Model Uncertainty in the Cross Section

Jiantao Huang   Ran Shi *

October 28, 2021

Abstract

We develop a transparent Bayesian approach to quantify uncertainty in linear stochastic discount factor (SDF) models. We show that, for a Bayesian decision maker, posterior model probabilities increase with maximum in-sample Sharpe ratios and decrease with model dimensions. Entropy of posterior probabilities represents model uncertainty. We apply our approach to quantify the time series of model uncertainty in North American, European, and Asian Pacific equity markets. Model uncertainty is countercyclical in these markets before the 2008 financial crisis, but remains high afterwards. It predicts investors' asset allocation decisions across equity and fixed-income funds. In survey data, 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.

---

*Jiantao Huang: j.huang27@lse.ac.uk. Ran Shi: r.shi1@lse.ac.uk. Both authors are at Department of Finance, London School of Economics. We thank Svetlana Bryzgalova, Thummim Cho, Vicente Cuñat, Christian Julliard, Péter Kondor, Dong Lou, Ian Martin, and Cameron Peng for their comments.

Introduction

Recent literature has provided a wide spectrum of real and financial uncertainty measures.[1] They display pronounced time-series variation, and their innovations appear to be associated with investment decisions.

In equity markets, uncertainty may have ambiguous implications to investors' asset allocation decisions. The conventional wisdom is that investors respond to uncertainty by curtailing risk exposures (or even purchasing assets that offer countercyclical payoffs such as options). However, uncertainty may rise in periods of "Schumpeterian growth", during which investors chase glamour stocks (which tend to be risker) in search of the new Eldorado.

Existing equity market uncertainty measures focus on second moments of major index returns such as realized and implied volatilities. These uncertainty measures do not take into account a crucial challenge equity investors face: a phenomenon dubbed as the "factor zoo." If we interpret existing uncertainty measures as time-series uncertainty, an important dimension they neglect is the cross section.

We attempt to bridge this gap by creating a cross-sectional uncertainty measure and explore its implications to investors' asset allocation decisions. Specifically, we take the perspective of Bayesian investors adopting 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.

The first key innovation is that we generalize the g-prior of Zellner (1986), from which Bayesian investors update their posterior beliefs. As originally exposited in Zellner (1986), the g-prior is a natural outcome from an uninformative prior in a sequential decision making setup. In the meantime, it induces well-defined posteriors conformative to the criteria emphasized by Chib, Zeng, and Zhao (2020). Under this prior, posterior model probabilities have simple closed-form solutions, which increase with maximal in-sample Sharpe ratios and decrease with model dimensions. The result crystallizes two competing forces when forming beliefs regarding one particular model: higher in-sample profits (on paper) and model simplicity.

We define cross-sectional uncertainty regarding linear SDF models as the entropy of posterior model probabilities. The intuition is straightforward. Suppose that there are only two candidate factor models, and we are uncertain about which one is true ex-ante. One extreme case is that the first model dominates the other with a high posterior probability,

[1] Bloom (2009) measures macroeconomic uncertainty using jumps in the VIX index and investigates their real impacts. Ludvigson, Ma, and Ng (2015) and Jurado, Ludvigson, and Ng (2015) construct and compare real 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 of the Wall Street Journal articles to construct long-history uncertainty measures.

i.e., 99%. Under this scenario, the entropy is close to its lower bound zero (and we are clearly facing low uncertainty). On the contrary, if the posterior probabilities are 50-50 for the two models, the entropy reaches its maximum (a coin tossing exercise is needed to pick one model). To sum up, the higher the entropy is, the more uncertain Bayesian investors are about the factor models.

We document four sets of empirical findings based on our cross-sectional model uncertainty measures, summarized as follows.

First, we measure uncertainty regarding 14 popular factor strategies in the US stock market. Model uncertainty is time-varying and exhibits countercyclical behaviors, as in Figure 1. 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 differences). Under extreme market conditions, investors do not only face higher second-moment (volatility) and third-moment (skewness) risk, 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. We repeat the exercise in European and

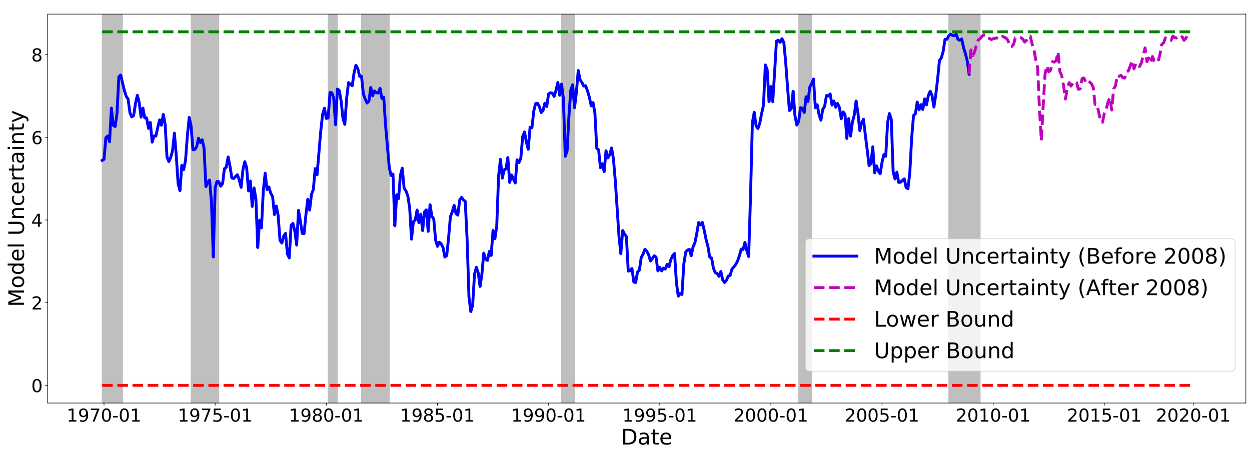


Figure 1: Time-Series of Model Uncertainty (3-Year Rolling Window)

The figure plots the time-series of model uncertainty of linear stochastic discount factor (SDF) formed from a small set of tradable factors. At the end of each month, we compute the posterior probability of each model by using the daily data of factor returns in the past 36 months. We quantify the model uncertainty in the cross section by entropy (see equation (13)). The sample ranges from January 1967 to December 2018. Since we use 3-year rolling window, the model uncertainty index starts from December 1969. The red line and green lines in the figure show the lower (0) and upper bounds (8.55) of the model uncertainty. Shaded areas are NBER-based recession periods for the United States from the period following the peak through the trough.

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 behaviors. For example, model uncertainty in this market is extremely high during 1997 Asian financial crisis.

Second, our cross-sectional uncertainty differs from conventional time-series uncertainty measures. The VIX index, macro uncertainty indices by Bloom (2009),Jurado, Ludvigson, and Ng (2015) and Ludvigson, Ma, and Ng (2015), and the economic policy uncertainty indices by Baker, Bloom, and Davis (2016) do not explain time-series variation in our uncertainty measure. This confirms that uncertainty regarding the cross-section of (expected) returns conveys additional information. It is also worthwhile noting that, our entropy-based measure enjoys a distinct property: its lower and upper bounds are always known and allow straightforward interpretations. This property makes our measure more like a barometer (which always comes with a range).

Third, we find high cross-sectional model uncertainty predicts (one-month ahead) outflows from equity mutual funds and inflows to government bond mutual funds. We treat mutual fund flows as proxies for investors' asset allocation decisions in aggregation. Moreover, we find increased cross-sectional model uncertainty predicts future outflows from "middle and small-cap" funds. We *do not* observe this predictive relationship for "large-cap" funds. These fund flows patterns do not emerge when using volatility-driven uncertainty indices such as the VIX. Investors' asset allocation decisions tend to respond to our uncertainty measure, in a way consistent with the conventional wisdom: facing high cross-sectional uncertainty, they reduce risky asset positions, especially in small-cap stocks, and reallocate proceeds into safe assets such as government bonds.

Fourth, we find that high cross-sectional model uncertainty is associated with investor sentiment. We quantify investor sentiment using surveys from American Association of Individual Investors (AAII) and 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.

Related Literature. This article mainly relates to three strands of literature. The first is research on cross-sectional asset pricing, especially those under the linear SDF framework. One of the related papers is Kozak, Nagel, and Santosh (2019). They use Lasso and ridge penalties to train a large set of characteristic-based long-short portfolios under the linear SDF specification and find that a characteristics-sparse SDF cannot explain the cross-section of asset returns well in the out-of-sample test. Another look at their conclusion is that many different factor models in their factor zoo achieve a similar maximal Sharpe ratio, so we cannot easily differentiate them empirically, so the model uncertainty is high. This implies that when we calculate the Bayesian model average, the coefficients of most characteristic factors are not close to zero. An SDF comprised of only a few factors absolutely cannot achieve superior out-of-sample performance.

Second, there is an increasing interest in developing uncertainty measures of both asset markets and economic activities. Bloom (2009) identifies 17 jumps in stock market volatility and VXO index and uses them as proxies for uncertainty shock. He further shows in a VAR analysis that a positive uncertainty shock predicts that future industrial production, productivity, and employment will decrease. Ludvigson, Ma, and Ng (2015) aims at measuring the macroeconomic uncertainty rather than asset market uncertainty. They show that their indices spike up in major economic recessions, but there is no apparent increase in macro uncertainty during some market crashes, such as the 1987 flash crash. Jurado, Ludvigson, and Ng (2015) further propose real and financial uncertainty indices. These two papers approximate the uncertainty by the conditional volatility of prediction errors in their regression analysis. Finally, Baker, Bloom, and Davis (2016) develop economic policy uncertainty indices based on news coverage. Unlike their measures, our goal is to quantify how uncertain investors are about the true model in the cross-section or the mean-variance efficient portfolio equivalently. Therefore our measure is conceptually novel.

Third, our paper also contributes to the literature on Bayesian inference for factor models and Bayesian portfolio choice. The main idea behind our g-prior is the implied "imaginary" prior sample with size related to g. Similar ideas of specifying priors are adopted in the past literature (e.g. Kandel and Stambaugh (1996) and Avramov (2002)). However, we also point out the potential Barlett's paradox (see Bartlett (1957)) in g-prior. We avoid Barlett's paradox by proposing a hyper-prior on g, following Liang, Paulo, Molina, Clyde, and Berger (2008). According to our knowledge, we are the first one to adopt this prior in the finance literature. Some other papers, such as Barillas and Shanken (2018a), focus more on the Bayesian inference in time-series regression, while our paper studies the cross-section of asset returns. Finally, although past literature has introduced model uncertainty under the portfolio choice specification (e.g., Avramov (2002), Barillas and Shanken (2018b)), we are the first one to formally develop a measure to quantify model uncertainty and discuss its time-series variation and other properties.

The paper proceeds as follows. Section I introduces the Bayesian econometric framework used in our paper. Section II reports all important empirical results. Section III provides additional evidence on the robustness of our model uncertainty measure. Finally, section IV concludes our paper.

I Methodology

Throughout our analysis, we focus on *excess* returns and study their risk premia in the cross-section. Denote by $\boldsymbol{R}$, a random vector of dimension N, the excess returns under consideration[2]. Out of these excess returns, some would be regarded as factors in a linear factor model. Common examples include the market excess return in the CAPM and long-short portfolios in empirical multi-factor asset pricing models. In terms of notation, we denote by $\boldsymbol{f}$, a subset of $\boldsymbol{R}$ with dimension p, the factors under consideration[3]. A linear factor model for these excess returns in the discount factor 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

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

where m is the stochastic discount factor that prices assets, i.e., it is such that the prices of excess returns all equal zero. Since the pricing equation (2) is scale-invariant, we normalize the constant term in the SDF to one. The covariance term, $\mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}]$, is an $N \times p$ matrix[4]. Its entry in the ith row and jth column is the covariance between excess return R_i and factor f_j.

