Model Uncertainty in the Cross Section *

Jiantao Huang† Ran Shi‡

May 2023

Abstract

We develop a transparent Bayesian framework to quantify uncertainty in linear stochastic discount factor (SDF) models. In our framework, the posterior probability of a model increases with its associated in-sample Sharpe ratio and decreases with the model dimension. The entropy of model probabilities represents model uncertainty. We provide theoretical guarantees to ensure consistent interpretation of our model uncertainty measure, even for misspecified models with omitted factors. Empirically, surging model uncertainty coincides with major market events. Combining SDF models improves mean-variance efficiency only during periods of high model uncertainty. Positive shocks to model uncertainty predict persistent outflows from US equity funds and inflows to Treasury funds.

Keywords: Model Uncertainty, Linear Stochastic Discount Factor, Bayesian Inference
JEL Classification Codes: C11, G11, G12.

* For helpful comments and discussions, we thank Svetlana Bryzgalova, Thummim Cho, Guanhao Feng, 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, and NBER-NSF SBIES 2022. Any errors or omissions are the responsibility of the authors.

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

Introduction

Empirical finance research has proposed various factor models for the cross-section of expected asset returns. Which model(s) should academics and practitioners use? For a specific model, one can test whether the model is rejected in the data. From a spectrum of candidates, dominant models can be selected based on specific criteria. However, it is unclear to what extent decision-makers can rely on their model testing and selection procedures. For instance, should we take a factor model as the ground truth if the data fail to reject it? Should we ignore models that are overshadowed by others? Should we treat factors excluded from our favorite models as useless? In other words, all model testing and selection processes involve certain degrees of uncertainty, and linear factor models in asset pricing are no exception.

In this paper, we develop a transparent Bayesian framework to define and quantify model uncertainty about linear stochastic discount factors (SDFs). Our framework yields closed-form posterior model probabilities, and we use the (information) entropy of these probabilities to measure model uncertainty. Accompanying theoretical support is provided to help interpret the model uncertainty measure, even under model misspecification. 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, picking a model boils down to an exercise of tossing a fair coin. Model uncertainty is the highest in this case, and correspondingly, the entropy reaches its maximum.

In the data, we document sizable model uncertainty in the US equity market, which also exhibits pronounced time-series variation. In particular, heightened model uncertainty coincides with major market events, touching its upper limit before the technology bubble

and during the early onset of the global financial crisis. In these periods, investors will find it challenging to navigate the equity market as all models seem equally "wrong." However, we find that the "average" SDF across models can still deliver significantly high out-of-sample Sharpe ratios: All models are also equally useful in turbulent markets. Departing from normative assessments of mean-variance efficiency, to a descriptive end, we find that model uncertainty is closely related to investors' asset allocation decisions on mutual fund products. Shocks that escalate model uncertainty predict persistent outflows from US equity mutual funds and inflows to Treasury funds.

Our econometric approach to quantifying model uncertainty takes the perspective of a Bayesian investor. Investors are not clairvoyant as they do not know the "true" SDF. Instead, they learn *both* model parameters (i.e., market prices of risk) and model specifications (i.e., factors that enter the SDF) through Bayesian updating.

The investor assigns an informative prior to the factors' market prices of risk. For factors excluded from the SDF, their risk prices are strictly zero. For factors entering the SDF, the baseline specification assumes that their risk prices are drawn from Zellner's g-priors (Zellner, 1986), which are "benchmark" priors for Bayesian inference in the context of model uncertainty (Fernández, Ley, and Steel, 2001).[1] In our setting, however, we show that g-priors suffer from model selection inconsistency. Even after observing a long history of return data generated from a fixed true model (*ex ante* the lowest uncertainty of being zero), the investor always fails to assign a high probability to this model *ex post*. Model uncertainty (defined as the entropy of model probabilities) will be artificially inflated above zero under the g-prior setting.

To address the abovementioned issue, we adapt the mixture of g-priors proposed by

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

Liang, Paulo, Molina, Clyde, and Berger (2008) to the linear SDF framework. Perhaps the most appealing feature of this adaptation is that model uncertainty becomes consistently interpretable, even under model misspecification due to omitted factors. When there are no omitted factors, the modified prior restores model selection consistency. In theory, the Bayesian investor can assign a probability of one to the true model and the resulting model uncertainty measure converges to zero (as it is supposed to because a constantly true model exists). When the investor ignores certain factors in the true SDF, our model uncertainty measure will be bounded far from its maximum with high probability. These properties imply that if the investor observes high degrees of model uncertainty, even calculated from a fairly limited set of factors, she should not treat her favorite model as the ground truth.

The posterior model probabilities derived under the mixture of g-priors have intuitive closed-form expressions, and they increase with model-implied maximal in-sample Sharpe ratios and decrease with model dimensions. This result crystallizes two contradicting criteria for an asset pricing model to be chosen by our Bayesian investor: higher mean-variance efficiency (in her backtesting exercises) and model simplicity.

Applying our framework to equity return data, we document four empirical findings. First, model uncertainty in the US equity market, as shown in Figure 1, displays considerable time-series variation and surges before major recession periods. Mounting model uncertainty appears before the early 1980s and 1990s recessions. It reaches the upper bound during the technology bubble and the global financial crisis. In these tumultuous markets, investors have to brace for not only conventional sources of risk (e.g., volatility, comovement, and skewness) but also the difficulties of finding a reliable model they can use to navigate out of the storm. We repeat the exercise for the European and Asia-Pacific stock markets and observe similar patterns. Interestingly, model uncertainty in the Asia-Pacific market is exceptionally high when the 1997 financial crisis swept through this region.

Second, in the US data, no individual factor demonstrates consistently high probabilities of entering the SDF (namely, *factor probabilities*). In general, factor probabilities tend

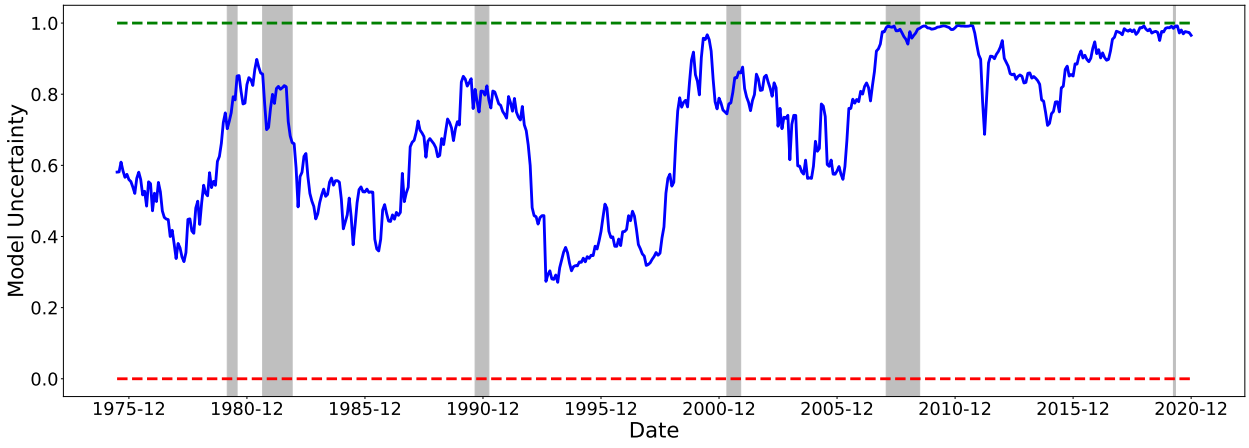


Figure 1: The Time-Series of Model Uncertainty in the US Equity Market

This figure presents the time-series of uncertainty about linear stochastic discount factor (SDF) models in the US equity market. The red and green dashed lines are the lower and upper bounds of model uncertainty. Shaded areas mark the NBER recession periods.

to decline during economic downturns. In normal times, the factor probabilities of the market, value, profitability, and beta anomaly (following the betting-against-beta (BAB) construction of Frazzini and Pedersen (2014)) are almost one before 2008. After the global financial crisis, only the market and beta anomaly prevail. On average, the probability of finding factors that enter the SDF has declined in recent years, especially under the high model uncertainty regime. Surprisingly, high model uncertainty does not preclude the existence of useful factors. For example, right before 2000, BAB and the post–earnings-announcement-drift (PEAD) factors both have probabilities of over 90% to enter the SDF, while model uncertainty culminates.

Third, when model uncertainty is high, mean-variance investors can significantly benefit from averaging different SDF models. The portfolio implications of aggregating models using Bayesian methods have been examined in the literature (see Avramov and Zhou (2010) for a comprehensive review). We revisit the question under our setting. 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

dividing 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. Ignoring the weakly dominated models can be particularly detrimental to mean-variance-efficient portfolio choices.

Fourth, shocks to model uncertainty predict aggregate flows to domestic equity and Treasury funds, which we treat as proxies for investors' asset allocation decisions among different classes of fund products. Uncertainty in general has ambiguous implications for investors' asset holdings. The conventional wisdom of "flight-to-quality" predicts that investors curtail risk exposures or hoard liquid assets in response to positive uncertainty shocks.[2] On the contrary, instead of seeking safety or liquidity, investors may chase risky assets such as glamour stocks during certain high-uncertainty periods, in search of a new El Dorado.[3]

We estimate the dynamic responses of fund flows to model uncertainty shocks using the vector autoregression (VAR) models. Throughout our VAR exercises, we investigate two possibilities: Model uncertainty can be an exogenous "cause" or merely a propagation channel. Under both settings, positive model uncertainty shocks forecast sharp outflows from US equity funds and inflows to Treasury funds, with effects persisting for approximately three years. Domestic equity investors withdraw only from small-cap and style funds (funds that commit to investment styles such as growth or value), not large-cap or sector funds (funds dedicated to certain industries). These findings are more consistent with the "flight-to-quality" prediction. Facing high model uncertainty in the equity market, investors tend to reduce risky asset positions (by withdrawing from small-cap and style funds) and invest in safe assets such as government bonds.

[2]Models featuring these effects are based on institutional redemption pressures (Vayanos, 2004), preferences featuring "robustness" concerns (Caballero and Krishnamurthy, 2008), and asymmetric information (Guerrieri and Shimer, 2014).

[3]Exemplary events include the "railway mania" in the mid 1840s and the "technology bubble" in the late 1990s. Such phenomena have been explained in the literature on learning and growth options (for example, Abel (1983); Pástor and Veronesi (2006, 2009)).

Literature. Our paper primarily contributes to the literature on Bayesian inferences about asset pricing models and Bayesian portfolio choices. The Bayesian approach to testing factor models dates back to Shanken (1987) and Harvey and Zhou (1990), under the null hypothesis of absolute zero *pricing errors*, namely *alphas*. Relaxing the zero-alpha dogma, Pástor (2000) and Pástor and Stambaugh (2000) investigate the portfolio implications of different factor models, when investors update beliefs about model alphas from Gaussian priors. Barillas and Shanken (2018) extend this prior setting and derive closed-form criteria to compare asset pricing models. Chib et al. (2020) modify the prior specification of Barillas and Shanken (2018) to resolve posterior indeterminacy issues in model comparison. Like our paper, Avramov, Cheng, Metzker, and Voigt (2023) also define and analyze uncertainty about factor models. They work under a conditional beta-pricing setup instead of the SDF framework that we focus on. Their paper studies the contribution of model uncertainty to volatility and return comovement estimates, while we examine uncertainty about which factors should enter the SDF. In addition, our model uncertainty measure is anchored within lower and upper bounds (making it easier to interpret), accompanied with theoretical guarantees.

Our Bayesian method makes inferences about factors' risk prices (which determine whether certain factors enter the SDF), instead of risk premia (which determine whether factors carry priced risk). To this end, our paper is related to recent methodological advancements under the SDF framework such as Kozak, Nagel, and Santosh (2020), Chib and Zeng (2020), Chib, Zhao, and Zhou (2023), and Bryzgalova, Huang, and Julliard (2023). Our key contribution along this line is proposing a new prior to facilitate Bayesian inferences about SDFs under model uncertainty. We conduct thorough econometric analysis and simulation studies to highlight favorable properties, such as closed-form posteriors, invariance to rotated testing assets and model selection consistency, of our proposed method.

There is increasing interest in developing time-series of uncertainty measures for both real and financial activities (Bloom, 2009; Jurado, Ludvigson, and Ng, 2015; Baker, Bloom,

and Davis, 2016; Manela and Moreira, 2017; Ludvigson, Ma, and Ng, 2021). On the cross-sectional dimension, Dew-Becker and Giglio (2021), borrowing information across firms, present an uncertainty measure from option prices. Hassan, Hollander, Van Lent, and Tahoun (2019) create a panel of firm-specific political uncertainty measures with detailed topic attributions. Compared with these papers, our model uncertainty measure is conceptually new. It quantifies equity investors' uncertainty about which asset pricing models better describe the cross-section of expected returns.

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

1 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.[4] 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):

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

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

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

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

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

1.1 A Framework Incorporating Model Uncertainty

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

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

and the expected excess returns are such that

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

The two equations above are counterparts of (1) and (2) after incorporating model uncertainty. We define $p_{\boldsymbol{\gamma}} = \sum_{j=1}^{p} I_{\{\gamma_j = 1\}}$, the number of factors that are included in model $\mathcal{M}_{\boldsymbol{\gamma}}$. $\boldsymbol{f}_{\boldsymbol{\gamma}}$ is a $p_{\boldsymbol{\gamma}}$-dimensional vector concatenating all factors that are included under $\mathcal{M}_{\boldsymbol{\gamma}}$; elements of $\boldsymbol{b}_{\boldsymbol{\gamma}}$ are market prices of risk for the incorporated factors. Under our notation, factors that have zero risk prices are excluded, and all elements in $\boldsymbol{b}_{\boldsymbol{\gamma}}$ are nonzero. As in Cochrane (2005, Page 261), the market prices of risk address the question of "should I include factor j given the other factors?" If $b_j = 0$, the answer would be "no", which maps directly

into our model uncertainty framework.

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

1.2 Prior Specification and Bayesian Inference

We now present a Bayesian framework to understand and quantify model uncertainty in the cross-section of expected stock returns, under the linear SDF setting. We observe excess return data $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^T$ and make inference about the posterior probability of each model $\mathcal{M}_{\boldsymbol{\gamma}}$ *conditional* on the data, denoted as $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$. Bayesian inference offers a natural way of computing these posterior model probabilities.

Based on Equation (4), we assume that the observed excess returns are generated from the following model:

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

where $\boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0},\ \boldsymbol{\Sigma})$; $\boldsymbol{C}_{\boldsymbol{\gamma}} = \text{cov}[\boldsymbol{R}_t, \boldsymbol{f}_{\boldsymbol{\gamma},t}]$ is a principal submatrix of $\boldsymbol{\Sigma}$ (since $\boldsymbol{f}_{\boldsymbol{\gamma}} \subseteq \boldsymbol{R}$, the factors are a subset of test assets). For inference about model configurations $\mathcal{M}_{\boldsymbol{\gamma}}$ (and the associated market prices of risk $\boldsymbol{b}_{\boldsymbol{\gamma}}$), we adopt an empirical Bayes method. Specifically, we only assign priors to $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and treat the variance-covariance matrix $\boldsymbol{\Sigma}$ as known when deriving the posterior probability of a model $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$. Then, we replace $\boldsymbol{\Sigma}$ with its consistent estimator.[5] Section 1.2.2 provides additional details of this substitution and its associated econometric properties.

Now we introduce and discuss our baseline prior specification for the risk prices. Moti-

[5]Empirical Bayes approaches use data to facilitate prior assignments. 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).

vated by Zellner's g-priors (Zellner, 1986), we assume that *conditional* on model $\mathcal{M}_{\boldsymbol{\gamma}}$,

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

where T is the sample size of the observed excess returns.

The parameter g in (6) controls the level of uncertainty about an "imaginary sample" (relative to the observed return data) in the language of Zellner (1986). Before making inference about different linear SDF models based on the observed return data $\mathcal{D}$, we consider an "imaginary" sample of size T', denoted by $\mathcal{D}' = \{\boldsymbol{R}_t'\}_{t=1}^{T'}$. The sample size is allowed to be different from T by a scalar g such that $T' = T/g$.[6] Under model $\mathcal{M}_{\boldsymbol{\gamma}}$, excess returns observed in this sample are distributed as follows: $\boldsymbol{R}_1', \ldots, \boldsymbol{R}_{T'}' \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{C}_{\boldsymbol{\gamma}}\boldsymbol{b}_{\boldsymbol{\gamma}},\, \boldsymbol{\Sigma})$. Assigning a noninformative prior on $\boldsymbol{b}_{\boldsymbol{\gamma}}$, which is a constant everywhere, we can derive the "imaginary posterior" of $\boldsymbol{b}_{\boldsymbol{\gamma}}$ under this conceptual data sample as $[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}'] \sim \mathcal{N}\left(\boldsymbol{b}_{\boldsymbol{\gamma}}',\, g/T \times \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right)$. Setting $\boldsymbol{b}_{\boldsymbol{\gamma}}' = \mathbf{0}$ as suggested by Zellner (1986), we have the prior specification in (6).

