k\left[\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right]\right\} \mathrm{d}k \\
&= \mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}] \exp\left(\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right) \left(\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right)^{-\frac{p_{\boldsymbol{\gamma}}+2}{2}} \int_0^{\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2} t^{\frac{p_{\boldsymbol{\gamma}}}{2}} e^{-t} \mathrm{d}t \\
&= \mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}] \underbrace{\exp\left(\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right) \left(\frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right)^{-\frac{p_{\boldsymbol{\gamma}}+2}{2}} \underline{\Gamma}\left(\frac{p_{\boldsymbol{\gamma}}+2}{2}, \frac{T}{2}\text{SR}_{\boldsymbol{\gamma}}^2\right)}_{\text{BF}(\boldsymbol{\gamma}, \mathbf{0})}.
\end{aligned}$$

To prove that the Bayes factor is increasing in $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ and decreasing in $p_{\boldsymbol{\gamma}}$, we notice that

$$\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0}) = \int_0^\infty (1+g)^{-\frac{p_{\boldsymbol{\gamma}}+4}{2}} \exp\left\{\frac{g}{1+g}\left[\frac{T}{2}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right]\right\} \mathrm{d}g.$$

Taking the first-order derivatives with respect to $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ and $p_{\boldsymbol{\gamma}}$:

$$\frac{\partial \mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})}{\partial \mathrm{SR}_{\boldsymbol{\gamma}}^2} = \int_0^\infty \frac{gT}{2(1+g)}(1+g)^{-\frac{p_{\boldsymbol{\gamma}}+4}{2}} \exp\left\{\frac{g}{1+g}\left[\frac{T}{2}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right]\right\} \mathrm{d}g > 0,$$

$$\frac{\partial \mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})}{\partial p_{\boldsymbol{\gamma}}} = \int_0^\infty -\frac{\log(1+g)}{2}(1+g)^{-\frac{p_{\boldsymbol{\gamma}}+a}{2}} \exp\left\{\frac{g}{1+g}\left[\frac{T}{2}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right]\right\} \mathrm{d}g < 0.$$

□

IA.2.4 Proof of Proposition 4

Proof. Before starting the proof of this proposition, we state and prove two lemmas.

Lemma IA.2. *When the return data are generated from model $\mathcal{M}_{\boldsymbol{\gamma}_0}$, $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 = T\delta_{\boldsymbol{\gamma}} + p_{\boldsymbol{\gamma}} + O_p(\sqrt{T})$.*

Proof. According to the distribution of $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ given in (IA.4), and applying the formula for the variance of non-central $\mathcal{F}$ distributions (Johnson et al., 1995, Page 481),

$$\mathrm{var}\left[\frac{T-p_{\boldsymbol{\gamma}}}{p_{\boldsymbol{\gamma}}} \times \widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2\right] = \frac{2(T\delta_{\boldsymbol{\gamma}}+p_{\boldsymbol{\gamma}})^2 + (4T\delta_{\boldsymbol{\gamma}}+2p_{\boldsymbol{\gamma}})(T-p_{\boldsymbol{\gamma}}-2)}{(T-p_{\boldsymbol{\gamma}}-2)^2(T-p_{\boldsymbol{\gamma}}-4)}\left(\frac{T-p_{\boldsymbol{\gamma}}}{p_{\boldsymbol{\gamma}}}\right)^2.$$

Thus, $\mathrm{var}[\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2] = O(1/T)$, i.e., $\mathrm{var}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2] = O(T)$. From Proposition IA.2, $\mathbb{E}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2] = T\delta_{\boldsymbol{\gamma}} + p_{\boldsymbol{\gamma}} + O(1)$. A simple application of the chebyshev's inequality yields $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 = \mathbb{E}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2] + O_p\left(\sqrt{\mathrm{var}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2]}\right)$, which proves the stated lemma. □

Lemma IA.3. *If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, that is, factors of model $\mathcal{M}_{\boldsymbol{\gamma}}$ subsume all factors in $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ that define the true model, we have $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 - T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2 = p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0} + O_p(1)$.*

Proof. Based on the derivations in equation (IA.1),

$$T\mathrm{SR}_{\boldsymbol{\gamma}}^2 - T\delta_{\boldsymbol{\gamma}} = \boldsymbol{z}_{\boldsymbol{\gamma}}^\top \boldsymbol{z}_{\boldsymbol{\gamma}} + \left(2\sqrt{T}\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}} {\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}}^{-\frac{1}{2}}\right)\boldsymbol{z}_{\boldsymbol{\gamma}}, \qquad \boldsymbol{z}_{\boldsymbol{\gamma}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{p_{\boldsymbol{\gamma}}}).$$

Similarly,

$$T\mathrm{SR}^2_{\boldsymbol{\gamma}_0} - T\delta_{\boldsymbol{\gamma}_0} = \boldsymbol{z}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{z}_{\boldsymbol{\gamma}_0} + \left(2\sqrt{T}\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}_0}{}^{\frac{1}{2}}\right)\boldsymbol{z}_{\boldsymbol{\gamma}_0}, \qquad \boldsymbol{z}_{\boldsymbol{\gamma}_0} \sim \mathcal{N}(\mathbf{0},\, \boldsymbol{I}_{p_{\boldsymbol{\gamma}_0}}).$$

Since $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, we can write $\boldsymbol{f}_{\boldsymbol{\gamma}}$ as $(\boldsymbol{f}_{\boldsymbol{\gamma}_0}^\top,\, \hat{\boldsymbol{f}}^\top)^\top$, then $\boldsymbol{z}_{\boldsymbol{\gamma}}$ can be expressed as $(\boldsymbol{z}_{\boldsymbol{\gamma}_0}^\top,\, \hat{\boldsymbol{z}}^\top)^\top$ where $\hat{\boldsymbol{z}} \sim \mathcal{N}(\mathbf{0},\, \boldsymbol{I}_{(p_{\boldsymbol{\gamma}}-p_{\boldsymbol{\gamma}_0})})$. According to Lemma IA.1, if $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$ (in parallel to $\boldsymbol{f} \subseteq \boldsymbol{R}$),

$$\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} = \begin{pmatrix} \boldsymbol{I}_{p_{\boldsymbol{\gamma}_0}} \\ \mathbf{0}_{(p_{\boldsymbol{\gamma}}-p_{\boldsymbol{\gamma}_0})} \end{pmatrix}.$$

As a result,

$$\left(2\sqrt{T}\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}{}^{-\frac{1}{2}}\right)\boldsymbol{z}_{\boldsymbol{\gamma}} = 2\sqrt{T}\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top (\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}{}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0})^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}{}^{\frac{1}{2}}\boldsymbol{z}_{\boldsymbol{\gamma}} = \left(2\sqrt{T}\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}_0}{}^{\frac{1}{2}}\right)\boldsymbol{z}_{\boldsymbol{\gamma}_0}.$$

According to equation (IA.3), $\delta_{\boldsymbol{\gamma}_0} - \delta_{\boldsymbol{\gamma}} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\boldsymbol{b}_{\boldsymbol{\gamma}_0}$. When $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, we have $\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right] = 0$. Thus, $\delta_{\boldsymbol{\gamma}} = \delta_{\boldsymbol{\gamma}_0}$. Combining the results above, we have

$$T\mathrm{SR}^2_{\boldsymbol{\gamma}} - T\mathrm{SR}^2_{\boldsymbol{\gamma}_0} = \boldsymbol{z}_{\boldsymbol{\gamma}}^\top \boldsymbol{z}_{\boldsymbol{\gamma}} - \boldsymbol{z}_{\boldsymbol{\gamma}_0}^\top \boldsymbol{z}_{\boldsymbol{\gamma}_0} \sim \chi^2(p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0}) = p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0} + O_p(1).$$

Finally, turning back to the sample version squared Sharpe ratios, from part 2 of Proposition IA.2, $(T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}) = [1 + O_p(1/\sqrt{T})](T\mathrm{SR}^2_{\boldsymbol{\gamma}})$ for all $\boldsymbol{\gamma}$, thus

$$T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} - T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}_0} = [1 + O_p(1/\sqrt{T})][p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0} + O_p(1)] = p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0} + O_p(1).$$

