Model Uncertainty in the Cross Section*

Jiantao Huang †
University of Hong Kong

Ran Shi‡
University of Colorado Boulder

December 2024

Abstract

We develop a transparent Bayesian framework to measure uncertainty in asset pricing models. By assigning a modified class of g-priors to the risk prices of asset pricing factors, our method quantifies the trade-off between mean-variance efficiency and parsimony for asset pricing models to achieve high posterior probabilities. Model uncertainty is defined as the entropy of these posterior probabilities. We prove the model selection consistency property of our procedure, which is missing from the classic g-priors. Acknowledging the possibility of omitting true factors in real applications, we also characterize the maximum degree of contamination that the omitted factors can introduce to our model uncertainty measure. Empirically, we find that model uncertainty escalates during major market events and carries 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

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, SoFiE 2023 Conference, and 2024 European meeting of the Econometric Society. Any errors or omissions are the responsibility of the authors.

†Faculty of Business and Economics, the University of Hong Kong, Pok Fu Lam, Hong Kong; email: huangjt@hku.hk.

‡Leeds School of Business, University of Colorado Boulder, 995 Regent Dr., Boulder, CO 80309, USA; email: ran.shi@colorado.edu.

1 Introduction

Uncertainty is a crucial concern of market participants. Existing research, exemplified by works such as Bloom (2009) and Ludvigson, Ma, and Ng (2021), measures uncertainty in financial markets by focusing on the volatilities of macroeconomic or financial variables. However, another type of uncertainty arises when investors decide on which asset pricing model to use for the cross-section of expected asset returns. This source of uncertainty, referred to as model uncertainty, raises several questions: To what extent does model uncertainty exist in equity markets? Does it contain useful information related to investors' asset allocation decisions? Is there a risk premium associated with model uncertainty?

To answer these questions, we propose a Bayesian approach to measuring uncertainty in asset pricing models. We start by adapting Zellner's g-priors (Zellner, 1986)—the "benchmark" priors for Bayesian inference in the context of model uncertainty (Fernández, Ley, and Steel, 2001)—to linear stochastic discount factor (SDF) models.[1] Under no arbitrage, an SDF connects the expected return of an asset to the linear combination of its covariances with respect to various common factors. The coefficients are the factors' risk prices. We assign our adapted version of g-priors to the risk prices of factors entering the SDF and let the risk prices of other redundant factors be zero.

We then derive 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

[1]The use of g-priors in the empirical 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). Cremers (2002), citing the seminal work of Kass and Raftery (1995), has also pointed out this issue. The g-priors are widely use in linear regression settings in the existing literature. We modify these priors and make analytical inferences about linear SDFs by exploiting their asset pricing implications. Our prior specification is motivated by Liang, Paulo, Molina, Clyde, and Berger (2008) and enjoys several favorable theoretical properties (more on this later).

reaches its maximum.

Of course, any prior specification will induce a set of posterior probabilities for asset pricing models and give us an entropy-based measure of model uncertainty. Perhaps the most appealing property of our prior is its model selection consistency: As the asset return sample becomes large, a posterior probability of one will be assigned to the true SDF.

The existence of a true asset pricing model is a commonly adopted null hypothesis about model uncertainty in the cross-section of returns. It is implicitly regarded as true when *a* particular asset pricing model is in use. It commands that all 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. With model selection consistency, our model uncertainty measure will report a zero entropy (of posterior model probabilities) when the set of candidate models includes the truth.

Another important property of our model uncertainty measure is its robustness against model misspecification. When none of the models under consideration is true, our measure will not be zero. However, it will still remain bounded within a specific limit, far from the theoretical maximum. This limit quantifies the maximum level of contamination that can result from model misspecification.[2]

In particular, our analysis starts by examining the classic g-priors in the context of linear SDF models. This prior induces 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 a Bayesian decision-maker: Higher mean-variance efficiency and model simplicity.

Despite its clean and intuitive implications, the default g-priors suffer from model selection inconsistency, leading to artificially inflated model uncertainty measures. We shed light on the default g-priors' failure through the lens of asset pricing theory. The parameter g controls the prior variance of the SDF, and to generate sizable risk premia observed in asset returns, the SDF must be sufficiently volatile (Hansen and Jagannathan, 1991). A direct implication is

[2]To clarify, our method is not designed for correcting the biases induced by the misspecified set of 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.

that g must scale at the rate of sample sizes to satisfy this economic restriction. However, a g parameter growing at this rate will lead to posterior probabilities that are *biased* toward dense models. We further show that the root cause of this bias is that the classic g-priors select the true factors at the cost of spuriously including the redundant ones.

We modify the g-priors by assigning a heavy-tailed (hyper-)prior distribution to the g parameter. Effectively, we are assuming that the ratio $g/(1+g)$ follows a uniform distribution on (0, 1). This ratio controls how much the observed in-sample Sharpe ratios contribute to the *posterior* SDF volatility through Bayesian learning. Our uniform prior specification reflects an agnostic view on this Bayesian updating process due to the lack of theory guidance. The resulting prior specification is a special case of the mixture of g-priors proposed by Liang, Paulo, Molina, Clyde, and Berger (2008). The implied SDF satisfies the volatility restriction, and, most importantly, the posterior model selection consistency is restored.

Empirically, 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.

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 a unit standard deviation. Then we apply the methodology of Giglio and Xiu (2021) to address omitted variable bias when estimating the risk premia of nontradable factors. We find that, in the US market, the annualized risk premium of model uncertainty shocks is about -3.9%, which is statistically significant.

Moving from pricing implications to quantities, we find that 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 concern only 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:[3] 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 VIX index and financial uncertainty proposed by Ludvigson, Ma, and Ng (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.[4]

Finally, 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 literature on Bayesian inference about asset pricing models (Harvey and Zhou, 1990; Barillas and Shanken, 2018; Chib, Zeng, and Zhao, 2020; Chib and Zeng, 2020; Bryzgalova, Huang, and Julliard, 2023; Avramov, Cheng, Metzker, and Voigt, 2023) and Bayesian portfolio choices (Shanken, 1987; Pástor, 2000; An-

[3]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).

[4]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.

derson and Cheng, 2016; Feng and He, 2022).[5] We contribute to this literature by proposing a new prior that incorporates meaningful economic restrictions on the prior SDF variance and their posterior updating schemes. We highlight some important properties of our approach—including the interpretable posterior probabilities, model selection consistency, and robustness to model misspecification—through posterior analysis, asymptotics, and simulation studies.

Our prior specification that establishes model selection consistency is adapted from Liang, Paulo, Molina, Clyde, and Berger (2008). Our methodological contribution can be viewed as an extension to the panel data setting under which asset returns exhibit cross-sectional dependence. There are other well-known Bayesian methods for model selection (for example, Ishwaran and Rao (2005); Carvalho, Polson, and Scott (2010); Johnson and Rossell (2012)). These methods, as well as Bryzgalova, Huang, and Julliard (2023) in asset pricing applications, approximate the posterior through the computationally intensive Markov chain Monte Carlos. In contrast, our analytical posteriors are highly interpretable and do not require numerical approximations.

Factor models are widely used in economics and finance. For example, in the empirical macro literature since the seminal work of Sargent and Sims (1977), Stock and Watson (1989) and Forni, Hallin, Lippi, and Reichlin (2000) use factor models to examine the shock propagation mechanisms. Stock and Watson (2002a), Stock and Watson (2002b) and Giannone, Reichlin, and Small (2008) apply factor models to macroeconomic forecasts. A common feature of this literature is that factors are *latent variables*, which ought to be estimated in order to capture the commonality among time-series. Latent factor models are also widely used in the study of the term-structure of interest rates and their macroeconomic determinants (Ang and Piazzesi, 2003; Dai and Singleton, 2002; Duffee, 2002, 2011; Ludvigson and Ng, 2009). Unlike this latent factor literature, asset pricing factors are all observable long-short portfolio returns in our analysis. Expected asset returns—and in particular their cross-sectional variation among assets—are linked to their covariances with respect to these observable factors.

Uncertainty in latent factor models mostly arises from choosing the number of factors (Lewbel, 1991; Connor and Korajczyk, 1988, 1993; Forni, Hallin, Lippi, and Reichlin, 2000; Bai and Ng, 2002). For our analysis of a chosen set of observable factors, model uncertainty comes from

[5] Avramov, Cheng, Metzker, and Voigt (2023) study the contribution of model uncertainty to covariance estimation in the context of beta pricing models. We focus on uncertainty about the specification of linear SDFs and developing an interpretable measure of this uncertainty.

the uncertainty about what are the right factors that enter the discount factor, such that it prices the cross-section of asset returns.

There is an increasing interest in developing time-series uncertainty measures for real activities, financial markets, or economic policies (Bloom, 2009; Jurado, Ludvigson, and Ng, 2015; Baker, Bloom, and Davis, 2016; Manela and Moreira, 2017; Hassan, Hollander, Van Lent, and Tahoun, 2019; Dew-Becker and Giglio, 2023; Ludvigson, Ma, and Ng, 2021). In comparison, our model uncertainty measure is conceptually new. It quantifies the equity investors' uncertainty about which asset pricing models better describe the cross-section of expected returns.

2 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, that is, $\mathbb{E}[\boldsymbol{R} \cdot m] = \boldsymbol{0}$,[7] elements of $\boldsymbol{b}$ are market prices of risk for the factors, and the portfolio $\boldsymbol{b}^\top \boldsymbol{f}$ defines the tangency portfolio.

The asset pricing model implied by the SDF is then

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

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

We now formalize the concept of model uncertainty. Without knowing which of the p factors

[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] A linear SDF could be interpreted as the projection of the inter-temporal marginal rate of substitution (IMRS) of a marginal investor onto the span of traded assets. Here, we are not assuming that m is *the* SDF for all security markets, acknowledging the possibility of market segmentation and cross-market arbitrage limits.

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 $\gamma_j = 1$ signifying 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

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

and the resulting asset pricing model becomes

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

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 $\mathcal{M}_{\boldsymbol{\gamma}}$, $\boldsymbol{\lambda} = \mathrm{cov}[\boldsymbol{f},\ \boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}}$. 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).

2.1 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), as follows:

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

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

We work under an empirical Bayes framework and treat $\boldsymbol{\Sigma}$ as known initially. Thus, all follow-up definitions of "posteriors" should be regarded as conditional on this covariance matrix.

We then replace $\boldsymbol{\Sigma}$ with its consistent estimator in the expressions of posterior probabilities. Section IA.1 in the Internet Appendix provides additional details of this approach and its associated econometric properties. We also design and implemented a fully-Bayesian version of our approach by assigning a standard inverse-Wishart prior on $\boldsymbol{\Sigma}$ and use Markov chain Monte Carlos (MCMCs) to integrate it out (detailed in the Internet Appendix), which yields very similar empirical findings.

2.1.1 Baseline specification: the g-priors

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

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

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.[8]

Under our g-prior specification, we can integrate out the risk prices $\boldsymbol{b}_\gamma$ and calculate the marginal likelihood in closed form, as summarized in the following proposition.

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

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

where $\mathrm{SR}^2_{\max}$ is the maximal squared Sharpe ratio of all testing assets $\boldsymbol{R}$,[9] *SR^2_γ is the maximal squared Sharpe ratio of the factors $\boldsymbol{f}_\gamma$ included under model $\mathcal{M}_\gamma$. These two Sharpe ratios are both in-sample quantities.*

[8]One might attempt to assign uninformative priors, such as the Jeffreys priors, to the risk prices. These priors, often being improper ones, can be assigned only to *common* parameters across models. Otherwise, posterior probabilities can be indeterminate. This has been noted in the finance literature (Cremers, 2002; Chib, Zeng, and Zhao, 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.

[9]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, when it does not cause confusion.

2.1.2 Economic implications

To help build intuitions, 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, as follows:

$$\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}') = \frac{\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}]}{\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}'}]} = \exp\left\{ \frac{gT}{2(1+g)} \left(\mathrm{SR}^2_{\boldsymbol{\gamma}} - \mathrm{SR}^2_{\boldsymbol{\gamma}'} \right) - \frac{\log(1+g)}{2} (p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}'}) \right\}. \tag{8}$$

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 constant (characterizing a "risk-neutral" economy), as follows:

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

Although the marginal likelihood in Proposition 1 depends on test assets through $\mathrm{SR}^2_{\max}$, 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, e.g., Cochrane (2005, pages 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, e.g., Gibbons, Ross, and Shanken (1989) and Barillas, Kan, Robotti, and Shanken (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 defines investors' belief updating schemes about the volatility of the SDF, under a given model $\mathcal{M}_{\gamma}$, summarized in Proposition 2.

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

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

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

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

The volatility of the SDF has clear economic implications. As the pricing model (4) is equivalent to $\mathbb{E}[R] = \text{cov}[m_{\gamma}, R]$ for any excess return R, we can see that[10]

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

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 should be “deals” that are too good to be true (Cochrane and Saa-Requejo, 2000; Kozak, Nagel, and Santosh, 2020). This motivates us to further impose the good-deal bound that $\text{var}[m_{\gamma}] \leq h^{2}$, 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 $[\text{SR}_{\infty}, h]$.[11]

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 $\boldsymbol{b}_{\gamma}$ 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 $g p_{\gamma}/T \in [\text{SR}_{\infty}, h]$. As a result, we must have $g = O(T)$.

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

[11] A direct outcome of this restriction is that $g p_{\gamma}/T \in [\text{SR}_{\infty}, h]$ and thus $g = O(T)$, since the total number of factors p ($\geq p_{\gamma}$) is fixed. For conducting valid model selection, the prior information in (6) cannot vanish as $T \to \infty$, because otherwise the prior become irrelevant in large samples and the data likelihood will favor richer sets of factors. We thank an anonymous referee for pointing out this intuition.

