Model Uncertainty in the Cross Section *

Jiantao Huang† Ran Shi‡

August 2022

Abstract

We develop a transparent Bayesian approach to quantify uncertainty about linear stochastic discount factor models. Under our framework, posterior model probabilities increase with maximum in-sample Sharpe ratios and decrease with model dimensions. The entropy of posterior probabilities represents model uncertainty. We show that model uncertainty displays massive independent variations from popular uncertainty proxies, and episodes of heightened model uncertainty coincide with major market events. These observations hold not only in US markets but also in European and Asian Pacific equity markets. Moreover, positive model uncertainty shocks relate to sharp outflows from US equity mutual funds but significant inflows to government bond funds, with effects persisting for three years. Finally, the survey data suggests that investors tend to be more pessimistic about equity performance during periods of higher model uncertainty.

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

* Any errors or omissions are the responsibility of the authors. For helpful comments, discussions, and suggestions, we thank Svetlana Bryzgalova, Thummim Cho, Christian Julliard, Dong Lou, Ian Martin, Cameron Peng, Paul Schneider, Gustavo Schwenkler, Fabio Trojani, and conference participants at LSE, SFS Cavalcade 2022, and Asian Meeting of Econometric Society 2022.

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

Introduction

Recent literature has provided a wide spectrum of real and financial uncertainty measures.[1] They display pronounced time-series variations, and their innovations are associated with business cycle fluctuations through impacts on firms' investment and hiring activities. A relatively understudied direction is the measure of uncertainty that directly matters to firms' investors.

This paper provides a novel measure of cross-sectional uncertainty in equity markets and examines its implications for investors' asset allocation decisions. Existing equity market uncertainty measures focus on volatilities (as well as their unexpected innovations) of aggregate market returns. These measures do not take into account a fundamental challenge equity investors face – identifying factors that determine the cross-section of expected returns. This challenge is more demanding nowadays as investors cohabit with the "factor zoo:" too many factors have been proposed.[2] If we interpret existing financial uncertainty measures as "time-series uncertainty," a missing dimension is the cross-section: In addition to uncertainty about which direction the market is going, there is also enormous uncertainty about which factor strategies will outperform.

Uncertainty has ambiguous implications for investors' asset allocation decisions. The conventional wisdom of "flight-to-safety" or "flight-to-liquidity" claims that investors respond to large uncertainty shocks by curtailing risk exposures or hoarding liquid assets.[3] However, uncertainty may also arise endogenously during periods of "Schumpeterian growth," when new technologies lead to unforeseeable disruptive industry dynamics. Instead of seeking safety or liquidity, investors chase glamour stocks in search of a new El Dorado. Examples include the "railway mania" in the mid 1840s and the "tech bubble" in the late 1990s.[4]

This paper addresses these questions taking the perspective of a Bayesian investor. Specif-

[1] Bloom (2009) constructs uncertainty shocks using jumps in the (price) of stock market volatilities. Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2021) construct and compare real, macro and financial uncertainty indices. Baker, Bloom, and Davis (2016) develop economic policy uncertainty indices based on news coverage. Manela and Moreira (2017) use textual analysis and machine learning methods to extrapolate the VIX index back in history.

[2] See Cochrane (2011), Harvey et al. (2016).

[3] The underlying drivers include institutional redemption pressures (Vayanos, 2004), preferences featuring robustness concerns (Caballero and Krishnamurthy, 2008), and asymmetric information (Guerrieri and Shimer, 2014).

[4] This argument is related to the literature on "growth options." See, for example, Abel (1983); Pástor and Veronesi (2006, 2009).

ically, we adopt the linear stochastic discount factor (SDF) models to price assets. Investors are not clairvoyant as they do not know the "true" model. Instead, they learn *both* model parameters and specifications through Bayesian updating. One benefit of the Bayesian framework is that not only do we infer the key economic quantities (i.e. factors' risk premia) but also measure uncertainty about the "true" SDF model.

We define cross-sectional uncertainty regarding linear SDF models as the entropy of posterior model probabilities. We dub this measure *model uncertainty*. The intuition is straightforward. Suppose that there are only two candidate factor models, and we are uncertain about which one is true. One extreme case is that the first model dominates the other with a high posterior probability, i.e., 99%. Under this scenario, entropy is close to its lower bound zero (and we are clearly facing low uncertainty). On the contrary, if the two models' posterior probabilities are 50-50, the entropy reaches its maximum (picking a model boils down to the exercise of coin tossing). To sum up, the higher the entropy measure is, the more uncertain Bayesian investors are about factor models.

Our key innovation is a generalized g-prior along the line of Zellner (1986), from which Bayesian investors update their posterior beliefs. As originally exposited in Zellner (1986), the g-prior is a natural prior choice in a sequential decision-making setup. We revisit Zellner's original arguments and tailor his g-prior for linear SDF models.

We emphasize three features of our Bayesian inference. First, our prior specification naturally rules out extremely high Sharpe ratio investment opportunities, guaranteeing the absence of (near) arbitrage as pointed out in Ross (1976); Cochrane and Saa-Requejo (2000); Kozak, Nagel, and Santosh (2018). Moreover, drawing inferences based on our prior leads to well-defined posterior probabilities of asset pricing models. These derived posteriors address the indeterminacy issue induced by "flat" priors as emphasized by Chib, Zeng, and Zhao (2020). Finally, the posterior model probabilities have intuitive closed-form solutions: They increase with model-implied (in-sample) Sharpe ratios and decrease with model dimensions. The result crystallizes two competing forces for an asset pricing model to be (optimally) chosen by a Bayesian investor: higher in-sample profits (on paper or in back tests) and model simplicity.

We then examine the behaviors and implications of our cross-sectional model uncertainty measure. We document four sets of empirical findings, summarized as follows.

First, we measure uncertainty regarding 14 popular factor strategies in the US stock mar-

ket. As Figure 1 indicates, model uncertainty displays considerable time-series variations and countercyclical behaviours. We also use regression analyses to show that the variations in model uncertainty are significantly independent of commonly used proxies for uncertainty. Particularly, model uncertainty increases *before* stock market crashes and peaks under tumultuous market conditions. It reaches its upper bound at the bust of the dot-com bubble and the 2008 global financial crisis. In other words, posterior model probabilities are almost equalized during these two periods: all models are wrong (or right, which does not make any difference). Under extreme market conditions, investors do not only encounter higher second-moment (volatility) and third-moment (skewness) risk, but they are also confronted with higher (if not the highest) model uncertainty, i.e., they are incredibly uncertain about which model can help navigate them out of the storm.

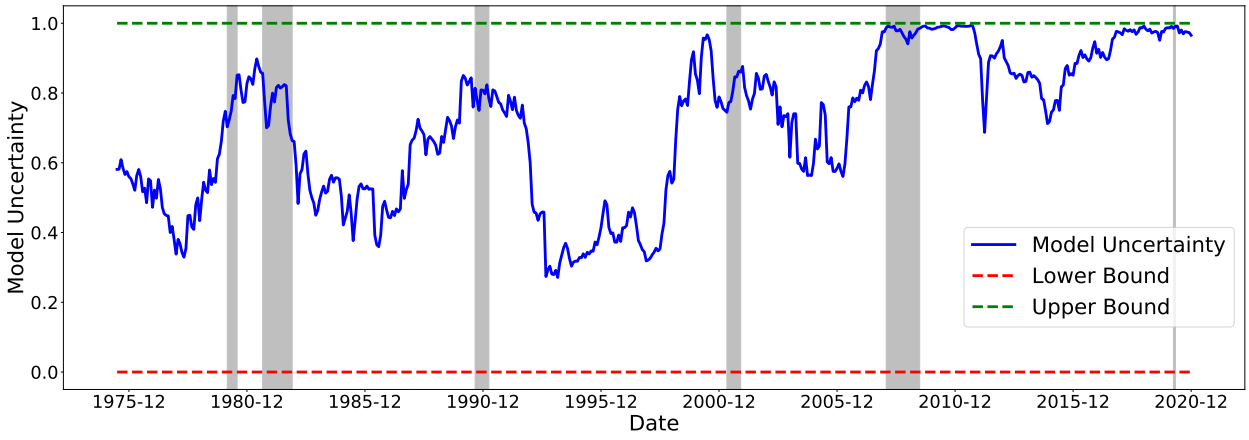


Figure 1: Time-Series of Model Uncertainty (3-Year Rolling Window)

The figure plots the time-series of model uncertainty about the linear stochastic discount factor (SDF). We consider 14 prominent factors from the past literature (see Section 2 for details). 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. The sample ranges from July 1972 to December 2020. Since we use a three-year rolling window, the model uncertainty index starts from June 1975. The red line and green lines show the lower (0) and upper bounds (1) of model uncertainty. Shaded areas are NBER-based recession periods for the US.

We repeat the exercise in European and Asian Pacific stock markets. While the time-series pattern in Europe is roughly the same as the US stock market, the Asian Pacific equity market displays certain unique behaviours. For example, model uncertainty in this market is exceptionally high during the 1997 Asian financial crisis.

Second, we show the time-varying importance of Bayesian model averaging (BMA) in

portfolio choice. Following past literature (e.g., Barillas and Shanken (2018)), we use as the criterion the out-of-sample (OOS) Sharpe ratio implied by factor models. We split the full sample into three equal subsamples based on model uncertainty and denoted them as low, middle, and high model uncertainty dates. In particular, we compare BMA with the top one model ranked by posterior model probabilities. The critical observation is that BMA outperforms the top model only in high model uncertainty dates, whereas they have almost identical performance in other periods. These results underscore the importance of aggregating the information over the space of all models instead of selecting a specific high probability model when model uncertainty is relatively heightened.

Third, model uncertainty is a crucial determinant of mutual fund flows, regardless of being an exogenous cause or a merely propagating mechanism. We adopt the canonical Vector Autoregression (VAR) model to study the dynamic responses of fund flows to uncertainty shocks. Most strikingly, model uncertainty innovations induce sharp outflows from the US equity funds and inflows to US government bond funds, with effects persisting for around three years. These outflows mainly come from small-cap and style funds but not large-cap or sector funds. In addition, we do not observe significant inflows to money market funds, so there is little evidence of "flight-to-liquidity" following high model uncertainty. Hence, investors' asset allocation decisions tend to respond to our uncertainty measure consistent with the conventional wisdom of "flight-to-safety": Facing high cross-sectional uncertainty, they reduce risky asset positions, especially in small-cap stocks and actively-managed (style) funds, and reallocate proceeds into safe assets such as government bonds.

It is also worth noting that similar fund flows patterns do not emerge when using volatility-driven uncertainty indices such as VXO and financial uncertainty. We document some evidence that VXO and financial uncertainty innovations relate to future inflows to money market funds, consistent with "flight-to-liquidity." However, dynamic responses of fund flows to these two uncertainty measures tend to be transitory and sensitive to identification assumptions. In contrast, those to model uncertainty shocks are persistent and robust. As a final caveat, while we find that uncertainty measures are associated with large fund inflows/outflows, we caution that our results are silent on whether uncertainty is the cause or effect of such fund flows.

Fourth, we find that high cross-sectional model uncertainty is associated with investors' expectations and confidence about the stock market. We quantify investors' expectations

using surveys from the American Association of Individual Investors (AAII) and their confidence levels using the Investor Behavior Project at Yale University. When our uncertainty measure goes up, both individual and institutional investors become more pessimistic about the stock market. More intriguingly, individual investors tend to "react" more aggressively (in terms of pessimism) to our cross-sectional uncertainty measure.

Literature. Our paper belongs to three strands of literature. The first is on new methodologies for factor models, with a particular focus on adapting and refining methods from the machine learning and high-dimensional statistics literature (see Giglio, Kelly, and Xiu (2021) for a review of recent advancements). The most related paper to ours is Kozak, Nagel, and Santosh (2020), which carefully examines sparsity under the linear SDF framework. They conclude that a sparse linear SDF model constructed from characteristics-sorted factors cannot explain the cross-sectional variation out-of-the-sample. Complementing their findings, our model uncertainty index further rules out the possibility of any dominant model (even "dense" ones with many factors). However, we also find that a smaller number of parsimonious models do perform well in explaining the cross-section of expected returns within specific sample periods.

There is an increasing interest in developing uncertainty measures for both real (e.g., Bloom (2009); Baker et al. (2016); Jurado et al. (2015)) and financial activities (e.g., Manela and Moreira (2017)). Dew-Becker and Giglio (2021) propose a cross-sectional uncertainty measure using the long history of option prices, which can be interpreted as uncertainty about general firm outcomes. Our contribution to the literature is to propose a conceptually new index capturing equity investors' uncertainty about factor models describing the cross-section of returns.

Our paper also contributes to the literature on Bayesian inferences about factor models and Bayesian portfolio choices (Kandel and Stambaugh, 1996; Barberis, 2000; Pástor, 2000; Avramov, 2002; Barillas and Shanken, 2018; Chib et al., 2020). The use of Zellner's g-priors for parameter uncertainties in time-series return regressions first appears in Kandel and Stambaugh (1996). Avramov (2002) extends their framework to account for both parameter and model uncertainty. In comparison, we assign the g-prior to factor loadings of linear SDF models. We develop an approach for hyper-prior tuning, addressing the issue of artificially favouring sparse models that may appear in existing Bayesian methods. Our posteriors are analytically tractable, which depend only on in-sample maximal Sharpe ratios and model

dimensions, recasting the classical GRS test intuition (Gibbons, Ross, and Shanken, 1989) in light of the factor zoo. Empirically, the existing Bayesian factor model and portfolio choice literature focus on normative questions: what should a Bayesian investor do (in terms of choosing factors and building portfolios)? On the other hand, we take a positive perspective and examine the implications of the Bayesian view (on model uncertainties) for the aggregate flows of funds.

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.[5] A subset of these excess returns would be regarded as asset pricing factors that drive the whole cross section of $\mathbb{E}[\boldsymbol{R}]$. Common examples of the factors are the market excess return in the CAPM and long-short portfolios in multi-factor asset pricing models. We use the notation $\boldsymbol{f}$ ($\boldsymbol{f} \subset \boldsymbol{R}$), a random vector of dimension p ($p \leq N$), to represent these factors.[6] A linear factor model for excess returns in the stochastic discount factor (SDF) form can be written as (see Chapter 13 of Cochrane (2005) for a detailed exposition):

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

$$\mathbb{E}[\boldsymbol{R} \times m] = \boldsymbol{0}, \tag{2}$$

or equivalently,

$$\mathbb{E}[\boldsymbol{R}] = \text{cov}[\boldsymbol{R}, \boldsymbol{f}]\boldsymbol{b}, \tag{3}$$

where m is an SDF such that the prices of excess returns all equal zero. Since the pricing equation (2) is scale-invariant, we normalize the constant term in equation (1) for the linear SDF to one. The covariance term, $\text{cov}[\boldsymbol{R}, \boldsymbol{f}]$, is an $N \times p$ matrix. Its entry in the ith row and jth column is the covariance between excess return R_i ($i = 1, \ldots, N$) and factor f_j ($j = 1 \ldots, p$).

[5] Excess returns in our context can be returns on risky assets less the risk-free rate, and more generally, returns on long-short portfolio positions with zero initial costs.

[6] We intentionally let the factors $\boldsymbol{f}$ be a subset of excess returns $\boldsymbol{R}$ to enforce that factors themselves are correctly priced, that is, their price being zero, by the factor models we write down next.

1.1 A framework incorporating model uncertainty

Now we would like to formalize the concept of model uncertaixnty. A priori, we do not know factors that enter the SDF. As a result, for a given set of p factors $\boldsymbol{f} = [f_1, \ldots, f_p]^\top$, a total number of 2^p models for the linear SDF are 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 into the linear SDF, while with $\gamma_j = 0$ 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

$$m_{\boldsymbol{\gamma}} = 1 - (\boldsymbol{f}_{\boldsymbol{\gamma}} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}}])^\top \boldsymbol{b}_{\boldsymbol{\gamma}} \tag{4}$$

and the expected excess returns are such that

$$\mathbb{E}[\boldsymbol{R}] = \operatorname{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}}. \tag{5}$$

The two equations above are counterparts of (1) and (3) after incorporating model uncertainty. We define $p_{\boldsymbol{\gamma}} = \sum_{j=1}^{p} I_{\{\gamma_j = 1\}}$, which is number of factors that are included under 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, p. 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 interests is the factor risk premia $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_p]^\top$. Under model $\mathcal{M}_{\boldsymbol{\gamma}}$, $\boldsymbol{\lambda} = \operatorname{cov}[\boldsymbol{f}, \boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}}$. Clearly, for factors that *do not* enter the SDF (their market prices of risk being zero), their risk premia are not necessarily zero. There is no clear theory guidance to introducing the latent variable $\boldsymbol{\gamma}$ for the risk premia. Knowing whether factors' risk premia equal zero or not *does not* help distinguish factor models.

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. With observed data

for excess returns, denoted by $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^T$, our primary goal is to evaluate the probability of each model $\mathcal{M}_{\boldsymbol{\gamma}}$ given the observed data $p[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$. Bayesian inference offers a natural way of computing these posterior model probabilities.

We (as have many others) assume that the observed excess returns are generated from a multivariate Gaussian distribution:

$$\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{6}$$

The linear SDF model $\mathcal{M}_{\boldsymbol{\gamma}}$ then sets a restriction on this distribution through the following moment condition:

$$\boldsymbol{\mu} = \boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{b}_{\boldsymbol{\gamma}}, \tag{7}$$

where $\boldsymbol{C}_{\boldsymbol{\gamma}} = \text{cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]$ consists of a subset of columns in $\boldsymbol{\Sigma}$. We adopt an empirical Bayes strategy by treating the variance-covariance matrix $\boldsymbol{\Sigma}$ as known initially to derive the posterior model probability $p[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$, and then substituting this matrix with a moment estimator[7].

Now we proceed to assign priors for $\boldsymbol{b}_{\boldsymbol{\gamma}}$. Our prior specification is motivated by the g-prior proposed by Arnold Zellner (see Zellner (1986)). We assume that *conditional* on choosing model $\mathcal{M}_{\boldsymbol{\gamma}}$,

$$\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}} \sim \mathcal{N}\left(\boldsymbol{0}, \frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right), \quad g > 0 \tag{8}$$

where T is the sample size for the observed excess returns. The parameter g is related to the effective sample size or level of uncertainty for an "conceptual or imaginary sample" according to Zellner (1986).

Following the reasoning of Zellner (1986), we generalize the original g-prior and adapt it to our specific setting. Before making inference about different linear SDF models using the observed excess return data $\mathcal{D}$, we consider an "imaginary" sample of size T', denoted by $\mathcal{D}' = \{\boldsymbol{R}'_t\}_{t=1}^{T'}$, where the sample size is allowed to be different from T by a scalar g such that $T' = T/g$. This parameter g also governs level of uncertainty about our imaginary sample

[7]Empirical Bayes approaches use data to facilitate prior assignments. Here although the matrix $\boldsymbol{\Sigma}$ is a likelihood parameter, it also enters the prior for $\boldsymbol{b}_{\boldsymbol{\gamma}}$, as will become clear next when we introduce our prior specification. Thus we are still using data to pin down (hyper)parameters in the priors. The use of moment estimators to replace parameters in the prior distributions dates all the way back to the seminal James-Stein estimator (James and Stein (1961)). For a monograph on modern empirical Bayes methods, see Efron (2012).

relative to the data sample we have [8]. 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 non-informative prior on $\boldsymbol{b}_{\boldsymbol{\gamma}}$, which is flat everywhere[9], we can derive the "posterior" of $\boldsymbol{b}_{\boldsymbol{\gamma}}$ given this conceptual data sample as $[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}'] \sim \mathcal{N}\left(\boldsymbol{b}'_{\boldsymbol{\gamma}}, g/T \times \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right)$, where the posterior mean $\boldsymbol{b}'_{\boldsymbol{\gamma}}$ is related to the particular hypothetical data set $\mathcal{D}'$ in mind, while the posterior variance is not (a celebrated result for conditional normal distributions). This leaves the posterior mean $\boldsymbol{b}'_{\boldsymbol{\gamma}}$ largely undetermined for we can have infinite degrees of freedom "imagining" the data set $\mathcal{D}'$. If we would like to use this posterior as our prior for $\boldsymbol{b}_{\boldsymbol{\gamma}}$, resorting to the Bayesian philosophy that "today's posterior is tomorrow's prior" (Lindley, 2000), we at lease need to find a way of determining $\boldsymbol{b}'_{\boldsymbol{\gamma}}$, the current posterior mean.

Zellner (1986) relies on the rational expectation hypothesis to pin down $\boldsymbol{b}'_{\boldsymbol{\gamma}}$. Suppose that we have an anticipatory value for $\boldsymbol{b}_{\boldsymbol{\gamma}}$, denoted by $\boldsymbol{b}^a_{\boldsymbol{\gamma}}$, in addition to the imaginary sample $\mathcal{D}'$ (as well as the initial diffuse prior for $\boldsymbol{b}_{\boldsymbol{\gamma}}$). The rational expectation hypothesis says that $\boldsymbol{b}^a_{\boldsymbol{\gamma}} = \mathbb{E}[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}'] = \boldsymbol{b}'_{\boldsymbol{\gamma}}$. Now we have a reference informative prior distribution that does not depend on the hypothetical sample, which is

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

To determine whether a model $\mathcal{M}_{\boldsymbol{\gamma}}$ is sensible or not, we are basically testing $H_0 : \boldsymbol{b}_{\boldsymbol{\gamma}} = \mathbf{0}$ versus $H_1 : \boldsymbol{b}_{\boldsymbol{\gamma}} \in \mathbb{R}^{p_{\boldsymbol{\gamma}}}$. These tests help us distinguish between different models as model $\mathcal{M}_{\boldsymbol{\gamma}}$ already impose the condition that $\boldsymbol{b}_{-\boldsymbol{\gamma}} = \mathbf{0}$. Following the suggestion of Zellner (1986), we set $\boldsymbol{b}^a_{\boldsymbol{\gamma}} = \mathbf{0}$, that is, the anticipatory expectations are the values under the null. This finally gives us the prior specification in (8).