□

Now we prove Proposition 4. We start with properties of Bayes factors in light of the two lemmas. Under the g-prior, the Bayes factor comparing model $\mathcal{M}_{\boldsymbol{\gamma}}$ against the true model $\mathcal{M}_{\boldsymbol{\gamma}_0}$ is computed as

$$\mathrm{BF}(\boldsymbol{\gamma},\, \boldsymbol{\gamma}_0) = \frac{\mathrm{BF}(\boldsymbol{\gamma},\, \mathbf{0})}{\mathrm{BF}(\boldsymbol{\gamma}_0,\, \mathbf{0})} = \exp\left\{\frac{g}{2(1+g)}\left(T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} - T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}_0}\right) - \frac{\log(1+g)}{2}(p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0})\right\}.$$

Case I. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, that is, there are factors from the true model that are not considered under model $\mathcal{M}_{\boldsymbol{\gamma}}$. From lemma IA.2,

$$T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} - T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}_0} = T(\delta_{\boldsymbol{\gamma}} - \delta_{\boldsymbol{\gamma}_0}) + p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0} + O_p(\sqrt{T}).$$

According to equation IA.3, $\delta_{\boldsymbol{\gamma}} - \delta_{\boldsymbol{\gamma}_0} = -\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right] \boldsymbol{b}_{\boldsymbol{\gamma}_0} < 0$ if $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}_0 \neq \mathbf{0}$ (here we also implicitly assume that every pair of factors is not perfectly correlated). Then $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 - T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2 = -(\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right] \boldsymbol{b}_{\boldsymbol{\gamma}_0})T + O(1) + o_p(T) \xrightarrow{p} -\infty$ as $T \to \infty$. As a result, $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) \xrightarrow{p} 0$.

Case II. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}$, that is, all factors from the true model are considered under model $\mathcal{M}_{\boldsymbol{\gamma}}$. Lemma IA.3 implies that

$$\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) = \exp\left\{\frac{g}{2(1+g)} O_p(1) - \frac{f(g)}{2}(p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0})\right\},$$

where $f(g) = \log(1+g) - g/(1+g)$. For any $g < \infty$, $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) > 0$ with probability one.[6]

We are now ready to prove the main results of the proposition.

Factor selection consistency. For the jth factor such that $\gamma_{0,j} = 1$,

$$\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] = 1 - \mathbb{P}[\gamma_j = 0 \mid \mathcal{D}] = 1 - \sum_{\boldsymbol{\gamma}:\, \gamma_j = 0} \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] = 1 - \sum_{\boldsymbol{\gamma}:\, \gamma_j = 0} \frac{\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0)}{\sum_{\boldsymbol{\gamma}'} \mathrm{BF}(\boldsymbol{\gamma}', \boldsymbol{\gamma}_0)} \xrightarrow{p} 1$$

The last step is because, for any $\boldsymbol{\gamma}$ such that $\gamma_j = 0$, $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$ must hold. As a result, $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) \xrightarrow{p} 0$ for all such $\boldsymbol{\gamma}$s.

Model selection inconsistency. The posterior probability of the true model $\mathcal{M}_{\boldsymbol{\gamma}_0}$ satisfies

$$\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}_0} \mid \mathcal{D}] = 1 - \sum_{\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0} \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] = 1 - \sum_{\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0} \frac{\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0)}{\sum_{\boldsymbol{\gamma}'} \mathrm{BF}(\boldsymbol{\gamma}', \boldsymbol{\gamma}_0)} < 1 - \sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}} \frac{\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0)}{\sum_{\boldsymbol{\gamma}'} \mathrm{BF}(\boldsymbol{\gamma}', \boldsymbol{\gamma}_0)} < 1,$$

with probability one. The last step is because $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) > 0$ almost surely for every $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}$. □

[6] Noticing that if $g \to \infty$ as $T \to \infty$, the Bayes factor can converge to zero in probability at the rate $\log(g)$, which implies that $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}_0} \mid \mathcal{D}] \xrightarrow{p} 1$ (as our later proof will illustrate). However, if $g \to \infty$, the Bayes factor comparing any model against the null model $\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})$ also approaches zero: Bayes factors will strongly favor the most parsimonious model. This phenomenon is closely related to the Bartlett's paradox in Bayesian inference.

IA.2.5 Proof of Proposition 5

Proof. Under the mixture of g-priors specification, the Bayes factor comparing model model $\mathcal{M}_{\boldsymbol{\gamma}}$ against the true model $\mathcal{M}_{\boldsymbol{\gamma}_0}$ is calculated as

$$\mathrm{BF}(\boldsymbol{\gamma},\,\boldsymbol{\gamma}_0) = \underbrace{\exp\left(\frac{T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 - T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2}{2}\right)}_{I_1} \underbrace{\left(\frac{\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2}{\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2}\right)^{-\frac{p_{\boldsymbol{\gamma}}+2}{2}}}_{I_2} \underbrace{\left(\frac{T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2}{2}\right)^{\frac{p_{\boldsymbol{\gamma}_0}-p_{\boldsymbol{\gamma}}}{2}}}_{I_3} \underbrace{\frac{\underline{\Gamma}\left(\frac{p_{\boldsymbol{\gamma}}+2}{2},\, T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2/2\right)}{\underline{\Gamma}\left(\frac{p_{\boldsymbol{\gamma}_0}+2}{2},\, T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2/2\right)}}_{I_4}. \tag{IA.7}$$

If follows trivially that $T \to \infty$,

$$\frac{\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2}{\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2} \xrightarrow{p} \frac{\delta_{\boldsymbol{\gamma}}}{\delta_{\boldsymbol{\gamma}_0}},$$

thus $I_2 = O_p(1)$.

For any $s > 0$, $x > 0$, $\underline{\Gamma}(s,x) = \Gamma(s)\mathbb{P}[\nu \le x]$ where $\Gamma(\cdot)$ is the standard Gamma function, and $\nu \sim \mathrm{Gamma}(s,1)$, a Gamma distribution with shape and scale parameters being s and 1 respectively. Thus, as $T \to \infty$, $I_4 = O_p(1)$ because

$$I_4 \xrightarrow{p} \begin{cases} (p_{\boldsymbol{\gamma}}/2)(p_{\boldsymbol{\gamma}}/2-1)(p_{\boldsymbol{\gamma}}/2-2)\cdots(p_{\boldsymbol{\gamma}_0}/2+2)(p_{\boldsymbol{\gamma}_0}/2+1), & \text{if } p_{\boldsymbol{\gamma}} \ge p_{\boldsymbol{\gamma}_0}+1 \\ \left[(p_{\boldsymbol{\gamma}_0}/2)(p_{\boldsymbol{\gamma}_0}/2-1)(p_{\boldsymbol{\gamma}_0}/2-2)\cdots(p_{\boldsymbol{\gamma}}/2+2)(p_{\boldsymbol{\gamma}}/2+1)\right]^{-1}, & \text{if } p_{\boldsymbol{\gamma}} \le p_{\boldsymbol{\gamma}_0}-1 \\ 1, & \text{if } p_{\boldsymbol{\gamma}} = p_{\boldsymbol{\gamma}_0} \end{cases}$$

For the two remaining items I_1 and I_3 in the Bayes factor, we discuss their behavior under two cases.

Case I. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, according to our proof for Proposition 4,

$$T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2 = \delta_{\boldsymbol{\gamma}_0} T + o_p(T), \qquad T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 - T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2 = -(\delta_{\boldsymbol{\gamma}_0} - \delta_{\boldsymbol{\gamma}}) \times T + o_p(T), \qquad \delta_{\boldsymbol{\gamma}_0} > \delta_{\boldsymbol{\gamma}} > 0.$$

As a result, it always holds that $I_1 \xrightarrow{p} 0$ under this scenario.

The behavior of the third item I_3 is discussed as follows. On the one hand, if $p_{\boldsymbol{\gamma}} > p_{\boldsymbol{\gamma}_0}$, $I_3 \xrightarrow{p} 0$ and obviously $\mathrm{BF}(\boldsymbol{\gamma},\,\boldsymbol{\gamma}_0) \xrightarrow{p} 0$.

If $p_{\boldsymbol{\gamma}} \le p_{\boldsymbol{\gamma}_0}$, $I_3 = O_p(T^{(p_{\boldsymbol{\gamma}_0}-p_{\boldsymbol{\gamma}})/2})$. As $I_1 T^{(p_{\boldsymbol{\gamma}_0}-p_{\boldsymbol{\gamma}})/2} = o_p(1)$, we have $I_1 I_3 = o_p(1)$ and $\mathrm{BF}(\boldsymbol{\gamma},\,\boldsymbol{\gamma}_0) \xrightarrow{p} 0$. That is, the Bayes factor still converges to zero in probability.