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 $g = O(T)$. 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.

2.1.3 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, Paulo, Molina, Clyde, and Berger (2008). If the parameter $a = 4 + 2p_{\boldsymbol{\gamma}}/(TC^2)$ for some $C \in [\text{SR}_\infty, h]$, it is easy to show that the prior volatility of the SDF will be exactly C, satisfying the economic restriction that $\sigma(m_{\boldsymbol{\gamma}})$ should fall in $[\text{SR}_\infty, h]$.

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 $g/(1+g) \sim \text{Uniform}(0, 1)$.

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.e.

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, Paulo, Molina, Clyde, and Berger (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*

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

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 $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ 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 in Section 2.3, 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.

2.2 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,[12] normalized by its upper bound (achieved when all models have the same posterior

[12]Past literature (e.g., Stutzer (1995); Almeida and Garcia (2012); Ghosh, Julliard, and Taylor (2016)) often uses entropy to diagnose asset pricing models and estimate the SDFs non-parametrically in the spirit of relative entropy minimization. Our paper, instead, considers the entropy of the posterior model probabilities as a

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.

2.3 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. All asymptotic properties below are based on the conditions that p is a fixed integer and the eigenvalues of $\boldsymbol{\Sigma}$ are all bounded.

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 $\gamma_{0,j} = 1$, that is, the true model includes factor j, the posterior marginal probability of choosing factor j converges to one in probability: $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] = \sum_{\{\boldsymbol{\gamma}:\ \gamma_j = 1\}} \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 1$[13];*

2. *(Model selection inconsistency) The posterior probability of the true model will always be strictly smaller than one; that is, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}_0} \mid \mathcal{D}] < 1$ with probability one.*

---

measure to quantify the dispersion in the model space. With this regard, our entropy measure is conceptually different from those used in the previous literature.

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

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 by-product from our proof of Proposition 4). 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. Table 1 tabulates the results. Both g-priors and the mixture of g-priors consistently select true factors with posterior probabilities of about one, even in moderately sized return samples. However, inferences using g-priors are plagued by redundant factors. The probability of mistakenly including factors into the SDF remains positive (with average posterior marginal probabilities of about 0.3 to 0.5) and essentially unchanged even as the sample size increases. In contrast, the same probability approaches zero in larger samples under the mixture of g-priors.

By guarding against redundant factors, the mixture of g-priors assign a posterior probability that approaches one to the true model in large samples. The resulting model uncertainty measure $\mathcal{E}$ should converge in probability to zero in simulations where a true model exists, which is confirmed in Table 1. In finite samples, the model uncertainty measure is the smallest under the mixture of g-priors.

Table 1: Simulation Study: Posterior Properties in Finite Samples

Prior	three-year				five-year				50-year			
	$g=2$	$g=4$	$g=16$	mix.g	$g=2$	$g=4$	$g=16$	mix.g	$g=2$	$g=4$	$g=16$	mix.g
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:												
$\gamma_{0,j} = 1$	0.98	0.98	0.98	0.98	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
	(0.03)	(0.03)	(0.03)	(0.04)	(0.01)	(0.01)	(0.02)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{0,j} = 0$	0.49	0.45	0.34	0.25	0.48	0.44	0.34	0.21	0.48	0.44	0.33	0.10
	(0.07)	(0.09)	(0.10)	(0.10)	(0.07)	(0.08)	(0.10)	(0.10)	(0.07)	(0.08)	(0.09)	(0.07)
Posterior Probabilities of Models $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$:												
$\mathcal{M}_{\boldsymbol{\gamma}} = \mathcal{M}_{\boldsymbol{\gamma}_0}$	0.07	0.09	0.18	0.29	0.07	0.10	0.19	0.39	0.08	0.10	0.20	0.67
	(0.04)	(0.06)	(0.12)	(0.17)	(0.04)	(0.06)	(0.12)	(0.19)	(0.04)	(0.06)	(0.12)	(0.22)
$\mathcal{M}_{\boldsymbol{\gamma}} \supset \mathcal{M}_{\boldsymbol{\gamma}_0}$	0.05	0.05	0.05	0.04	0.06	0.06	0.05	0.04	0.06	0.06	0.05	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
$\mathcal{M}_{\boldsymbol{\gamma}} \not\supseteq \mathcal{M}_{\boldsymbol{\gamma}_0}$	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 $\mathcal{E}$:												
	0.42	0.40	0.36	0.31	0.39	0.37	0.33	0.26	0.38	0.36	0.32	0.14
	(0.04)	(0.04)	(0.05)	(0.05)	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)	(0.04)

This table presents simulation results on posterior probabilities of models being correct and factors entering the SDF. Factors under consideration $\boldsymbol{f}$ include the market, SMB, HML, MOM, RMW, CMA, QMJ, FIN, PEAD, and BAB. We follow the assumptions in Propositions 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 of repeating 0.5). The true factors $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ are $\{\text{market}, \text{SMB}, \text{HML}, \text{MOM}, \text{RMW}, \text{CMA}\}$; the redundant factors are $\{\text{QMJ}, \text{FIN}, \text{PEAD}, \text{BAB}\}$. 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 $\widehat{\boldsymbol{\Sigma}}$ is estimated from our sample of US daily 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 three, five, and 50 years. Since there are 252 trading days per year, they represent $T = 756$, $1,260$, and $12,600$ days. For each of the $1,000$ Monte Carlo samples, we calculate the posterior probability of each factor $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$, the posterior probability of each model $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$, 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 parentheses are averages across Monte Carlo samples; numbers in parentheses are standard deviations across Monte Carlo samples.

2.3.1 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 a number of true factors will never be under her consideration. As a result, no prior specification for the risk prices can deliver model selection consistency.

Even when certain true factors are omitted, 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* $\gamma_{0,j} = 1$, $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] \xrightarrow{p} 1$.

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* $\mathcal{E} \leq (p - p_{\boldsymbol{\gamma}_0})/p$ *with probability one, as* $T \to \infty$.

Proposition 7 indicates that if we observe remarkably high model uncertainty $\mathcal{E}$, say $\mathcal{E} \approx 1$, only two possibilities exist: (1) $p_{\boldsymbol{\gamma}_0} = 0$, which implies that the set of factors we are working with does not cover *any* of the true factors; (2) from the perspective of a Bayesian investor, the observed data is entirely uninformative about the true 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 appears quixotic.

The upper bound for model uncertainty in Proposition 7 tends to be loose. The proof of Proposition 7 in the Internet Appendix shows that the upper limit is binding only when all models subsuming factors in $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ (true factors that are *not* omitted) have exactly the same (posterior) probability. To further explore the behaviors of model uncertainty under model misspecification, we consider various simulation setups in Table 2. In particular, we assume that the true SDF model consists of Fama and French (1993) three factors plus an additional factor this is always omitted. The model uncertainty measures, with different pricing factors

Table 2: Simulation Study: The Model Uncertainty Measure under Misspecification

	MOM	RMW	CMA	FIN	PEAD	QMJ	BAB
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:							
$\gamma_{0,j} = 1$	0.93	1.00	1.00	1.00	1.00	0.99	1.00
	(0.18)	(0.03)	(0.01)	(0.01)	(0.01)	(0.07)	(0.01)
$\gamma_{0,j} = 0$	0.47	0.33	0.24	0.41	0.25	0.33	0.35
	(0.34)	(0.27)	(0.21)	(0.32)	(0.22)	(0.26)	(0.27)
Model Uncertainty Measure $\mathcal{E}$:							
	0.44	0.43	0.42	0.41	0.43	0.46	0.45
	(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

The table presents additional simulation results on the model uncertainty measure under model misspecification. We follow the assumptions in Propositions 6 and 7, treating the true SDF as $1 - (\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}$, where $\boldsymbol{b} = 0.5 \times \mathbf{1}$. True factors are the Fama-French three factors plus one additional factor (the column names). This additional factor will always be omitted. Factors under consideration, namely $\boldsymbol{f}$, are the factors in $\{\text{market}, \text{SMB}, \text{HML}, \text{MOM}, \text{RMW}, \text{CMA}, \text{QMJ}, \text{FIN}, \text{PEAD}, \text{BAB}\}$ as in our empirical work, *excluding* the factor defined by each column name (the omitted factor). We simulate $1,000$ return samples according to $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}})$ for three-year data ($T = 756$ 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 $\boldsymbol{\mu}_0 = \widehat{\text{cov}}(\boldsymbol{R}, \boldsymbol{f}_0)\boldsymbol{b}$. For each of the $1,000$ Monte Carlo samples, we calculate the posterior probability of each factor $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$, and the corresponding entropy measure $\mathcal{E}$, under the mixture of g-priors specification. Numbers without parentheses are averages across Monte Carlo samples; numbers in parentheses are standard deviations across Monte Carlo samples.

being omitted, are consistently less than 0.50 and significantly below the theoretical upper bound $(p - p_{\boldsymbol{\gamma}_0})/p$.

3 Data

In our primary empirical implementation, we investigate 14 prominent factors from the past literature: Fama-French five factors (Fama and French, 2016), the momentum factor (Jegadeesh and Titman, 1993), the size, investment, and profitability factors from Hou, Xue, and Zhang (2015), the short-term and long-term behavioral factors from Daniel, Hirshleifer, and Sun (2020), the HML devil (Asness and Frazzini, 2013), quality-minus-junk (Asness, Frazzini, and Pedersen, 2019), and betting-against-beta (Frazzini and Pedersen, 2014) factors from the AQR library. Appendix A presents a detailed description of these factors.

We obtain other uncertainty measures and economic variables from multiple sources. Specifically, we consider indices of economic policy uncertainty (EPU) in Baker, Bloom, and Davis (2016) and three uncertainty measures developed in Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2021). 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 10-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.[14] Since high-quality investment objective data are unavailable until 1990, our sample begins afterward.

4 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 previous three years to compute the posterior model probabilities and model uncertainty measures based on equations (12) and (13).

Bayes factors in these equations are calculated under the mixture of g-priors as in Proposition 3. The behavioral factors in Daniel, Hirshleifer, and Sun (2020) are available only from July 1972, and we use 36-month data in the estimation, so the model uncertainty measure begins in June 1975.

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

[14]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.

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 10 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. The average (median) model uncertainty is 0.70 (0.75), with the first and third quartiles equal to 0.53 and 0.87. Model uncertainty in the cross-section also fluctuates significantly over time—it varies from the lowest value of 0.27 to the highest of 0.99, with a standard deviation of 0.21.

Our model uncertainty measure is countercyclical. In particular, the mid 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 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 impossible to distinguish between models from asset return data. The 2008 crisis is noteworthy also because model uncertainty remains at a high level for a prolonged period until recently. In contrast, it declines shortly after other downturns. Specifically, from 2015 to 2020, model uncertainty has increased from 0.7 to almost 1 (its theoretical maximum). Figure IA.1 of the Internet Appendix extends the model uncertainty measure to 2023. The extremely high uncertainty prevails after the COVID-19 pandemic.

Our findings of overall high and countercyclical model uncertainty are robust to various perturbations in methodology or factors under consideration. In the Appendix, Figure A1 repeats the analysis in Figure 1 using a fully-Bayesian approach which internalizes the uncertainty about covariance matrix estimation; Figure A2 presents various model uncertainty measures calculated from the factors used in Jensen, Kelly, and Pedersen (2023). In the Internet Appendix, Figure IA.2 presents our measure based on four- and five-year rolling windows (instead

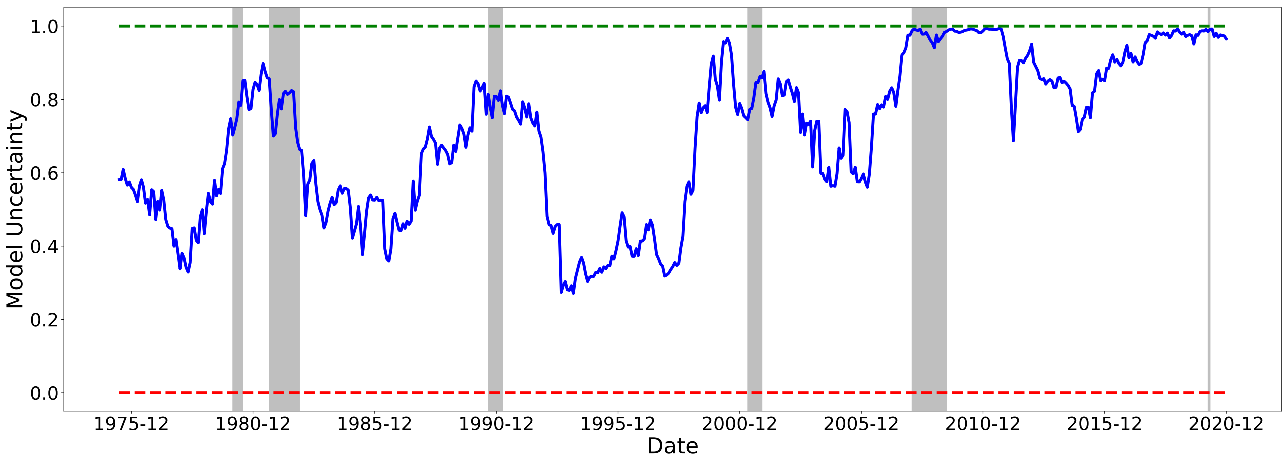


Figure 1: Time Series of Model Uncertainty in the US Equity Market

The figure plots the time series of model uncertainty in the linear factor model. We consider 14 prominent factors in the literature and apply our framework to calculate uncertainty. The red and green lines show the lower and upper bounds of model uncertainty. Shaded areas mark the NBER recession periods.