One might attempt to assign an objective or uninformative prior, such as the Jeffreys prior, to $\boldsymbol{b}_{\boldsymbol{\gamma}}$. However, uninformative priors can only be assigned to *common* parameters across models, which is clearly not the case for the model-specific risk prices $\boldsymbol{b}_{\boldsymbol{\gamma}}$. Otherwise, posterior probabilities can be indeterminate. This has also been noted in the finance literature (Cremers, 2002; Chib et al., 2020).

Our prior specification for risk prices enjoys three favorable properties. First, it leads to analytically tractable posteriors $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$. Second, asymptotic analysis confirms that these posterior distributions consistently select factors (and models after modifications, detailed in Section 1.2.4). Third, unlike Gaussian or double-Laplacian priors (i.e., Bayesian versions of ridge and LASSO regressions), our prior is "rotation-invariant:" Inference about model

[6] In Zellner (1986), the scalar g is used to capture the fact that the variance of the hypothetical sample can be different from the variance of the sample under study. These two arguments (effective sample size vs. variance of the hypothetical dataset) lead to the same g-prior specification. Our sample-size-based arguments echo the ideas of factional or intrinsic Bayes factors (O'Hagan, 1995; Berger and Pericchi, 1996), which are intended to "transform" improper into proper priors. Similar ideas for prior specifications are adopted by Kandel and Stambaugh (1996) in the return predictability literature.

uncertainty will not be affected by standard procedures such as principal component analysis applied to the testing assets.

1.2.1 Posterior inference: Analytical marginal likelihood and Bayes factors

We can integrate out the risk prices $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and calculate the (marginal) likelihood of observing the excess return data $\mathcal{D}$ based on each asset pricing model $\mathcal{M}_{\boldsymbol{\gamma}}$. The following proposition summarizes our analytical results.

Proposition 1. *(analytical marginal likelihood functions under g-priors) The likelihood of observing excess return data $\mathcal{D}$ under model $\mathcal{M}_{\boldsymbol{\gamma}}$ is*

$$\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] = \exp\left\{-\frac{T}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) - \frac{T}{2}\left(\mathrm{SR}^2_{\max} - \frac{g}{1+g}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right)\right\}\frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}}, \tag{7}$$

where $\boldsymbol{S} = 1/T\sum_{t=1}^{T}(\boldsymbol{R}_t - \overline{\boldsymbol{R}})(\boldsymbol{R}_t - \overline{\boldsymbol{R}})^\top$ and $\overline{\boldsymbol{R}} = 1/T\sum_{t=1}^{T}\boldsymbol{R}_t$, both of which are the method of moment estimators. $\mathrm{SR}^2_{\max}$ is the maximal squared Sharpe ratio achievable from forming portfolios using all testing assets $\boldsymbol{R}$; $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ is the maximal squared Sharpe ratio from combining all factors $\boldsymbol{f}_{\boldsymbol{\gamma}}$ under model $\mathcal{M}_{\boldsymbol{\gamma}}$. These two Sharpe ratios are both ex post in-sample values and $\mathrm{SR}^2_{\boldsymbol{\gamma}} \leq \mathrm{SR}^2_{\max}$ for all $\boldsymbol{\gamma}$.[7]

The proof of this proposition, as well as all subsequent propositions are relegated to Appendix D.

Based on this proposition, we can calculate the Bayes factor of two linear factor models, namely $\mathcal{M}_{\boldsymbol{\gamma}}$ and $\mathcal{M}_{\boldsymbol{\gamma}'}$, as

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

A large Bayes factor $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ favors model $\mathcal{M}_{\boldsymbol{\gamma}}$ over model $\mathcal{M}_{\boldsymbol{\gamma}'}$. A special case is comparing $\mathcal{M}_{\boldsymbol{\gamma}}$ against the null model $\mathcal{M}_{\mathbf{0}}$, under which risk prices are all zeros, and the SDF is a

[7] Throughout the rest of the paper, we may use "Sharpe ratios" to refer these maximal in-sample squared Sharpe ratios for ease of exposition.

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

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

Although the marginal likelihood in Proposition 1 depends on test assets through $\mathrm{SR}_{\max}^2$, the Bayes factors do not. They are determined solely by the set of factors entering our asset pricing models. This observation is driven by the condition that factors must be a subset of test assets. In other words, factors that define the SDF must correctly price themselves. This phenomenon is reminiscent of the observation that the efficient GMM objective function for estimating factor models assigns zero weights to test assets except for the factors entering the SDF (see, for example, Cochrane (2005, Page 244-245)).

Models favored by the Bayes factors should feature large Sharpe ratios $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ and small dimensions $p_{\boldsymbol{\gamma}}$. In other words, Bayes factors under our framework favor parsimonious factor models generating large in-sample squared Sharpe ratios.

When model dimensions are fixed, factor models associated with large *ex post* Sharpe ratios are always preferred. This echoes the intuitions behind the GRS tests (see, for example, Gibbons et al. (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.

The penalty term on the model dimension appears under model uncertainty. It is important because models associated with large $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ can be artificially dense: Adding additional assets into one's portfolio always leads to higher achievable Sharpe ratios. Proposition 2 formally demonstrates the econometric property of $\mathrm{SR}_{\boldsymbol{\gamma}}^2$. It also shows that when the return sample is large, the improvement of Sharpe ratios by including additional assets diminishes.

Proposition 2. *(the expectations of the Sharpe ratios) If there exists a true linear factor SDF model* $\mathcal{M}_{\boldsymbol{\gamma}_0} : m_{\boldsymbol{\gamma}_0} = 1 - (\boldsymbol{f}_{\boldsymbol{\gamma}_0} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}_0}])^\top \boldsymbol{b}_{\boldsymbol{\gamma}_0}$ *generating the observed return data, that is,* $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T \overset{\text{iid}}{\sim} \mathcal{N}(\mathrm{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}_0}]\boldsymbol{b}_{\boldsymbol{\gamma}_0}, \boldsymbol{\Sigma})$*, then the maximal in-sample Sharpe ratio of* $\boldsymbol{f}_{\boldsymbol{\gamma}}$*,*

namely factors under consideration for model $\mathcal{M}_{\boldsymbol{\gamma}}$, satisfies

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

According to Proposition 2, 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 spanning the true set of factors ($\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]$ is small). If $\mathcal{M}_{\boldsymbol{\gamma}}$ includes all true factors in $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$, the conditional variance $\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]$ is zero. Under this scenario, $\mathbb{E}\left[\mathrm{SR}_{\boldsymbol{\gamma}}^2\right] = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \left(\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0}\right]\right) \boldsymbol{b}_{\boldsymbol{\gamma}_0} + p_{\boldsymbol{\gamma}}/T \geq \mathbb{E}\left[\mathrm{SR}_{\boldsymbol{\gamma}_0}^2\right]$ (for $p_{\boldsymbol{\gamma}} \geq p_{\boldsymbol{\gamma}_0}$). Without the penalty on model sizes, model comparison based on Sharpe ratios will artificially favor dense models.

1.2.2 Empirical Bayes: the plug-in covariance matrix estimator

Thus far, we have treated the covariance matrix $\boldsymbol{\Sigma}$ of asset returns as known. When calculating Bayes factors defined in Equation (8) and (9), we adopt an empirical Bayes method and replace $\boldsymbol{\Sigma}$ with its method of moment estimator (namely $\boldsymbol{S}$ in Proposition 1). 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 $\boldsymbol{\Sigma}$.[8] Proposition 3 summarizes properties of $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$.

Proposition 3. *(the econometric properties of the Sharpe ratios under empirical Bayes) When the covariance matrix of returns $\boldsymbol{\Sigma}$ is replaced by its method of moments estimator, the corresponding maximal squared in-sample Sharpe ratio under model $\mathcal{M}_{\boldsymbol{\gamma}}$, denoted by $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$, satisfies*

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

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

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

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

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

Based on Proposition 3, the expected value of $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ converges to that of $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ (presented in Proposition 2), as the sample size T becomes large (and the model dimension does not scale with T). According to the second part of Proposition 3, the distribution of $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ concentrates heavily around that of $\mathrm{SR}_{\boldsymbol{\gamma}}^2$. A byproduct of this result is that $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ must be a consistent estimator. Without further clarification, $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ will be replaced with $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ throughout our empirical studies; all theoretical results involving $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ will be derived accounting for this replacement (i.e., we only present sample-based instead of population-based results.)

1.2.3 Posterior model probabilities 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

$$\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] = \frac{\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] \times \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}}]}{\sum_{\boldsymbol{\gamma}'} \mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}'}] \times \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}'}]} = \frac{\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})}{\sum_{\boldsymbol{\gamma}'} \mathrm{BF}(\boldsymbol{\gamma}', \mathbf{0})}, \tag{10}$$

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

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

$$\mathcal{E} = -\frac{1}{p \log 2} \sum_{\boldsymbol{\gamma}} \left(\log \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]\right) \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]. \tag{11}$$

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 (with the others' being

zero). When $\mathcal{E} = 1$, all models have the same posterior probability: they must be equally right or equally wrong.

Under the g-prior specification, we have the following asymptotic results for posterior probabilities (in light of a frequentist setup under which a fixed true model and set of parameters exist), summarized in Proposition 4.

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

1. *(factor selection consistency) if* $\gamma_{0,j} = 1$*, i.e., the true model includes factor* j*, the posterior marginal probability of choosing factor* j *converges to one in probability:* $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] = \sum_{\{\boldsymbol{\gamma}:\ \gamma_j = 1\}} \mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 1$;[9]

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.*

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 $\gamma_{0,j} = 0$ (which is a byproduct from our proof of Proposition 4 and validated through simulation studies tabulated in Table A3 and Table A4 of the Appendix). Combining this posterior behavior with the factor selection consistency result, we conclude that the g-priors can avoid discarding true factors, at the cost of incorporating redundant factors.

Proposition 4 highlights the limitation of assigning g-priors to the market prices of risk. Even if the observed return data are generated from a fixed true model, an econometrician

[9] 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."

can never identify this model with high probability, regardless of how much return data have been accumulated. As a result, the model uncertainty measure defined in (11) is biased. Although there is no model uncertainty in the data-generating process, the unfavorable property of the g-priors still artificially inflates $\mathcal{E}$ above zero.

1.2.4 Restoring model selection consistency: the mixture of g-priors

Model selection inconsistency under the g-priors will lead to biased model uncertainty measures. To resolve this issue, we adapt the mixture of g-priors specification proposed in Liang et al. (2008) to further assume $g \sim \pi(g)$ (a hyper-prior distribution for the g parameter) where

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

Under model $\mathcal{M}_{\boldsymbol{\gamma}}$, the risk prices now follow a scale mixture of normal distributions:

$$\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}} \sim \frac{1}{G} \int_0^\infty \mathcal{N}\left(\mathbf{0}, \frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right) \pi(g)\, \mathrm{d}g, \tag{12}$$

where G is a normalizing constant. Under the mixture of g-priors, we have the following analytical result for the Bayes factor, $\mathrm{BF}(\boldsymbol{\gamma}, \mathbf{0})$, summarized in Proposition 5.

Proposition 5. *The Bayes factor for comparing model $\mathcal{M}_{\boldsymbol{\gamma}}$ with the null model $\mathcal{M}_{\mathbf{0}}$ under the mixture of g-priors is:*

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

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

Plugging the expressions for $\mathrm{BF}(\boldsymbol{\gamma},\, \mathbf{0})$ in Proposition 5 back into Equations (10) and (11), we can calculate the posterior model probabilities and model uncertainty measures under the new specification.

The key benefit of using the mixture of g-priors is to achieve posterior model selection consistency, as summarized in Proposition 6.

Proposition 6. *(posterior property: 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 6, we perform simulation studies examining the finite-sample behavior of our methods. Appendix Table A3 tabulates the results. Both g-priors and the mixture of g-priors consistently select true factors, even based on moderately sized return samples (e.g., three years of daily observations). However, inferences based on g-priors are plagued by redundant factors. The probability of mistakenly including factors in the SDF remains positive and largely unchanged as the sample size increases. In contrast, the same probability shrinks toward zero in larger samples under the mixture of g-priors. This family of newly proposed priors, by guarding against redundant factors, assigns a posterior probability of approximately one to the true model.

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

1.2.5 A misspecified set of factors

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

because it omits factors in the true SDF.

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

Even when there are true factors that are omitted from $\boldsymbol{f}$, the mixture of g-priors still maintains (partial) factor selection consistency: True factors that are included in $\boldsymbol{f}$ can always be selected according to their marginal posterior probabilities. Proposition 7 describes this property.

Proposition 7. *(posterior property with omitted factors) Assume that the observed return data are generated from a true linear SDF* $m_0 = 1 - (\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}_0$. *Let* $\boldsymbol{f}_{\boldsymbol{\gamma}_0} = \boldsymbol{f}_0 \cap \boldsymbol{f}$*; that is,* $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ *is the subset of* $\boldsymbol{f}$ *(factors under consideration) that includes the largest number of factors in the true model with no redundancy. As* $T \to \infty$*, for all* j *such that* $\gamma_{0,j} = 1$*,* $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] \xrightarrow{p} 1$.

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

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

Proposition 8. *(model uncertainty under misspecification) Under the assumption of Proposition 7, 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 8 indicates that if we observe very high model uncertainty $\mathcal{E}$, say $\mathcal{E} \approx 1$, only two possibilities exist: (1) $p_{\gamma_0} = 0$, which implies that the set of factors we are working with does not cover *any* of the true factors; (2) the assumptions in Propositions 7 and 8 are tenuous – the observed returns are not generated from a unique linear SDF model. By investigating factor probabilities $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$, we can rule out possibility (1) under the presence of "strong" factors ($\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}] \approx 1$). Under the second possibility, the quest for one dominant linear factor model seems empirically quixotic.

The upper bound for model uncertainty in Proposition 8 tends to be loose. The proof of Proposition 8 in the Appendix mandates that the upper limit is binding only when all models subsuming factors in $\boldsymbol{f}_{\gamma_0}$ (true factors that are *not* omitted) have exactly the same posterior probability. Simulation results under misspecified $\boldsymbol{f}$ (omitting the momentum factor) reported in Table A5 illustrate that the gap between $\mathcal{E}$ and $(p - p_{\gamma_0})/p$ becomes substantially larger as p increases. Table A6 further demonstrates that the gap is significant if other factors are ignored when calculating the model uncertainty measure.

2 Data

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

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

Finally, we obtain mutual fund data from the Center for Research in Security Prices (CRSP) survivorship-bias-free mutual fund database. In particular, we are interested in monthly mutual fund flows, so we download the monthly total net assets, monthly fund returns, and the codes of fund investment objectives. In addition, we download the total market value of all US-listed stocks from CRSP.

3 Model Uncertainty in the US Equity Market

We construct a monthly time-series of model uncertainty based on the proposed framework from June 1975 to December 2020.[10] At the end of each month, we use daily asset returns in the past three years to compute the posterior model probabilities $\mathbb{P}[\mathcal{M}_\gamma \mid \mathcal{D}]$ and the model uncertainty measure based on Equations (10) and (11).[11] Bayes factors appearing in these

[10] The behavioral factors in Daniel et al. (2020) are available only from July 1972, and we use 36-month data in the estimation, so the model uncertainty measure starts from June 1975.

[11] We also repeat our calculation with four-year and five-year rolling windows. Figure A1 in the Appendix confirms that our cross-sectional model uncertainty measure exhibits similar time-series patterns under alternative specifications of window widths.

equations are calculated under the mixture of g-priors (as in Proposition 5).

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

Figure 1 plots the time-series of our linear factor model uncertainty measure. The red and green dashed lines are the lower and upper bounds of this measure. The lower bound is zero, i.e., when there exists a true model. The upper bond is one, which is achieved when the posterior model probabilities are all equal.

The model uncertainty index has several interesting features that could shed light on the nature of uncertainty in the cross-section of expected returns. First, we observe a surprisingly high level of model uncertainty. The average (median) model uncertainty is 0.70 (0.75), with the first and third quartiles equal to 0.53 and 0.87, respectively. Second, model uncertainty in the cross-section is a dynamic phenomenon: It fluctuates significantly over time. In particular, the index ranges from 0.27 to 0.99, with a standard deviation of 0.21.

According to Figure 1, our model uncertainty measure is countercyclical. In particular, the 1990s, often remembered as a period of strong economic conditions and high stock returns, exhibits 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 those of the others. Hence, it is relatively straightforward to identify the true SDF model in this period.