Case II. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}$, Lemma IA.3 implies that $I_1 = O_p(1)$. As it is always true that $p_{\boldsymbol{\gamma}} > p_{\boldsymbol{\gamma}_0}$ when $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}$, $I_3 \xrightarrow{p} 0$. As a result, $\mathrm{BF}(\boldsymbol{\gamma},\,\boldsymbol{\gamma}_0) \xrightarrow{p} 0$.

Summing up, $\mathrm{BF}(\boldsymbol{\gamma},\,\boldsymbol{\gamma}_0) \xrightarrow{p} 0$ for any $\mathcal{M}_{\boldsymbol{\gamma}}$ as long as $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$. As a result,

$$\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}_0} \mid \mathcal{D}] = 1 - \sum_{\boldsymbol{\gamma}\neq\boldsymbol{\gamma}_0} \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] = 1 - \sum_{\boldsymbol{\gamma}\neq\boldsymbol{\gamma}_0} \frac{\mathrm{BF}(\boldsymbol{\gamma},\,\boldsymbol{\gamma}_0)}{\sum_{\boldsymbol{\gamma}'} \mathrm{BF}(\boldsymbol{\gamma}',\,\boldsymbol{\gamma}_0)} \xrightarrow{p} 1.$$

□

IA.2.6 Proof of Proposition 6

Proof. Now the true model is defined by factors in $\boldsymbol{f}_0$ and $\boldsymbol{f}_{\boldsymbol{\gamma}_0} = \boldsymbol{f}_0 \cap \boldsymbol{f}$. We define $\boldsymbol{C}_{\boldsymbol{\gamma}} = \mathrm{cov}\left[\boldsymbol{f}_{\boldsymbol{\gamma}},\,\boldsymbol{f}_0\right]$, $\boldsymbol{\Sigma}_0 = \mathrm{var}[\boldsymbol{f}_0]$. Applying Lemma IA.2, we have that for any model $\mathcal{M}_{\boldsymbol{\gamma}}$,

$$T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 = T\delta_{\boldsymbol{\gamma}}^{(0)} + p_{\boldsymbol{\gamma}} + O_p(\sqrt{T})$$

where $\delta_{\boldsymbol{\gamma}}^{(0)} = \boldsymbol{b}_0^\top \boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{b}_0$.

We now state and prove the following lemma:

Lemma IA.4. *Under the assumptions of Proposition 6, for any $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$ as $T \to \infty$.*

Proof. For the ease of exposition, we first define $\widetilde{\boldsymbol{f}}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\gamma}_0} \cap \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{f}}_{\boldsymbol{\gamma}_0} = \boldsymbol{f}_{\boldsymbol{\gamma}_0} \backslash \widetilde{\boldsymbol{f}}_{\boldsymbol{\gamma}}$. Now consider the model, namely $\mathcal{M}_{\boldsymbol{\gamma}'}$, defined by factors $\boldsymbol{f}_{\boldsymbol{\gamma}'}^\top = (\boldsymbol{f}_{\boldsymbol{\gamma}}^\top,\,\widehat{\boldsymbol{f}}_{\boldsymbol{\gamma}_0}^\top)$. Noticing that

$$\begin{aligned}
\delta_{\boldsymbol{\gamma}'}^{(0)} - \delta_{\boldsymbol{\gamma}}^{(0)} &= \boldsymbol{b}_0^\top \boldsymbol{C}_{\boldsymbol{\gamma}'}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}'} \boldsymbol{b}_0 - \boldsymbol{b}_0^\top \boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{b}_0 \\
&= \boldsymbol{b}_0^\top \boldsymbol{C}_{\boldsymbol{\gamma}'}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}'} \boldsymbol{b}_0 - \boldsymbol{b}_0^\top \boldsymbol{\Sigma}_0 \boldsymbol{b}_0 + \boldsymbol{b}_0^\top \boldsymbol{\Sigma}_0 \boldsymbol{b}_0 - \boldsymbol{b}_0^\top \boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_0^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{b}_0 \\
&= -\mathrm{var}[\boldsymbol{b}_0^\top \boldsymbol{f}_0 \mid \boldsymbol{f}_{\boldsymbol{\gamma}'}] + \mathrm{var}[\boldsymbol{b}_0^\top \boldsymbol{f}_0 \mid \boldsymbol{f}_{\boldsymbol{\gamma}}] > 0,
\end{aligned}$$

The last inequality is due to the fact that $\widehat{\boldsymbol{f}}_{\boldsymbol{\gamma}_0} \neq \emptyset$ when $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, which implies $\boldsymbol{f}_{\boldsymbol{\gamma}} \subset \boldsymbol{f}_{\boldsymbol{\gamma}'}$.

The Bayes factor $\mathrm{BF}(\boldsymbol{\gamma},\,\boldsymbol{\gamma}')$, according to equation (IA.7) (replacing $\boldsymbol{\gamma}_0$ with $\boldsymbol{\gamma}'$), must satisfy:

1. $I_1 = \exp\left(-(\delta_{\boldsymbol{\gamma}'}^{(0)} - \delta_{\boldsymbol{\gamma}}^{(0)})T + o_p(T)\right)$;
2. $I_2 \xrightarrow{p} \delta_{\boldsymbol{\gamma}}^{(0)}/\delta_{\boldsymbol{\gamma}'}^{(0)}$, that is, $I_2 = O_p(1)$;
3. $I_3 = O_p\left(T^{\frac{p_{\boldsymbol{\gamma}'} - p_{\boldsymbol{\gamma}}}{2}}\right)$;
4. $I_4 \xrightarrow{p} \left[(p_{\boldsymbol{\gamma}'}/2)(p_{\boldsymbol{\gamma}'}/2-1)(p_{\boldsymbol{\gamma}'}/2-2)\cdots(p_{\boldsymbol{\gamma}'}/2+2)(p_{\boldsymbol{\gamma}'}/2+1)\right]^{-1}$, that is, $I_4 = O_p(1)$.

As a result, $I_1 = o_p(T^{-\alpha})$ for any finite positive number α, and we have $I_1 I_2 = o_p(1)$. Then $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}') \xrightarrow{p} 0$ as $T \to \infty$. Now, for any $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$,

$$\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] = \frac{\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')}{\sum_{\widetilde{\boldsymbol{\gamma}}} \mathrm{BF}(\widetilde{\boldsymbol{\gamma}}, \boldsymbol{\gamma}')} \xrightarrow{p} 0.$$

□

Going back to Proposition 6, for the marginal probability of selecting the jth factor when $\gamma_{0,j} = 1$, we have

$$\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] = 1 - \mathbb{P}[\gamma_j = 0 \mid \mathcal{D}] = 1 - \sum_{\boldsymbol{\gamma}:\, \gamma_j = 0} \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 1.$$

The last step for convergence is because, for any $\boldsymbol{\gamma}$ such that $\gamma_j = 0$, it must be that $\boldsymbol{f}_{\boldsymbol{\gamma}} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, which implies $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$. □

IA.2.7 Proof of Proposition 7

Proof. Based on Lemma IA.4, for any $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$, which is equivalent to

$$\lim_{T \to \infty} \mathrm{Prob}_0 \left\{ \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] < \varepsilon \right\} = 1, \qquad \forall 0 < \varepsilon < 1, \tag{IA.8}$$

where Prob_0 is a measure defined based on the sample distribution of returns under the true SDF $m_0 = 1 - (\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}_0$, while $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$ is the posterior calculated using the return sample (a random variable).