Remark. Linear factor characterization of SDFs relates to the results of Hansen and Jagannathan (1991): Assuming no arbitrage, an SDF within the space spanned by all the excess returns under consideration can be written as

$$m = 1 - (\boldsymbol{R} - \mathbb{E}[\boldsymbol{R}])^{\top} (\mathrm{Var}[\boldsymbol{R}])^{-1} \mathbb{E}[\boldsymbol{R}].$$

Clearly, the equation $\mathbb{E}[m \times \boldsymbol{R}] \equiv 0$ always holds under the specification above. This corresponds to the case where factors under consideration are all the excess returns, i.e., $\boldsymbol{f} = \boldsymbol{R}$ and $\boldsymbol{b} = (\mathrm{Var}[\boldsymbol{R}])^{-1} \mathbb{E}[\boldsymbol{R}]$ in equation (3).

---

[2]Our definition of excess returns is in a relatively broader sense, which means that they can be returns on assets less the risk-free rate, and more generally, returns on long-short portfolio positions with zero initial costs.

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

[4]$\mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}] = \mathbb{E}\left[(\boldsymbol{R} - \mathbb{E}[\boldsymbol{R}])(\boldsymbol{f} - \mathbb{E}[\boldsymbol{f}])^{\top}\right] = \mathbb{E}\left[\boldsymbol{R}(\boldsymbol{f} - \mathbb{E}[\boldsymbol{f}])^{\top}\right]$.

I.1 A simple framework for incorporating model uncertainty

Now we would like to formalize our notion of model uncertainty. In practice, we do not know exactly which factors contribute to the pricing of assets given the other ones. From a model choice perspective, we are uncertain about which subset of factors to include into our linear SDF specification. Under our setting, given the 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 $\boldsymbol{\gamma}$ thus defines a model for the SDF[5], denoted by $\mathcal{M}_{\boldsymbol{\gamma}}$, as follows: Under $\mathcal{M}_{\boldsymbol{\gamma}}$, the linear SDF is

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

and the expected excess returns are such that

$$\mathbb{E}[\boldsymbol{R}] = \mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}} \tag{5}$$

where $\boldsymbol{f}_{\boldsymbol{\gamma}}$ is a $p_{\boldsymbol{\gamma}}$-dimensional vector that contains all the factors that are included under the current model[6]; $\boldsymbol{b}_{\boldsymbol{\gamma}}$ is a $p_{\boldsymbol{\gamma}}$-dimensional vector of nonzero factor loading; $\mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]$ is now an $N \times p_{\boldsymbol{\gamma}}$ covariance matrix. The two equations above are counterparts of (1) and (3) after introducing model uncertainty.

Models in economics and finance set restrictions on variables under investigation, most commonly through moment conditions. The linear factor SDF under model $\mathcal{M}_{\boldsymbol{\gamma}}$ does so for the distribution of all excess returns $\boldsymbol{R}$ (conditional on vector $\boldsymbol{\gamma}$), according to equation (5). The expectations of this random vector $\boldsymbol{R}$ are linked to a block of its variance-covariance matrix, namely $\mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]$[7], through a vector of coefficients $\boldsymbol{b}_{\boldsymbol{\gamma}}$.

We choose to study model uncertainty under the linear SDF specification mainly for three reasons. First, this specification enables us to focus only on the cross section of expected excess returns. Adding in the time-series dimension, model uncertainty has been introduced to panel regressions of *realized* returns on multiple factors in the literature (see

[5] For notation simplicity, we use "$-\boldsymbol{\gamma}$" to denote the set of factors that are excluded from now on. That is, it is always the case that elements in vector $\boldsymbol{f}$ are unions of elements in $\boldsymbol{f}_{\boldsymbol{\gamma}}$ and $\boldsymbol{f}_{-\boldsymbol{\gamma}}$, and the intercept of elements in $\boldsymbol{f}_{\boldsymbol{\gamma}}$ and $\boldsymbol{f}_{-\boldsymbol{\gamma}}$ is empty.

[6] $p_{\boldsymbol{\gamma}} = \sum_{j=1}^{p} I[\gamma_j = 1]$ is the total number of factors that are included under model $\mathcal{M}_{\boldsymbol{\gamma}}$.

[7] Recall that under our setting, factors are a predetermined subset of excess returns, that is, $\boldsymbol{f}_{\boldsymbol{\gamma}} \subseteq \boldsymbol{f} \subseteq \boldsymbol{R}$. Thus $\mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]$ is a sub-block of the full $N \times N$ variance-covariance matrix $\mathrm{Var}[\boldsymbol{R}]$. It is in fact an $N \times p_{\boldsymbol{\gamma}}$ matrix consisting of $p_{\boldsymbol{\gamma}}$ columns of $\mathrm{Var}[\boldsymbol{R}]$. These columns are ones such that the corresponding elements in $\boldsymbol{\gamma}$ are equal to one, just as what we have done for indexing the vector $\boldsymbol{b}$.

Avramov (2002) and Barillas and Shanken (2018a)). Factor models in these panel regressions are purely statistical, just as they are assumptions (instead of results) in Ross's arbitrage pricing theory Ross (1976). Bringing in the no arbitrage condition using a linear SDF imposes moment restrictions for the *expected* excess returns as (5). What we would like to explore is the model uncertainty after imposing these sensible restrictions, not model uncertainty based only statistical assumptions.

Second, linear factor models in the SDF form enable us to ask the following question: Does one set of factors drive out another? To understand which set of factors survive in presence of the others in terms of explaining the cross sectional variations, we should study whether the parameters in vector $\boldsymbol{b}$ are zeros or not. The latent variable $\boldsymbol{\gamma}$ for model uncertainty should be introduced to elements in $\boldsymbol{b}$, not the factor risk premia or the factor loadings (those betas). This is because, given the other factors, we may not need to include one new factor (its b is zero) even if it is priced (its market price of risk λ is not zero).

Specifically, if elements in $\boldsymbol{f}$ are all regarded as "common risk factors" à la Fama and French (1993), the vector $\boldsymbol{b}_{\boldsymbol{\gamma}}$ is related to market prices of risk, because from equation (5)

$$\begin{aligned}\mathbb{E}[\boldsymbol{R}] &= \mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}_{\boldsymbol{\gamma}}]\{\mathrm{Var}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\}^{-1}\mathrm{Var}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}} \\ &= \boldsymbol{B}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\lambda}_{\boldsymbol{\gamma}},\end{aligned}$$

where $\boldsymbol{B}_{\boldsymbol{\gamma}} = \{\mathrm{Var}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\}^{-1}\mathrm{Cov}[\boldsymbol{f}_{\boldsymbol{\gamma}}, \boldsymbol{R}]$ are the "beta" risks and $\boldsymbol{\lambda}_{\boldsymbol{\gamma}} = \mathrm{Var}[\boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}}$ are the factor risk premiums.

Noticing the link between $\boldsymbol{b}_{\boldsymbol{\gamma}}$ and $\boldsymbol{\lambda}_{\boldsymbol{\gamma}}$, one may consider introducing the latent variable $\boldsymbol{\gamma}$ for the risk premiums instead of for the coefficients in $\boldsymbol{b}$ like we do. However, this specification can lead to outcomes that are hard to interpret. If we arrange $\boldsymbol{f}$ as $\boldsymbol{f}^{\top} = [\boldsymbol{f}_{\boldsymbol{\gamma}}^{\top}, \boldsymbol{f}_{-\boldsymbol{\gamma}}^{\top}]$, then from equation (5),

$$\mathbb{E}[\boldsymbol{R}] = \boldsymbol{B}^{\top}\begin{bmatrix}\boldsymbol{\lambda}_{\boldsymbol{\gamma}} \\ \mathrm{Cov}[\boldsymbol{f}_{-\boldsymbol{\gamma}}, \boldsymbol{f}_{\boldsymbol{\gamma}}]\boldsymbol{b}_{\boldsymbol{\gamma}}\end{bmatrix},$$

where $\boldsymbol{B} = \{\mathrm{Var}[\boldsymbol{f}]\}^{-1}\mathrm{Cov}[\boldsymbol{f}, \boldsymbol{R}]$. As a result, if we let $\boldsymbol{\lambda}^{\top} = \left[\boldsymbol{\lambda}_{\boldsymbol{\gamma}}^{\top}, \boldsymbol{b}_{\boldsymbol{\gamma}}^{\top}\mathrm{Cov}[\boldsymbol{f}_{\boldsymbol{\gamma}}, \boldsymbol{f}_{-\boldsymbol{\gamma}}]\right]$, it is always the case that $\mathbb{E}[\boldsymbol{R}] = \boldsymbol{B}^{\top}\boldsymbol{\lambda}$ regardless of which model $\mathcal{M}_{\boldsymbol{\gamma}}$ is under consideration. That is, the full model including all factors always holds. Thus, we introduce the latent model index parameter for the coefficients in vector $\boldsymbol{b}$, which can help distinguish among different linear SDF models without ambiguity.

The third reason is due to parameter stability concerns. In equilibrium models, the vector $\boldsymbol{b}$ tends to concatenate deep structural parameters, while parameters such as factor loadings (the "beta"s) and factor risk premia are more likely to be driven by additional variables that could be time-varying. For example, under the setting of the CAPM, this coefficient equals the risk premium on the tangency portfolio over its variance. With a representative agent holding the market, this ratio is the risk-aversion parameter in the mean-variance utility. Thus, the b coefficient in this single factor model can be regarded as the (average) level of risk aversion. Another example looks at the consumption-based models with the Epstein-Zin preferences. According to the results in Epstein and Zin (1991), the linear SDF for this type of models can be approximated (using one plus the log SDF) as

$$m \approx \text{constant} + \frac{\gamma - 1}{\psi - 1}[\text{log consumption growth}] + \frac{1 - \psi\gamma}{\psi - 1}[\text{log return on wealth}],$$

where γ and ψ are the relative risk aversion and elasticity of intertemporal substitution parameters respectively. In this case, the two ratios, $(\gamma - 1)/(\psi - 1)$ and $(1 - \psi\gamma)/(\psi - 1)$, consist the vector $\boldsymbol{b}$, which is determined only by parameters in the preference, and is not changing across time.

I.2 Prior specification and empirical Bayes 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[8].

[8]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

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

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

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 relative to the data sample we have [9]. Under model $\mathcal{M}_\gamma$, excess returns observed in this sample are distributed as follows: $\boldsymbol{R}'_1, \ldots, \boldsymbol{R}'_{T'} \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{C}_\gamma \boldsymbol{b}_\gamma,\, \boldsymbol{\Sigma})$. Assigning a non-informative prior on $\boldsymbol{b}_\gamma$, which is flat everywhere[10], we can derive the "posterior" of $\boldsymbol{b}_\gamma$ given this conceptual data sample as $[\boldsymbol{b}_\gamma \mid \mathcal{M}_\gamma, \mathcal{D}'] \sim \mathcal{N}\left(\boldsymbol{b}'_\gamma,\, g/T \times \left(\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_\gamma\right)^{-1}\right)$, where the posterior mean $\boldsymbol{b}'_\gamma$ 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 $\boldsymbol{b}'_\gamma$ 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 $\boldsymbol{b}_\gamma$, 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 $\boldsymbol{b}'_\gamma$, the current posterior mean.