In addition, episodes of heightened model uncertainty tend to coincide with economic downturns and stock market crashes. For example, our model uncertainty measure touches the upper bound during the global financial crisis. The blue dots in Figure 2, showing the

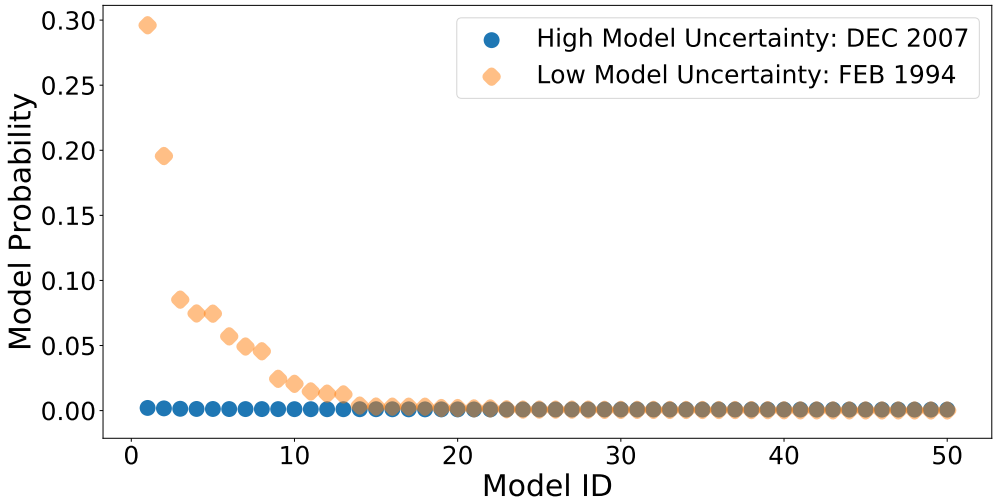


Figure 2: Posterior Probabilities of Top 50 models: High vs. Low Model Uncertainty

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

posterior probabilities of the top 50 models in December 2007, lie on a horizontal line – the posterior probabilities of the models are equalized. It is virtually infeasible to distinguish models based on the observed data. The 2008 crisis is also noteworthy because model uncertainty remains at a high level for a prolonged period, until recently. In the last five years, model uncertainty has slowly increased from 0.70 to 0.99 at the end of 2020. In contrast, it declines shortly after other crisis periods.[12]

3.1 Model Uncertainty and Maximal In-Sample Sharpe Ratios

Proposition 5 shows that posterior model probabilities increase in the squared Sharpe ratios SR_γ^2 and decrease in model dimensions p_γ. Thus, model uncertainty, as the entropy of these posterior distributions, is low only when a small number of (parsimonious) models carry

[12]We also measure model uncertainty in the dataset with two mispricing factors. The orange dashed line in Figure A2 shows the corresponding time-series, and we confirm that including mispricing factors leaves the main empirical patterns nearly unchanged.

sizable SR^2_γ. We start this section by exploring the dispersion of SR^2_γ among factor models.

In Figure A3, we plot the difference between the maximum (delivered by the full model) and $90th$-quantile/median of SR^2_γ. The dispersion of SR^2_γ among models decreases sharply before economic downturns and remains low during the bear markets. For example, the distance between the highest and medium squared Sharpe ratios is over five times larger in 1997 than in 2000. After 2000, factor models are similar in terms of SR^2_γ: Including more factors cannot meaningfully improve mean-variance efficiency. In general, small dispersion of Sharpe ratios is marginally linked to high model uncertainty.

Since SR^2_γ is determined by the mean, volatility, and correlation structure of factor returns, we compare these three simple statistics with our model uncertainty measure in Figures A4 and 3. Figures A4a and A4b show that model uncertainty is clearly different from the average and volatility of factor returns. Interestingly, from 2010 to 2020, factors yield lower average returns (compared with their historical benchmark); the same period also exhibits high model uncertainty. Figure 3 plots the average pairwise correlations among factors under consideration. The factor correlations demonstrate a certain degree of association with model uncertainty. In theory, when factors are more correlated on average (smaller benefits from diversification), the dispersion among different models' SR^2_γs is likely to decrease (higher model uncertainty). Overall, none of these three statistics can entirely convey the exacts information in model uncertainty.

3.2 Model Uncertainty and Other Economic Variables

We explore whether our model uncertainty measure is related to other uncertainty indices in recent literature, including the VIX, economic policy uncertainty (EPU) in Baker et al. (2016), and three uncertainty measures from Jurado et al. (2015) and Ludvigson et al. (2021). Table 1 reports the results from regressing our model uncertainty measure $\mathcal{E}_t$ on these indices after controlling for lagged terms. The regression coefficients describe contemporaneous

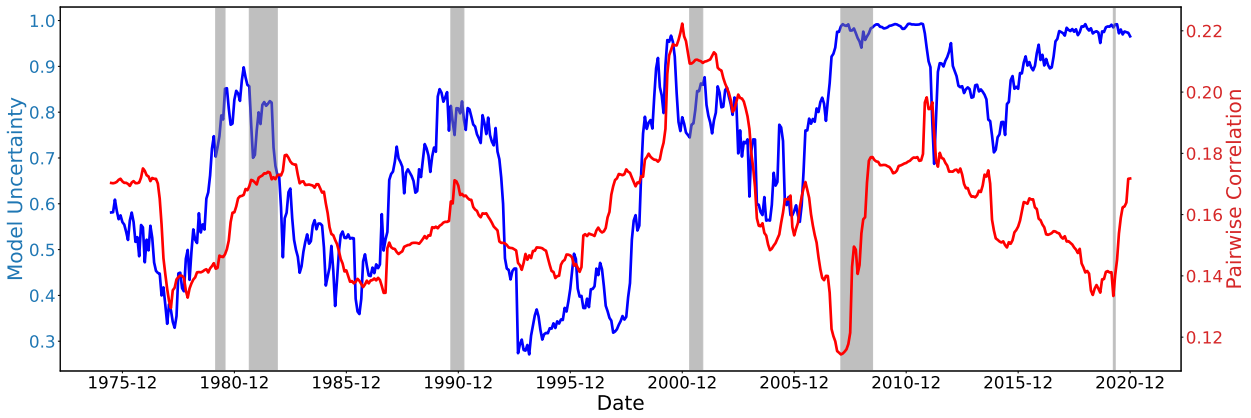


Figure 3: Time-Series of Average Pairwise Correlation of 14 Factors

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

associations, with no intention to study causal relationships. Our model uncertainty measure is only significantly associated with the financial uncertainty index of Ludvigson et al. (2021) and the VIX, both of which also rely heavily on asset prices.

We further compare model uncertainty with financial variables that are known to be related to aggregate fluctuations, including the term spread (the difference between 10-year and 3-month Treasury yields), and the credit spread (the yield difference between BAA and AAA bonds). The term spread is negatively associated with model uncertainty.

Remark. One favorable property of our measure that is missing from other uncertainty indices in the literature is that ours is always bounded. When model uncertainty hits the upper bound, we can conclude that a "true" model driving the cross-section of expected returns is elusive, under the theoretical guidance of Proposition 8.

3.3 Factor Uncertainty and Model Uncertainty

Our model uncertainty measure quantifies the difficulties of choosing asset pricing models. When the measure indicates that no linear SDF clearly dominates the others, one might postulate that there must be high uncertainty about which factors should enter the SDF, namely,

Table 1: Regressions of Model Uncertainty on Contemporaneous Variables

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

The table reports results from the following regression:

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

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

factor uncertainty. To quantify factor uncertainty, we compute the posterior (marginal) probability of selecting each individual factor according to definitions in Proposition 6 and plot the time-series in Figure A5.

According to Figure A5, posterior marginal probabilities of factor selection demonstrate pronounced time-series variation. During economic downturns, factors' chances of being selected tend to decline. The market, size, value, profitability, betting-against-beta (BAB), and short-term behavioral factors (PEAD) all undergo extended periods of being selected with probability one. Before 2000, only the market, value and profitability factors show consistently high probabilities of entering the SDF. After 2000, only the market and betting-against-beta factors pass the same scrutiny.

A surprising finding from the plots in Figure A5 is that high model uncertainty does not preclude the existence of factors that have high probabilities of entering the SDF (low factor uncertainty). For example, our model uncertainty measure almost reaches its maximum in late 1999; moreover, 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 8 calls for devoting special attention to the phenomenon above. It is

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

However, since 2008, especially during the long periods of high model uncertainty after the global financial crisis, all factors' odds of entering the SDF are no better than flipping a fair coin: Factor uncertainty and model uncertainty are both high in these periods.

4 Portfolio Choice: Incorporating Model Uncertainty

How should investors take into account 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 ($\boldsymbol{b}_{-\boldsymbol{\gamma}}=\mathbf{0}$) and induce different posteriors for $\boldsymbol{b}_{\boldsymbol{\gamma}}$. Our proposed BMA estimator of $\boldsymbol{b}$ is then

$$\boldsymbol{b}_{bma} := \mathbb{E}[\boldsymbol{b} \mid \mathcal{D}] = \sum_{\boldsymbol{\gamma}} \mathbb{E}[\boldsymbol{b} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}] \times \mathbb{P}(\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}). \tag{13}$$

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, deviance, and information criteria) to select a single model and presume that the selected model is correct.

We evaluate the benefits of aggregating models by examining 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 2 presents the OOS Sharpe ratios of the tangency portfolio, con-

Table 2: Out-of-Sample Model Performance

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

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

structed from the BMA estimates of risk prices. In comparison, we tabulate in Columns (2)-(7) the OOS Sharpe ratios from: (1) the model with the highest posterior probability (Top 1), (2) the full model that always includes all factors under consideration (All), (3) Carhart (1997) four-factor model (Carhart4), (4) Fama and French (2016) five-factor model (FF5), (5) Hou et al. (2015) q-factor model (HXZ4), and (6) Daniel et al. (2020) behavioral factor model (DHS3). We use the nonparametric bootstrap to test the null hypothesis that BMA and the other model deliver an identical Sharpe ratio, i.e., $H_0 : \mathrm{SR}^2_{bma} = \mathrm{SR}^2_{\boldsymbol{\gamma}}$.[13]

We start with describing the full-sample performance, as shown in the first row of Table 2. First, BMA outperforms traditional factor models out-of-sample. The top Bayesian

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

model (see Column (2)) has an OOS Sharpe ratio of 1.75, which is 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 divide the full sample into three equal subsamples. Consistent with past literature, the performance of factor models tends to decline over time, and the drops in Sharpe ratios are sizable 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 is significantly higher than those other models except that composed of all 14 factors.

The last three rows of Table 2 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 approximately 2.57. In other words, when the data overwhelmingly support 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, making it significantly more profitable than any other specification.

In summary, Table 2 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.

5 Aggregate Mutual Fund Flows and Model Uncertainty

Results from Section 4 have provided normative guidance on what mean-variance-efficient investors should do to account for model uncertainty. In reality, how do market participants respond to the changing model uncertainty in equity markets? In this section, we examine the association between investors' asset allocation decisions across different types of mutual

funds and model uncertainty.

Our analyses focus on aggregate fund flows to different asset classes and investment styles. We rely on the objective codes provided by the CRSP mutual fund database to categorize fund products and aggregate their flows. Following the literature (Chevalier and Ellison, 1997; Sirri and Tufano, 1998), we calculate the net dollar 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. Then, we aggregate individual fund flows in each period across all funds belonging to a specific category (e.g., large-cap funds, Treasury funds) and scale the aggregate flows by the lagged total market capitalization of all stocks in CRSP:

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

in which the superscript $\mathcal{O}$ specifies a specific investment objective. Since high-quality investment objective data are unavailable until 1990, our sample begins thereafter.

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

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

where l denotes the lag order, $(\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_l, \boldsymbol{B}_{\boldsymbol{x}})$ are the matrices of coefficients, $\boldsymbol{Y}_t$ is a $k \times 1$ vector concatenating both aggregate fund flows and our model uncertainty measure, $\boldsymbol{X}_t$ is a vector of control variables, and $\boldsymbol{u}_t$ is a $k \times 1$ vector of reduced-form innovations with the covariance matrix $\boldsymbol{\Sigma}_u$. We assume that the innovations $\boldsymbol{u}_t$ can be related to the structural shocks to the system, namely $\boldsymbol{\epsilon_t}$ (the elements of which are orthogonal to each other), via a linear transformation $\boldsymbol{u_t} = \boldsymbol{S}\boldsymbol{\epsilon_t}$, where $\boldsymbol{S}$ is a $k \times k$ structural impact matrix.

Identifying the structural impact matrix $\boldsymbol{S}$ is crucial to credibly quantify the impulse response functions (IRFs), which we use to understand the dynamic effects of model un-

certainty shocks on fund flows. A straightforward approach to estimating $\boldsymbol{S}$ is performing Cholesky decomposition of $\boldsymbol{\Sigma}_u$. Then, the ordering of variables in $\boldsymbol{Y_t}$ will determine the structural impacts. In other words, the identification of IRFs depends on whether model uncertainty is an exogenous source of fluctuations in fund flows or an endogenous propagating channel. As our goal is to investigate the dynamic relationship between fund flows and model uncertainty (instead of identifying the structural shocks), we do not take a firm stance on either of these possibilities. We will consistently report IRFs under both scenario throughout this section.

5.1 Aggregate Flow Responses: Equity vs. Fixed-Income Funds

In the baseline analysis, we consider the aggregate mutual fund flows to the US fixed-income (FI) and equity (EQ) mutual funds. The vector $\boldsymbol{Y_t}$ of Equation (16) includes $\{\mathcal{E}_t,\ \text{Flow}_t^{FI},\ \text{Flow}_t^{EQ}\}$. We present the results based on two different identification assumptions. In the first case, we place model uncertainty first in the VAR. The implicit identification assumption is that fund flows react to contemporaneous model uncertainty shocks, while model uncertainty does not respond to shocks to fund flows in the current period. In the second case, we place model uncertainty as the last element in $\boldsymbol{Y_t}$. Under this configuration, shocks to fund flows have impacts on contemporaneous model uncertainty, which further propagates these shocks to future fund flows.

Table 3 reports the results from the VAR regression. The sample ranges from January 1991 to December 2020. The lag is chosen by the Bayesian information criterion (BIC) and equals one. In addition, we standardize all economic variables such that they have unit variances. We also include the lagged market return and VIX index as control variables. The reported t-statistics are based on the Newey-West estimates of the covariance matrix with 36 lags. Several results are noteworthy. First, model uncertainty relates only to its lag. Second, model uncertainty negatively forecasts equity fund flows, and the coefficient estimate

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

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

This table reports the results from the VAR estimation in equation (16), where $\boldsymbol{Y_t}$ includes $\{\mathcal{E}_t, \text{Flow}_t^{FI}, \text{Flow}_t^{EQ}\}$. $\mathcal{E}_t$ is the model uncertainty measure, and Flow_t^{FI} (Flow_t^{EQ}) is the aggregate flows to the domestic fixed-income (equity) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (15)). The lag is chosen by the Bayesian information criterion and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged market return, lagged fixed-income fund return, and VIX index in each regression. The sample spans from January 1991 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags.

is sizable in both economic and statistical senses. In particular, a one-standard-deviation increase in model uncertainty predicts 34% of a standard deviation in equity fund outflows.

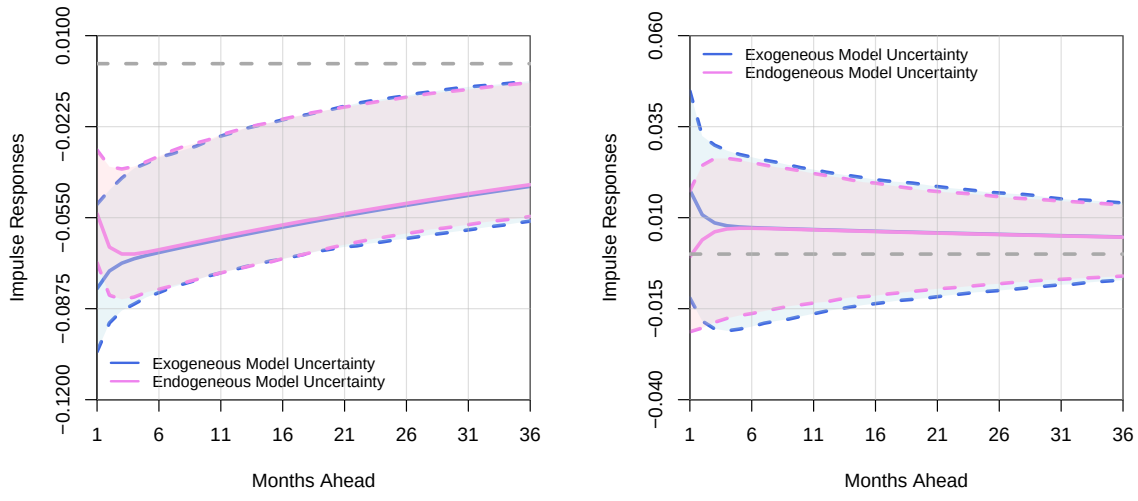
Figure 4 plots the dynamic responses of fund flows to model uncertainty shocks in VAR(1). Strikingly, model uncertainty innovations sharply induce fund outflows from the US equity market, with the effects persisting even after 36 months, as depicted in Panel (a). The IRFs start from −0.08 in period one and slowly decline to −0.05 in period 36, a finding that is significant and negative based on the 90% standard error bands. In contrast, model uncertainty has negligible effects on fixed-income fund flows (see Panel (b)).

5.2 Heterogeneity in Flow Responses: Equity Funds

We further study the heterogeneous responses of different equity mutual funds to model uncertainty shocks. In particular, we separate equity mutual funds into four categories: (a) style funds that specialize in factor investing, (b) sector funds that invest in specific industries, (c) small-cap funds that invest in small stocks,[14] and (d) large-cap funds.