According to the definition of $\mathcal{E}$, we have

$$\begin{aligned}
\mathcal{E} &= -\frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}} \log \left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] \right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] \\
&= -\frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log \left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] \right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] - \frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log \left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] \right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] \\
&\leq -\frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log \left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] \right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] - \frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log \left(\frac{1}{2^{p - p_{\boldsymbol{\gamma}_0}}} \right) \left(\frac{1}{2^{p - p_{\boldsymbol{\gamma}_0}}} \right) \\
&= -\frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log \left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] \right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D} \right] + \frac{p - p_{\boldsymbol{\gamma}_0}}{p}.
\end{aligned}$$

Thus,

$$\begin{aligned}\text{Prob}_0\left[\mathcal{E} \geq \frac{p-p_{\boldsymbol{\gamma}}}{p}\right] &\leq \text{Prob}_0\left[\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] \leq 0\right] \\ &= 1 - \text{Prob}_0\left[\lim_{T\to\infty}\left\{\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] > -\frac{1}{T}\right\}\right] \\ &\leq 1 - \text{Prob}_0\left[\lim_{T\to\infty}\left\{\min_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] > -\frac{1}{(2^p - 2^{p_{\boldsymbol{\gamma}}})T}\right\}\right]\end{aligned}$$

Consider function $\phi(x) = x \log x$. For any $0 < \varepsilon < e^{-1}$ and $0 < x < \varepsilon$, $\phi(x) < 0$ and is continuously decreasing. As a result, for any $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$,

$$\text{Prob}_0\left[\lim_{T\to\infty}\left\{\log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right) \mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] > -\frac{1}{(2^p - 2^{p_{\boldsymbol{\gamma}}})T}\right\}\right] = 1,$$

according to IA.8. Then, as $T \to \infty$, we must have

$$\text{Prob}_0\left[\mathcal{E} \geq \frac{p-p_{\boldsymbol{\gamma}}}{p}\right] = 0,$$

that is, $\mathcal{E} \leq (p - p_{\gamma_0})/p$ with probability one. □

IA.3 Bayesian Inference about the Risk Premium $\lambda_{\mathcal{E}}$

We assume that $\boldsymbol{f}_t \in \mathbb{R}^p$ follows a multivariate normal distribution:

$$\boldsymbol{f}_t \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{f}}, \boldsymbol{\Sigma}_{\boldsymbol{f}}).$$

By equation (14),

$$\mathcal{E}_t^{ar1} \mid \boldsymbol{f}_t \overset{\text{iid}}{\sim} \mathcal{N}(\eta_0 + \boldsymbol{\eta}^\top \boldsymbol{f}_t, \sigma_e^2),$$

where σ_e^2 is the variance of e_t. The full conditional distribution is

$$\prod_{t=1}^{T} \mathbb{P}\left[\boldsymbol{f}_t, \mathcal{E}_t^{ar1} \mid \boldsymbol{\mu}_{\boldsymbol{f}}, \boldsymbol{\Sigma}_{\boldsymbol{f}}, \eta_0, \boldsymbol{\eta}, \sigma_e^2\right] = \prod_{t=1}^{T} \mathbb{P}\left[\boldsymbol{f}_t \mid \boldsymbol{\mu}_{\boldsymbol{f}}, \boldsymbol{\Sigma}_{\boldsymbol{f}}\right] \times \mathbb{P}\left[\mathcal{E}_t^{ar1} \mid \eta_0, \boldsymbol{\eta}, \sigma_e^2\right];$$

hence, as long as the priors for $(\boldsymbol{\mu_f}, \boldsymbol{\Sigma_f})$ and $(\eta_0, \boldsymbol{\eta}, \sigma_e^2)$ are independent, we can make inference about these two sets of variables separately. For simplicity, we assign the Jefferys priors: $\pi(\boldsymbol{\mu_f}, \boldsymbol{\Sigma_f}) \propto |\boldsymbol{\Sigma_f}|^{-\frac{p+1}{2}}$ and $\pi(\eta_0, \boldsymbol{\eta}, \sigma_e^2) \propto \sigma_e^{-2}$.

According to these prior specifications, the posterior distribution of $(\boldsymbol{\mu_f}, \boldsymbol{\Sigma_f})$ is normal-inverse-Wishart:

$$\boldsymbol{\Sigma_f} \mid \{\boldsymbol{f}_t\}_{t=1}^T \sim \mathcal{W}^{-1}\left(T-1,\ \sum_{t=1}^{T}(\boldsymbol{f}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{f}})(\boldsymbol{f}_t - \hat{\boldsymbol{\mu}}_{\boldsymbol{f}})^\top\right), \quad \boldsymbol{\mu_f} \mid \boldsymbol{\Sigma_f}, \{\boldsymbol{f}_t\}_{t=1}^T \sim \mathcal{N}\left(\hat{\boldsymbol{\mu}}_{\boldsymbol{f}},\ \frac{\boldsymbol{\Sigma_f}}{T}\right),$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{f}} = 1/T \sum_{t=1}^T \boldsymbol{f}_t$ and $\mathcal{W}^{-1}(\cdot,\cdot)$ denotes the inverse-Wishart distribution.

The posterior distributions of $(\eta_0, \boldsymbol{\eta}, \sigma_e^2)$ are normal-inverse-gamma:

$$\sigma_e^2 \mid \{\boldsymbol{f}_t,\, \mathcal{E}_t^{ar1}\}_{t=1}^T \sim \mathcal{IG}\left(\frac{T-p-1}{2},\ \frac{\sum_{t=1}^T(\mathcal{E}_t^{ar1} - \hat{\eta}_0 - \hat{\boldsymbol{\eta}}^\top \boldsymbol{f}_t)^2}{2}\right),$$

$$(\eta_0, \boldsymbol{\eta}) \mid \sigma_e^2,\, \{\boldsymbol{f}_t,\, \mathcal{E}_t^{ar1}\}_{t=1}^T \sim \mathcal{N}(\boldsymbol{\xi},\, \sigma_e^2(\boldsymbol{F}^\top \boldsymbol{F})^{-1}),$$

where $\boldsymbol{\xi} = (\hat{\eta}_0,\, \hat{\boldsymbol{\eta}})$ is the OLS estimates of $(\eta_0, \boldsymbol{\eta})$; $\boldsymbol{F}$ is a $T \times (p+1)$ matrix of $\mathbf{1}_T$ and the factor realizations $\{\boldsymbol{f}_t\}_{t=1}^T$. Based on these analytical conditional distributions, we use Gibbs sampling to draw samples from the posterior distribution $\left[\boldsymbol{\mu}_f,\, \boldsymbol{\Sigma_f}, \eta_0, \boldsymbol{\eta}, \sigma_e^2 \mid \boldsymbol{f}_t, \mathcal{E}_t^{ar1}\right]$. For each posterior draw of $(\boldsymbol{\eta}, \boldsymbol{\mu_f})$, we calculate the risk premia $\lambda_{\mathcal{E}} = \boldsymbol{\eta}^\top \boldsymbol{\mu_f}$. Credible intervals and the Bayesian "t-statistics" are computed using the posterior samples of $\lambda_{\mathcal{E}}$.