Zellner (1986) relies on the rational expectation hypothesis to pin down $\boldsymbol{b}'_\gamma$. Suppose that we have an anticipatory value for $\boldsymbol{b}_\gamma$, denoted by $\boldsymbol{b}^a_\gamma$, in addition to the imaginary sample

---

James-Stein estimator (James and Stein (1961)). For a monograph on modern empirical Bayes methods, see Efron (2012).

[9] 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).

[10] 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 $\boldsymbol{b}_\gamma$ 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 $\pi(\boldsymbol{b}_\gamma, \boldsymbol{\Sigma})$ is proportional to $\boldsymbol{\Sigma}^{-\frac{N+2}{2}}$. Some existing work (e.g. Barillas and Shanken (2018a)) specifies a prior such that $\pi(\boldsymbol{b}_\gamma, \boldsymbol{\Sigma}) \propto \boldsymbol{\Sigma}^{-\frac{N+1}{2}}$, which is the so-called independence Jeffreys prior (*not* the original Jeffreys-rule prior) imposing the assumption that $\boldsymbol{b}_\gamma$ and $\boldsymbol{\Sigma}$ are independent at the prior level.

$\mathcal{D}'$ (as well as the initial diffuse prior for $\boldsymbol{b}_{\boldsymbol{\gamma}}$). The rational expectation hypothesis says that $\boldsymbol{b}_{\boldsymbol{\gamma}}^{a} = \mathbb{E}[\boldsymbol{b}_{\boldsymbol{\gamma}} \mid \mathcal{M}_{\boldsymbol{\gamma}}, \mathcal{D}'] = \boldsymbol{b}'_{\boldsymbol{\gamma}}$. Now we have a reference informative prior distribution that does not depend on the hypothetical sample, which is

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

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 $\boldsymbol{b}_{\boldsymbol{\gamma}}^{a} = \mathbf{0}$, that is, the anticipatory expectations are the values under the null. This finally gives us the prior specification in (8).

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 *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*

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

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

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

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; $\mathrm{SR}^2_{\max}$ *is the maximal squared Sharpe ratio achievable from forming portfolios using all excess returns under consideration;* $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ *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* $\mathrm{SR}^2_{\boldsymbol{\gamma}} \leq \mathrm{SR}^2_{\max}$ *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 $\mathrm{SR}^2_{\boldsymbol{\gamma}}$ 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

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

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

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

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 $\mathrm{BF}_{\boldsymbol{\gamma},\boldsymbol{\gamma}'}(g)$ 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, Ross, and Shanken (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 $\mathrm{SR}_{\boldsymbol{\gamma}}$ increases as additional assets are added into the factor portfolio. Thus the penalty term on model dimensionality $p_{\boldsymbol{\gamma}}$ 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 $\mathrm{SR}_{\boldsymbol{\gamma}}$ 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 ($p_{\boldsymbol{\gamma}} - p_{\boldsymbol{\gamma}'} = 1$), 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,

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

However, it is always the case that $T\left(\mathrm{SR}^2_{\boldsymbol{\gamma}} - \mathrm{SR}^2_{\boldsymbol{\gamma}'}\right) \leq T\mathrm{SR}^2_{\max}$. 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), $\mathrm{BF}_{\boldsymbol{\gamma}}(g) \to 0$. 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.

I.3 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, Paulo, Molina, Clyde, and Berger (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, Paulo, Molina, Clyde, and Berger (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[11], 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 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 $a = 4 + 2/(TR_f^2 \text{SR}_\infty^2)$. Under this choice, the prior expectation for the shrinkage factor is still strictly smaller than one half, but can be very close (the ratio $2/(TR_f^2 \text{SR}_\infty^2)$ is usually very small).

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

I.4 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*

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

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

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

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

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

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:

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

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.

II Empirical Results

II.1 Data Description

In our main empirical implementation, we compute the entropy of posterior model probabilities (see equation (13)) using 14 tradable factors. We refrain the size of the factor zoo from being too large because of the computational difficulty in finding posterior model probabilities in high-dimensional cases. There is no golden rule in selecting candidate factors, but we enforce the following standards: (a) well-known factors such as Fama-French 3 factors should be considered with priority; (a) the anomalies should be robust in different periods[12]. Consequently, we select 14 factors in our analysis, and their definition is in the appendix A.

In addition, we obtain mutual fund data from the Center for Research in Security Prices (CRSP) survivorship-bias-free mutual fund database. More specifically, 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 normalize the aggregate fund flows, we divide total fund flows across all funds within a particular investment objective (e.g., large-cap funds) by the total market capitalization of all listed companies in CRSP. Therefore we download the total market value of all US-listed stocks from CRSP.

We also study the relationship between our model uncertainty measure and investors' sentiments, proxied by their expectations in several surveys. In our paper, 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.

We further compare our model uncertainty measure with other uncertainty measures proposed in the past literature. Bloom (2009) uses jumps in stock market volatility (VXO index) as the stock market uncertainty shock. We download the time-series of VXO and VIX indices from Wharton Research Data Services (WRDS). Baker, Bloom, and Davis (2016) develop indices of economic policy uncertainty (EPU) based on news coverage of keywords (e.g., regulation, white house) in 10 large newspapers, tax code expiration data, and the survey data of professional forecasters from the Federal Reserve Bank of Philadelphia. We

[12]As McLean and Pontiff (2016) points out, the out-of-sample performance of many investment anomalies decline after publication.

download the time-series of EPU measures from Nick Bloom's website. Finally, we obtain the macro, real and financial uncertainty index in Ludvigson, Ma, and Ng (2015) and Jurado, Ludvigson, and Ng (2015) from the Macro Finance Society database in WRDS.

II.2 Countercyclical Model Uncertainty

The first empirical analysis is to construct the time series of model uncertainty. At the end of the month t, we use all daily factor returns in the past 3 years to estimate the posterior model probability, $p[\mathcal{M}_{\boldsymbol{\gamma}} \mid \mathcal{D}]$, and compute the entropy as in equation (13). In the benchmark case, we choose the hyper-parameter a to be 4. We also present the results obtained from alternative rolling windows and other choices of a in robustness check (see section III). The four factors in Hou, Xue, and Zhang (2015) are available only from January 1967, and we use 36-month data in the estimation, so the model uncertainty measure starts from December 1969. Since some factors are highly correlated, we only consider models that contain at most one version of the factors 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 essentially 10 effective factors, and they are market, size, profitability, value, investment, short-term reversal, long-term reversal, momentum, QMJ, and BAB.

The blue line in figure 1 plots the time series of model uncertainty of linear SDFs consisting of 14 tradable factors described in section II.1, and the sample period spans from December 1969 to December 2018. The red and green dotted lines show the lower and upper bounds of the model uncertainty, respectively. The lower entropy bound is always 0, for example, when there is one dominant model with a posterior model probability of 100%. On the contrary, uncertainty is maximized when the posterior model probabilities are equalized across different models. Because we have 14 factors and only one of the categorical factors could be selected into the true model, there are 5184 different candidate models.[13] The upper bound of model uncertainty is around 8.55.[14]

The key observation in figure 1 is that model uncertainty varies significantly over time. More importantly, the model uncertainty is countercyclical: it is low in normal times but increases rapidly just before stock market crashes and remains high during bear markets. We will describe the time-series behaviors of model uncertainty in detail.

[13] 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 $2^6 \times 3^4 = 5184$.

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

At the beginning of the 1970s, the model uncertainty is high. There were two famous stock market crashes: November 1968 to May 1970 and Jan 1973 to Oct 1974[15]. High model uncertainty also coincides with NBER-based Recession periods: according to NBER, "January 1970 to November 1970" and "December 1973 to March 1975" were two recession periods. After 1975, the model uncertainty declines gradually until the beginning of 1980.

The second peak of model uncertainty appears between 1980 and 1982[16]. Specifically, the model uncertainty starts to increase at the end of 1979 and keeps climbing up to the peak at 1982. After a decade of high inflation in the 1970s, the Federal Reserve (Fed) decided to raise the nominal interest rate to nearly 20%. Accompanied by the unprecedented high interest rate, the US economy entered recessions from February 1980 to July 1980 and from August 1981 to November 1982.

Surprisingly, uncertainty is not exceptionally high during the flash crash in November 1987. The potential reason is that the 1987 market crash 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 recovered rapidly[17]. Instead, the leading cause was synchronous program trading, illiquidity in the market, and the subsequent market panic. Since our uncertainty measure makes use of past-3-year daily data in its construction[18], the impact of short-term market chaos is averaged out and does not particularly affect the Bayesian inference in the cross-section.

The US stock market entered another bear market between 1989 and 1991. The failed leveraged buyout of United Airlines led to the first market crash. The subsequent increase in oil price due to the Iraqi invasion of Kuwait in July 1990 caused the Dow to drop by 18%. The US economy also entered a recession from August 1990 to March 1991. The 1990s was a special 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, so investors at that time were confident about the true model.

Slightly before the dot-com crash in 2000, the model uncertainty begins to surge, from less than 2 in December 1999 to around 8 in March 2000. According to NBER, the economy entered a recession from April 2001 to November 2001. During this bubble, the largest drop

[15]The largest drop in the S&P 500 index was 36.1% between November 1968 to May 1970, and the bear market lasted 18 months. Similarly, another stock market crash took place between Jan 1973 and Oct 1974, during which S&P 500 index plummeted from the highest 119.87 to the lowest 62.28.

[16]At the same time, The S&P 500 index dropped from 140.52 to the lowest, 101.44, so market investors lost 27.8% of their financial wealth in the worst case.

[17]A new bull market started in December 1987.

[18]To be more specific, we use past-36-month daily data to compute the mean returns and variance matrix of 14 factors. If the stock market recovers rapidly, the average performance of factor strategies could remain roughly the same as in normal times.

in S&P 500 was around 50%, and many high-tech listed companies went into bankruptcy.

During our sample period, model uncertainty is the highest in the 2008 global financial crisis. Investors and households experienced the most catastrophic stock market crash and economic recession since the Great Depression between 1929 and 1933. The model uncertainty measure almost touches its upper bound. This period is also unique in the sense that model uncertainty stays at a high level for a long time, while it tends to decline soon after the market crash in other crises. The model uncertainty drops considerably at the beginning of 2011. However, it suddenly goes up to a high level, possibly because of the downgrading of US sovereign debts in August 2011 when the market index plunged by more than 20% in a few months.

Finally, model uncertainty remains very high after the 2008 financial crisis, even though it drops slightly between 2011 and 2014. In recent years (after 2015), the economic and stock market prospects have been uncertain. For example, the market index, be it the Dow or S&P 500 index, experienced a 20% plummet at the end of 2015 and 2018. Also, the trade war between the US and China started in 2018 has contributed to additional uncertainty.

In conclusion, our model uncertainty measure exhibits a countercyclical pattern: it is particularly high during stock market crashes and economic recessions. The stock market crisis that only lasts for a short period, such as the flash crash in 1987, is not reflected in our model uncertainty measure. Furthermore, the cyclical behavior of model uncertainty implies another layer of investment risk: when investors experience bear stock markets, they are also more uncertain about the true model in the cross-section, or equivalently, which "alpha" strategies still work and should be adopted. Therefore, a natural hypothesis is that model uncertainty affects investors' portfolio choice and their market expectations. We investigate these topics in section II.5 and II.6.