Table 4 reports the estimation results of Equation (16), where the vector $\boldsymbol{Y_t}$ incluces

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



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

Figure 4: Impulse Responses of Equity and Fixed-Income Mutual Fund Flows to Model Uncertainty Shocks

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

$\{\mathcal{E}_t, \text{Flow}_t^{style}, \text{Flow}_t^{sector}, \text{Flow}_t^{small}, \text{Flow}_t^{large}\}$. The lag of VAR is chosen by the BIC and equals one. Since the cap-based investment objective code is available after 1997, the sample begins in January 1998. Model uncertainty negatively forecasts style and small-cap fund flows, and the coefficients are sizable. If model uncertainty rises by one standard deviation, style (small-cap) fund flows tend to decline by 26% (12%) of one standard deviation over the next period. In contrast, we do not discover a significant relationship between model uncertainty and sector/large-cap fund flows.

Figure 5 shows the dynamic responses of the four types of equity fund flows to model uncertainty shocks. The amount of capital withdrawn from style equity funds is remarkably large when model uncertainty is heightened (see Panel (a)). We also observe significantly negative IRFs of small-cap fund flows (see Panel (c)), although the effects are less persistent

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

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

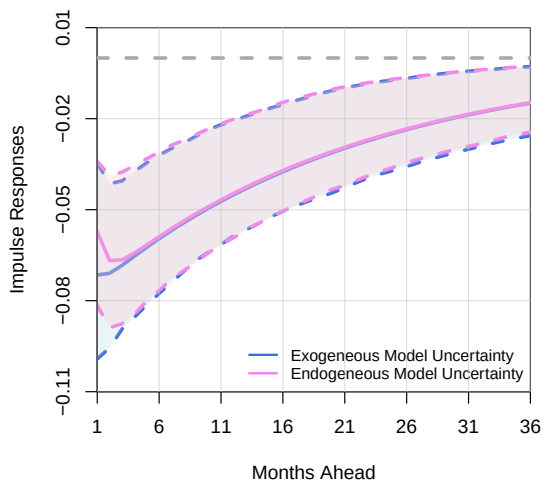
This table reports the results from the VAR estimation in equation (16), where $\boldsymbol{Y_t}$ incluces $\{\mathcal{E}_t, \text{Flow}_t^{style}, \text{Flow}_t^{sector}, \text{Flow}_t^{small}, \text{Flow}_t^{large}\}$. $\mathcal{E}_t$ is the model uncertainty measure, and Flow_t^{style} (Flow_t^{sector}, Flow_t^{small}, Flow_t^{large}) is the aggregate flows to the domestic style (sector, small-cap, large-cap) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (15)). The lag is chosen by the Bayesian information criterion and equals one. We standardize all economic variables such that they have unit variances. We also control for the lagged fund returns of each type and VIX index in each regression. The sample spans from January 1998 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags.

than those for style funds. On the contrary, sector and large-cap funds do not respond to model uncertainty shocks.

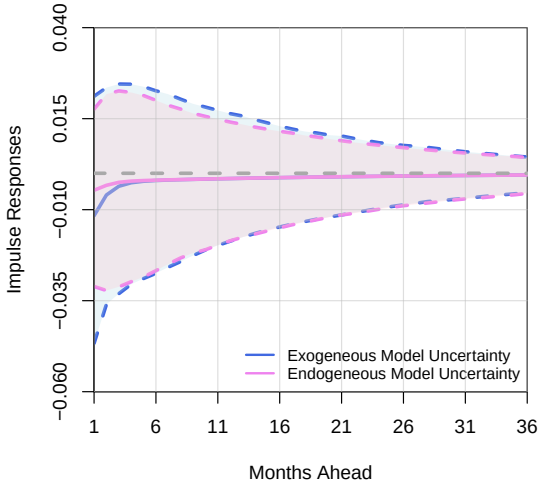
5.3 Heterogeneity in Flow Responses: Fixed-Income Funds

Now we turn our attention to the heterogeneity among flows to fixed-income funds. We divide all fixed-income mutual funds into four categories: (a) government bond funds, (b) money market funds, (c) corporate bond funds, and (d) municipal bond funds. This subsection repeats a similar VAR estimation and investigates the dynamic responses of fixed-income fund flows to model uncertainty shocks.

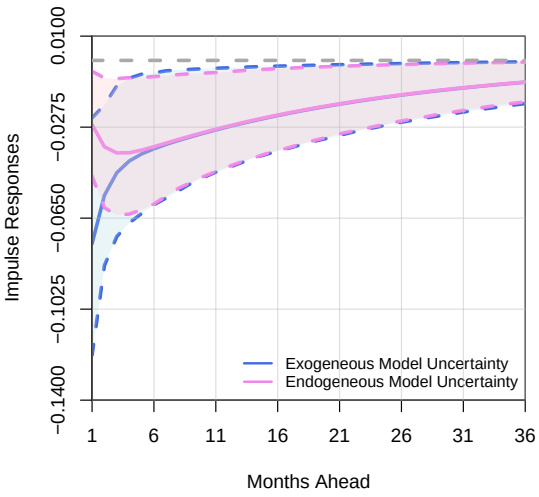
Table 5 shows the VAR(1) estimation results. Model uncertainty positively predicts the flows into US government bond funds (t-stat.=2.5) and negatively predicts the flows into corporate bond funds (t-stat.=−1.9), although the statistical significance is marginal. The negligible effects of model uncertainty on *total* fixed-income fund flows are mainly due to the netting of these two opposite impacts. In addition, we do not observe the same significant



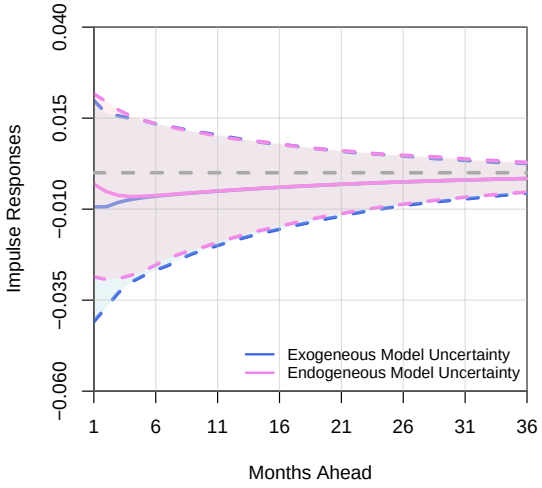
(a) Style Fund Flows



(b) Sector Fund Flows



(c) Small-Cap Fund Flows



(d) Large-Cap Fund Flows

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

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

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

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

This table reports the results from the VAR estimation in equation (16), where $\boldsymbol{Y_t}$ includes $\{\mathcal{E}_t, \text{Flow}_t^{gov}, \text{Flow}_t^{money}, \text{Flow}_t^{corp}, \text{Flow}_t^{muni}\}$. $\mathcal{E}_t$ is the model uncertainty measure, and Flow_t^{gov} (Flow_t^{money}, Flow_t^{corp}, Flow_t^{muni}) is the aggregate flows to the domestic government bond (money market, corporate bond, and municipal bond) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (15)). The lag is chosen by the Bayesian information criterion and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged fund returns of each type and VIX index in each regression. The sample spans from January 1998 to December 2020. We report both coefficient estimates and t-statistics, calculated using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

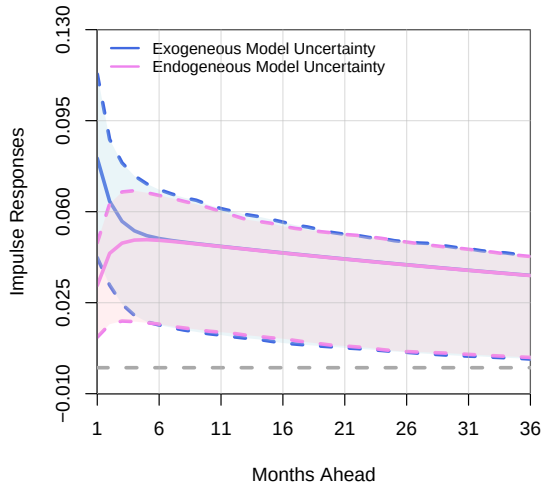
predictive relationship for the flows into municipal bond funds and money market funds.

The IRFs of different fixed-income fund flows are plotted in Figure 6. We consistently find dynamic inflows to government bond funds in response to model uncertainty shocks, under the two different identification assumptions. As Panel (a) suggests, a unit model uncertainty shock corresponds to over 3% of a standard deviation increase in government bond fund inflows over the next month, and the dynamic response persists for more than 36 periods. On the contrary, the IRFs of other fixed-income fund flows are insignificant.

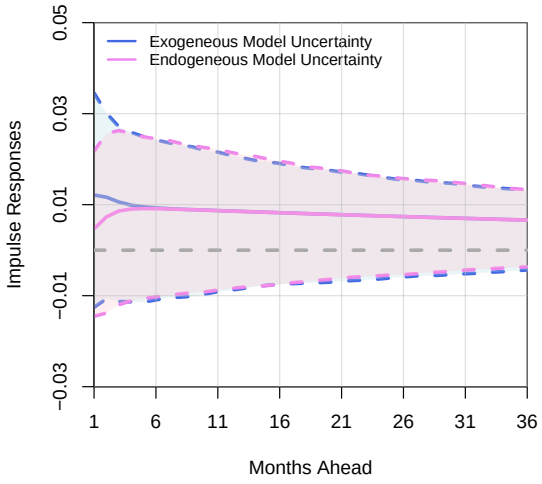
5.4 Fund Flow Responses to Other Uncertainty Measures

For comparison, this section examines how other uncertainty measures reported in the literature affect aggregate fund flows. We consider the CBOE VIX index and the financial uncertainty index in Ludvigson et al. (2021), both of which also target asset markets and are associated with our model uncertainty index, as tabulated in Table 1.

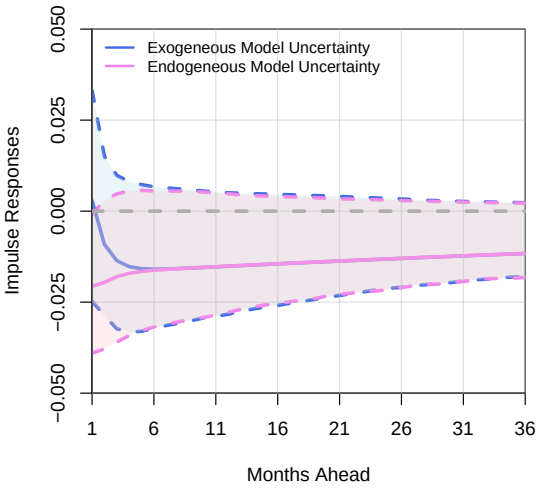
Figures A8 and A9 plot the dynamic responses of four different types of equity fund flows



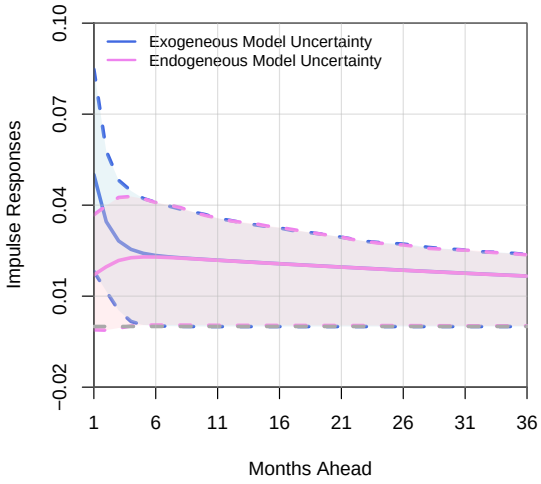
(a) Government Bond Fund Flows



(b) Money Market Fund Flows



(c) Corporate Bonds Fund Flows



(d) Municipal Bond Fund Flows

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

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

to VIX and financial uncertainty shocks. Consistent with the previous analyses, we consider two identification assumptions, placing the uncertainty measure as the first (exogenous shock) or last (endogenous response) variable in the VAR. We also control for lagged model uncertainty in each regression. Unlike the responses to model uncertainty shocks, we do not observe significant IRFs of equity fund flows to VIX or financial uncertainty shocks. This evidence highlights our model uncertainty measure in terms of its unique associations with investors' asset allocation decisions among different equity mutual funds.

We present the dynamic responses of fixed-income fund flows in Figures A10 and A11. All four types of fixed-income fund flows do not respond to financial uncertainty shocks. Interestingly, we find significant flows into money market funds after positive VIX shocks (6% of one standard deviation increase in response to a unit VIX shock). This effect, however, is transient. In contrast, our model uncertainty measure is associated with persistent inflows into government bond fund, while has no impacts on flows into money market funds.

In summary, our model uncertainty measure induces some unique dynamic responses of fund flows, and notably, they are distinct from uncertainty measures in the literature. In particular, we observe significant inflows to government bond funds and outflows from style and small-cap equity funds only in response to our model uncertainty measure.

6 Model Uncertainty in International Equity Markets

This section presents the time-series of model uncertainty in European and Asia Pacific stock markets. Instead of using all 14 factors as in the US stock market, we include only nine of them because of the limited data availability. Specifically, the size (ME), profitability (ROE), and investment (IA) factors in Hou et al. (2015) and short-term and long-term behavioral factors in Daniel et al. (2020) are excluded because they are unavailable in international markets. Since the AQR library provides the QMJ factor from July 1993, and we use a three-year rolling window, our model uncertainty measure starts in June 1996.

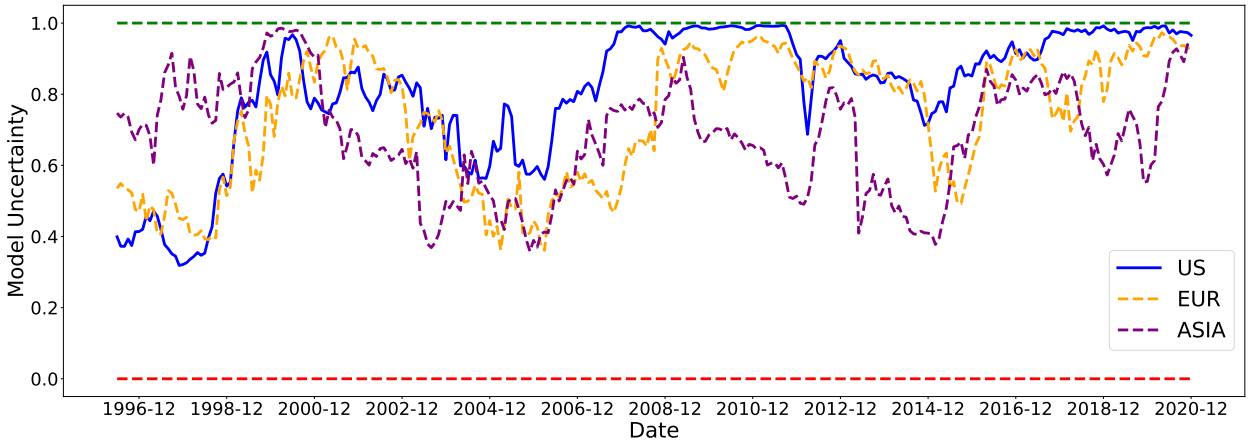


Figure 7: Model Uncertainty in European and Asia Pacific Markets

The figure plots the time series of model uncertainty about the linear stochastic discount factor (SDF) in European and Asian Stock Markets. The construction of model uncertainty is the same as in Figure 1 except that we use only nine factors to calculate the posterior model probabilities. Details about used factors could be found in section 6. The sample ranges from July 1993 to December 2020. Since we use 3-year rolling window, the model uncertainty index starts from June 1996. The red line and green lines in the figure show the lower (0) and upper bounds (1) of model uncertainty. As comparison, we include the US model uncertainty index (blue solid line) in Figure 1 during the same period.

The orange line in Figure 7 plots the time-series of model uncertainty in the European stock market from June 1996 to December 2020. The time-series patterns in European markets are in general similar to the US stock market.[15] The main exception is that, during the 2008 global financial crisis, culminating model uncertainty in these markets appears almost one year later than in the US market.

The purple line in Figure 7 presents our results for the Asia Pacific markets.[16] Model uncertainty is uniquely high in these markets starting in 1997, in correspondence to the profound 1997 Asian financial crisis, even surpassing the peak level during the 2008 crisis.

In short, model uncertainty in the European and Asia Pacific markets is also high in many periods, far above its lower bound. Second, it fluctuates significantly over time and coincides with major events in corresponding asset markets.

[15]European markets include the following countries: Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, UK, Greece, Ireland, Italy, the Netherlands, Norway, Portugal, and Sweden.

[16]The Asia Pacific market refers to the stock markets in Australia, Hong Kong, New Zealand, and Singapore.

7 Conclusions

We introduce a new metric of model uncertainty in cross-sectional asset pricing. This metric is the information entropy of posterior probabilities for SDF models under a Bayesian econometric framework. The variation of our model uncertainty measure is associated with the fluctuations in asset markets and the broader economy. We delve into both normative and descriptive implications of model uncertainty, shedding light on its significance for investors' asset allocation decisions.

In our concluding remarks, we circle back to the three initial questions posed in the introduction. Our measure is anchored within the range of zero to one. When a true asset pricing model exists and no factors are omitted, model uncertainty converges to zero in probability. Even in cases where omitted factors exist, our measure is bounded far from one (almost surely). As a result, if we observe high model uncertainty in the data, it serves as evidence against the possibility of identifying a true model.

In scenarios where model uncertainty is elevated, even factor models with low probabilities can convey nontrivial information about the cross-section of expected returns. We demonstrate that averaging all models benefits mean-variance investors in high-uncertainty environments. However, disregarding asset pricing models that are overshadowed by others does not necessarily incur significant costs, particularly when model uncertainty is low.