[8] 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 v.s. variance of the hypothetical data set) are isomorphic because they will lead to the same g-prior specification. Our sample-size based arguments echo the ideas of factional and intrinsic Bayes factor in the mid 90's (see O'Hagan (1995) and Berger and Pericchi (1996)), which aim to "transform" improper priors to proper ones. Similar ideas for specifying priors are adopted in the paper by Shmuel Kandel and Robert F. Stambaugh in the finance literature to discipline the specification of informative priors Kandel and Stambaugh (1996).

[9] This flat prior is non-informative in the sense that it is a Jeffreys prior, a common notion of prior objectiveness or non-informativeness in Bayesian analysis Jeffreys (1946). Under our setting, we treat $\boldsymbol{\Sigma}$ as known. As a result, Jeffreys prior for $\boldsymbol{b}_{\boldsymbol{\gamma}}$ is proportional to a constant, i.e., it is flat. Of remark, this flatness outcome is not true if the covariance matrix is unknown, under which the Jeffreys prior would specify that the joint density of $\pi(\boldsymbol{b}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma})$ is proportional to $\boldsymbol{\Sigma}^{-\frac{N+2}{2}}$. Some existing work (e.g. Barillas and Shanken (2018)) specifies a prior such that $\pi(\boldsymbol{b}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}) \propto \boldsymbol{\Sigma}^{-\frac{N+1}{2}}$, which is the so-called independence Jeffreys prior (*not* the original Jeffreys-rule prior) imposing the assumption that $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\Sigma}$ are independent at the prior level.

Remark. One might attempt to assign an objective prior, such as the Jeffreys prior, to $\boldsymbol{b}_{\boldsymbol{\gamma}}$. In this case, it is an improper flat prior as we have discussed early on. This would be desirable without model uncertainty, for it will lead to proper posterior distributions. However, with model uncertainty, improper priors can only be assigned to *common* parameters across models, which is clearly not the case for $\boldsymbol{b}_{\boldsymbol{\gamma}}$. Otherwise, posterior model probability would be indeterminate. This is a well-known result in Bayesian statistics and has also been pointed out in the finance literature (e.g., Cremers (2002)).

Our g-prior specification in (8) leads to a surprisingly simple expression for the variance of the SDF, which is summarized in Proposition 1.

Proposition 1. *(prior property: implications to the variance of SDF) Under model $\mathcal{M}_{\boldsymbol{\gamma}}$, in which $m_{\boldsymbol{\gamma}} = 1 - (\boldsymbol{f}_{\boldsymbol{\gamma}} - \mathbb{E}[\boldsymbol{f}_{\boldsymbol{\gamma}}])^{\top} \boldsymbol{b}_{\boldsymbol{\gamma}}$, the g-prior specification for $\boldsymbol{b}_{\boldsymbol{\gamma}}$ implies that*

$$\operatorname{var}[m_{\boldsymbol{\gamma}} \mid g] = \frac{g p_{\boldsymbol{\gamma}}}{T}.$$

According to Proposition 1, volatility of the SDF ($= \sqrt{g p_{\boldsymbol{\gamma}}/T}$) under a certain model is determined by the conditionality of that model, at least at the prior level. The renowned Hansen-Jagannathan bound states that this volatility (times the gross risk-free rate) sets an upper bounds on any achievable Shape ratios in the economy Hansen and Jagannathan (1991); Cochrane and Saa-Requejo (2000) regards portfolio positions with high Sharpe ratios as deals that are too good to be realized in the market. These arguments imply that models with too many factors are not likely to be realistic *a priori*.

The g-prior offers us an analytically tractable framework to make posterior inference. Under the g-prior, we can integrate out $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and calculate the marginal likelihood of observing the excess return data $\mathcal{D}$ based on each model. All these marginal likelihoods are available in closed form and results are collected in Proposition 2.

Proposition 2. *The likelihood of observing excess return data $\mathcal{D}$ under model $\mathcal{M}_{\boldsymbol{\gamma}}$ is*

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

where

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

is the in-sample variance-covariance matrix for the excess returns; SR^2_{max} *is the maximal squared Sharpe ratio achievable from forming portfolios using all excess returns under consideration;* $\text{SR}^2_{\boldsymbol{\gamma}}$ *is the maximal squared Sharpe ratio from combining all factors under model* $\mathcal{M}_{\boldsymbol{\gamma}}$*. These two Sharpe ratios are both in-sample values and it is always the case that* $\text{SR}^2_{\boldsymbol{\gamma}} \leq \text{SR}^2_{\text{max}}$ *for all* $\boldsymbol{\gamma}$*.*

Proposition 2 has a couple of implications. To begin with, we can calculate the marginal likelihood for a very special model, the null model, in which $\boldsymbol{\gamma} = \mathbf{0}$. SDF $m_{\boldsymbol{\gamma}}$ in this case is a constant, characterizing a risk-neutral market. Under this setup, $p_{\boldsymbol{\gamma}}$ equals zero because no factors are included, and the maximal squared Sharpe ratio $\text{SR}^2_{\boldsymbol{\gamma}}$ is also zero. Plugging these two quantities into the expression in Proposition 2, we have $p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}, g] \equiv p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}]$, because the posterior marginal likelihood under the null model does not depend on the scalar g. The Bayes factor, which is the ratio between marginal likelihoods under two different models, that compares model $\mathcal{M}_{\boldsymbol{\gamma}}$ with the null model $\mathcal{M}_{\mathbf{0}}$ is

$$\begin{aligned} \text{BF}_{\boldsymbol{\gamma}}(g) &= \frac{p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]}{p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}]} \\ &= \exp\left\{\frac{Tg}{2(1+g)}\text{SR}^2_{\boldsymbol{\gamma}} - \frac{p_{\boldsymbol{\gamma}}}{2}\log(1+g)\right\}. \end{aligned} \tag{9}$$

This Bayes factor can be regarded as evidence of model $\mathcal{M}_{\boldsymbol{\gamma}}$ against the null model. To further compare two arbitrary models $\mathcal{M}_{\boldsymbol{\gamma}}$ and $\mathcal{M}_{\boldsymbol{\gamma}'}$, we can calculate the Bayes factor

$$\begin{aligned} \text{BF}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}(g) &= \frac{\text{BF}_{\boldsymbol{\gamma}}(g)}{\text{BF}_{\boldsymbol{\gamma}'}(g)} \\ &= \exp\left\{\frac{Tg}{2(1+g)}\left(\text{SR}^2_{\boldsymbol{\gamma}} - \text{SR}^2_{\boldsymbol{\gamma}'}\right) - \frac{p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}'}}{2}\log(1+g)\right\}, \end{aligned} \tag{10}$$

which is, by definition, the (marginal) likelihood ratio $p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]/p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}'}, g]$. A large Bayes factor $\text{BF}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}(g)$ lends evidence to favor model $\mathcal{M}_{\boldsymbol{\gamma}}$ against model $\mathcal{M}_{\boldsymbol{\gamma}'}$.

A first observation based on equation (10) is that although the marginal likelihood in Proposition 2 depends on the test assets (the pre-specified set of excess returns that define $\boldsymbol{R}$), the Bayes factors do not. The Bayes factors are only determined by the in-sample time series of the factors that enter the linear SDF, through the maximal Sharpe ratios and the number of factors. A key assumption driving this outcome is that factors are a subset of the testing assets. In other words, the linear factor SDF model must price the factors themselves

correctly. This finding is reminiscent of the observation that, when estimating factor risk premia in linear factor models, the efficient GMM objective function assigns zero weights to the testing assets except for the factors entering the SDF (See for example, (Cochrane, 2005, Page 244-245)).

The Bayes factor above illustrates a clear trade-off when comparing models. With the number of factors fixed, models in which factors can generate larger in-sample Sharpe ratios are always preferred. This echoes the intuitions behind the GRS tests in Gibbons et al. (1989), which show the link between time-series tests of the factor models and the mean-variance efficiency of factor portfolios. Under our setting, when the factor portfolios deliver large maximal Sharpe ratios, it is evidence that they are more likely to span the excess return space, thus favoring the linear SDF constructed from these factors. On the other hand, it is a simple mechanical phenomenon that maximal Sharpe ratio SR_γ increases as additional assets are added into the factor portfolio. Thus the penalty term on model dimensionality p_γ imposed by the g-prior plays an key role in preventing the Bayes factor to favor large models blindly. In order to properly penalize large models, g cannot be too small, as SR_γ always increases after one augments the linear SDF.

Perhaps the most desirable feature of our Bayes factor calculation in equation (10) is that it helps us understand the aforementioned trade-off quantitatively. When model dimension is increased by one ($p_\gamma - p_{\gamma'} = 1$), the maximal squared Sharpe ratio (times the sample size T) of the factor portfolio has to increase by at least $(1+g)/g \times \log(1+g)$ to lend support to the augmented model, that is,

$$T\left(\text{SR}^2_\gamma - \text{SR}^2_{\gamma'}\right) > \frac{1+g}{g}\log(1+g).$$

However, it is always the case that $T\left(\text{SR}^2_\gamma - \text{SR}^2_{\gamma'}\right) \leq T\text{SR}^2_{\text{max}}$. Then for g large enough, the inequality above will always be violated, as the function $(1+g)/g \times \log(1+g)$ is monotonically increasing and unbounded. As a result, smaller models will always be supported by the Bayes factor. Under the extreme case that $g \to \infty$, from equation (9), $\text{BF}_\gamma(g) \to 0$. Paradoxically, the most favorable model will always be the null model. The case under which $g \to \infty$ corresponds to the conventional diffuse priors; and the fact that, with model uncertainty, diffuse priors always support the null model is sometimes called the Bartlett's paradox (Bartlett (1957)). Of note, this paradox poses another refutation to the use of improper diffuse priors under model uncertainty, in addition to posterior indeterminacy that

has been pointed out earlier.

1.3 Posterior properties

1.4 A prior for the parameter g

Discussions above point to the subtlety of choosing the parameter g. Instead of plugging in particular numbers for g, a natural way under our Bayesian framework is to integrate out g with a proper prior for it. A prior on g, namely $\pi[g]$, is equivalent to assigning a scale-mixture of g priors for $\boldsymbol{b}_{\boldsymbol{\gamma}}$. This idea is adapted from Liang et al. (2008), who argues that this type of mixture priors provides more robust posterior inference. As a result, our g prior specification will be modified to

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

where the prior for g is such that

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

This prior $\pi[g]$ is improper when $a \leq 2$. A special case when $a = 2$ corresponds to the Jeffreys prior in Liang et al. (2008). Because the marginal likelihood of the null model does not depend on g (recall that $p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}, g] \equiv p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}]$), improper priors will lead to indeterminacy in the ratio

$$\begin{aligned} \mathrm{BF}_{\boldsymbol{\gamma}} &= \frac{\int_0^\infty p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]\pi[g]\,\mathrm{d}g}{\int_0^\infty p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}]\pi[g]\,\mathrm{d}g} \\ &= \int_0^\infty \frac{p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]}{p[\mathcal{D} \mid \mathcal{M}_{\mathbf{0}}]}\pi[g]\,\mathrm{d}g \end{aligned} \tag{12}$$

up to an arbitrary constant, which is the Bayes factor under the new mixture of g prior specification. Thus we force $a > 2$.

This additional prior on g also leads to refinements on the volatility of the SDF. Based on the result from Proposition 1, the unconditional volatility of the SDF for model $\mathcal{M}_{\boldsymbol{\gamma}}$ must satisfy

$$\mathrm{var}[m_{\boldsymbol{\gamma}}] \geq \mathbb{E}[\mathrm{var}[m_{\boldsymbol{\gamma}} \mid g]] = \frac{p_{\boldsymbol{\gamma}}}{T}\mathbb{E}[g].$$

The prior $\pi[g]$ is such that $\mathbb{E}[g] = \infty$ if $a \leq 4$, and that $\mathbb{E}[g] = 2/(a-4)$ if $a > 4$. To make sure that the variance of the SDF does not explode, we need $a > 4$. And if we follows the argument of Cochrane and Saa-Requejo (2000) to set an upper limit on the maximal achievable Sharpe ratio in the economy[10], denoted by SR_∞, then

$$R_f^2 \text{SR}_\infty^2 = \text{var}[m_{\boldsymbol{\gamma}}] \geq \mathbb{E}[\text{var}[m_{\boldsymbol{\gamma}} \mid g]] = \frac{2p_{\boldsymbol{\gamma}}}{T(a-4)},$$

where R_f represents the risk-free rate. For the investor in the economy to be not risk-neutral, the SDF must include at least one factor, that is, $p_{\boldsymbol{\gamma}} \geq 1$ (for example, under the CAPM world). As a result, we will require that

$$a \geq 4 + \frac{2}{TR_f^2 \text{SR}_\infty^2}.$$

Another way of looking at our prior for g is that it is equivalent to

$$\frac{g}{1+g} \sim \text{Beta}\left(1, \frac{a}{2} - 1\right).$$

This ratio is crucial in that it determines the contribution of data evidence when making posterior inferences. It is sometimes referred to as the "shrinkage factor". To see this more clearly, we can calculate the posterior of the cross-sectional expected return $\boldsymbol{\mu} = \boldsymbol{C}_{\boldsymbol{\gamma}} \boldsymbol{b}_{\boldsymbol{\gamma}}$, which is given as follows

$$\mathbb{E}[\boldsymbol{\mu} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g, \mathcal{D}] = \frac{g}{1+g} \boldsymbol{C}_{\boldsymbol{\gamma}} \left\{\text{var}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\right\}^{-1} \left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{f}_{\boldsymbol{\gamma},t}\right).$$

Under all models, the posterior mean of expected returns are scaled by a fixed factor $g/(1+g) \in (0,1)$. Our prior specification is equivalent to a Beta distribution for this shrinkage factor, and the prior mean for it is

$$\mathbb{E}\left[\frac{g}{1+g}\right] = \frac{2}{a} \leq \frac{1}{2 + \left(TR_f^2 \text{SR}_\infty^2\right)^{-1}}.$$

In order to give enough credit to the data-driven estimates and avoid over-shrinkage, we

[10] Note that this must be larger than the maximal in-sample Sharpe ratio of portfolios formed using excess returns under our consideration, denoted by SR_{max} in Proposition 2.

choose the smallest possible a such that $\mathbb{E}\left[g/(1+g)\right]$ is as large as possible *a priori*, that is, we pick $a = 4 + 2/(TR_f^2 \mathrm{SR}_\infty^2)$. Under this choice, the prior expectation for the shrinkage factor is still strictly smaller than one half, but can be very close (the ratio $2/(TR_f^2 \mathrm{SR}_\infty^2)$ is usually very small).

1.5 Posterior probability of models

We can integrate out the parameter g according to equation (12) to find the Bayes factors under the mixture of g-priors. Proposition 3 presents the results.

Proposition 3. *The Bayes factor for comparing model $\mathcal{M}_{\boldsymbol{\gamma}}$ with the null model $\mathcal{M}_{\mathbf{0}}$ is*

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

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 263); the scalar $s_{\boldsymbol{\gamma}}$ is defined as

$$s_{\boldsymbol{\gamma}} = \frac{p_{\boldsymbol{\gamma}} + a}{2} - 1.$$

This Bayes factor is always increasing in $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ always decreasing in $p_{\boldsymbol{\gamma}}$.

The Bayes factor that compares any two models can be computed as

$$\mathrm{BF}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'} = \frac{\mathrm{BF}_{\boldsymbol{\gamma}}}{\mathrm{BF}_{\boldsymbol{\gamma}'}},$$

which is the same as what we have done earlier. Bayes factors decide the posterior odds of one model against another:

$$\frac{\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]}{\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}'} \mid \mathcal{D}]} = \frac{\pi[\mathcal{M}_{\boldsymbol{\gamma}}]}{\pi[\mathcal{M}_{\boldsymbol{\gamma}'}]} \times \mathrm{BF}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}.$$

Equivalently, the posterior odds give us the posterior model probabilities: for model $\mathcal{M}_{\boldsymbol{\gamma}}$, its posterior probability given the excess return data is

$$\mathbb{P}[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}] = \frac{\mathrm{BF}_{\boldsymbol{\gamma}} \pi[\mathcal{M}_{\boldsymbol{\gamma}}]}{\sum_{\boldsymbol{\gamma}} \mathrm{BF}_{\boldsymbol{\gamma}} \pi[\mathcal{M}_{\boldsymbol{\gamma}}]},$$

which is a direct outcome of the Bayes' rule. We can then define a model uncertainty measure as the entropy of the posterior model probabilities:

$$\text{entropy}[\mathcal{M}_\gamma \mid \mathcal{D}] = \sum_\gamma \log(p[\mathcal{M}_\gamma \mid \mathcal{D}])p[\mathcal{M}_\gamma \mid \mathcal{D}]. \tag{13}$$

Roughly speaking, larger entropy corresponds to higher model uncertainty. For example, suppose that we have only two candidate models. If one of them has a posterior model probability of 99%, we should be confident about this high-probability model. Actually, the model uncertainty is almost zero in this scenario. However, if the posterior probability of each model is around 50%, then choosing the true model is equivalent to flipping a fair coin. In this case, the model uncertainty in equation (13) is maximized.

2 Data Description

In our primary empirical implementation, we combine 14 prominent factors from the past literature and measure model uncertainty in this small zoo of factors. First, we include notable Fama-French five factors (Fama and French (2016)) plus the momentum factor (Jegadeesh and Titman (1993)). In addition, we consider the q-factor model from Hou et al. (2015) and include their size, investment, and profitability factors. The factor models mentioned earlier are based on rational asset pricing theory. Taking the insights from behavioural models, Daniel et al. (2020) propose a three-factor model consisting of the market factor, the short-term behavioural factor (PEAD), and the long-term behavioural factor (FIN). Finally, we include the HML devil, the quality-minus-junk factor, and the betting-against-beta factor from the AQR library. Appendix A presents the detailed description of these factors.

Table A1 reports the annualised mean returns and Sharpe ratios of 14 factors. First, most of them (except for two size factors) have enormous Sharpe ratios in the full sample from July 1972 to December 2020. In particular, the short-term behavioural factor (PEAD) seems to be the most profitable historically. Furthermore, I split the entire sample into two equal subsamples. Consistent with past literature (e.g., McLean and Pontiff (2016)), the performances of many factor strategies decline significantly from subsample one to two. Most strikingly, the annualised Sharpe ratio of the value factors has plunged from above 0.9 to nearly zero in the second subsample. This observation suggests that we should focus on

the out-of-sample instead of the in-sample Sharpe ratio in evaluating factor models.

With the estimate of model uncertainty, we next compare it with other uncertainty measures and economic variables. Bloom (2009) uses the jumps in VXO/VIX indices as the stock market uncertainty shocks. We download the time-series of VXO/VIX indices from Wharton Research Data Services (WRDS). Baker et al. (2016) develop indices of economic policy uncertainty (EPU), which can be downloaded from Nick Bloom's website. Other uncertainty measures that we use include the macro, real and financial uncertainty measures in Ludvigson et al. (2021) and Jurado et al. (2015). We download them from the authors' websites. In addition, we compare our model uncertainty with the intermediary factor from He et al. (2017), the term yield spread (the yield on ten-year government bonds minus the yield on three-month treasury bills), and the credit spread (the yield on BAA corporate bonds minus the yield on AAA corporate bonds). We download the intermediary factor from the authors' websites and the bond yields from the Federal Reserve Bank of St. Louis.

Moreover, 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. To normalise the aggregate fund flows, we divide the equity (fixed-income) fund flows across all funds within a specific investment objective by the total market capitalisation of all listed companies in CRSP (US GDP). In addition, we download the total market value of all US-listed stocks from CRSP.

Finally, we study the relationship between our model uncertainty measure and investors' expectations about future stock market performance. Specifically, we use the survey data from the American Association of Individual Investors (AAII) survey and Shiller's survey conducted by the International Center for Finance at the University of Yale. We download the related data from their official websites.

3 Measuring Model Uncertainty

We now adopt the perspective of Bayesian investors and construct the time series of model uncertainty. At the end of each month, we use all daily factor returns in the past three years to estimate the posterior model probabilities, $p[\mathcal{M}_\gamma \mid \mathcal{D}]$, and compute the entropy as in equation (13). We choose the hyper-parameter a to be four in the benchmark case. We also

present the results obtained from alternative rolling windows and consider other choices of a in robustness checks (see Section 7).

The behavioural 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. 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); (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 behavioural factors, momentum, QMJ, and BAB.

The blue line in Figure 1 plots the time series of model uncertainty about linear SDFs, and the sample period spans from June 1975 to December 2020. The red and green dotted lines show the lower and upper bounds of model uncertainty, respectively. The lower entropy bound is always zero, i.e., when there is one dominant model with the posterior model probability of 100%. On the contrary, uncertainty is maximized when the posterior model probabilities are equalized across all models. Because we have 14 factors, and only one of the categorical factors could be selected into the true model, there are 5,184 different candidate models.[11] This implies that the upper bound of model uncertainty is around 8.55.[12] To normalize the model uncertainty index, we divide it by 8.55. Hence, the upper bound is one in Figure 1.

The model uncertainty index has several interesting features that could shed light on the nature of uncertainty about the linear SDF. First, we observe a surprisingly high level of model uncertainty. Specifically, the average (median) model uncertainty is around 0.70 (0.75), with the first and third quartiles equal to 0.53 and 0.87, respectively. Hence, most of the time, Bayesian investors are not confident about the true SDF model. Second, model uncertainty fluctuates significantly over time. In particular, the index varies from the lowest value of 0.27 to the highest 0.99, representing economic states in which Bayesian investors find it almost unlikely to determine the true SDF model. The standard deviation of the index

[11] The model in our framework is indexed by $\boldsymbol{\gamma}$: $\gamma_j \in \{0, 1\}$ and $\gamma_j = 1$ implies that the factor j should be included into true SDF. We do not have restrictions on the market, short-term reversal, long-term reversal, momentum, QMJ, and BAB, so the number of models for these 6 factors is 2^6. For SMB and ME, we only allow three cases: (0,0), (1,0) or (0,1). Therefore, each categorical factor has 3 (instead of 4) possibilities. The total number of candidate models equals $2^6 \times 3^4 = 5184$.

[12] upper bound $= -\sum_\gamma \frac{1}{5184} \times \log(\frac{1}{5184}) = \log(5184) \simeq 8.55$..

is 0.21. Overall, model uncertainty is a dynamic phenomenon. Finally, model uncertainty is persistent by construction since we use a rolling window of 36 months in the estimation. The first-order autocorrelation is 0.98, and the autocorrelation coefficients strictly decrease in time lags, with insignificant autocorrelations after 30 lags.