II.3 Decomposing Model Uncertainty

The posterior model probabilities (see Proposition 3) are closely related to maximum in-sample Sharpe ratio squared achieved by investing in a set of factors, SR^2_{γ}.[19] As more factors are included, in-sample SR^2_{γ} always rises up. Only when a few factor models obviously dominate others can we be confident about the true model. In other words, when the distances in SR^2_{γ} are huge across different factor model $\boldsymbol{\gamma}$, we can easily differentiate them and end up with low model uncertainty. In contrast, when factor models have similar SR^2_{γ}, the model uncertainty will be high.

Figure 2 plots the time-series of distances in SR^2_{γ} during our sample period. More precisely, we show the difference between the maximal SR^2_{γ} and the $90th$-quantile of SR^2_{γ}, as

[19] SR^2_{γ} equals $\mathbb{E}_T[\boldsymbol{f}_{\gamma}]^{\intercal}\boldsymbol{V}_{\gamma}^{-1}\mathbb{E}_T[\boldsymbol{f}_{\gamma}]$.

well as the difference between the maximal SR^2_γ and medium SR^2_γ. Strikingly, the difference in SR^2_γ decreases obviously 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 around 1997 and 1998, but it plunges to almost 0 from 1998 to 2000. After the tech bubble in the early 2000s, factor models are becoming more similar with each other in terms of in-sample SR^2_γ.

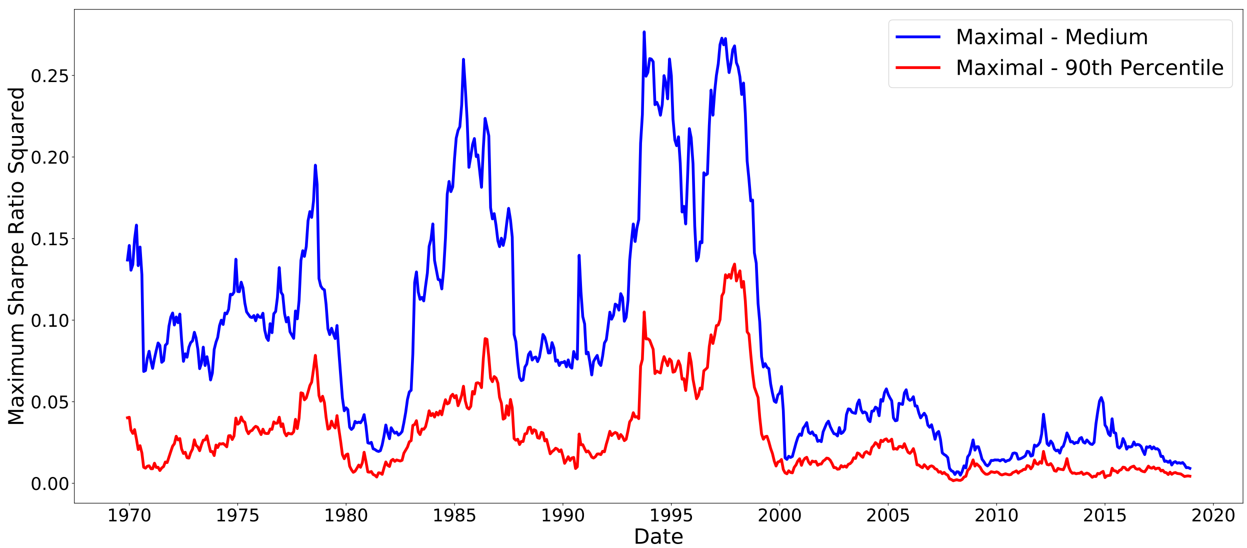


Figure 2: Time-Series of Maximum Sharpe Ratio Squared SR^2_γ (3-Year Rolling Window)

The figure plots the time series of distances in SR^2_γ over our sample period. 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 maximum in-sample Sharpe ratio squared achieved by investing in a set of factors, and it is equal to $\mathbb{E}_T[\boldsymbol{f}_\gamma]^\intercal \boldsymbol{V}_\gamma^{-1}\mathbb{E}_T[\boldsymbol{f}_\gamma]$. $\mathbb{E}_T[\boldsymbol{f}_\gamma]$ and $\boldsymbol{V}_\gamma$ are estimated using the daily factor returns in the past 36 months.

Theoretically, SR^2_γ of model $\boldsymbol{\gamma}$ is determined by mean returns of factors and covariance structure among factors. We further analyze the maximum in-sample Sharpe ratio squared SR^2_γ by dipping into three parts: (a) average daily returns of factors 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 3 plots these time series.

In figure 3a, we show that the average daily return of all 14 factors[20] is very volatile, but surprisingly, it is always positive. The average factor return also exhibits a cyclical pattern: it declines during the run-ups of stock markets and plummets to the bottom in the market

[20] At the end of each month t, we use daily factor returns from month $t-35$ to month t to compute the average daily returns.

crash, but it recovers gradually after the bear markets. In the recent two most influential market crashes (dot-com bubble and 2008 global financial crisis), the average factor returns decline to nearly 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 3b plots the average volatility of 14 factors over time. Even though the average return dispersion 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 long-short portfolios 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 3b that model uncertainty does not have the same time-series pattern as the average volatility of factors.

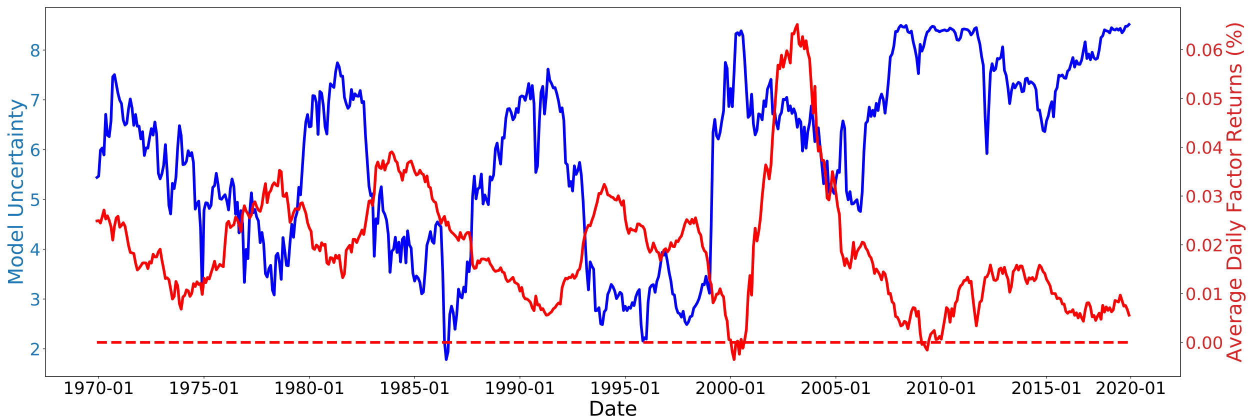
During market crashes, it is highly likely that arbitrageurs who invest in these "alpha" strategies will exit the market simultaneously, which causes high comovement among factors. Since the correlation matrix of factors determines the extent to which investors can diversify their investment, it could potentially influence the distances in factor models' profitability, SR^2_γ. To illustrate this point, we plot the time-series of average pairwise correlation[21] of 14 factors. The average correlation exhibits a similar cyclical pattern as model uncertainty. However, there are two key differences: (a) the average correlation is decreasing before the 2008 crisis while our model uncertainty starts to climb up at the end of 2006; (b) The model uncertainty is the highest during the 2008 crisis and remains at a high level after then, but the average correlation is at the peak during the dot-com bubble.

To sum up, model uncertainty is high when the distances in SR^2_γ among different factor models are low. Since in-sample SR^2_γ always increases with more factors included, we are uncertain about whether to include an additional factor if the benefit of including it is only marginal. Furthermore, the model uncertainty about linear SDFs increases dramatically during the run-ups and stands at the peak during bear markets because different factor models are extremely similar, so the posterior marginal probability of each factor, i.e., $p[\gamma_j \mid \mathcal{D}]$, is close to 50%.

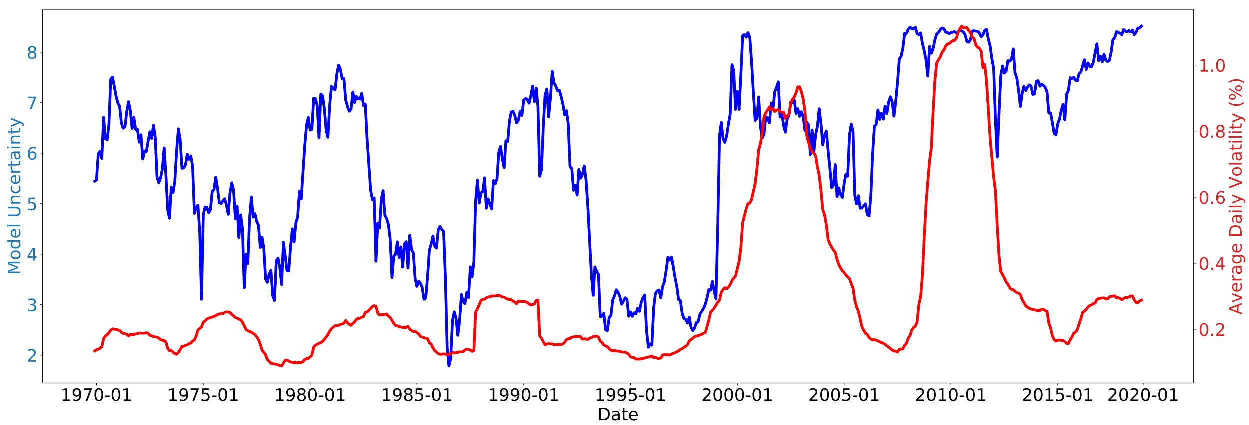
II.4 Comparison with Other Uncertainty Measures

There are several famous uncertainty measures proposed in the literature. In this section, we compare our model uncertainty measure with them. The first measure is VXO (used in

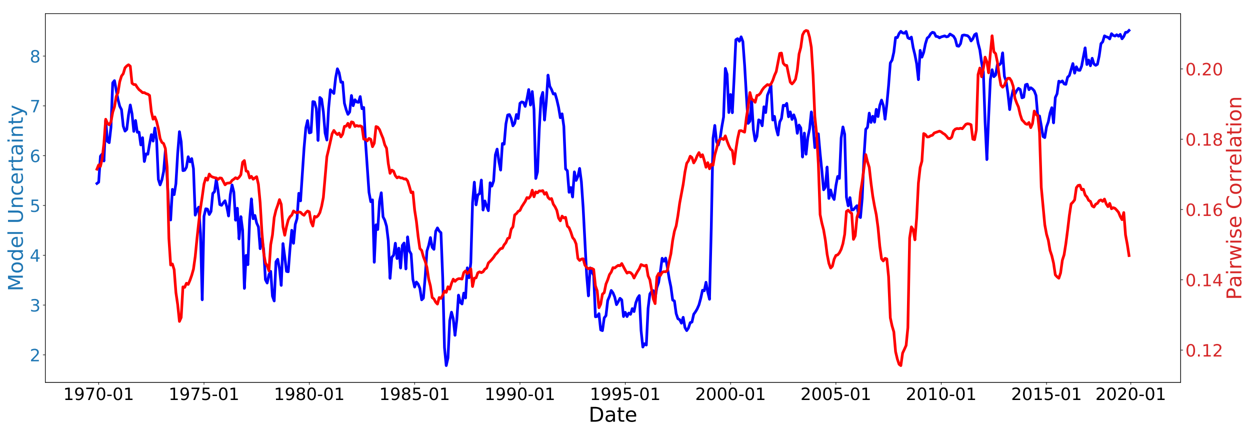
[21] At the end of each month t, we use daily factor returns from month $t-35$ to month t to compute the pairwise correlation between any two factors, denoted as ρ_{ij}. The average is computed as $\frac{1}{N\times(N-1)}\sum_{i\neq j}|\rho_{ij}|$.