In our framework, we make a clear distinction between the probability of an asset pricing model and the probability of an individual factor entering the SDF (conditional on the observed return data). Even when the data do not strongly favor any specific models, they can still offer compelling evidence in favor of particular factors. Although there is generally high model uncertainty in the cross-section of equity returns, it does not necessarily negate the existence of useful asset pricing factors.

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*, Volume 55. Courier Corporation.

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

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

AVRAMOV, D., S. CHENG, L. METZKER, and S. VOIGT (2023). Integrating factor models. *The Journal of Finance 78*(3), 1593–1646.

Avramov, D. and G. Zhou (2010). Bayesian portfolio analysis. *Annual Review of Financial Economics 2*(1), 25–47.

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

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.

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.

Chevalier, J. and G. Ellison (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy 105*(6), 1167–1200.

Chib, S. and X. Zeng (2020). Which factors are risk factors in asset pricing? a model scan framework. *Journal of Business & Economic Statistics 38*(4), 771–783.

Chib, S., X. Zeng, and L. Zhao (2020). On comparing asset pricing models. *Journal of Finance 75*(1), 551–577.

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

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

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

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

Dew-Becker, I. and S. Giglio (2023). Cross-sectional uncertainty and the business cycle: evidence from 40 years of options data. *American Economic Journal: Macroeconomics 15*(2), 65–96.

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

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

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

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

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

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.

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

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

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

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

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

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

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

Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: an asset-allocation perspective. *Journal of Finance 51*(2), 385–424.

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

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

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

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

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

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

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

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

Pástor, L. and P. Veronesi (2006). Was there a nasdaq bubble in the late 1990s? *Journal of Financial Economics 81*(1), 61–100.

Pástor, L. and P. Veronesi (2009). Technological revolutions and stock prices. *American Economic Review 99*(4), 1451–83.

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.

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

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

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

Appendices

Appendix A Factor Details

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

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

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

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

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

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

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

Mispricing factors. Stambaugh and Yuan (2017) propose two mispricing factors that ag-

gregate information across 11 prominent anomalies discovered in past literature. We download the data from Robert Stambaugh's website.

Appendix B Additional Tables

Table A1: Summary Statistics of 14 Factors

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

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

Table A2: The Autocorrelation of Different Uncertainty Measures

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

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

Table A3: Simulation Study: Posterior Properties Without Misspecification

	$T = 750$				$T = 1500$				$T = 15000$			
scenario	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:												
$\gamma_{0,j} = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{0,j} = 0$	0.48	0.44	0.33	0.22	0.48	0.44	0.33	0.18	0.47	0.43	0.32	0.08
	(0.13)	(0.16)	(0.18)	(0.17)	(0.14)	(0.16)	(0.18)	(0.17)	(0.13)	(0.15)	(0.18)	(0.12)
Posterior Probabilities of Models $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$:												
$\mathcal{M}_{\boldsymbol{\gamma}} = \mathcal{M}_{\boldsymbol{\gamma}_0}$	0.28	0.32	0.45	0.61	0.28	0.32	0.45	0.67	0.28	0.33	0.46	0.85
	(0.10)	(0.13)	(0.17)	(0.19)	(0.10)	(0.13)	(0.18)	(0.20)	(0.10)	(0.13)	(0.17)	(0.16)
$\mathcal{M}_{\boldsymbol{\gamma}} \supset \mathcal{M}_{\boldsymbol{\gamma}_0}$	0.24	0.23	0.18	0.13	0.24	0.23	0.18	0.11	0.24	0.22	0.18	0.05
	(0.09)	(0.11)	(0.13)	(0.13)	(0.10)	(0.11)	(0.13)	(0.13)	(0.09)	(0.11)	(0.13)	(0.10)
$\mathcal{M}_{\boldsymbol{\gamma}} \not\supseteq \mathcal{M}_{\boldsymbol{\gamma}_0}$	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 $\mathcal{E}$:												
	0.38	0.37	0.32	0.26	0.38	0.36	0.32	0.23	0.38	0.37	0.32	0.12
	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)	(0.06)	(0.03)	(0.03)	(0.04)	(0.06)

This table presents simulation results on posterior probabilities of models being correct and factors entering the SDF. The full set of factors under consideration $\boldsymbol{f}$ are the Carhart four factors. We follow the assumptions in Proposition 4 and 6, treating the true SDF as $1 - (\boldsymbol{f}_{\boldsymbol{\gamma}_0} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}_0}])^\top \boldsymbol{b}$ where $\boldsymbol{b} = 0.5 \times \mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$). The true factors $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ are the Fama-French three factors.

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

For each of the 1,000 Monte Carlo sample, we calculate the posterior probability of each factor $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}], j = 1, \ldots, 4$, the posterior probability of each model $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$, 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 parenthesis are average across Monte Carlo samples; numbers in parenthesis are standard deviations across Monte Carlo samples.

Table A4: Simulation Study: Posterior Properties With Omitted Factors

	$T = 750$				$T = 1500$				$T = 15000$			
scenario	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g	$g = 2$	$g = 4$	$g = 16$	mix.g
Panel (A): true model (Carhart4)/ factors under consideration (FF5)												
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:												
$\gamma_{0,j} = 1$	0.91	0.91	0.89	0.88	0.95	0.95	0.94	0.92	1.00	1.00	1.00	1.00
	(0.17)	(0.18)	(0.22)	(0.25)	(0.13)	(0.13)	(0.16)	(0.21)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{0,j} = 0$	0.59	0.56	0.48	0.42	0.65	0.63	0.56	0.46	0.75	0.74	0.69	0.57
	(0.21)	(0.24)	(0.28)	(0.31)	(0.24)	(0.26)	(0.32)	(0.36)	(0.27)	(0.29)	(0.35)	(0.45)
Model Uncertainty Measure $\mathcal{E}$:												
	0.46	0.44	0.40	0.37	0.38	0.36	0.34	0.31	0.18	0.18	0.16	0.08
	(0.07)	(0.07)	(0.07)	(0.08)	(0.06)	(0.07)	(0.07)	(0.08)	(0.03)	(0.03)	(0.03)	(0.05)
Panel (B): true model (FF5)/ factors under consideration (Carhart4)												
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:												
$\gamma_{0,j} = 1$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\gamma_{0,j} = 0$	0.52	0.49	0.39	0.30	0.57	0.55	0.46	0.31	0.96	0.97	0.96	0.89
	(0.17)	(0.19)	(0.23)	(0.23)	(0.19)	(0.22)	(0.26)	(0.27)	(0.08)	(0.08)	(0.10)	(0.21)
Model Uncertainty Measure $\mathcal{E}$:												
	0.23	0.22	0.20	0.18	0.22	0.21	0.20	0.16	0.04	0.03	0.03	0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)

This table presents simulation results on factors' posterior probabilities of entering the SDF, under model misspecification. We follow the assumptions in Proposition 7, treating the true SDF as $1 - (\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}$ where $\boldsymbol{b} = 0.5 \times \mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_0$). True factors $\boldsymbol{f}_0$ and factors under consideration $\boldsymbol{f}$ are described accordingly in the titles of table panels. $\boldsymbol{f}_{\boldsymbol{\gamma}_0} = \boldsymbol{f} \cap \boldsymbol{f}_0$ is the subset of factors under consideration that enter the SDF.

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

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

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

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

This table presents simulation results on the model uncertainty measure under model misspecification. We follow the assumptions in Proposition 7 and 8, treating the true SDF as $1-(\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}$ where $\boldsymbol{b} = 0.5\times\mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_0$). True factors $\boldsymbol{f}_0$ are described accordingly in the titles of table panels. When $p = 4$, four factors are under consideration, namely $\boldsymbol{f} = \{\text{market}, \text{SMB}, \text{HML}, \text{RMW}\}$. For larger p, five additional factors are sequentially added into $\boldsymbol{f}$, in the order of $\{\text{CMA}, \text{QMJ}, \text{FIN}, \text{PEAD}, \text{BAB}\}$.

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

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

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

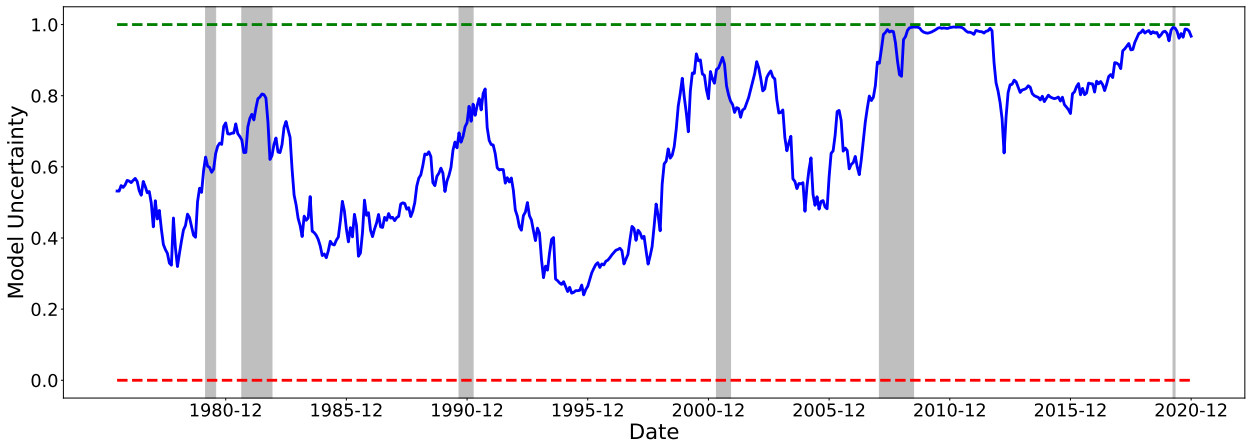
	MOM	RMW	CMA	FIN	PEAD	QMJ	BAB
Posterior Probabilities of Factors $\mathbb{P}[\gamma_j = 1 \mid \mathcal{D}]$:							
$\gamma_{0,j} = 1$	0.94	1.00	1.00	1.00	1.00	0.99	1.00
	(0.17)	(0.02)	(0.00)	(0.01)	(0.01)	(0.06)	(0.01)
$\gamma_{0,j} = 0$	0.49	0.35	0.27	0.43	0.29	0.36	0.38
	(0.33)	(0.27)	(0.21)	(0.31)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

This table presents additional simulation results on the model uncertainty measure under model misspecification. We follow the assumptions in Proposition 7 and 8, treating the true SDF as $1-(\boldsymbol{f}_0 - \mathbb{E}[\boldsymbol{f}_0])^\top \boldsymbol{b}$ where $\boldsymbol{b} = 0.5 \times \mathbf{1}$ (a vector repeating 0.5, the dimension of which is determined by the number of true factors in $\boldsymbol{f}_0$). True factors are the Fama-French three factors plus one additional factor (the name of which is at the top of each column). This additional factor will always be omitted. (The rationale behind only considering one omitted factor is due to the *linear* SDF assumption. If there are multiple omitted factors, a linear combination of them with their risk prices as coefficients create one omitted factors.) Factors under consideration, namely $\boldsymbol{f}$, are the factors in $\{\text{market}, \text{SMB}, \text{HML}, \text{MOM}, \text{RMW}, \text{CMA}, \text{QMJ}, \text{FIN}, \text{PEAD}, \text{BAB}\}$ *excluding* the factor defined by each column name.

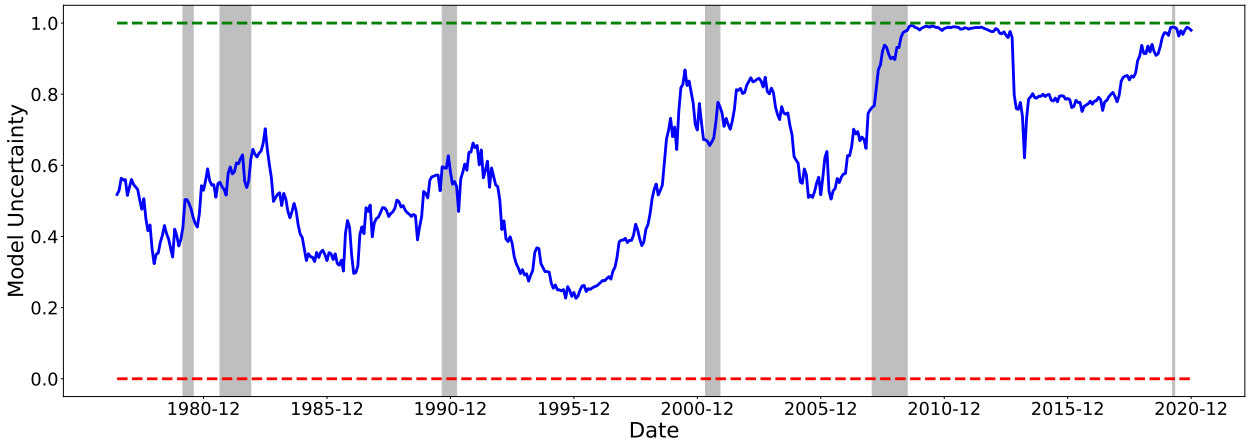
We simulate $1,000$ copies of return samples according to $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_0, \widehat{\boldsymbol{\Sigma}})$, where $T = 750$ days, and $\widehat{\boldsymbol{\Sigma}}$ is estimated from our sample of US equity returns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

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

Appendix C Additional Figures



(a) Model Uncertainty in 4-Year Rolling Window



(b) Model Uncertainty in 5-Year Rolling Window

Figure A1: Alternative Rolling Windows

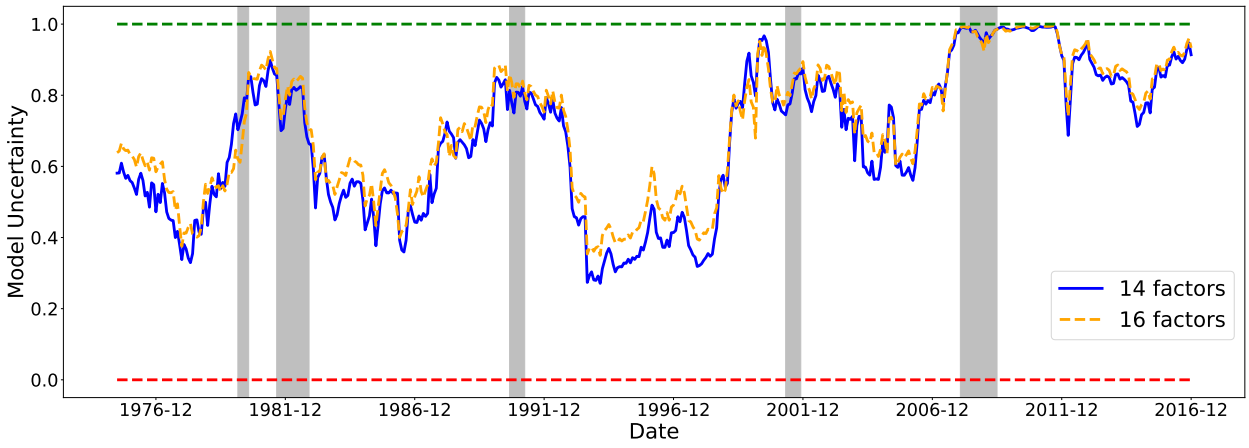


Figure A2: Time-Series of Model Uncertainty: Including two mispricing factors

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.