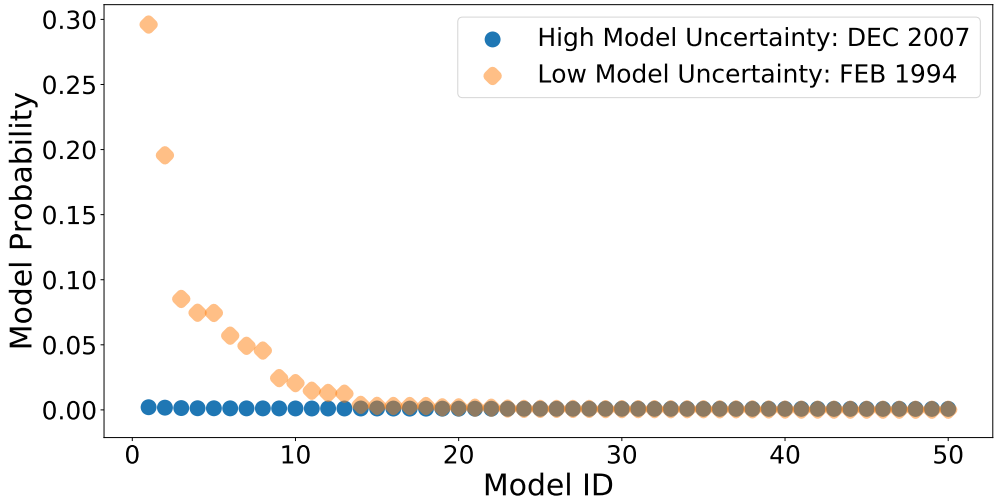


Figure 2: Posterior Probabilities of Top 50 models: High vs. Low Model Uncertainty

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).

Figure 1 also suggests the countercyclical nature of model uncertainty. In particular, the 1990s was a remarkable period: It was remembered as a period of strong economic growth, low inflation and unemployment rate, and high stock returns. During the 1990s, model uncertainty is the lowest across our sample. As the orange diamonds in Figure 2 suggest, posterior probabilities of the top two models are significantly larger than others. Hence, it is relatively straightforward to figure out the true SDF model in this period.

In addition, episodes of heightened model uncertainty tend to coincide with major events in the US stock markets and economy. Important examples include the dot-com crash in 2000 and the global financial crisis in 2008 when model uncertainty almost touches its upper bound. Specifically, the blue dots in Figure 2 show that posterior probabilities of the top 50 models, in December 2007, are almost equalized. In other words, it is virtually infeasible to distinguish models based on the observed data. The 2008 crisis is noteworthy because model

uncertainty stays at a high level for a prolonged period. In contrast, it declines shortly after other crises/recessions. In the recent five years, model uncertainty has slowly increased from 0.7 to 1 at the end of 2020.

Interestingly, we do not observe a spike in model uncertainty during the 1987 flash crash. One potential reason is that this event was not long-lasting. Even though S&P 500 index declined by more than 20% in one day, the crisis was not caused by any economic recession, and the market rebounded rapidly. Instead, the leading cause was synchronous program trading, illiquidity in the market, and the subsequent market panic. Since our uncertainty measure is based on past-three-year daily data, the impact of short-term market chaos is averaged out.

In conclusion, our model uncertainty measure displays considerable time-series variations: It is incredibly sizable in bad economic states. The stock market crash that lasts only for a short period, such as the 1987 flash crash, is not captured by model uncertainty. Furthermore, the cyclical behaviours of model uncertainty imply another layer of investment risk: When investors experience bear stock markets, they are also the most uncertain about the true model in the cross-section, or equivalently, which portfolio of factor strategies they should hold. Therefore, a natural hypothesis is that model uncertainty relates to investors' portfolio choices and expectations. We investigate these topics in section 4 and 5.

3.1 Does Model Uncertainty Matter?

Should investors take into account model uncertainty in the cross-section? A natural hypothesis is that model uncertainty plays a more critical role when it is more sizable. The logic is as follow. When model uncertainty is relatively low, the factor model with the highest model probability dominates others, such as the orange diamonds in Figure 2. Hence, investors are more willing to trust the top model ranked by the Bayesian posterior probabilities. In contrast, the top model is not informative if model uncertainty is heightened, such as during market crashes. In this case, they may prefer to aggregate the information over the space of all models.

The Bayesian model averaging (BMA) is one common approach to aggregating models. It enables us to flexibly model investors' uncertainty about potentially relevant factors. In

the SDF model, we are interested in the risk prices, $\boldsymbol{b}$. The BMA of $\boldsymbol{b}$ is defined as

$$\boldsymbol{b}_{bma} := \mathbb{E}[\boldsymbol{b} \mid \mathcal{D}] = \sum_{\gamma} \mathbb{E}[\boldsymbol{b} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}] \times P(\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}). \tag{14}$$

Rather than considering the expectation of $\boldsymbol{b}$ conditional on a specific model, we take the weighted average of the model-implied expectations, where the weights are posterior model probabilities. Intuitively, models with high probabilities are more influential in BMA.

BMA deviates sharply from the traditional model selection, in which researchers always use a particular criterion (e.g., adjusted R^2, model probabilities, etc.) to select a single model and presume that the selected model is correct. Past literature also shows the importance of model averaging in asset pricing (e.g., Avramov (2002), Bryzgalova et al. (2021), and Avramov et al. (2021)).

We now compare the performance of BMA with the top Bayesian model. The performance metric that we use is the out-of-sample (OOS) Sharpe ratio of factor models. We also compare our Bayesian procedure with several candidate models: (1) All 14 factors (All), (2) Carhart (1997) four-factor model (Carhart4), (3) Fama and French (2016) five-factor model (FF5), (4) Hou et al. (2015) q-factor model (HXZ4), and (5) Daniel et al. (2020) behavioural factor model (DHS3).

For each factor model $\boldsymbol{\gamma}$ in month t, we estimate the risk prices of $\boldsymbol{f}_{\boldsymbol{\gamma}}$ via the standard GMM estimation: $\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}} = (\text{var}[\boldsymbol{f}_{\boldsymbol{\gamma}}])^{-1}(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{f}_{\boldsymbol{\gamma t}})$, where the covariance matrix and mean returns of $\boldsymbol{f}_{\boldsymbol{\gamma}}$ are estimated using the data from month $t-35$ to month t, consistent with Figure 1. The tangency portfolio conditional on model $\boldsymbol{\gamma}$ is $\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}}^{\top} \boldsymbol{f}_{\boldsymbol{\gamma},\boldsymbol{t+1}}$, and the BMA tangency portfolio is $\boldsymbol{b}_{bma}^{\top} \boldsymbol{f}_{\boldsymbol{t+1}}$.[13] We update the tangency portfolio each month.[14]

In addition, we use the non-parametric Bootstrap to test the null hypothesis that BMA and the model $\boldsymbol{\gamma}$ have an identical Sharpe ratio, i.e., $H_0 : \text{SR}_{bma}^2 = \text{SR}_{\boldsymbol{\gamma}}^2$. Under H_0, the expected return of the tangency portfolio implied by the model $\boldsymbol{\gamma}$ is linear in that of BMA: $\mathbb{E}[R_t^{\gamma}] = \mathbb{E}[R_t^{bma}]\sigma(R_t^{\gamma})/\sigma(R_t^{bma})$. We adjust the average return of R_t^{γ} using the previous equality and draw 100,000 sample paths of $\{R_{t^\star}^{\gamma}, R_{t^\star}^{bma}\}_{t^\star=1}^{T}$ with replacement, where T is the sample size in 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

[13] For model $\boldsymbol{\gamma}$, we scale the tangency weights $\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}}$ each month such that the target monthly portfolio volatility is 1% based on historical data from month $t-35$ to month t.

[14] Moreover, the top Bayesian model (with the highest model probability) is time-varying.

datasets, we claim that H_0 is rejected by the data at 10% (5%, 1%) significance level.

Table 1: Out-of-Sample Model Performance

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

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

We start with describing the full-sample performance, as shown in the first row of Table 1. First, our Bayesian procedure successfully selects the model that outperforms traditional factor models in the out-of-sample. The top Bayesian model (see column (2)) has an OOS Sharpe ratio of 1.75, which is virtually comparable to the model composed of all 14 factors (see column (3)). Second, BMA beats the top Bayesian model, but the distinction is merely marginal in the economic sense.

One may be concerned that these 14 factors are data-mined, so choosing the top model only reflects data snooping rather than the outperformance of our Bayesian procedure. We further split the whole sample into three equal subsamples to tackle this concern. Consistent with past literature, the performance of factor models tends to decline over time, and the drops in Sharpe ratios are particularly enormous from subsample II (September 1990 - October 2005) to subsample III (November 2005 - December 2020). In addition, BMA is more valuable in the third subsample: its Sharpe ratio (1.106) is significantly higher than other models except for the one composed of all 14 factors.

Whether the performance of factor models is related to model uncertainty? The short answer is yes. On average, the performance of factor models declines as model uncertainty increases. Specifically, when model uncertainty is low, both the top model and BMA have similar Sharpe ratios of around 2.57, which are exceptionally high. In other words, investors should be confident about the top model chosen by our Bayesian procedure in low uncertainty states. 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, significantly more profitable than any other specifications.

In summary, Table 1 reveals two important findings. First, our Bayesian procedure is competent in picking the model that has satisfactory OOS performance. Second, it underscores the importance of considering model uncertainty when it is relatively high. In this scenario, BMA, which aggregates the information across all models, is salient for real-time portfolio choice.

3.2 Decomposing Model Uncertainty

We show in Proposition 3 that posterior model probabilities are closely related to the model-implied squared Sharpe ratio, SR^2_γ. Simply speaking, when several candidate models have SR^2_γ considerably more sizable than others, we can easily differentiate them and observe low model uncertainty. In contrast, when factor models have similar SR^2_γ, model uncertainty is heightened. Hence, the cross-sectional distribution of SR^2_γ determines the level of model uncertainty.

Figure 3 plots the time series of cross-sectional distances in SR^2_γ. More precisely, we show the difference between the maximal SR^2_γ and the $90th$-quantile of SR^2_γ, as well as the difference between its maximum and median. It is worth noting that the cross-sectional dispersion in SR^2_γ decreases sharply before the stock market crashes and remains at a low level during the bear markets. For example, the distance between the highest and medium in-sample SR^2_γ is close to 0.2 (daily) in 1997, but it plunges to almost 0 from 1998 to 2000. After the tech bubble, factor models have been becoming more similar in terms of in-sample SR^2_γ. Overall, the evidence in Figure 3 confirms that model uncertainty is directly related to the cross-sectional variability of SR^2_γ.

Theoretically, SR^2_γ is determined by the mean returns of factors and their covariance structure. We further analyze SR^2_γ by dipping into three parts: (a) average daily factor

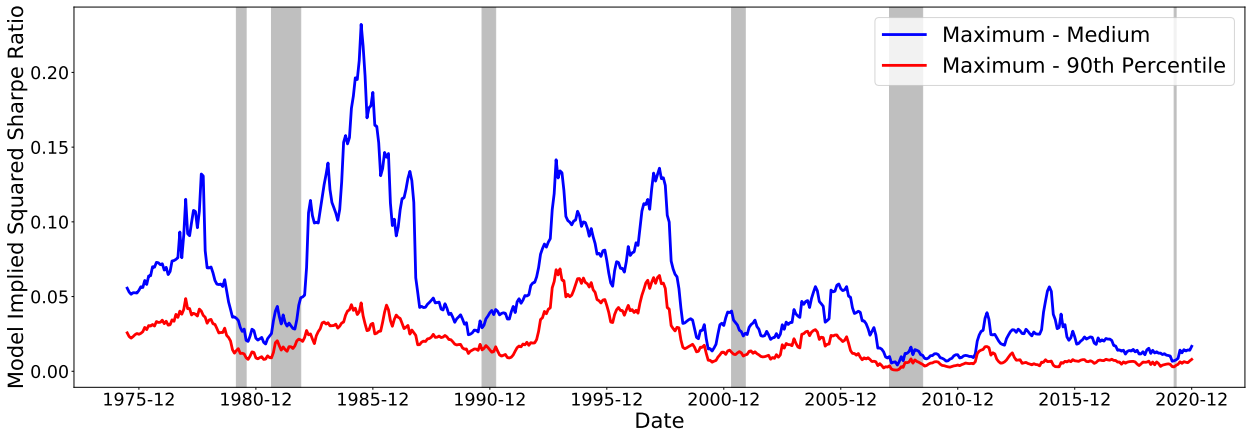


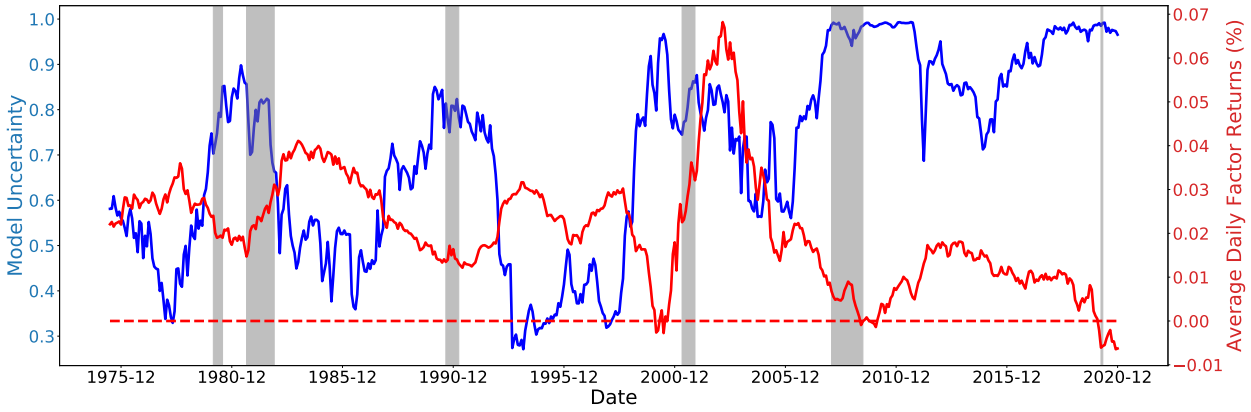
Figure 3: Time Series of Model-Implied Squared Sharpe Ratio (3-Year Rolling Window)

The figure plots the time series of distances in SR^2_γ from June 1975 to December 2020. We present the difference between the highest SR^2_γ and the 90th-quantile of SR^2_γ, as well as the difference between the highest SR^2_γ and medium SR^2_γ. SR^2_γ is the model-implied squared Sharpe ratio, $\mathbb{E}_T[\boldsymbol{f}_\gamma]^\intercal \boldsymbol{V}_\gamma^{-1} \mathbb{E}_T[\boldsymbol{f}_\gamma]$. $\mathbb{E}_T[\boldsymbol{f}_\gamma]$ and $\boldsymbol{V}_\gamma$ are estimated using the daily factor returns in the past 36 months.

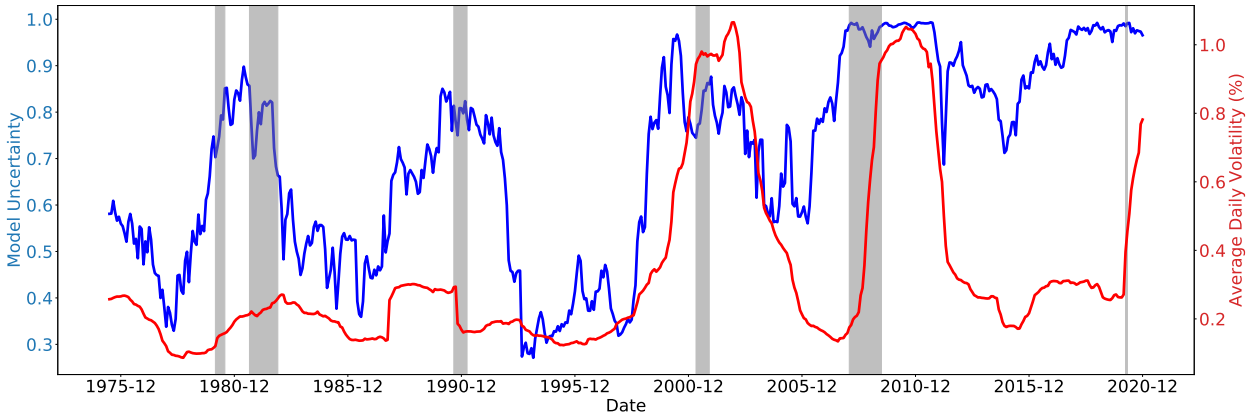
returns in the past three years; (b) average daily factor volatility in the past three years; (c) average pairwise correlation among daily factor returns in the past three years. Figure 4 plots these time series.

In Figure 4a, we show that the average daily return of all 14 factors is incredibly volatile. The average daily return also exhibits cyclical patterns. Specifically, it declines during the run-ups of stock markets, plummets to the bottom during the market crash, and recovers gradually after the bear markets. In the recent three most influential market crashes (dot-com bubble, 2008 global financial crisis, and the Covid-19), the average factor return declines to near zeros. In the past decade, the profitability of these 14 factors is no longer comparable to their historical performance. One potential reason is that more investors implement the same investment strategies after the publication of these factors (see McLean and Pontiff (2016)).

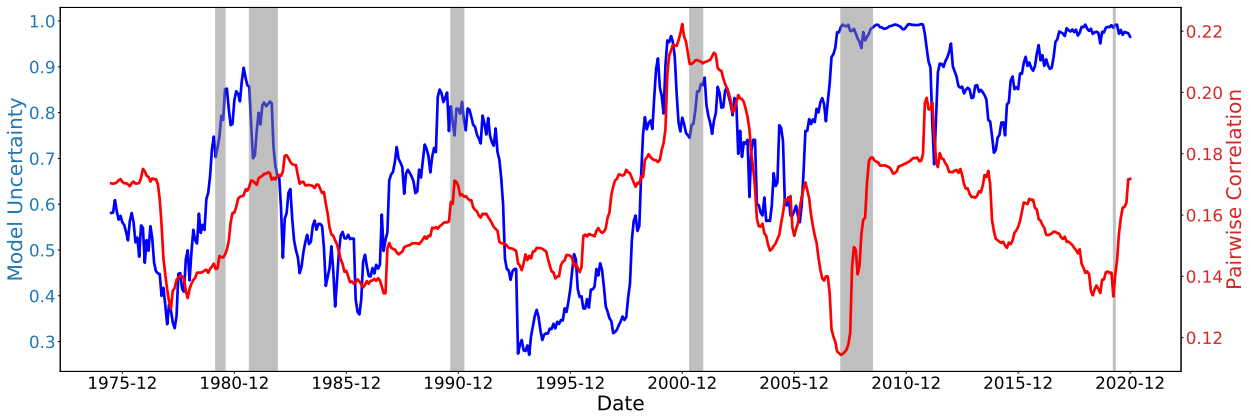
Figure 4b plots the average volatility of 14 factors. Even though the average factor volatility generally increases in the bear markets, the factor returns before the dot-com bubble are not as volatile as after 2000. Typically, the average standard deviation of 14 factors is between 0.2% and 0.4%. During the dot-com bubble and recent global financial crisis, it surges to higher than 1% daily. However, it is evident from figure 4b that model



(a) Time-Series of Average (Daily) Return of 14 Factors



(b) Time-Series of Average (Daily) Volatility of 14 Factors



(c) Time-Series of Average Pairwise Correlation of 14 Factors

Figure 4: Decomposing the Model Uncertainty

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

uncertainty does not have the same time-series patterns as the average factor volatility.

We further explore the correlation structure of factor strategies in Figure 4c. During market crashes, arbitrageurs investing in these factor strategies will exit the market simultaneously, thus driving up comovements among factors. Since the correlation matrix of factors determines the extent to which investors can diversify their investment, it indirectly influences the cross-sectional distances in SR^2_γ. To illustrate this point, we plot the time series of the average pairwise correlation of 14 factors.[15] The average correlation exhibits a similar cyclical pattern as model uncertainty. However, two key differences are noteworthy: (a) the average correlation decreases before the 2008 crisis while our model uncertainty starts to climb up from 2006, and (b) model uncertainty increases from 2015 to 2019, while the average correlation among factors declines during the same period.

To sum up, model uncertainty is considerable when the cross-sectional distances in SR^2_γ among different factor models are low. Since the in-sample SR^2_γ always increases with more factors included, we are uncertain about whether to include an additional factor conditioned that the benefit of including it is only marginal. Furthermore, model uncertainty about linear SDFs increases dramatically during the run-ups and stands at the peak during bear markets because different factor models are highly analogous.

3.3 Correlation with Other Economic Variables

Figure 1 indicates that model uncertainty increases during times of extreme uncertainty in the financial markets and economy. A natural question is how our model uncertainty index correlates with a number of key financial and macroeconomic variables known as capturing critical financial and economic fluctuations.

There are several notable uncertainty measures in the literature. The first measure, used in Bloom (2009), is VXO/VIX index,[16] which quantifies forward-looking market volatility. Subsequent to Bloom (2009), Jurado et al. (2015) and Ludvigson et al. (2021) develop the real, macro and financial uncertainty measures by exploiting a large set of macro and financial variables.[17] Baker et al. (2016) use the coverage of economic or policy-related keywords in

[15] At the end of each month t, we use daily factor returns from month $t-35$ to month t to compute the pairwise correlation between any two factors, denoted as ρ_{ij}. The average is computed as $\frac{1}{N\times(N-1)}\sum_{i\neq j}|\rho_{ij}|$.

[16] VIX and VXO index are essentially the same: the correlation between them is higher than 0.98.

[17] They quantify the h-period ahead uncertainty by the extent to which a particular set of economic variables (either real, macro, or financial) become more or less predictable from the perspective of economic agents. Suppose there is a set of economic indicators, $\boldsymbol{Y_t} = (y_{1t}, ..., y_{Lt})^\intercal$. For each variable, they find the

the media as proxies for economic policy uncertainty.

In addition to uncertainty measures, we compare model uncertainty with the intermediary factor from He et al. (2017), the term yield spread (the yield on ten-year government bonds minus the yield on three-month treasury bills), and the credit spread (the yield on BAA corporate bonds minus the yield on AAA corporate bonds).