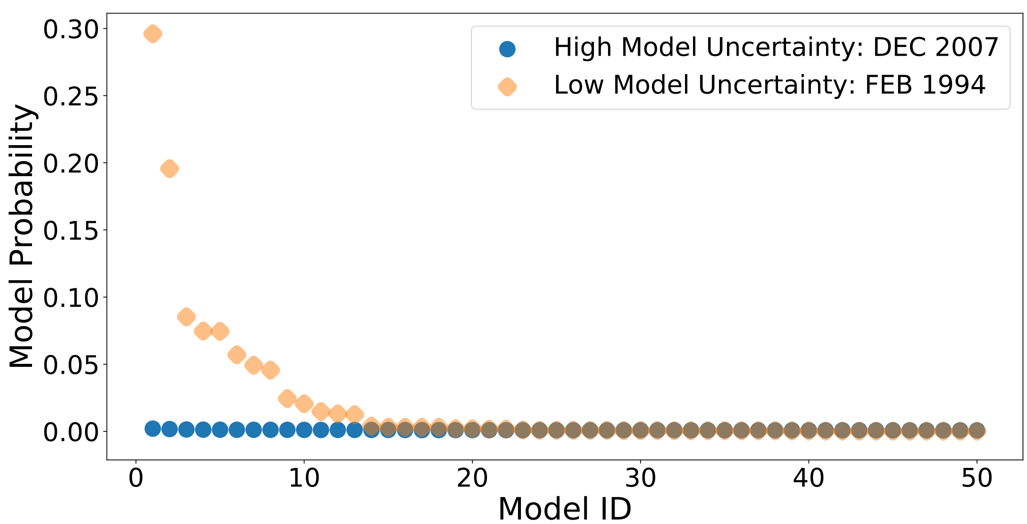


Figure 2: Posterior Probabilities of Top 50 Models: High vs. Low Model Uncertainty

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 previous 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).

of three). Figure IA.4 shows the corresponding time series to Figure 1, when the two mispricing factors proposed by Stambaugh and Yuan (2017) are added to our calculation. In both panels of

Figure IA.3, we rule out factor redundancy (e.g., the two size factors SMB and ME demonstrate considerably high correlations, so including both of them together might artificially inflate model uncertainty). In particular, we include the market, short-term and long-term behavioral factors, momentum, QMJ, and BAB. However, unlike Figure 1, we include only (1) SMB, HML, RMW, and CMA in Panel A and (2) ME, ROE, IA, and HML Devil in Panel. As a result, no closely correlated factors are under consideration simultaneously. Summing up, similar patterns of model uncertainty emerge across all these different analysis.

4.1 International Evidence

International equity market evidence supports the external validity of our findings. We present in Figure 3 the time series of model uncertainty in European and Asia-Pacific stock markets,[15] alongside the US evidence (the 1996-2020 data from Figure 1). Model uncertainty in European markets is more similar to that of the US markets. In the Asia-Pacific markets, model uncertainty is uniquely high in the beginning of 1997, echoing the 1997 Asian financial crisis. Overall, our findings of high and countercyclical model uncertainty in equity markets are supported by the international evidence. Heightened model uncertainty also coincides with major events in corresponding asset markets.

4.2 The Informational Contents of Model Uncertainty

In this subsection, we examine whether our model uncertainty measure is related to other uncertainty indices in recent literature, including the VIX, EPU in Baker, Bloom, and Davis (2016), and three uncertainty measures from Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2021). Table 3 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 establish causal relationships. Our model

[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. We only include nine out of the 14 factors in the US markets since the size (ME), profitability (ROE), and investment (IA) factors in Hou, Xue, and Zhang (2015) and short-term and long-term behavioral factors in Daniel, Hirshleifer, and Sun (2020) are not available for these markets. Since the AQR library provides the QMJ factor from July 1993 and we use a three-year rolling window, our model uncertainty measure begins in June 1996.

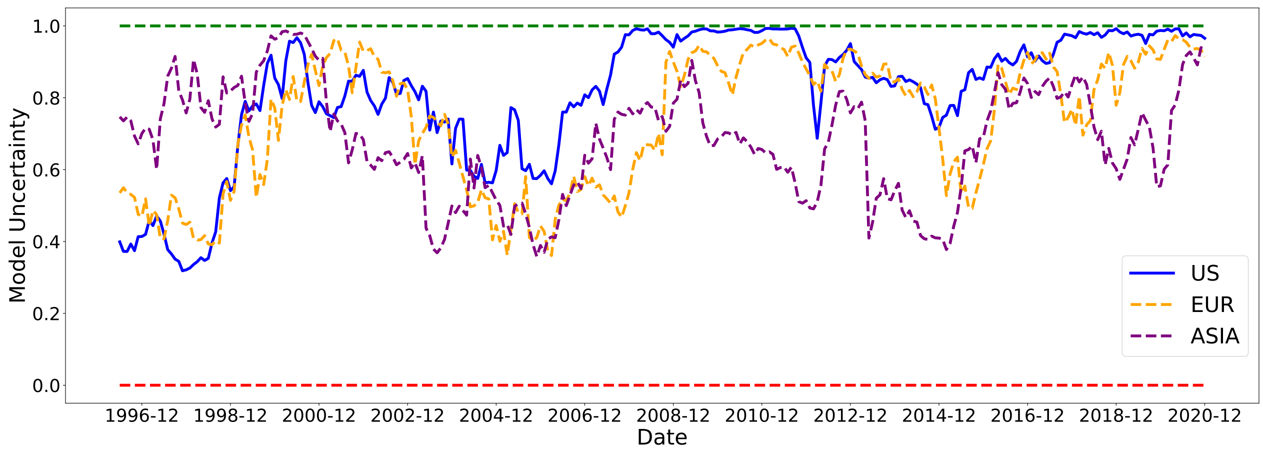


Figure 3: Model Uncertainty in European and Asia Pacific Markets

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. The sample ranges from July 1993 to December 2020. Since we use 3-year rolling window, the model uncertainty index begins in June 1996. The red and green lines in the figure show the lower (0) and upper bounds (1) of model uncertainty, respectively. As a comparison, we include the US model uncertainty index (solid blue line) in Figure 1 during the same period.

uncertainty measure is only significantly associated with the financial uncertainty index of Ludvigson, Ma, and Ng (2021) and the VIX.

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 three-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 IA.5 in the Internet Appendix indicates, model uncertainty displays significant independent variation different from these variables from the existing literature.

4.3 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.

Table 3: 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, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2021) (Fin U, Macro U, and Real U), b) two EPU indices from Baker, Bloom, and Davis (2016) (EPU I and EPU II), c) the CBOE VIX index (VIX), d) the term spread between 10-year and three-month treasuries (TS), and e) the default spread between BAA and AAA corporate bond yields (DS). We standardize all variables to have unit variances. The t-statistics in parentheses are computed based on Newey-West standard errors with 36 lags.

According to Figure IA.6 in the Internet Appendix, posterior marginal probabilities of factor selection demonstrate pronounced time-series variation. During economic downturns, all 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 a probability of 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 BAB factors pass the same scrutiny.

A surprising finding in Figure IA.6 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; that is, the set of factors under consideration does not include any true factors ($p_{\gamma_0} \approx 0$ in the proposition). Low uncertainty factors in the context of high model uncertainty help mitigate this concern.

5 The Market Price of Model Uncertainty

In this section, we investigate whether model uncertainty is priced in the equity markets. If investors become more risk-averse in high model uncertainty time and tilt away from risky assets, model uncertainty shocks could command a negative risk premium in the cross-section of stocks. Conversely, investors may chase risky assets—such as growth stocks as in the literature of learning and growth options (Abel, 1983; Pástor and Veronesi, 2006, 2009)—in certain high-uncertainty periods. Then a positive risk premium of model uncertainty shocks may appear. We take an agnostic view and estimate the risk premium from data.

We extract the AR(1) innovations, namely $\mathcal{E}_t^{ar1}$, to the model uncertainty measure, normalize these shocks to a unit standard deviation, and estimate their risk premiums. Giglio and Xiu (2021) point out that risk premium estimates based on a smaller number of prespecified factors will be subject to the canonical omitted variable bias. To overcome this issue, they propose to estimate risk premiums based on the principal component (PC) factors of a large cross-section of asset returns. We apply their method and estimate the risk premiums of model uncertainty shocks in the US equity market using a large cross-section of 275 characteristic-sorted portfolios.[16]

The three-pass estimator of Giglio and Xiu (2021) can be interpreted in two dimensions. First, it is a two-pass regression estimator as in Fama and MacBeth (1973) using the PC factors as controls. Alternatively, it is the average return of a mimicking portfolio constructed from the PC factors.

Table 4 presents the risk premium estimates. Model uncertainty shocks command a significantly negative risk premium of -0.066, which implies that investors are willing to pay a premium to hedge against heightened model uncertainty using stocks. For robustness, we report results based on different numbers of PCs (extracted from the 275 base portfolios). The risk premium estimates are remarkably stable. Recall that we have normalized the model uncertainty shocks to a unit standard deviation; hence, the monthly risk premium $\lambda_{\mathcal{E}} = -0.066$ is effectively a Sharpe ratio. For the ease of interpretation, if we equate the standard deviation

[16] We consider the 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, and (11) short-term reversal.

Table 4: Risk Premia of Model Uncertainty Shocks: Monthly Estimates

Number of PCs:	5	6	7	8	9	10
$\lambda_{\mathcal{E}}$	-0.066***	-0.067***	-0.067***	-0.065***	-0.062***	-0.060***
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%

The table reports the risk premia estimates of model uncertainty shocks ($\mathcal{E}_t^{ar1}$) based on the three-pass method of Giglio and Xiu (2021). In all estimations, we standardize $\mathcal{E}_t^{ar1}$ to have a unit variance. In particular, we project $\mathcal{E}_t^{ar1}$ onto the space of large PCs of 275 Fama-French characteristic-sorted portfolios in the US market. The number of latent factors ranges from five to 10. 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. Sample: 1975/07 - 2020/12.

of the model uncertainty shocks to that of the market portfolio (17% per year in our sample), the annualized risk premium equals $(\sqrt{12})\lambda_{\mathcal{E}} \times 17\% \approx -3.9\%$, about a half of the market risk premium in magnitude.

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 serve only as simple yet imperfect benchmarks. Second, we do not claim that model uncertainty shocks are necessary to *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.

6 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.

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

$$\text{Flow}_{i,t} = \text{TNA}_{i,t} - \text{TNA}_{i,t-1} \times (1 + R_{i,t}), \tag{14}$$

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

$$\text{Flow}_t^{\mathcal{O}} = \frac{\sum_{i \in \mathcal{O}} \text{Flow}_{i,t}}{\text{CRSP-Market-Cap}_{t-1}}, \tag{15}$$

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 following reduced-form VAR(l) model:

$$\boldsymbol{Y}_t = \boldsymbol{B}_0 + \boldsymbol{B}_1 \boldsymbol{Y}_{t-1} + \cdots + \boldsymbol{B}_l \boldsymbol{Y}_{t-l} + \boldsymbol{B}_{\boldsymbol{x}} \boldsymbol{X}_t + \boldsymbol{u}_t, \tag{16}$$

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; that is, $\boldsymbol{u_t} = \boldsymbol{S\epsilon_t}$, where $\boldsymbol{S}$ is a $k \times k$ non-singular matrix, and $\boldsymbol{\epsilon_t}$ is a $k \times 1$ vector of structural shocks. 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.

6.1 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 (16). We next use impulse response functions (IRFs) to better understand the dynamic effects and propagating mechanisms of uncertainty shocks.

IRFs depend heavily on the identification assumption, that is, 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, whereas 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.

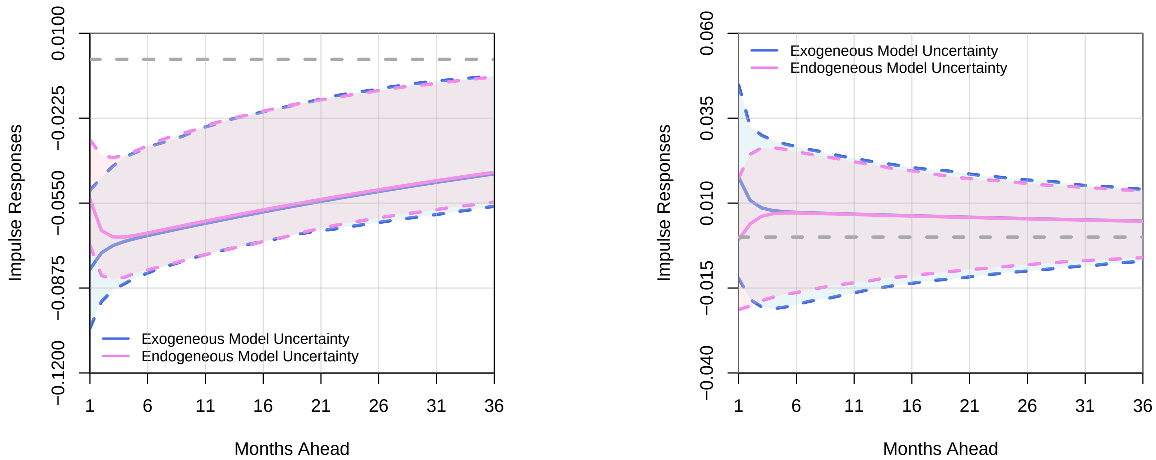
We present the empirical results based on two identification assumptions. In the first case, we place model uncertainty first in the VAR (exogenous assumption). Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does 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).

Figure 4 plots the dynamic responses of aggregate fund flows to model uncertainty shocks. The sample ranges from January 1991 to December 2020. In the VAR regression, we include the lagged market return and VIX index as control variables, and the lag of VAR is chosen by BIC and equals one. 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 begin at 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)).