IA.4 Additional Figures

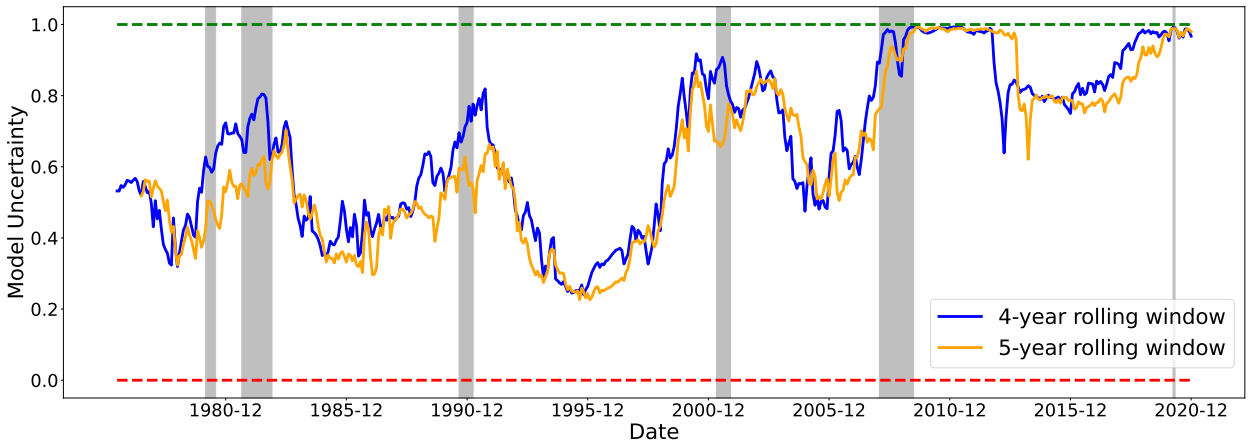


Figure IA.1: Alternative Rolling Windows

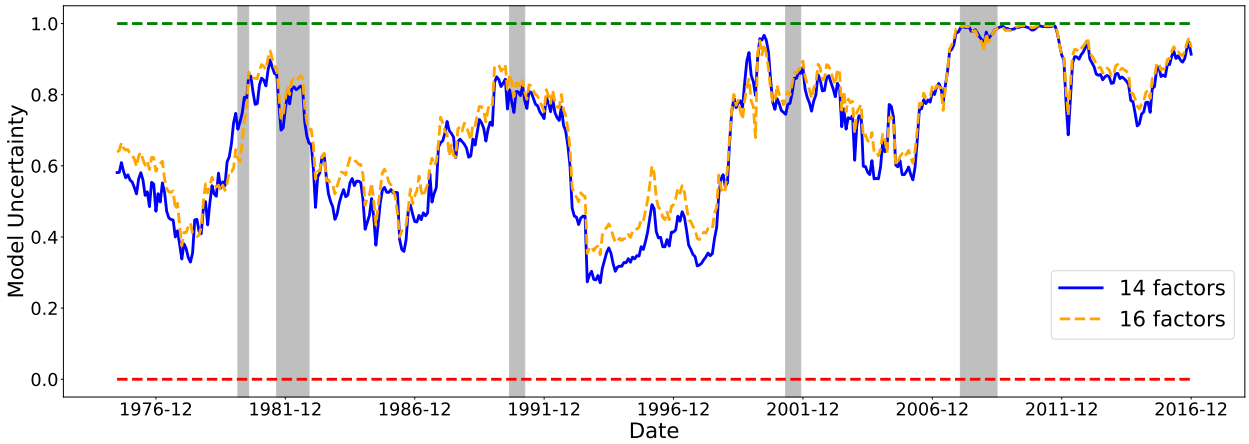
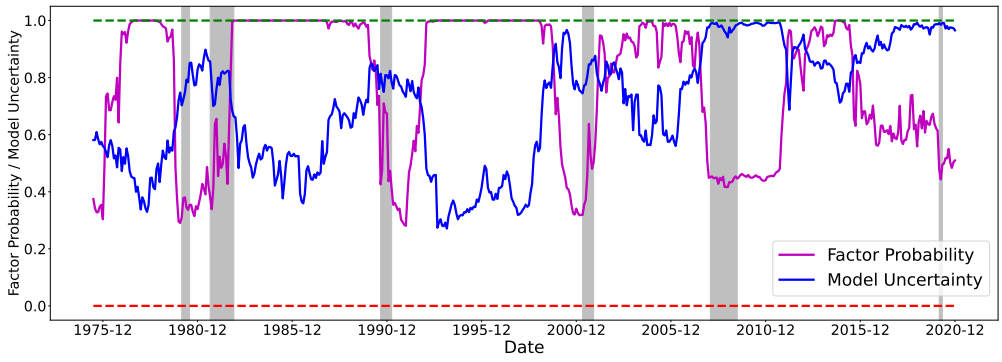
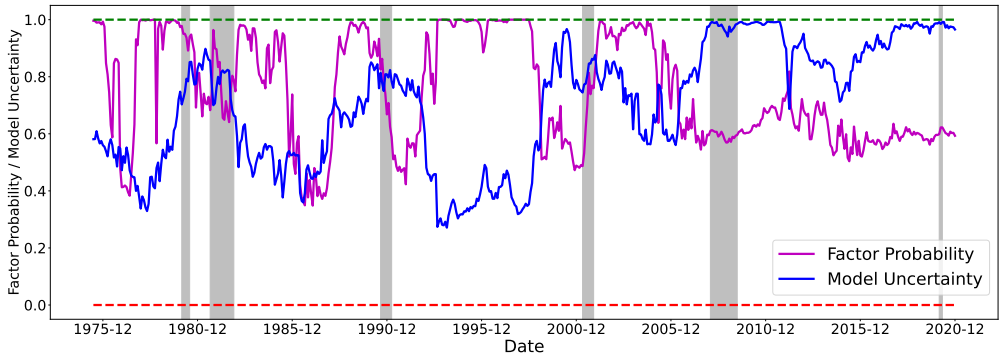


Figure IA.2: Time-Series of Model Uncertainty: Including two mispricing factors

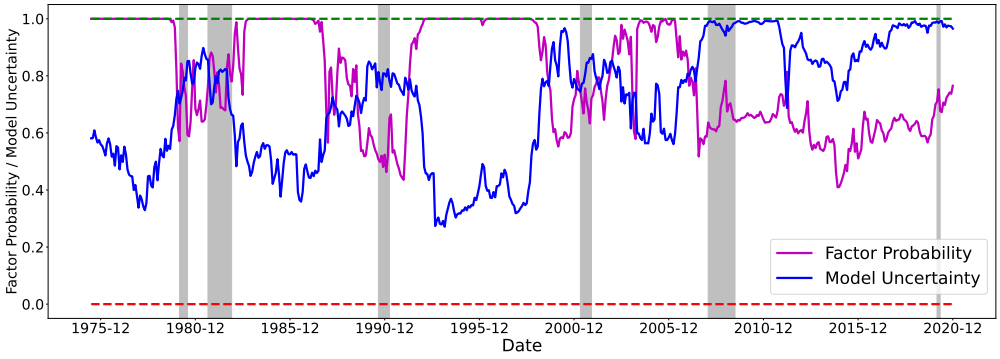
The figure plots the time series of model uncertainty about the linear SDF. Different from Figure 1, we further include two mispricing factors in Stambaugh and Yuan (2017), hence leading to 16 factors.



(a) MKT

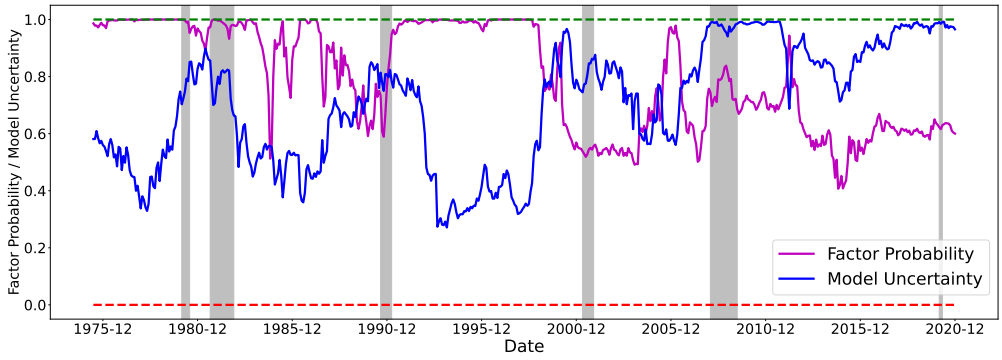


(b) Size (SMB or ME)

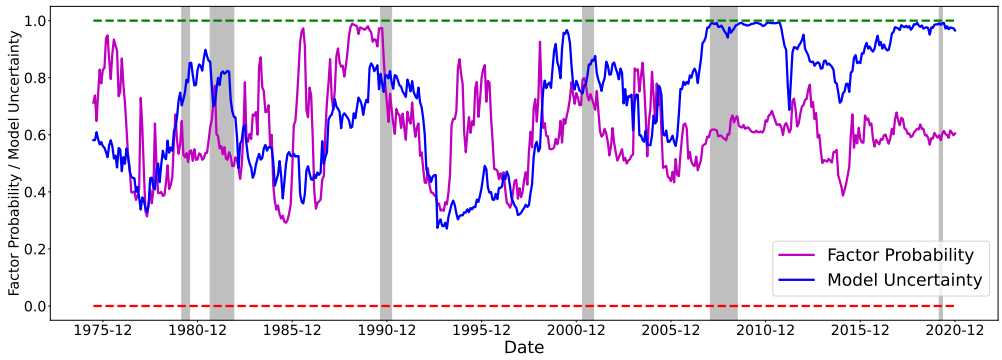


(c) Value (HML or HML devil)

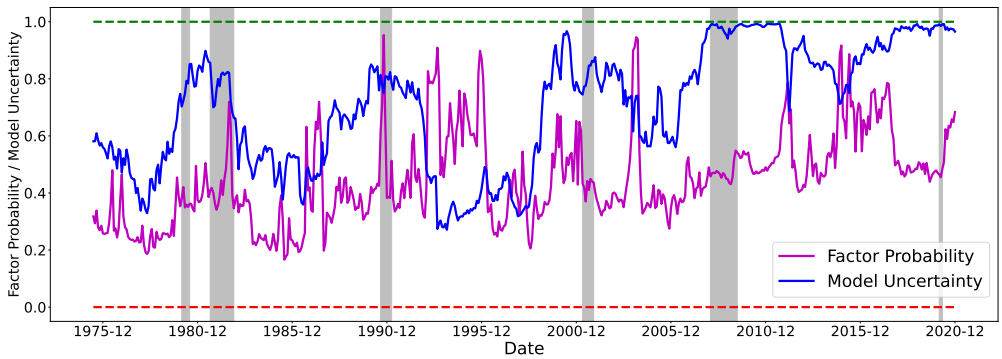
Figure IA.3: Time Series of Posterior Factor Probabilities