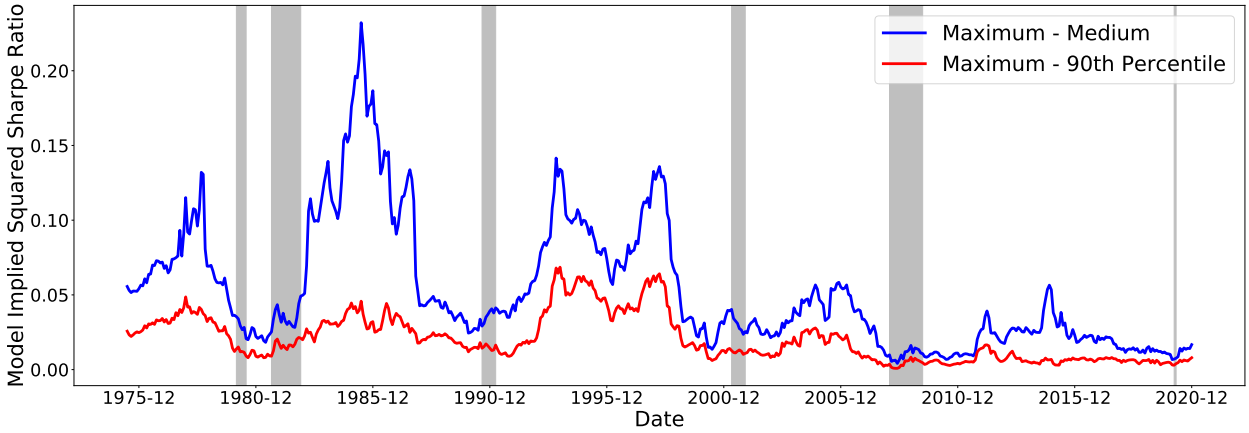
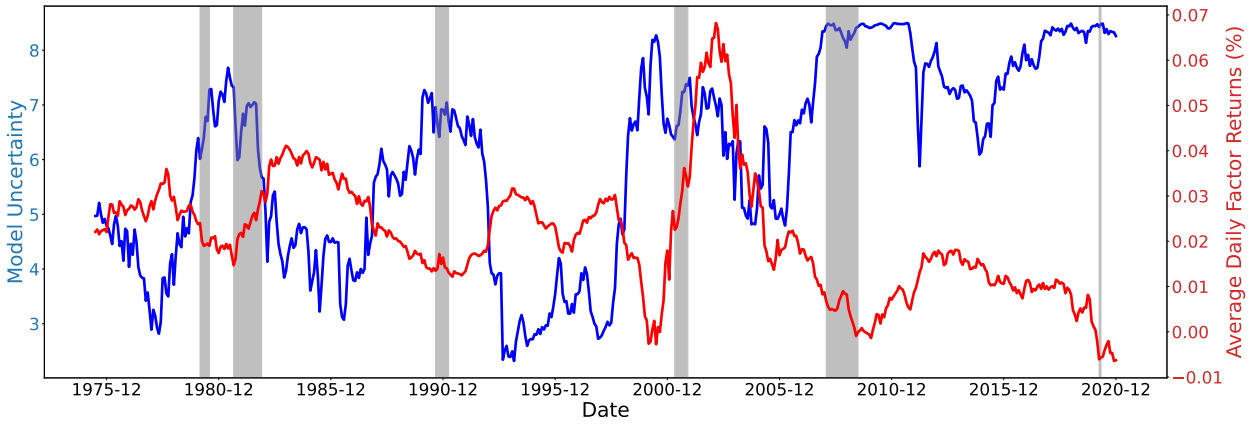
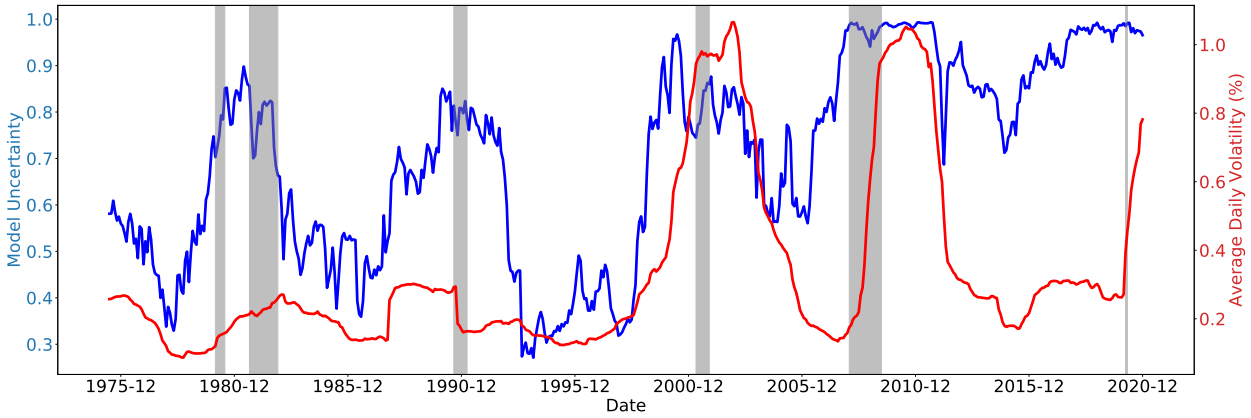


Figure A3: Time Series of Model-Implied Squared Sharpe Ratio (3-Year Rolling Window)

The figure plots the time series of distances in SR_γ^2 from June 1975 to December 2020. We present the difference between the highest SR_γ^2 and the 90th-quantile of SR_γ^2, as well as the difference between the highest SR_γ^2 and medium SR_γ^2. SR_γ^2 is the model-implied squared Sharpe ratio, $\overline{\boldsymbol{f}}_\gamma^\top \boldsymbol{S}_\gamma^{-1} \overline{\boldsymbol{f}}_\gamma$. The average factor returns $\overline{\boldsymbol{f}}_\gamma$ and variance-covariance matrix estimates $\boldsymbol{S}_\gamma$ are computed from the daily factor returns in the past 36 months.



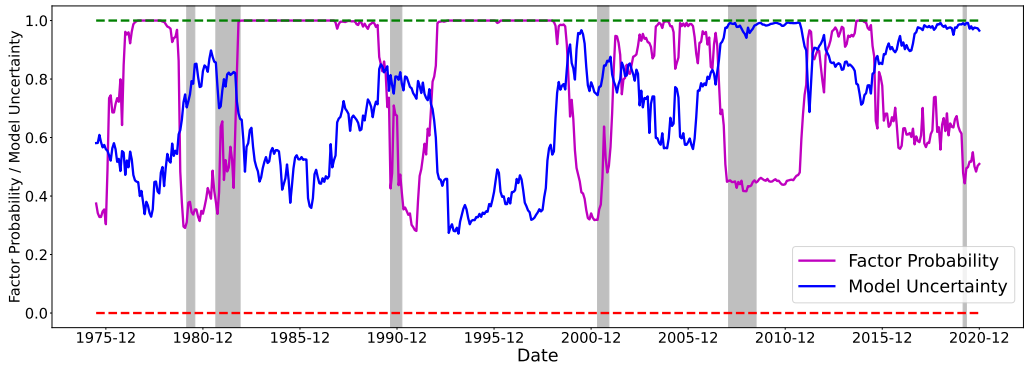
(a) Average (Daily) Return



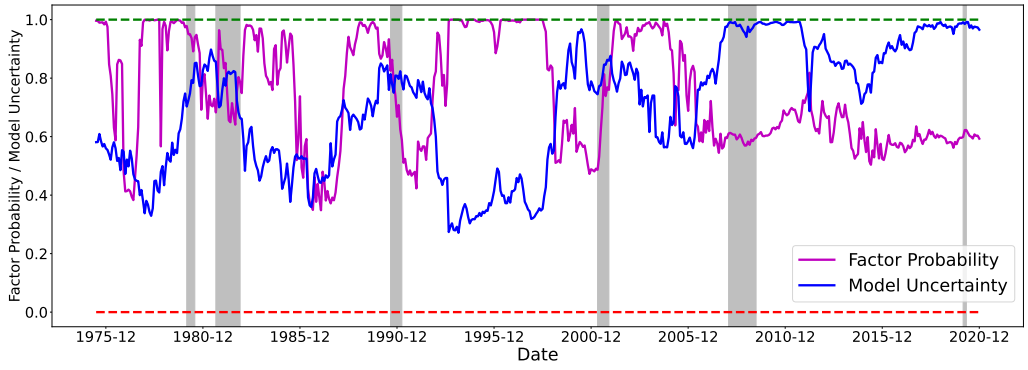
(b) Average (Daily) Volatility

Figure A4: Time-Series of Average (Daily) Mean Return and Volatility of 14 Factors

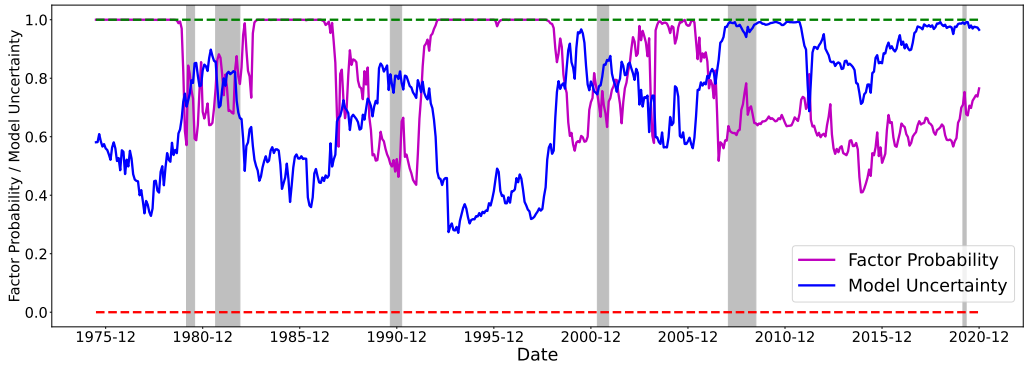
The figures plot the time-series of (a) average daily returns of factors and (b) average daily factor volatility, and these statistics are estimated using the daily factor returns in the past 36 months.



(a) MKT

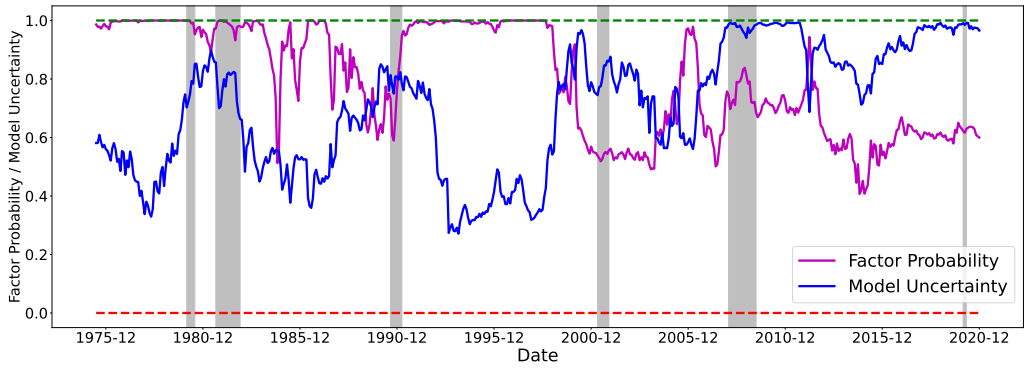


(b) Size (SMB or ME)

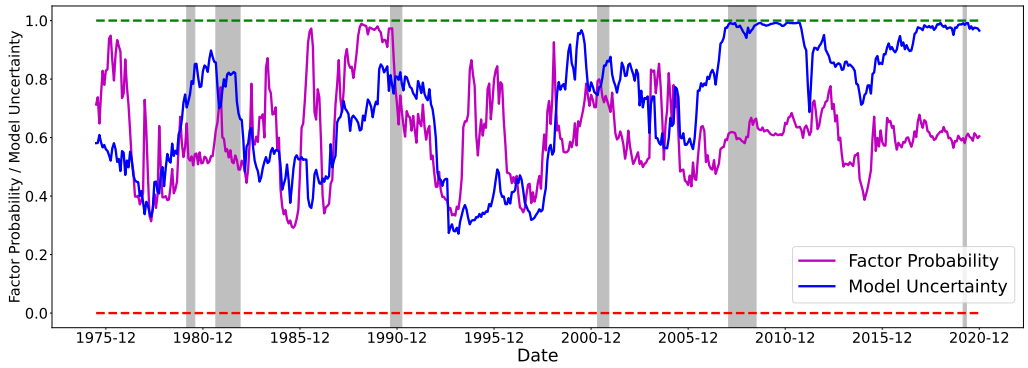


(c) Value (HML or HML devil)

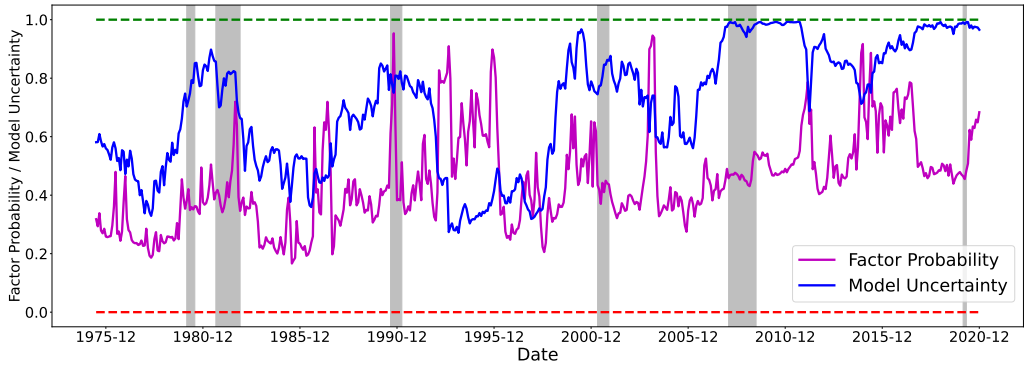
Figure A5: Time Series of Posterior Factor Probabilities



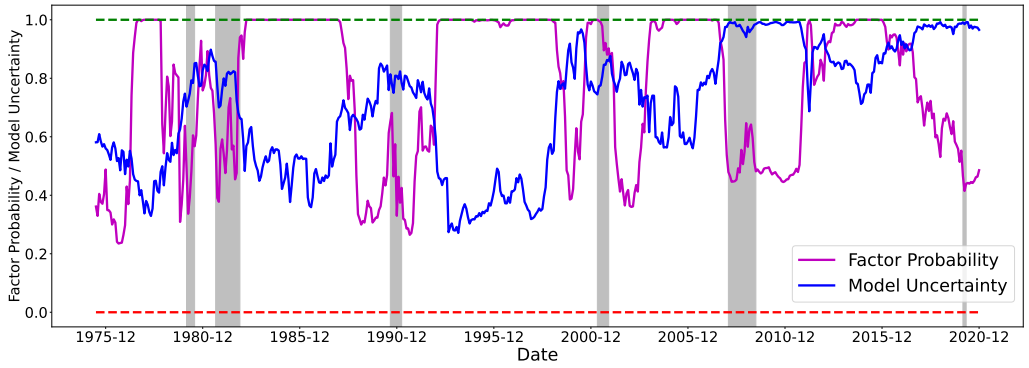
(d) Profitability (ROE or RMW)



(e) Investment (IA or CMA)

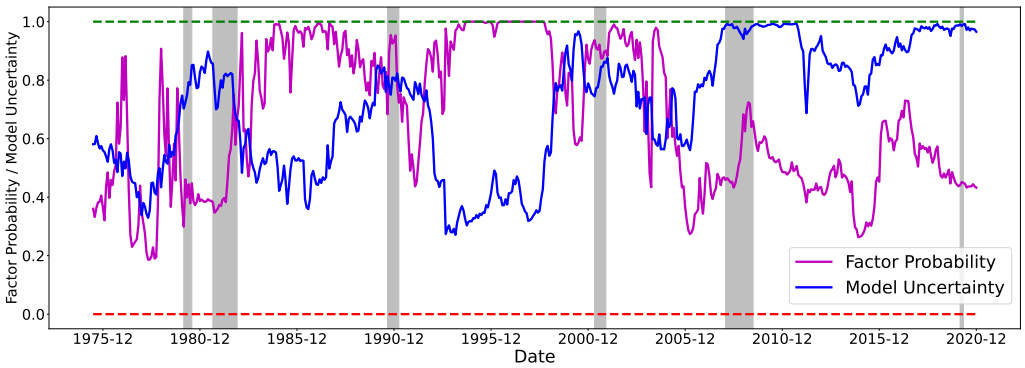


(f) Momentum

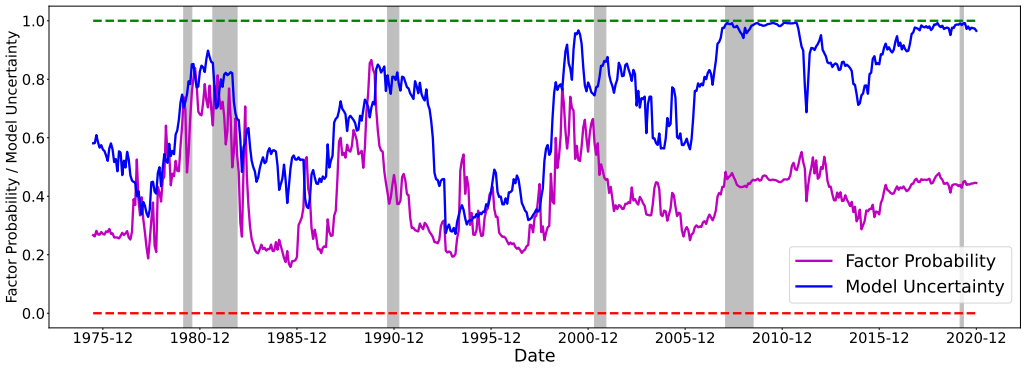


(g) BAB

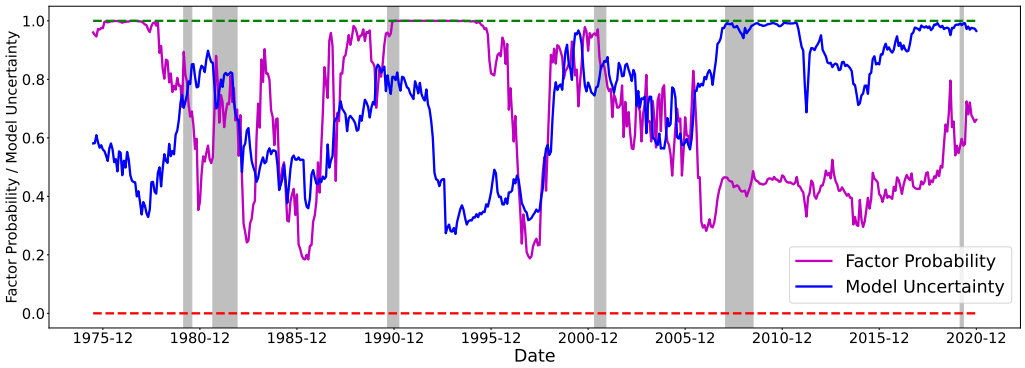
Figure A5: Time Series of Posterior Factor Probabilities (Continued)