(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 3: 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.

Bloom (2009)) or VIX[22] index. It quantifies forward-looking market volatility and mainly describes the dispersion of future market returns, e.g., how likely the market will increase or decrease by a certain amount. Subsequent to Bloom (2009), Ludvigson, Ma, and Ng (2015) and Jurado, Ludvigson, and Ng (2015) develop the real, macro and financial uncertainty measures by exploiting a large set of macro and financial factors. 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[23]. Baker, Bloom, and Davis (2016) use the coverage of economic or policy-related keywords in the media as a proxy for economic policy uncertainty. The details of the construction in the uncertainty mentioned above measures could be found in their papers.

Our goal is to test whether other uncertainty sequences can explain the time-variation in our model uncertainty measure. However, our model uncertainty is intrinsically persistent since we use daily data of a 3-year rolling window to compute the entropy at the end of each month. When we run the first-order autoregression (AR(1)): $entropy_t = \rho \times entropy_{t-1} + \epsilon_t$, the OLS estimate of ρ is 0.98 (see table (6)). Therefore, $entropy_t$ is a persistent time series, and we need to be careful in statistical inference. Since our monthly model uncertainty measure is constructed in the 3-year rolling window, we use Newey-West standard errors (see Newey and West (1987)) with 36 lags in all monthly regressions.

Strong persistence of the time-series process is ubiquitous in other uncertainty measures. Table (6) 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 0.83. Only the second economic policy uncertainty measure (EPU_2) suffers less from massive autocorrelations.

In table 1, we regress model uncertainty $entropy_t$ on other uncertainty measures. Since both dependent and independent variables have persistent trends, we are worried about the spurious regressions. To relieve our concerns, we control the one-month lagged value of model uncertainty ($entropy_{t-1}$) in all regressions in addition to using Newey-West standard errors. After we control one-period lagged entropy, the coefficient estimates of all six uncertainty measures are insignificant. Therefore, they could not explain our model uncertainty measures, and more importantly, it indicates that our measure conveys different and additional information beyond the other six measures.

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

[23]Suppose there is a set of economic indicators, $\boldsymbol{Y_t} = (y_{1t}, ..., y_{Lt})^\intercal$. For each variable, they find the 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.

Table 1: Correlation with Other Uncertainty Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.0103	-0.0083	-0.0017	0.1207	0.1081*	0.0238
	(0.12)	(-0.08)	(-0.01)	(1.64)	(1.76)	(0.37)
$Entropy_t$	0.9792***	0.9776***	0.9797***	0.9880***	0.9876***	0.9856***
	(137.63)	(129.45)	(143.44)	(148.85)	(145.43)	(144.20)
$Financial_{t+1}$	0.1213					
	(1.14)					
$Macro_{t+1}$		0.2068				
		(1.19)				
$Real_{t+1}$			0.1871			
			(0.62)			
$EPU_{1,t+1}$				-0.0004		
				(-0.67)		
$EPU_{2,t+1}$					-0.0002	
					(-0.55)	
VXO_{t+1}						0.0034
						(1.48)
N	588	588	588	408	408	396

t statistics in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The table reports empirical results in the following regression: $Entropy_{t+1} = \beta_0 + \beta_1 Entropy_t + \rho Uncertainty_{t+1} + \epsilon_{t+1}$. The uncertainty measures $Uncertainty_{t+1}$ include financial, macro and real uncertainty measures from Ludvigson, Ma, and Ng (2015) and Jurado, Ludvigson, and Ng (2015), two economic policy uncertainty (EPU) from sequences Baker, Bloom, and Davis (2016), as well as VXO. 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 N is the number of observations.

Comments. Conceptually, our model uncertainty measure quantifies different layers of uncertainty from others. Stock market volatility, proxied by the VXO index, measures the second-moment investment risk. Three uncertainty measures in Ludvigson, Ma, and Ng (2015) and Jurado, Ludvigson, and Ng (2015) are essentially volatility of prediction errors. In other words, they express the dispersion of unexpected change in related economic indicators. Two economic policy uncertainty sequences in Baker, Bloom, and Davis (2016) are essentially quantifying public attention to economic policy. In contrast, our paper quantifies the model uncertainty of linear SDFs or equivalently the uncertainty of mean-variance efficient portfolio. Since we know the lower and upper bound of entropy, we can easily detect the degree of model uncertainty in the cross-section. For example, during some special periods, the model uncertainty could even reach its upper bound, implying that different models' posterior probabilities are almost equal. In short, the uncertainty measures developed in the past cannot replace our model uncertainty. In other words, ours provides a new angle of analyzing investment uncertainty.

II.5 Mutual Fund Flows

If investors consider the model uncertainty an important source of investment risk, a natural prediction is that their portfolio choice should be related to our model uncertainty measure. The difficulty in empirical tests arises due to the lack of observation in their complete portfolio choice. However, it is well known that the fraction of institutional holdings in the US stock market has been increasing in the past 20 years, and the mutual fund sector is one of the most important financial intermediaries through which individual investors participate in stock markets. This section focuses on the mutual fund flows and uses them as proxies for investors' portfolio rebalancing.

The data is available on CRSP survivor-bias-free US mutual fund database. The database includes investment style or objective codes from three different sources over the whole life of the database. From 1962 to 1993, Wiesenberger objective codes are used. Strategic insight objective codes are populated between 1993 and 1998. Lipper objective codes start from 1998. Instead of using the three measures mentioned above directly, CRSP builds its objective codes based on them. The CRSP style code consists of up to four letters. For example, a fund with style "EDYG" means that this fund mainly invests in domestic equity (E = Equity, D = Domestic) markets, and it has a specific investment style "Growth" (Y = Style, G = Growth).[24] The quality of data before 1990 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 1990. 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 after 1990.

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

$$Flow_{i,t} = TNA_{i,t} - TNA_{i,t-1} \times (1 + RET_{i,t}) \tag{14}$$

where $TNA_{i,t}$ and $RET_{i,t}$ are total net assets and gross returns of fund i in period t. Then we aggregate the individual fund flows in each period across all funds in a group (e.g. all large-cap funds) and scale it by total market capitalization of all stocks in CRSP.

Since our model uncertainty measure is constructed from factors in the US, we delete all foreign funds. We further split the sample into domestic equity and fixed-income funds. More specifically, we consider the following empirical settings: (a) all domestic equity mutual funds; (b) all domestic fixed-income funds; (c) domestic "style", "sector", "large-cap" and "middle and small-cap" equity funds; (d) finally, domestic government bond funds. The

[24] More details could be found in the handbook of CRSP survivor-bias-free US mutual fund database.

time-series regression that we run is

$$Flow_{g,t+1} = \beta_{g,0} + \beta_{g,1} Flow_{g,t} + \beta_{g,2} Flow_{g,t-1} + \beta_{g,3} Flow_{g,t-2} + \rho Entropy_t + \gamma X_t + \epsilon_{t+1} \quad (15)$$

where $Flow_{g,t}$ is the aggregate mutual fund flows in group g (e.g. large-cap funds) in period t, normalized by total market value of all CRSP stocks; $Entropy_t$ is the model uncertainty measure in period t; and X_t includes other control variables.

In the first regression, we consider all domestic equity funds. The dependent variable is the aggregate fund flows in all domestic equity funds in each month. Since fund flows are persistent, we control three lagged fund flows. Column (1) in the table (2) shows that model uncertainty negatively predicts the one-month ahead aggregate domestic fund flows, and the coefficient estimate is significant in both statistical and economic sense. The p-value constructed by 36-lag Newey-West standard errors is less than 1%. One unit increase in entropy predicts that the aggregate domestic fund flows decreases by -0.0119%. Since we normalize the fund flows by the aggregate market value of all CRSP listed firms, this percentage is economically substantial. Although we cannot interpret the regression results as causal, we still find that investors in domestic equity mutual funds tend to decrease their holdings when the model uncertainty increases. We further control lagged VIX index and lagged aggregate return on the domestic equity funds, but they do not influence the relationship between domestic equity fund flows and model uncertainty.

The last two columns in the table (2) show that model uncertainty cannot predict future aggregate fund flows in domestic fixed-income funds. It is expected since the underlying risk factors in aggregate fixed-income markets are quite different from the stock market.

We further group the domestic equity funds based on their CRSP investment objective codes. The "style" funds refer to the growth, income, growth & income and "hedged" funds. They are more likely to rely on "alpha" strategies similar to factors used in constructing model uncertainty than other funds. Therefore, the fund outflows of "style" equity funds should be particularly severe when the model uncertainty is high. Column (1) in table 3 proves our hypothesis. The coefficient estimate of model uncertainty is -0.008% and is statistically significant at the significance level of 1%.

In contrast, columns (2) and (3) in the table 3 conclude that aggregate fund flows in sector funds[25] and large-cap funds are not sensitive to the change in model uncertainty - the coefficient estimate of model uncertainty in the regression of large-cap funds is almost zero.

We also test whether middle and small-cap funds experience outflows with higher model uncertainty. Column (4) in the table (3) gives us a positive answer: middle and small-cap

[25] In CRSP, sector funds refer to those funds that invest in a particular industry.

Table 2: Regression of Mutual Fund Flows on Model Uncertainty: All Domestic Equity and Fixed-Income Funds

	Domestic Equity			Domestic Fixed-Income	
	(1)	(2)	(3)	(4)	(5)
Intercept	0.1063***	0.1206***	0.1092***	0.0273	-0.0245
	(6.06)	(4.70)	(4.67)	(0.82)	(-0.95)
$Entropy_t$	-0.0119***	-0.0121***	-0.0118***	0.0017	-0.0002
	(-5.39)	(-5.55)	(-5.48)	(0.29)	(-0.03)
VIX_t		-0.0006	-0.0002		0.0033**
		(-1.11)	(-0.37)		(2.19)
RET_t			0.2146		
			(1.33)		
$Flows_t$	0.2294***	0.2153***	0.1778***	0.1686***	0.1592***
	(3.85)	(3.38)	(3.27)	(4.20)	(4.01)
$Flows_{t-1}$	0.0325	0.0257	0.0544	0.1265**	0.1185**
	(0.35)	(0.27)	(0.62)	(2.36)	(2.21)
$Flows_{t-2}$	0.1306***	0.1260***	0.1404***	0.2164***	0.2104***
	(2.86)	(2.74)	(2.94)	(4.45)	(4.12)
N	348	348	348	348	348
R^2	0.2775	0.2797	0.2859	0.1316	0.1386

t statistics in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The table reports empirical results in regression: $Flow_{g,t+1} = \beta_{g,0} + \beta_{g,1}Flow_{g,t} + \beta_{g,2}Flow_{g,t-1} + \beta_{g,3}Flow_{g,t-2} + \rho Entropy_t + \gamma X_t + \epsilon_{t+1}$, where $Flow_{g,t}$ is the aggregate mutual fund flows normalized by total market value of all CRSP stocks in group g (e.g. large-cap funds) in period t, $Entropy_t$ is the model uncertainty measure in period t, and X_t includes other control variables. We use aggregate fund flows in domestic equity (column (1) to (3)) and fixed-income (column (4) and (5)) mutual funds in the regressions. 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 N is the number of observations.

funds tend to experience fund outflows over the next period when model uncertainty is high in the present. Let us consider the investment in middle and small-cap funds as investors implementing size factor SMB or ME. The results in the table (3) indicate that model uncertainty in the cross-section absolutely influences investors' holdings in small firms.