We report in Table 2 the results from the regressions of model uncertainty on its one-period lag and some contemporaneous economic variables. By running these regressions, we do not intend to study the causal relationship between model uncertainty and other economic variables. Instead, our objective is to describe the contemporaneous relation between them. We also want to point out that model uncertainty is persistent[18] since it is constructed in a rolling window of 36 months. Therefore, we need to be cautious in statistical inference. In all following tables, we use Newey-West standard errors (see Newey and West (1987)) with 36 lags.

As Table 2 shows, a number of economic variables are significantly related to model uncertainty, even after we control one-period lagged entropy. For example, model uncertainty is positively correlated with financial uncertainty and the VXO index but almost orthogonal to real, macro, and two economic policy uncertainty measures. This finding is intuitive since model uncertainty mainly refines information in financial markets. In addition, the intermediary factor and term yield spread negatively relate to model uncertainty. In column (10), we run horse racing among the VXO index, the intermediary factor, and term yield spread: While the coefficient estimates of the VXO index and term yield spread still remain significant, the intermediary factor becomes inconsequential.

Comments. Conceptually, model uncertainty quantifies a different layer of uncertainty from other measures. The stock market volatility, proxied by the VXO index, measures the second-moment investment risk. Three uncertainty measures in Jurado et al. (2015)

conditional volatility of the prediction errors: $u_{jt}(h) = \sqrt{E[(y_{j,t+h} - E[y_{j,t+h}|I_t])^2|I_t]}$. The aggregate uncertainty is quantified by the average conditional volatility of the prediction error of each economic indicator: $u_t(h) = \sum_{j=1}^{L} \omega_j u_{jt}(h)$, where ω_j is the weight on the j-th economic indicator. The detailed econometric framework could be found in the original papers. Our paper considers their one-period ahead uncertainty measures.

[18]Strong persistence of the time-series process is ubiquitous in other uncertainty measures. Table A2 shows the AR(1) coefficients of the other six uncertainty sequences, and we find that the real, macro and financial uncertainty measures also have AR(1) coefficients less than but close to 1. It is well-known that the volatility of asset returns tends to cluster. When we run the AR(1) for the VXO index, the coefficient estimate of ρ is 0.812. Only the second economic policy uncertainty measure (EPU_2) suffers less from massive autocorrelations.

Table 2: Regressions of Model Uncertainty on Contemporaneous Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged Entropy	0.979***	0.982***	0.983***	0.985***	0.983***	0.983***	0.986***	0.985***	0.986***	0.983***
	(128.85)	(142.76)	(146.97)	(106.55)	(105.75)	(129.25)	(161.37)	(150.06)	(154.04)	(131.19)
Financial Uncertainty	0.212*									
	(1.95)									
Macro Uncertainty		0.174								
		(1.53)								
Real Uncertainty			0.140							
			(1.20)							
EPU I				0.000						
				(0.33)						
EPU II					0.000					
					(1.07)					
VIX/VXO						0.005**				0.004**
						(2.20)				(2.34)
Intermediary Factor							-0.503**			-0.196
							(-2.01)			(-0.71)
Term Spread								-0.034***		-0.033**
								(-3.44)		(-2.44)
Default Spread									-0.003	
									(-0.09)	
Sample size	546	546	546	432	432	420	546	546	546	420

The table reports the results from the regression of model uncertainty on its one-period lag and some contemporaneous economic variables (X_{t+1}):

$$Entropy_{t+1} = \beta_0 + \beta_1 Entropy_t + \rho X_{t+1} + \epsilon_{t+1}.$$

X_{t+1} include a) financial, macro, and real uncertainty measures from Jurado et al. (2015) and Ludvigson et al. (2021) in columns (1) - (3), b) two economic policy uncertainty (EPU) indices from Baker et al. (2016) in columns (4) and (5), c) VXO index in column (6), d) the intermediary factor from He et al. (2017) in column (7), e) term spread in column (8), f) default spread in column (9). The t-statistics are computed using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

and Ludvigson et al. (2021) are essentially volatilities of prediction errors. Two economic policy uncertainty indices in Baker et al. (2016) are to quantify public attention to economic policy. In contrast, our paper quantifies the cross-sectional model uncertainty regarding linear SDFs. One benefit of our measure is that the lower and upper bounds of entropy are known. For example, model uncertainty reaches its upper bound in some periods, implying that different models' posterior probabilities are almost identical. Hence, the economic interpretation of model uncertainty is straightforward. In short, our model uncertainty index is complementary to other uncertainty measures developed in the past literature.

4 Mutual Fund Flows

If investors consider model uncertainty a crucial source of investment risk, a natural hypothesis is that their portfolio choice decisions are related to our model uncertainty measure. We test this hypothesis using mutual fund flows as proxies for investors' portfolio choices.

Our paper differs from past literature studying mutual fund flows. Earlier analyses focus on the relationship between mutual fund flows and performance.[19] In addition, a large and growing body of literature is concerned with how mutual fund flows determine the returns and covariance structure of factor strategies.[20] Unlike these papers, we investigate how investors react to uncertainty shocks.

The data is available on CRSP survivor-bias-free US mutual fund database. The database includes investment styles or objective codes from three different sources over the whole life of the database.[21] The CRSP style code consists of up to four letters. For example, a fund with the style "EDYG" means that i) this fund mainly invests in domestic equity markets (E = Equity, D = Domestic), and ii) it has a specific investment style "Growth" (Y = Style, G = Growth).[22] The quality of data before 1991 is low because the CRSP investment objective code is incomplete. For example, only domestic equity "style" funds and mixed fixed income and equity funds are recorded before 1991. Also, the market values of institutional holdings proportional to the total market value of all stocks (in CRSP) were tiny. Therefore, we focus on the sample from January 1991 to December 2020.

To begin with, we define the aggregate mutual fund flows. Following the literature (see Lou (2012)), we calculate the net fund flows to each fund i in period t as

$$Flow_{i,t} = TNA_{i,t} - TNA_{i,t-1} \times (1 + R_{i,t}) \tag{15}$$

where $TNA_{i,t}$ and $R_{i,t}$ are total net assets and gross returns of fund i in period t. Next, we aggregate individual fund flows in each period across all funds in a specific group (e.g., all

[19] See Chevalier and Ellison (1997), Sirri and Tufano (1998), Barber et al. (2016), Berk and Van Binsbergen (2016), etc.

[20] See Lou (2012), Huang et al. (2021), Li (2021), Ben-David et al. (2021), Koijen and Yogo (2019), Gabaix and Koijen (2021), etc.

[21] From 1962 to 1993, Wiesenberger objective codes are used. Strategic insight objective codes are populated between 1993 and 1998. Lipper objective codes start in 1998. Instead of using the three measures mentioned above directly, CRSP builds its objective codes based on them.

[22] More details are in the handbook of CRSP survivor-bias-free US mutual fund database.

large-cap funds) and scale the aggregate flows by the lagged total market capitalization of all stocks in CRSP:

$$Flows_t^Y = \frac{\sum_{i \in Y} Flow_{i,t}}{\text{CRSP-Market-Cap}_{t-1}}, \tag{16}$$

where Y specifies a certain investment objective, such as small-cap funds.

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

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

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

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

4.1 Aggregate Equity vs Fixed-Income Funds

Since our model uncertainty measure is based on factors in the US, we delete all foreign mutual funds. In the baseline analysis, we consider the aggregate mutual fund flows to the entire US equity and fixed-income markets. That is, we study the VAR regression in equation (17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows_t^{FI}, Flows_t^{Equity})$. We next use impulse response functions (IRFs) to better understand the dynamic effects and propagating mechanisms of uncertainty shocks.

IRFs greatly depend on the identification assumption, i.e., whether model uncertainty is an exogenous source of fluctuations in fund flows or an endogenous response. In the first case, model uncertainty is a cause of fund flows, while it acts as a propagating mechanism in the latter case. Without taking a strong stance on the identification assumption, we aim to investigate the dynamic relationship between fund flows and several uncertainty measures, either as a cause or propagating mechanism. To make as few assumptions as possible, we focus on the dynamic responses to uncertainty shocks and are silent on how innovations in

fund flows affect model uncertainty. This simplification allows us to ignore the ordering of other economic variables beyond model uncertainty.

In the benchmark case, we place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to fund flows in the current period. We consider a different identification assumption in robustness checks in Section 7, where we put model uncertainty as the last element in $\boldsymbol{Y_t}$. As shown below, the IRFs to model uncertainty shocks are essentially robust to the alternative identification strategy, whereas the IRFs to other uncertainty measures are not.

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

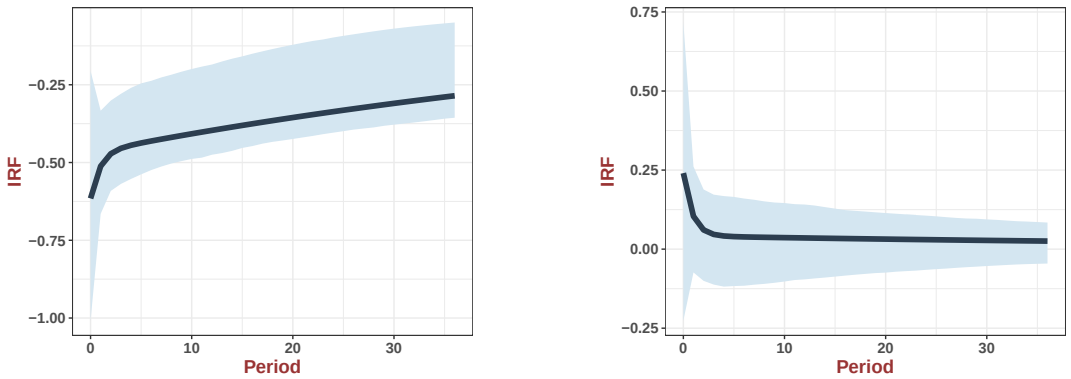
	$Entropy_{t+1}$		$Flows^{FI}_{t+1}$		$Flows^{Equity}_{t+1}$	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	0.042	1.266	-0.064	-0.259	1.615***	8.197
$Entropy_t$	0.985***	140.610	-0.012	-0.178	-0.344***	-8.106
$Flows^{FI}_t$	0.009	1.198	0.247***	3.856	-0.081	-1.329
$Flows^{Equity}_t$	-0.003	-0.331	-0.093	-1.500	0.240***	4.044
MKT_t	-0.008	-0.980	-0.054	-0.642	0.062	0.970
VXO_t	0.006	0.483	0.170**	2.115	-0.010	-0.251

This table reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows^{FI}_t, Flows^{Equity}_t)$. $Entropy_t$ is the model uncertainty measure, and $Flows^{FI}_t$ ($Flows^{Equity}_t$) is the aggregate flows to the domestic fixed-income (equity) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged market return (MKT_t) and VXO index (VXO_t) 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. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

Table 3 reports the results from the VAR regression. The sample ranges from January 1991 to December 2020. The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also include the lagged market return and VXO index as control variables in each regression. 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 only relates to its lag. Second, the VXO index positively predicts the aggregate flows to fixed-income funds: One standard deviation increase in VXO predicts 0.17 standard deviation inflows to fixed-income funds. Third, model uncertainty negatively forecasts equity fund flows, and the coefficient estimate is sizable

in both economic and statistical senses. In particular, one standard deviation increase in model uncertainty implies 0.34 standard deviation equity fund outflows. Although we cannot interpret the regression results as causal, we still find that investors in domestic equity mutual funds tend to decrease their exposures when model uncertainty increases.

Figure 5 plots the dynamic responses of fund flows to model uncertainty shocks in VAR-1. Most 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 impulse response functions (IRFs) start from around -0.6 in period zero and slowly decline to -0.35 in period 36, significantly negative based on the 90% standard error bands. In contrast, model uncertainty has negligible effects on fixed-income fund flows (see Panel (b)).



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

Figure 5: Impulse Responses of Equity and Fixed-Income Mutual Fund Flows using Entropy as Uncertainty

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 normalize the IRFs such that the model uncertainty shock increases one standard deviation model uncertainty. We place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to mutual funds in the current period. The data are monthly and span the period 1991:01 - 2020:12.

4.2 Different Equity Mutual Funds

We further study the heterogeneous responses of different equity mutual funds to model uncertainty shocks. In particular, we split equity mutual funds into four categories: (a) style funds that specialize in factor investing, (b) sector funds that invest in specific industries

(e.g., gold, oil, etc.), (c) small-cap funds that invest in relatively small stocks,[23] and (d) large-cap funds that invest in large stocks.

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

	$Entropy_{t+1}$		$Flows^{style}_{t+1}$		$Flows^{sector}_{t+1}$		$Flows^{small}_{t+1}$		$Flows^{large}_{t+1}$	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	0.268***	3.270	1.582***	5.195	0.180	0.974	0.533	1.525	0.347	0.746
$Entropy_t$	0.952***	67.064	-0.261***	-5.672	-0.021	-0.525	-0.121**	-1.967	-0.014	-0.209
$Flows^{style}_t$	-0.012	-1.048	0.211***	2.936	-0.056	-1.034	-0.003	-0.054	0.003	0.034
$Flows^{sector}_t$	0.031	1.553	-0.056	-1.089	0.254*	1.686	-0.059	-0.664	-0.123**	-2.266
$Flows^{small}_t$	-0.001	-0.035	0.010	0.169	0.039	0.541	0.424***	6.081	0.089	1.225
$Flows^{large}_t$	0.019*	1.682	0.062	1.181	-0.043	-0.661	-0.107*	-1.731	0.092	1.164
R^{style}_t	0.191	0.987	0.627	0.682	0.401	0.944	-0.192	-0.215	-1.891*	-1.652
R^{sector}_t	0.043	0.900	0.121	0.956	0.367	0.957	-0.210	-1.115	0.010	0.056
R^{small}_t	-0.165**	-2.566	-0.212	-0.983	-0.126	-0.363	0.535**	1.967	0.383	1.527
R^{large}_t	-0.099	-0.741	-0.437	-0.576	-0.605	-1.610	-0.022	-0.034	1.468*	1.653
VXO_t	0.006	0.510	-0.027	-0.467	0.067	1.022	0.093*	1.957	-0.013	-0.151

This table reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows^{style}_t, Flows^{sector}_t, Flows^{small}_t, Flows^{large}_t)$. $Entropy_t$ is the model uncertainty measure, and $Flows^{style}_t$ ($Flows^{sector}_t$, $Flows^{small}_t$, $Flows^{large}_t$) is the aggregate flows to the domestic style (sector, small-cap, large-cap) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the lagged fund returns of each type and VXO 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.

Table 4 reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows^{style}_t, Flows^{sector}_t, Flows^{small}_t, Flows^{large}_t)$. The lag of VAR is chosen by BIC and equals one. Since the cap-based investment objective code is available after 1997, the sample begins in January 1998. First, after controlling its lag, model uncertainty is negatively predicted by large-cap fund flows and small-cap fund returns. Second, model uncertainty negatively forecasts style and small-cap fund flows, and the coefficients are sizable. Specifically, if model uncertainty rises by one standard deviation, style (small-cap) fund flows tend to drop by 0.26 (0.12) standard deviation over the next period. In contrast, we do not discover a significant relationship between model uncertainty and sector (large-cap) fund flows.

Different from model uncertainty, the traditional volatility-based uncertainty measure (VXO) plays a limited role in the VAR regression. It can marginally predict small-cap fund flows, but the sign of coefficient estimate is counter-intuitive: When uncertainty goes up,

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

investors invest more in small-cap funds. Instead, we observe a negative response of small-cap funds when using entropy as the uncertainty measure. Therefore, our model uncertainty index captures an essential source of investment risk for equity investors, which is omitted by the traditional VXO index.

Figure 6 shows the dynamic responses of four different types of equity fund flows to model uncertainty shocks in VAR-1. Consistent with Table 4, model uncertainty shocks reduce future style fund flows, and the effects are long-lasting (see Panel (a)). This observation is intuitive. Style funds refer to the growth, income, growth & income, and "hedged" funds, so they mostly rely on the factor strategies used in constructing model uncertainty. Therefore, the outflows from style equity funds are remarkably enormous when model uncertainty is heightened.

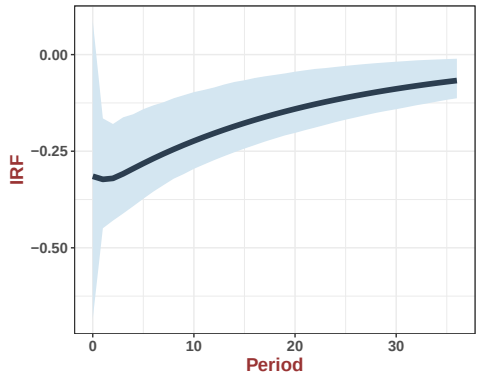
Moreover, we observe significantly negative IRFs of small-cap funds (see Panel (c)), although the effects are not as persistent as of style funds. This observation is reasonable since we include two size factors in model uncertainty. On the contrary, sector and large-cap funds do not respond to model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

4.3 Different Fixed-Income Mutual Funds

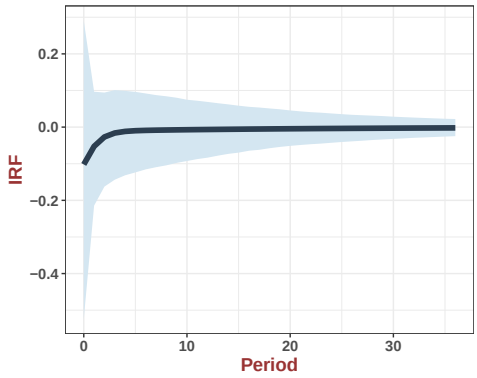
Similar to the previous section, we divide all fixed-income mutual funds into four categories: (a) government bond funds, (b) money market funds, (c) corporate bond funds, and (d) municipal bond funds. This subsection repeats a similar VAR estimation and investigates the dynamic responses of fixed-income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

Table 5 shows the results from the VAR-1 regression. According to columns (3) and (4), model uncertainty positively predicts the aggregate fund flows in US government bonds. US government bonds are notable for their superior safety over other asset classes. Hence, investors tend to allocate more wealth to safe assets when model uncertainty is more substantial. In contrast, model uncertainty negatively forecasts corporate fund flows, so mutual fund investors reduce their exposure to corporate bonds following high model uncertainty.

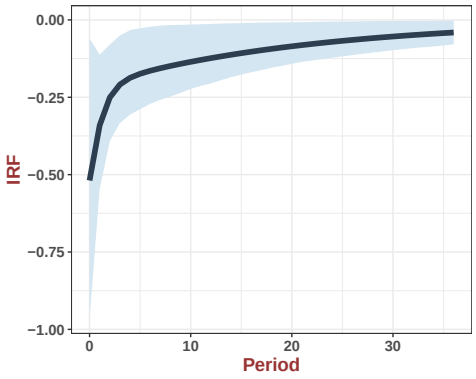
Next, we report the IRFs of different fixed-income funds to entropy shocks in Figure 7. Not surprisingly, we document sharp dynamic inflows to government bond funds. As Panel (a) suggests, one standard deviation increase in model uncertainty corresponds to



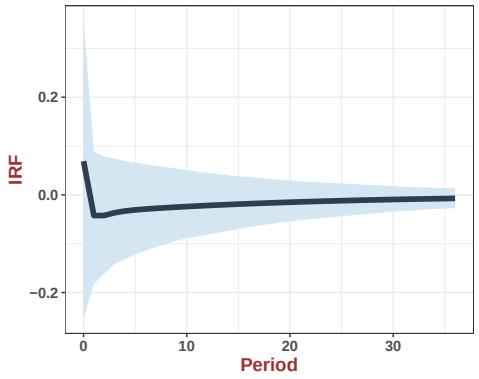
(a) Style Fund Flows to Entropy



(b) Sector Fund Flows to Entropy



(c) Small-Cap Fund Flows to Entropy



(d) Large-Cap Fund Flows to Entropy

Figure 6: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes using Entropy as Uncertainty

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 place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to mutual funds in the current period. The data are monthly and span the period 1998:01 - 2020:12.