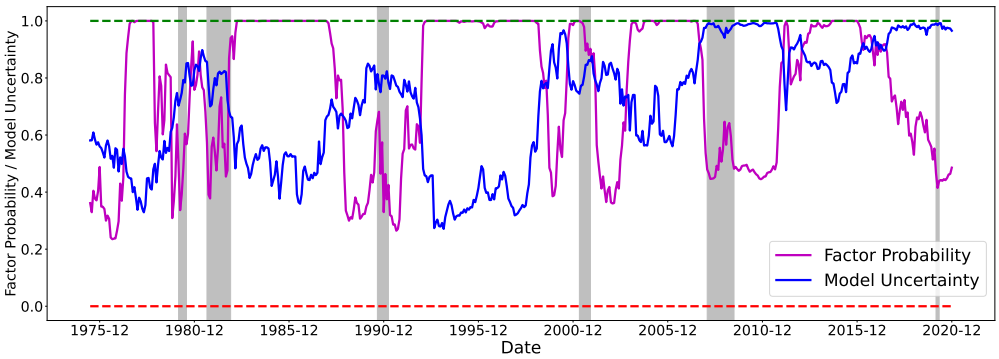
(d) Profitability (ROE or RMW)



(e) Investment (IA or CMA)

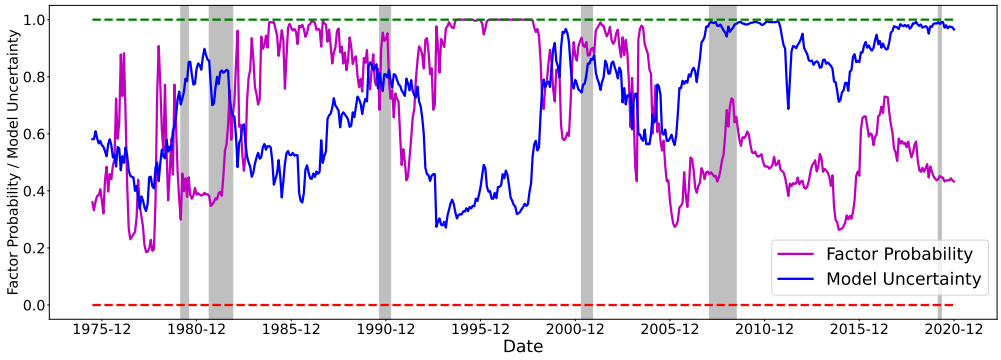


(f) Momentum

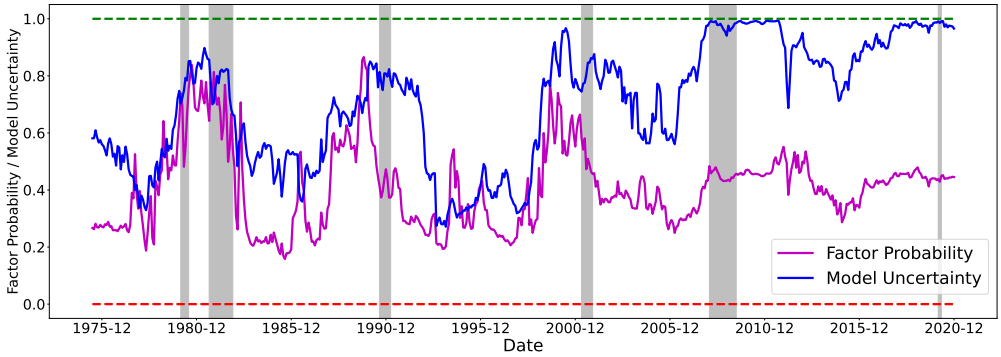


(g) BAB

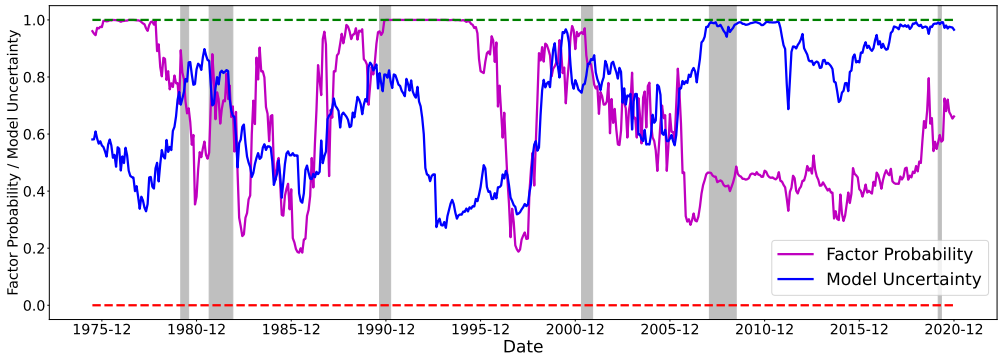
Figure IA.3: Time Series of Posterior Factor Probabilities (Continued)



(h) QMJ



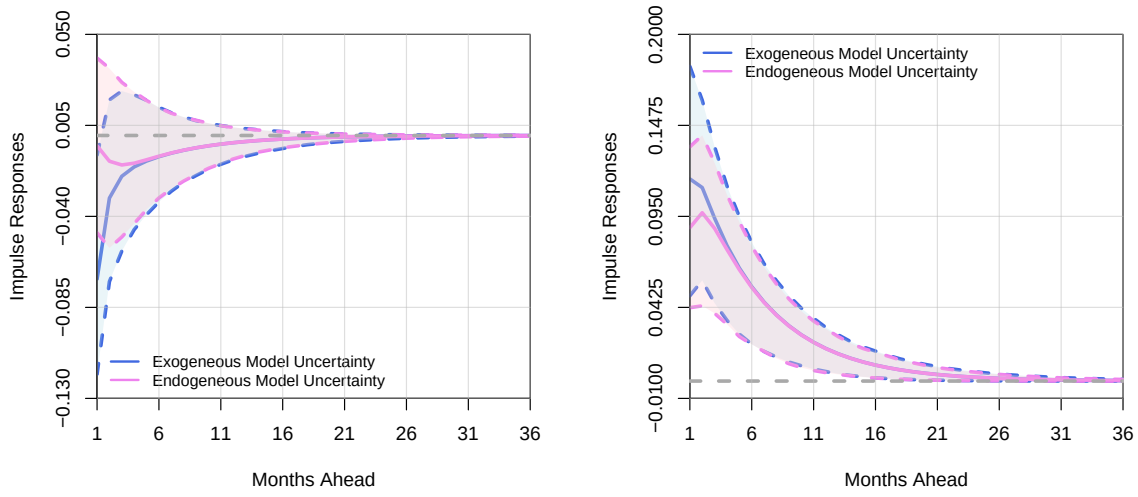
(i) FIN



(j) PEAD

Figure IA.3: Time Series of Posterior Factor Probabilities (Continued)

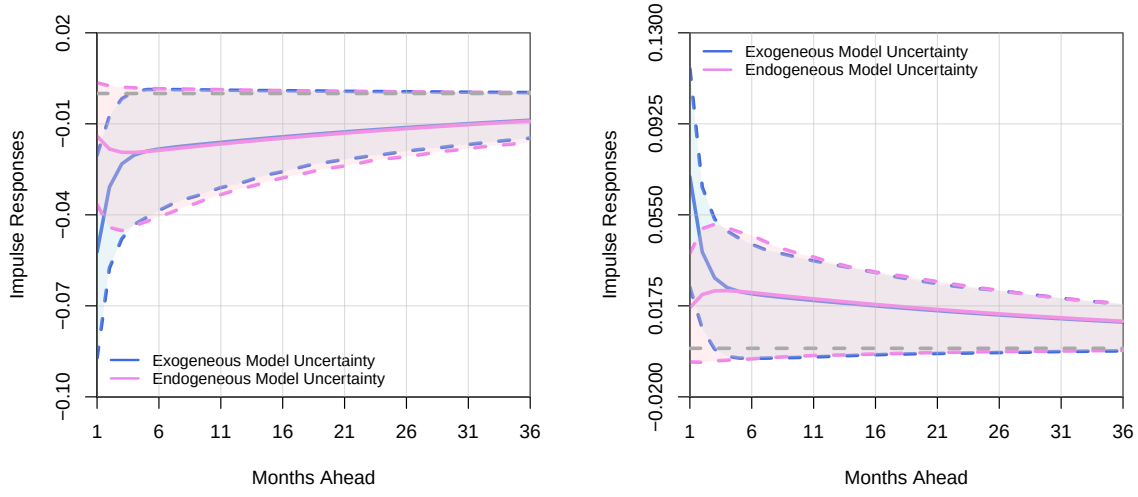
The figures plot the time series of posterior marginal probabilities of 14 factors. At the end of each month, we estimate models using the daily factor returns in the past three years. The sample ranges from July 1972 to December 2020. Since we use a three-year rolling window, the time series of factor probabilities starts from June 1975. Shaded areas are NBER-based recession periods for the US.