6.2 Heterogeneity in Flow Responses

We first study the heterogeneous responses of different equity mutual funds to model uncertainty shocks. In particular, we divide equity funds into four categories: (a) style funds that specialize in factor investing, (b) sector funds that invest in specific industries (e.g., gold and oil), (c) small-cap funds that invest in small stocks, and (d) large-cap funds. Panel (a) in Figure 5 shows that model uncertainty shocks reduce future style fund flows, and the effects are long-lasting. This observation is intuitive because style funds rely mostly on factor strategies used in constructing model uncertainty. Moreover, we observe significantly negative IRFs of small-cap funds (see Panel (c)), although the effects are less persistent than those of 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.

Next, we divide all fixed-income mutual funds into four categories: (a) government bond



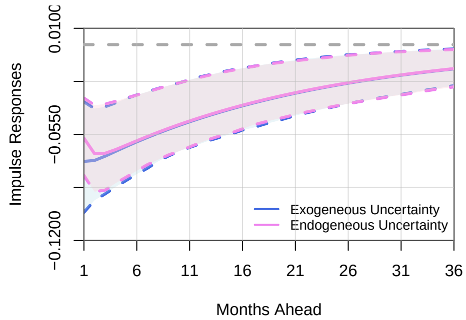
(a) IRFs of Equity Fund Flows **(b)** IRFs of Fixed-Income Fund Flows

Figure 4: Impulse Responses of Aggregate Fund Flows to Model Uncertainty Shocks

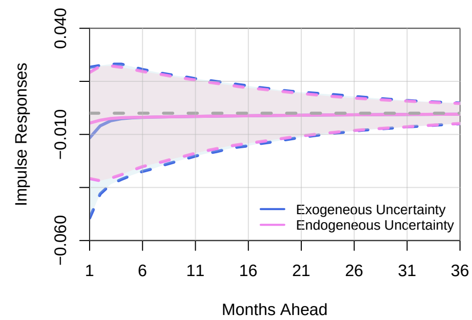
The figure shows the dynamic impulse response functions (IRFs) of fund flows to model uncertainty shocks in VAR(1). The shaded area denotes the 90 percent standard error bands. We consider mutual fund flows to aggregate equity and fixed-income markets in the US. We consider two identification assumptions, (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 of 1991:01–2020:12.

funds, (b) money market funds, (c) corporate bond funds, and (d) municipal bond funds. In Panel (e) of Figure 5, we document sharp dynamic fund inflows to 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. On the contrary, the IRFs of other fixed-income fund flows are insignificant.

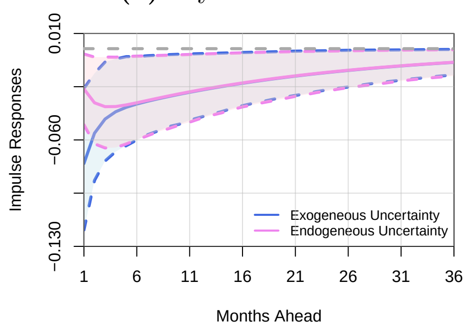
What is the uniqueness of model uncertainty? Previous analyses show that model uncertainty is significantly correlated with VIX and financial uncertainty, so these alternative uncertainty measures are likely to capture similar impulse responses of mutual fund flows. For comparison, Figures IA.7 and IA.8 in the Internet Appendix plot the dynamic responses of equity or fixed-income fund flows to VIX and financial uncertainty shocks. We control the lagged model uncertainty in each VAR 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. Interestingly, we document significant inflows to money market funds after positive VIX shocks, but fixed-income fund flows do not respond



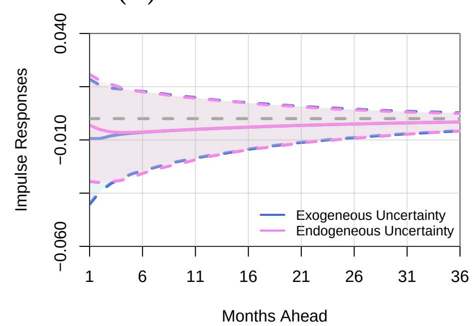
(a) Style Fund Flows



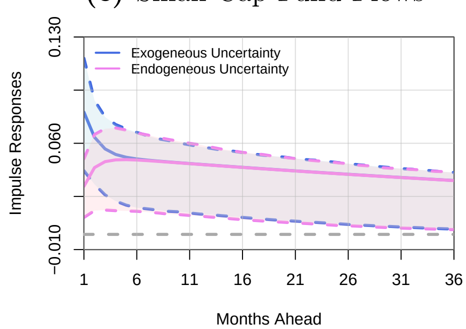
(b) Sector Fund Flows



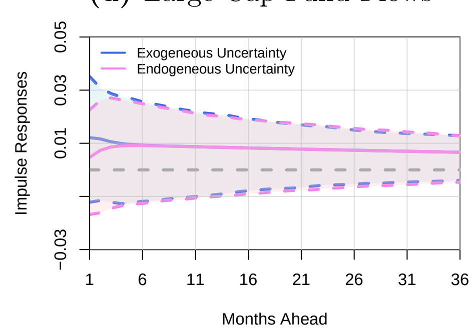
(c) Small-Cap Fund Flows



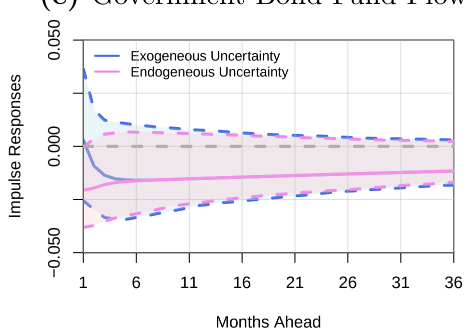
(d) Large-Cap Fund Flows



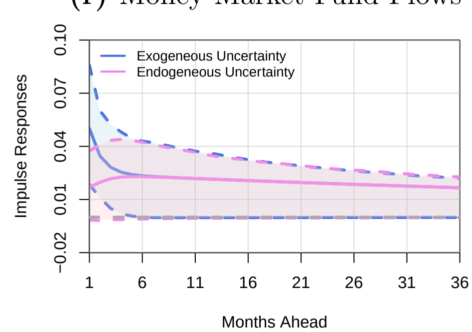
(e) Government-Bond Fund Flows



(f) Money-Market Fund Flows



(g) Corporate-Bond Fund Flows



(h) Municipal-Bond Fund Flows

Figure 5: Impulse Responses of Fund Flows with Different Investment Objective Codes to Model Uncertainty Shocks

The 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. In Panels (a)–(d), we consider equity fund flows with different investment objective codes (style, sector, small-cap, and large-cap). In Panels (e)–(h), 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 of 1998:01–2020:12.

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 unique dynamic responses of fund flows, and notably, they are distinct from traditional volatility-based measures, including VIX and financial uncertainty.

7 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

$$\boldsymbol{b}_{bma} := \mathbb{E}[\boldsymbol{b} \mid \mathcal{D}] = \sum_{\boldsymbol{\gamma}} \mathbb{E}[\boldsymbol{b} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}] \times \mathbb{P}(\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}). \tag{17}$$

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 or information criteria) 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 5 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) the Carhart (1997) four-factor model (Carhart4), (4) the Fama and French (2016) five-factor model (FF5), (5) the Hou, Xue, and Zhang (2015) q-factor model (HXZ4), and (6) the Daniel, Hirshleifer, and Sun (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; that is, $H_0 : \text{SR}^2_{bma} = \text{SR}^2_{\gamma}$.[17]

Table 5: 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, Xue, and Zhang (2015) q-factor model, and (7) DHS3: the market factor plus two behavioural factors in Daniel, Hirshleifer, and Sun (2020). We also report the results on testing the null hypothesis that the Sharpe ratio of BMA is equal to the model γ, i.e., $H_0 : \text{SR}^2_{bma} = \text{SR}^2_{\gamma}$. 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 5. First, BMA outperforms traditional factor models in the 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

[17] We draw $100,000$ sample paths of $\{R_{\gamma,t^\star}, R_{bma,t^\star}\}_{t^\star=1}^{T}$ with replacement, where T is the sample size of the observed dataset. If the difference in Sharpe ratios between BMA and model γ 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 the 10% (5%, 1%) significance level.

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 5 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 5 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.

8 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 has 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 that the observed data is not informative about a fixed "true" asset pricing model.

Model uncertainty displays countercyclical 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 using stocks.

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 carry significant weight on the decision-makers' final choices.

References

Abel, A. B. (1983): "Optimal investment under uncertainty," *American Economic Review*, 73(1), 228–233.

Abramowitz, M., and I. A. Stegun (1965): *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, vol. 55. US Government printing office, Washington, D.C.

Almeida, C., and R. Garcia (2012): "Assessing misspecified asset pricing models with empirical likelihood estimators," *Journal of Econometrics*, 170(2), 519–537.

Anderson, E. W., and A.-R. Cheng (2016): "Robust Bayesian portfolio choices," *Review of Financial Studies*, 29(5), 1330–1375.

Ang, A., and M. Piazzesi (2003): "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables," *Journal of Monetary Economics*, 50(4), 745–787.

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.*

Bai, J., and S. Ng (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70(1), 191–221.

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.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010): "The horseshoe estimator for sparse signals," *Biometrika*, 97(2), 465–480.

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.

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, Princeton, New Jersey.

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.

Connor, G., and R. A. Korajczyk (1988): "Risk and return in an equilibrium APT: Application of a new test methodology," *Journal of Financial Economics*, 21(2), 255–289.

——— (1993): "A test for the number of factors in an approximate factor model," *Journal of Finance*, 48(4), 1263–1291.

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

Dai, Q., and K. J. Singleton (2002): "Expectation puzzles, time-varying risk premia, and affine models of the term structure," *Journal of Financial Economics*, 63(3), 415–441.

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 (2023): "Cross-sectional uncertainty and the business cycle: Evidence from 40 years of options data," *American Economic Journal: Macroeconomics*, 15(2), 65–96.

Duffee, G. R. (2002): "Term premia and interest rate forecasts in affine models," *Journal of Finance*, 57(1), 405–443.

——— (2011): "Information in (and not in) the term structure," *Review of Financial Studies*, 24(9), 2895–2934.

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

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

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

Fama, E. F., and J. D. MacBeth (1973): "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, 81(3), 607–636.

Feng, G., and J. He (2022): "Factor investing: A Bayesian hierarchical approach," *Journal of Econometrics*, 230(1), 183–200.

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

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The generalized dynamic-factor model: Identification and estimation," *Review of Economics and Statistics*, 82(4), 540–554.

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

Ghosh, A., C. Julliard, and A. Taylor (2016): "What is the Consumption-CAPM missing? An Information-Theoretic Framework for the Analysis of Asset Pricing Models," *Review of Financial Studies*, 30, 442–504.

Giannone, D., L. Reichlin, and D. Small (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55(4), 665–676.

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.

Ishwaran, H., and J. S. Rao (2005): "Spike and slab variable selection: Frequentist and Bayesian strategies," *Annals of Statistics*, 33(2), 730 – 773.

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*, pp. 361–379. University of California Press, Berkeley, Calif.

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.

Jensen, T. I., B. Kelly, and L. H. Pedersen (2023): "Is there a replication crisis in finance?," *Journal of Finance*, 78(5), 2465–2518.

Johnson, N. L., S. Kotz, and N. L. Johnson (1995): *Continuous Univariate Distributions, Second Edition*, vol. 2. John Wiley & Sons, Inc.

Johnson, V. E., and D. Rossell (2012): "Bayesian model selection in high-dimensional settings," *Journal of the American Statistical Association*, 107(498), 649–660.

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.

Lewbel, A. (1991): “The rank of demand systems: theory and nonparametric estimation,” *Econometrica*, pp. 711–730.

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.

Ludvigson, S. C., and S. Ng (2009): “Macro factors in bond risk premia,” *Review of Financial Studies*, 22(12), 5027–5067.

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 P. Veronesi (2006): “Was there a Nasdaq bubble in the late 1990s?,” *Journal of Financial Economics*, 81(1), 61–100.

——— (2009): “Technological revolutions and stock prices,” *American Economic Review*, 99(4), 1451–83.

Sargent, T. J., and C. A. Sims (1977): “Business cycle modeling without pretending to have too much a priori economic theory,” Working Papers 55, Federal Reserve Bank of Minneapolis.

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

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.

Stock, J. H., and M. W. Watson (1989): “New Indexes of Coincident and Leading Economic Indicators,” *NBER Macroeconomics Annual*, pp. 351–394.

——— (2002a): “Forecasting using principal components from a large number of predictors,” *Journal of the American Statistical Association*, 97(460), 1167–1179.

——— (2002b): “Macroeconomic forecasting using diffusion indexes,” *Journal of Business & Economic Statistics*, 20(2), 147–162.

Stutzer, M. (1995): “A Bayesian Approach to Diagnosis of Asset Pricing Models,” *Journal of Econometrics*, 68(2), 367 – 397.

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, Cambridge, UK.

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

Appendix A Factor Details

Fama-French five factors. Fama and French (2016) introduce a five-factor model that includes 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 previous three to 12 months continue their performance in the following three to 12 months. We download the momentum (MOM) factor from Ken French's data library.

The q-factor model (three additional). Hou, Xue, and Zhang (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 (two additional). Daniel, Hirshleifer, and Sun (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, Frazzini, and Pedersen (2019) define high-quality stocks as ones with higher profits, faster growth, lower betas/volatilities, and a 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.