Finally and interestingly, column (5) in the table (3) displays that model uncertainty positively predicts the aggregate fund flows in US government bonds. US government bonds are well-known for their superior safety over all other asset classes. Combined with the findings mentioned above in equity mutual funds, we can interpret this empirical result as the fact that investors tend to allocate more of their wealth into safe assets when the model uncertainty in the stock market is more pronounced.

Table 3: Regression of Mutual Fund Flows on Model Uncertainty: Different Investment Objective Codes

	Style	Sector	Large-Cap	Middle and Small Cap	Government Bonds
	(1)	(2)	(3)	(4)	(5)
Intercept	0.0633***	0.0016	0.0090	0.0045	-0.0055**
	(4.02)	(0.73)	(0.68)	(0.91)	(-2.39)
$Entropy_t$	-0.0080***	-0.0001	0.0000	-0.0015*	0.0009*
	(-4.97)	(-0.59)	(0.00)	(-1.88)	(1.96)
VIX_t	-0.0001	0.0001	0.0000	0.0004***	0.0001
	(-0.27)	(0.87)	(0.01)	(2.63)	(0.79)
$Flows_t$	0.2253***	0.1815	0.1672*	0.3580***	0.3105***
	(4.28)	(1.50)	(1.68)	(5.11)	(4.84)
$Flows_{t-1}$	0.1069	0.2360***	-0.1248	0.1139**	0.1570**
	(1.59)	(3.17)	(-1.46).	(2.14)	(1.98)
$Flows_{t-2}$	0.1809***	0.0998**	0.0795	0.1077**	0.1826***
	(3.81)	(2.42).	(1.26)	(2.34)	(2.84)
RET_t	0.1106	0.0285**	-0.0874	0.0879***	
	(1.09)	(2.20)	(-1.21)	(2.81)	
N	348	345	249	335	345
R^2	0.4742	0.1584	0.0433	0.3499	0.3281

t statistics in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The table reports empirical results in regression: $Flow_{g,t+1} = \beta_{g,0} + \beta_{g,1}Flow_{g,t} + \beta_{g,2}Flow_{g,t-1} + \beta_{g,3}Flow_{g,t-2} + \rho Entropy_t + \gamma X_t + \epsilon_{t+1}$, where $Flow_{g,t}$ is the aggregate mutual fund flows normalized by total market value of all CRSP stocks in group g (e.g. large-cap funds) in period t, $Entropy_t$ is the model uncertainty measure in period t, and X_t includes other control variables. We use aggregate fund flows in "style", "sector", "large-cap" and "middle and small-cap" equity funds (see column (1) to (4)) and domestic government bond funds (see column (5)) in the regressions. 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 N is the number of observations.

II.6 Investors' Sentiments

We further investigate whether our model uncertainty measure correlates with investors' sentiments, that is, their expectations on the stock markets. The first measure of sentiment is from the sentiment 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. Since our model uncertainty measure is monthly, we use the sentiment index in the last week of each month as a proxy for investors' sentiment in that month.

We also consider Shiller's stock market confidence indices. The survey was conducted by the International Center for Finance at the University of Yale. In our research, 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%. Roughly speaking, the higher the indices are, the more confident individual or institutional investors are.

The time-series regression is

$$Sentiment_{t+1} = \beta_0 + \gamma Entropy_t + \psi X_t + \epsilon_{t+1} \tag{16}$$

where $Sentiment_{t+1}$ is the one-period ahead sentiment 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 sentiment indices, VIX and etc. Since all sentiment indices are autocorrelated, we control their one and two-period lags in all regressions[26]. 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 4, we regress AAII sentiment indices on model uncertainty in order to shed light on how individual investors change their attitudes towards the stock market in response to variation in model uncertainty. In column (1), $entropy_t$ cannot predict the one-period ahead percentage of bullish investors even without further controls. On average, bullish investors remain optimistic about the market even when the model uncertainty in the cross-section goes up. Columns (2) and (3) regress the percentage of neutral investors on lagged model uncertainty. Surprisingly, the percentage of being neutral declines if entropy rises. The next question is, towards which direction do neutral investors change their attitude? Columns (4) and (5) indicate that investors are more likely to be bearish following an increase in model uncertainty. Our interpretation is that a higher level of model uncertainty causes some neutral investors to become bearish. Finally, we regress the difference between fractions of bullish and bearish investors on the entropy. The coefficient estimate of entropy is negative and statistically significant at the significance level of 10%. Overall, when the model uncertainty goes up, people tend to be more pessimistic about the stock market performance in the future.

Table 5 regresses Shiller's confidence indices on entropy. Different from the AAII sentiment index, we also have the expectations of institutional investors. The results are generally similar to table 4; investors tend to be more pessimistic about the stock market when the model uncertainty goes up. They also believe that a market crash is more likely to occur with higher model uncertainty. One interesting empirical fact is: the coefficient estimates

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

Table 4: AAII Sentiment Index

	Bullish	Neutral		Bearish		Bullish - Bearish	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Entropy_t$	-0.0020	-0.0037**	-0.0034**	0.0058***	0.0056***	-0.0072*	-0.0068*
	(-0.57)	(-2.10)	(-2.05)	(2.73)	(2.65)	(-1.87)	(-1.76)
VIX_t		0.0004		-0.0001		-0.0001	
		(0.49)		(-0.13)		(-0.11)	
$Sentiment_t$		0.4782***	0.4731***	0.3522***	0.3302***	0.3619***	0.3201***
		(7.84)	(8.05)	(8.51)	(6.82)	(7.05)	(5.59)
$Sentiment_{t-1}$		0.2144***	0.2154***	0.1859***	0.1973***	0.1178***	0.1504***
		(5.17)	(5.28)	(5.62)	(5.16)	(2.94)	(3.19)
RET_t			-0.1978		-0.1253		0.4427*
			(-1.64)		(-1.03)		(1.79)
RET_{t-1}			0.1536		0.0161		-0.0969
			(1.38)		(0.22)		(-0.47)
RET_{t-2}			-0.0616		0.0476		-0.0050
			(-0.91)		(0.40)		(-0.02)
RET_{t-3}			0.0809		-0.0767		0.0979
			(1.41)		(-0.74)		(0.53)
RET_{t-4}			0.0609		0.0112		-0.0806
			(0.70)		(0.15)		(-0.54)
RET_{t-5}			0.1373*		0.0111		-0.1637
			(1.89)		(0.12)		(-0.77)
N	379	377	377	377	377	377	377
r2	0.0018	0.4402	0.4618	0.2749	0.2793	0.2052	0.2172

t statistics in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The table reports empirical results in regression: $Sentiment_{t+1} = \beta_0 + \gamma Entropy_t + \psi X_t + \epsilon_{t+1}$, where $Sentiment_{t+1}$ is the one-period ahead sentiment 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 sentiment indices, VIX and etc. Since all sentiment indices are autocorrelated, we control their one and two-period lags in all regressions. Since investors' expectations on the market are extrapolative, we further control lagged market returns (S&P 500 index) in the regression. Sentiment indices come from the sentiment survey conducted by 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 of percentages of respondents who are bearish, neutral or bullish in each week. Since our model uncertainty measure is monthly, we use the sentiment index in the last week of each month as a proxy for investors' sentiment in that month. 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 N is the number of observations.

of $entropy_t$ in the regression of individual investors' sentiment indices are always more negative than institutional investors. Since individuals are more likely to be sentiment-driven investors, they seem to react more dramatically to the change in model uncertainty.

In short, this section shows that higher model uncertainty generally predicts that investors in the survey, be it individual or institutional, will become more pessimistic about the future stock market returns.

Table 5: Shiller's Confidence Index

	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)
$Entropy_t$	-0.3836***	-0.2971***	-0.4522***	-0.3169***	-0.6762***	-0.7350***	-0.8117***	-0.8720***
	(-3.17)	(-2.61)	(-2.87)	(-2.76)	(-4.42)	(-5.20)	(-4.22)	(-4.94)
VIX_t	0.0313**		0.0291		-0.0978***		-0.0621***	
	(2.34)		(1.21)		(-3.30)		(-3.24)	
$Sentiment_t$	1.1528***	1.1636***	0.9788***	0.9891***	1.0744***	1.0846***	1.0178***	0.9409***
	(16.87)	(17.51)	(13.19)	(12.10)	(15.89)	(18.56)	(14.44)	(16.18)
$Sentiment_{t-1}$	-0.3004***	-0.2960***	-0.0530	-0.0480	-0.2567***	-0.2384***	-0.2006***	-0.1150***
	(-3.81)	(-3.62)	(-0.70)	(-0.60)	(-3.90)	(-3.73)	(-3.35)	(-2.94)
RET_t		-3.4684		1.7876		14.2780***		8.7997***
		(-1.29)		(0.58)		(4.23)		(4.48)
RET_{t-1}		6.7243**		2.7331		3.6080		10.9416***
		(2.17)		(0.82)		(0.91)		(4.09)
RET_{t-2}		0.5931		-0.7076		7.1782		6.6587***
		(0.17)		(-0.25)		(1.59)		(5.69)
RET_{t-3}		-1.0368		-4.4327**		9.6582		4.5690
		(-0.49)		(-2.27)		(1.59)		(1.09)
RET_{t-4}		-4.6652*		5.9696		2.3029		4.6544*
		(-1.95)		(1.33)		(0.80)		(1.79)
RET_{t-5}		0.4433		-4.1218*		-2.4099		1.8127
		(0.13)		(-1.81)		(-0.74)		(0.70)
N	208	208	208	208	208	208	208	208
r2	0.8367	0.8389	0.9372	0.9383	0.8989	0.9017	0.9214	0.9278

t statistics in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The table reports empirical results in regression: $Sentiment_{t+1} = \beta_0 + \gamma Entropy_t + \psi X_t + \epsilon_{t+1}$, where $Sentiment_{t+1}$ is the one-period ahead sentiment 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 sentiment indices, VIX and etc. Since all sentiment indices are autocorrelated, we control their one and two-period lags in all regressions. Since investors' expectations on the market are extrapolative, we further control lagged market returns (S&P 500 index) in the regression. Sentiment 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 individual or institutional investors expecting an increase in the Dow in a year, while 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, respectively and N is the number of observations.

II.7 More Evidence on Model Uncertainty

Suppose our model uncertainty measure can capture cyclical uncertainty in linear SDFs or mean-variance efficient portfolios. In that case, we should be able to observe similar countercyclical patterns in international stock markets. In this section, we present the time series of model uncertainty in European and Asian Pacific stock markets. Instead of using all 14 factors as in the US stock market, we include only nine of them because of the limited availability of factor data. We ignore the size (ME), profitability (ROE), and investment (IA) in Hou, Xue, and Zhang (2015) because related factors are not available in international markets, and we believe that Fama-French 5 factors are essentially representing the same kinds of systematic risks as them. Short-term and long-term reversals are also excluded, so we end up with 9 candidates: market, SMB, HML, RMW, CMA, momentum, QMJ, BAB, and HML devil. Either HML or HML devil could enter the true SDF. Since the AQR library provides QMJ factor only from July 1993 and we use a 3-year rolling window, our measure of model uncertainty starts from June 1996.