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

	$Entropy_{t+1}$		$Flows^{gov}_{t+1}$		$Flows^{money}_{t+1}$		$Flows^{corp}_{t+1}$		$Flows^{muni}_{t+1}$	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
Intercept	0.381***	2.812	-0.574***	-3.305	-0.191	-0.724	1.033***	4.198	-0.132	-0.576
$Entropy_t$	0.983***	97.964	0.182**	2.535	-0.049	-0.816	-0.189***	-2.712	0.093	1.621
$Flows^{gov}_t$	0.016**	2.407	0.341***	4.580	0.090	1.325	0.094	1.448	0.136**	1.991
$Flows^{money}_t$	0.011	1.528	-0.017	-0.369	0.252***	3.125	-0.064	-1.072	0.016	0.299
$Flows^{corp}_t$	-0.012	-0.990	0.008	0.235	0.029	0.709	0.161**	2.264	0.166***	2.591
$Flows^{muni}_t$	-0.022**	-2.067	0.131**	2.064	-0.095	-1.380	0.200	1.455	0.193	1.336
VXO_t	0.006	0.590	0.044	0.623	0.191**	2.298	-0.007	-0.084	-0.094	-1.422

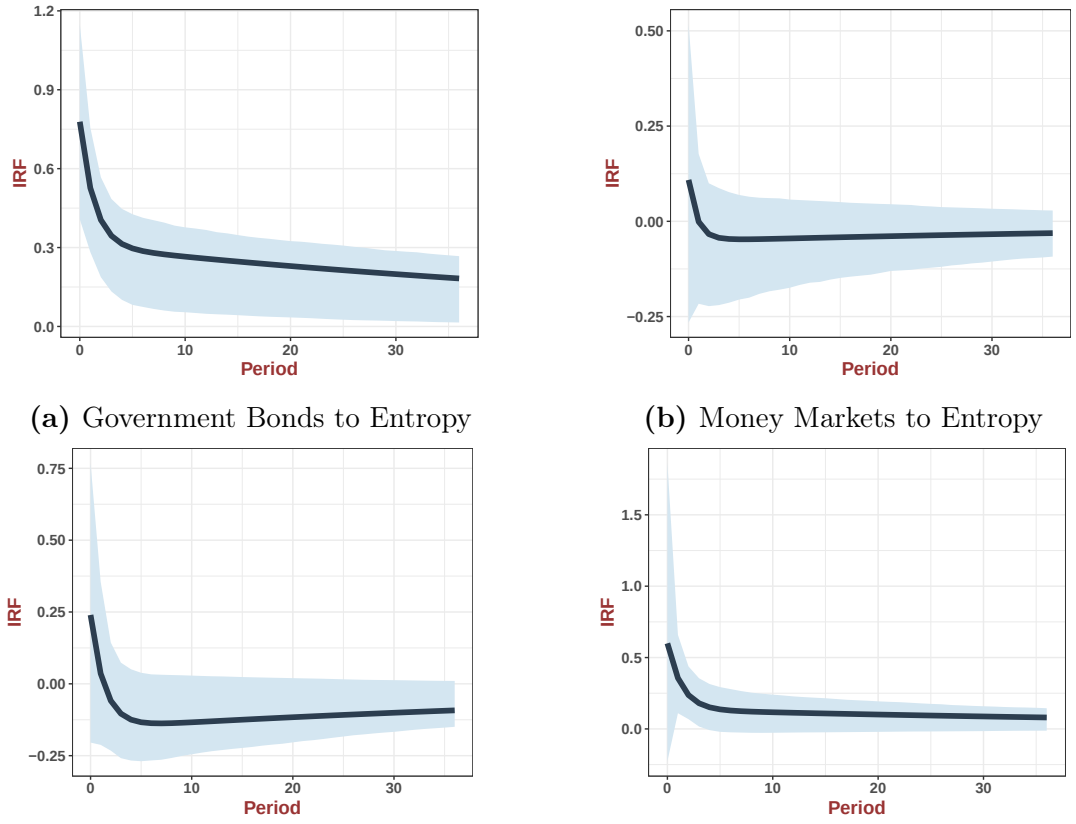
This table reports the results from the VAR estimation in equation (17), where $\boldsymbol{Y_t}^\top = (Entropy_t, Flows^{gov}_t, Flows^{money}_t, Flows^{corp}_t, Flows^{muni}_t)$. $Entropy_t$ is the model uncertainty measure, and $Flows^{gov}_t$ ($Flows^{money}_t$, $Flows^{corp}_t$, $Flows^{muni}_t$) is the aggregate flows to the domestic government bond (money market, corporate bond, and municipal bond) mutual funds, normalized by the lagged total market capitalization of all stocks in CRSP (see equation (16)). The lag is chosen by BIC and equals one. In addition, we standardize all economic variables such that they have unit variances. We also control for the VXO 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.

more than 0.7 standard deviation increase in government bond fund inflows at time zero, and the dynamic response persists for more than 36 periods. On the contrary, the IRFs of other fixed-income fund flows are insignificant.

It is also worth noting that we do not observe a significant relationship between model uncertainty and money market funds. The difference between money market and government bond funds is that the first type has a smaller duration and more liquid, while the latter consists of government bonds of different maturities. Unlike model uncertainty, the VXO index significantly predicts positive inflows to money market funds. We interpret these facts as evidence that high model uncertainty induces "flight to safety", whereas high VXO relates to "flight to liquidity".

4.4 Comparison with Other Uncertainty Measures

One major concern about the previous analyses is that model uncertainty is correlated with other uncertainty indicators, so the dynamic responses of mutual fund flows to model uncertainty shocks are confounded by them. Hence, we study how other uncertainty measures affect mutual fund flows in this section and compare their dynamic responses with the previous results. We consider the VXO index and financial uncertainty in Jurado et al. (2015) since these two measures are closely associated with our model uncertainty measure, as we



(a) Government Bonds to Entropy

(b) Money Markets to Entropy

(c) Corporate Bonds to Entropy

(d) Municipal Bonds to Entropy

Figure 7: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes using Entropy as Uncertainty

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 normalize the IRFs such that the model uncertainty shock increases one standard deviation model uncertainty. We place model uncertainty first in the VAR. Hence, the implicit identification assumption is that fund flows react to the contemporaneous uncertainty shocks, while model uncertainty does not respond to the shocks to mutual funds in the current period. The data are monthly and span the period 1991:01 - 2020:12.

show in Table 2.

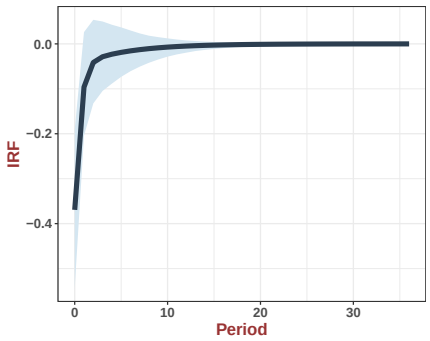
Figure 8 plots the dynamic responses of four different types of equity fund flows to VXO or financial uncertainty shocks in VAR-1. Consistent with the previous identification assumption, We place VXO or financial uncertainty first in the VAR. We also control the lagged model uncertainty in each regression. First, as Panels (a) and (b) indicate, style funds experience massive outflows when VXO or financial uncertainty increases. However, these effects are temporary; that is, the IRFs of fund flows reverse back to zeros immediately after time zero. On the contrary, model uncertainty shocks are followed by persistent outflows from style funds even beyond 36 periods. Similarly, the dynamic responses of fund flows to sector/small-cap/large-cap funds are also transitory and not significant (except for Panel (c) at period zero).

We further consider the dynamic responses of fixed-income funds in Figure 9. When VXO or financial uncertainty goes up, government bond funds tend to experience massive inflows, although these effects are less than 50% of those following model uncertainty shocks (see Figure 7(a)). Most strikingly, we document massive inflows to money market funds after positive VXO and financial uncertainty shocks. In contrast, model uncertainty is not essential in money market funds. In other words, model uncertainty shocks primarily induce "flight to safety," while other volatility-based uncertainty measures are mainly related to "flight to liquidity."

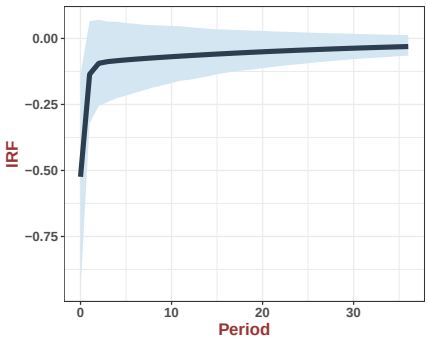
In summary, our model uncertainty measure captures some unique dynamic responses of fund flows, and notably, they are distinct from traditional volatility-based measures, such as VXO and financial uncertainty. In particular, we observe significant fund inflows to government bond funds and outflows from style and small-cap equity funds. In contrast, VXO and financial uncertainty shocks fail to generate similar dynamic responses. Finally, as we will show in Section 7, the IRFs of fund flows to model uncertainty shocks are virtually robust to an alternative identification assumption, whereas the effects of VXO or financial uncertainty shocks are fairly sensitive.

5 Investors' Expectations

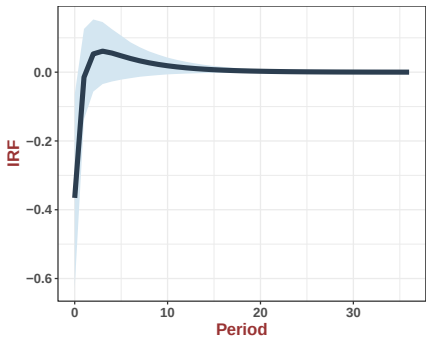
This section investigates whether our model uncertainty measure correlates with investors' expectations of the stock markets. The first measure is from the American Association of



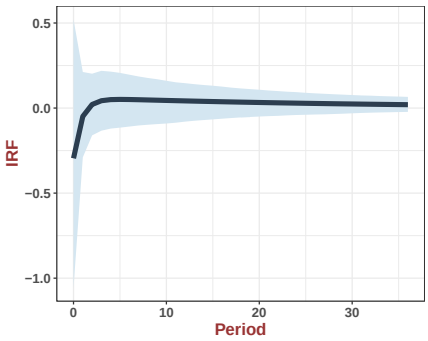
(a) Style Fund Flows to VXO



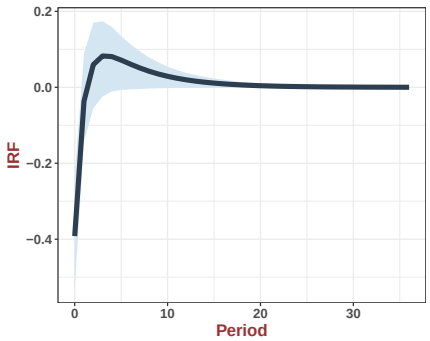
(b) Style Fund Flows to Financial Unc



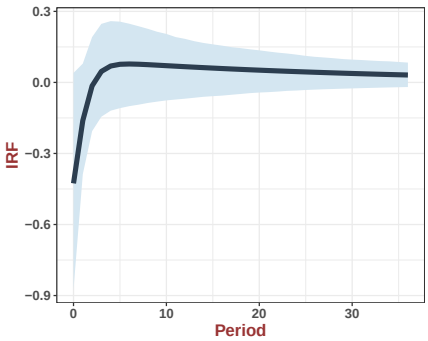
(c) Sector Fund Flows to VXO



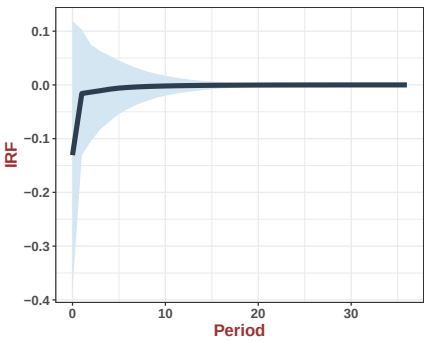
(d) Sector Fund Flows to Financial Unc



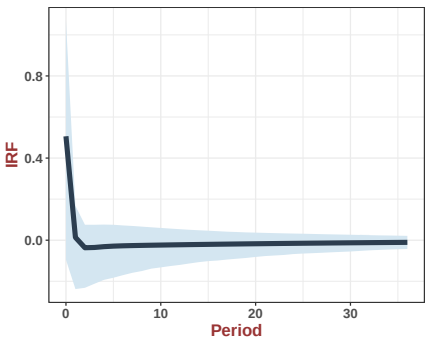
(e) Small-Cap Fund Flows to VXO



(f) Small-Cap Fund Flows to Financial Unc



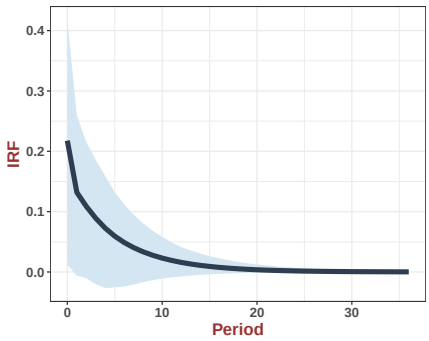
(g) Large-Cap Fund Flows to VXO



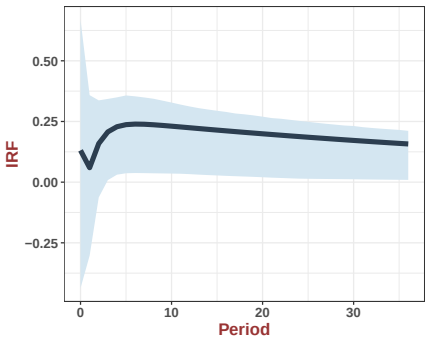
(h) Large-Cap Fund Flows to Financial Unc

Figure 8: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes using VXO and Financial Uncertainty as Uncertainty Measures

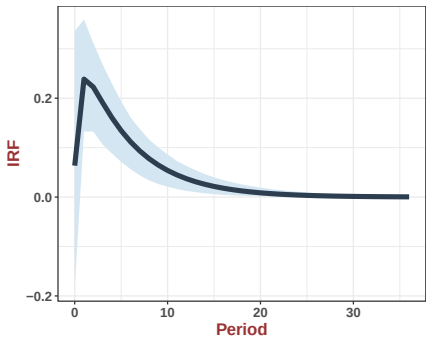
This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to VXO and financial uncertainty shocks in VAR-1. Other details can be found in the footnote of Figure 6.



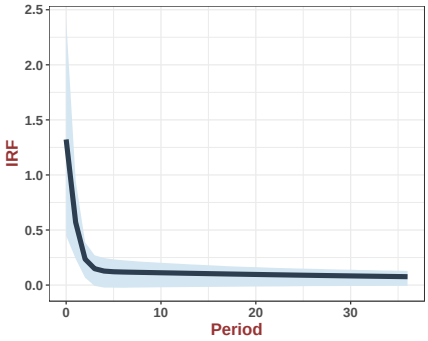
(a) Government Bonds to VXO



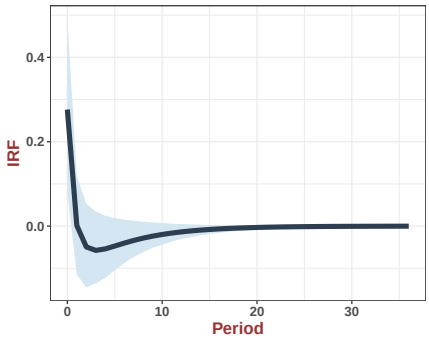
(b) Government Bonds to Financial Unc



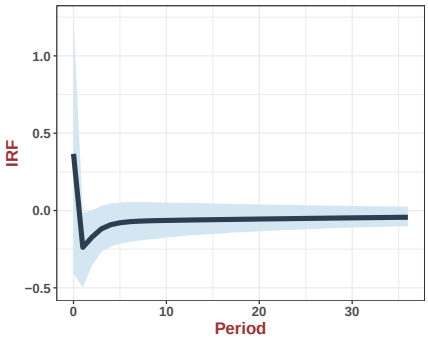
(c) Money Markets to VXO



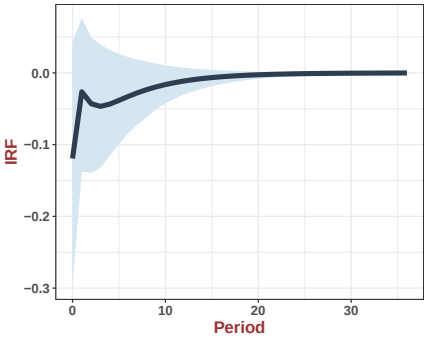
(d) Money Markets to Financial Unc



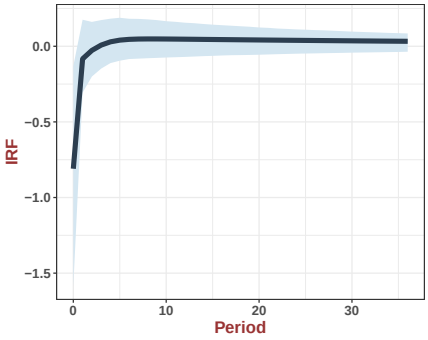
(e) Corporate Bonds to VXO



(f) Corporate Bonds to Financial Unc



(g) Municipal Bonds to VXO



(h) Municipal Bonds to Financial Unc

Figure 9: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes using VXO and Financial Uncertainty as Uncertainty Measures

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

Individual Investors (AAII). The survey is completed weekly by registered members of AAII, and it asks the investors whether they are bearish, neutral, or bullish on the stock market for the next six months. Since our model uncertainty measure is monthly, we use the expectation measures in the last week of each month.

We also consider Robert Shiller's stock market confidence indices from the survey conducted by the International Center for Finance at the University of Yale. Our paper focuses on the US one-year confidence index and US crash confidence index. Specifically, the one-year confidence index is the percentage of the individual or institutional investors expecting an increase in the Dow in a year. In contrast, the crash confidence index is the percentage of individual or institutional investors who believe the probability of a catastrophic stock market crash in the next six months is lower than 10%. Roughly speaking, the higher the indices are, the more confident individual or institutional investors are about the stock market.

We consider the following time-series regression:

$$Exp_{t+1} = \beta_0 + \gamma Entropy_t + \psi X_t + \epsilon_{t+1} \tag{18}$$

where Exp_{t+1} is the one-period ahead expectation measure, $Entropy_t$ is the model uncertainty measure in period t, and X_t includes other control variables up to time t, such as lagged expectation indices, VXO and etc. Since all expectation indices are autocorrelated, we control their one and two-period lags in all regressions.[24] We further control lagged market returns (S&P 500 index) in the regression for investors' expectations on the market are extrapolative (see Greenwood and Shleifer (2014)).

In table 6(a), we regress AAII sentiment indices on model uncertainty to explore how individual investors change their attitudes towards the stock market in response to variation in model uncertainty. To increase the interpretability of our results, we standard model uncertainty to have unit variance, so coefficient estimates of $Entropy_t$ are interpreted as the increases in the percentages of bullish/neutral/bearish investors when model uncertainty grows by one standard deviation.

In columns (1) and (2), $Entropy_t$ cannot predict the next-period percentage of bullish investors. Specifically, the average investors become less bullish if model uncertainty in the cross-section goes up, but this prediction is not sharp. Columns (3) and (4) regress the

[24]The coefficient estimate of 3-period lagged variable is close to zero and insignificant, so we include only the first two lags.

Table 6: Investors' Expectations, Confidence Indices, and Model Uncertainty

	Panel (a). AAII Sentiment Index							
Exp_{t+1} =	Bullish		Neutral		Bearish		Bullish - Bearish	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Entropy_t$	-0.280	-0.374	-0.605**	-0.434**	1.043**	1.036***	-1.511*	-1.574**
	(-0.683)	(-1.122)	(-2.121)	(-2.102)	(2.499)	(2.656)	(-1.826)	(-2.127)
VXO_t	0.022	0.079	0.016	-0.009	-0.008	-0.034	0.016	0.118
	(0.311)	(1.500)	(0.211)	(-0.161)	(-0.169)	(-0.500)	(0.157)	(1.060)
Exp_t	0.418***	0.373***	0.487***	0.452***	0.367***	0.335***	0.373***	0.331***
	(8.593)	(6.954)	(9.709)	(10.249)	(9.238)	(7.155)	(7.325)	(5.983)
Exp_{t-1}	0.098**	0.158***	0.213***	0.253***	0.182***	0.208***	0.118***	0.160***
	(2.434)	(3.531)	(6.103)	(6.209)	(5.676)	(5.850)	(3.151)	(3.623)
Lagged Market Returns	N	Y	N	Y	N	Y	N	Y
Sample Size	400	396	400	396	400	396	400	396
R^2_{adj}	21.76%	22.53%	43.11%	44.98%	27.24%	26.79%	20.92%	21.01%

	Panel (b). Shiller's Confidence Indices							
Exp_{t+1} =	1-Year Confidence Index (Institution)		1-Year Confidence Index (Individual)		Crash Confidence Index (Institution)		Crash Confidence Index (Individual)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Entropy_t$	-0.365***	-0.379***	-0.546***	-0.682***	-0.562***	-0.635***	-0.754***	-0.754***
	(-2.727)	(-2.952)	(-2.733)	(-5.405)	(-3.265)	(-3.335)	(-5.790)	(-6.048)
VXO_t	0.025*	0.030	0.044*	0.080***	-0.058**	-0.034	-0.046***	-0.001
	(1.767)	(0.829)	(1.705)	(4.204)	(-2.153)	(-1.066)	(-2.712)	(-0.047)
Exp_t	1.133***	1.165***	0.931***	0.949***	1.068***	1.065***	1.086***	1.071***
	(16.820)	(18.984)	(11.730)	(15.898)	(19.165)	(21.126)	(16.459)	(13.272)
Exp_{t-1}	-0.270***	-0.304***	-0.015	-0.045	-0.217***	-0.219***	-0.241***	-0.208***
	(-3.603)	(-4.449)	(-0.212)	(-0.823)	(-3.540)	(-4.000)	(-4.268)	(-3.078)
Lagged Market Returns	N	Y	N	Y	N	Y	N	Y
Sample Size	232	228	232	228	232	228	232	228
R^2_{adj}	82.70%	83.38%	93.17%	93.25%	87.44%	87.01%	92.24%	92.82%

The table reports empirical results in regression: $Exp_{t+1} = \beta_0 + \gamma Entropy_t + \psi X_t + \epsilon_{t+1}$, where Exp_{t+1} is the one-period ahead expectation/confidence index, $Entropy_t$ is the model uncertainty measure in period t, and X_t includes other control variables up to time t, such as lagged expectation/confidence indices, VXO and etc. Since all expectation/confidence indices are autocorrelated, we control their one and two-period lags (Exp_t and Exp_{t-1}) in all regressions. We further control lagged market returns in the regression (we include six lags). In Panel (a), expectation indices come from the survey conducted by the American Association of Individual Investors (AAII). The survey is completed weekly by registered members of AAII, and it asks the investors whether they are bearish, neutral or bullish on the stock market for the next six months. Therefore, we have the data regarding the percentages of bearish, neutral or bullish respondents each week. Since our model uncertainty measure is monthly, we use the expectation index in the final week of each month. In Panel (b), confidence indices come from Shiller's survey. We focus on the US one-year confidence index and US crash confidence index. The one-year confidence index is the percentage of the individual or institutional investors expecting an increase in the Dow in a year. In contrast, the crash confidence index is the percentage of individual or institutional investors who think that the probability of a catastrophic stock market crash in the next six months is lower than 10%. The t-statistics are computed using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level.

percentage of neutral investors on lagged model uncertainty: If model uncertainty increases by one standard deviation, the fraction of neutral investors declines by 0.605% or 0.434%, depending on the regression setup.

The next question is, in which direction do bullish investors change their attitudes?

Columns (5) and (6) indicate that investors are more likely to be bearish following an increase in model uncertainty. Our interpretation is that some neutral investors become bearish after observing a higher level of model uncertainty. Finally, we regress the difference between fractions of bullish and bearish investors on entropy. The coefficient estimate of entropy is negative and significant at the 10% level. Overall, when model uncertainty goes up, market participants tend to be more pessimistic about the future stock market performance.

Table 6(b) regresses Shiller's confidence indices on entropy. Unlike the AAII sentiment index, we also observe the expectations of institutional investors. The results are generally similar to table 6: Investors tend to be more pessimistic about the stock market when model uncertainty increases. They also believe that a market crash is more likely to occur following higher model uncertainty. One interesting empirical fact is that the coefficient estimates of $Entropy_t$ in the regressions of individual investors' confidence indices are always more negative than institutional investors. Hence, individual investors react more aggressively to the changes in model uncertainty than institutional ones.