Table IA.I reports the annualized mean returns and Sharpe ratios of the above 14 factors. The sample runs 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 through recent years, we exclude these two factors in the main analysis. We download the data from Robert Stambaugh's website.

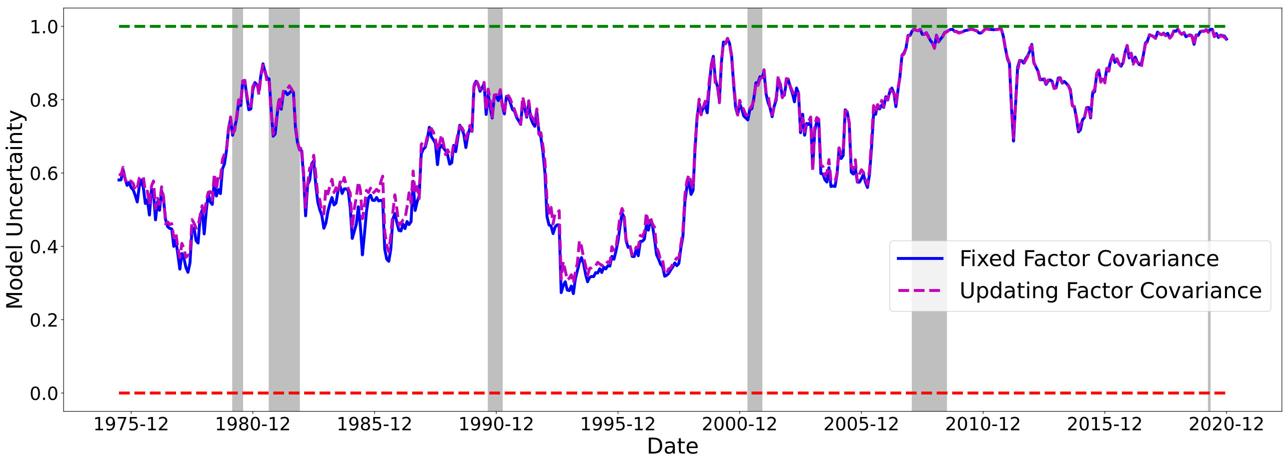
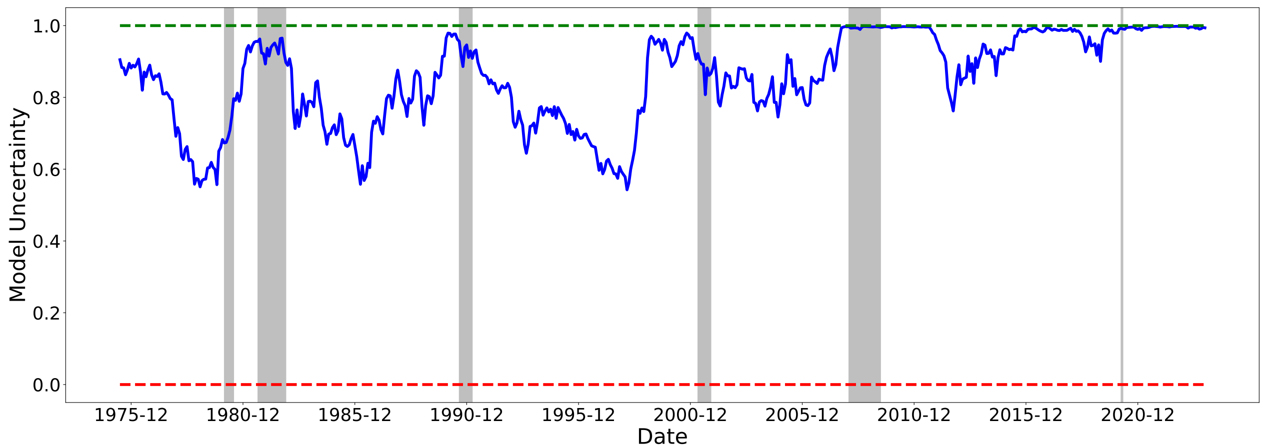
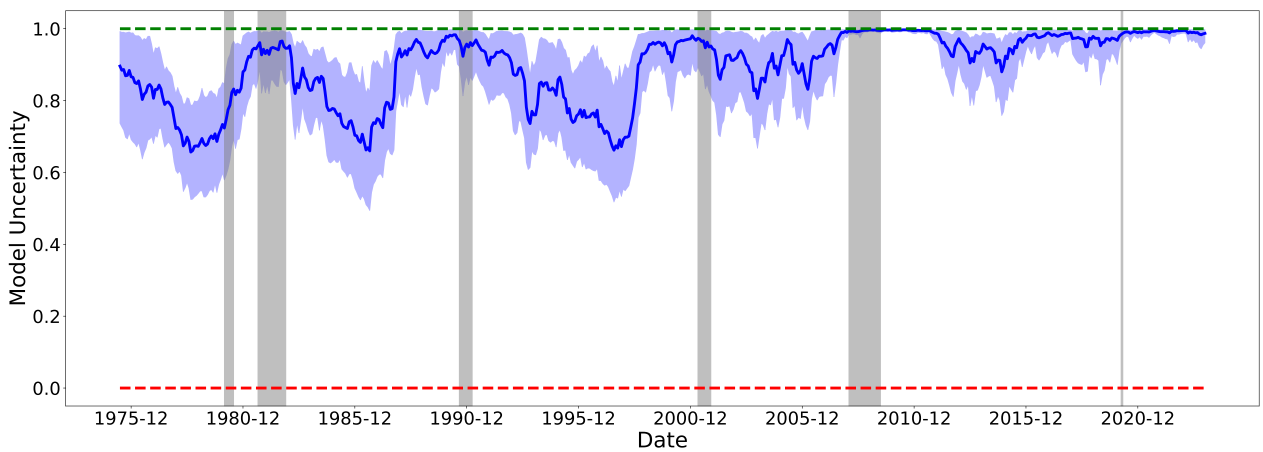


Figure A1: Time Series of Model Uncertainty in the US Equity Market: Accounting for Uncertainty in $\boldsymbol{\Sigma}$

The figure plots the time series of model uncertainty in the linear factor model using a three-year rolling window. We consider 14 prominent factors in the literature and apply our framework to calculate uncertainty. The blue solid line shows the model uncertainty measure conditional on the sample covariance matrix of factor returns, whereas the purple dotted line presents the measure based on the Metropolis-Hasting algorithm, in which we account for the estimation uncertainty in $\boldsymbol{\Sigma}$. The red and green lines show the lower and upper bounds of model uncertainty. Shaded areas mark the NBER recession periods.



(a) market plus 13 cluster factors



(b) market plus 13 randomly selected factors (1,000 randomly chosen sets of factors)

Figure A2: Time Series of Model Uncertainty, 153 Factors in Jensen, Kelly, and Pedersen (2023)

The figure plots the time series of model uncertainty in the 153 factors provided by Jensen, Kelly, and Pedersen (2023). In Panel (a), we use the 13 "cluster factors" in Jensen, Kelly, and Pedersen (2023) (the average returns of all factors under each theme of clusters), as well as the market portfolio, to measure model uncertainty in the factor model. In Panel (b), we randomly select one factor from each of the 13 clusters in Jensen, Kelly, and Pedersen (2023). The model uncertainty is then measured based on the market factor plus the 13 randomly selected factors. We repeat this exercise 1,000 times. The blue solid line in Panel (b) denotes the average model uncertainty measure across 1,000 randomly chosen sets of factors, with the 5th and 95th quantiles given by the shaded blue area. The red and green lines show the lower and upper bounds of model uncertainty. Shaded areas mark the NBER recession periods.

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.

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 $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ 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

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

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

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

[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).

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,* $\boldsymbol{R}_1, \dots, \boldsymbol{R}_T \overset{\text{iid}}{\sim} \mathcal{N}(\text{cov}[\boldsymbol{R},\, \boldsymbol{f}_{\boldsymbol{\gamma}0}]\boldsymbol{b}_{\boldsymbol{\gamma}0},\, \boldsymbol{\Sigma})$*, then the maximal in-sample Sharpe ratio of* $\boldsymbol{f}_{\boldsymbol{\gamma}}$*, namely factors under consideration for model* $\mathcal{M}_{\boldsymbol{\gamma}}$*, satisfies*

$$\mathbb{E}\left[\text{SR}^2_{\boldsymbol{\gamma}}\right] = \boldsymbol{b}_{\boldsymbol{\gamma}0}^\top \left(\text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0}\right] - \text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\right) \boldsymbol{b}_{\boldsymbol{\gamma}0} + \frac{p_{\boldsymbol{\gamma}}}{T}$$

Proof. When the data are generated from model $\mathcal{M}_{\boldsymbol{\gamma}0}$, for any $\boldsymbol{\gamma}$,

$$\boldsymbol{f}_{\boldsymbol{\gamma},t} \overset{\text{iid}}{\sim} \mathcal{N}\left(\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}0}\boldsymbol{b}_{\boldsymbol{\gamma}0},\ \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\right), \quad t = 1, \dots, T,$$

where $\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}0} = \text{cov}[\boldsymbol{f}_{\boldsymbol{\gamma}}, \boldsymbol{f}_{\boldsymbol{\gamma}0}]$ is a $p_{\boldsymbol{\gamma}} \times p_{\boldsymbol{\gamma}0}$ matrix. As a result,

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

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

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

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

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

Now define $\delta_{\boldsymbol{\gamma}0} = \boldsymbol{b}_{\boldsymbol{\gamma}0}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}0}\boldsymbol{b}_{\boldsymbol{\gamma}0}$. By definition,

$$\delta_{\boldsymbol{\gamma}0} - \delta_{\boldsymbol{\gamma}} = \boldsymbol{b}_{\boldsymbol{\gamma}0}^\top \left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}0} - \boldsymbol{C}_{\boldsymbol{\gamma}0,\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}0}\right)\boldsymbol{b}_{\boldsymbol{\gamma}0} = \boldsymbol{b}_{\boldsymbol{\gamma}0}^\top \text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\boldsymbol{b}_{\boldsymbol{\gamma}0}. \tag{IA.3}$$

As a result,

$$\delta_{\boldsymbol{\gamma}} = \delta_{\boldsymbol{\gamma}0} - \boldsymbol{b}_{\boldsymbol{\gamma}0}^\top \text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\boldsymbol{b}_{\boldsymbol{\gamma}0} = \boldsymbol{b}_{\boldsymbol{\gamma}0}^\top \left(\text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0}\right] - \text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\right)\boldsymbol{b}_{\boldsymbol{\gamma}0}.$$

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 ($\text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]$ is small). If $\mathcal{M}_{\boldsymbol{\gamma}}$ includes all true factors in $\boldsymbol{f}_{\boldsymbol{\gamma}0}$, the conditional variance $\text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]$ is zero. Under this scenario, $\mathbb{E}\left[\text{SR}^2_{\boldsymbol{\gamma}}\right] = \boldsymbol{b}_{\boldsymbol{\gamma}0}^\top \left(\text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}0}\right]\right)\boldsymbol{b}_{\boldsymbol{\gamma}0} +$

$p_{\boldsymbol{\gamma}}/T \geq \mathbb{E}\left[\mathrm{SR}^2_{\boldsymbol{\gamma}_0}\right]$ (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 $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}$ 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 $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}$, satisfies*

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

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

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 $1 - 2e^{-\psi^2/2}$,*

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

Based on Proposition IA.2, the expected value of $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}$ converges to that of $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ (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 $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}$ concentrates heavily around that of $\mathrm{SR}^2_{\boldsymbol{\gamma}}$. A byproduct of this result is that $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}$ must be a consistent estimator. Without further clarification, $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ will be replaced with $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}$ throughout our empirical studies; all theoretical results involving $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ 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 $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} = \overline{\boldsymbol{f}}^{\top}_{\boldsymbol{\gamma}} \boldsymbol{S}^{-1}_{\boldsymbol{\gamma}} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}$, 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

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

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 $T\delta_{\gamma}$.[3] Equivalently,

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

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

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

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

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

Simple algebra gives the formula in bullet point one.

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

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

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

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

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

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

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

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

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

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

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

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

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

The condition $1/(T-3) < \eta < 3/2$ implies $\sqrt{T-1}/(T-3) < \psi + \sqrt{p_{\gamma}} < 3/2\sqrt{T-1}$. 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* $\boldsymbol{\Sigma}_{\gamma} = \mathrm{var}[\boldsymbol{f}_{\gamma}]$, $\boldsymbol{C}_{\gamma} = \mathrm{cov}[\boldsymbol{R}, \boldsymbol{f}_{\gamma}]$, *and* $\boldsymbol{\Sigma} = \mathrm{var}[\boldsymbol{R}]$, *then*

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

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

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

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

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

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,

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

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

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

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

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

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

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

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

[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.