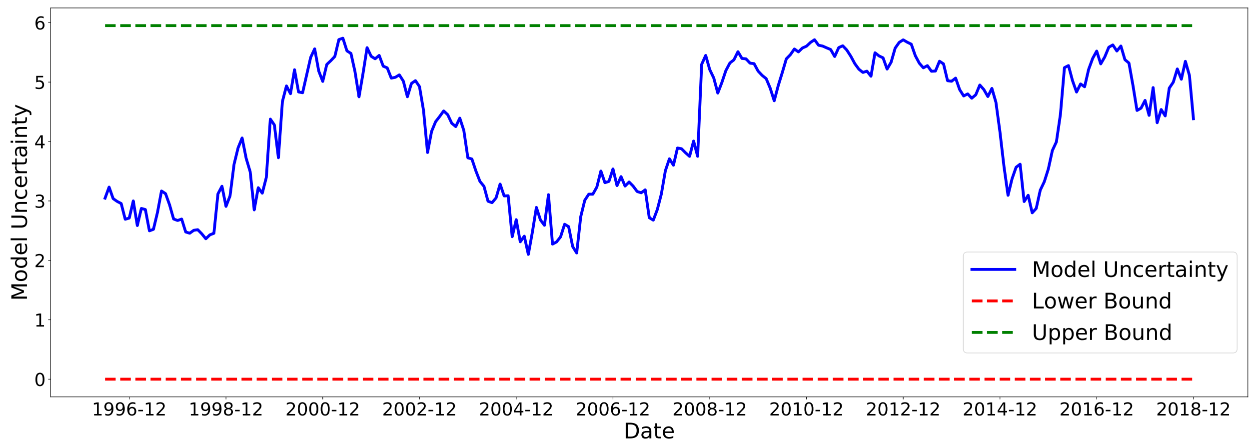
Figure 4a plots the time-series of model uncertainty in the European stock market from June 1996 to December 2018. The red and green lines show the lower and upper bound[27]. The time-series variation in the European market[28] is very similar to the US stock market. The model uncertainty starts to go up from 1999 and reaches its first peak between 2000 and 2001 because of the dot-com bubble burst. The model uncertainty almost touches its upper 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. From 2015, the uncertainty index shoots up again, possibly due to investors' lack of confidence in the stock market.

The findings in Asian Pacific market[29] are more interesting than the US and European markets since we observe some unique time-series variation in the Asian stock market. According to figure 4b, the model uncertainty is high starting from 1997 due to the profound 1997 Asian financial crisis. Asian stock markets were over-heated, and market crashes appeared in almost every Asian country. 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 afterward. 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. We argue that this observation is intriguing: the 1997 Asian financial crisis combined with the burst of the dot-com bubble in 2000 should be more destructive from the perspectives of Asian investors 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, the model uncertainty surges from the beginning of 2015. In the second half of 2015, market investors in Hong Kong and Singapore lost a huge amount of wealth because of markets crashes, mainly caused by the drop in US stock market and Chinese

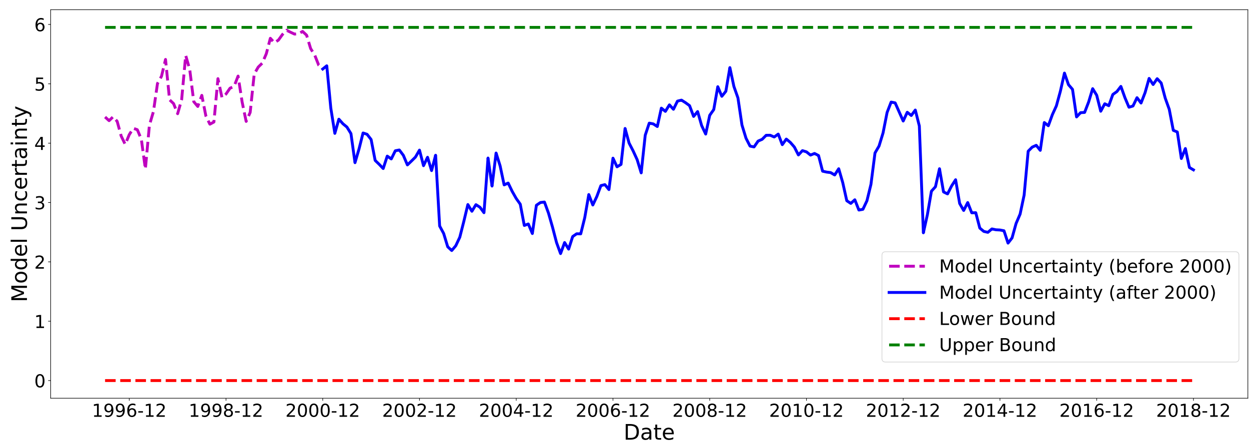
[27]The minimum is zero. Since there are $2^7 \times 3 = 384$ models, the largest entropy is approximately 5.95.

[28]European market includes stock markets in the following countries: Austria, Belgium, Switzerland, Germany, Denmark, Spain, Finland, France, UK, Greece, Ireland, Italy, Netherlands, Norway, Portugal, and Sweden.

[29]By saying 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 4: Model Uncertainty in European and Asian Pacific Markets

The figures plot the time-series of model uncertainty of 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 are only using 9 factors to calculate the posterior model probabilities. Details about used factors could be found in section II.7. The sample ranges from July 1993 to December 2018. 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 (5.95) of the model uncertainty.

market crash[30].

Overall, our model uncertainty measure captures the uncertainty in European and Asian Pacific stock markets, confirming our measure's usefulness and importance.

[30]Between June 2015 and January 2016, the SSE Composite Index of the Shanghai Stock Exchange decreased by more than 40% compared to its peak.

III Robustness Check

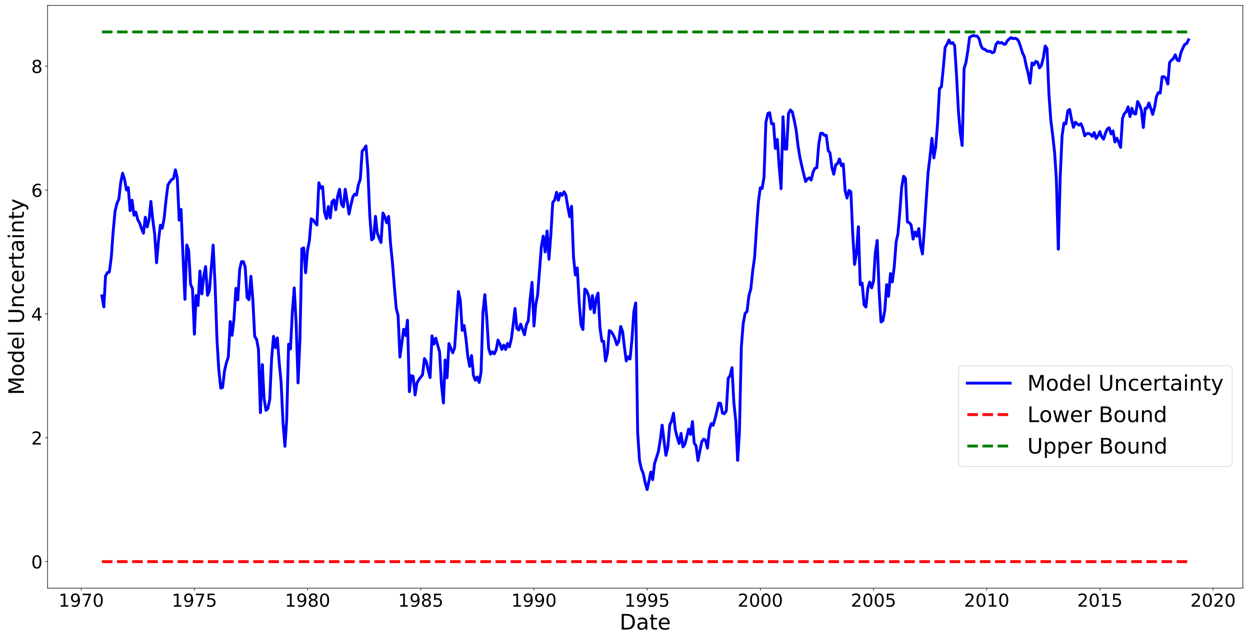
One may concern that our model uncertainty measure is sensitive to the choice of rolling window or the value of hyper-parameter in our priors. We address this issue in this section by investigating time-series variation in model uncertainty under different setups.

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 not large enough since it is well-known that factor expected returns and covariance matrix are difficult to pin down. On the other hand, a larger sample size is not always desirable. First, the main objective of this paper is to capture cyclical behaviors of model uncertainty. A long estimation period, such as 10 or 20 years, will average most of the valuable information contained in the cyclical behaviors factors. For example, we have shown that the average factor return in a 3-year rolling window exhibits a counter-cyclical pattern (see figure 3a). More importantly, it is not entirely reasonable to assume that the performance of factor strategies is stationary over a long period. For example, many research (e.g. McLean and Pontiff (2016)) show that factors' performance deteriorates post-publication.

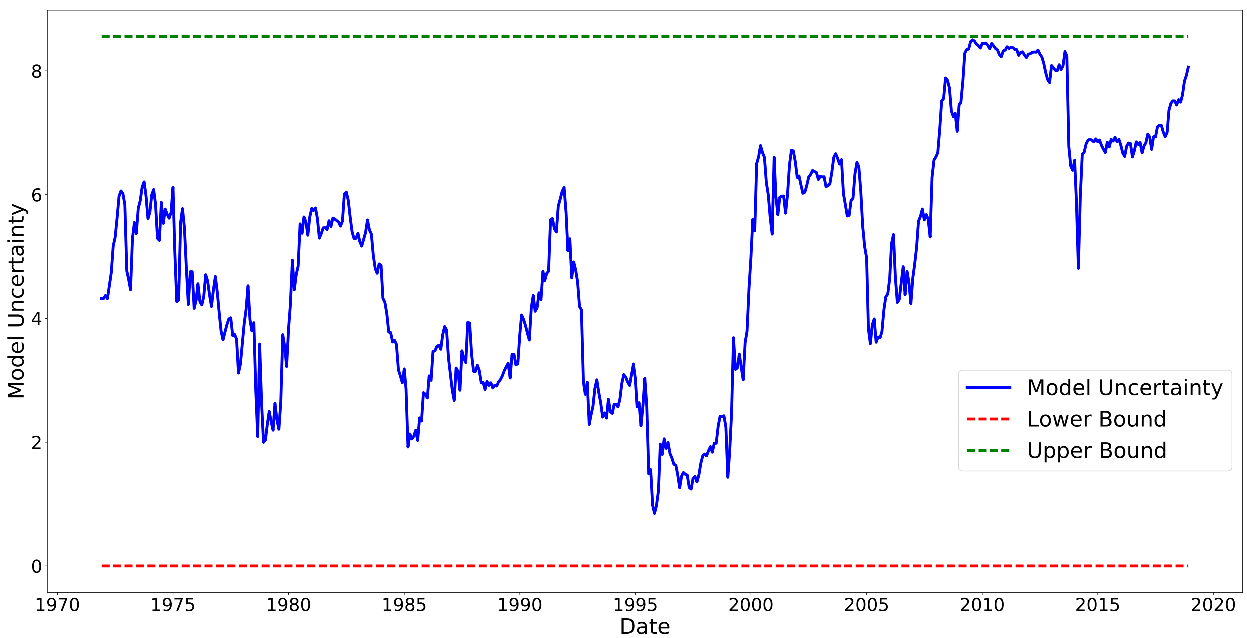
Figure 5a and 5b present the model uncertainty measures under 4-year and 5-year rolling windows, respectively. The patterns are similar to those found in a 3-year rolling window. There are two tiny differences: (a) the model uncertainty during the burst of the dot-com bubble is not as high as in the 2008 financial crisis; (b) the model uncertainty plummets after 2012 and remains at a low level that is far away from the upper bound for a few years before it increases again at the end of 2017. However, the model uncertainty under a 3-year rolling window starts to go up from the beginning of 2015.