6 Evidence in European and Asian Pacific Markets

This section presents the time series of model uncertainty in European and Asian Pacific stock markets. Instead of using all 14 factors in the US stock market, we include only nine of them because of the limited data availability. Specifically, short-term and long-term behavioural factors are excluded because they are unavailable in international markets. For the same reason, we ignore the size (ME), profitability (ROE), and investment (IA) in Hou et al. (2015). Finally, we end up with nine candidates: MKT, SMB, HML, RMW, CMA, MOM, QMJ, BAB, and HML devil. Either HML or HML devil can enter the true SDF. Since the AQR library provides the QMJ factor from July 1993, and we use a three-year rolling window, our model uncertainty measure starts from June 1996.

Figure 10a plots the time series of model uncertainty in the European stock market from June 1996 to December 2020. Several results stand out. The time-series patterns in European markets[25] are remarkably similar to the US stock market. In particular, model uncertainty increases from 1999 and reaches its first peak between 2000 and 2001 because of the dot-com bubble burst. During these periods, model uncertainty almost touches its upper

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

bound. After 2002, model uncertainty declines gradually and remains relatively low until the start of the 2008 global financial crisis. During this long-lasting economic and stock market crisis, model uncertainty stays close to the upper bound from 2008 to 2012 and only declines gradually after 2012. Finally, the uncertainty index shoots up again after 2015, similar to what we observe in the US market.

We next turn to discuss the findings in Asian Pacific markets.[26] It is worth noting that we observe some unique time-series variations in Asian stock markets. According to figure 10b, model uncertainty is heightened starting from 1997 due to the profound 1997 Asian financial crisis. The dot-com bubble in 2000 led to another peak in model uncertainty, which almost reaches the upper bound. However, the Asian markets recovered quickly after 2000, so the model uncertainty index declines afterwards.

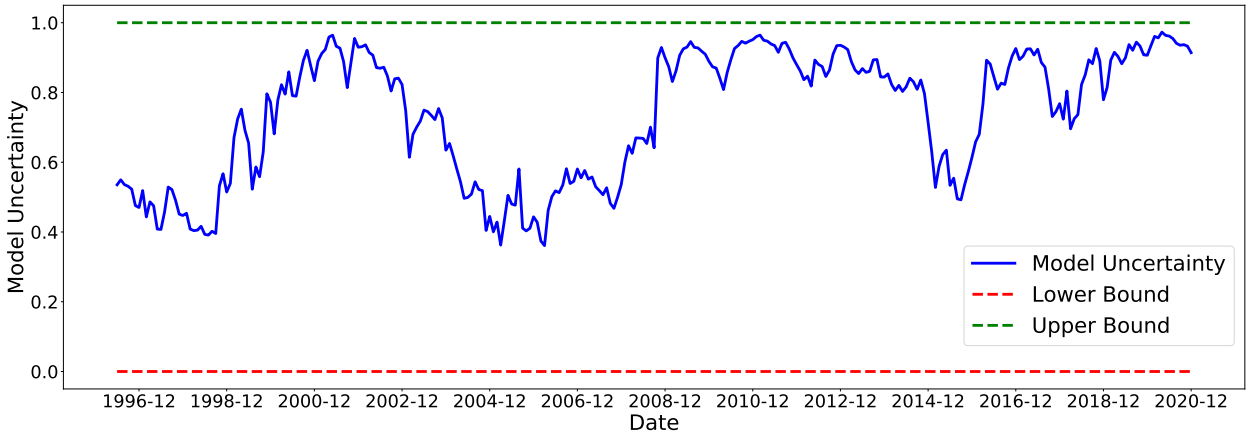
Another steady increase in model uncertainty appears before and during the 2008 crisis, but the entropy is not as high as in the late 1990s and drops immediately from 2009. This particular pattern is unlike the US and European markets, in which we observe the highest model uncertainty in the 2008 crisis. One potential explanation is that the 1997 Asian financial crisis, combined with the burst of the dot-com bubble in 2000, was more destructive than the 2008 financial crisis. There is a short-term upward jump in model uncertainty between 2011 and 2012 when the US government bonds were downgraded. Similar to US and European markets, model uncertainty surges from the beginning of 2015.

In short, the international market evidence in this section lends further support to the time-varying nature of model uncertainty. First, model uncertainty is enormous in many periods, way above its lower bound. Second, it fluctuates significantly over time and coincides with major events in corresponding asset markets. However, model uncertainty is not all alike. For example, Asian markets display unique behaviours that distinguish them from the US and European markets.

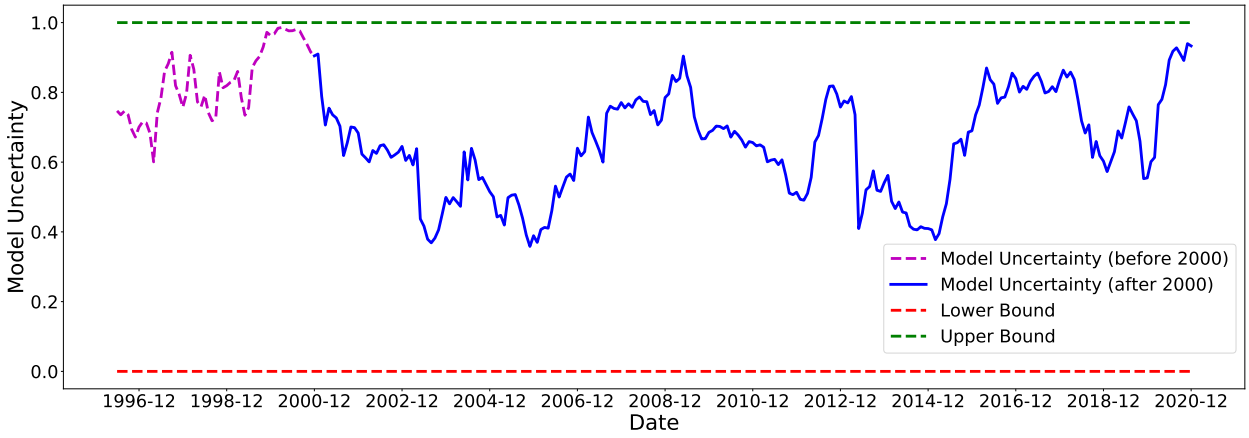
7 Robustness Checks

This section considers several robustness checks, including alternative hyper-parameter a in estimating factor models, alternative rolling windows in constructing the time series of

[26] By saying the Asian Pacific market, we refer to the stock markets in Australia, Hong Kong, New Zealand, and Singapore.



(a) European Stock Markets



(b) Asian Pacific Stock Markets

Figure 10: Model Uncertainty in European and Asian 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.

model uncertainty, and a different identification assumption under which we re-estimate the dynamic responses of fund flows to uncertainty shocks.

7.1 Alternative Hyper-Parameter a

One important choice in our Bayesian inference is the value of hyper-parameter a. In the benchmark case, we assign a to be 4. Just as Section 1 shows, a higher a implies a stronger shrinkage for factors' risk prices, $\boldsymbol{b}$.

Figure A2 plots the time series of model uncertainty using different values of a, including 3, 8, 16. Several findings stand out. First, we find that the time-series patterns in model uncertainty are not sensitive to the choice of a. In fact, the sequences under different values of a are virtually identical. Second, model uncertainty is increasing in a. This observation is not surprising since a larger a mechanically shrinks all candidate models to the null model, rendering factor models to become more similar and driving up model uncertainty. Overall, the time-series patterns in model uncertainty are driven by the data rather than the prior.

7.2 Alternative Rolling Windows

There is a trade-off in choosing the length of the rolling window. On the one hand, we prefer a larger time-series sample to achieve higher precision in estimating model parameters. The one-year or two-year daily sample is insufficient since estimating factors' expected returns and their covariance matrix is challenging. On the other hand, larger sample size is not always desirable since it implicitly assumes that factor models remain constant and robust over a long period. As many research (e.g. McLean and Pontiff (2016)) suggest, factors' performances deteriorate post-publication. Moreover, a long estimation period of 10 or 20 years will average valuable information concerning factors' cyclical behaviours.

Motivated by the above discussion, we consider four-year and five-year rolling windows in Figure A3. There is one tiny difference: Model uncertainty tends to be smoother in longer rolling windows, especially the five-year window. Beyond that, the time-series properties are similar to those found in a three-year rolling window.

7.3 Alternative Identification Assumption in VAR

Another robustness check concerns the identification assumption in our VAR analysis. In Section 4, we put uncertainty measures first in $\boldsymbol{Y_t}$. We now consider an alternative setup, in which uncertainty measures are the last variables in $\boldsymbol{Y_t}$. In other words, we allow uncertainty measures to correlate with contemporaneous shocks to mutual fund flows, but uncertainty

shocks do not affect mutual fund flows simultaneously. Although model uncertainty is an endogenous response to innovations in fund flows under this assumption, it is still worth investigating whether model uncertainty is a key player to propagate those exogenous shocks over a long-lasting period.

Figures A5 and A6 plot the IRFs of fund flows to three uncertainty measures. Under the current assumption, the IRFs are zeros at period zero by construction. The first column shows the dynamic responses to model uncertainty shocks. Similar to the observations in Figures 6 and 7, an increase in model uncertainty relates to persistent outflows from style and small-cap funds but sharp inflows to government bond funds. The dynamic effects are bounded well away from zero even beyond 36 months, although they decline slowly over time. Hence, the main results in Figures 6 and 7 are largely robust.

The second and third columns show the IRFs of fund flows using VXO and financial uncertainty. Surprisingly, VXO shocks imply positive inflows to small-cap funds. On average, one standard deviation increase in the VXO index corresponds to more than 0.1 standard deviation fund inflows, and these positive dynamic responses last for around 20 months. However, the 90% confidence interval of IRFs covers zero effects, so they are on the edge of being consequential. Beyond that, the IRFs in other panels are virtually zeros, so there is little evidence that mutual fund investors react to VXO or financial uncertainty shocks.

Finally, we observe significant inflows to money market funds following positive VXO shocks, and the dynamic responses have similar economic sizes to those in Figure 9. The key difference under the new identification assumption is that the IRFs of money market funds to financial uncertainty shocks are no longer significant. In other words, the dynamic responses to financial uncertainty shocks in Figure 9 are driven mainly by the identification assumption.

To conclude, model uncertainty has robust and persistent effects on mutual fund flows, particularly the style, small-cap, and government bond funds. We argue that model uncertainty is a crucial determinant of mutual fund flows, regardless of being an exogenous cause or a merely propagating mechanism. On the contrary, the dynamic responses of fund flows to volatility-based measures, be it VXO or financial uncertainty, are more or less sensitive to different identification assumptions. In fact, there is little evidence that equity mutual fund investors respond to VXO or uncertainty shocks.

8 Conclusions

We develop a new measure of model uncertainty in cross-sectional asset pricing. The measure is based on the entropy of Bayesian posterior probabilities for all possible factor models. The critical observation is that model uncertainty is countercyclical: it begins to climb up right before the stock market crashes and remains at its peaks during bear markets. Since we can calculate the lower and upper bound of entropy, we can easily discern when model uncertainty is abnormally high or low. In contrast, uncertainty measures in past literature do not satisfy this property. We find that model uncertainty almost touches its upper bounds in the burst of the dot-com bubble and the 2008 financial crisis.

If investors consider model uncertainty as another source of investment risk, their portfolio choice and expectations of the stock market should be naturally related to model uncertainty. Our second key observation is that model uncertainty can predict the next-period mutual fund flows, even after controlling past fund flows, VXO, and the past performance of mutual funds. In particular, investors reduce their investment in style and small-cap mutual funds but allocate more of their wealth to safer US government bond funds. Model uncertainty is also closely related to investors' expectations and confidence. We document that investors in the survey, no matter individual or institutional investors, are more pessimistic about the stock market when confronted with higher model uncertainty.

As model uncertainty is an essential source of investment risk, future theoretical research on portfolio choice should incorporate it into the model. Even though a few partial equilibrium models have considered model uncertainty about mean-variance portfolios, no such a general equilibrium model exists, at least according to our knowledge. Future research could attempt to endogenize model uncertainty in the general equilibrium model and explain its countercyclical behaviours.

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. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics 64*(3), 423–458.

Avramov, D., S. Cheng, L. Metzker, and S. Voigt (2021). Integrating factor models. *Available at SSRN 3924337*.

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

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

Barberis, N. (2000). Investing for the long run when returns are predictable. *Journal of Finance 55*(1), 225–264.

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

Bartlett, M. (1957). A comment on d.v. lindley's statistical paradox. *Biometrika 44*(3-4), 533–533.

Ben-David, I., J. Li, A. Rossi, and Y. Song (2021). Ratings-driven demand and systematic price fluctuations. *Fisher College of Business Working Paper* (2020-03), 026.

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

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

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

Bryzgalova, S., J. Huang, and C. Julliard (2021). Bayesian solutions for the factor zoo: We just ran two quadrillion models. *Available at SSRN 3481736*.

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

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

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

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

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

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

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

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

Fama, E. F. and K. R. French (1992, Jun). The cross-section of expected stock returns. *Journal of Finance 47*, 427–465.

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

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

Gabaix, X. and R. S. Koijen (2021). In search of the origins of financial fluctuations: The inelastic markets hypothesis. Technical report, National Bureau of Economic Research.

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

Giglio, S., B. Kelly, and D. Xiu (2021+). Factor models, machine learning, and asset pricing. *Annual Review of Financial Economics forthcoming.*

Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *Review of Financial Studies 27*(3), 714–746.

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

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

Harvey, C. R., Y. Liu, and H. Zhu (2016). ... and the cross-section of expected returns. *Review of Financial Studies 29*(1), 5–68.

He, Z., B. Kelly, and A. Manela (2017). Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics 126*(1), 1–35.

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

Huang, S., Y. Song, and H. Xiang (2021). Noise trading and asset pricing factors. *Available at SSRN 3359356*.

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.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 186*(1007), 453–461.

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.

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.

Koijen, R. S. and M. Yogo (2019). A demand system approach to asset pricing. *Journal of Political Economy 127*(4), 1475–1515.

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

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

Li, J. (2021). What drives the size and value factors? *Available at SSRN 2909960*.

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.

Lindley, D. V. (2000). The philosophy of statistics. *Journal of the Royal Statistical Society: Series D (The Statistician) 49*(3), 293–337.

Lintner, J. (1965, December). Security prices, risk, and maximal gains from diversification. *Journal of Finance 20*, 587–615.

Lou, D. (2012). A flow-based explanation for return predictability. *Review of Financial Studies 25*(12), 3457–3489.

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.

McLean, R. D. and J. Pontiff (2016). Does academic research destroy stock return predictability? *Journal of Finance 71*(1), 5–32.

Newey, W. K. and K. D. West (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–08.

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

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

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

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

Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory 13*(3), 341–360.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance 19*(3), 425–42.

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

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

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 Description of Factors

CAPM. The CAPM in Sharpe (1964) and Lintner (1965) is the pioneer of linear factor models. The only factor in CAPM is the excess return on the market portfolio (MKT). The data comes from Ken French's website.

Fama-French Five-factor model. Fama and French (1992) extend CAPM by introducing SMB and HML, where SMB is the return difference between portfolios of small and large stocks, and HML is the return difference between portfolios of stocks with high and low book-to-market ratios. Fama and French (2016) further include a profitability factor (RMW) and one investment factor (CMA). Again, the data comes from Ken French's website.

Momentumn. 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. Therefore, investors can outperform the market by buying past winners and selling past losers. We download the momentum (MOM) factor from Ken French's data library.

q-factor model. Hou et al. (2015) introduce a four-factor model that includes market excess return (MKT), a new size factor (ME), an investment factor (IA), and finally, the profitability factor (ROE).[27]

Behavioral Factors. Daniel et al. (2020) propose a three-factor model consisting of the market factor and two theory-based behavioural factors. The short-term behavioural factor is based on the post-earnings announcement drift (PEAD) and captures the underreaction to quarterly earnings announcements in the short horizon. Instead, the long-term behavioural factor (FIN) is based on the one-year net and five-year composite share issuance.

Quality-minus-junk. Asness et al. (2019) groups the listed companies into the quality and junk stocks. They find that a quality-minus-junk (QMJ) strategy generate high positive abnormal returns. We download the QMJ factor from the AQR data library.

Betting-against-beta. One of the most prominent failures of CAPM is that the security market line is too flat, so the risk premia of high-beta stocks are not as substantial as CAPM suggests. 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

[27] We are grateful to the authors for sharing the data with us.

the BAB factor from the AQR data library.

HML Devil. Asness and Frazzini (2013) propose an alternative way to construct the value factor, which relies on more timely market value information. We download the HML Devil factor from the AQR data library.

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

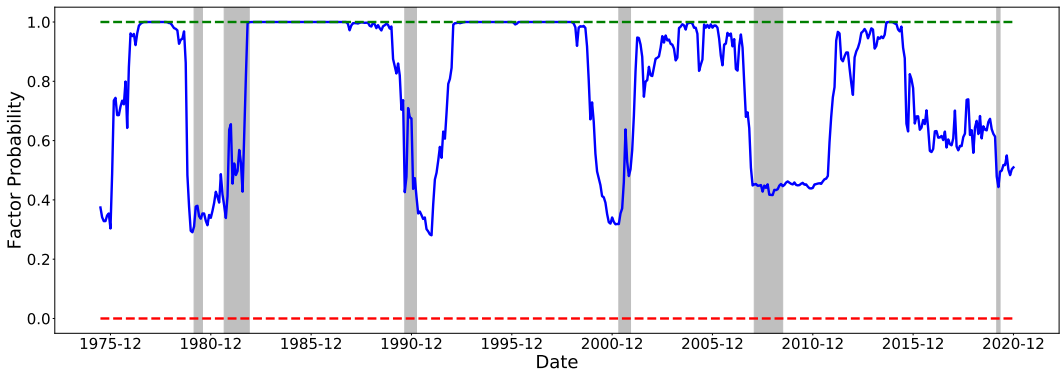
This table reports the annualised mean returns and Sharpe ratios of 14 factors listed in Appendix A. The full sample starts from July 1972 to December 2020.

Table A2: Summary of First-Order Autoregression

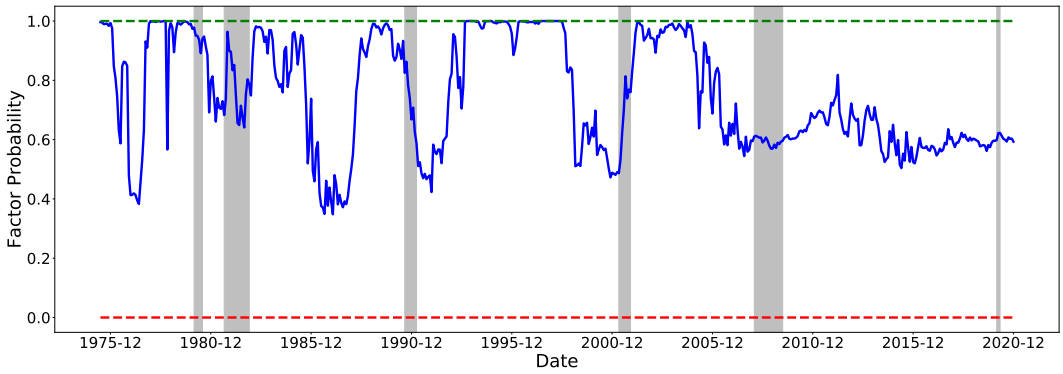
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Entropy	Financial	Macro	Real	EPU_1	EPU_2	VXO
AR(1)	0.986***	0.977***	0.985***	0.984***	0.844***	0.700***	0.812***
	(158.08)	(98.78)	(73.92)	(46.84)	(24.64)	(14.30)	(23.40)
Sample size	546	546	546	546	431	431	419
R^2	0.9697	0.9523	0.9667	0.9514	0.6929	0.5945	0.6586

The table reports empirical results in the first-order autoregression of seven uncertainty measures: $y_{t+1} = \alpha + \rho y_t + \epsilon_{t+1}$. Entropy is our model uncertainty measure. Financial, macro and real uncertainty measures come from Ludvigson et al. (2021) and Jurado et al. (2015). EPU_1 and EPU_2 are two economic policy uncertainty sequences from Baker et al. (2016). VXO is the forward-looking market volatility traded in CME. The t-statistics are computed using Newey-West standard errors with 36 lags. *, ** and *** denote significance at the 90%, 95%, and 99% level, respectively.