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),

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

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

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

thus, it is trivial to show that

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

The volatility of the SDF can be computed as

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

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

$$\text{var}[m_\gamma] = \frac{gp_\gamma}{T}$$

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

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

□

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 $\text{SR}_{\max}^2$, $\boldsymbol{\Sigma}$, and $\boldsymbol{S}$ in (7) are all independent of model $\mathcal{M}_{\boldsymbol{\gamma}}$, when $\boldsymbol{\gamma} \neq \mathbf{0}$:

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

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

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

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

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

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

□

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 $\mathcal{M}_{\gamma_0}$, $T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} = T\delta_{\boldsymbol{\gamma}} + p_{\boldsymbol{\gamma}} + O_p(\sqrt{T})$.*

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

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

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

Lemma IA.3. *If $\boldsymbol{f}_{\gamma_0} \subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, that is, factors of model $\mathcal{M}_{\boldsymbol{\gamma}}$ subsume all factors in $\boldsymbol{f}_{\gamma_0}$ that define the true model, we have $T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} - T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}_0} = p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0} + O_p(1)$.*

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

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

Similarly,

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

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 $\hat{\boldsymbol{z}} \sim \mathcal{N}(\boldsymbol{0},\, \boldsymbol{I}_{(p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}_0})})$. According to Lemma IA.1, if $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$ (in parallel to $\boldsymbol{f} \subseteq \boldsymbol{R}$),

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

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

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

Case I. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \not\subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, 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, $\delta_{\boldsymbol{\gamma}} - \delta_{\boldsymbol{\gamma}_0} = -\boldsymbol{b}^\top_{\boldsymbol{\gamma}_0}\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\boldsymbol{b}_{\boldsymbol{\gamma}_0} < 0$ if $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \not\subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}_0 \neq \mathbf{0}$ (here we also implicitly assume that every pair of factors is not perfectly correlated). Then $T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} - T\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}_0} = -(\boldsymbol{b}^\top_{\boldsymbol{\gamma}_0}\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\boldsymbol{b}_{\boldsymbol{\gamma}_0})T + O(1) + o_p(T) \xrightarrow{p} -\infty$ 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 $\gamma_j = 0$, $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \not\subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$ 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 $\mathcal{M}_{\gamma}$ against the true model $\mathcal{M}_{\gamma_0}$ is calculated as

$$\mathrm{BF}(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)=\underbrace{\exp\left(\frac{T\widehat{\mathrm{SR}}_{\gamma}^2-T\widehat{\mathrm{SR}}_{\gamma_0}^2}{2}\right)}_{I_1}\underbrace{\left(\frac{\widehat{\mathrm{SR}}_{\gamma}^2}{\widehat{\mathrm{SR}}_{\gamma_0}^2}\right)^{-\frac{p_\gamma+2}{2}}}_{I_2}\underbrace{\left(\frac{T\widehat{\mathrm{SR}}_{\gamma_0}^2}{2}\right)^{\frac{p_{\gamma_0}-p_\gamma}{2}}}_{I_3}\underbrace{\frac{\underline{\Gamma}\left(\frac{p_\gamma+2}{2},T\widehat{\mathrm{SR}}_{\gamma}^2/2\right)}{\underline{\Gamma}\left(\frac{p_{\gamma_0}+2}{2},T\widehat{\mathrm{SR}}_{\gamma_0}^2/2\right)}}_{I_4}. \quad \text{(IA.7)}$$

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

$$\frac{\widehat{\mathrm{SR}}_{\gamma}^2}{\widehat{\mathrm{SR}}_{\gamma_0}^2}\xrightarrow{p}\frac{\delta_\gamma}{\delta_{\gamma_0}},$$

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

$$I_4\xrightarrow{p}\begin{cases}(p_\gamma/2)(p_\gamma/2-1)(p_\gamma/2-2)\cdots(p_{\gamma_0}/2+2)(p_{\gamma_0}/2+1), & \text{if } p_\gamma\ge p_{\gamma_0}+1\\ \left[(p_{\gamma_0}/2)(p_{\gamma_0}/2-1)(p_{\gamma_0}/2-2)\cdots(p_\gamma/2+2)(p_\gamma/2+1)\right]^{-1}, & \text{if } p_\gamma\le p_{\gamma_0}-1\\ 1, & \text{if } p_\gamma=p_{\gamma_0}\end{cases}$$

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

Case I. If $\boldsymbol{f}_{\gamma_0}\not\subseteq\boldsymbol{f}_\gamma$, according to our proof for Proposition 4,

$$T\widehat{\mathrm{SR}}_{\gamma_0}^2=\delta_{\gamma_0}T+o_p(T),\qquad T\widehat{\mathrm{SR}}_{\gamma}^2-T\widehat{\mathrm{SR}}_{\gamma_0}^2=-(\delta_{\gamma_0}-\delta_\gamma)\times T+o_p(T),\qquad \delta_{\gamma_0}>\delta_\gamma>0.$$

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 $p_\gamma>p_{\gamma_0}$, $I_3\xrightarrow{p}0$ and obviously $\mathrm{BF}(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)\xrightarrow{p}0$.

If $p_\gamma\le p_{\gamma_0}$, $I_3=O_p(T^{(p_{\gamma_0}-p_\gamma)/2})$. As $I_1T^{(p_{\gamma_0}-p_\gamma)/2}=o_p(1)$, we have $I_1I_3=o_p(1)$ 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 $\boldsymbol{f}_{\gamma_0}\subset\boldsymbol{f}_\gamma$, Lemma IA.3 implies that $I_1=O_p(1)$. As it is always true that $p_\gamma>p_{\gamma_0}$ when $\boldsymbol{f}_{\gamma_0}\subset\boldsymbol{f}_\gamma$, $I_3\xrightarrow{p}0$. 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 $\mathcal{M}_\gamma$ as long as $\boldsymbol{\gamma}\neq\boldsymbol{\gamma}_0$. As a result,

$$\mathbb{P}[\mathcal{M}_{\gamma_0}\mid\mathcal{D}]=1-\sum_{\gamma\neq\gamma_0}\mathbb{P}[\mathcal{M}_\gamma\mid\mathcal{D}]=1-\sum_{\gamma\neq\gamma_0}\frac{\mathrm{BF}(\boldsymbol{\gamma},\boldsymbol{\gamma}_0)}{\sum_{\gamma'}\mathrm{BF}(\boldsymbol{\gamma}',\boldsymbol{\gamma}_0)}\xrightarrow{p}1.$$

□

IA.2.6 Proof of Proposition 6

Proof. Now the true model is defined by factors in $\boldsymbol{f}_0$ and $\boldsymbol{f}_{\gamma_0} = \boldsymbol{f}_0 \cap \boldsymbol{f}$. We define $\boldsymbol{C}_{\boldsymbol{\gamma}} = \text{cov}\left[\boldsymbol{f}_{\boldsymbol{\gamma}},\, \boldsymbol{f}_0\right]$, $\boldsymbol{\Sigma}_0 = \text{var}[\boldsymbol{f}_0]$. Applying Lemma IA.2, we have that for any model $\mathcal{M}_{\boldsymbol{\gamma}}$,

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

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 $\boldsymbol{f}_{\gamma_0} \not\subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$ as $T \to \infty$.*

Proof. For the ease of exposition, we first define $\widetilde{\boldsymbol{f}}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\gamma_0} \cap \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{f}}_{\gamma_0} = \boldsymbol{f}_{\gamma_0} \backslash \widetilde{\boldsymbol{f}}_{\boldsymbol{\gamma}}$. Now consider the model, namely $\mathcal{M}_{\boldsymbol{\gamma}'}$, defined by factors $\boldsymbol{f}_{\boldsymbol{\gamma}'}^\top = (\boldsymbol{f}_{\boldsymbol{\gamma}}^\top,\, \widehat{\boldsymbol{f}}_{\gamma_0}^\top)$. Noticing that

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

The last inequality is due to the fact that $\widehat{\boldsymbol{f}}_{\gamma_0} \neq \emptyset$ when $\boldsymbol{f}_{\gamma_0} \not\subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, which implies $\boldsymbol{f}_{\boldsymbol{\gamma}} \subset \boldsymbol{f}_{\boldsymbol{\gamma}'}$.

The Bayes factor $\text{BF}(\boldsymbol{\gamma},\, \boldsymbol{\gamma}')$, 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 $\text{BF}(\boldsymbol{\gamma},\, \boldsymbol{\gamma}') \xrightarrow{p} 0$ as $T \to \infty$. Now, for any $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\gamma_0} \not\subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$,

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

□

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 $\boldsymbol{f}_{\gamma_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$, which is equivalent to

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

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

$$\begin{aligned}
\mathcal{E} &= -\frac{1}{p\log 2}\sum_{\boldsymbol{\gamma}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] \\
&= -\frac{1}{p\log 2}\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] - \frac{1}{p\log 2}\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\subseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] \\
&\le -\frac{1}{p\log 2}\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] - \frac{1}{p\log 2}\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\frac{1}{2^{p-p_{\gamma_0}}}\right)\left(\frac{1}{2^{p-p_{\gamma_0}}}\right) \\
&= -\frac{1}{p\log 2}\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] + \frac{p - p_{\gamma_0}}{p}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\text{Prob}_0\left[\mathcal{E} \ge \frac{p-p_{\boldsymbol{\gamma}}}{p}\right] &\le \text{Prob}_0\left[\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] \le 0\right] \\
&= 1 - \text{Prob}_0\left[\lim_{T\to\infty}\left\{\sum_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] > -\frac{1}{T}\right\}\right] \\
&\le 1 - \text{Prob}_0\left[\lim_{T\to\infty}\left\{\min_{\boldsymbol{\gamma}:\, \boldsymbol{f}_{\gamma_0}\nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}} \log\left(\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right]\right)\mathbb{P}\left[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}\right] > -\frac{1}{(2^p - 2^{p_{\boldsymbol{\gamma}}})T}\right\}\right]
\end{aligned}$$

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 $\boldsymbol{f}_{\gamma_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$,

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

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

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

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

IA.3 Additional Tables

Table IA.I: Summary Statistics of the 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

The table reports the annual mean returns and Sharpe ratios of the 14 factors listed in Appendix A. The full sample runs from July 1972 to December 2020. Subsample I (II) starts from July 1972 to September 1996 (October 1996 to December 2020).

Table IA.II: 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 are from Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2021). EPU I and EPU II are two economic policy uncertainty sequences from Baker, Bloom, and Davis (2016). VIX is the CBOE volatility index. The t-statistics are computed using Newey-West standard errors with 36 lags.

IA.4 Additional Figures

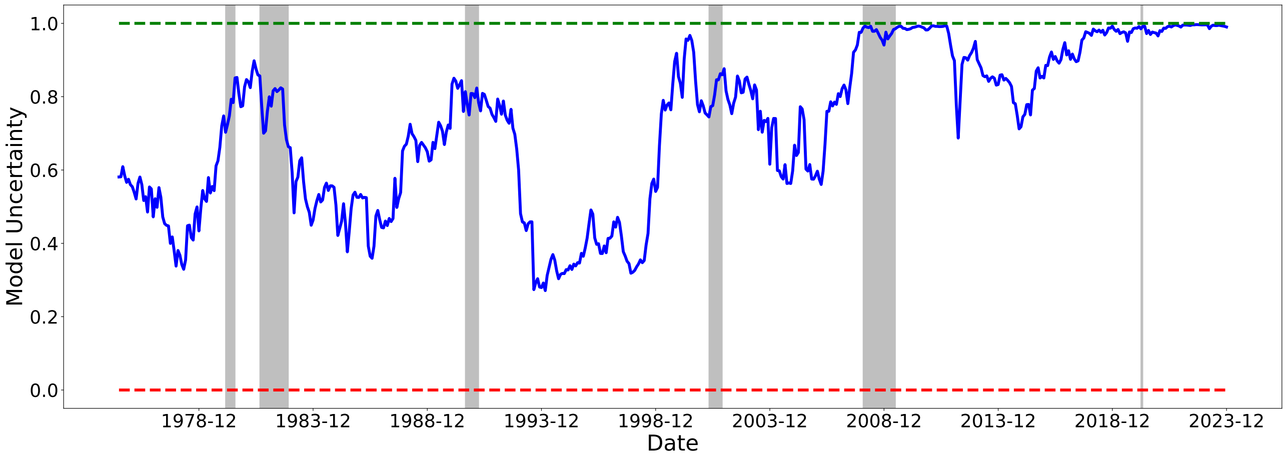


Figure IA.1: Time Series of Model Uncertainty in the US Equity Market

The figure plots the time series of model uncertainty in the linear factor model using a three-year rolling window. The sample starts from 1975 to 2023. We consider 14 prominent factors in the literature and apply our framework to calculate uncertainty. The red and green lines show the lower and upper bounds of model uncertainty. Shaded areas mark the NBER recession periods.