Another 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 I shows, higher a acts as a larger shrinkage for Bayesian model average of $\boldsymbol{b}$. Conditional on a given model $\boldsymbol{\gamma}$ and g, the posterior mean of $\boldsymbol{b}_{\boldsymbol{\gamma}}$ is $\frac{g}{1+g}\boldsymbol{V}_{\boldsymbol{\gamma}}^{-1}\mathbb{E}_T[\boldsymbol{f}_{\boldsymbol{\gamma}\boldsymbol{t}}]$ and the prior average of $\frac{g}{1+g}$ equals $\frac{2}{a}$, so higher a is actually shrinking $\boldsymbol{b}$ towards zeros. However, when we assign different values to a, including 8, 12, 16, we find that the time-series patterns in model uncertainty are not sensitive to the choice of a.[31]

[31] Since the time-series plots are extremely similar to figure 1 and it is hard to find obvious difference, we do not provide the graphs here.



(a) Model Uncertainty in 4-Year Rolling Window



(b) Model Uncertainty in 5-Year Rolling Window

Figure 5: Alternative Rolling Windows

The figures plot the time-series of model uncertainty of linear stochastic discount factor (SDF) under alternative rolling windows. The construction of model uncertainty is the same as in figure 1 except that we are using 4-year and 5-year rolling windows.

IV Conclusions

We develop a new measure of model uncertainty in the cross-sectional asset pricing under the linear SDF specification. Roughly speaking, the measure is based on the entropy of Bayesian posterior probabilities for all possible factor models. The main observation is that the model uncertainty in the cross-section is countercyclical: it begins to climb up just before stock market crashes and remains at its peak during bear markets. Since we can calculate the lower and upper bound of entropy, we can easily discern when the model uncertainty is exceptionally high or low, while other uncertainty measures in the literature do not have this satisfactory property. We find that model uncertainty almost touches its upper bounds in the burst of the dot-com bubble and 2008 financial crisis.

If investors consider model uncertainty as another source of investment risk, their portfolio choice and expectations on the stock market should be naturally related to model uncertainty. Our second observation is that the model uncertainty can predict one-period ahead mutual fund flows, even after controlling past fund flows, VIX, and past performance of mutual funds. More precisely, investors seem to reduce their investment in "style" and "middle 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' sentiments. We discover that investors in the survey, no matter individual or institutional investors, are more pessimistic about the stock market when confronted with higher model uncertainty. We further find similar countercyclical behaviors of model uncertainty in European and Asian Pacific stock markets.

As model uncertainty in the cross-section is an important 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 of mean-variance portfolios, no such a general equilibrium model exists up to now, at least according to our knowledge. Future research could try to endogenize the model uncertainty in the general equilibrium model and explain its countercyclical behaviors.

References

ABRAMOWITZ, M., AND I. A. STEGUN (1965): *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, vol. 55. Courier Corporation.

ASNESS, C., AND A. FRAZZINI (2013): "The devil in HML's details," *The 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.

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

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

——— (2018b): "Real-time Portfolio Choice Implications of Asset Pricing Models," .

Bartlett, M. (1957): "A comment on D.V. Lindley's statistical paradox," *Biometrika*, 44(3-4), 533–533.

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

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

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., 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.

De Bondt, W. F., and R. Thaler (1985): "Does the stock market overreact?," *The Journal of finance*, 40(3), 793–805.

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

Epstein, L. G., and S. E. Zin (1991): "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis," *Journal of Political Economy*, 99(2), 263–286.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Stock Returns," *The Journal of Finance*, 47, 427–465.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *The Journal of Financial Economics*, 33, 3–56.

Fama, E. F., and K. R. French (2016): "Dissecting anomalies with a five-factor model," *The 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.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989): "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57(5), 1121–1152.

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

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.

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

James, W., and C. Stein (1961): "Estimation with Quadratic Loss," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pp. 361–379, Berkeley, Calif. 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. (1990): "Evidence of predictable behavior of security returns," *The Journal of finance*, 45(3), 881–898.

Jegadeesh, N., and S. Titman (1993): "Returns to buying winners and selling losers: Implications for stock market efficiency," *The 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.

Kozak, S., S. Nagel, and S. Santosh (2019): "Shrinking the cross-section," *Journal of Financial Economics*.

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): "Security Prices, Risk, and Maximal Gains from Diversification," *Journal of Finance*, 20, 587–615.

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

Ludvigson, S. C., S. Ma, and S. Ng (2015): "Uncertainty and business cycles: exogenous impulse or endogenous response?," Discussion paper, National Bureau of Economic Research.

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?," *The 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.

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.

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

Appendices

A Description of Factors

CAPM. CAPM (see Sharpe (1964) and Lintner (1965)) is the pioneer of linear factor model. The only factor in CAPM is the excess return on market portfolio, which is proxied by a value-weighted portfolio of all CRSP firms incorporated in US and listed on the NYSE, AMEX or NASDAQ. We use 1-month Treasury rate as a proxy for risk-free rate. The data comes from Ken French's website.

Long-term reversal. De Bondt and Thaler (1985) argues that investors in the stock market tend to overreact, so an investment strategy that buys stocks performing poorly in prior 13 to 60 months and sells those performing well in the past could be profitable. The long-term reversal factor is downloaded from Ken French's data library.

Short-term reversal. Jegadeesh (1990) constructs a reversal strategy of buying and short selling stocks based on their prior-month stock return performance and rebalancing portfolios monthly, and he finds that the abnormal return on this long-short portfolio is as high as 2.49% per month. We download the data from Ken French's data library.

Momentumn. In contrast to reversal strategies, Jegadeesh and Titman (1993) find that stocks that perform well or poorly in the past 3 to 12 months continue their performance in the next 3-month to 12-month holding periods, so investors can outperform the market by buying past winners and selling past losers. We download the momentumn factor from Ken French's data library.

Fama-French 5 factor model. Fama and French (1992) extend CAPM by introducing two additional factors, SMB and HML, where SMB is the return difference between portfolios of stocks with small and large market values, and HML is the return difference between portfolios of stocks with high and low book-to-market ratio. Fama and French (2016) further include a profitability factor (RMW) and one investment factor (CMA). Again, the data comes from Ken French's website.

q-factor model. Hou, Xue, and Zhang (2015) introduce a four-factor model that includes market excess return, a new size factor (ME), proxied by the return difference between large and small stocks, an investment factor (I/A), proxied by the return difference of stocks with high and low investment-to-asset ratio, and finally the profitability factor (ROE), created by sorting stocks based on their return-on-equity ratio.[32]

Quality-minus-junk. Asness, Frazzini, and Pedersen (2019) groups the listed companies into two types: quality stocks - those that are profitable, growing and well-managed, and

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

junk stocks - those that are unprofitable, stagnant or poorly-managed. They find that a quality-minus-junk (QMJ) strategy, in which investors go long the quality and short the junk stocks, generate a high positive abnormal returns. We download the QMJ factor from 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 premium 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 the BAB factor from AQR data library.

HML Devil. Asness and Frazzini (2013) propose alternative way, which relies on more timely market value information but maintain suitable use of lagged book value, in constructing the value factor. We download the HML Devil factor from AQR data library.

A.1 Additional Tables

Table 6: Summary of First-Order Autoregression

	(1) Entropy	(2) Financial	(3) Macro	(4) Real	(5) EPU_1	(6) EPU_2	(7) VIX
AR(1)	0.9808***	0.9802***	0.9887***	0.9699***	0.7998***	0.6737***	0.8313***
	(150.74)	(110.15)	(92.30)	(98.45)	(16.04)	(16.77)	(26.64)
N	588	588	588	588	413	413	395
R^2	0.9589	0.9603	0.9773	0.9424	0.6371	0.4475	0.6894

t statistics in parentheses: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

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, Ma, and Ng (2015) and Jurado, Ludvigson, and Ng (2015). EPU_1 and EPU_2 are two economic policy uncertainty sequences from Baker, Bloom, and Davis (2016). VIX 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 and N is the number of observations.

B Proofs

Lemma 1 *Suppose that the vector of test assets, $\boldsymbol{R}$, includes all candidate factors, $\boldsymbol{f}$. Let $\boldsymbol{V}_\gamma = \mathrm{Var}[\boldsymbol{f}_\gamma]$, $\boldsymbol{C}_\gamma = \mathrm{Cov}[\boldsymbol{R}, \boldsymbol{f}_\gamma]$, and $\boldsymbol{\Sigma}$ denotes the covariance matrix of $\boldsymbol{R}$. The following two equalities always hold:*

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_\gamma = \begin{pmatrix} \boldsymbol{I}_{p_\gamma} \\ \boldsymbol{0}_{(N-p_\gamma)} \end{pmatrix}, \tag{17}$$

$$\boldsymbol{C}_\gamma^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{C}_\gamma = \boldsymbol{V}_\gamma. \tag{18}$$

Proof. Without loss of generality, the vector $\boldsymbol{R}_t$, $t = 1, \ldots, T$, can be arranged as

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

where $\boldsymbol{r}_t^e$ is a vector of test assets excluding candidate factors $\boldsymbol{f}_t$. Under this specification,

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

where

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

Then

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

or equivalently

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

Thus

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

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

B.1 Proof of Proposition 1

Proof. As in section I.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

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

This completes the proof of Proposition 1. ■

B.2 Proof of Proposition 2 and 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

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

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

where $\overline{\boldsymbol{R}}^{\intercal}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\left(\boldsymbol{C}_{\boldsymbol{\gamma}}^{\top}\boldsymbol{\Sigma}^{-1}\boldsymbol{C}_{\boldsymbol{\gamma}}\right)\boldsymbol{C}_{\boldsymbol{\gamma}}^{\intercal}\boldsymbol{\Sigma}^{-1}\overline{\boldsymbol{R}} = \overline{\boldsymbol{f}}_{\boldsymbol{\gamma}}^{\intercal}\boldsymbol{V}_{\boldsymbol{\gamma}}^{-1}\overline{\boldsymbol{f}}_{\boldsymbol{\gamma}} = SR_{\boldsymbol{\gamma}}^2$, 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

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

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,

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

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:

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

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[33]. Therefore, the Bayes factor could be defined as

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

The prior for g is such that $\pi[g] = \frac{a-2}{2}(1+g)^{-\frac{a}{2}}$. We calculate the marginal likelihood of

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

data only conditional on model $\mathcal{M}_{\boldsymbol{\gamma}}$ by integrating out g in equation (22).

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

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 $\mathrm{SR}_{\boldsymbol{\gamma}}^2$ always decreasing in $p_{\boldsymbol{\gamma}}$, we use the original representation of Bayes Factor, that is,

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

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

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

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

This completes the proof of Proposition 3. ∎