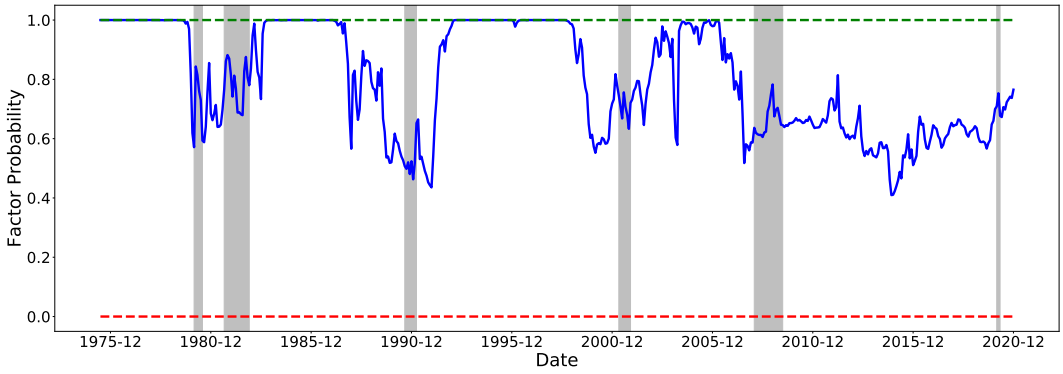
Appendix C Additional Figures



(a) MKT

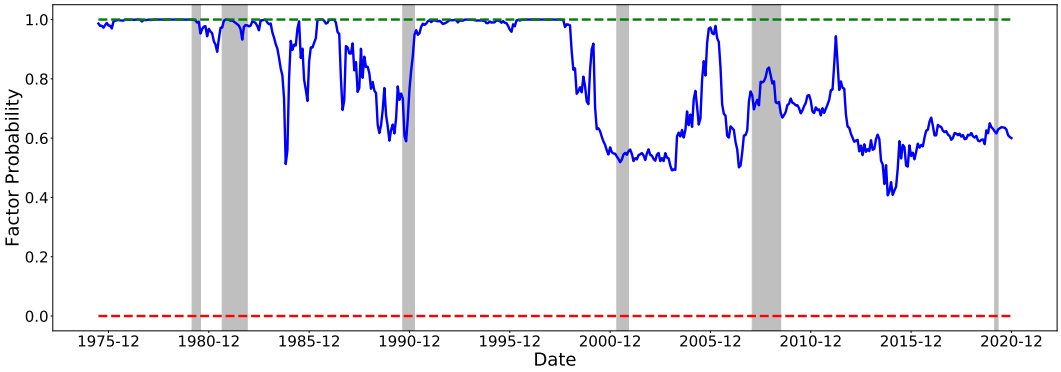


(b) Size (SMB or ME)

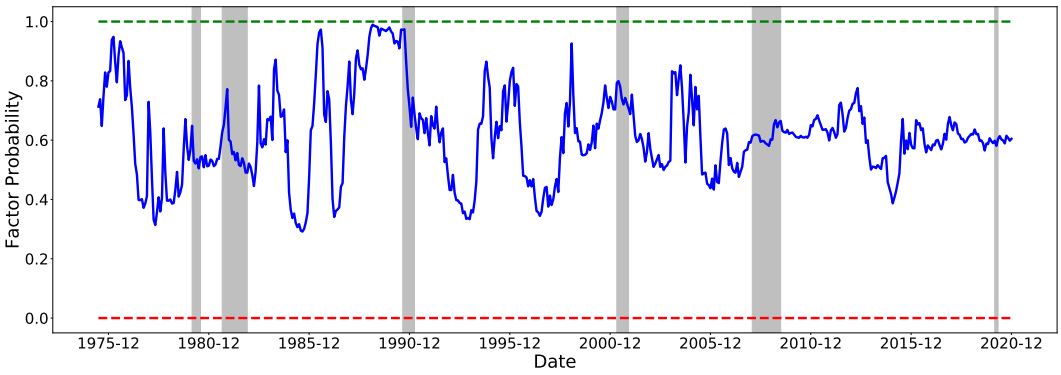


(c) Value (HML or HML_devil)

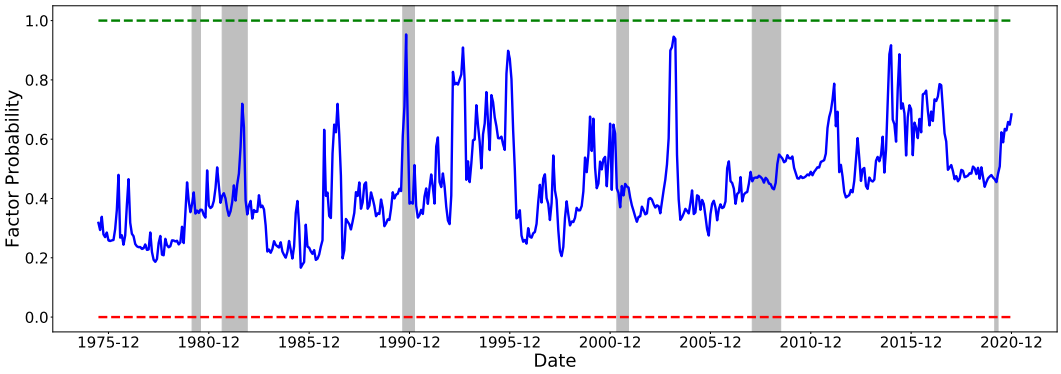
Figure A1: Time Series of Posterior Factor Probabilities



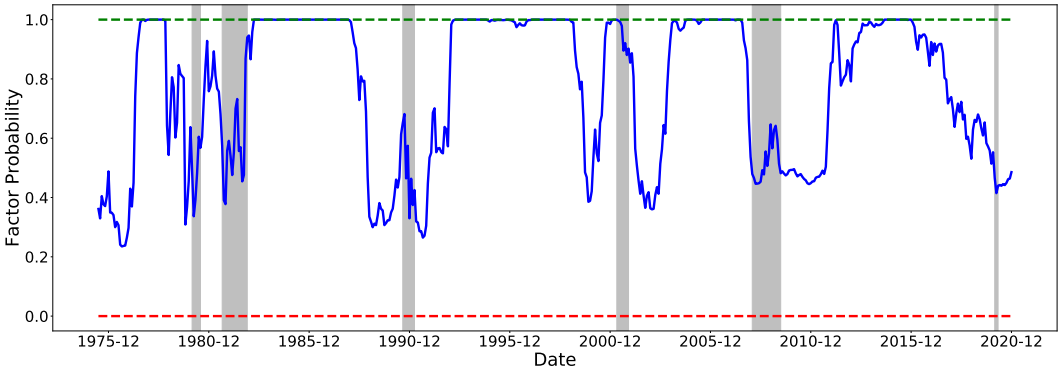
(d) Profitability (ROE or RMW)



(e) Investment (IA or CMA)

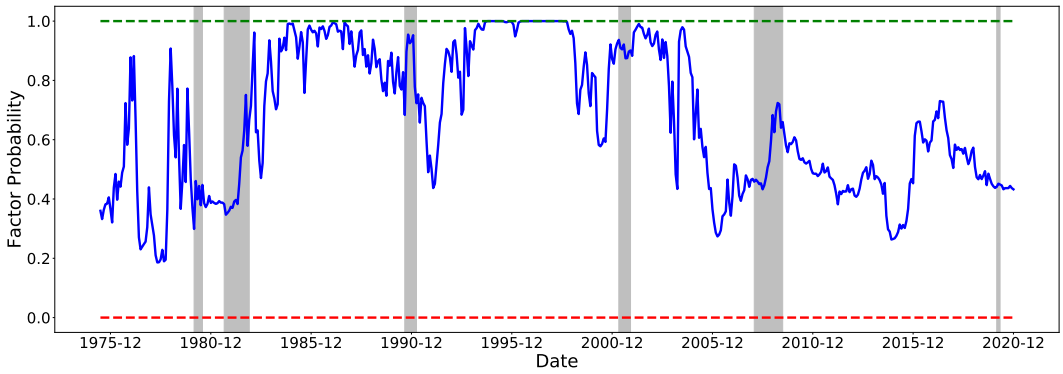


(f) Momentum

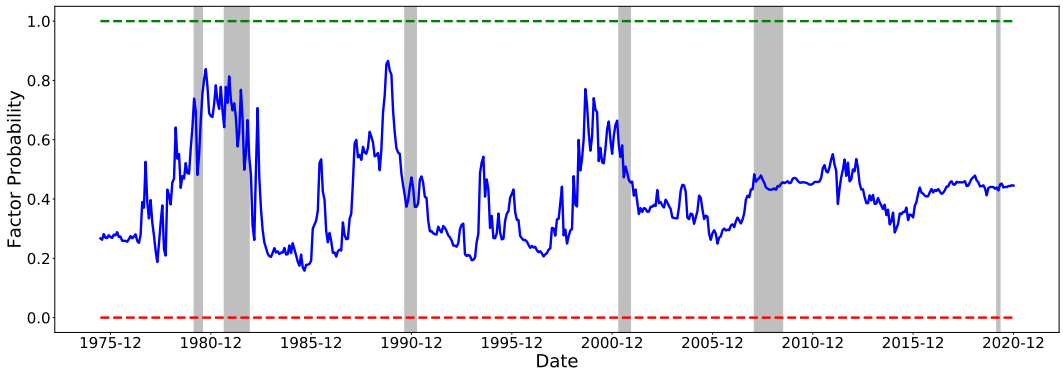


(g) BAB

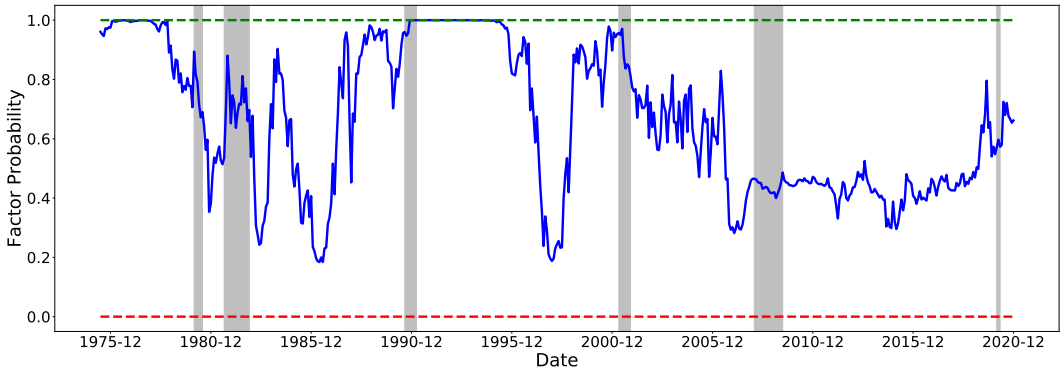
Figure A1: Time Series of Posterior Factor Probabilities (Continued)



(h) QMJ



(i) FIN



(j) PEAD

Figure A1: 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 start from June 1975. Shaded areas are NBER-based recession periods for the US.

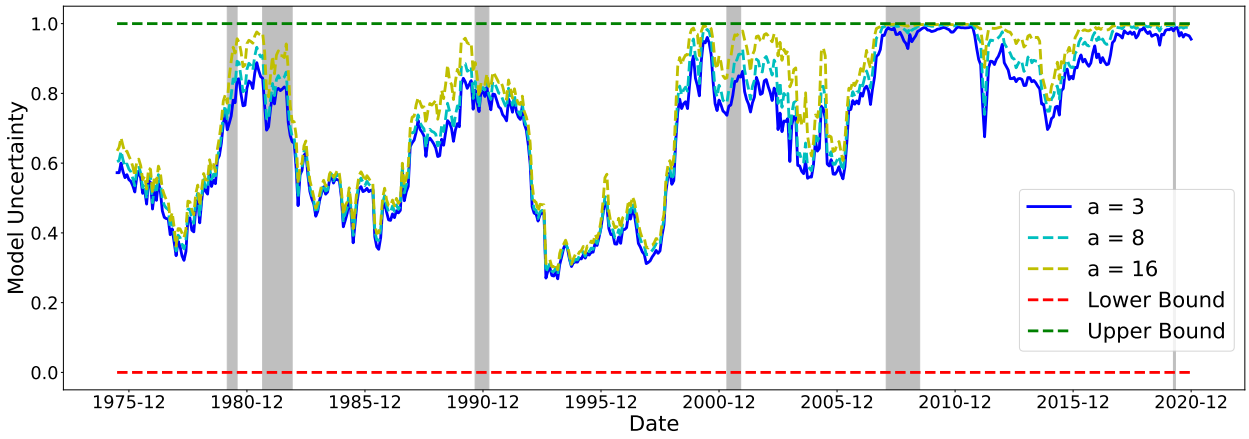
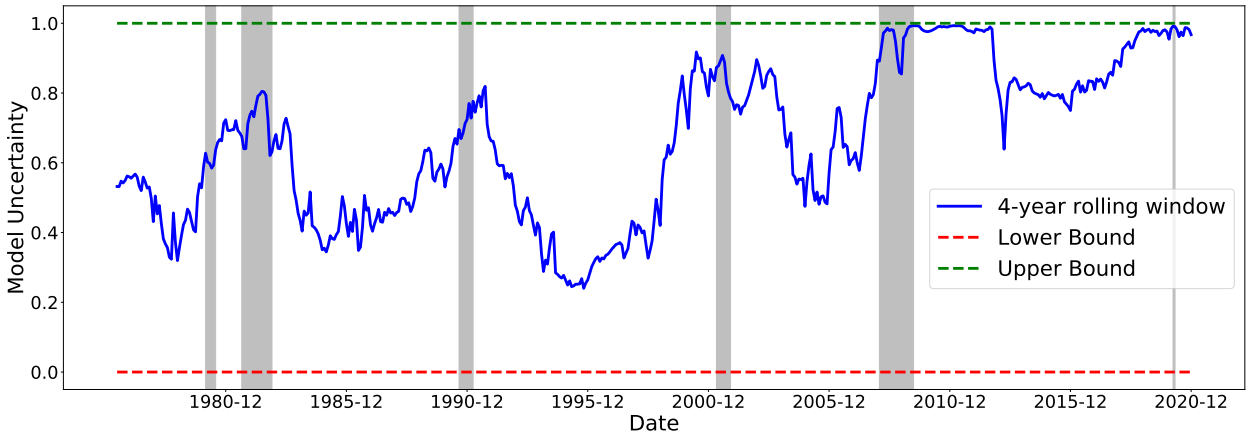
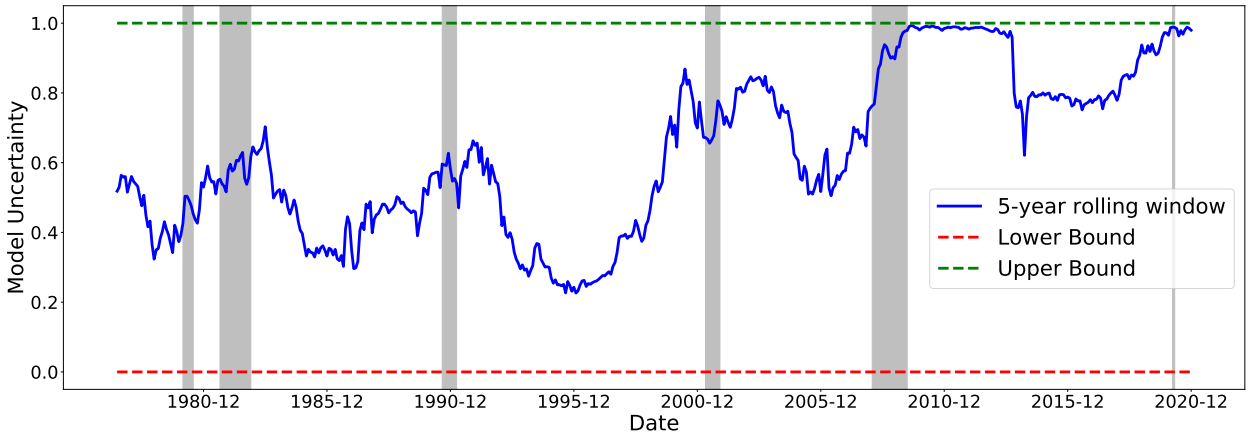


Figure A2: Time-Series of Model Uncertainty (3-Year Rolling Window) using different values of the hyper-parameter, $a \in \{3, 8, 16\}$

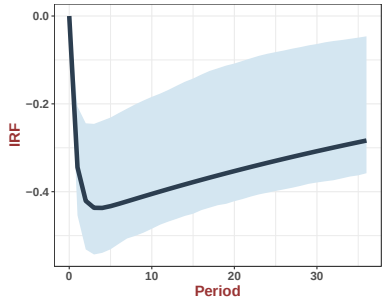


(a) Model Uncertainty in 4-Year Rolling Window

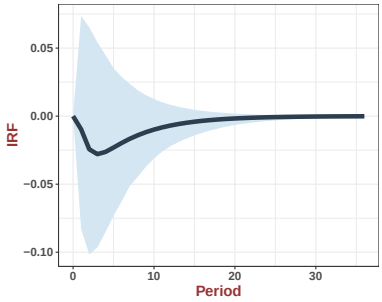


(b) Model Uncertainty in 5-Year Rolling Window

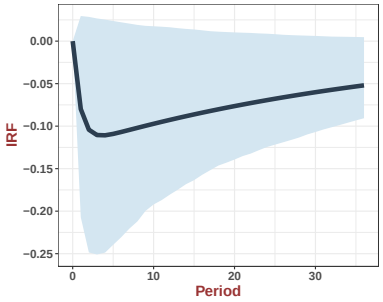
Figure A3: Alternative Rolling Windows



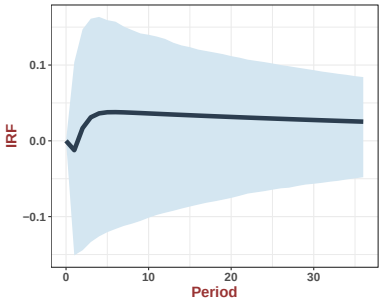
(a) Equity Fund Flows to Entropy



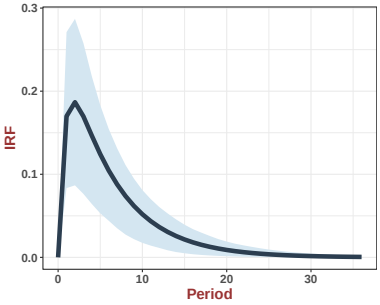
(b) Equity Fund Flows to VXO



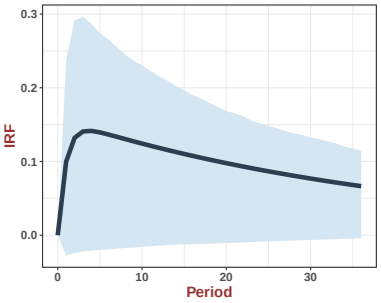
(c) Equity Fund Flows to Financial Uncertainty



(d) Fixed-Income Fund Flows to Entropy



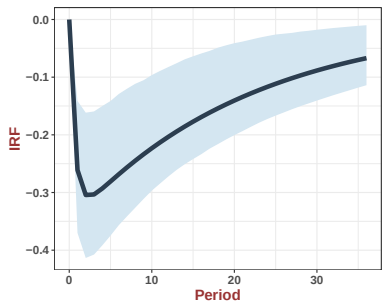
(e) Fixed-Income Fund Flows to VXO



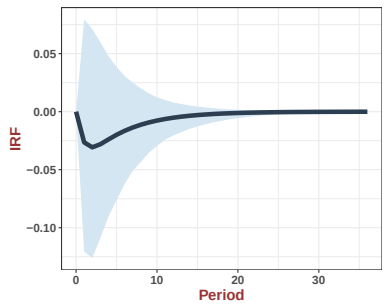
(f) Fixed-Income Fund Flows to Financial Uncertainty

Figure A4: Robustness Check: Impulse Responses of Equity and Fixed-Income Fund Flows under Alternative Identification Assumption

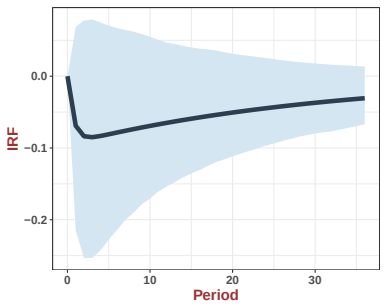
This figure shows the dynamic impulse response functions (IRFs) of equity and fixed-income fund flows to uncertainty shocks in VAR-1. We identity the IRFs by putting uncertainty last in VAR.



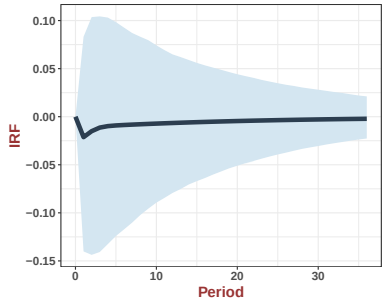
(a) Style Fund Flows to Entropy



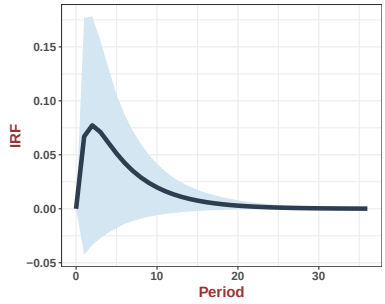
(b) Style Fund Flows to VXO



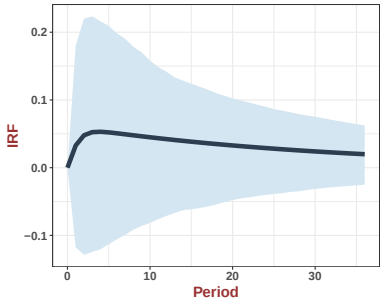
(c) Style Fund Flows to Financial Uncertainty



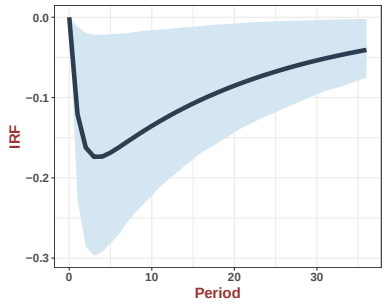
(d) Sector Fund Flows to Entropy



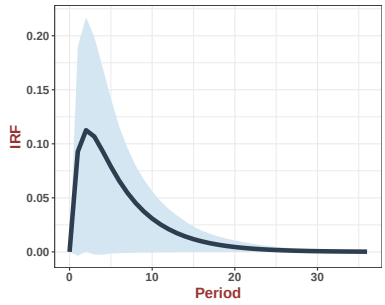
(e) Sector Fund Flows to VXO



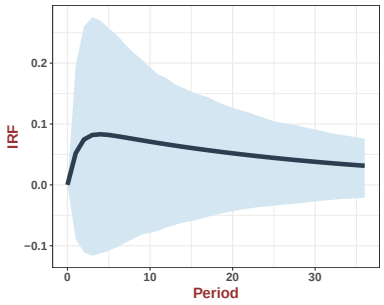
(f) Sector Fund Flows to Financial Uncertainty



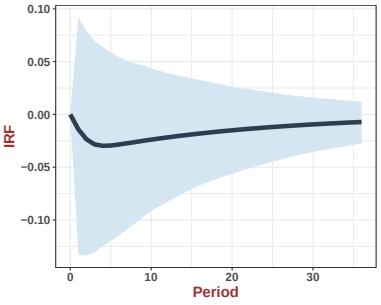
(g) Small-Cap Fund Flows to Entropy



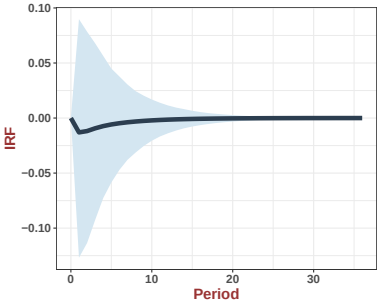
(h) Small-Cap Fund Flows to VXO



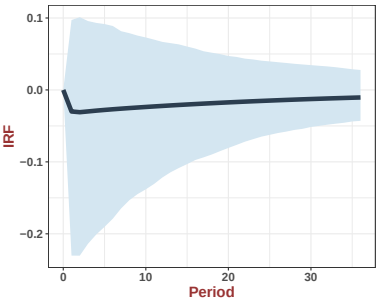
(i) Small-Cap Fund Flows to Financial Uncertainty



(j) Large-Cap Fund Flows to Entropy



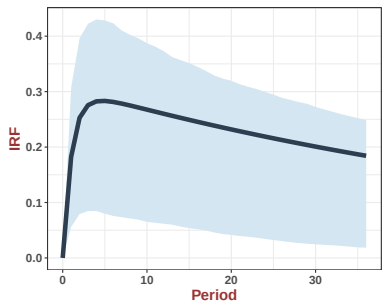
(k) Large-Cap Fund Flows to VXO



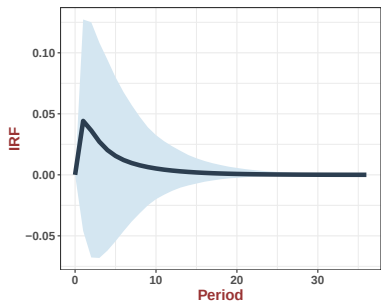
(l) Large-Cap Fund Flows to Financial Uncertainty

Figure A5: Robustness Check: Impulse Responses of Equity Fund Flows with Different Investment Objective Codes under Alternative Identification Assumption

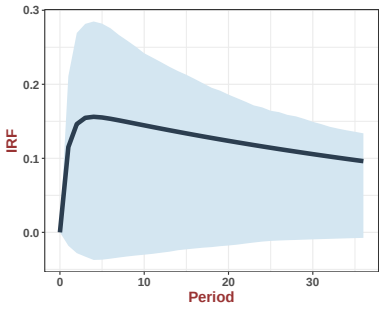
This figure shows the dynamic impulse response functions (IRFs) of equity fund flows to uncertainty shocks in VAR-1. We identity the IRFs by putting uncertainty last in VAR.