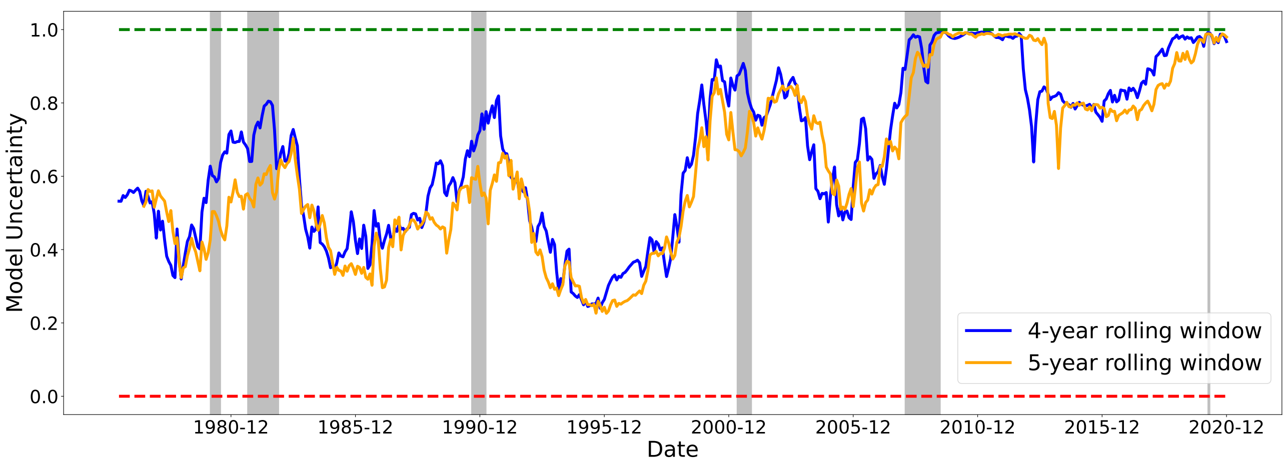
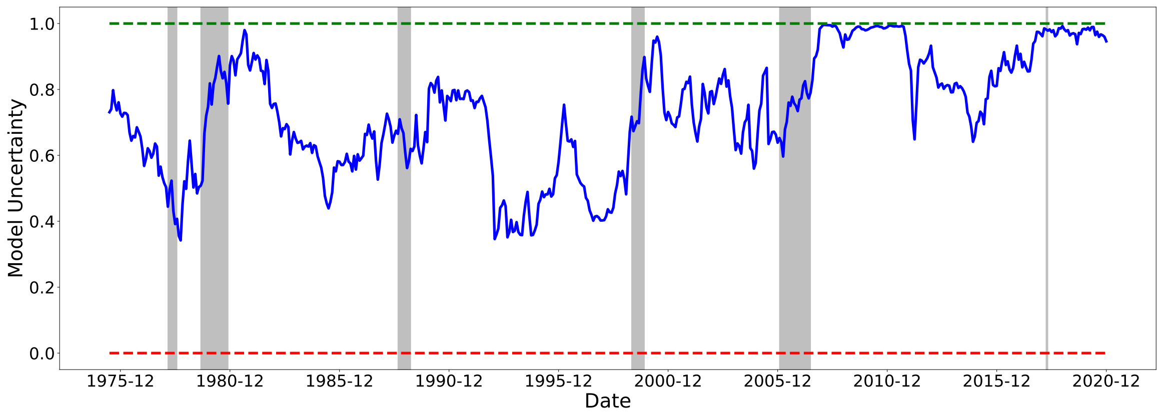
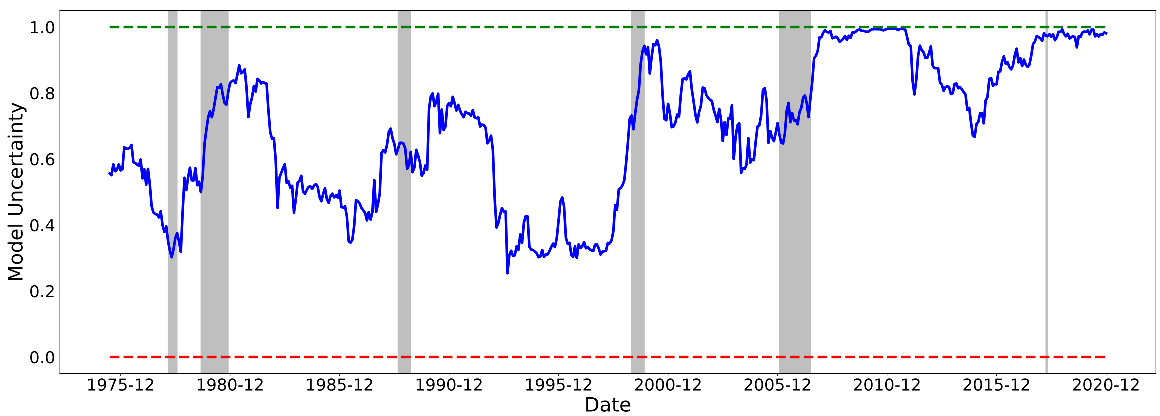


Figure IA.2: Alternative Rolling Windows



(a) SMB, HML, RMW, and CMA



(b) ME, ROE, IA, and HML Devil

Figure IA.3: Time Series of Model Uncertainty in the US Equity Market: Exclude Potentially Redundant Factors

The figure plots the time series of model uncertainty in the linear factor model. In both Panels A and B, we always include market, short-term and long-term behavioral factors, momentum, QMJ, and BAB. Unlike Figure 1, we include (1) SMB, HML, RMW, and CMA in Panel A, and (2) ME, ROE, IA, and HML Devil in Panel B. The red and green lines show the lower and upper bounds of model uncertainty. Shaded areas mark the NBER recession periods.

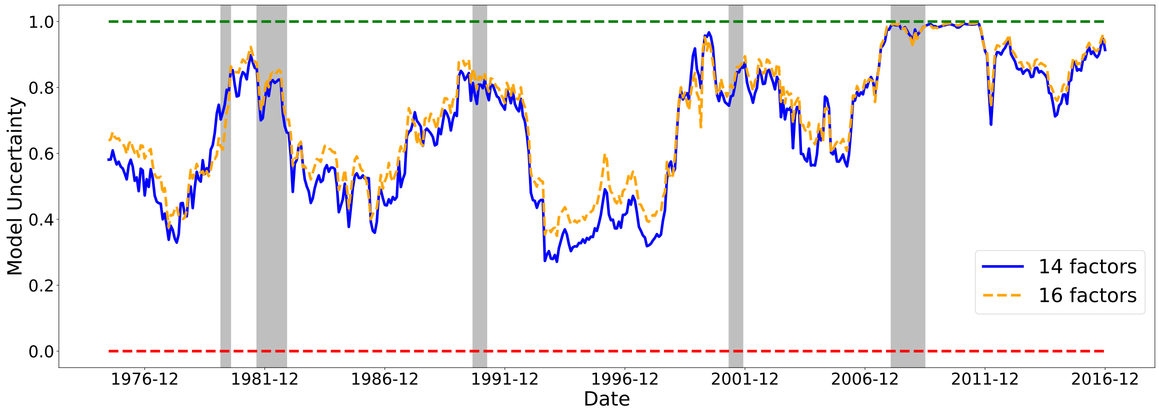


Figure IA.4: Time-Series of Model Uncertainty: Including two mispricing factors

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.

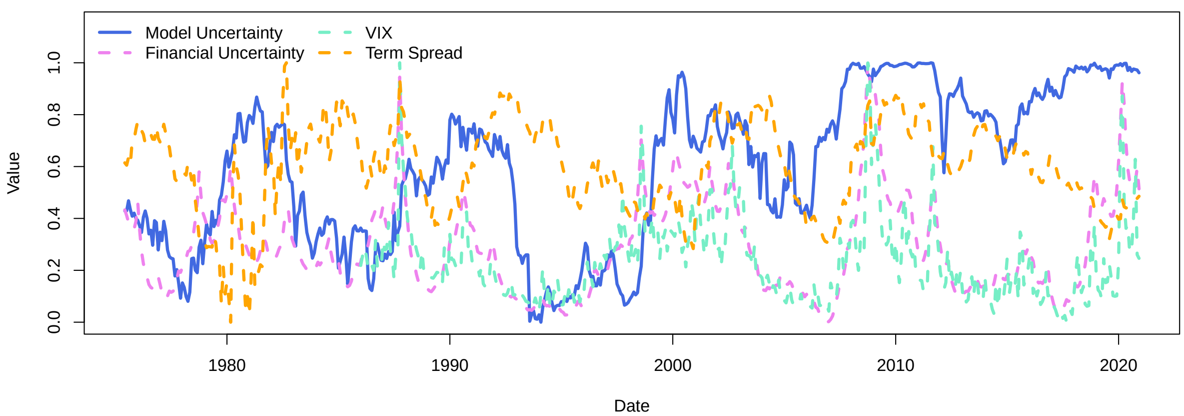
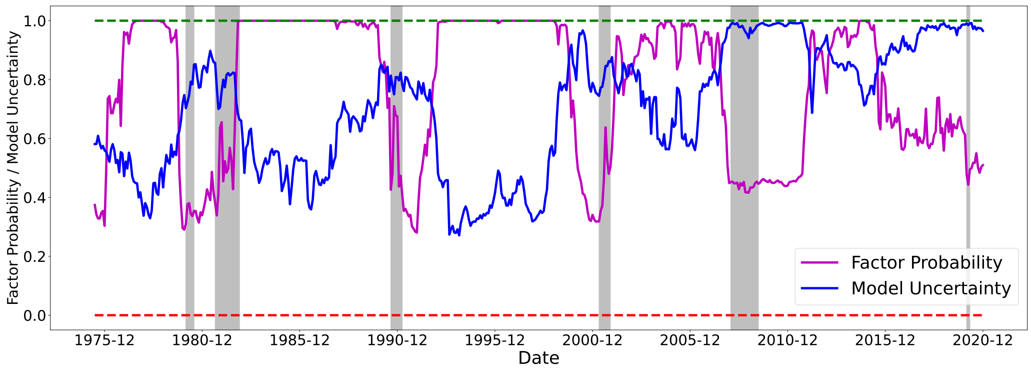
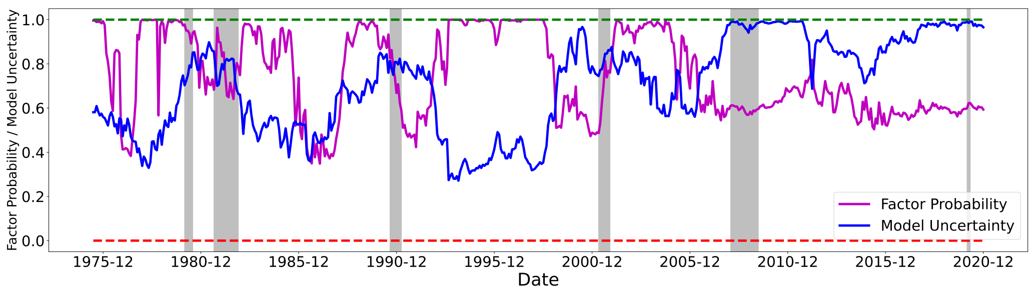


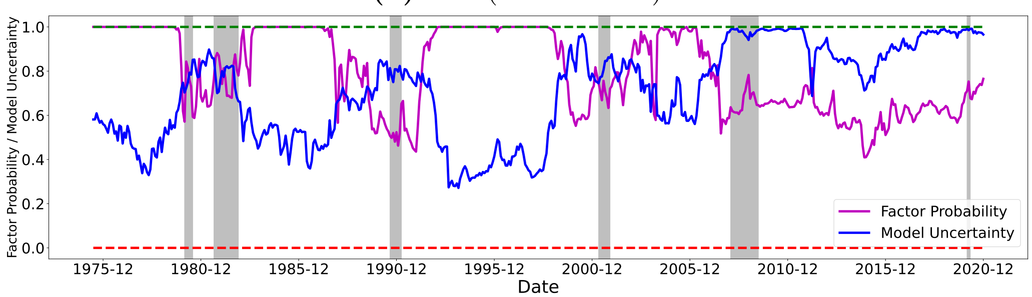
Figure IA.5: Time-Series of Model Uncertainty, Financial Uncertainty, VIX, and Term Spread



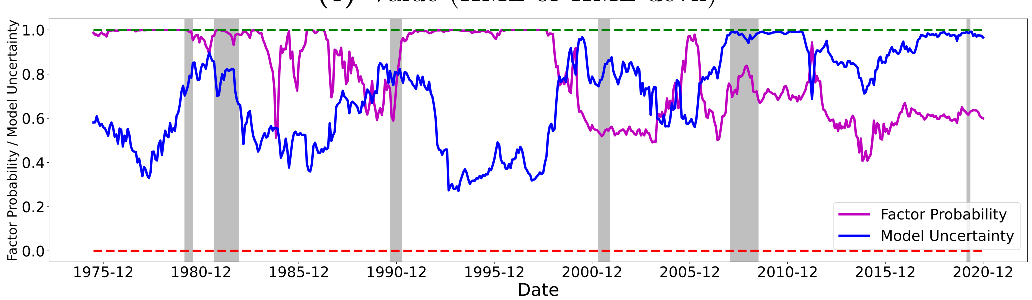
(a) MKT



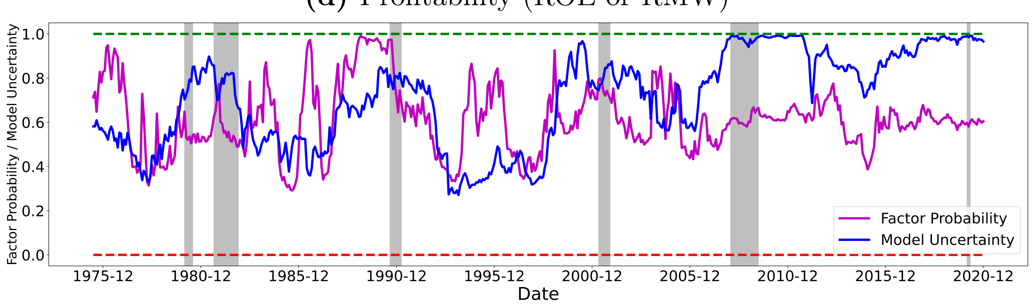
(b) Size (SMB or ME)



(c) Value (HML or HML devil)

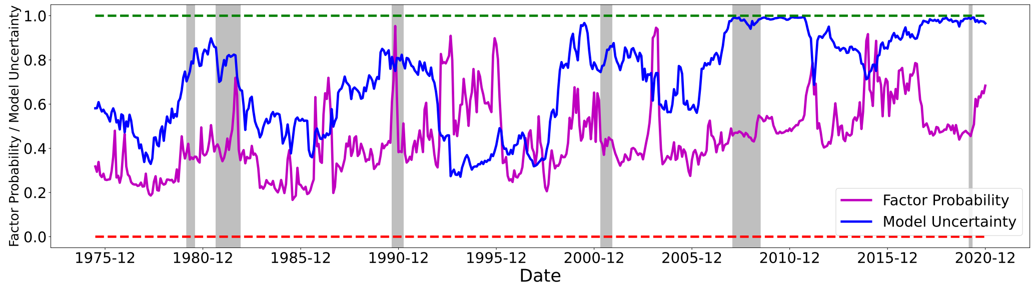


(d) Profitability (ROE or RMW)