(h) QMJ



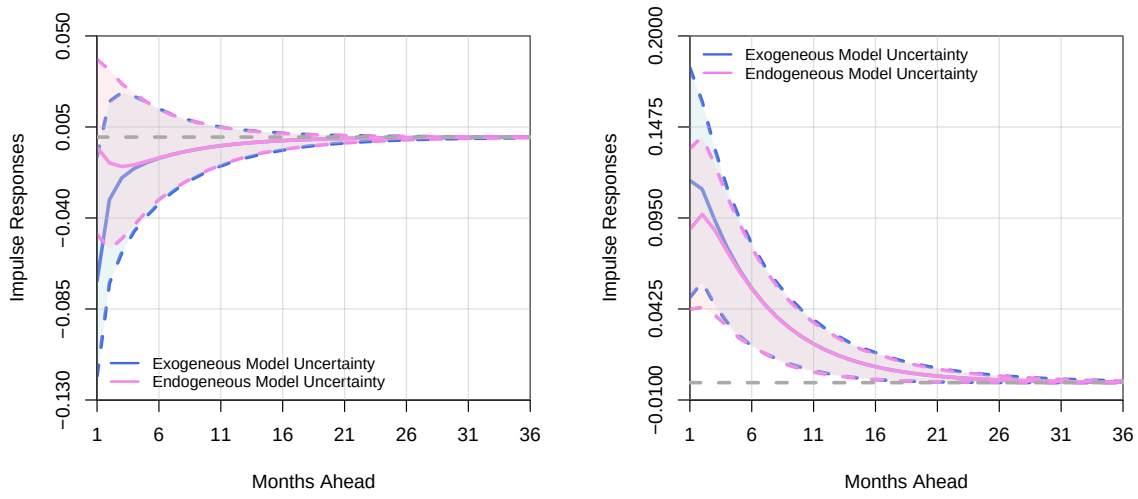
(i) FIN



(j) PEAD

Figure A5: Time Series of Posterior Factor Probabilities (Continued)

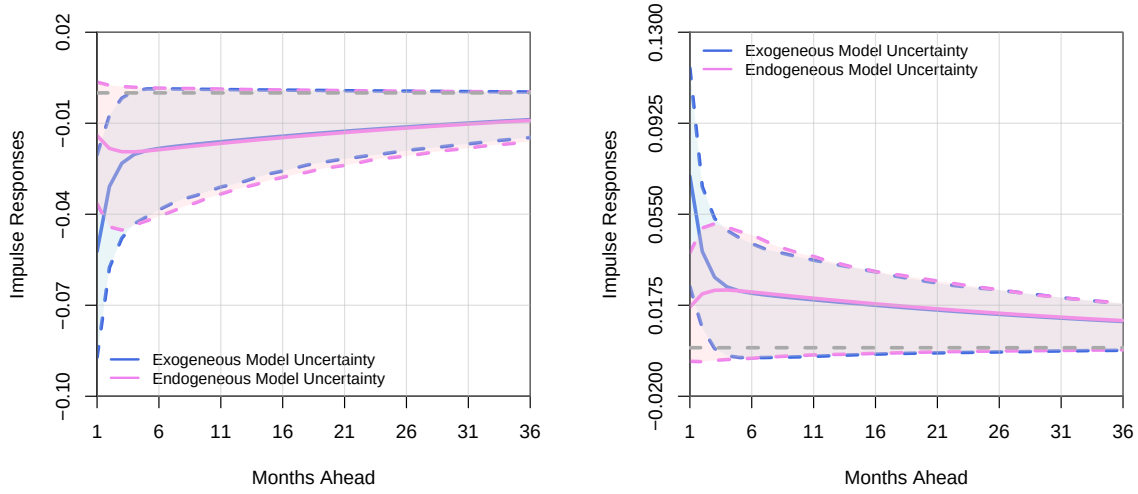
The figures plot the time series of posterior marginal probabilities of 14 factors. At the end of each month, we estimate models using the daily factor returns in the past three years. The sample ranges from July 1972 to December 2020. Since we use a three-year rolling window, the time series of factor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



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

Figure A6: Impulse Responses of Equity and Fixed-Income Fund Flows to VIX Shocks

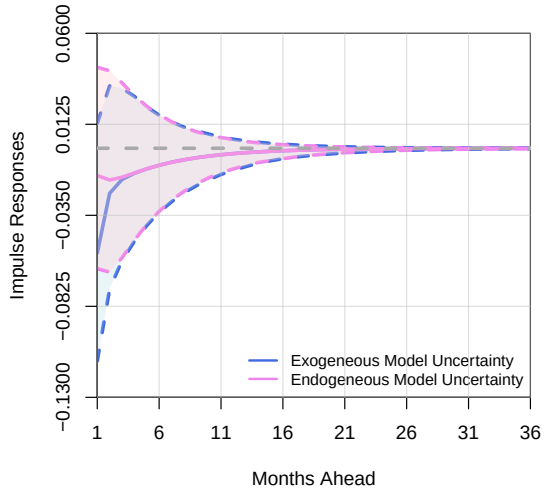
This figure shows the dynamic impulse response functions (IRFs) of equity and fixed-income fund flows to VIX shocks in VAR(1).



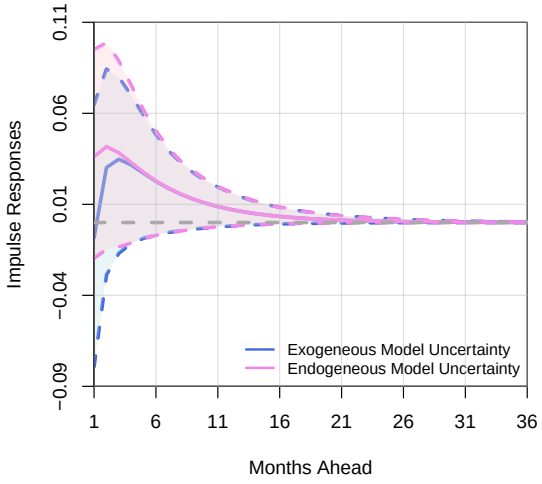
(a) Equity Fund Flows **(b)** Fixed-Income Fund Flows

Figure A7: Impulse Responses of Equity and Fixed-Income Mutual Fund Flows to Financial Uncertainty Shocks

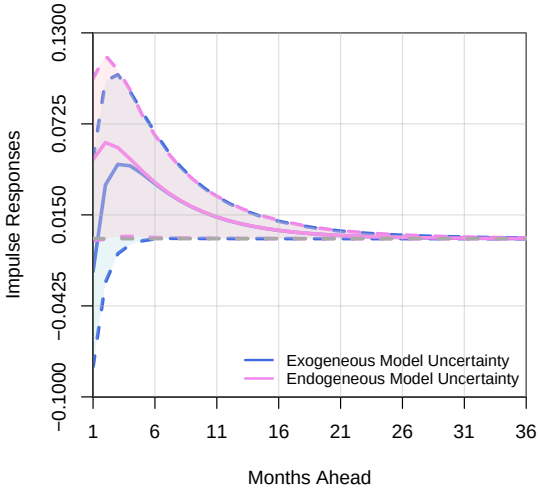
This figure shows the dynamic impulse response functions (IRFs) of equity and fixed-income fund flows to financial uncertainty shocks in VAR(1).



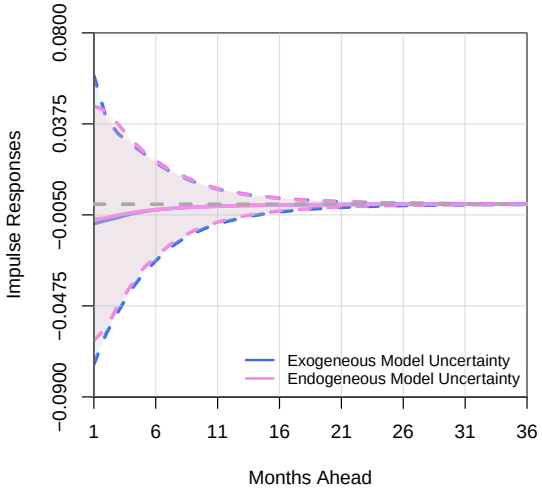
(a) Style Fund Flows



(b) Sector Fund Flows



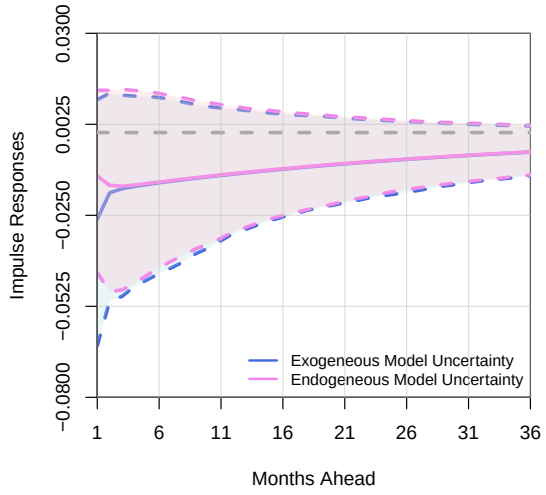
(c) Small-Cap Fund Flows



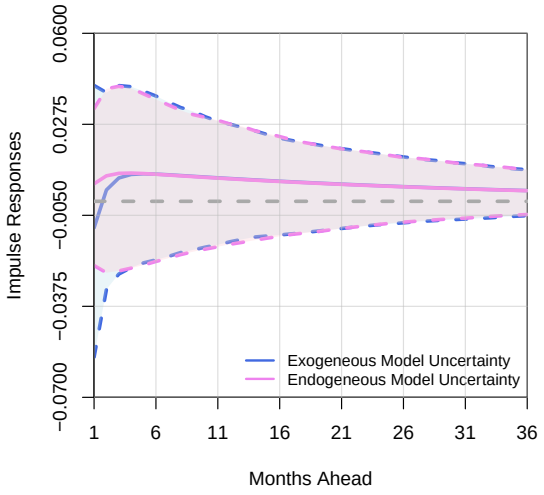
(d) Large-Cap Fund Flows

Figure A8: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes to VIX Shocks

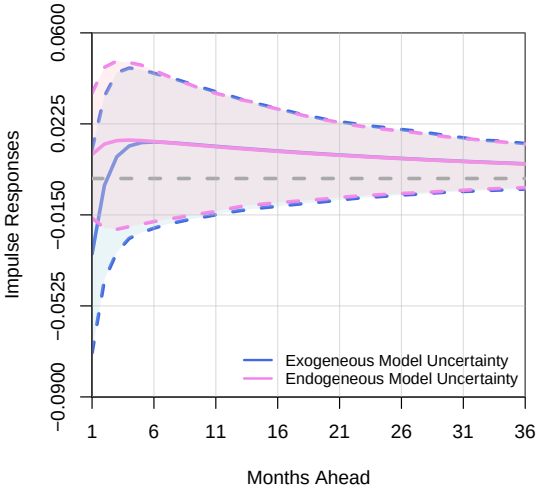
This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to VIX shocks in VAR(1). Other details can be found in the footnote of Figure 5.



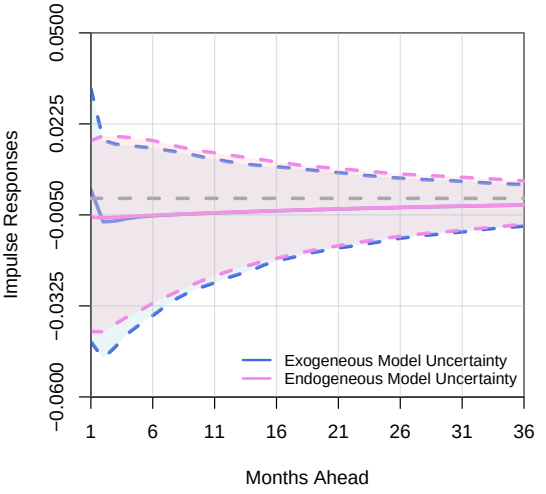
(a) Style Fund Flows



(b) Sector Fund Flows



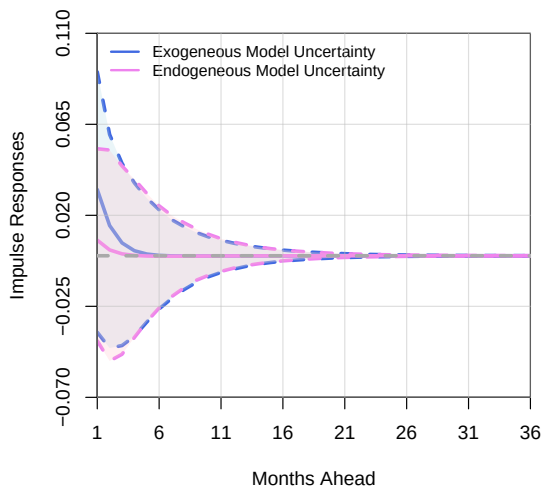
(c) Small-Cap Fund Flows



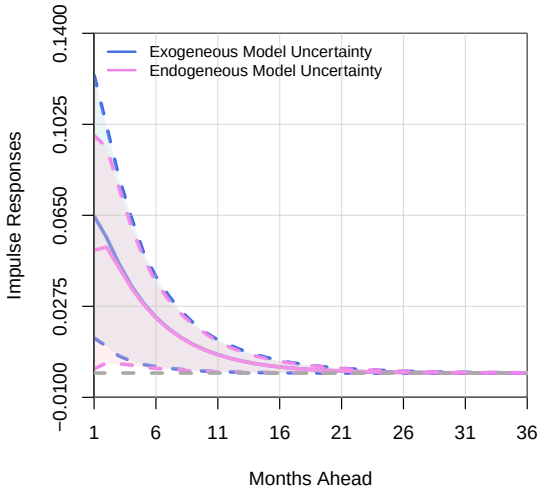
(d) Large-Cap Fund Flows

Figure A9: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes to Financial Uncertainty Shocks

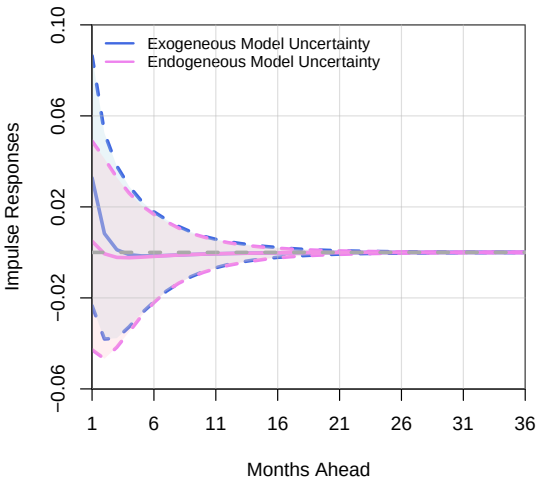
This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to financial uncertainty shocks in VAR(1). Other details can be found in the footnote of Figure 5.



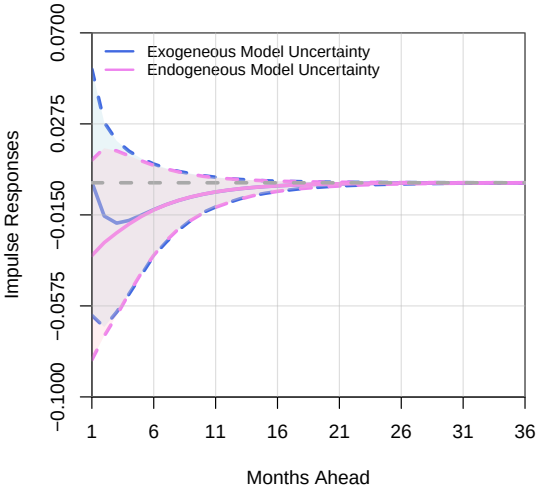
(a) Government Bond Fund Flows



(b) Money Market Fund Flows



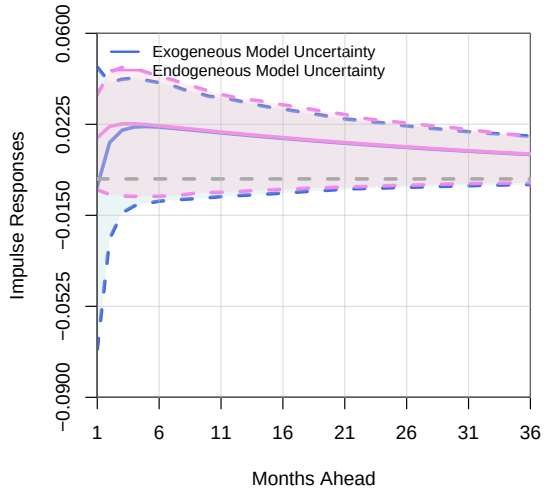
(c) Corporate Bond Fund Flows



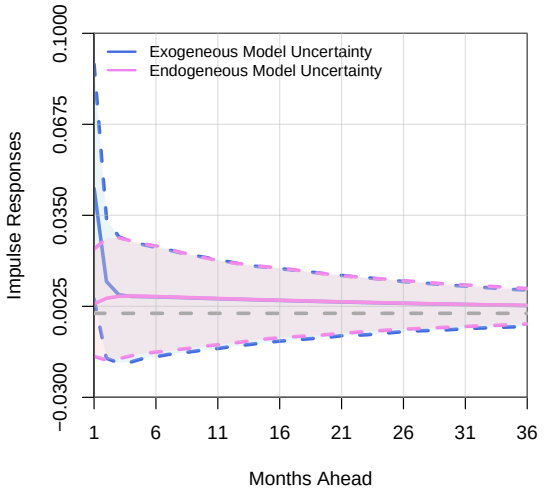
(d) Municipal Bond Fund Flows

Figure A10: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes to VIX Shocks

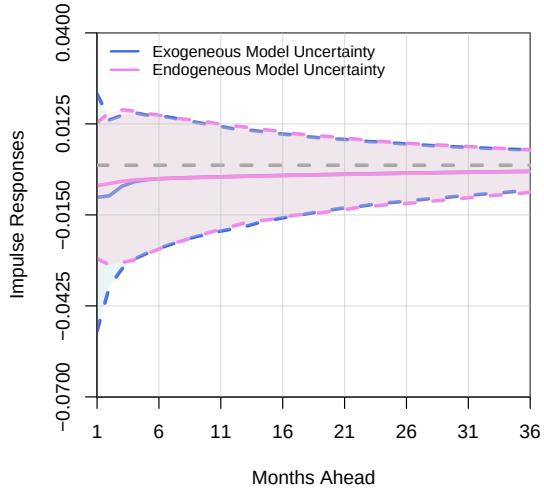
This figure shows the dynamic impulse response functions (IRFs) of fixed-income fund flows to VIX shocks in VAR(1). Other details can be found in the footnote of Figure 6.