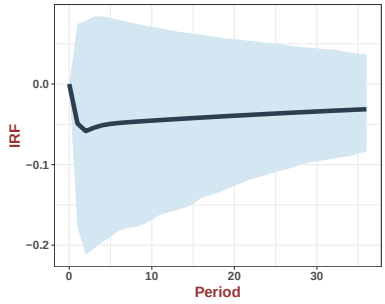
(a) Government Bonds to Entropy



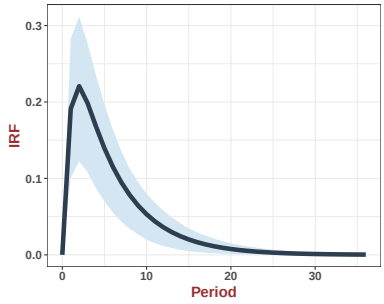
(b) Government Bonds to VXO



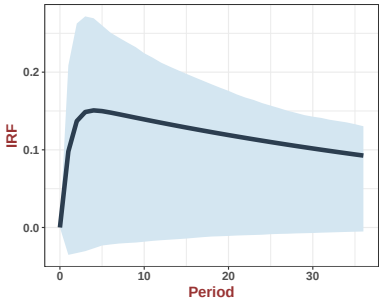
(c) Government Bonds to Financial Uncertainty



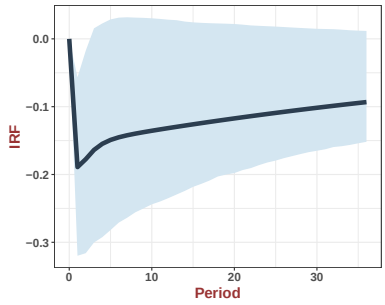
(d) Money Markets to Entropy



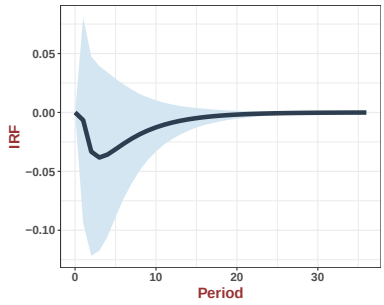
(e) Money Markets to VXO



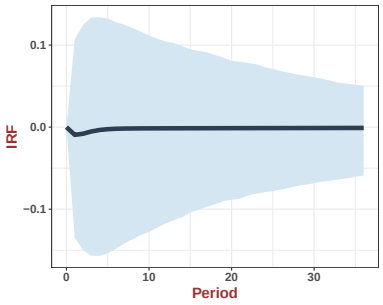
(f) Money Markets to Financial Uncertainty



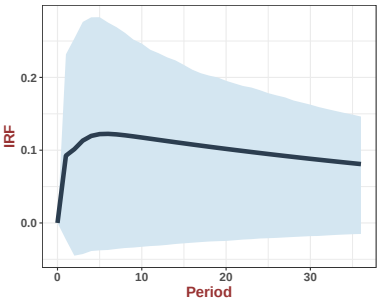
(g) Corporate Bonds to Entropy



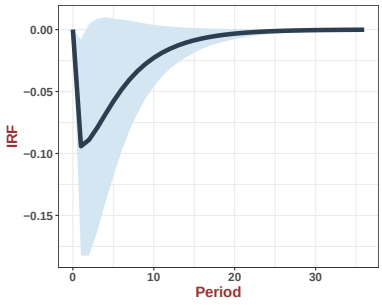
(h) Corporate Bonds to VXO



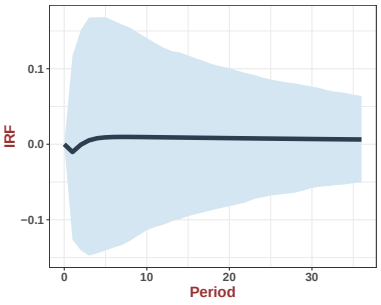
(i) Corporate Bonds to Financial Uncertainty



(j) Municipal Bonds to Entropy



(k) Municipal Bonds to VXO



(l) Municipal Bonds to Financial Uncertainty

Figure A6: Robustness Check: Impulse Responses of Fixed-Income Fund Flows with Different Investment Objective Codes under Alternative Identification Assumption

This figure shows the dynamic impulse response functions (IRFs) of fixed-income fund flows to uncertainty shocks in VAR-1. We identity the IRFs by putting uncertainty last in VAR.

Appendix D Proofs

D.1 Proof of Proposition 1

Proof. As in section 1.2, we assign g-prior for $\boldsymbol{b}_{\boldsymbol{\gamma}}$: $\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g \sim \mathcal{N}\left(\mathbf{0}, \frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right)$. From lemma 1, $\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{V}_{\boldsymbol{\gamma}}$, so the prior distribution for $\boldsymbol{b}_{\boldsymbol{\gamma}}$ is simplified as $\mathcal{N}\left(\mathbf{0}, \frac{g}{T}\boldsymbol{V}_{\boldsymbol{\gamma}}^{-1}\right)$. Thus, the variance of linear SDF $m_{\boldsymbol{\gamma}}$, conditioned that g and $\boldsymbol{V}_{\boldsymbol{\gamma}}$ are known, is

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

This completes the proof of Proposition 1. □

D.2 Proof of Proposition 2

We begin the proof of Proposition 2 with the following lemma.

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

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

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 $\boldsymbol{r}_t^e$ is a vector of test assets that are excess returns themselves but are excluded from

factors under consideration (i.e., $\boldsymbol{f}$). Then

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

where

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

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

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

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

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

Thus

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

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

Under this lemma, we prove Proposition 2 as follows.

Proof. Since

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

under our distributional assumption and

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

under our g-prior specification, we can integrate out $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and reach the following distributional results for the observed dataset $\mathcal{D} = \{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T\}$:

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

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

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

By the Sherman-Morrison-Woodbury formula,[28]

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

By the generalized Sylvester's theorem for determinants,[29]

$$\begin{aligned}
&\left|\boldsymbol{I}_T \otimes \boldsymbol{\Sigma} + \frac{g}{T}(\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}}) \left(\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1} (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^\top\right| \\
=&\frac{|Tg^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}} + (\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})^\top (\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1})(\boldsymbol{1}_T \otimes \boldsymbol{C}_{\boldsymbol{\gamma}})|}{|Tg^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}| \times |\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}|} \\
=&\frac{|(g^{-1}+1)\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}|}{|g^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}}| \times |\boldsymbol{I}_T \otimes \boldsymbol{\Sigma}^{-1}|} \\
=&\frac{(1+g)^{p_{\boldsymbol{\gamma}}}}{|\boldsymbol{\Sigma}^{-1}|^T},
\end{aligned}$$

the last equation of which is due to the fact that $\boldsymbol{C}_{\boldsymbol{\gamma}}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{V}_{\boldsymbol{\gamma}}$ according to the lemma.

[28] $(A+UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$ for "well-behaved" matrices A, U, C, V.
[29] $|X + ACB| = |X| \times |C| \times |C^{-1} + BX^{-1}A|$ for "well-behaved" matrices A, B, C, X.

Plugging the two results above back to our original formula of $\mathbb{P}[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]$, we get

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

where $\overline{\boldsymbol{R}} = \left(\sum_{t=1}^{T}\boldsymbol{R}_t\right)/T$, $\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} = \left(\overline{\boldsymbol{f}}_{\boldsymbol{\gamma},t}\right)/T$; the second equation is due to the fact that $\boldsymbol{R}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{f}_{\boldsymbol{\gamma}}$ and that $\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}} = \boldsymbol{V}_{\boldsymbol{\gamma}}$ as demonstrated in the lemma. □

D.3 Proof of Proposition 3

Proof. Now we prove proposition 2 and 3. We assume that the observed excess returns are generated from a multivariate Gaussian distribution:

$$
\boldsymbol{R}_1, \ldots, \boldsymbol{R}_T \mid \mathcal{M}_{\boldsymbol{\gamma}}, b, g \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{C}_{\boldsymbol{\gamma}}\boldsymbol{b}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}). \tag{19}
$$

The likelihood function of observed data $\mathcal{D} = \{\boldsymbol{R}_t\}_{t=1}^{T}$ is

$$
p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, b, g] = (2\pi)^{-\frac{NT}{2}}|\boldsymbol{\Sigma}|^{-\frac{T}{2}}\exp\{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{R}_t - \boldsymbol{C}_{\boldsymbol{\gamma}}\boldsymbol{b}_{\boldsymbol{\gamma}})^{\intercal}\boldsymbol{\Sigma}^{-1}(\boldsymbol{R}_t - \boldsymbol{C}_{\boldsymbol{\gamma}}\boldsymbol{b}_{\boldsymbol{\gamma}})\}. \tag{20}
$$

In order to find the posterior model probabilities, we need to derive the marginal likelihood of data $\mathcal{D}$ conditional on model $\mathcal{M}_{\boldsymbol{\gamma}}$. First of all, we find $p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]$ by integrating

out $\boldsymbol{b}_{\boldsymbol{\gamma}}$. We assign g-prior for $\boldsymbol{b}_{\boldsymbol{\gamma}}$: $\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}} \sim \mathcal{N}\left(\mathbf{0},\, \frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}\right)$, thus

$$
\begin{aligned}
p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}},\, g] &= \int p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}},\, b,\, g]\pi[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}},\, g]d\boldsymbol{b}_{\boldsymbol{\gamma}} \\
&= \int (2\pi)^{-\frac{NT}{2}}|\boldsymbol{\Sigma}|^{-\frac{T}{2}} \exp\{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{R}_t - \boldsymbol{C}_{\boldsymbol{\gamma}}\boldsymbol{b}_{\boldsymbol{\gamma}})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{R}_t - \boldsymbol{C}_{\boldsymbol{\gamma}}\boldsymbol{b}_{\boldsymbol{\gamma}})\} \\
&\qquad (2\pi)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}|\frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}|^{-\frac{1}{2}} \exp\big\{-\frac{T}{2g}\boldsymbol{b}_{\boldsymbol{\gamma}}^{\mathsf{T}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{b}_{\boldsymbol{\gamma}}\big\}d\boldsymbol{b}_{\boldsymbol{\gamma}} \\
&= (2\pi)^{-\frac{p_{\boldsymbol{\gamma}}+NT}{2}}|\boldsymbol{\Sigma}|^{-\frac{T}{2}}|\frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}|^{-\frac{1}{2}} \exp\big\{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{R}_t^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_t\big\} \\
&\qquad \int exp\big\{-\frac{T}{2}[\frac{1+g}{g}\boldsymbol{b}_{\boldsymbol{\gamma}}^{\mathsf{T}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{b}_{\boldsymbol{\gamma}} - 2\boldsymbol{b}_{\boldsymbol{\gamma}}^{\mathsf{T}}C_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}]\big\}d\boldsymbol{b}_{\boldsymbol{\gamma}},
\end{aligned}
$$

where $\overline{\boldsymbol{R}} = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{R}_t$. Let

$$
\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}} = \frac{g}{1+g}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}C_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}},
$$

$$
\hat{\boldsymbol{\Sigma}}_{\boldsymbol{b}} = \frac{1}{T}\frac{g}{1+g}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1},
$$

so the posterior distribution of $\boldsymbol{b}$ conditional on $(\mathcal{D},\, g,\, \mathcal{M}_{\boldsymbol{\gamma}})$ is

$$
\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{D},\, g,\, \mathcal{M}_{\boldsymbol{\gamma}} \sim \mathcal{N}(\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{b}}),
$$

$$
\boldsymbol{b}_{-\boldsymbol{\gamma}} \mid \mathcal{D},\, g,\, \mathcal{M}_{\boldsymbol{\gamma}} = 0.
$$

We further simplify the integral term in $p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}},\, g]$:

$$
\begin{aligned}
&\int exp\big\{-\frac{T}{2}[\frac{1+g}{g}\boldsymbol{b}_{\boldsymbol{\gamma}}^{\mathsf{T}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{b}_{\boldsymbol{\gamma}} - 2\boldsymbol{b}_{\boldsymbol{\gamma}}^{\mathsf{T}}C_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}]\big\}d\boldsymbol{b}_{\boldsymbol{\gamma}} \\
&= \exp\{\frac{1}{2}\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}}^{\top}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{b}}^{-1}\hat{\boldsymbol{b}}_{\boldsymbol{\gamma}}\}\int exp\{-\frac{1}{2}(\boldsymbol{b}_{\boldsymbol{\gamma}} - \hat{\boldsymbol{b}}_{\boldsymbol{\gamma}})^{\top}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{b}}^{-1}(\boldsymbol{b}_{\boldsymbol{\gamma}} - \hat{\boldsymbol{b}}_{\boldsymbol{\gamma}})\}d\boldsymbol{b}_{\boldsymbol{\gamma}} \\
&= \exp\{\frac{gT}{2(1+g)}\overline{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{C}_{\boldsymbol{\gamma}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}\}(2\pi)^{\frac{p_{\boldsymbol{\gamma}}}{2}}|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{b}}|^{\frac{1}{2}} \\
&= (2\pi)^{\frac{p_{\boldsymbol{\gamma}}}{2}}|\frac{1}{T}\frac{g}{1+g}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}|^{\frac{1}{2}} \exp\{\frac{gT}{2(1+g)}\overline{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{C}_{\boldsymbol{\gamma}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}\} \\
&= (1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}(2\pi)^{\frac{p_{\boldsymbol{\gamma}}}{2}}|\frac{g}{T}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)^{-1}|^{\frac{1}{2}} \exp\{\frac{gT}{2(1+g)}\overline{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{C}_{\boldsymbol{\gamma}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}\}
\end{aligned}
$$

where $\overline{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{C}_{\boldsymbol{\gamma}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}} = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^{\mathsf{T}}\boldsymbol{V}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} = SR_{\boldsymbol{\gamma}}^2$, the in-sample maximal squared Sharpe ratio that can be achieved by investing in the factors under model $\mathcal{M}_{\boldsymbol{\gamma}}$. Plug it into the expression of $p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]$ above, we have

$$
\begin{aligned}
p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g] &= \frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}}\exp\Big\{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{R}_t^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_t + \frac{gT}{2(1+g)}\overline{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{C}_{\boldsymbol{\gamma}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}}\Big\} \\
&= \frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}}\exp\Big\{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{R}_t^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_t + \frac{gT}{2(1+g)}SR_{\boldsymbol{\gamma}}^2\Big\}
\end{aligned}
$$

To make $p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g]$ more transparent, we rewrite $\sum_{t=1}^{T}\boldsymbol{R}_t^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_t$:

$$
\begin{aligned}
\sum_{t=1}^{T}\boldsymbol{R}_t^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{R}_t &= \sum_{t=1}^{T}(\boldsymbol{R}_t - \overline{\boldsymbol{R}} + \overline{\boldsymbol{R}})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{R}_t - \overline{\boldsymbol{R}} + \overline{\boldsymbol{R}}) \\
&= \sum_{t=1}^{T}(\boldsymbol{R}_t - \overline{\boldsymbol{R}})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{R}_t - \overline{\boldsymbol{R}}) + T\overline{\boldsymbol{R}}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}} \\
&= tr(\boldsymbol{\Sigma}^{-1}\sum_{t=1}^{T}(\boldsymbol{R}_t - \overline{\boldsymbol{R}})(\boldsymbol{R}_t - \overline{\boldsymbol{R}})^{\mathsf{T}}) + TSR_{max}^2
\end{aligned}
$$

Finally, we end up with the formula in Proposition 2, that is,

$$
\begin{aligned}
p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g] &= \exp\Big\{-\frac{1}{2}tr(\boldsymbol{\Sigma}^{-1}\sum_{t=1}^{T}(\boldsymbol{R}_t - \overline{\boldsymbol{R}})(\boldsymbol{R}_t - \overline{\boldsymbol{R}})^{\mathsf{T}}) - \frac{T}{2}\left(\mathrm{SR}_{\max}^2 - \frac{g}{1+g}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right)\Big\}\frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}} \\
&= \exp\left\{-\frac{T-1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) - \frac{T}{2}\left(\mathrm{SR}_{\max}^2 - \frac{g}{1+g}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right)\right\}\frac{(1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}}{(2\pi)^{\frac{NT}{2}}|\boldsymbol{\Sigma}|^{\frac{T}{2}}}
\end{aligned}
\tag{21}
$$

However, when we compare different models, the common factor unrelated to $(\mathcal{M}_{\boldsymbol{\gamma}}, g)$ can be ignored, so we simplify the marginal likelihood of data as following:

$$
p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}, g] \propto (1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}}\exp\left\{\frac{gT}{2(1+g)}\mathrm{SR}_{\boldsymbol{\gamma}}^2\right\} \tag{22}
$$

An equivalent way to think about equation (22) is to treat it as the Bayes factor of model $\mathcal{M}_{\boldsymbol{\gamma}}$ relative to $\mathcal{M}_0$. One amazing fact is that $p[\mathcal{D} \mid \mathcal{M}_0, g]$ does not depend on g[30].

[30] $p[\mathcal{D} \mid \mathcal{M}_0, g] = (2\pi)^{-\frac{NT}{2}}|\boldsymbol{\Sigma}|^{-\frac{T}{2}}\exp\left\{-\frac{T-1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right) - \frac{T}{2}\mathrm{SR}_{\max}^2\right\}$.

Therefore, the Bayes factor could be defined as

$$\mathrm{BF}_{\boldsymbol{\gamma}}(g) = \frac{p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}},\, g]}{p[\mathcal{D} \mid \mathcal{M}_0,\, g]} = (1+g)^{-\frac{p_{\boldsymbol{\gamma}}}{2}} \exp\left\{\frac{gT}{2(1+g)}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right\}$$

The prior for g is such that $\pi[g] = \frac{a-2}{2}(1+g)^{-\frac{a}{2}}$. We calculate the marginal likelihood of data only conditional on model $\mathcal{M}_{\boldsymbol{\gamma}}$ by integrating out g in equation (22).

$$\begin{aligned}
p[\mathcal{D} \mid \mathcal{M}_{\boldsymbol{\gamma}}] &\propto \frac{a-2}{2}\int_0^\infty (1+g)^{-\frac{p_{\boldsymbol{\gamma}}+a}{2}} \exp\left\{\frac{g}{1+g}\left[\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right]\right\} \mathrm{d}g \\
&= \frac{a-2}{2}\exp\left\{\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right\}\int_0^\infty (1+g)^{-\frac{p_{\boldsymbol{\gamma}}+a}{2}} \exp\left\{-\frac{1}{1+g}\left[\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right]\right\} \mathrm{d}g \\
&= \frac{a-2}{2}\exp\left\{\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right\}\int_0^1 k^{\frac{p_{\boldsymbol{\gamma}}+a}{2}-2} \exp\left\{-k\left[\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right]\right\} \mathrm{d}k \\
&= \frac{a-2}{2}\exp\left\{\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right\}\left(\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right)^{1-\frac{p_{\boldsymbol{\gamma}}+a}{2}}\int_0^{\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}} t^{\frac{p_{\boldsymbol{\gamma}}+a}{2}-2}e^{-t}\,\mathrm{d}t \\
&= \frac{a-2}{2}\exp\left\{\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right\}\left(\frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right)^{-s_{\boldsymbol{\gamma}}}\underline{\Gamma}\left(s_{\boldsymbol{\gamma}},\ \frac{T}{2}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right)
\end{aligned}$$

where $\underline{\Gamma}(s,x) = \int_0^x t^{s-1}e^{-t}\,\mathrm{d}t$ is the lower incomplete Gamma function; the scalar $s_{\boldsymbol{\gamma}}$ is defined as $s_{\boldsymbol{\gamma}} = \frac{p_{\boldsymbol{\gamma}}+a}{2} - 1$. We have proved the formula of Bayes factor $\mathrm{BF}_{\boldsymbol{\gamma}}$ in Proposition 3. To prove that the Bayes factor is always increasing in $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ always decreasing in $p_{\boldsymbol{\gamma}}$, we use the original representation of Bayes Factor, that is,

$$\mathrm{BF}_{\boldsymbol{\gamma}} = \frac{a-2}{2}\int_0^\infty (1+g)^{-\frac{p_{\boldsymbol{\gamma}}+a}{2}} \exp\left\{\frac{gT}{2(1+g)}\mathrm{SR}^2_{\boldsymbol{\gamma}}\right\} \mathrm{d}g$$

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

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

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

This completes the proof of Proposition 3. □