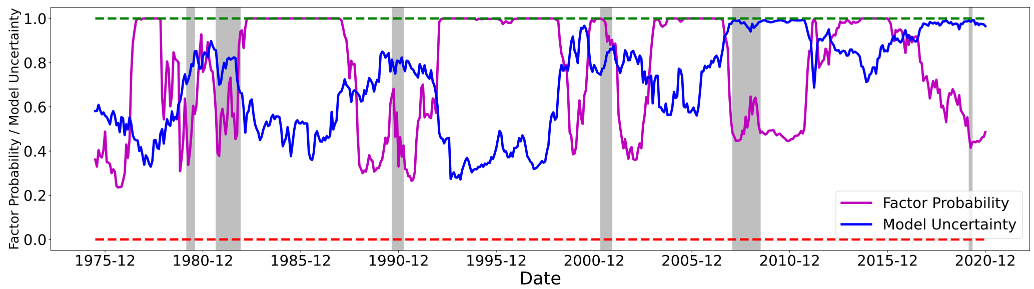


(e) Investment (IA or CMA)

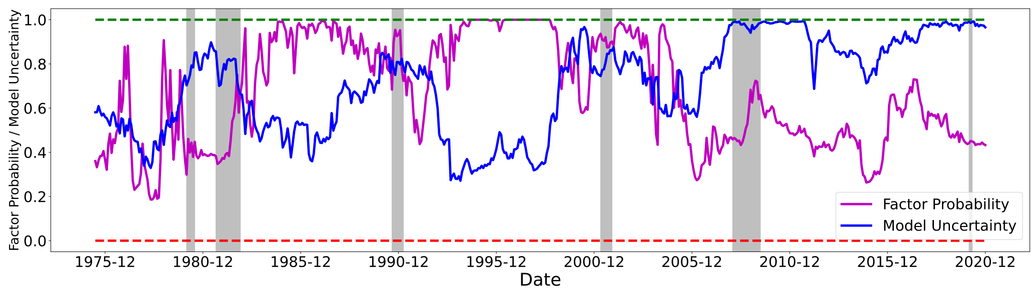
Figure IA.6: Time Series of Posterior Factor Probabilities



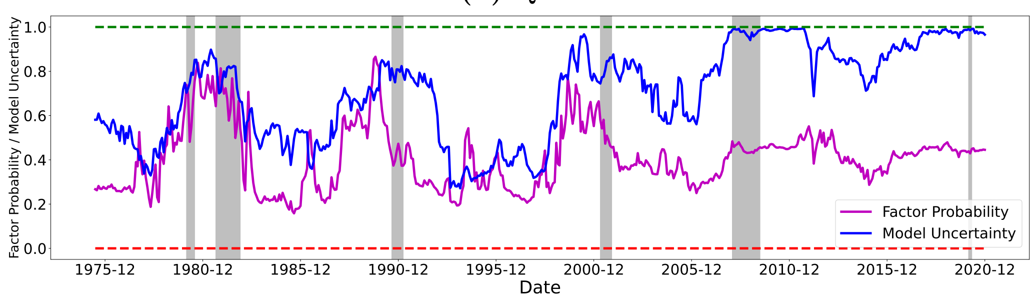
(f) Momentum



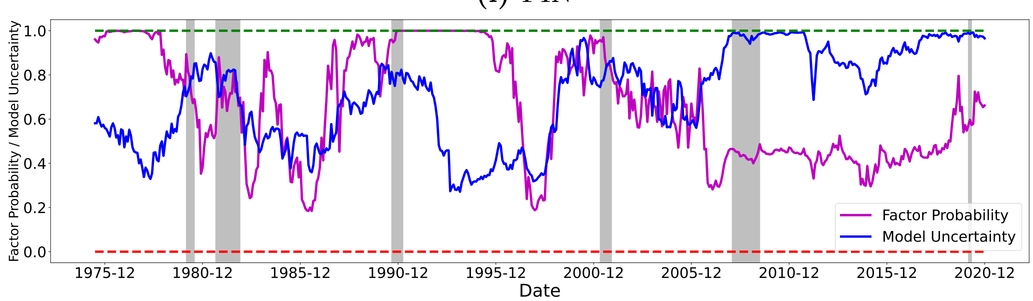
(g) BAB



(h) QMJ



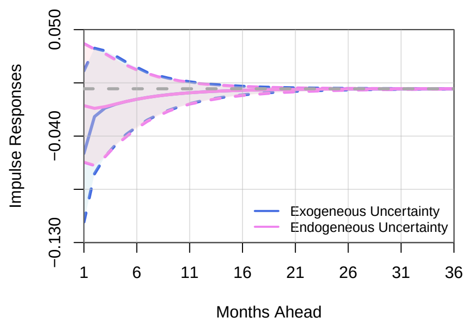
(i) FIN



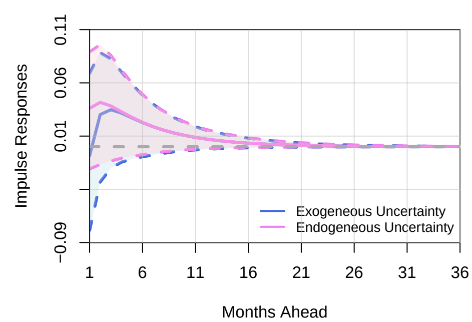
(j) PEAD

Figure IA.6: Time Series of Posterior Factor Probabilities (Continued)

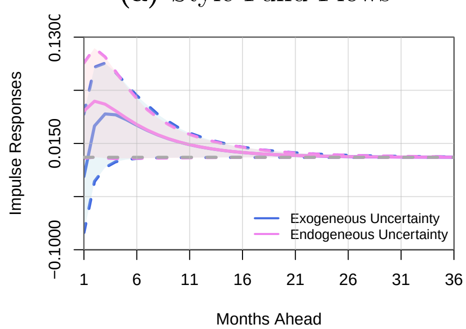
The figures plot the time series of posterior marginal probabilities of 14 factors. The time series of factor probabilities starts from June 1975. Shaded areas are NBER-based recession periods for the US.



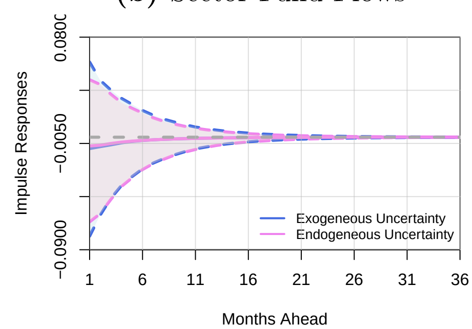
(a) Style Fund Flows



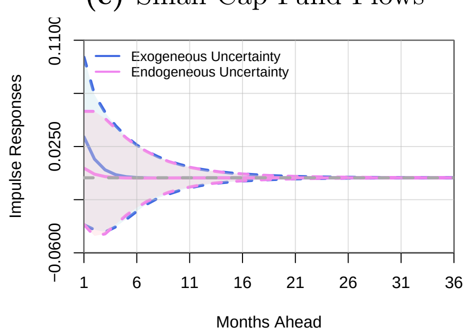
(b) Sector Fund Flows



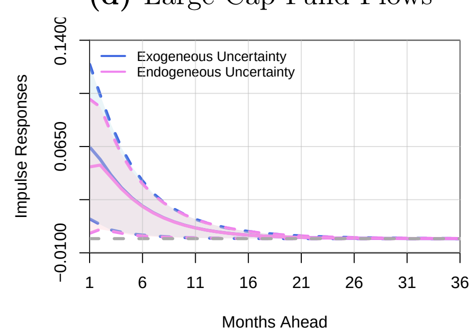
(c) Small-Cap Fund Flows



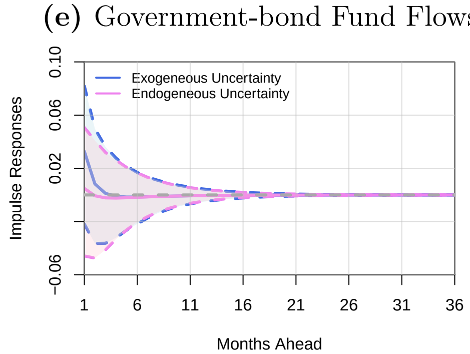
(d) Large-Cap Fund Flows



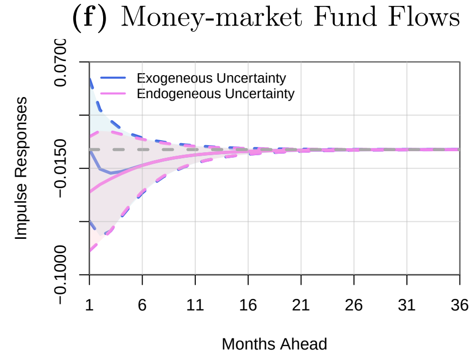
(e) Government-bond Fund Flows



(f) Money-market Fund Flows



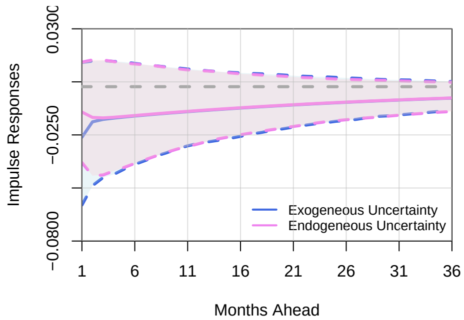
(g) Corporate-bond Fund Flows



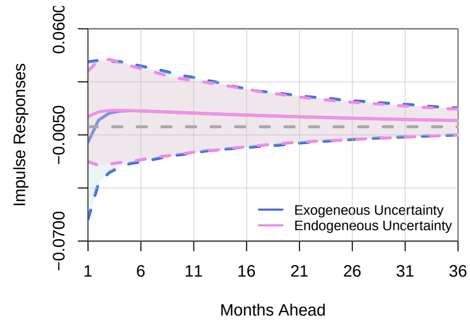
(h) Municipal-bond Fund Flows

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

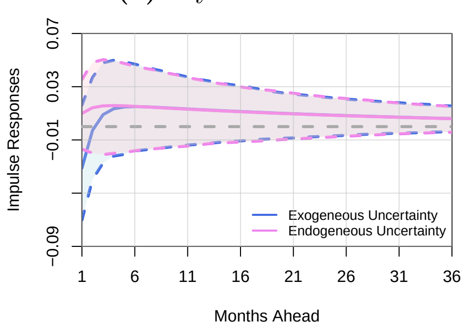
This figure shows the dynamic impulse response functions (IRFs) of fund flows to VIX shocks in VAR(1). Other details can be found in the footnote of Figure 5.



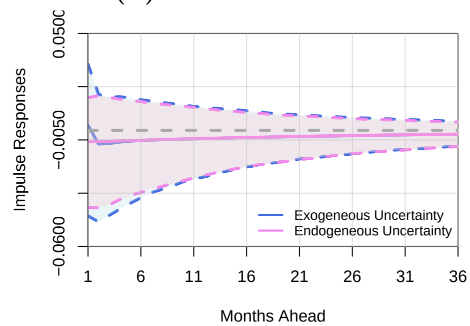
(a) Style Fund Flows



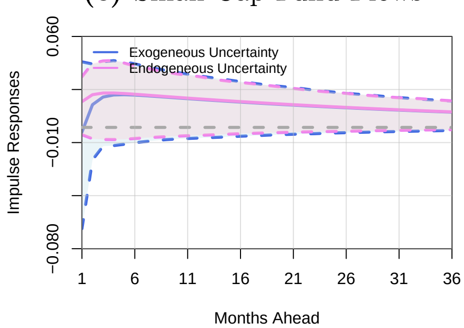
(b) Sector Fund Flows



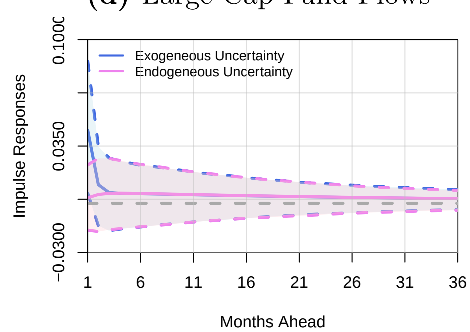
(c) Small-Cap Fund Flows



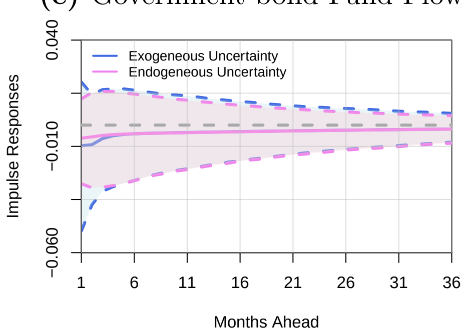
(d) Large-Cap Fund Flows



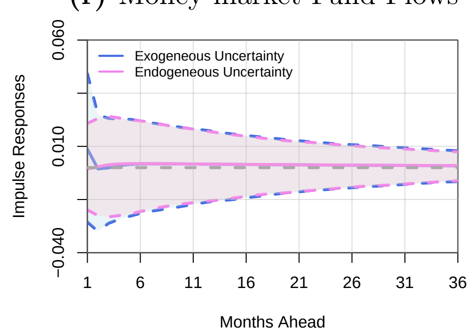
(e) Government-bond Fund Flows



(f) Money-market Fund Flows



(g) Corporate-bond Fund Flows



(h) Municipal-bond Fund Flows

Figure IA.8: Impulse Responses of Fund Flows with Different Investment Objective Codes to Financial Uncertainty Shocks

This figure shows the dynamic impulse response functions (IRFs) of fund flows to financial uncertainty shocks in VAR(1). Other details can be found in the footnote of Figure 5.