(a) Equity Fund Flows **(b)** Fixed-Income Fund Flows

Figure IA.4: Impulse Responses of Equity and Fixed-Income Fund Flows to VIX Shocks

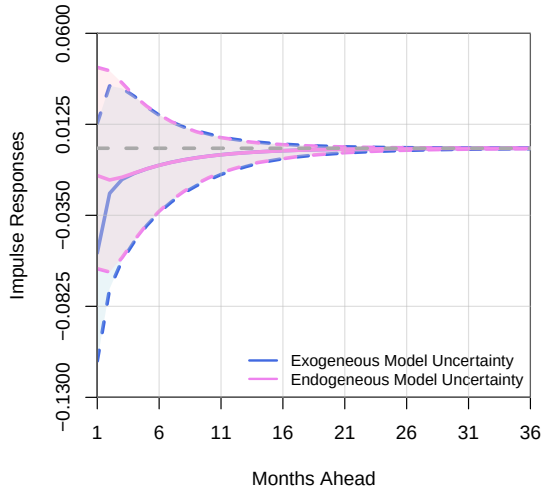
This figure shows the dynamic impulse response functions (IRFs) of equity and fixed-income fund flows to VIX shocks in VAR(1).



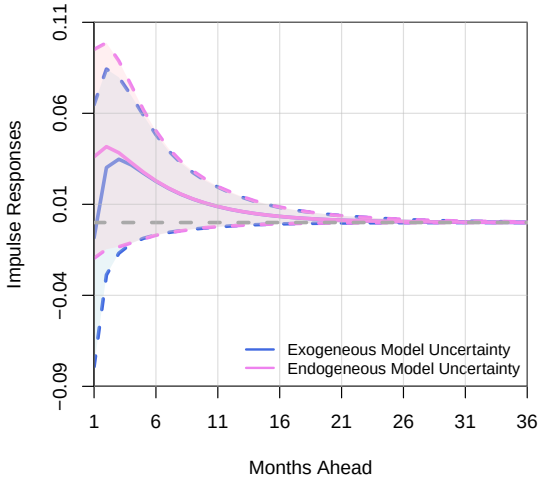
(a) Equity Fund Flows **(b)** Fixed-Income Fund Flows

Figure IA.5: Impulse Responses of Equity and Fixed-Income Mutual Fund Flows to Financial Uncertainty Shocks

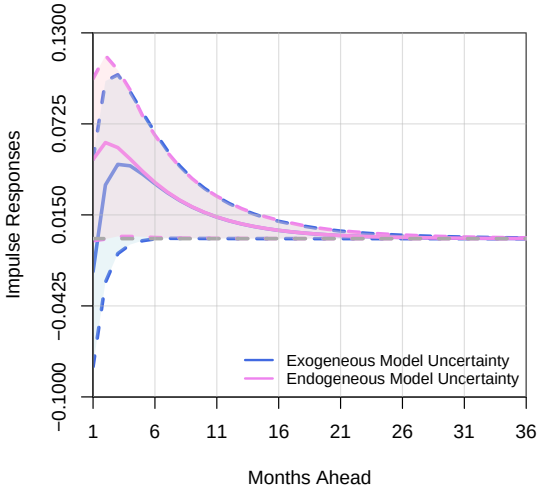
This figure shows the dynamic impulse response functions (IRFs) of equity and fixed-income fund flows to financial uncertainty shocks in VAR(1).



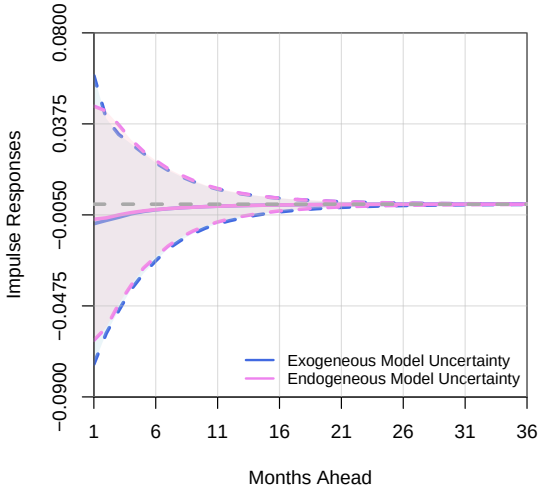
(a) Style Fund Flows



(b) Sector Fund Flows



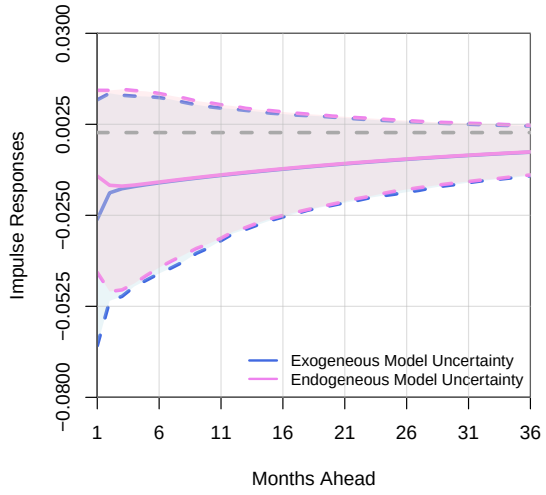
(c) Small-Cap Fund Flows



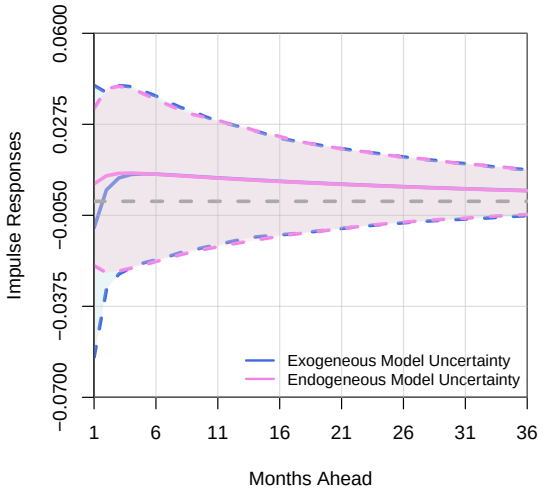
(d) Large-Cap Fund Flows

Figure IA.6: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes to VIX Shocks

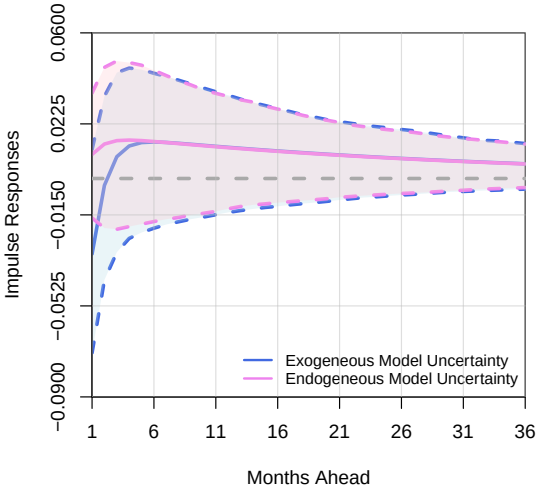
This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to VIX shocks in VAR(1). Other details can be found in the footnote of Figure 6.