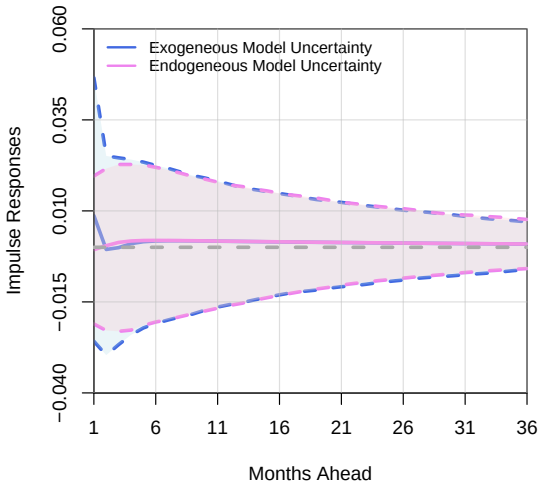
(a) Government Bond Fund Flows



(b) Money Market Fund Flows



(c) Corporate Bond Fund Flows



(d) Municipal Bond Fund Flows

Figure A11: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes to Financial Uncertainty Shocks

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

Appendix D Proofs

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

D.1 Proof of Proposition 1

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

Lemma A1. *Define* $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \text{var}[\boldsymbol{f}_{\boldsymbol{\gamma}}]$, $\boldsymbol{C}_{\boldsymbol{\gamma}} = \text{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]$, *and* $\boldsymbol{\Sigma} = \text{var}[\boldsymbol{R}]$, *then*

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

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

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

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

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

where

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

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,

$$\begin{aligned} \mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] = &\exp\left\{-\frac{1}{2} \boldsymbol{R}_{[1:T]}^\top \left[\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1} + \frac{g}{T} (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}}) \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1} (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^\top\right]^{-1} \boldsymbol{R}_{[1:T]}\right\} \\ &\times \left|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1} + \frac{g}{T} (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}}) \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1} (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^\top\right|^{-\frac{1}{2}} (2\pi)^{-\frac{NT}{2}}. \end{aligned} \tag{A1}$$

To simplify (A1), first, by the Sherman-Morrison-Woodbury formula,[17]

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

Second, by the generalized Sylvester's theorem for determinants,[18]

$$
\begin{aligned}
&\left|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma} + \frac{g}{T}(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}}) \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1} (\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^\top\right| \\
=&\frac{|Tg^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} + (\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^\top (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})|}{|Tg^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}| \times |\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}|} \\
=&\frac{|(g^{-1}+1)\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}|}{|g^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}| \times |\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}|} \qquad (\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}} \text{ according to Lemma A1}) \\
=&\frac{(1+g)^{p_{\boldsymbol{\gamma}}}}{|\boldsymbol{\Sigma}^{-1}|^T}.
\end{aligned}
$$

Plugging these two results above back to equation (A1), we get

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

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

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

where $\overline{\boldsymbol{R}} = \left(\sum_{t=1}^{T} \boldsymbol{R}_t\right)/T$, $\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} = \left(\sum_{t=1}^{T} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma},t}\right)/T$, $\boldsymbol{S} = \sum_{t=1}^{T}(\boldsymbol{R}_t - \overline{\boldsymbol{R}})(\boldsymbol{R}_t - \overline{\boldsymbol{R}})^\top/T$; the second equation above relies on the fact that $\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\gamma}}$ and that $\boldsymbol{C}_{\boldsymbol{\gamma}}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$ according to Lemma A1. □

D.2 Proof of Proposition 2

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

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

where $\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0} = \text{cov}[\boldsymbol{f}_{\boldsymbol{\gamma}}, \boldsymbol{f}_{\boldsymbol{\gamma}_0}]$ 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

$$\begin{aligned}
T\text{SR}_{\boldsymbol{\gamma}}^2 - T\delta_{\boldsymbol{\gamma}} =& T\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top\boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0}\boldsymbol{b}_{\boldsymbol{\gamma}_0} \\
=& T\left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0}\boldsymbol{b}_{\boldsymbol{\gamma}_0}\right)^\top\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0}\boldsymbol{b}_{\boldsymbol{\gamma}_0}\right) + 2T\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top\boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} - \boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0}\boldsymbol{b}_{\boldsymbol{\gamma}_0}\right) \\
=& \underbrace{\boldsymbol{z}_{\boldsymbol{\gamma}}^\top\boldsymbol{z}_{\boldsymbol{\gamma}}}_{\sim\chi^2(p_{\boldsymbol{\gamma}})} + \underbrace{\left(2\sqrt{T}\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top\boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}}{\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}}^{-\frac{1}{2}}\right)\boldsymbol{z}_{\boldsymbol{\gamma}}}_{\sim\mathcal{N}(0,\,4T\delta_{\boldsymbol{\gamma}})}. \qquad \text{(A2)}
\end{aligned}$$

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

$$\mathbb{E}[\text{SR}_{\boldsymbol{\gamma}}^2] = \delta_{\boldsymbol{\gamma}} + \frac{p_{\boldsymbol{\gamma}}}{T}. \qquad \text{(A3)}$$

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

$$\delta_{\boldsymbol{\gamma}_0} - \delta_{\boldsymbol{\gamma}} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top\left(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_0} - \boldsymbol{C}_{\boldsymbol{\gamma}_0,\boldsymbol{\gamma}}^\top\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma},\boldsymbol{\gamma}_0}\right)\boldsymbol{b}_{\boldsymbol{\gamma}_0} = \boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top\text{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right]\boldsymbol{b}_{\boldsymbol{\gamma}_0}. \qquad \text{(A4)}$$

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 (A3), the proof is completed. □

D.3 Proof of Proposition 3

Proof. Let $\boldsymbol{S}_{\boldsymbol{\gamma}}$ be the sample counterpart of $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$, indexed from the estimator $\boldsymbol{S}$. Then $\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^\top \boldsymbol{S}_{\boldsymbol{\gamma}}^{-1} \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}$, and $T\boldsymbol{S}_{\boldsymbol{\gamma}} \sim \mathcal{W}_{p_{\boldsymbol{\gamma}}}(\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}, T-1)$, a Wishart distribution with $(T-1)$ degrees of freedom and a scale matrix $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}$. Recall from the proof of Proposition 2, $\sqrt{T}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} \sim \mathcal{N}\left(\sqrt{T}\boldsymbol{C}_{\boldsymbol{\gamma},\gamma_0}\boldsymbol{b}_{\gamma_0}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}\right)$, and it is well-known that $\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} \perp \boldsymbol{S}_{\boldsymbol{\gamma}}$, thus

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

a non-central Hotelling's $\mathcal{T}^2$ distribution with degree-of-freedom parameters $p_{\boldsymbol{\gamma}}$ and $(T-1)$, as well as a non-centrality parameter $T\delta_{\boldsymbol{\gamma}}$.[19] Equivalently,

$$\frac{T-p_{\boldsymbol{\gamma}}}{p_{\boldsymbol{\gamma}}} \times \widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}} \sim \mathcal{F}(p_{\boldsymbol{\gamma}}, T-p_{\boldsymbol{\gamma}}; T\delta_{\boldsymbol{\gamma}}), \tag{A5}$$

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

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

Solving for $\mathbb{E}[\widehat{\mathrm{SR}}^2_{\boldsymbol{\gamma}}]$ and comparing the formula with equation (A3),

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

Simple algebra gives the formula in bullet point one.

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

[19] $\delta_{\boldsymbol{\gamma}}$ is defined in the proof of Proposition 2. We will use this notation without referencing back to the previous proofs from now on.

variables. Thus, $\boldsymbol{S}_{\boldsymbol{\gamma}} = \boldsymbol{\Omega}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Omega}_{\boldsymbol{\gamma}}/T = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{\frac{1}{2}}\boldsymbol{Z}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{Z}_{\boldsymbol{\gamma}}\boldsymbol{\Sigma}_{\boldsymbol{\gamma}}^{\frac{1}{2}}/T$, and

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

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

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

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

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

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

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

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

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

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

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

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

D.4 Proof of Proposition 4

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

Lemma A2. *When the return data are generated from model $\mathcal{M}_{\boldsymbol{\gamma}_0}$, $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 = T\delta_{\boldsymbol{\gamma}} + p_{\boldsymbol{\gamma}} + O_p(\sqrt{T})$.*

Proof. According to the distribution of $\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2$ given in (A5), and applying the formula for the variance of non-central $\mathcal{F}$ distributions (Johnson et al., 1995, Page 481),

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

Thus, $\mathrm{var}[\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2]=O(1/T)$, i.e., $\mathrm{var}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2]=O(T)$. From Proposition 3, $\mathbb{E}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2]=T\delta_{\boldsymbol{\gamma}}+p_{\boldsymbol{\gamma}}+O(1)$. A simple application of the chebyshev's inequality yields $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2=\mathbb{E}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2]+O_p\left(\sqrt{\mathrm{var}[T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2]}\right)$, which proves the stated lemma. □

Lemma A3. *If $\boldsymbol{f}_{\boldsymbol{\gamma}_0}\subseteq\boldsymbol{f}_{\boldsymbol{\gamma}}$, that is, factors of model $\mathcal{M}_{\boldsymbol{\gamma}}$ subsume all factors in $\boldsymbol{f}_{\boldsymbol{\gamma}_0}$ that define the true model, we have $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2-T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2=p_{\boldsymbol{\gamma}}-p_{\boldsymbol{\gamma}_0}+O_p(1)$.*

Proof. Based on the derivations in equation (A2),

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

Similarly,

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

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 A1, 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 (A4), $\delta_{\boldsymbol{\gamma}_0}-\delta_{\boldsymbol{\gamma}}=\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top\mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0}\mid\boldsymbol{f}_{\boldsymbol{\gamma}}\right]\boldsymbol{b}_{\boldsymbol{\gamma}_0}$. 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

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

Finally, turning back to the sample version squared Sharpe ratios, from part 2 of Propo-

sition 3, $(T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2) = [1 + O_p(1/\sqrt{T})](T\mathrm{SR}_{\boldsymbol{\gamma}}^2)$ for all $\boldsymbol{\gamma}$, thus

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

□

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

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

Case I. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, that is, there are factors from the true model that are not considered under model $\mathcal{M}_{\boldsymbol{\gamma}}$. From lemma A2,

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

According to equation A4, $\delta_{\boldsymbol{\gamma}} - \delta_{\boldsymbol{\gamma}_0} = -\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right] \boldsymbol{b}_{\boldsymbol{\gamma}_0} < 0$ if $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\gamma}_0 \neq \mathbf{0}$ (here we also implicitly assume that every pair of factors is not perfectly correlated). Then $T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 - T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2 = -(\boldsymbol{b}_{\boldsymbol{\gamma}_0}^\top \mathrm{var}\left[\boldsymbol{f}_{\boldsymbol{\gamma}_0} \mid \boldsymbol{f}_{\boldsymbol{\gamma}}\right] \boldsymbol{b}_{\boldsymbol{\gamma}_0})T + O(1) + o_p(T) \xrightarrow{p} -\infty$ as $T \to \infty$. As a result, $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) \xrightarrow{p} 0$.

Case II. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}$, that is, all factors from the true model are considered under model $\mathcal{M}_{\boldsymbol{\gamma}}$. Lemma A3 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.[20]

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

[20] 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.

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

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

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

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

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

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

D.5 Proof of Proposition 5

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 $\mathrm{SR}^2_{\max}$, $\boldsymbol{\Sigma}$, 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 $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ and $p_{\boldsymbol{\gamma}}$:

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

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

□

D.6 Proof of Proposition 6

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

$$\mathrm{BF}(\boldsymbol{\gamma},\, \boldsymbol{\gamma}_0) = \underbrace{\exp\left(\frac{T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2 - T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2}{2}\right)}_{I_1} \underbrace{\left(\frac{\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2}{\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2}\right)^{-\frac{p_{\boldsymbol{\gamma}}+2}{2}}}_{I_2} \underbrace{\left(\frac{T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2}{2}\right)^{\frac{p_{\boldsymbol{\gamma}_0}-p_{\boldsymbol{\gamma}}}{2}}}_{I_3} \underbrace{\frac{\underline{\Gamma}\left(\frac{p_{\boldsymbol{\gamma}}+2}{2},\, T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}}^2/2\right)}{\underline{\Gamma}\left(\frac{p_{\boldsymbol{\gamma}_0}+2}{2},\, T\widehat{\mathrm{SR}}_{\boldsymbol{\gamma}_0}^2/2\right)}}_{I_4}. \tag{A6}$$

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

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

thus $I_2 = O_p(1)$.

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

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

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

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

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

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

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

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

Case II. If $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}$, Lemma A3 implies that $I_1 = O_p(1)$. As it is always true that $p_{\boldsymbol{\gamma}} > p_{\boldsymbol{\gamma}_0}$ when $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \subset \boldsymbol{f}_{\boldsymbol{\gamma}}$, $I_3 \xrightarrow{p} 0$. As a result, $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) \xrightarrow{p} 0$.

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

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

□

D.7 Proof of Proposition 7

Proof. Now the true model is defined by factors in $\boldsymbol{f}_0$ and $\boldsymbol{f}_{\boldsymbol{\gamma}_0} = \boldsymbol{f}_0 \cap \boldsymbol{f}$. We define $\boldsymbol{C}_{\boldsymbol{\gamma}} = \mathrm{cov}\left[\boldsymbol{f}_{\boldsymbol{\gamma}}, \boldsymbol{f}_0\right]$, $\boldsymbol{\Sigma}_0 = \mathrm{var}[\boldsymbol{f}_0]$. Applying Lemma A2, we have that for any model $\mathcal{M}_{\boldsymbol{\gamma}}$,

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

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

We now state and prove the following lemma:

Lemma A4. *Under the assumptions of Proposition 7, for any $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$ as $T \to \infty$.*

Proof. For the ease of exposition, we first define $\widetilde{\boldsymbol{f}}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\gamma}_0} \cap \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{f}}_{\boldsymbol{\gamma}_0} = \boldsymbol{f}_{\boldsymbol{\gamma}_0} \backslash \widetilde{\boldsymbol{f}}_{\boldsymbol{\gamma}}$. Now consider

the model, namely $\mathcal{M}_{\boldsymbol{\gamma}'}$, defined by factors $\boldsymbol{f}_{\boldsymbol{\gamma}'}^\top = (\boldsymbol{f}_{\boldsymbol{\gamma}}^\top, \widehat{\boldsymbol{f}}_{\boldsymbol{\gamma}_0}^\top)$. Noticing that

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

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

The Bayes factor $\mathrm{BF}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$, according to equation (A6) (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. $I_4 \xrightarrow{p} [(p_{\boldsymbol{\gamma}'}/2)(p_{\boldsymbol{\gamma}'}/2-1)(p_{\boldsymbol{\gamma}'}/2-2)\cdots(p_{\boldsymbol{\gamma}'}/2+2)(p_{\boldsymbol{\gamma}'}/2+1)]^{-1}$, that is, $I_4 = O_p(1)$.

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

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

□

Going back to Proposition 7, 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 $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$. □

D.8 Proof of Proposition 8

Proof. Based on Lemma A4, for any $\boldsymbol{\gamma}$ such that $\boldsymbol{f}_{\boldsymbol{\gamma}_0} \nsubseteq \boldsymbol{f}_{\boldsymbol{\gamma}}$, $\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] \xrightarrow{p} 0$, which is equivalent to

$$\lim_{T\to\infty} \text{Prob}_0 \left\{\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] < \varepsilon\right\} = 1, \qquad \forall 0 < \varepsilon < 1, \tag{A7}$$

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

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

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

Thus,

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

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

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

according to A7. Then, as $T \to \infty$, we must have

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

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