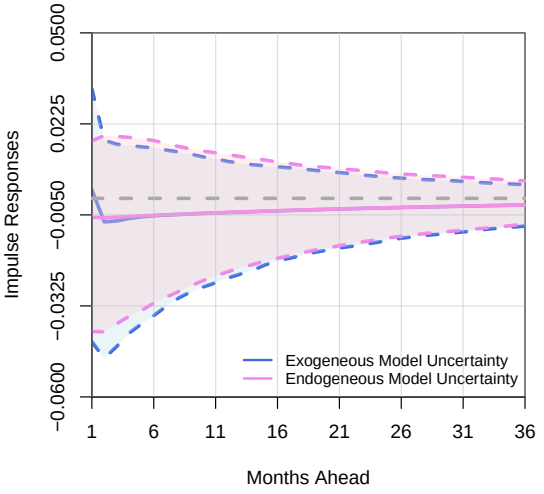
(a) Style Fund Flows



(b) Sector Fund Flows



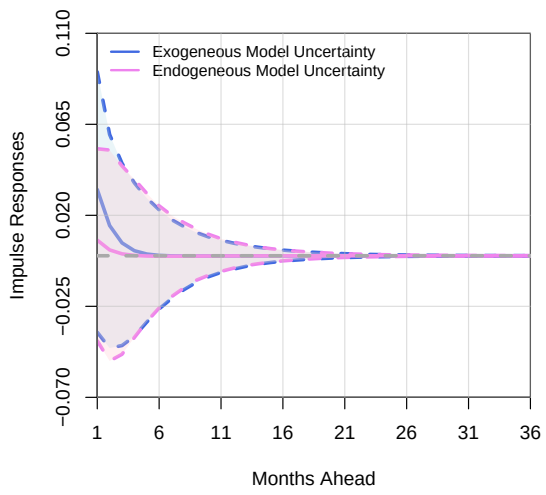
(c) Small-Cap Fund Flows



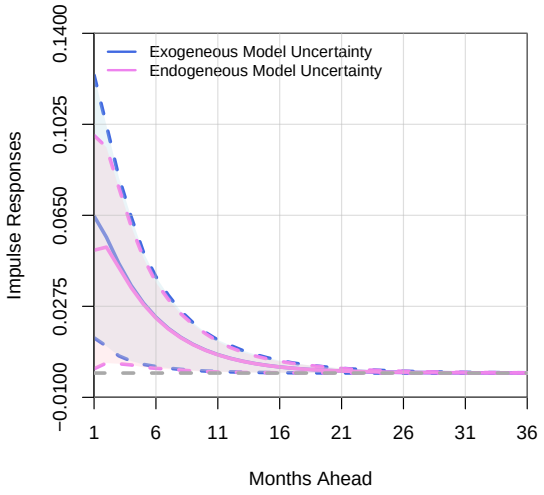
(d) Large-Cap Fund Flows

Figure IA.7: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes to Financial Uncertainty Shocks

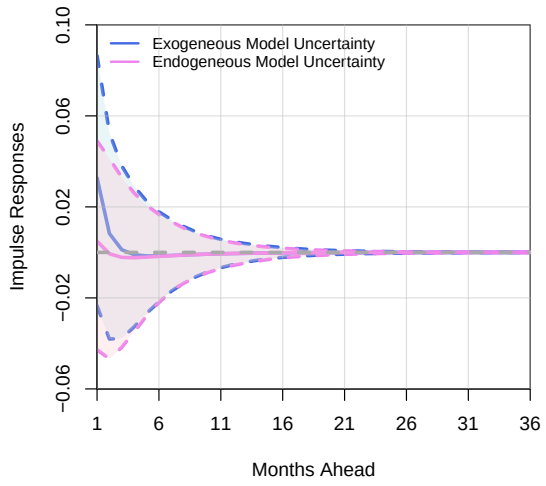
This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to financial uncertainty shocks in VAR(1). Other details can be found in the footnote of Figure 6.



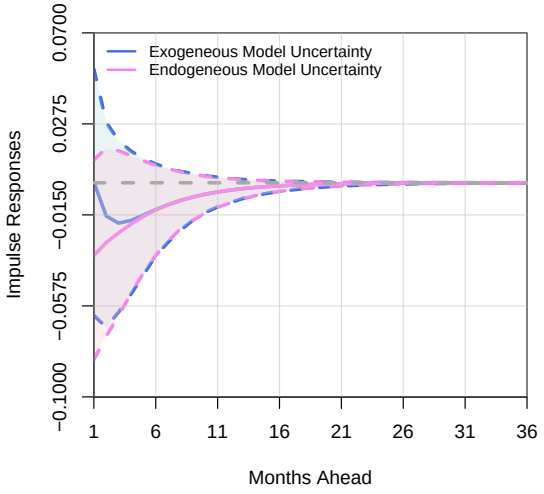
(a) Government Bond Fund Flows



(b) Money Market Fund Flows



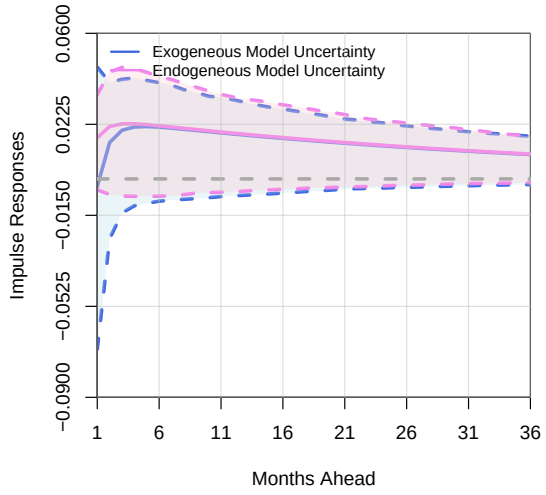
(c) Corporate Bond Fund Flows



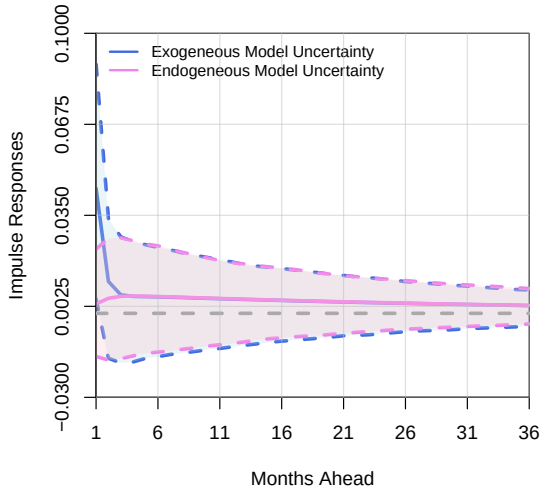
(d) Municipal Bond Fund Flows

Figure IA.8: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes to VIX Shocks

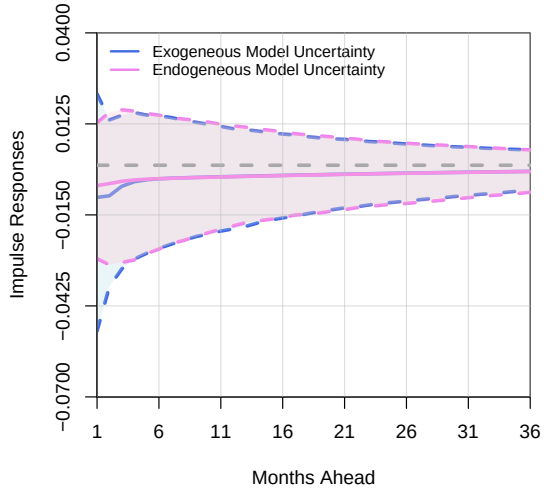
This figure shows the dynamic impulse response functions (IRFs) of fixed-income fund flows to VIX shocks in VAR(1). Other details can be found in the footnote of Figure 7.



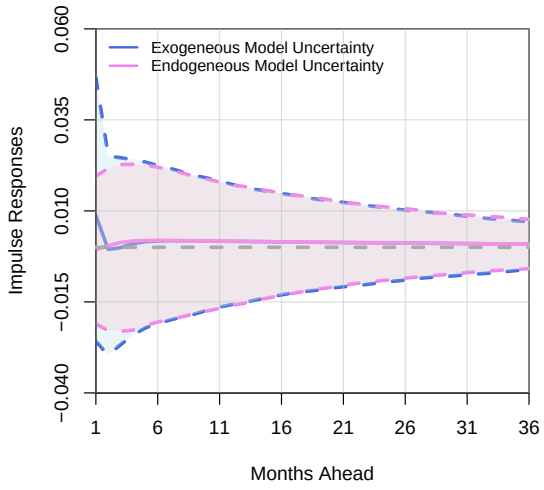
(a) Government Bond Fund Flows



(b) Money Market Fund Flows



(c) Corporate Bond Fund Flows



(d) Municipal Bond Fund Flows

Figure IA.9: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes to Financial Uncertainty Shocks

This figure shows the dynamic impulse response functions (IRFs) of fixed-income fund flows to financial uncertainty shocks in VAR(1). Other details can be found in the footnote of Figure 7.